Jelena Zdravkovic · Jānis Grabis
Selmin Nurcan · Janis Stirna (Eds.)

Perspectives in Business Informatics Research

LNBIP 330

17th International Conference, BIR 2018
Stockholm, Sweden, September 24–26, 2018
Proceedings

Springer

Lecture Notes in Business Information Processing 330

More information about this series at http://www.springer.com/series/7911

Jelena Zdravkovic · Jānis Grabis
Selmin Nurcan · Janis Stirna (Eds.)

Perspectives in Business Informatics Research

17th International Conference, BIR 2018
Stockholm, Sweden, September 24–26, 2018
Proceedings

Springer

Editors
Jelena Zdravkovic
Stockholm University
Kista
Sweden

Jānis Grabis
Riga Technical University
Riga
Latvia

Selmin Nurcan
Université de Paris 1 Panthéon - Sorbonne
Paris
France

Janis Stirna
Stockholm University
Kista
Sweden

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-319-99950-0 ISBN 978-3-319-99951-7 (eBook)
https://doi.org/10.1007/978-3-319-99951-7

Library of Congress Control Number: 2018952476

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Business informatics is an integrative academic discipline combining information and communication technology (ICT), informatics, and management. It is about empowering enterprises to, by utilizing advantages of existing ICT solutions, create and introduce new solutions corresponding to ever-changing business needs and conditions.

The Business Informatics Research (BIR) conference series was established 17 years ago as the result of the collaboration of several Swedish and German universities. The goal was to create a forum where researchers in business informatics, senior as well as junior, could meet and hold discussions. Over the years, the conference has expanded to many other countries, and some workshops as well as a doctoral consortium accompany the main conference. The conference series has a Steering Committee, to which one or two persons from each appointed organizer are invited.

The 17th BIR conference was held in Stockholm during September, 24–26, 2018, at the Department of Computer and Systems Sciences (DSV) of Stockholm University. The theme of the conference was "Business Resilience – Organizational and Information System Resilience in Congruence." Resilience refers to an entity's (e.g., society, organization, and information system) ability to continuously deliver the intended outcome despite adverse events. The management of resilience is often handled on one level only, without a holistic approach to the communication of problems. Moreover, both systems and organizations have often been designed without built-in mechanisms that can steer adjustments according to quickly changing application situations. To this end, our field should strengthen the focus on addressing the resilience of businesses and their information systems in an integrated manner, which would allow the concept of business resilience to be established.

This year the BIR conference attracted 50 submissions from 16 countries. They were rigorously reviewed by the Program Committee. As a result, 17 high-quality papers were selected for publication in this volume and for presentation at the conference. They cover different aspects of the conference's main topic as well as of business informatics research in general; as such they were distributed in six thematic sessions: Business, People, and System Resilience; Digital Transformation; Business and IS Development; Resilient Collaboration; Enterprise Systems and Data Analytics; and IT Management.

The conference program included two keynotes: by Prof. Oscar Pastor from the Universidad Politècnica de València (Spain) on supporting business resilience by conceptual modeling and MDD of information systems; and by Prof. Ulrich Frank from the University of Duisburg-Essen (Germany) on improvements for coping with dynamic business environments by adequate language engineering and enterprise modeling.

We would like to thank everyone who contributed to the BIR 2018 conference. We thank the authors for contributing and presenting their research, we appreciate the invaluable contribution of the members of the Program Committee and external reviewers, and we thank all members of the local organization team from Stockholm

University for ensuring the smooth processing of the conference. We acknowledge the EasyChair development team for providing such a convenient tool for preparing these proceedings and the Springer publishing team for their collaboration. Last but not the least, we thank the Steering Committee for their support and we hope that BIR 2018 was a valuable addition to the further development of the BIR conference series.

July 2018

Jelena Zdravkovic
Jãnis Grabis
Janis Stirna
Selmin Nurcan

Organization

Program Committee

Jan Aidemark	Linnaeus University, Sweden
Gundars Alksnis	Riga Technical University, Latvia
Said Assar	Institut Mines-Telecom, France
Eduard Babkin	Higher School of Economics Nizhny Novgorod, Russia
Per Backlund	University of Skövde, Sweden
Robert Andrei Buchmann	Babeş-Bolyai University of Cluj Napoca, Romania
Rimantas Butleris	Kaunas University of Technology, Lithuania
Albertas Caplinskas	Vilnius University, Lithuania
Witold Chmielarz	University of Warsaw, Poland
Sybren De Kinderen	University of Luxembourg, Luxembourg
Hans-Georg Fill	University of Bamberg, Germany
Peter Forbrig	University of Rostock, Denmark
Jānis Grabis	Riga Technical University, Latvia
Darek Haftor	Linnaeus University, Sweden
Amin Jalali	Stockholm University, Sweden
Christian Janiesch	University of Würzburg, Germany
Manfred Jeusfeld	University of Skövde, Sweden
Björn Johansson	Lund University, Sweden
Gustaf Juell-Skielse	Stockholm University, Sweden
Dimitris Karagiannis	University of Vienna, Austria
Marite Kirikova	Riga Technical University, Latvia
Manuele Kirsch-Pinheiro	Université Paris 1 Panthéon-Sorbonne, France
Nikolaos Korfiatis	University of East Anglia, UK
Janne Korhonen	Aalto University, Finland
Birger Lantow	University of Rostock, Denmark
Michael Le Duc	Mälardalen University, Sweden
Ginta Majore	Vidzeme University of Applied Sciences, Latvia
Raimundas Matulevicius	University of Tartu, Estonia
Laila Niedrite	University of Latvia, Latvia
Jyrki Nummenmaa	University of Tampere, Finland
Jacob Nørbjerg	Copenhagen Business School, Denmark
Geert Poels	Ghent University, Belgium
Henderik Proper	Public Research Centre Henri Tudor, Luxembourg
Vaclav Repa	University of Economics, Czech Republic
Irina Rychkova	Université Paris 1 Panthéon-Sorbonne, France
Kurt Sandkuhl	University of Rostock, Denmark
Rainer Schmidt	Munich University of Applied Sciences, Germany
Ulf Seigerroth	Jönköping University, Sweden

Manuel Serrano	ALARCOS Research Group, University of Castilla–La Mancha, Spain
Monique Snoeck	Katholieke Universiteit Leuven, Belgium
Janis Stirna	Stockholm University, Sweden
Gianluigi Viscusi	EPFL, CDM-CSI, Switzerland
Hans Weigand	Tilburg University, The Netherlands
Anna Wingkvist	Linnaeus University, Sweden
Stanislaw Wrycza	University of Gdansk, Poland
Jelena Zdravkovic	Stockholm University, Sweden
Milan Zdravković	University of Niš, Serbia

Steering Committee

Marite Kirikova (Chair)	Riga Technical University, Latvia
Eduard Babkin	State University - Higher School of Economics, Nizhny Novgorod, Russia
Rimantas Butleris	Kaunas Technical University, Lithuania
Sven Carlsson	Lund University, Sweden
Peter Forbrig	Rostock University, Germany
Bjorn Johansson	Lund University, Sweden
Andrzej Kobyliñski	Warsaw School of Economics, Poland
Charles Møller	Aalborg University, Denmark
Lina Nemuraite	Kaunas Technical University, Lithuania
Jyrki Nummenmaa	University of Tampere, Finland
Raimundas Matulevicius	Tartu University, Estonia
Vaclav Repa	Economic University Prague, Czech Republic
Kurt Sandkuhl (Chair)	Rostock University, Germany
Janis Stirna	Stockholm University, Sweden
Benkt Wangler	University of Skövde, Sweden
Stanislaw Wrycza	University of Gdansk, Poland
Joze Zupancic	University of Maribor, Slovenia

Additional Reviewers

Akram, Asif	Imgrund, Florian	Su, Yanhui
Bader, Nikolina	Kampars, Jānis	Sudzina, Frantisek
Boissier, Fabrice	Kunnummel, Vimal	Svatoš, Oleg
Bork, Dominik	Lehmann, Holger	Walch, Michael
Danenas, Paulius	Reitemeyer, Benedikt	Wanner, Jonas

Contents

Information Technology Management

Enterprise Systems and Data Analytics

Business and Information Systems Development

Business, People and System Resilience

External Cybersecurity Incident Reporting for Resilience

Annika Andreasson(✉) and Nicole Fallen

Cybercom Sweden, P.O. Box 7574, 103 93 Stockholm, Sweden
{annika.andreasson,nicole.fallen}@cybercom.com

Abstract. Reporting cybersecurity incidents to external authorities is a newer requirement mandated by several complex and multi-layered laws. It is non-trivial, however, to determine what constitutes a reportable incident, the reporting timeframe, report recipients, and which data to include in the report, as it varies by country, organizational size and industry sector.

This research aims to help organizations navigate the various external cybersecurity incident reporting (ECIR) requirements, both to help them avoid penalties and to assist international cybersecurity efforts. This research focuses on EU and Swedish legal acts, and addresses which EU and Swedish laws govern the external incident reporting requirements of organizations located in Sweden, including the details of reportable incidents, report contents, recipients and timeframes.

A survey research strategy based on document analysis was used to synthesize the regulatory landscape for ECIR. 16 laws were found governing ECIR within Sweden; nine at the EU level and seven at the Swedish level (plus three pending at the Swedish level). The answers to the research questions are presented along with a discussion of the complexity of the legislation and double-reporting. Further research avenues are suggested.

Keywords: Incident reporting · Data breach reporting · Regulatory compliance ECIR · Cybersecurity

1 Introduction

The increasing danger posed by hackers, cyber-criminal organizations, state-sponsored hacking organizations, and even enemy states, has led to information security and cybersecurity initiatives being prioritized at every level of government, all aimed at improving overall cyber resilience. Cyber resilience is an organization's ability to deliver intended outcomes even under adverse cyber circumstances where normal delivery mechanisms cannot be used, whether that organization be at the micro level, such as an individual organization's IT department, or the macro level, such as nation states or international alliances (Björck et al. 2015). But attaining cyber resilience requires organizations to gather and analyze high quality intelligence about both the cyber landscape and organizational circumstances. As a result, sweeping legislation has been introduced, attempting to combat both cybercrime and privacy infringements and

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 3–17, 2018.
https://doi.org/10.1007/978-3-319-99951-7_1

thereby attain cyber resilience, and nowhere has this been more evident than in the European Union (EU).

Since the turn of the new millennium, several legislative acts have been passed in the EU that deal with IT. These legislative acts aim to improve the cybersecurity, cyber resilience, and privacy rights of EU nations and their citizens and residents by setting guidelines for how organizations work with information and IT, respond to threats and incidents, learn from the mistakes they make, and cooperate with others to improve and to prevent future incidents, all in line with the EU's Cybersecurity Strategy (European Commission 2013). To gather the intelligence necessary to achieve resilience, these new EU laws are mandating high quality incident reports from the IT department level to be passed up to the national level and finally to the EU level, as overseen by the European Union Agency for Network and Information Security (ENISA) (European Commission and High Representative of the European Union for Foreign Affairs and Security Policy 2013). The laws also provide increased incentives to assist with this international cybersecurity effort, generally in the form of penalties and fines (Voss 2018).

The goal of this research is to help organizations navigate the various external cybersecurity incident reporting (ECIR) requirements, both to help organizations avoid the negative consequences of a failure to comply, and to assist international efforts in achieving a higher level of cyber resilience. To realize this goal, the research aims to answer the research question of which EU and Swedish laws govern the ECIR requirements of organizations located in Sweden, and what ECIR they mandate, by specifically determining: What are the in-scope European Union and Swedish laws that require ECIR, what constitutes a reportable incident under these different laws, and what data must be reported, to whom, and in what time frames?

A survey research strategy was chosen in order to provide a full census, and document analysis was chosen as the research method, as the law is meant to be the final authority on legally mandated requirements such as those for ECIR. The MaxQDA software tool was used to mark and code data, assist in qualitative data analysis, and maintain research logs. The iterative nature of the analysis was assisted by having all the relevant documents available to both researchers within the tool. This was especially important in gathering the relevant laws, as there was no full listing of ECIR-mandating laws for the research to draw upon. As laws referenced other laws, guidelines referenced laws, and even email discussions referenced laws, these resources were added to the tool and coded. This continued until the researchers determined that saturation had been reached, and no further references to laws or guidelines containing new information were found. Analysis was then undertaken, where the coded data were reviewed in depth and interpreted.

There were two main issues with this approach. First, the legal acts mandating ECIR are generally complex and rarely contain comprehensive guidelines for compliance. To mitigate this, data triangulation was used both through consultation with subject-matter experts and analyzing secondary (documentary) sources for further guidance. These secondary sources were all chosen based on recency, topicality and reputation/mandate. ENISA and Article 29 Working Party compliance guidelines were used for EU laws, while guidelines published by the relevant supervisory/regulatory

authorities were utilized for both Swedish and EU laws. Prior research was used to corroborate the researchers' understanding of legal texts.

Secondly, as the current cybersecurity legal framework continues to mature, some of the relevant legal acts were either becoming obsolete or not yet completed. The most recent version (as of 10 April 2018) of laws slated to go into effect during 2018 were included in the study, as were older laws that were deemed likely to remain in effect at least throughout 2018. The new Swedish Data Protection Act (SFS 2018:218 2018), which was approved on 24 April, 2018 was also included in the study.

This research focuses on the EU laws with the most relevance to cybersecurity and data privacy protection, their transpositions and successors in Swedish law, and Swedish laws that also have cybersecurity and data privacy as the goal of their ECIR. Left out were reporting outside the realm of cybersecurity; non-mandatory ECIR; and highly specialized areas of law, such as maritime and military. (For a graphical overview of the research scope see https://bit.ly/2Gmx6mN.) Although this research is limited to the EU and Swedish legal acts governing ECIR of organizations within their compliance scope, its findings regarding EU law will be useful to other organizations bound by those laws. Further, the findings within Sweden can serve as an illustrative example of how ECIR laws at the EU and national level interact.

While internal incident handling is a necessary component of ECIR, the researchers have taken the approach that any adequate incident handling routine will be sufficient to inform an ECIR report and so will not treat incident handling in this paper.

The remainder of the paper is structured as follows: Sects. 2 and 3 provide background on the evolution of cybersecurity law and of how ECIR fits within regulatory compliance management. Section 4 presents the results and analysis. Section 5 presents the discussion. Section 6 presents research limitations and suggestions for future research. Finally, Sect. 7 provides concluding remarks. A practical methodology for utilizing the extended results of this research, along with figures and other supplementary materials, is available at https://bit.ly/2Gmx6mN.

2 Evolution of Cybersecurity Law

2.1 The EU Perspective

In Europe, the idea that personal data is something in need of legal protection has evolved since the 1970s (Voss 2018). Voss (2018) splits up this legislative protection into three phases, beginning with Western European countries adopting national laws for protection of data, like Sweden's SFS 1973:289 Data Act of 1973. The second phase was the attempt by the European Union to harmonize Member State laws through Directive 95/46/EC (DP Directive) (1995). During this phase, the e-Privacy directive (2002) was also passed, strengthening privacy protections specifically in the context of public telecommunications providers (Porcedda 2018). In 2009, the e-Privacy Directive was amended to include ECIR, and an amendment to the Framework Directive, Directive 2009/140/EC, was also introduced, which mandated ECIR as well as strengthening other cybersecurity measures taken by telecommunications providers (Porcedda 2018). Finally, in 2013, a commission regulation, 611/2013 (CR611) was

passed in order to harmonize the ECIR mandated by the rest of the Telecommunications Package (Porcedda 2018).

We have now entered the third phase: trans-EU legislative harmonization (Voss 2018). The adoption of the eIDAS Regulation in 2014, in part to ensure improved cybersecurity for e-identification and e-trust services by harmonizing these laws across the EU, was likely a first, very early step in this phase (Porcedda 2018). The phase has now become firmly established, however, with the GDPR repealing the DP Directive, and obviating the need for the e-Privacy Directive (Voss 2018; Porcedda 2018). The NIS Directive, while not a regulation, has also increased the overall cybersecurity requirements on providers of network and information systems and critical infrastructure of all kinds. While the new e-Privacy Regulation was initially slated to enter into force the on same date as GDPR, it remains a proposal and is unlikely to be completed before 2019 (Splittgerber and Schonhofen 2017). When it does come into force, however, it will replace the e-Privacy Directive.

Porcedda (2018) specifically outlined the evolution of ECIR within the EU, though she divides it into data privacy laws, as outlined above, and security breaches. The latter she characterized as starting with the Telecom package – the collective name for three directives regarding telecommunications providers. Directive 2009/140/EC mandates telecommunications companies to notify supervisory authorities in case of an interruption of their services with "significant impact" (Directive 2009/140/EC 2009, Art 13 (a)(3)). With the introduction of GDPR, NIS and eIDAS, the scope of mandatory ECIR has widened well beyond the telecommunications industry.

2.2 The Swedish Perspective

Within Sweden, there are and have been many laws that touch on privacy and cybersecurity. For 28 years, the Data Act of 1973 (SFS 1973:289 1973) was the core of Swedish data protection. It regulated the creation and maintenance of personal data registries (Datainspektionen 2018). Because of the EU DP Directive, the Swedish Data Act was replaced in 1998 by the Swedish Data Protection Act (PuL) (SFS 1998:204 1998), which had a broader application to all personal data, not just registries, and became enforceable in 2001 (Datainspektionen 2018). The e-Privacy and Framework Directives, along with other related laws, were transposed into the Electronic Communications Act (LEK) (SFS 2003:389 2003), plus the Electronic Communications Ordinance (FEK) (SFS 2003:396 2003) and the attendant authority regulations specifying further details of the reporting (PTSFS 2012:1 (now obsolete) and 2) (PTSFS 2012:1 2012; PTSFS 2012:2 2012; PTSFS 2014:2 2014).

On 25 May 2018, GDPR will become enforceable across the EU, replacing PuL in Sweden. Because of the aforementioned delays in the e-Privacy Regulation, the e-Privacy Directive remains in force, as does LEK. The NIS Directive mandates transposition into Member States' laws by 9 May 2018, and identification of "operators of essential services" by 9 November 2018 (Directive 2016/1148/EU 2016, Art. 25(1), Art. 5(1)), but the Swedish transposition has been proposed to enter into force 1 August 2018 (Regeringskansliet 2018).

3 Regulatory Compliance Management and Compliance Support

The field of regulatory compliance management (RCM) "deals with ensuring that a given enterprise is in accordance with a set of regulatory guidelines" (El Kharbili 2012, p. 23). These guidelines can be external requirements such as legislative acts and third-party contractual requirements, or internal policies and guidelines (El Kharbili 2012). RCM seeks to manage all requirements placed on an organization, especially from a Business Process Management (BPM) perspective. This includes implementing internal processes in a way that complies with all necessary rules as well as creating proof of this compliance suitable for audit.

Most research within RCM does not deal with ECIR, as ECIR is relatively niched and separate from most business processes. So, while RCM compliance support systems could certainly support ECIR, for most organizations that will need to comply with ECIR regulations (like SMEs), RCM is unnecessarily complex and would be out of reach.

One study within RCM was found that treated ECIR. In his document analysis, Voss (2018) examined compliance mechanisms as mandated by the GDPR, as well as placing them in historical perspective. He treated all aspects of mandated compliance, though his discussion of Data Breach Notification (DBN) was most salient to the current research. Here, Voss pointed out that the GDPR is the first legislative act within the EU that is applicable to all persons dealing with information relating to EU subjects, creating a new DBN requirement for many organizations (Voss 2018). Voss outlined three categories of DBN specified in the GDPR:

- Controllers notifying the Supervisory Authority (SA), which must occur "without undue delay" (quoted from GDPR within (Voss 2018, p. 20)), within 72 h of finding out about the breach, if feasible.
- Controllers notifying affected data subjects (DS).
- Processors notifying controllers, so that controllers can comply with their obligations.

The latter two categories must also occur "without undue delay" (quoted from GDPR within (Voss 2018, p. 20)). He also outlined what the GDPR specifies must be included in the notification.

4 Results and Analysis

Over 100 documents were analyzed for this research, including both primary sources (legal acts) and secondary sources (research reports, standards, guidelines and other publications from both governmental and non-governmental sources). (Please see https://bit.ly/2Gmx6mN for additional information.) Legal acts were analyzed for the definitive answers to the research questions while secondary sources were analyzed to fill in the details.

4.1 The Relevant EU and Swedish Laws Requiring ECIR

The legislative framework applicable to organizations within Sweden is a mix of EU regulations, directives that have been transposed into Swedish law in various ways, and different levels of Swedish law. EU regulations and native Swedish laws are direct in their application, while the transposition of directives can be somewhat more complex. For example, even when only considering ECIR requirements, the Swedish Electronic Communications Act (LEK) transposes two different amended EU directives (SFS 2003:389 2003; Directive 2009/136/EC 2009; Directive 2009/140/EC 2009) and is additionally complemented by both a Swedish authority regulation (PTSFS 2012:2 2012) and an EU Commission Regulation (Commission Regulation 611/2013 2013) which previously replaced a second Swedish authority regulation (PTSFS 2012:1 2012).

Because many of the laws in the Mandatory ECIR legal framework are relatively new, the ancillary Swedish laws required either to implement them (in the case of EU directives) or complement them are not yet completed. The authors have therefore chosen to analyze both the original EU directives and their transpositions, even if the transpositions were only at the proposal stage. These relationships are indicated in Table 1 below, which presents all in-scope EU and Swedish laws mandating ECIR and their compliance scope (which legal entities the law is aimed at). Included in this list are one law that only mandates annual aggregated incident reporting (the MSB Steering document) and one law that only assists in defining what constitutes a reportable incident for the NIS Directive (Commission Implementing Regulation 2018/151 2018). Please note that scoping laws are sometimes highly complex (as in GDPR, for example), and the following table provides only simplified summaries.

4.2 Reportable Incidents

All the applicable laws prioritize either information security incidents/breaches, personal data breaches, or both. Some are more concerned with service availability (which falls under the rubric of information security more generally). While all personal data breaches are information security incidents (as are losses of information service availability), the opposite is not true (Article 29 Data Protection Working Party 2018). For example, a denial of service attack to a public agency's public-facing informational website would constitute an information security incident, as it denies availability of the service, but it is unlikely that personal data would have been posted on the site, and so it would likely not be a personal data breach. If, however, a hospital's internal network were to be hacked, rendering patient records unavailable for a time, this would constitute both a personal data breach and an information security incident in the more general sense. It would also specifically be a NIS incident at an essential service provider.

What follows is a very brief overview of the ECIR thresholds within scope. The full threshold analysis for each individual law is available at https://bit.ly/2Gmx6mN.

Akin to the division made in Porcedda (2018) between cybersecurity and data protection laws, ECIR laws can be divided into two general categories: personal data breach reporting and cybersecurity incident reporting. Fewer than half of the laws are concerned with personal data breaches, those being GDPR, CA Directive, and e-Privacy

Table 1. EU and Swedish Laws Requiring ECIR

Level	Legal act	Compliance scope
EU	Regulation EU 2016/679 General Data Protection Regulation (GDPR)	Any natural or legal person with an establishment in the EU who processes personal data (of living persons) at least partly by automated means, or data which are a part of – or intended to be part of – a filing system, whether or not the actual processing takes place in the EU. Controllers/processors with no establishment in the EU that process personal data of a living data subject in the EU are also within scope of the law if they offer goods or services to, or monitor the behavior (that takes place within the EU) of, EU data subjects. Exemptions: competent authorities carrying out policing duties (covered by Directive 2016/680/EU), Member States when carrying out activities of international cooperation (Chapter 2 of Title V of the TEU), Union institutions, bodies, offices and agencies, natural persons carrying out purely personal or household activities, and data processed during activities outside the scope of EU law (GDPR Art. 2) Within Sweden, SFS 2018:218 (the Swedish Data Protection Act) states that personal data breaches reported under the Swedish Protective Security Ordinance or any related authority regulations shall not be reported under GDPR (SFS 2018:218 2018, 1 Kap. 4 §).
	Regulation EU 910/2014 (eIDAS)	Trust service providers (TSPs) established or notified in the union, including notified e-identification providers
	Commission regulation (EU) 611/2013 e-privacy notification measures (CR611)	Providers of publicly available electronic communications services
	Commission implementing regulation (EU) 2018/151 further specifications for NIS	Digital service providers

(*continued*)

Table 1. (*continued*)

Level	Legal act	Compliance scope
	Directive (EU) 2016/1148 (NIS directive) (Swedish transposition at the proposal stage)	Operators of essential services (OESs) and digital services providers (DSPs). Excepted are those covered by eIDAS, the e-Privacy Directive, or other "sector-specific Union legal act[s]" that mandate ECIR at least to the level of the NIS Directive (Directive 2016/1148/EU 2016, Art. 1(3) and (7)) and digital service providers that are micro- or small enterprises as defined in Commission Recommendation 2004/36/EC
	Directive (EU) 2015/2366 (PSD2) (Swedish transposition at the proposal stage)	Legal entities providing payment services within the EU. Does not apply to commercial agents authorized via an agreement to negotiate/conclude the sale or purchase of goods/services on behalf of only the payer or only the payee; professional bank note transportation companies working in that capacity; non-professional cash collection for non-profit/charitable activities; and "cash-to-cash currency exchange operations where the funds are not held on a payment account" (Art. 2(1), 3(f))
	Directive EU 2016/680 data protection for competent authorities (CA directive) (The Swedish *Brottsdatalagen* is in process)	Applies to "the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security" (Art. 1(1)). This data must by processed at least partially by automatic means or be part of (or intended as part of) a filing system
	Directive EC 2002/58 and EC 2009/136 directive on privacy and electronic communications (e-privacy directive and 2009 amendments) (transposed into SFS 2003:389)	Publicly available electronic communications services in public communications networks in the Community, including public communications networks supporting data collection and identification devices

(*continued*)

Table 1. (*continued*)

Level	Legal act	Compliance scope
	Directive EC 2002/21 and EC 2009/140 Framework directive and 2009 amendments (Transposed into SFS 2003:389)	"Electronic communications services, electronic communications networks, associated facilities and associated services, and certain aspects of terminal equipment to facilitate access for disabled users" (Directive 2002/21/EC 2002)
Sweden	SFS 1996:633 (Protective security ordinance)	State, county and municipal agencies; corporations, private partnerships, associations and foundations over which the state, counties or municipalities exercise a legal authority and private companies if their operations are of importance for national security or specifically must be protected against terrorism
	SFS 2003:389 (LEK) (Electronic communications act) (implements both e-privacy and framework directives)	Electronic communications networks and services with their attendant installations and services as well as other radio use
	SFS 2003:396 (FEK) (Electronic communications ordinance) (complements SFS 2003:389)	Electronic communications networks and services with their attendant installations and services as well as other radio use
	PTSFS 2012:2	Those who run a public communications network or a publicly accessible electronic communications service
	SFS 2015:1052 (ordinance regarding crisis preparedness)	State authorities and agencies
	SFS 2008:1002 (instruction to the civil contingencies agency (MSB steering document))	Civil contingencies agency (MSB)
	MSBFS 2016:2 MSB's regulations on government authorities' reporting of IT incidents (complements SFS 2015:1052)	State authorities and agencies
(in progress)	Proposition 2017/18:205 information security for operators of essential services and digital service providers (NIS transposition)	OESs and DSPs, with likely carve-outs for those covered by the Protective Security Ordinance or sector-specific laws with requirements at least as strict as those for this law (8–9 §)

(*continued*)

Table 1. (*continued*)

Level	Legal act	Compliance scope
(in progress)	Proposition 2017/18:77 new rules for payment services (PSD2 transposition)	Payment service providers who provide payment services in Sweden that are carried out within the European economic community
(in progress)	Proposition 2017/18:89 (Protective security act) (to replace SFS 1996:627, the law complementing SFS 1996:633 above)	Those who have operations of significance to Swedish security or are covered by an international security initiative to which Sweden has been bound

(including part of LEK and CR611). The remainder are concerned with cybersecurity incidents of various types. eIDAS is the one interesting exception. While Porcedda (2018) listed it within the cybersecurity category, this law specifically addresses both data protection and cybersecurity breaches.

Nearly all of the relevant laws have exceedingly vague thresholds for reportable incidents, meaning how severe an incident must be before it is reported. Most refer to "significant" or "substantial" risks or disruptions or effects, depending on the focus of the law. "Non-negligible" and "major" are also seen in the Swedish laws. In fact, the only laws that specify anything concrete are the implementation-type regulations: CR611, CIR151, PTSFS 2012:2 and MSBFS 2016:2. The first three of these specify thresholds based on the magnitude of the incident while the last only specifies categories of incidents that must be reported.

For the laws that specify notification of affected individuals, the threshold is always higher than that for notification of relevant supervisory authorities and based on the risk of adverse effects. This means that technical and organizational measures that mitigate adverse consequences can help an organization lessen the burden of needing to notify individuals. All of the data protection-focused laws make specific mention of this. Of course, all of the laws require that technical and organizational cybersecurity measures are used, and successful use of these, based on a risk analysis, is likely to lessen the burden of reporting in all cases, as these measures reduce the risk (and lessen the severity) of cybersecurity incidents.

Only the GDPR, the CA Directive and the e-Privacy Directive, which all focus on personal data breaches, include any ECIR with no threshold (meaning all breaches must be reported to the relevant recipient). In addition, for all of the laws, the supervisory authority is always authorized to require reporting of an incident based on its own judgment.

4.3 Report Recipients, Timeframes and Contents

Overall, there was little variation in the report data and timeframes for the different laws, while the recipients varied more. What follows here is a summary overview. For full details, please see https://bit.ly/2Gmx6mN.

Recipients. All of the ECIR laws require reporting to at least one supervisory-level authority, either at the national or EU level, though the SA varies based on the law. All laws in the Swedish Telecom Framework require reporting to the Swedish Post and Telecom Authority (PTS), GDPR requires reporting to the Swedish Data Protection Authority (*Datainspektionen* or *Integritets-skyddsmyndigheten* (DI/ISM)), NIS reports go to the Swedish Civil Contingencies Agency (MSB)/CERT.se, and the (non-military-related) Protective Security Ordinance and Act reports go to the Security Police. For most of these supervisory/competent authorities, incident reports should be filed electronically, the Security Police being the exception. The applicable report recipients for which the information could be obtained are available in the online materials.

Other recipients include affected users, the European Banking Authority, other Member States, the public, ENISA, the Commission and the NIS Cooperation Group. Overall, the ECIR reporting ecosystem is quite complex, involving very many actors.

Timing. All summary reporting done to EU level authorities is done annually. For most other reporting, the vast majority of the laws list some variation of "without undue delay" as their timeframe. Notification of affected individuals is always done without undue delay, or as soon as possible. The GDPR and CA Directive, eIDAS, and the Swedish Telecom Framework, however, all provide actual timeframes for reporting to the SA, ranging from 24 to 72 h for the preliminary report and generally allowing some sort of phased reporting. The guidelines for PSD2 submitted by the EBA require much faster reporting, with initial reports due within 4 h, and Sweden has signed on to follow these guidelines (European Banking Authority (EBA) 2017a).

Report Details. The data to be reported, when specified, always includes:

- Contact information for the reporting organization's contact person.
- Information about the breach, including the type, time of occurrence, whether it is ongoing, time of becoming aware, root causes, and location, and a narrative-type description of the breach event.
- Information about how the breach affects the organization and its systems, assets, users, individuals, and any other relevant parties, including cross-border information. Possible damage, magnitude and scope.
- What mitigation measures have been taken and are planned.

There are minor variations based on the focus of the law (continuity vs. personal data breach vs. financial, for example). Templates/portals are provided by the relevant supervisory authorities for all but the newest laws (2016 and later).

5 Discussion

The analysis has shown that there are several different ways in which these laws interact regarding ECIR requirements.

- Directives must be transposed into local laws which specify details that are vague in the directives. Later, regulations may specify these details, sometimes overriding previous local laws.

The main example of this is the Swedish LEK transposition of the telecom directives (e-Privacy and Framework). (Please see the supplementary materials for a graphical overview of this Swedish Telecom Framework.) Here, the ECIR was mandated in 2009 and the laws specifying how to report (PTSFS 2012:1 and 2012:2) were published in 2012. In 2013, CR611 came out specifying the rules for ECIR under the amended e-Privacy Directive, and in 2014, PTSFS 2012:1 was repealed in response. Keeping track of requirements in the constantly shifting field of international laws is daunting.

- Different laws may overlap in their compliance scope.

Lawmakers are quite sensitive to this and take pains to avoid it when possible. For example, carve-outs in NIS allow those covered by eIDAS's reporting requirements (among others) to not be covered under NIS's reporting requirements (Directive 2016/1148/EU 2016, Art. 1(3)). But for organizations forced to report under one law for data protection breaches and another for cybersecurity breaches (as is the case for all DSPs and OESs, for example), the fact that these are two different types of breaches may not matter. The complexity of reporting under two different laws, to (at least) two different agencies, remains.

- The cybersecurity vs. data protection focus of the laws can mean that different laws can apply to the same incident if there are both personal data and other cybersecurity aspects to the breach, and that different agencies may be recipients of reports from the same incident. These are especially likely when an organization offers several types of services.

Many modern organizations offer a wide variety of products and services to their customers, and in the cyber-sphere, these different offerings can end up within the compliance scopes of different laws. For example, a hospital will be required to report to DI/ISM under GDPR for personal data breaches, to MSB under both the NIS Directive (for information security breaches) and under the Swedish Ordinance regarding crisis preparedness (for breaches affecting civil security), and to the Swedish security police under the Swedish Protective Security Ordinance (for any breaches that might affect national security). A single breach, such as a social engineering attack, can lead to compromised passcodes and unauthorized access to systems. The attacker might steal patient records, including data from high-ranking government officials, modify a donor list, then infect the system with ransomware. In such a disastrous scenario, the same incident would require notifying all of the above authorities.

While the NIS Directive and the Swedish proposal for a transposition make it clear that lawmakers are attempting to alleviate the need for double reporting by carving out exceptions to information security breach reporting based on the extent of regulation within the organization's sector (including those who report under eIDAS and the telecom framework) (Directive 2016/1148/EU 2016, Art. 1(3) and (7); Prop. 2017/18:205 2017, 8–9 §, pp. 27–28), this does not alleviate the problem with

GDPR. The new e-Privacy Regulation might, however, when it replaces the Directive within the next year or two, as that regulation (in its current proposal form) does not mandate ECIR. This would leave GDPR, the CA Directive, and eIDAS as the only EU laws governing data protection breach reporting; and would imply a subsequent repeal or serious overhaul of the Swedish LEK. With the Swedish NIS Directive carving out exceptions to specific Swedish ECIR requirements (like the Protective Security Ordinance), it is possible that the overlap will be limited to organizations that offer services in different sectors. Until the legal framework is completed, however, this is only speculation.

6 Limitations and Future Research

The foremost limitation for this study was the still-changing legal framework. In addition to those propositions listed in the results, several pending laws have the possibility to affect the ECIR framework within the EU and Sweden, including the upcoming e-Privacy Regulation, a new European Electronic Communications Code Directive, and Swedish transpositions of, and complementary authority regulations for, both NIS and the CA Directive. Another limitation, however, was the amount of time required to do the research, despite limiting the scope to just Sweden. If similar ECIR law censuses were to be undertaken in other EU countries, this might eventually allow a comparison of ECIR laws across the EU.

The legal framework will continue to evolve as technology advances. Future research might look into whether the legal overlap/double reporting issues have been resolved as a result of maturing legislation. It would also be instructive to compare the threshold levels for data protection breach laws and information security breach laws to determine whether a single threshold calculation for privacy laws and another for cybersecurity incident laws would make sense. Developing a standard incident reporting template, or one for each of the two types of incident, would also be helpful as it would simplify reporting for organizations while allowing for easier aggregation at the EU level and beyond. Finally, the design and testing of a tool to allow organizations to more efficiently route reports to the correct external recipients and maintain traceability would also be worthwhile, especially one that could link to other compliance and/or incident handling tools.

7 Conclusions

This study undertook a census of all laws applicable in Sweden that mandate ECIR, in order to determine which EU and Swedish laws govern the ECIR requirements of organizations located in Sweden, and what ECIR they mandate. The scope was narrowed to exclude mandatory reporting that did not have cybersecurity as its purpose, highly specialized law like maritime and military law, and non-mandatory ECIR. Nine EU and seven Swedish laws were found, with some of the Swedish laws overlapping the EU laws. The thresholds for reporting, report recipients, report contents and timeframes for reporting were analyzed.

The authors hope that the current research will both help organizations comply with current ECIR requirements and illuminate the need for simplification in this area. Hopefully, future work will lead to legal and technical support for compliance so that the necessary reports are filed and can be analyzed to help achieve cyber resilience.

References

Article 29 Data Protection Working Party: Guidelines on personal data breach notification under regulation 2016/279 (2018)

Björck, F., Henkel, M., Stirna, J., Zdravkovic, J.: Cyber resilience - fundamentals for a definition. In: Rocha, A., Correia, A., Costanzo, S., Reis, L. (eds.) New Contributions in Information Systems and Technologies. Advances in Intelligent Systems and Computing, vol. 353, pp. 311–316. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16486-1_31

Commission Implementing Regulation (EU) 2018/151 of 30 January 2018. Official Journal of the European Union, L26, 48–51 (2018)

Commission Regulation (EU) No 611/2013 of 24 June 2013. Official Journal of the European Union, L173, 2–8 (2013)

Datainspektionen: Datainspektionen 1973–2018. Retrieved from Datainspektionen: https://www.datainspektionen.se/om-oss/historik/ (2018)

Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 (NIS Directive). Official Journal of the European Union, L194, 1–30 (2016)

Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002. Official Journal of the European Union, L108, 33–50 (2002)

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002. Official Journal of the European Union, L201, 37–47 (2002)

Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC. Official Journal of the European Union, L337, 11–36 (2009)

Directive 2009/140/EC of the European Parliament and of the Council of 25 November 2009 amending Directives 2002/21/EC. Official Journal of the European Union, L337, 37–69 (2009)

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. Official Journal of the European Union, L281, 31–59 (1995)

El Kharbili, M.: Business process regulatory compliance management solution frameworks: a comparative evaluation. In: Proceedings of the Eighth Asia-Pacific Conference on Conceptual Modelling (APCCM 2012), vol. 130, pp. 23–32, Melbourne (2012)

European Banking Authority (EBA): Guidelines compliance table (EBA/GL/2017/10). European Banking Authority (2017a)

European Banking Authority (EBA): Guidelines on major incident reporting under Directive (EU) 2015/2366 (PSD2) (2017b, July 27)

European Commission and High Representative of the European Union for Foreign Affairs and Security Policy: European cyber security strategy: an open, safe and secure cyberspace (2013)

European Commission: Joint communication to the European parliament, the council, the European economic and social committee and the committee of the regions: cybersecurity strategy of the European Union. European Commission, Brussels (2013)

Lagrådsremiss: Brottsdatalag. Retrieved from http://www.regeringen.se/492bb7/contentassets/be861421338643ccb72f10076581e9d8/brottsdatalag.pdf (2018)

Porcedda, M.G.: Regulation of data breaches in the European Union: private companies in the driver's seat of cybersecurity? In: Bures, O., Carapaccio, H. (eds.) Security Privatization: How Non-Security-Related Private Businesses Shape Security Governance, pp. 275–299. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-63010-6_12

Proposition 2017/18:105 Ny dataskyddslag. Stockholm (2018)
Proposition 2017/18:205 Informationssäkerhet för samhällsviktiga och digitala tjänster. Stockholm (2018)
Proposition 2017/18:77 Nya regler om betaltjänster. Stockholm (2018)
Proposition 2017/18:89 Ett modernt och stärkt skydd för Sveriges säkerhet - ny säkerhetsskyddslag. Stockholm (2018)
PTSFS 2012:1 Post- och telestyrelsens föreskrifter och allmänna råd om underrättelse om integritetsincidenter samt innehållet i förteckning över integritetsincidenter (2012)
PTSFS 2012:2 Post- och telestyrelsens föreskrifter och allmänna råd om rapportering av störningar eller avbrott av betydande omfattning. Post- och telestyrelsen (2012)
PTSFS 2014:2 Post- och telestyrelsens föreskrifter om upphävande av föreskrifter (PTSFS 2012:1) och allmänna råd om underrättelse om integritetsincidenter (2014)
Regeringskansliet. Stärkt skydd av informationssystem. Retrieved from Regeringskansliet: http://www.regeringen.se/pressmeddelanden/2018/02/starkt-skydd-av-informationssystem/, 15 February 2018
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88 (2016)
Regulation (EU) No 1093/2010 of the European Parliament and of the Council of 24 November 2010. Official Journal of the European Union, 24 November 2010
Regulation (EU) No 526/2013 of the European Parliament and of the Council of 21 May 2013. Official Journal of the European Union, L165, 41–58 (2013)
Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 (eIDAS). Official Journal of the European Union, L257, 73–114 (2014)
SFS 1973:289 Datalag. Svensk Författningssamling. Justitiedepartementet, Stockholm (1973)
SFS 1996:627 Säkerhetsskyddslag. Svensk Författningssamling. Justitiedepartementet, Stockholm (1996)
SFS 1996:633 Säkerhetsskyddsförordning. Svensk Författningssamling. Justitiedepartementet, Stockholm (1996)
SFS 1998:204 Personuppgiftslag. Svensk Författningssamling. Justitiedepartementet, Stockholm (1998)
SFS 2003:389 Lag om elektronisk kommunikation. Svensk Författningssamling. Näringsdepartementet, Stockholm (2003)
SFS 2003:396 Förordning om elektronisk kommunikation. Svensk Författningssamling. Näringsdepartementet, Stockholm (2003)
SFS 2008:1002 Förordning med instruktion för Myndigheten för samhällsskydd och beredskap. Svensk Författningssamling. Justitiedepartementet, Stockholm (2008)
SFS 2015:1052 Förordning om krisberedskap och bevakningsansvariga myndigheters åtgärder vid höjd beredskap. Svensk Författningssamling. Justitiedepartementet, Stockholm (2015)
SFS 2018:218 Lag med kompletterande bestämmelser till EU:s dataskyddsförordning, 24 April 2018
SOSFS 2008:1 Föreskrifter om användning av medicintekniska produkter i hälso- och sjukvården. Socialstyrelsens författningssamling. Socialstyrelsen, Stockholm (2008)
Splittgerber, A., Schonhofen, S.: Pre-Christmas Update on the ePrivacy Regulation. Retrieved from Technology Law Dispatch: https://www.technologylawdispatch.com/2017/12/privacy-data-protection/pre-christmas-update-on-the-eprivacy-regulation/, 13 December 2017
Voss, W.G.: Internal compliance mechanisms for firms in the EU General Data Protection Regulation. Revue juridique Thémis de l'Université de Montréal **50**(3), 783–820 (2018)

Conceptual Representation of the GDPR: Model and Application Directions

Jake Tom(✉), Eduard Sing, and Raimundas Matulevičius

University of Tartu, 50409 Tartu, Estonia
{jaketom,ecbyu,rma}@ut.ee

Abstract. With the GDPR coming into force in 2018, organizations require techniques to assess and improve the current state of their data processing activities with regards to personal data. Although current research has explored GDPR compliance to some degree, the work is either highly generalized or focused on specific principles within the legislation. This paper presents the current state of a model of the GDPR that provides a concise visual overview of the associations between entities defined in the legislation and their constraints. We also discuss its application directions towards the development of an overall approach to organizational privacy management - as a visual representation and as a tool to aid the definition of privacy policy.

Keywords: GDPR · Privacy management · Regulation compliance

1 Introduction

The rapid advancement of digital technologies has enabled the development of online innovations that have changed the way we conduct our daily lives and communicate with each other. However, such rapid innovations come with certain drawbacks. One of which is that average users are not often in control of how their personal data may be harvested, stored and processed by the providers of these services. Personal data processing has become more important in consumer's minds as privacy-invasive technologies become increasingly sophisticated. In April 2018, it came to public knowledge that Facebook was aggregating information of users that had not been explicitly given on its platform and in some cases, had not even signed up on Facebook [12]. This was made possible when users started sharing their address books to Facebook which inadvertently allowed the company to process information of those contacts without their consent. To emphasize the breadth of this issue, this was discovered during the Cambridge Analytica scandal [13] when the company was coming under fire for allowing third party application developers on its platform to access user information without any controls in place to mitigate the abuse of this data.

Events such as this have prompted consumers to be more wary of divulging their sensitive information to corporate entities. The TRUSTe consumer privacy

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 18–28, 2018.
https://doi.org/10.1007/978-3-319-99951-7_2

index of 2016 [8] has highlighted the fact that 69% of US internet users do not understand how companies share their personal information and that 37% say their top concern is how companies collect and share personal information with each other. 23% wished for procedures to be in place regarding the deletion of collected information. However, due to weak privacy protection regulations in many countries [14], companies have little incentive to establish privacy guarantees for their users.

A powerful step in the direction of personal data protection is the General Data Protection Regulation (GDPR) [6] that has come into effect from May, 2018. The regulation guarantees specific privacy rights to consumers (data subjects) by levying hefty fines on corporations that violate them. This includes the principle of *consent* which ensures that the data subject has been informed of the intentions behind the collection of personal data in a manner clearly distinguishable from other matters. The regulation also introduces the *right to be forgotten* which means that users can request a company to delete all information collected about them. Additionally, companies will now be required to implement appropriate technical measures to ensure that their systems are adequately protected from a privacy standpoint and must inform affected users if their systems have been breached.

Companies that wish to comply with the GDPR will need techniques to assess the state of their existing data processing activities and compare it against the regulation. As is often the case in legislative text, the GDPR provides generic rules and principles that might be difficult to interpret and then implement. In this work, we present a model of the GDPR that captures the explicit declarations within it. Such a representation can aid the development of organizational privacy policy and techniques to assess its compliance to the GDPR. Secondarily, it provides a concise visual overview of the regulation that can be used to understand its important aspects by anyone who understands the UML notation. This paper also discusses the steps required for its completion and its application directions.

The paper is structured as follows. Section 2 presents related work on the subject. Section 3 introduces the GDPR model. Section 4 talks about how the model can be applied to aid the creation of organizational privacy policy through the extraction of compliance rules. Section 5 concludes the paper.

2 Related Work

Apart from the GDPR, some other governments have already passed similar data protection laws such as the 1998 UK data protection act [9], the 1988 Privacy Act of Australia [10] and the 2013 Personal Data Protection Act in Malaysia [11]. DLA Piper [14] presents an overview of the strength of data protection laws around the world. While there has been some research done into modeling and implementing regulatory requirements, the approaches are often rather generalized. Current research [2,5] related to the GDPR has focused on specific principles of the regulation while its representation has received little

attention. We feel that such a representation would be useful from multiple perspectives as it would provide a visual overview of the regulation while also finding use in the development of organizational privacy policy.

In [2], Robol et al. present a goal-based modeling language to model the social aspects of the GDPR and relationships between the different actors like the data subjects, controllers and processors. The approach uses the Socio-Technical Security (STS) method in the method's formal language, STS-ml, which the authors extend to accomodate the GDPR. A method to formally represent and perform a verification of the privacy policies is also described. The authors state that while relevant to the aims of modeling the regulation's social aspects, further work is required to capture its other constraints.

Gjermundrød et al. [1] present the privacyTracker framework that is specifically tailored to track the data disclosure activities that the GDPR protects against. The framework makes it possible to identify the entities that receive and share a consumer's personal data by constructing a tree-like data structure that maintains traceability as disclosures occur. A prototype that implements the framework is also presented.

Diamantopoulou et al. [5] propose a meta-model to derive privacy level agreements (PLAs) from the concepts mentioned in the GDPR. PLAs are similar to service level agreements (SLAs) specifically tailored towards the privacy domain. The authors argue that PLA adoption will enhance consumer confidence and trust due to the guarantees and transparency they provide up front with regards to personal data processing and usage. This work is suited to capture the principle of consent as defined in the GDPR while other aspects such as technical measures to ensure data security are not explicitly captured.

In [7], Becker et al. introduce a meta-design for integrating regulatory requirements into the information system development process to ensure legal compliance. The meta-design aims to be applicable to any such regulations and is represented as a four-field matrix that describe four perspectives that must be considered in order to account for regulations. The meta-design is instantiated on a use case and demonstrated. The authors believe the approach could ease the communication challenges between information system experts and legal experts when translating regulations into system requirements.

In [3], Islam et al. provide a framework to elicit security requirements from a regulation during the requirements stage of the software development lifecycle. The framework is developed bearing in mind traceability of the requirements to specific areas of the regulation. The framework is expressed in the Secure Tropos goal modeling language that the authors believe provides a translatory medium to move from legal requirements to security requirements. Additionally, the framework supports the analysis of potential non-compliances and provides remedial measures for the same.

3 A Model of the GDPR

Figures 1 and 2 illustrate a preliminary version of the GDPR model. The regulation text has been analyzed to identify key aspects, properties and the

associations between them. Figure 1 represents the entities (human or otherwise), actions and artifacts defined in the GDPR. Associations of the rights of data subjects to the elements of Fig. 1 are shown in Fig. 2. Table 1 shows the article that the class or association corresponds to within the GDPR. This table is constructed to assess the completeness of the model in terms of capturing the regulation text. It is the aim of this research to assess each article of the GDPR for relevance to its representation within the model. This table can provide a measure of completeness.

Personal data is the foremost consideration of the legislation defined by Article 4(1), *"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); ... by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;"*. We have represented it with the class PersonalData possessing the attribute category that describes the type of data collected using the enumeration, DATA_CATEGORY. This attribute takes values that correspond to categories mentioned in the GDPR. Additionally, the value CRIMINAL_RECORD is used to assess the constraint defined in Article 10 which states, *"Processing of personal data relating to criminal convictions ... shall be carried out only under the control of official authority..."*.

Data processing as defined in Articles 4(2), 4(4) and 4(5) (Processing, profiling and pseudonymization) is captured by the DataProcessing class. Additionally, Article 4(23) describes cross-border processing. The attributes member_states and main_establishment have been added to represent this in case personal data is processed in more than one EU member state. Main establishment is the place of the organization's central administration within the EU. In addition, Article 28(1) requires that the organization implement *appropriate* technical measures to meet the regulation requirements. This motivates the inclusion of the TechnicalMeasures class linked to DataProcessing via the association implements. This class has two attributes, category and stereotype. These attributes are based on a taxonomy [4] that categorizes privacy-enhancing technologies based on their technology group (such as data protection or communication protection) and their general privacy goal, also called stereotype (such as secure channel or privacy-enhanced computation).

Similarly, other key aspects such as consent and the different actors (processors, controllers, supervisory authorities, etc) have been represented. The GDPR defines several roles that we have generalized with the Actor class. Controllers can also be processors which is why class Controller has the boolean attribute is_processor. Consent is given for a specific Purpose and is described with several attributes that represent conditions under which the consent is valid. This is further discussed in Sect. 4.2. The regulation also describes certain artifacts that can be used to validate compliance. For example, Articles 30(1) and 30(2) state that the controller and processor must maintain a record of all processing activities. This is represented through the artifact ProcessingLog.

Rights of data subjects have been captured in Fig. 2. Associations between them and the classes they impact have also been added. For example, Article 16 defines the right of the data subject to have his or her personal data corrected when relevant. This right is further linked to the notification obligation placed on the controller if personal data has been rectified as described in Article 19. As such, an act of Rectification can trigger a Notification.

The next section provides a discussion on how this model can aid in the development of organizational policy and rules to evaluate GDPR compliance.

Table 1. Articles in the GDPR and the model entities that represent them

Article	Related entity	Related association
4(1)	DataSubject	
4(11), 7	Consent	
7(1,2,3)		Consent <<manifests>> ConsentAgreement
4(13,14,15)	DATA_CATEGORY	
4(1)	PersonalData	
4(9)	Recipient	
4(21)	SupervisoryAuthority	
4(6)	FilingSystem	
4(2,5,4,23)	DataProcessing	
4(7)	Controller	
4(8)	Processor	
4(10)	ThirdParty	
37(1)	DataProtectionOfficer	
28(1)	TechnicalMeasures	
30(1,2)		DataProcessing <<manifests>> TechnicalMeasures
15(1)	Access	
4(3), 18(1,2)	ProcessingRestriction	
19	Notification	
20(1,2)	Portability	
16	Rectification	
17(1,2)	Erasion	
7(3)		ProcessingRestriction <<withdraws>>Consent

4 Application Directions

The model of the GDPR is an artifact upon which practical approaches to achieving regulation compliance can be developed. This section introduces some practical directions.

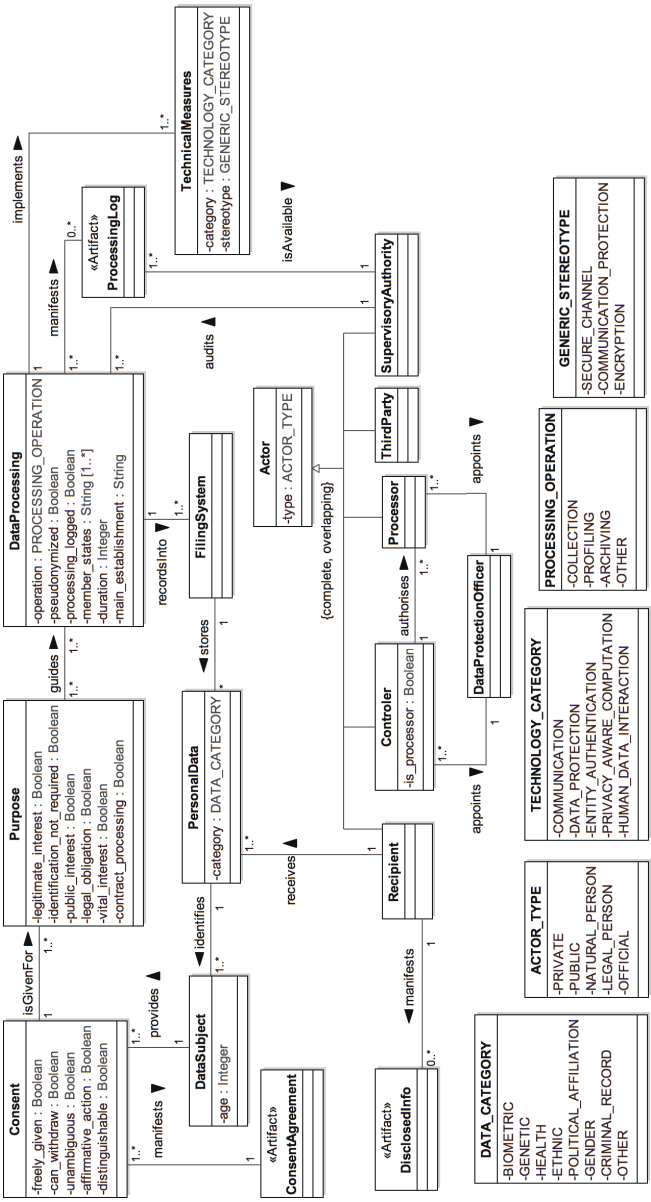


Fig. 1. GDPR entities and their associations

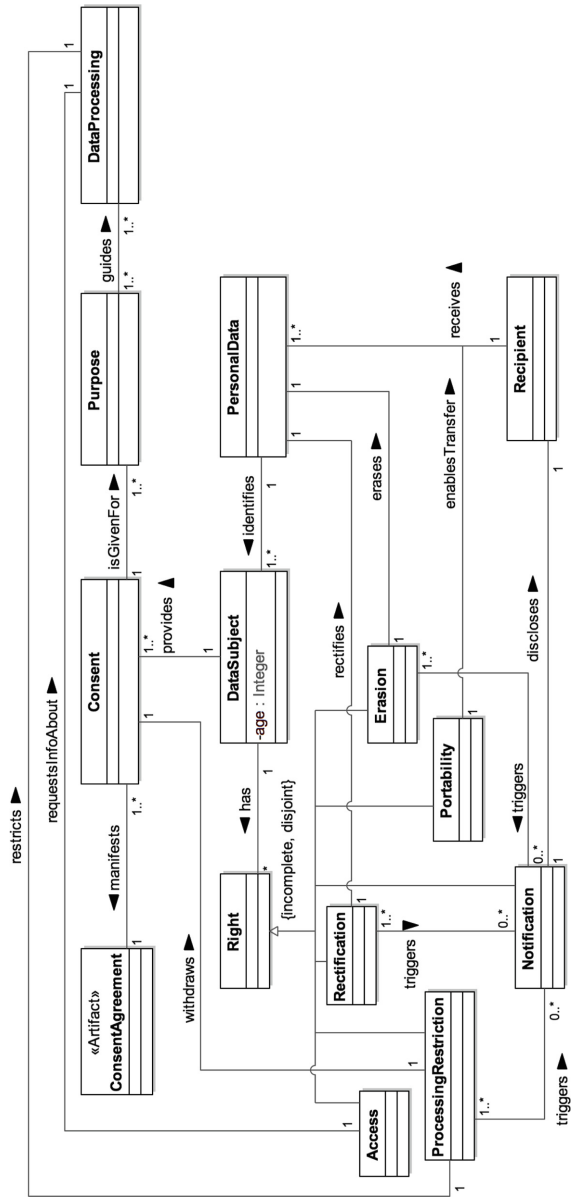


Fig. 2. GDPR rights and their associations to entities

4.1 From Model to Organizational Privacy Policy

It is important to remember that the model is a tool to aid development of organizational privacy policy. We see it being used as a frame of reference to drive the definition of privacy rules as it provides a convenient, yet concrete overview of the GDPR. By understanding the existing state of data processing activities in an organization and instantiating them within the model, areas of partial compliance or non-compliance can be identified from which internal policies can be constructed and even strengthened. In [15], Sing presents a software tool based on an earlier iteration of the model that uses business process models drawn in BPMN as an input. The model of the GDPR is instantiated in terms of the process models to carry out a semi-automated analysis of GDPR compliance violations. A meaningful analysis, of course, requires that the model should adequately cover the important aspects of the GDPR. Additionally, it is important to determine whether there are articles or clauses in the regulation that cannot be represented in the model but are important considerations to avoid the hefty GDPR fines. This motivates its critical evaluation at different stages. Primarily, we are interested in knowing whether the GDPR can be represented via the model or whether it requires an addendum.

For example, in Fig. 1, the attributes of the Consent class captures conditions stated in Article 7. The third clause of this article states that the data subject has the right to withdraw consent at any time and that it must be easy to withdraw consent. While the presence of a consent withdrawal mechanism can be verified using the attribute can_withdraw, the subjective nature of its ease of removal leads to the question of whether trying to represent the latter part of this clause is possible or even meaningful. Similar considerations will have to be made on other subjective texts within the regulation.

4.2 Compliance Rule Extraction

After checking the completeness of the model, the next step would be annotating constraints defined in the GDPR within the model to provide a basis for compliance rule extraction [15]. Constraints are the conditions that determine applicability of the regulation's articles within a given scenario. While the associations between classes are useful for comprehension of the main entities and relations between them, the constraints are what will define the rules around them. For example, the regulation places several conditions for determining the validity of consent or its need: (*i*) Article 6(1) and 11(1) provide several conditions upon which consent need not be collected (*ii*) Article 7 provides conditions under which obtained consent is considered valid and (*iii*) Article 8(1) states that the consent obtained is valid when the data subject is younger than 16 years only if it is provided by the holder of parental responsibility of the child.

Regarding the first condition, the class Purpose captures these conditions as attributes. If any of them are true upon instantiation within a scenario, consent need not be collected for the processing to be lawful. The second condition has been represented as attributes in class Consent. All of these attributes need to be

set to true for the obtained consent to be valid. The third condition is yet to be expressed through the model. These conditions enable us to extract compliance rules (CR) from the model that can be stated as follows:

CR1: Consent is not required if any of the attributes of class Purpose are true
CR2: Consent is only valid if all attributes of class Consent are true

Similar rules can be constructed from the attributes of other classes and the conditions described in the GDPR.

4.3 Modeling Language Compliance

The GDPR mandates the implementation of privacy by design principles when carrying out data processing activities. This means that organizations should implement *technical* and *organizational* measures to protect data privacy. From this perspective, achieving compliance can be broken down into two categories of improvement factors - technological factors and process factors. For example, pseudonymization would be a technological factor while the collection of consent would be a process factor.

This brings us to another potential application of this model which lies in the development of GDPR-compliant extensions to modeling languages [16]. If compliance rules extracted from the model can be expressed using an extension to a modeling language such as Business Process Model and Notation or the i^* framework, it stands to reason that models developed within those languages would provide the privacy guarantees required to achieve regulation compliance.

5 Concluding Remarks

In this paper, we presented a preliminary version of a GDPR model. This representation is intended to provide a simpler, visual overview to aid process implementers in understanding the associations between different entities in the GDPR. Related literature was reviewed to identify gaps that position this contribution. An approach to use this model as a tool to develop organizational privacy policy was also described along with an illustration on compliance rule extraction.

Future work includes determining how much of the legislation can be justifiably captured within such a representation for the purpose of practical application. Also, constraints are yet to be captured within the model. For example, the Consent class should indicate that consent obtained from a data subject is only valid if the data subject is at least 16 years of age. Formalization of the compliance rules in Object Constraint Language can also be explored. The model has been developed by a team with expertise in conceptual modeling and business process management. A significant threat to its validity is that legal expertise was not involved during its development so its correctness cannot yet be asserted. It should be noted that the development of this model is a sub-goal of broader

research aimed at developing an approach towards GDPR compliance management using business process modeling techniques. The approach has currently been validated on real scenarios (such as [15,17]) which will help refine the model even further.

Acknowledgment. This research is partly supported by the Estonian Research Council (grant IUT20-55).

References

1. Gjermundrød, H., Dionysiou, I., Costa, K.: privacyTracker: a privacy-by-design GDPR-compliant framework with verifiable data traceability controls. In: Casteleyn, S., Dolog, P., Pautasso, C. (eds.) ICWE 2016. LNCS, vol. 9881, pp. 3–15. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46963-8_1
2. Robol, M., Salnitri, M., Giorgini, P.: Toward GDPR-compliant socio-technical systems: modeling language and reasoning framework. In: Poels, G., Gailly, F., Serral Asensio, E., Snoeck, M. (eds.) PoEM 2017. LNBIP, vol. 305, pp. 236–250. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70241-4_16
3. Islam, S., Mouratidis, H., Wagner, S.: Towards a framework to elicit and manage security and privacy requirements from laws and regulations. In: Wieringa, R., Persson, A. (eds.) REFSQ 2010. LNCS, vol. 6182, pp. 255–261. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14192-8_23
4. Pullonen, P., Matulevičius, R., Bogdanov, D.: PE-BPMN: privacy-enhanced business process model and notation. In: Carmona, J., Engels, G., Kumar, A. (eds.) BPM 2017. LNCS, vol. 10445, pp. 40–56. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65000-5_3
5. Diamantopoulou, V., Angelopoulos, K., Pavlidis, M., Mouratidis, H.: A metamodel for GDPR-based privacy level agreements. In: ER Forum 2017 and the ER 2017 Demo Track (2017)
6. EU General Data Protection Regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG
7. Becker, J., Knackstedt, R., Braeuer, S., Heddier, M.: Integrating regulatory requirements into information systems design and implementation. In: 35th International Conference on Information Systems (2014)
8. TRUSTe Consumer Privacy Index (2016). https://www.trustarc.com/resources/privacy-research/ncsa-consumer-privacy-index-us/
9. UK Data Protection Act (1998). https://www.legislation.gov.uk/ukpga/1998/29/contents
10. Australian Privacy Act (1988). https://www.legislation.gov.au/Details/C2014C00076
11. Malaysian Personal Data Protection Act. http://www.pdp.gov.my/index.php/en/
12. Facebook Processes Personal Information Obtained Without Consent. https://www.cnet.com/news/shadow-profiles-facebook-has-information-you-didnt-hand-over/
13. Facebook and Cambridge Analytica Scandal. https://www.cnet.com/news/facebook-cambridge-analytica-data-mining-and-trump-what-you-need-to-know/
14. Strength of Data Protection Laws Around the World. https://www.dlapiperdataprotection.com/

15. Sing, E.: A Meta-model Driven Method for Establishing Business Process Compliance to GDPR. Master thesis, University of Tartu (2018). http://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=62090&year=2018
16. Çelebi, I.: Privacy Enhanced Secure Tropos (PESTOS): A Privacy Modeling Language for GDPR Compliance. Master thesis, University of Tartu (2018). http://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=62167&year=2018
17. Abbasi, A.: GDPR Implementation in an Airline's Contact Center. Master thesis, University of Tartu (2018). http://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=62295&year=2018

Trust Handling Framework for Networks in Cyber Physical Systems of Industry 4.0

Marina Harlamova and Marite Kirikova(✉)

Riga Technical University, 1 Kalku, Riga 1658, Latvia
marina.harlamova@edu.rtu.lv, marite.kirikova@rtu.lv

Abstract. Industry 4.0 is a concept of automation and data exchange in a manufacturing environment that combines advances of cyber physical systems, Internet of Things, cloud computing, and cognitive computing. It is a highly networked system where only trustworthy elements are allowed to participate in the network. Therefore, the possibility of having tools for indicating whether a particular node can or cannot be trusted in the network becomes an essential issue in the successful functioning of cyber physical systems in Industry 4.0. The paper proposes a trust handling framework as a base for such tools. The framework is based on knowledge derived from related work in social sciences, information technology and computer sciences, and also distributed sensing systems. The framework consists of several components that can handle initial and continuous trust issues at the level of nodes and at the level of network in cyber physical system of Industry 4.0.

Keywords: Trust · Trust framework · Cyber physical systems
Industry 4.0

1 Introduction

Industry 4.0 [1] is a concept of automation and data exchange in a manufacturing environment that combines advances of cyber physical systems (CPS) [2], the Internet of Things (IoT), cloud computing and cognitive computing. Components of CPS in the Industry 4.0 context typically would be industrial machines, automated guided vehicles, robots, mobile robots, sensors and human workers [3]. It is expected that CPS technology will change the approach of human interaction with engineered systems, similarly to how the Internet has changed the approach of human interaction with information [4]. The highly networked and flexible environment of Industry 4.0 requires that different network participants (human beings as well as technical devices (e.g., sensors)) must be able to know whether they can trust other nodes in the network. Therefore the research in progress presented in this paper aims at answering the following research question "What is the trust handling framework appropriate for CPS of Industry 4.0?".

To answer this question, the Industry 4.0 smart production line installed on premises of Aalborg University was investigated. A case study [5] of a similar experimental setup revealed underlying security issues and industry needs. The gathered requirements were considered during the design of the conceptual framework. To identify theoretical trust forming factors, the concept of trust was analyzed from

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 29–43, 2018.
https://doi.org/10.1007/978-3-319-99951-7_3

different perspectives. Based on the knowledge amalgamated in the analysis, a conceptual framework for trust handling was constructed.

The paper is organized as follows. The related works used in the analysis are listed in Sect. 2. The proposed framework and its components are presented in Sect. 3. The initial theoretical evaluation of the framework is provided in Sect. 4. Brief conclusions and further research directions are given in Sect. 5.

2 Related Work

To gain an understanding of the trust concept, its interpretations in domains of social science, information technology (IT) and computer science, and ad-hoc distributed sensing systems were investigated. A study of a trust concept across these domains contributed to our understanding of trust needs in CPS of Industry 4.0 and the trust models for all three above-mentioned domains were built [6]. These models served as the basis for the proposed framework for trust handling in the networks of CPS of Industry 4.0 that is presented and discussed in the next section. The trust factors revealed in the analysis of trust concept [6] and corresponding related works are reflected in Table 1. The requirements for a trust framework were also extracted from expert narratives collected by the Manufacturing Academy of Denmark (MADE) [5] and general network security requirements, such as CIA triad [7]. More details on related work are available in [6].

Table 1. Trust factors.

Trust factor	Literature
Social Sciences	
• Volition	[8, 9]
• Competence	[8, 10]
• Consistency	[10–12]
• Integrity	[8, 10–13]
• Benevolence	[8, 11]
• Emotion	[8, 13, 14]
• Reliability	[8–13, 15]
• External trust metrics	[10, 14]
Computer Science	
• Trust metrics in performance dimension	[16–20]
• Trust metrics in process dimension	[16, 18, 20–23]
• Trust metrics in purpose dimension	[16, 17, 19–21, 23]
• External trust metrics	[15, 17, 18, 22]
Distributed Sensing Systems	
• Trust metrics in cooperation dimension	[24–29]
• Trust metrics in data integrity dimension	[29–33]
• Trust metrics in availability dimension	[29, 34, 35]
• Trust metrics in access control dimension	[29, 33]
• Trust metrics in communication integrity	[32, 33]

3 Conceptual Trust Handling Framework

In this section we propose a conceptual framework for trust handling (see Fig. 1).

As already mentioned, the framework amalgamates knowledge obtained during the analysis of related work that was organized in conceptual models for trust handling in different domains [6]. The models revealed the main trust dimensions in each domain. After analysis of these models a conceptual framework for trust handling in CPS of Industry 4.0 presented in Fig. 1 was created.

The components of the proposed conceptual framework, as well as certain trust functions are described in the remainder of this section. The framework has six main components: the node initial trust module, the node continuous trust module, the node trust engine, the network trust engine, the alarm system module, and the graphical user output module. Each module concerns one or more trust dimensions with corresponding trust functions. Only a limited number of trust functions are proposed in this work. Depending on production environment operation specifics, the framework module components can be extended with other suitable trust functions.

Node Initial Trust Module: When a new node is added to the network, initially it is considered as a "not trusted" participant until it fulfils the trust criteria defined in the initial trust module. In the proposed framework two trust dimensions are assessed in this component – access control and initial credibility. The trust assessment process for the initial trust module is presented in detail in Fig. 2.

The access control trust dimension is composed of identification, authentication and authorization steps. If the node completes the steps required by the access control dimension, it is authorized to be in the network, thus fulfilling the access control requirements. Another part of the trust value is gained from initial credibility. *Initial credibility*, the second dimension of the initialization module, is composed of vendor data validation, software validation and firmware validation functions. As a new node joins the network, its properties, vendor attributes and code version are scanned. Attributes of validated and functioning elements are loaded to the version control repository (see (4) in Fig. 2). If scanned properties correspond to standard values from the element storage, the node has fulfilled the initial credibility dimension criteria. A similar credibility check is also reinitiated during network operation time in order to identify unexpected element or firmware changes since the last check.

The access control dimension is assessed using the higher-level function T_{ac} (1). This function is not composed of lower level functions; it simply checks whether the node has the necessary privileges to join the network. The node has to pass identity, authentication and authorization steps to join. The *initial credibility* dimension, on the other hand, is represented by a higher-level function T_{ic} (2) that is derived from lower-level functions t_1, t_2, t_3 (3). The version control repository component (4) is used to retrieve defined firmware and software version parameters, attributes and versions for lower-level functions. Next, higher-level trust values are passed to the trust aggregation component (5). At this point a total initial trust value is calculated. The value received from the access control dimension is either 0 or 1 (the node is either authorized or excluded). If value is 0, the trust evaluation process is terminated and any further actions cancelled. If value is 1 and the node is authorized, the next *initial credibility*

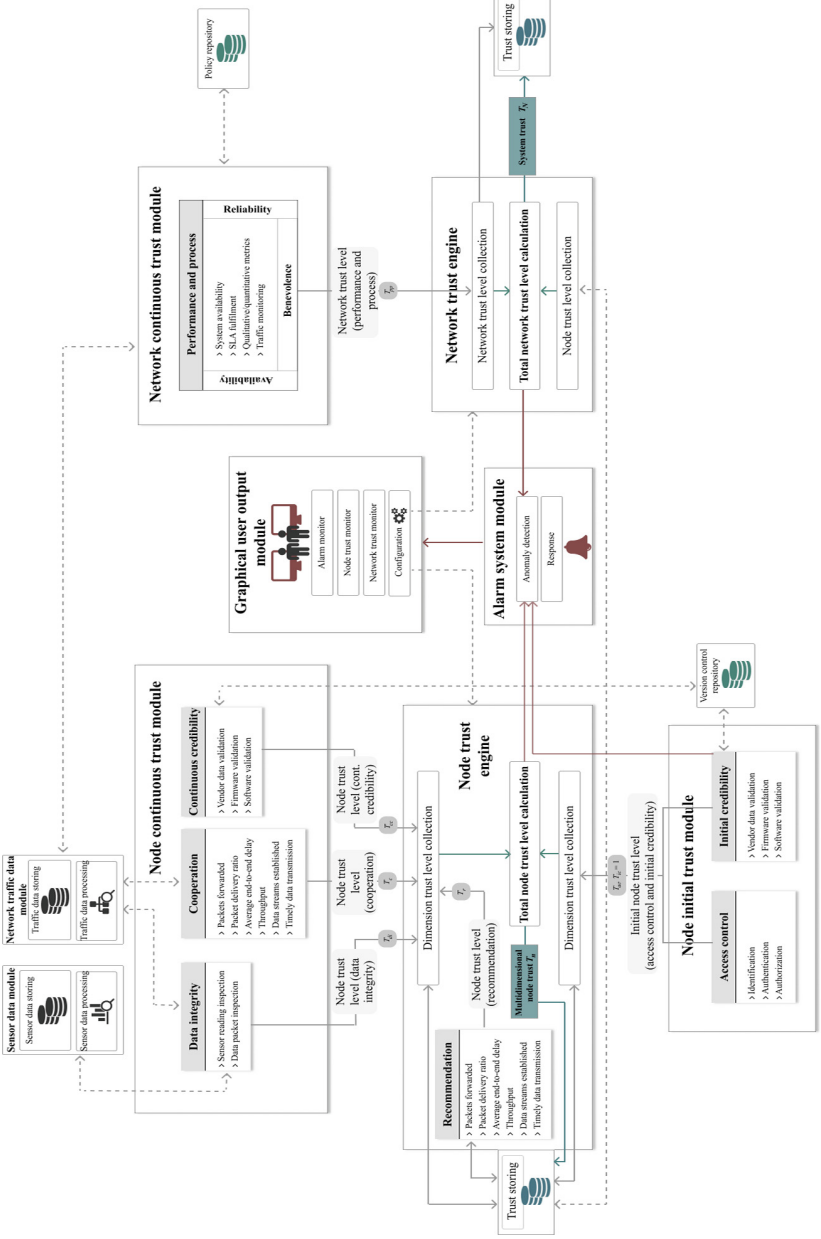


Fig. 1. Conceptual trust handling framework for the networks of CPS of Industry 4.0.

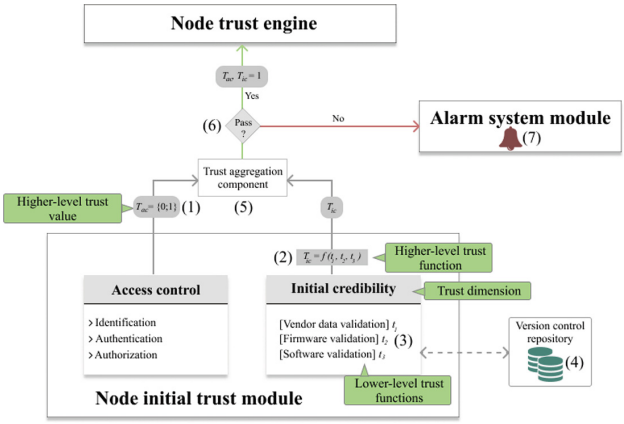


Fig. 2. Functions in the node initial trust module.

trust value is added to the equation. If it is higher than a specified credibility trust threshold at decision point (6), total initial trust is set to 1; the node's total trust value is passed to the trust engine for storage and the element can join the network. In a scenario, when the node has not gained enough *initial credibility,* the system alarm is triggered in the alarm system (7) because of a potential malicious element attempting to join the system.

Node Continuous Trust Module: Node-level continuous trust module amalgamates trust dimensions that express trust level of a certain node at any given time. Nodes that enter the module have gained the initial trust by the dimensions of *access control* and *initial credibility*. The continuous trust module has dimensions of *data integrity*, *co-operation* and *continuous credibility*. In wireless sensor networks, due to their distributed nature, trust is often calculated and stored locally in network nodes. Limited node computational power and battery lifetime necessitate that often just one trust function is chosen for trust evaluation. For instance, monitoring whether the packet has been forwarded and received from a node would require less processing than to check payload integrity and accuracy at every incoming event. For CPS of Industry 4.0 under investigation, it is assumed to have central elements – a node trust processing unit and a node trust storage unit. A dedicated computer or server, connected to a constant electrical supply and having appropriate storage would not suffer energy loss from being required to calculate several functions at once, while for a tiny sensor node it would result in a dramatic decrease of its lifetime. Therefore, many different trust functions can be enabled in this module, depending on application specifics and computational limitations of the central processing element. Node behaviour and attributes are constantly monitored in the node continuous trust module that handles *data integrity, cooperation* and *continuous credibility* dimensions. A trust assessment process example is presented in detail in Fig. 3. The module is integrated with additional data modules (1) and (2) where variables needed for lower-level functions t_1, t_2 (3) are stored and processed. Each dimension is represented with a higher-level function T_{di} (4) that aggregates lower-level dimension functions. An additional

repository component (5) is used for *continuous credibility* dimension derivation, similar to the initial trust module represented in Fig. 2. For each dimension a trust level calculation component uses the higher-level trust function to form the trust value of each dimension. Resulting values are carried to the node trust engine to calculate and store the total trust level of an element.

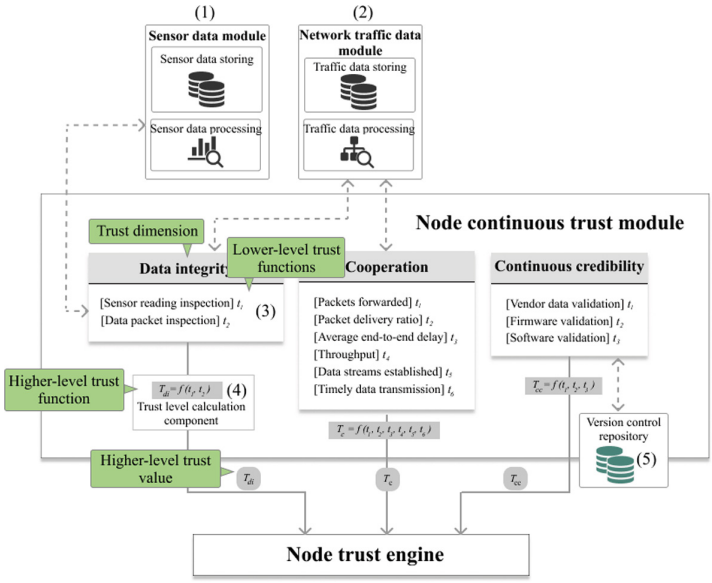


Fig. 3. Functions in the node continuous trust module.

The *data integrity* aspect ensures that collected, transferred and processed data is accurate and not interfered with in a production line system. The *data integrity* dimension of trust utilizes sensor reading inspection and data packet inspection functions. For fulfilment of these functions, additional processing modules are required, namely, the sensor data module and network traffic data module. Sensor readings and exchanged network traffic between elements are stored and processed in the modules. Sensor data analysis helps to identify anomalous readings which could be a signal of network intrusion or node malfunctioning. Network analysis post-processing algorithms are used to determine if, for instance, data packet control and sequencing information is valid, source and destination network addresses are correct, there are no unusual errors, packet payload values are correct, etc. For the *data integrity* dimension of the continuous trust module the context-awareness aspect is crucial. Consider an example with a temperature measuring node. Temperature readings are collected for a particular sensing element. The setting where a node is installed is a robotic arm tool. In such a context, reporting a reading of −20 °C, given that the arm operates at room temperature, would be an anomaly. However, if a sensing node is moved to an actual environment, where operations occur at temperatures below 0 °C, this reading would

not be considered as anomalous. Context-specific information is significant for avoiding false alarms and invalid reading interpretation. For more complex CPS network elements, such as robotic arms or drilling machines, even more context-specific knowledge is required to evaluate the transferred data accuracy.

The cooperation trust dimension for CPS of Industry 4.0 in the node continuous trust module collects various lower-level functions that evaluate the connectivity and cooperability between close-distance elements. When a node needs to obtain the trust value of another node in the CPS network (i.e. by command of the central element), it checks the recorded list of neighbouring nodes. Cooperation for nodes is assessed based on the recorded list. Some or all of the following trust functions can be applied in this module: Data or control packets forwarded; Data packet delivery ratio; Average end-to-end delay; Throughput; Data streams established; and Timely data transmission.

To apply defined cooperation functions, the network traffic data module is used again for traffic collection and analysis. For instance, the "data packets forwarded" function measures how much packet loss occurs between two communicating elements, e.g. a module controller and a module screen microcomputer. If sensor clusters are installed in the factory network, the "average end-to-end delay" function is used to detect black holes where data is lost. Several functions can be used at once to obtain trustworthy results. Similar to the *data integrity* function, it is essential to define a context for each function and element. Measurements of the "data streams established" function will vary depending on element utilization and application. As the last dimension of continuous node trust, *continuous credibility* is proposed. Functions performed in this dimension are exactly the same as in the *initial credibility* dimension. These functions are vendor data validation, firmware validation, and software validation. This trust dimension is scheduled for assessment at specific time intervals. The functions check if network element meta- and vendor-specific data have remained intact since the last check. If any unexpected changes in software and firmware are detected, the trust score is decreased accordingly.

Node Trust Engine: The central node-oriented element of the proposed trust handling framework is the node trust engine. The node trust engine deals with recommendation dimension trust level calculation and total multidimensional trust calculation for a particular network element. Dimension trust levels, as well as multidimensional total trust score per node, are stored in the node trust engine module. The node trust engine is composed of the following functional components, also seen in Fig. 4: *Dimension trust level collection component* (1) – computed trust scores per dimension are collected and prepared for further total trust calculation; recently recalculated values are updated in the trust storage; older trust values can be retrieved from the same trust location; *Recommendation trust level calculation component* (2) – *recommendation* dimension is the last trust dimension proposed in the framework; it is placed in the node trust engine module for indirect trust value calculation; values for calculation are taken from the trust storage entity; *Total trust level calculation component* (3) – inputs provided from module-specific trust dimensions are used as variables for calculation of multidimensional trust of an element. Here, the six highest-level trust dimensions T_{ac}, T_{ic}, T_{di}, T_c, T_{cc} and T_r are combined in one function to determine the total multidimensional node trust value T_n. This value (4) is then stored in the trust storage until the

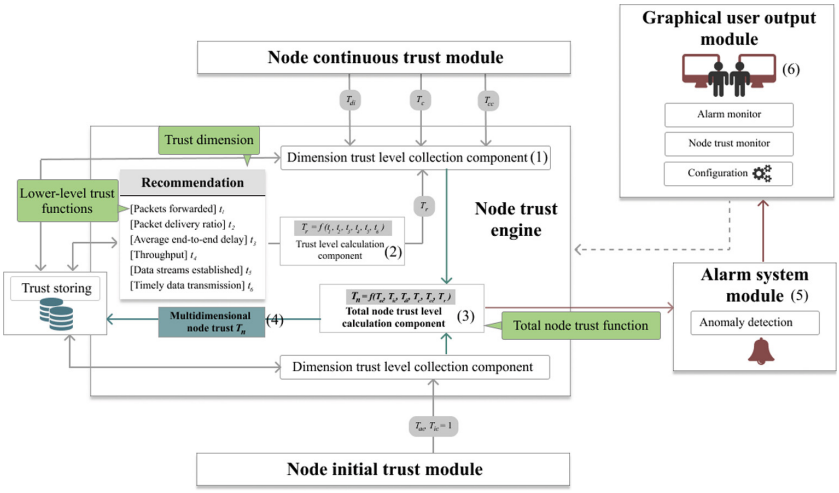


Fig. 4. Functions of the node trust engine.

next trust update iteration takes place. An element (node) for which the resulting trust score is below the specified threshold is likely to be compromised and should be excluded from the network. For this purpose, the node trust engine is integrated with an alarm system module (5). Untrusted nodes are reported to network operators. The graphical interface module (6) is installed to receive alarms, review statistics, and configure engine functions. Positive trust node values from the whole CPS are also communicated to the network trust engine. Trustworthiness of element sets contributes to overall trust evaluation of the network. It has already been stated that the *recommendation* trust dimension suggests the same functions as the *cooperation* dimension, considering remote node trust assessment of a particular node instead of neighbour node evaluation of a particular node. The same communication and network analytical functions, such as data or control packets forwarded, data packet delivery ratio, average end-to-end delay, etc., can be utilized for this purpose. A node recommends a more distant node if, historically, their cooperation has been successful. Due to larger trust coverage, this dimension is placed in the node trust engine with direct access to the trust storage for calculations. The combined dimensions of *cooperation* and *recommendation* form a node consensus, i.e. a composed trust feedback aggregated from nodes that are closer and further apart.

Network Continuous Trust Module: The network-level continuous trust module contributes to operational requirements or lower-level trust dimensions of *availability*, *benevolence* and *reliability* of the network. The module appraises trust by detecting unusual behaviour; but on a generalized network-level, instead of a more granular node level. Simply put, results of trust calculations on this level determine how trustworthy the network is from the perspective of a human user. For this module there are no detailed specifications; as here we focus more on the node-level trust dimensions. Network-level security mechanisms are the subject of a broad research topic: several

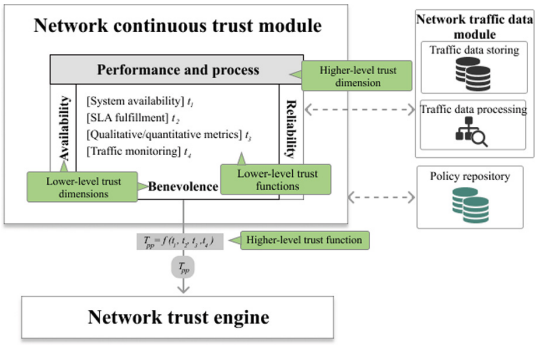


Fig. 5. Network continuous trust module functions.

efficient network monitoring approaches exist. Since network level monitoring is an essential part of an autonomous, dynamic system, it is included in the conceptual framework. Figure 5 shows that the network continuous trust module is composed of an extensive *performance and process* higher-level trust dimension. Formative lower-level dimensions of *performance and process* are network *availability, benevolence* and *reliability.* Lower-level trust functions t_1, t_2, t_3, t_4 form a higher-level trust function T_{pp}. Lower-level functions also use data from the policy repository element and network data traffic module.

For the *performance and process* higher-level trust dimension the following functions are suggested:

- System availability: This function measures whether a production line wireless network is up and running for a longer than normal period according to the agreed level of operational performance.
- SLA fulfilment: Service-level agreements are defined based on specific tasks performed by production lines in the factory.
- Qualitative/quantitative metrics (KPIs): Different metrics and key performance indicators can be used, depending on functions and operations of the factory.
- Traffic monitoring: Network traffic monitoring through bandwidth analysis and performance monitoring helps to evaluate network performance and detect suspicious behaviours.

Network Trust Engine: The network trust engine calculates the overall trustworthiness of a network, stores it in the trust storage, and reports the values to the graphical user interface. In Fig. 6 the network level trust assessment process is shown.

Firstly, network-level trust value T_{pp} from network continuous trust module is calculated and transferred to the network trust engine (1). Then, from the node trust engine storage element, individual node trust values T_{n1} … T_{nn} are collected (2). The trust dimensions T_{pp}, T_{n1}, …, T_{nn} are combined in one function (3) to determine the total network (system) trust value T_N (4). The total network trust value is stored in the trust storage. If the calculated value is below the required system trust level, an alarm is raised in the alarm system module (5) and displayed in the graphical output module for

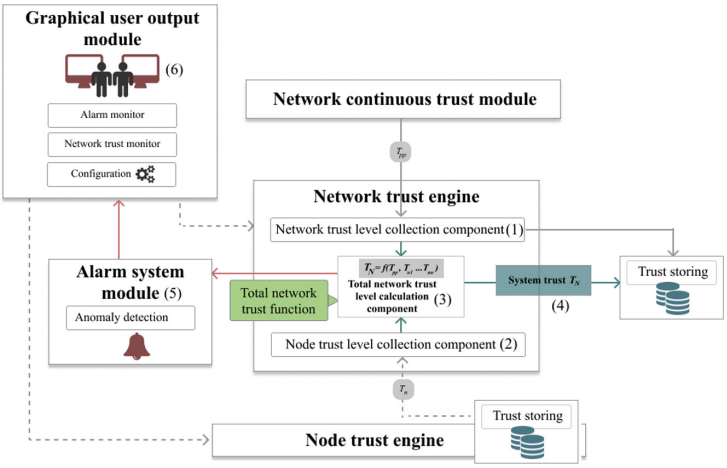


Fig. 6. Functions of the network trust engine.

network operators (6). Critical system trust values might imply that a system is under attack and an intervention is necessary. Additionally, network trust engine functions can be adjusted from the graphical user output module.

Alarm system module and Graphical user output module. An alarm module is included in the proposed trust framework to inform operators about detected anomalous events and potentially malicious activities in the CPS network. By itself the alarm module does not cover any trust dimensions. As soon as the calculated trust level falls below the specified threshold, an abnormal event is logged and the alarm is raised for further consideration of network users. It can happen in three cases:

- A newly added element fails to pass the initial trust module's dimension of initial credibility,
- An operating element trust level falls below a certain threshold level,
- Network level trust falls below a certain threshold level.

The graphical user output module is the interaction point between established trust mechanisms and human users of a CPS. The graphical user output module offers, but is not limited to, the following features: real-time network monitoring, business intelligence reporting, alarm, and event display.

4 Compliance with Established Requirements

Current technical setting of smart production lines, reported user stories and known network security requirements [5, 7] provide the basis for shaping the production facilities of the future. In this section, we verify whether or not the components of the proposed trust framework are in conflict with the features expected from future production facilities. This verification helps us to see to what extent the proposed framework supports the requirements of future networks of CPS of Industry 4.0. The verification results are summarized in Table 2.

Table 2. Fulfilment of industry and security requirements.

Origin	Requirement	Relevance	Relation to requirement	Fulfilment method	Primary contributing components
Required features from user stories	Data availability	High	Supportive	Sensor data transfer Network traffic transfer Access to trust scores	Sensor data module Network traffic data module Trust storage
	Centralization	High	Supportive	Centralized trust processing Centralized trust storing	Node trust engine Network trust engine Trust storage
	Wireless connectivity	Low	N/A	N/A	N/A
Underlying network challenges	Low latency	Low	N/A	N/A	N/A
	Low energy consumption	Medium	Supportive	Centralized trust processing	Node trust engine Network trust engine
	High capacity	Low	N/A	N/A	N/A
	Mobility and flexibility	Medium	Affective	*[trust score is affected by mobility factor]*	N/A
Security requirements	Availability	High	Supportive	Is a lower-level trust dimension	Network continuous trust module
	Integrity	High	Supportive	Is a higher-level trust dimension	Node continuous trust module
	Confidentiality	High	Supportive	Is implied by access control trust dimension	Node initial trust module
	Access control	High	Supportive	Is a higher-level trust dimension	Node initial trust module
	Accountability and auditing	High	Supportive	Anomaly reporting Access to trust scores SLA, KPI definition Software/firmware control	Alarm system module Trust storage Policy repository Version control repository

It was established that desired features that are to be implemented in a future Industry 4.0 CPS are *data availability, centralized code base* and *wireless connectivity* [5, 7].

Data availability stands for the need of on-demand access to data from any node of the network with rights to do so. It includes enabling both access to data storage points and mechanisms for efficient data sharing and processing. For this, a scenario of a common factory network can be considered, i.e. the scenario where data from sensors are made available for local and central controllers. The proposed framework is aligned with this requirement by utilizing data collected from sensing elements. Collected sensor data is mined, as seen in Fig. 3, in the sensor data processing unit.

The need for centralization is also covered in the proposed trust framework. The concentration of computing units under a single element, as seen in the framework for modules of the node trust engine and network trust engine, simplifies the network setup and technical architecture, as well as saves power resources (if compared to distributed architectures, where trust calculation is performed locally in a network node) and decreases the overhead. Naturally, network architecture with the main central element calls for additional fortification and security system implementation for the central element. Putting the central trust element in jeopardy could compromise the trust ratings of the whole network. As for the feature of wireless connectivity, the proposed trust framework is not limited to network implementation techniques per se. The solution is applicable to a production line design that uses Ethernet and bus connectivity, as well as future factory environments with wireless setup.

A set of network challenges, as the network undergoes transformation from wired to wireless connectivity, is also a factor for verification [5, 7]. Listed challenges are *latency, energy consumption, capacity,* and *mobility and flexibility*. The requirement of low latency is not relevant for the proposed trust framework, as the framework is conceptual and performance metrics of a network depend on technical implementation. Low energy consumption requirement is supported by the fact that computation in the proposed framework can be centralized. It is assumed that the central computational element for network or node trust levels is connected to a constant power source. Thus, sensor lifetime is not affected by mechanisms involved in the framework. Element trust can be accessed at any time without disturbing the element for which the trust score is assessed. Capacity, similar to low latency, is a matter of technical realization rather than conceptual design. Mobility and flexibility feature of future CPS networks is slightly more demanding when it comes to the trust framework application. The results of functions of such trust dimensions as cooperation (direct trust) and recommendation (indirect trust) will become outdated as soon as the element is moved within the network. The trust score in such dimensions needs to be reset and recalculation initiated, considering the feedback from new neighbouring nodes.

General security requirements, such as availability, integrity, and confidentiality are also important for verification. Consideration of security requirements in the proposed trust framework implies that these requirements are covered; trust assessment mechanisms are evaluating the level at which the requirements are fulfilled. Elements of CIA triad [7] are directly considered in the proposed trust framework. For instance, *availability* is a lower-level trust dimension of the *performance and process* higher-level trust dimension (rf. Fig. 5) in the network continuous trust module. *Integrity* is a trust

dimension in the node continuous trust module (as seen in Fig. 2). *Confidentiality* is not proposed as a separate dimension. However, it is indirectly implied by the *access control* trust dimension in the node initial trust module, as, by definition, it stands for data protection against unauthorized access. By extension, the same trust dimension covers the security requirement of having *access controls* in place. The list of verified security needs in concluded with *accountability and auditing* requirements. Several elements in the proposed framework contribute to common concepts of these requirements. For instance, the alarm system module of the proposed trust framework is designed to detect anomalous behaviour and catch unexpected events in the network; thus contributing to an accountable environment. Node and network trust monitoring features, as well as storage of network policies, SLA and KPI definitions, and previous node and network trust scores are essentially features of internal auditing.

5 Conclusions

The paper proposes a conceptual trust handling framework for wireless networks of CPS of Industry 4.0. Trust forming metrics and dimensions for the framework are derived from related literature, as well as network and security requirements and collected user stories. Namely, the main trust dimensions for trust establishment in Industry 4.0 environment involving human and technological actors (or nodes) are *Access control, Initial credibility, Data integrity, Cooperation, Continuous credibility* and *Recommendation.* Fusion of human and artificial trust in Industry 4.0 environment required two trust levels to be introduced in the framework: network-level trust and node-level trust. An approach to define initial trust evaluation and continuous trust evaluation was proposed. Thus, the proposed trust handling framework consists of the following functional components: *Node initial trust module, Node continuous trust module, Node trust engine, Network continuous trust module, Network trust engine, Alarm system module* and *Graphical user output module.* Brief descriptions and functional values of all components are presented in this work. The paper concludes with evaluation of the proposed framework. Reported user stories and known network security requirements are used as evaluation criteria (requirements). The contribution of the framework and its components is evaluated with relevance and relationship to each requirement. The fulfillment method and main contributing components of the framework are also given for each requirement where it is applicable.

It is a matter of further research to focus on the framework components in more detail. As the next step, more thorough analysis of the proposed framework could be conducted, with inclusion of expert survey of the framework's feasibility and efficiency. Ideas presented in this work should be implemented in a laboratory setting and practical results evaluated. Our further research roadmap also includes a separate trust function evaluation.

Acknowledgement. This research was partly done within the Erasmus+ Strategic Partnership "Improving Employability through Internationalisation and Collaboration" (EPIC) project, with the support of the Erasmus+ programme of the European Union.

References

1. Kagermann, H., Wahlster, W., Helbig, J.: Securing the future of German manufacturing industry – Recommendations for implementing the strategic initiative industrie 4.0. Acatech (2013)
2. i-SCOOP, 17 March 2018. https://www.i-scoop.eu/industry-4-0/
3. Li, X., Li, D., Wan, J., Vasilakos, A.V., Lai, C.F., Wang, S.: A review of industrial wireless networks in the context of Industry 4.0. Wirel. Netw. **23**(1), 23–41 (2017)
4. National Science Foundation, Cyber-Physical Systems (CPS), 27 April 2018. https://www.nsf.gov/pubs/2018/nsf18538/nsf18538.pdf
5. Denmark, MADE casekatalog. http://www.made.dk/media/1720/made-casekatalog-2016-one-page.pdf. Accessed 23 Feb 2018
6. Harlamova, M., Kirikova, M.: Towards the trust model for Industry 4.0. In: 13th International Baltic Conference on Databases and Information Systems (Baltic DB&IS 2018), 1–4 July 2018, Trakai, Lithuania (2018)
7. Andress, J.: What is Information Security? In: The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice, 2nd edn, pp. 1–22. Syngress (2014)
8. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manag. Rev. **20**(3), 709–734 (1995)
9. Rousseau, D.M., Sitkin, S.B., Burt, R.S., Camerer, C.: Not so different after all: a cross-discipline view of trust. Acad. Manag. Rev. **23**(3), 393–404 (1998)
10. Nooteboom, B.: Trust: Forms, Foundations, Functions, Failures and Figures. Edward Elgar, Cheltenham (2002)
11. Heimer, C.A.: Solving the problem of trust. In: Cook, K.S. (ed.) Trust in Society, pp. 40–88. Russell Sage Foundation, New York (2001)
12. Hardin, R.: Conceptions and explanations of trust. In: Cook, K.S. (ed.) Trust in Society, pp. 3–39. Russell Sage Foundation, New York (2001)
13. Schoorman, F.D., Mayer, R.C., Davis, J.H.: An integrative model of organizational trust: past, present, and future. Acad. Manag. Rev. **32**(2), 344–354 (2007)
14. Eisenegger, M.: Trust and reputation in the age of globalisation. In: Klewes, J., Wreschniok, R. (eds.) Reputation Capital, pp. 11–22. Springer, Berlin (2009). https://doi.org/10.1007/978-3-642-01630-1_2
15. Trček, D.: A brief overview of trust and reputation over various domains. In: Trček, D. (ed.) Trust and Reputation Management Systems. SIS, pp. 5–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-62374-0_2
16. Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., Leimeister, J.M.: Understanding the formation of trust in IT artifacts. In: Proceedings of the International Conference on Information Systems, Orlando, Florida (2012)
17. Noor, T.H., Sheng, Q.Z., Bouguettaya, A.: Trust Management in Cloud Services. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12250-2
18. Pawar, P.S., Rajarajan, M., Dimitrakos, T., Zisman, A.: Trust model for cloud based on cloud characteristics. In: Fernández-Gago, C., Martinelli, F., Pearson, S., Agudo, I. (eds.) IFIPTM 2013. IAICT, vol. 401, pp. 239–246. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38323-6_18
19. Palmer, G., Selwyn, A., Zwillinger, D.: The "Trust V": building and measuring trust in autonomous systems. In: Mittu, R., Sofge, D., Wagner, A., Lawless, W.F. (eds.) Robust Intelligence and Trust in Autonomous Systems, pp. 55–77. Springer, Boston, MA (2016). https://doi.org/10.1007/978-1-4899-7668-0_4

20. Vance, A., Elie-Dit-Cosaque, C., Straub, D.: Examining trust in information technology artifacts: the effects of system quality and culture. J. Manag. Inf. Syst. **24**(4), 73–100 (2008)
21. Talboom, S., Pierson, J.: Understanding trust within online discussion boards: trust formation in the absence of reputation systems. In: Fernández-Gago, C., Martinelli, F., Pearson, S., Agudo, I. (eds.) IFIPTM 2013. IAICT, vol. 401, pp. 83–99. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38323-6_6
22. Gefen, D., Karahanna, E., Staub, D.W.: Trust and TAM in online shopping: an integrated model. MIS Q. **27**(1), 51–90 (2003)
23. McKnight, D.H., Choudhury, V., Kacmarc, C.: The impact of initial consumer trust on intentions to transact with a web site: a trust building model. J. Strateg. Inf. Syst. **11**, 297–323 (2002)
24. Michiardi, P., Molva, R.: Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In: Jerman-Blažič, B., Klobučar, T. (eds.) Advanced Communications and Multimedia Security. ITIFIP, vol. 100, pp. 107–121. Springer, Boston, MA (2002). https://doi.org/10.1007/978-0-387-35612-9_9
25. Buchegger, S., Boudec, J.Y.L.: Performance analysis of the CONFIDANT protocol (Cooperation Of Nodes – Fairness In Dynamic Ad-hoc NeTworks). In: The 3rd ACM International Symposium Mobile Ad-Hoc Networking & Computing (MobiHoc 2002), Lausanne, CH (2002)
26. Feng, R., Che, S., Wang, X.: A credible routing based on a novel trust mechanism in ad hoc networks. Int. J. Distrib. Sens. Netw. **9**(4), 1–12 (2013)
27. Ganeriwal, S., Balzano, L., Srivastava, M.: Reputation-based framework for high integrity sensor networks. ACM Trans. Sens. Netw. **4**, 1–37 (2008)
28. Zhang, Q., Yu, T., Ning, P.: A framework for identifying compromised nodes in sensor networks. In: Securecomm and Workshops (2006)
29. Hur, J., Lee, Y., Yoon, H., Choi, D., Jin, S.: Trust evaluation model for wireless sensor networks. In: The 7th International Conference on Advanced Communication Technology. IEEE, Phoenix Park (2005)
30. Yao, Z., Kim, D., Doh, Y.: PLUS: parameterized and localized trust management scheme for sensor networks security. In: International Conference on Mobile Ad Hoc and Sensor Systems (2006)
31. Momani, M., Challa, S.: Survey of trust models in different network domains. Int. J. Ad Hoc Sens. Ubiquitous Comput. **1**(3), 1 (2010)
32. Cordasco, J., Wetzel, S.: Cryptographic versus trust-based methods for MANET routing security. Electron. Notes Theor. Comput. Sci. **197**(2), 131–140 (2008)
33. Janani, V.S., Manikandan, M.S.: Efficient trust management with Bayesian-evidence theorem to secure public key infrastructure-based mobile ad hoc networks. EURASIP J. Wirel. Commun. Netw. **2018**(25), 1–27 (2018)
34. Rajkumar, B., Narsimha, D.: Trust based certificate revocation for secure routing. Proc. Comput. Sci. **92**, 431–441 (2016)
35. Ye, Z., Wen, T., Liu, Z., Song, X., Fu, C.: An efficient dynamic trust evaluation model for wireless sensor networks. J. Sens. **2017**, 1–16 (2017)

Resilient Collaboration

Getting to Win-Win in Industrial Collaboration Under Coopetition: A Strategic Modeling Approach

Vik Pant[1(✉)] and Eric Yu[1,2]

[1] Faculty of Information, University of Toronto, Toronto, Canada
vik.pant@mail.utoronto.ca, eric.yu@utoronto.ca
[2] Department of Computer Science, University of Toronto, Toronto, Canada

Abstract. Interorganizational coopetition describes a relationship in which two or more organizations cooperate and compete simultaneously. Actors under coopetition cooperate to achieve collective objectives and compete to maximize their individual benefits. Such relationships are based on the logic of win-win strategies that necessitate decision-makers in coopeting organizations to develop relationships that yield favorable outcomes for each actor. We follow a strategic modeling approach that combines *i** goal-modeling to explore strategic alternatives of actors with Game Tree decision-modeling to evaluate the actions and payoffs of those players. In this paper, we elaborate on the method, illustrating one particular pathway towards a positive-sum outcome - through the introduction of an intermediary actor. This paper demonstrates the activation of one component in this guided approach of systematically searching for alternatives to generate a new win-win strategy. A hypothetical industrial scenario drawn from practitioner and scholarly literatures is used to explain this approach. This illustration focuses on the Industrial Data Space which is a platform that can help organizations to overcome obstacles to data sharing in a coopetitive ecosystem.

Keywords: Coopetition · Win-Win · Design · Modeling

1 Introduction

Coopetition refers to concomitant cooperation and competition among actors wherein actors "cooperate to grow the pie and compete to split it up" [1]. Actors under coopetition simultaneously manage interest structures that are partially congruent and partially divergent [2]. Partial congruence emerges from actors sharing in certain common objectives while partial divergence emanates from each actor's pursuit of self-interest. Coopetition has become "increasingly popular in recent years" [3] and is widely observed in various domains including business, politics, and diplomacy [4].

Coopetition is predicated on the rationale of positive-sum outcomes through which all actors are better off by coopeting rather than by purely competing or solely cooperating. This aspect of coopetition requires decision-makers in coopeting organizations to develop and analyze win-win strategies. We apply a synergistic approach that combines *i** goal-modeling with Game Tree decision-modeling to generate and

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 47–66, 2018.
https://doi.org/10.1007/978-3-319-99951-7_4

discriminate win-win strategies in a structured and systematic manner. In [32], we illustrated a win-win scenario arrived at by generating a new alternative for achieving a goal, using the means-ends reasoning supported by *i** goal-modeling. In this paper, we illustrate a different pathway to get to win-win by introducing a new actor within an existing relationship between two actors. We use a hypothetical industrial scenario adapted from practitioner and scholarly literatures to explain this approach.

Coopetition research originated in the field of economics where researchers applied concepts from game theory to explain the motivations of coopeting actors [5]. According to game theory, three types of results are possible in strategic relationships between players: positive-sum, zero-sum, and negative-sum [6]. In positive-sum outcomes all players are better off and in negative-sum outcomes all players are worse off [6]. In zero-sum outcomes the amount of gain by some players equals the amount of loss by other players.

These outcomes are correlated to distinct types of strategies that are adopted by players in coopetitive relationships: win-win, win-lose, and lose-lose. Win-win strategies are the only durable options for sustaining coopetitive relationships. Win-lose strategies are unsustainable in coopetitive relationships because some actors (i.e., those that are disadvantaged) will be worse off as a result and these actors are likely to withdraw from or abandon such relationships.

2 Motivating Example: Interorganizational Knowledge-Sharing in Pharmaceutical Industry

Drug discovery and biopharmaceutical development is characterized by long innovation cycles and high capital requirements. Pharmaceutical companies share knowledge with each other to accelerate "product development processes", "reduce costs", and increase "development productivity" [9]. Coopetitive relationships within research and development (R&D) alliances in the pharmaceutical industry are described in [9]. The complexity of interorganizational knowledge-sharing in the pharmaceutical industry is discussed in [10, 11].

Knowledge-sharing can expose members of R&D alliances to the risk of knowledge expropriation through knowledge leakage [10, 11]. This is because R&D alliances can be among firms that are competitors in the marketplace. Such firms are coopetitors because they cooperate in the R&D domain but compete for customers in the marketplace. Knowledge leakage occurs when a "focal firm's private knowledge is intentionally appropriated by or unintentionally transferred to partners beyond the scope of the alliance agreement" [12]. Knowledge expropriation is an opportunistic behavior [13, 14] that is motivated by the desire of firms to engage in 'learning races' [15, 16] to 'learn faster' [17, 18] than each other in the pursuit of 'competitive advantage' [19, 20]. Knowledge management researchers refer to this phenomenon as 'boundary paradox' and 'learning paradox' [48].

The potential for knowledge expropriation through knowledge leakage implies that simple knowledge-sharing under cooperation can lead to win-lose or lose-lose outcomes. In such a scenario, no immediate solutions might exist for the firms under coopetition to get to positive-sum outcomes. Subject matter experts (SMEs) and

domain specialists in such firms might contemplate different pathways for generating win-win strategies. For example, one option might be for coopeting firms to engage other actors, illustrated in Sect. 3 in this paper, into their relationship to help reduce opportunities for exploitation. Another option might be for coopeting firms to jointly develop and operate knowledge-sharing systems in-house that mitigate the risks of knowledge misappropriation. Yet another option might be for the actors to change their motivations to disincentivize opportunistic behavior through rewards and penalties.

The pathway selected by SMEs in coopeting firms will depend on the specifics of their firms as well as their relationships. In the real-world, the process of generating and discriminating among such options is complex and nontrivial due to two main reasons [7]. First, the decision space of each actor is constrained or enlarged by interdependencies with potential actions of other actors. Second, trade-offs between multiple competing objectives lead to different prioritization of alternatives by each actor due to the unique preference structure of that actor.

3 Modeling Win-Win Strategies Using *i** and Game Trees

3.1 Framework with *i** and Game Trees for Modeling Win-Win Strategies

In this paper, we illustrate the use of a mediating actor to get to win-win by applying the modeling approach that is depicted in Fig. 1. This process interleaves steps from *i** and Game Tree modeling in an incremental and iterative manner. It is useful for co-developing complementary models that jointly offer greater 'interpretability' and 'explainability' than either can individually. *i** (denoting distributed intentionality) is a *goal*- and *actor*-oriented modeling language that supports strategic reasoning. The semantics and notation of *i** are explained in [8].

Game Trees are decision trees that support representation of decisions and *payoffs* associated with *actors* in a *game*. In Game Theory, a *game* refers to any social situation in which two or more *players* are involved. A *player* is an active participant in a strategic relationship with one or more *players*. A *payoff* is the *reward* (positive) or *penalty* (negative) associated with a specific course of action. A course of action is a sequence of decisions and actions undertaken by the players in a *game*. Solving a game refers to selecting a *reward* maximizing or *penalty* minimizing strategy for one or more *players*. The characteristics and features of Game Trees are described in [6].

It is noted in [7] that, "while game trees support the depiction of payoffs they do not explicitly codify the reasons for those payoffs". However, "even though the internal intentional structure of an actor cannot be expressed directly in Game Trees it can be represented via *i** Strategic Rationale (SR) diagrams" [7]. In [7, 32], *i** SR diagrams are used to represent and reason about internal intentional structures of *actors* while Game Trees are used to express and evaluate decisions and *payoffs* of those players. Therefore, "Game Trees and actor modeling with *i** can be used together to achieve a deeper understanding of the decision space as well as to secure a stronger decision rationale" [7].

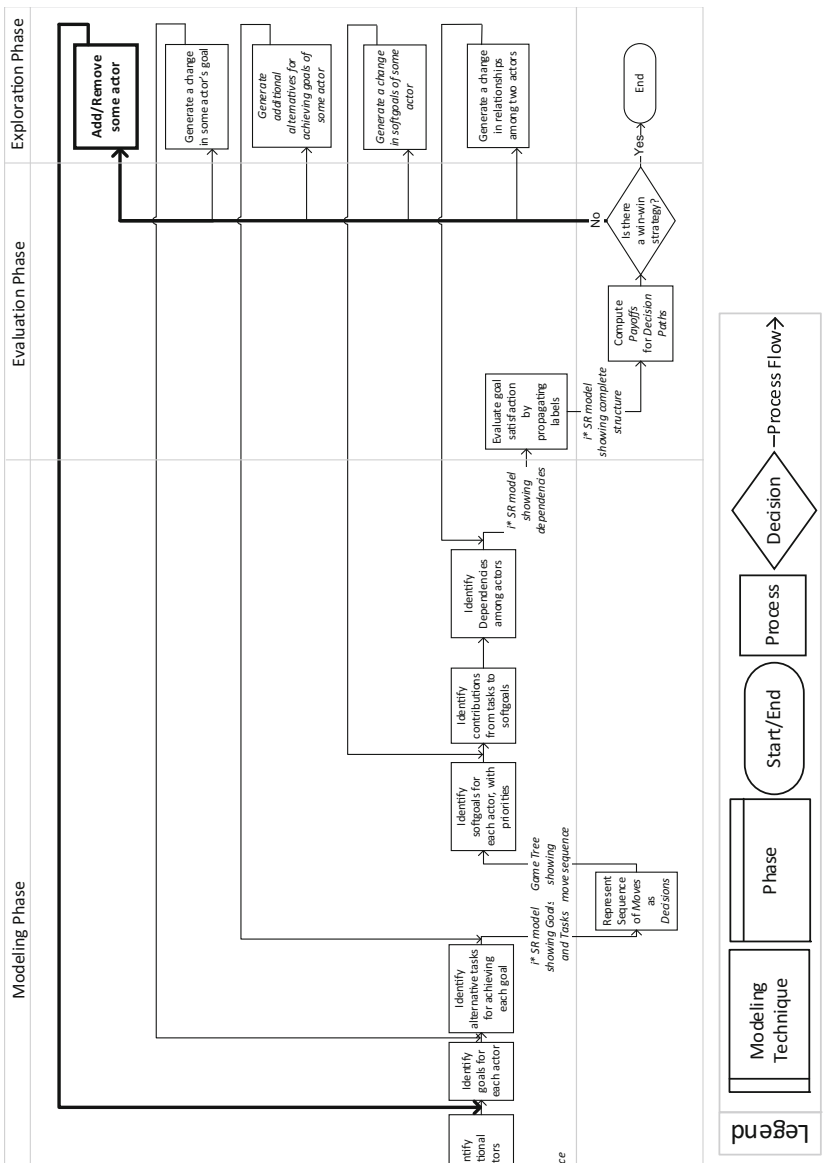


Fig. 1. Process steps for alternating between *i** and Game Tree modeling to get to win-win (Introduction of new **actor** is highlighted in **bold**, *softgoal* and *tasks* in *italics*)

This process, which is depicted in Fig. 1, comprises three phases: Modeling, Evaluation, and Exploration. In the Modeling phase, an *i** SR diagram and its corresponding Game Tree are instantiated and populated. In the Evaluation phase, the impacts of various choices on objectives are calculated to detect the presence of any extant win-win strategies. In the Exploration phase, a systematic search is performed to generate new alternatives that yield positive-sum outcomes. This process can be repeated to generate as many win-win strategies as necessary.

Modeling Phase: In this phase, strategic relationships among *actors* are modeled in terms of *goals*, *tasks*, *resources*, *softgoals*, and *dependencies* among them that are denoted in an *i** SR diagram. The sequence of decisions and *payoffs* of these *players* are codified in a Game Tree. An *actor* is an active entity that performs actions by applying its know-how to accomplish its *goals*. A *goal* is a state of affairs in the world that an *actor* wishes to achieve. *Task* is a concrete method for addressing a *softgoal* to satisfy some requirements. In *i**, *softgoals* denote quality objectives that do not have clear-cut satisfaction criteria. They are evaluated as being satisfied or denied from the subjective perspective of an *actor*.

A *task* is an activity that can be used to accomplish a *goal*. The relationship between a *goal* and its associated *tasks* is shown via *means-ends* links. A *goal* (the "end") is achieved when any of its associated *tasks* (the "means") are completed. A *task* can be decomposed into subsidiary *goals*, *tasks*, *softgoals*, and *resources*. A *resource* is a physical or informational entity that is necessary for completing a *task*. The relationship between a task and its subsidiary entities is depicted via a *task-decomposition* link. In *i**, a *depender* depends on a *dependee* for a *dependum*. A *dependum* can be a *goal* to be achieved, *task* to be completed, *softgoal* to be satisfied, or *resource* to be obtained.

Contribution links relate *tasks* to *softgoals* and *softgoals* to other *softgoals*. *Contribution links* can be of type *Help* (denoted by a green line accompanied with a plus symbol) or *Hurt* (denoted by a red line accompanied with a minus symbol). A *Help contribution link* contributes positively towards the achievement of a *softgoal*. A *Hurt contribution link* contributes negatively towards the achievement of a *softgoal*. Contributions can be intentional (denoted by a solid line) or incidental (denoted by a dashed line). Further details about i* modeling can be found in [8].

Evaluation Phase: In this phase, *Contribution links* are used to propagate and trace the impact of lower-level *tasks* and *softgoals* on higher-level *softgoals*. *Softgoals* can either be fully satisfied (denoted by a checkmark) or partially satisfied (denoted by a dot underneath a checkmark). Conversely, *softgoals* can either be fully denied (denoted by a cross) or partially denied (denoted by a dot underneath a cross).

Forward propagation of labels can be used to answer 'is this solution viable' type of questions. The process for forward propagation of satisfaction labels in goal models is explained in [27]. This process involves the iterative application of propagation rules to attach current values from each offspring to its parent and then resolving *softgoal* labels at the parent level [27]. We apply the rules for satisfaction analysis in goal models that are explained in [28, 29] in the Evaluation phase.

3.2 As-Is Scenario: Discriminating Win-Win Strategies with *i** and Game Trees

Modeling Phase: Figure 2 presents a goal model of an As-Is knowledge-sharing scenario between firms under coopetition. This goal model focuses on interdependencies among *softgoals*, and *tasks* that operationalize those *softgoals* while deferring consideration of relationships among actors. In this goal model, the nodes are *softgoals* or *tasks* while the edges are *contribution links*. Tables 1 and 2 expand on the meanings of these *softgoals* and *tasks*.

In this industry scenario, a firm has two top-level *softgoals* which are "No Leakage" of knowledge assets and "No Blocking" of knowledge transfers. "No Leakage" of knowledge assets is a *softgoal* because separate firms may judge the presence or absence of knowledge leakage differently. Similarly, "No Blocking" of knowledge transfers is another *softgoal* because different firms may use dissimilar criteria to determine whether or not knowledge-sharing is being blocked.

A firm can adopt a Strict knowledge-sharing policy or a Permissive knowledge-sharing policy. A Strict policy prioritizes minimization of knowledge-leakage over circumvention of knowledge-blocking. Conversely, a Permissive policy treats avoidance of knowledge-blocking with greater importance than prevention of knowledge-leakage.

In the knowledge sharing setting considered here, the same goal model applies equally to all sharing parties. In other settings, a separate goal model may be needed to represent the perspective of each actor.

Softgoals are operationalized by *tasks* (bottom of Fig. 2). For example, "Processing" involves generating machine-readable metadata for each knowledge asset. This makes it easier to distinguish among individual knowledge assets such as on the basis of their ownership. Therefore, "Processing" is a *task* that operationalizes the *softgoal* "Annotatable" asset ownership. Similarly, "Integrating" involves mixing together knowledge assets from various partners. This makes it simpler for each firm to avail of the knowledge of their partners. Therefore, "Integrating" operationalizes the *softgoal* "Available" partner assets.

In this example, we use the notation wherein the inclusion of a *task* in a Strict or Permissive policy is inscribed within each *task*. A circle inscribed with an S and a numerical identifier in the top left corner of a *task* denotes the inclusion of that *task* in a Strict policy. A square inscribed with a P and a numerical identifier in the top right corner denotes the inclusion of that *task* in a Permissive policy. For example, "Auditing" of knowledge transfers is a part of a Strict policy and "Integrating" of partner assets is a part of a Permissive policy.

A *task* can also be included simultaneously in Strict and Permissive policies while being implemented differently in each policy type. For instance, "Modularizing" the boundary of a knowledge asset is part of both Permissive as well as Strict policies even though modularization may be implemented differently in Strict and Permissive policies. It should be noted that these inscriptions (i.e., S with identifier in circle on top left of *task* and P with identifier in circle on top right of *task*) are specific to this example.

Table 1. *Softgoal* types and topics in As-Is scenario in Fig. 2

Softgoal type [Topic]	Description of *softgoal*
No Leakage [Knowledge Assets]	Assets should not be misappropriated by partners [10, 11]
No Blocking [Knowledge Transfers]	Transfers should be seamless and frictionless [42, 43]
Synergetic [Knowledge Assets]	Assets should be more valuable jointly than individually [21, 22]
Leveragability [Knowledge Assets]	Assets should be useful and usable to generate benefits [21, 22]
No Negative Cross Impact [A. Val.]	Sharing with partner should not reduce value of asset for self [21, 22]
Interdependence [Bus. Partners]	Sharing should take place among co-dependent partners [22]
Complementarity [Partner Assets]	Partner assets should enhance each other's asset value [23]
Transferability [Knowledge Assets]	Assets should be distributable to partners [40]
Appropriability [Knowledge Assets]	Assets should be receivable by partners [14]
Irreducible [Asset Value]	Benefits from asset should be indestructible and renewable [45]
Protectable [Knowledge Assets]	Assets should be containable and isolatable [41]
Mutuality [Partner Assets]	Sharing should encompass assets that are inter-reliant [44]
Annotatable [Asset Ownership]	Identity of the owner of each asset should be discernible [48]
Combinable [Partner Assets]	Assets should be integrable with other assets [50]
Compatible [Knowledge Assets]	Assets should function normally in conjunction with other assets [46]
Available [Partner Assets]	Assets should be easily reachable when needed [47]
Absorbable [Partner Assets]	Assets should be easily consumable when needed [14]
Dynamic [Knowledge Assets]	Content and functionality of asset should be changeable [45]
Concealable [Asset Content]	Asset contents should be capable of being hidden from partners [48]
Licensable [Knowledge Assets]	Assets should support deactivation and decommissioning [49]

Goal models aid in detecting and analyzing tradeoffs that exist among different *softgoals*. The goal model in Fig. 2 shows that various *tasks* impact *softgoals* differently. For instance, "Posting" a knowledge asset into an asset directory *Helps* to make that knowledge asset more "Combinable" (i.e., easier to integrate) with other knowledge assets. Conversely, "Modifying" the behavior of a knowledge asset can make it less "Compatible" with knowledge assets with which it is already interoperable (i.e., *Hurts* link). Specific combinations of *tasks* within a Strict or Permissive policy can also

Table 2. *Task* types and topics in As-Is scenario in Fig. 2

Task type [Topic]	Policy	Description of *task*
Auditing [Knowledge Transfers]	S	Reviewing actions performed by users and processes [48]
Processing [Asset Metadata]	S	Generating machine-readable metadata for each asset [54]
Exposing [Asset Interface]	P	Registering input and output parameters of an asset [51]
Documenting [Asset Schema]	P	Explaining types of entities and relationships in an asset [40]
Integrating [Partner Assets]	P	Commingling content from disparate partner assets [53]
Publishing [Asset Directory]	P	Advertising sharing of an asset via a repository [51]
Modifying [Asset Behavior]	S	Reprogramming the content and functionality of an asset [56]
Modularizing [Asset Boundary]	S, P	Setting perimeter of each asset specifying its scope [55]
Reconfiguring [Knwldg. Assets]	S	Asset should be packagable in many ways [52]

impact softgoals differently. For instance, "Auditing" is a *task* that is part of a Strict policy and operationalizes the *softgoal* "Mutuality" of partner assets. It also *Helps* the *softgoal* "Licensable" knowledge assets. Similarly, "Reconfiguring" of knowledge assets is a *task* that is also part of a Strict policy and operationalizes the *softgoal* "Licensable" knowledge assets. This *softgoal* "Licensable" knowledge assets is considered to be satisfied in a Strict policy since multiple *tasks* that are part of a Strict policy make positive contributions to it. Conversely, the *softgoal* "Dynamic" knowledge assets is only partially satisfied in a Strict policy due to the conflicting interaction of two *tasks* which are part of a Strict policy. These are "Modifying" asset behaviour and "Processing" asset metadata. While "Modifying" asset behavior operationalizes the *softgoal* "Dynamic" knowledge assets this *softgoal* is *Hurt* by "Processing" asset metadata.

In the real world, each actor assesses such trade-offs between *softgoals* in line with its preferences and prioritizes those *softgoals* differently depending on its proclivities. The goal model in Fig. 2 is instantiated in Fig. 3 to demonstrate this with respect to two *actors* in a coopetitive relationship. Figure 3 depicts co-developed *i** SR diagram and Game Tree of the As-Is scenario pertaining to two business partners in the pharmaceutical industry.

In this *i** diagram, Branded Pharmaceutical Company (BPC) and Generic Pharmaceutical Compounder (GPC) are two *actors*. BPC develops and markets prescription medicines based on its R&D initiatives as well as its protected intellectual property (IP) (not shown[1]). GPC manufactures ingredients that are used in BPC's medicines and

[1] In this instance, and in the remainder of this paper, certain aspects of the relationship between actors are not shown due to page limitations.

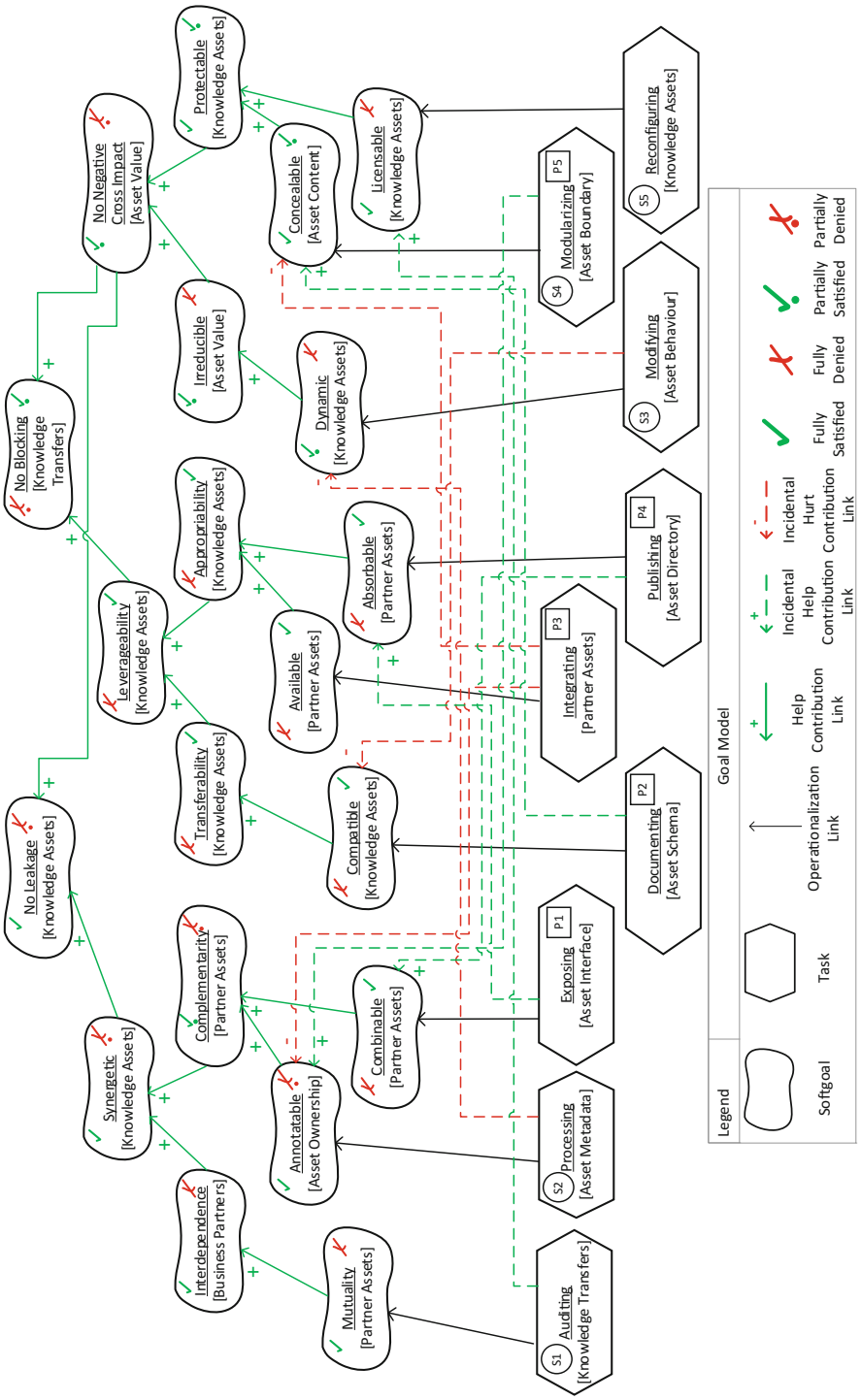
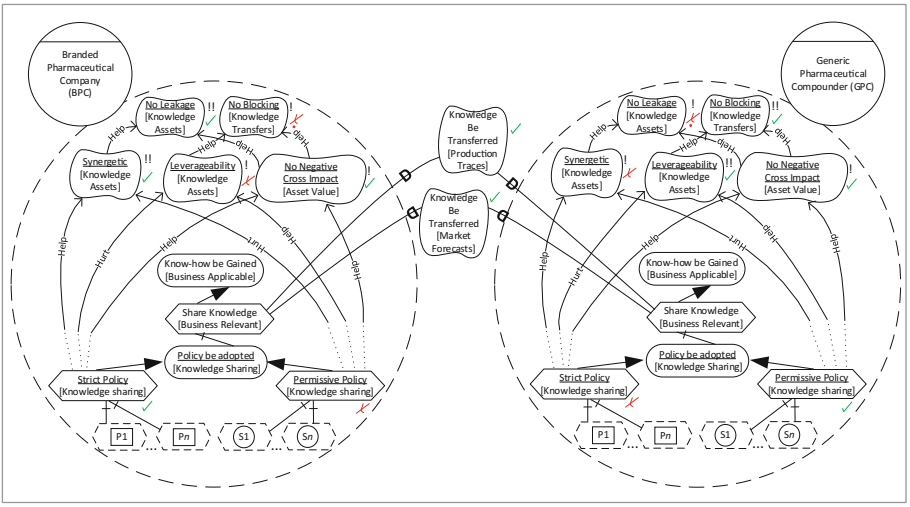
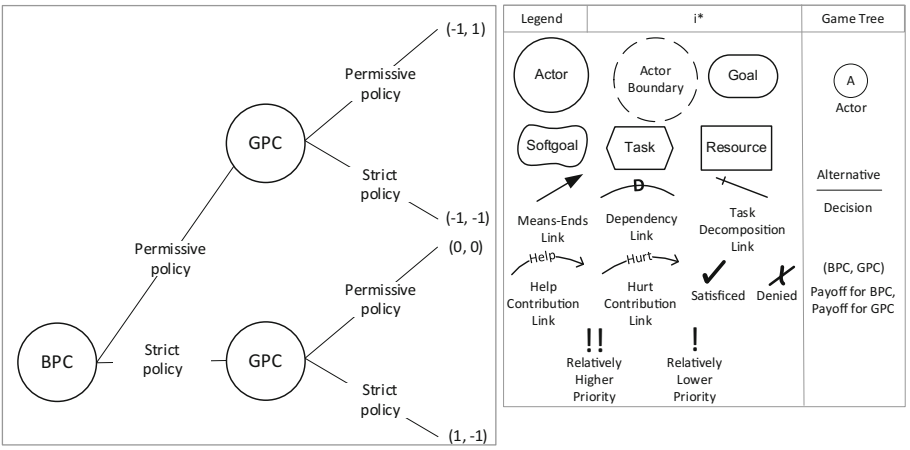


Fig. 2. Goal model of As-Is scenario representing knowledge sharing goals and potential tasks, synthesized from [21–26] – the As-Is scenario



(i) *i** SR diagram showing strategic goals of the two coopeting *actors*



(ii) Game Tree showing *payoffs* from possible moves by BPC followed by GPC moves

Fig. 3. As-Is scenario

produces medicines for BPC that BPC sells in the market (not shown, see footnote 1). GPC also sells generic medicines that are analogous to the prescription medicines sold by BPC only if their IP is not protected (not shown, see footnote 1).

The two *actors* depend on each other to meet their respective *goals* pertaining to "Know-how be Gained". GPC depends on "Market Forecasts" of BPC (shown) so that *GPC* can approximate the upcoming requirements of BPC (not shown, see footnote 1). This helps *GPC* to plan its production runs based on medicines that *BPC* will likely contract *GPC* to produce (not shown, see footnote 1). BPC depends on the "Production Traces of GPC" (shown) to verify that GPC is only manufacturing those quantities of ingredients of BPC's high margin medicines that are ordered by BPC (not shown, see

footnote 1). This helps BPC to verify that GPC is not manufacturing extra quantities of those ingredients to produce substitute medicines that GPC can sell by itself (not shown, see footnote 1).

Dependencies among BPC and GPC are shown as *softgoals* because each is satisficed from the perspective of the *depender*. Both *actors* can achieve their respective *goals* of "Know-how be Gained" by performing the *task* "Share Knowledge". Knowledge sharing *"Policy be adopted"* is a sub-*goal* of this *task* "Share Knowledge". This sub-*goal* is associated with two *tasks* which pertain to the adoption of either a "Strict" or a "Permissive" knowledge sharing policy. The *tasks* labeled "Strict Policy" and "Permissive Policy" for knowledge sharing in Fig. 3 map to the set of *tasks* in Fig. 2 with the inscriptions of S and P respectively. This is shown in Fig. 3 via the decomposition of two *tasks*, which are "Strict Policy" and "Permissive Policy", into their respective sub-tasks, which are denoted by "$P_1 \ldots P_n$" and "$S_1 \ldots S_n$". Contributions from the *tasks* labeled "Strict Policy" and "Permissive Policy" to *softgoals* labeled "Synergetic" knowledge assets, "Leveregeability" of knowledge assets, and "No negative-cross impact" of asset value are depicted indirectly via a partially dotted *contribution link*. This is done to hide the full intentional structure in the *i** SR diagram since the complete goal model in Fig. 2 contains these details.

Potential benefits from knowledge sharing serve as incentives for BPC and GPC to adopt "Permissive" policies. However, the countervailing threat of opportunism serve as motivations for BPC and GPC to adopt "Strict" policies. Since BPC and GPC are autonomous actors they are free to select either Permissive or Strict policy in line with their preferences and proclivities. In this example, as shown in Fig. 3, BPC prioritizes a "Strict" policy over a "Permissive" policy while GPC prioritizes a "Permissive" policy over a "Strict" policy. The selection of one policy over another in the real-world is likely to be the result of deliberation and contemplation by subject matter experts (SMEs) and domain specialists. This modeling approach complements and supplements their reasoning and analysis rather than substitute or obviate it.

Evaluation Phase: In the Evaluation phase, *payoffs* in the Game Tree are estimated by analyzing *softgoal* satisfaction in the *i** SR diagram. A preliminary analysis of *softgoal* satisfaction in the goal model in Fig. 2 reveals that neither Strict nor Permissive knowledge-sharing policies satisfy all top-level *softgoals* in the As-Is scenario. The *i** SR diagram in Fig. 3 shows that neither BPC nor GPC satisfy every *softgoal* through their chosen policies. For example, BPC is not able to satisfy one of its top-level *softgoals* of "No Blocking" of knowledge transfers by choosing a Strict policy while GPC is not able to satisfy one of its top-level *softgoals* of "No Leakage" of knowledge assets by choosing a Permissive policy. The *i** SR diagram in Fig. 3 can be used to calculate the relative *payoffs* for these *players* in the Game Tree.

On the Game Tree, in the first case, BPC and GPC select Permissive policies. Since GPC prioritizes a top-level *softgoal* that is satisfied when this type of policy is chosen then it earns a *payoff* of 1. However, BPC prioritizes a top-level *softgoal* that is denied when this policy is chosen then it earns a *payoff* of −1. In the second case, BPC selects a "Permissive" policy but GPC selects a "Strict" policy. In this case neither BPC nor GPC achieve their higher priority top-level *softgoals* and thus both earn payoffs of −1. In the third case, BPC selects a "Strict" policy but GPC selects a "Permissive"

policy. In this case while both BPC and GPC satisfy their higher priority top-level *softgoals* they do not satisfy some of their, albeit lower priority, *softgoals*. Thus, both earn payoffs of 0. In the fourth case, BPC and GPC select "Strict" policies. Since BPC prioritizes a top-level *softgoal* that is satisfied when this type of policy is chosen then it earns a *payoff* of 1. However, GPC prioritizes a top-level *softgoal* that is denied when this type of policy is chosen then it earns a *payoff* of −1.

These *payoffs* in the Game Tree can be used to detect the presence of any positive-sum outcomes. In the As-Is scenario, there are no win-win strategies since neither "Permissive" nor "Strict" policies allow BPC and GPC to satisfy each of their top level *softgoals*. This motivates their systematic search for new alternatives to generate positive-sum outcomes.

3.3 To-Be Scenario: Generating Win-Win Strategies with *i** and Game Trees

Exploration Phase: In the Exploration phase, an SME can pursue any of five non-deterministic lines of action incrementally and iteratively. As depicted in Fig. 1, they can add/remove some *actor*, generate additional alternatives for achieving *goals* of some *actor*, generate a change in relationships among some *actors*, generate a change in *softgoals* of some *actor*, or generate a change in some *actor's goals*. For example, as shown in the goal model in Fig. 4, new *softgoals* and *tasks* can be introduced that favorably impact (i.e., *Help*) top-level *softgoals*. These new *softgoals* and *tasks* can be used to satisfy previously denied top-level *softgoals*.

Figure 4 is a goal model of a hypothetical To-Be knowledge-sharing scenario between businesses under coopetition. Model elements, from the As-Is scenario in Fig. 2, that are unimpacted by new *softgoals* and *tasks* in Fig. 4 are greyed-out. This improves the presentation of the goal model to highlight the To-Be scenario. New *softgoals* and *tasks* in Fig. 4 are shown in blue color while existing *softgoals* that are impacted by new *softgoals* and *tasks* are shown in black color. New *contribution links* are shown in green (Help) and red (Hurt) colors while existing contribution links are greyed-out. We anticipate that, with tool support in the future, one would be able to collapse or expand portions of the model to hide or reveal details as necessary.

Loops in the process depicted in Fig. 1 indicate that any step in the Exploration phase of this modeling approach can trigger other steps. For example, in the pursuit of a win-win strategy, an SME may decide to generate new *tasks* to improve overall satisfaction of top-level *softgoals*. These new *tasks,* depicted in Fig. 4, may trigger the generation of new *softgoals*. Collectively, these additional tasks and softgoals represent new system requirements that expand the set of existing system requirements depicted in Fig. 2.

Tables 3 and 4 describe these new *softgoals* and *tasks*. However, their sources are not listed due to constraints on paper length. These new requirements can be fulfilled by performing certain activities in-house (i.e., generate additional alternatives for achieving goals of some actor). Alternatively, they can be fulfilled by including a new actor into the existing relationship (i.e., add/remove some actor). If needed, the pros and cons of each option in the Exploration phase can also be modeled with *i** separately.

Table 3. Softgoal types and topics and topics in To-Be scenario in Fig. 4

Softgoal type [Topic]	Description of *softgoal*
Balanced [Asset Sharing]	Quantity of contents transferred should be equal among partners
Reportable [Asset Sharing]	Quantity and quality of contents transferred should be auditable
Compliant [Knowledge Assets]	Format of assets should be consistent with third-party specifications
Redundant [Knowledge Assets]	Copies of assets should be stored for safeguarding

Table 4. Task types and topics in To-Be scenario in Fig. 4

Task type [Topic]	Policy	Description of *task*
Metering [Knowledge Transfers]	S	Measuring quantity of transfers between partners
External Tracking [Knowledge Transfers]	P	Surveilling content in transfers between partners
Canonical Template [Knowledge Model]	S	Establishing uniform format to be used by partners
Certifying [Asset Specification]	P	Attesting system specification by standards organization
Replicating [Knowledge Assets]	S, P	Creating multiple copies of asset

Evaluation Phase: The *i** SR diagram in Fig. 5 can be used to calculate the relative *payoffs* for the *players* in the Game Tree. In the first case, BPC and GPC select Permissive policies. Since all top-level *softgoals* of GPC are satisfied and it acts in accordance with its preference (i.e., adopts Permissive policy) then it earns a *payoff* of 2. Each top-level *softgoal* of BPC is also satisfied in this case but since it does not act in line with its preference (i.e., does not adopt Strict policy) then it earns a *payoff* of 1. In the second case, BPC selects a Permissive policy but GPC selects a Strict policy. In this case both BPC nor GPC achieve their higher priority top-level *softgoals* but neither acts according to their preferences and thus both earn payoffs of 1. In the third case, BPC selects a Strict policy but GPC selects a Permissive policy. In this case both BPC and GPC satisfy each of their higher priority top-level *softgoals* and act according to their preferences. Therefore, both earn payoffs of 2. In the fourth case, BPC and GPC select Strict policies. Since all top-level *softgoals* of BPC are satisfied and it acts in accordance with its preference (i.e., adopts Strict policy) then it earns a *payoff* of 2. Each top-level *softgoal* of BPC is also satisfied in this case but it does not act in line with its preferences (i.e., does not adopt Permissive policy) then it earns a *payoff* of 1.

The *i** SR diagram of the To-Be scenario shows that all the top-level *softgoals* of BPC and GPC are satisfied. This is due to their addition of new *softgoals* and *tasks* as well as the introduction of a new *actor*, which is IDS. Therefore, the *payoffs* associated with the To-Be scenario in the Game Tree reflect higher values than their corresponding

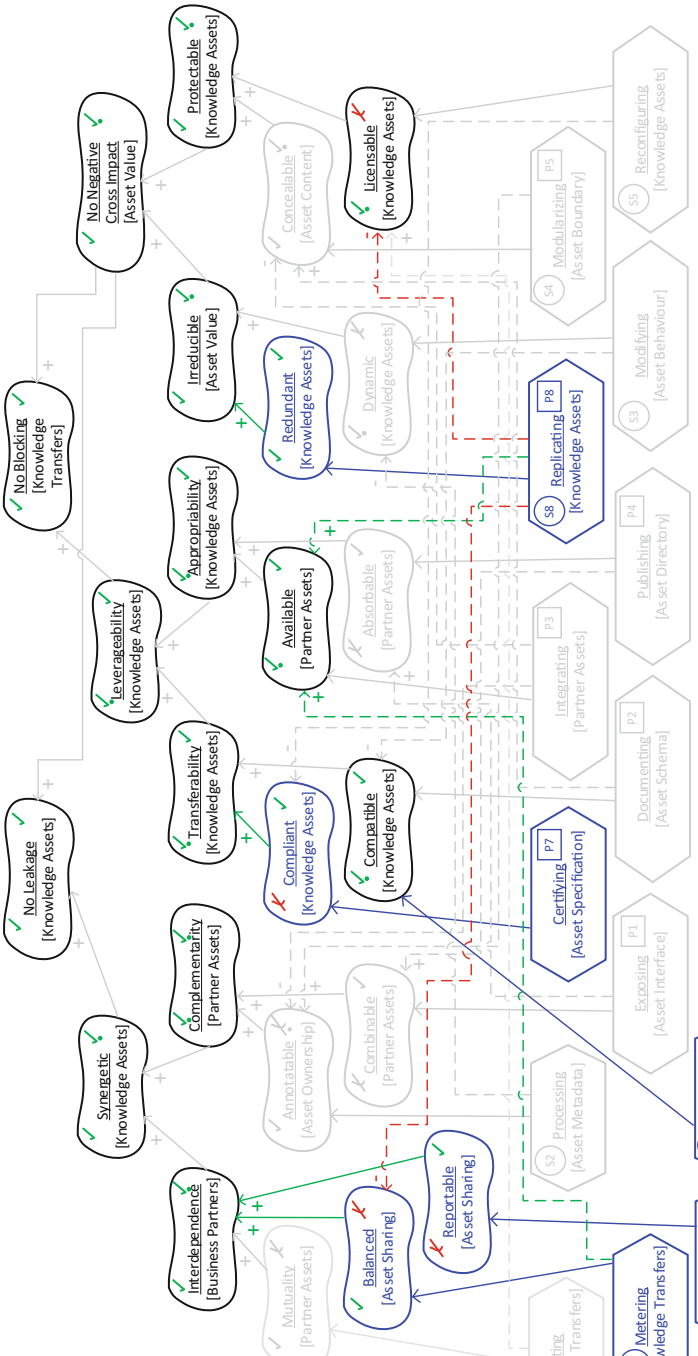
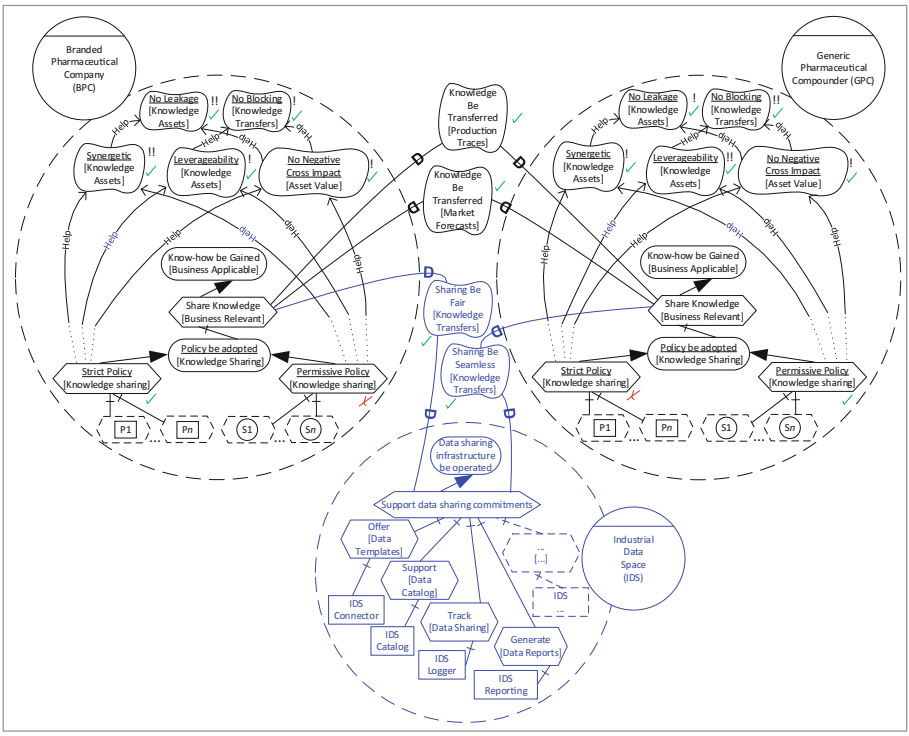
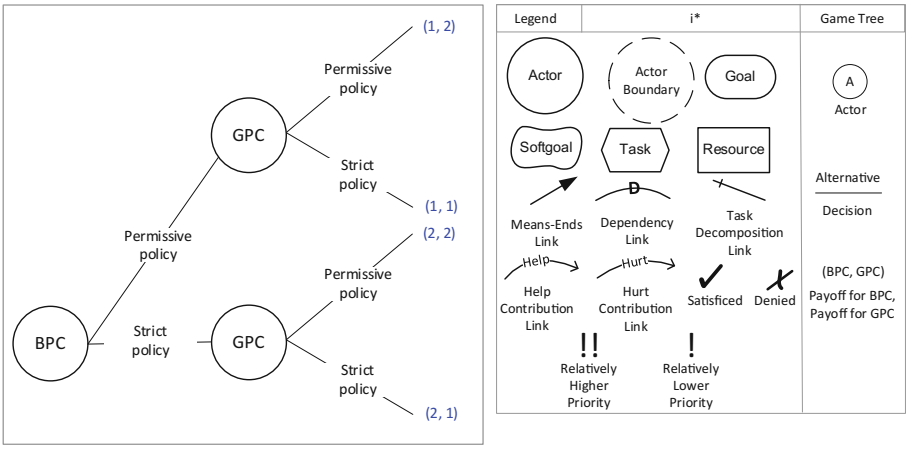


Fig. 4. Goal model of To-Be scenario representing knowledge sharing goals and potential tasks, synthesized from [21–26] – the To-Be scenario (Color figure online)



(i) *i** SR diagram showing strategic goals of the two coopeting *actors* and a mediating *actor*



(ii) Game Tree showing *payoffs* from possible moves by BPC followed by GPC moves

Fig. 5. To-Be scenario

options in the As-Is scenario. Following the process described in Fig. 1 shows that multiple win-win strategies can be created in an industrial collaboration scenario where none existed originally. A comparison of Figs. 2 and 3 with Figs. 4 and 5 highlights a

primary benefit of using this approach to co-develop *i** SR diagrams with Game Trees. In Figs. 3 and 5, the Game Trees are structurally similar but have different *payoffs* and, in the *i** SR diagrams, the internal intentional structure of BPC and GPC is identical except for certain *contribution links*. Figures 2 and 4 are crucial for understanding the reasons for these differences. The goal models in Figs. 2 and 4 explain the reasons for the differences in the *payoffs* on the Game Trees and the changes in the *contribution links* within the *i** SR diagrams.

4 Related Work

This paper contributes to the body of knowledge pertaining to intentional modeling of coopetition. Majority of the research on coopetition modeling has focused on game-theoretic approaches [4]. Such approaches encode the intentionality of the *players* within the *payoffs* thereby eliding their goal structures. Recent research in the enterprise modeling literature has focused on the intentionality of actors engaged in strategic coopetition. Requirements for enterprise modeling of strategic coopetition are described in [4, 33]. The intentional modeling approach that is applied in this paper was introduced in [7] and refined in [32].

In [32] a basic example of cake-cutting is presented to demonstrate the application of this process. That example of cake-cutting is drawn from game theory and is used to demonstrate the co-design and co-evolution of *i** SR diagrams and their corresponding Game Trees. That example shows the introduction of a new alternative in an ultimatum game between two *players* to generate a new win-win strategy when originally none existed. That pathway to win-win is further illustrated with a case of coopetition between software ecosystems of Apple and Adobe. Modeling of complementarity, which is a motivator of coopetition, and relevant in knowledge-sharing scenarios, is discussed in [34]. More broadly, this research paper also contributes to the scholarly literature on enterprise modeling of business strategy. Researchers in this domain have developed modeling techniques that incorporate strategic management concepts [35–39].

5 Conclusions and Future Work

We utilized a strategic modeling approach to systematically search for win-win strategies and generate new alternatives for organizations under coopetition. This integrative approach incrementally and iteratively elaborated and refined the *i** SR diagram and its corresponding Game Tree. No win-win strategies were detected in the As-Is scenario due to threats related to knowledge leakage and knowledge blocking. However, in the To-Be scenario, multiple win-win strategies were generated by applying this strategic modeling approach to the As-Is scenario. New *softgoals* and *tasks* were added that obviated the threats from knowledge leakage and knowledge blocking. These *softgoals* and *tasks* could be satisfied by the *actors* by themselves (e.g., by building a system that meets necessary requirements) or with the help of another

actor (e.g., by subscribing to a service that meets necessary requirements). In this paper we depicted the latter option.

This strategic modeling approach incorporates three practical and reasonable assumptions to ensure its usefulness in real-world applications [7]. However, the efficacy and viability of these assumptions needs to be tested via empirical investigation. Our future work will comprise achievement of three objectives that must be satisfied to encourage mainstream adoption of these models by industry professionals. Firstly, these models may need to be simplified to gain broader acceptance by practitioners. This would be done by developing collaboration patterns that represent common behaviors in the real world (e.g., collaborating to avoid common threat). Secondly, these models may need to support more sophisticated and nuanced methods for calculating payoffs. Game Theorists have proposed many methods for calculating payoffs under different circumstances and these methods could be supported by these models. Thirdly, these models may need to be commingled with existing processes that are used by organizations to manage coopetitive relationships. For example, organizations use contracts and legal agreements to set the terms and conditions of such relationships. These models could be used to support the contract negotiation and agreement formation processes. These areas of future work shall increase the value and utility of these models in the industry.

References

1. Brandenburger, A.M., Nalebuff, B.J.: The right game: use game theory to shape strategy. Harv. Bus. Rev. **73**(4), 57–71 (1995)
2. Castaldo, S., Dagnino, G.B.: Trust and coopetition: the strategic role of trust in interfirm coopetitive dynamics. In: Dagnino, G.B., Rocco, E. (eds.) Coopetition Strategy: Theory, Experiments and Cases, pp. 74–100. Routledge, New York (2009)
3. Gnyawali, D.R., Park, B.: Co-opetition and technological innovation in small and medium-sized enterprises: a multilevel conceptual model. J. Small Bus. Manag. **47**(3), 308–330 (2009)
4. Pant, V., Yu, E.: Modeling simultaneous cooperation and competition among enterprises. Bus. Inf. Syst. Eng. **60**(1), 39–54 (2018)
5. Brandenburger, A.M., Nalebuff, B.J.: Co-opetition. Doubleday, New York (1996)
6. Dixit, A.K., Nalebuff, B.: The Art of Strategy: A Game Theorist's Guide to Success in Business & Life. WW Norton & Company, New York (2008)
7. Pant, V., Yu, E.: Understanding strategic moves and reciprocity on Software Ecosystems: a strategic modeling approach. In: 9th International Workshop on Software Ecosystems (IWSECO 2017) (2017)
8. Yu, E., Giorgini, P., Maiden, N., Mylopoulos, J.: Social Modeling for Requirements Engineering. MIT Press, Cambridge (2011)
9. Baglieri, D., Carfì, D., Dagnino, G.B.: Asymmetric R&D alliances in the biopharmaceutical industry: a multi-stage coopetitive approach. Int. Stud. Manag. Organ. **46**(2–3), 179–201 (2016)
10. Lowman, M., Trott, P., Hoecht, A., Sellam, Z.: Innovation risks of outsourcing in pharmaceutical new product development. Technovation **32**(2), 99–109 (2012)
11. Diestre, L., Rajagopalan, N.: Are all 'sharks' dangerous? New biotechnology ventures and partner selection in R&D alliances. Strateg. Manag. J. **33**(10), 1115–1134 (2012)

12. Jiang, X., Li, M., Gao, S., Bao, Y., Jiang, F.: Managing knowledge leakage in strategic alliances: the effects of trust and formal contracts. Ind. Mark. Manag. **42**(6), 983–991 (2013)
13. Heiman, B.A., Nickerson, J.A.: Empirical evidence regarding the tension between knowledge sharing and knowledge expropriation in collaborations. Manag. Decis. Econ. **25**(6–7), 401–420 (2004)
14. Ritala, P., Olander, H., Michailova, S., Husted, K.: Knowledge sharing, knowledge leaking and relative innovation performance: an empirical study. Technovation **35**, 22–31 (2015)
15. Kale, P., Singh, H., Perlmutter, H.: Learning and protection of proprietary assets in strategic alliances: building relational capital. Strateg. Manag. J. **21**, 217–237 (2000)
16. Khanna, T., Gulati, R., Nohria, N.: The dynamics of learning alliances: competition, cooperation, and relative scope. Strateg. Manag. J. **19**, 193–210 (1998)
17. Carayannis, E.G., Alexander, J., Ioannidis, A.: Leveraging knowledge, learning, and innovation in forming strategic government–university–industry (GUI) R&D partnerships in the US, Germany, and France. Technovation **20**(9), 477–488 (2000)
18. Petts, N.: Building growth on core competences—a practical approach. Long Range Plan. **30** (4), 551–561 (1997)
19. Jashapara, A.: The competitive learning organization: a quest for the Holy Grail. Manag. Decis. **31**(8) (1993)
20. Jashapara, A.: Cognition, culture and competition: an empirical test of the learning organization. Learn. Organ. **10**(1) (2003)
21. Loebbecke, C., van Fenema, P.C.: Interorganizational knowledge sharing during coopetition. In: European Conference on Information Systems, pp. 1632–1639 (1998)
22. Loebecke, C., Van Fenema, P.C., Powell, P.: Co-opetition and knowledge transfer. ACM Goal modelMIS Database: DATABASE Adv. Inf. Syst. **30**(2), 14–25 (1999)
23. Ghobadi, S., D'Ambra, J.: Coopetitive knowledge sharing: an analytical review of literature. Electron. J. Knowl. Manag. **9**(4) (2011)
24. Yoong, P., Molina, M.: Knowledge sharing and business clusters. In: 7th Pacific Asia Conference on Information Systems (2003)
25. Lertpittayapoom, N., Paul, S., Mykytyn Jr., P.: A theoretical perspective on effective interorganizational knowledge. In: 40th Annual Hawaii International Conference on System Sciences, p. 187b. IEEE (2007)
26. Wilson, T.D.: Information sharing: an exploration of the literature and some propositions. Inf. Res.: Int. Electron. J **15**(4) (2010)
27. Horkoff, J., Yu, E.: Comparison and evaluation of goal-oriented satisfaction analysis techniques. Requir. Eng. **18**(3), 199–222 (2013)
28. Horkoff, J., Yu, E.: Analyzing goal models: different approaches and how to choose among them. In: Proceedings of the 2011 ACM Symposium on Applied Computing, pp. 675–682. ACM (2011)
29. Otto, B., et al.: Industrial Data Space: Digital Sovereignty Over Data. Fraunhofer-Gesellschaft, München (2016)
30. Jarke, M.: Data spaces: combining goal-driven and data-driven approaches in community decision and negotiation support. In: Schoop, M., Kilgour, D.M. (eds.) GDN 2017. LNBIP, vol. 293, pp. 3–14. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63546-0_1
31. Quix, C., Chakrabarti, A., Kleff, S., Pullmann, J.: Business process modelling for a data exchange platform. In: Proceedings of the CAiSE'17 Forum, at the 29th International Conference on Advanced Information Systems Engineering (CAiSE 2017) (2017)
32. Pant, V., Yu, E.: Generating win-win strategies for software businesses under coopetition: a strategic modeling approach. In: International Conference of Software Business. Springer, Cham (2018)

33. Pant, V., Yu, E.: Coopetition with frenemies: towards modeling of simultaneous cooperation and competition among enterprises. In: Horkoff, J., Jeusfeld, M.A., Persson, A. (eds.) PoEM 2016. LNBIP, vol. 267, pp. 164–178. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48393-1_12
34. Pant, V., Yu, E.: Modeling strategic complementarity and synergistic value creation in coopetitive relationships. In: Ojala, A., Holmström Olsson, H., Werder, K. (eds.) ICSOB 2017. LNBIP, vol. 304, pp. 82–98. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69191-6_6
35. Giannoulis, C., Zikra, I., Bergholtz, M., Zdravkovic, J., Stirna, J., Johannesson, P.: A comparative analysis of enterprise modeling approaches for modeling business strategy. In: 6th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling (PoEM 2013), Riga, Latvia, 6–7 November 2013, pp. 193–204 (2013)
36. López, L.C., Franch, X.G.: Applying business strategy models in organizations. In: Proceedings of the 7th International i* Workshop 2014, Thessaloniki, Greece, 16–17 June 2014 (2014)
37. Bleistein, S.J., Cox, K., Verner, J.: Modeling business strategy in e-business systems requirements engineering. In: Wang, S., et al. (eds.) ER 2004. LNCS, vol. 3289, pp. 617–628. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30466-1_57
38. Giannoulis, C., Petit, M., Zdravkovic, J.: Towards a unified business strategy language: a meta-model of strategy maps. In: van Bommel, P., Hoppenbrouwers, S., Overbeek, S., Proper, E., Barjis, J. (eds.) PoEM 2010. LNBIP, vol. 68, pp. 205–216. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16782-9_15
39. Osterwalder, A., Pigneur, Y., Tucci, C.L.: Clarifying business models: Origins, present, and future of the concept. Commun. Assoc. Inf. Syst. **16**(1), 1 (2005)
40. Renzl, B.: Trust in management and knowledge sharing: the mediating effects of fear and knowledge documentation. Omega **36**(2), 206–220 (2008)
41. Levy, M., Loebbecke, C., Powell, P.: SMEs, coopetition and knowledge sharing: the IS role. In: Proceedings of the European Conference on Information Systems (2001)
42. Cai, S., Goh, M., de Souza, R., Li, G.: Knowledge sharing in collaborative supply chains: twin effects of trust and power. Int. J. Prod. Res. **51**(7), 2060–2076 (2013)
43. Durst, S., Runar Edvardsson, I.: Knowledge management in SMEs: a literature review. J. Knowl. Manag. **16**(6), 879–903 (2012)
44. Van Baalen, P., Bloemhof-Ruwaard, J., Van Heck, E.: Knowledge sharing in an emerging network of practice: the role of a knowledge portal. Eur. Manag. J. **23**(3), 300–314 (2005)
45. Yeşil, S., Koska, A., Büyükbeşe, T.: Knowledge sharing process, innovation capability and innovation performance: an empirical study. Procedia-Soc. Behav. Sci. **75**, 217–225 (2013)
46. Ho, H., Ganesan, S.: Does knowledge base compatibility help or hurt knowledge sharing between suppliers in coopetition? The role of customer participation. J. Market. **77**(6), 91–107 (2013)
47. Du Plessis, M.: The role of knowledge management in innovation. J. Knowl. Manag. **11**(4), 20–29 (2007)
48. Manhart, M., Thalmann, S.: Protecting organizational knowledge: a structured literature review. J. Knowl. Manag. **19**(2), 190–211 (2015)
49. Bogers, M.: The open innovation paradox: knowledge sharing and protection in R&D collaborations. Eur. J. Innov. Manag. **14**(1), 93–117 (2011)
50. Cricelli, L., Grimaldi, M.: Knowledge-based inter-organizational collaborations. J. Knowl. Manag. **14**(3) (2010)
51. Mentzas, G., Kafentzis, K., Georgolios, P.: Knowledge services on the semantic web. Commun. ACM **50**(10), 53–58 (2007)

52. Ye, H.J., Kankanhalli, A.: Leveraging structural holes for innovation: the moderating effects of IT-enabled absorptive capacity. In: Proceedings of the 15th Pacific Asia Conference on Information Systems (2011)
53. Teo, T.S., Bhattacherjee, A.: Knowledge transfer and utilization in IT outsourcing partnerships: a preliminary model of antecedents and outcomes. Inf. Manag. **51**(2), 177–186 (2014)
54. Leung, N.K., Lau, S.K., Tsang, N.: An ontology-based collaborative interorganizational knowledge management network (Cik-Net). In: Proceedings of the 16th Pacific Asia Conference on Information Systems (2012)
55. von der Oelsnitz, D., Lorenz, M.: Interorganizational learning between convergence and cospecialization – a knowledge-based typology of strategic alliances. In: Kersten, W., Wittmann, J. (eds.) Kompetenz, Interdisziplinarität und Komplexität in der Betriebswirtschaftslehre, pp. 243–258. Springer, Wiesbaden (2013). https://doi.org/10.1007/978-3-658-03462-7_18
56. Choi, S.G., Johanson, J.: Knowledge translation through expatriates in international knowledge transfer. Int. Bus. Rev. **21**(6), 1148–1157 (2012)

Organizational Transformation for Virtual Team Integration – A Technological Perspective

Birgit Großer(✉) and Ulrike Baumöl

FernUniversität Hagen, Hagen, Germany
{birgit.grosser, ulrike.baumoel}@fernuni-hagen.de

Abstract. The ongoing digitalization bears chances and challenges for our everyday lives, including how we work, e.g. by enhancing mobility. Deploying virtual teams as workforce is a way to meet some of these chances and challenges. The performance of virtual teams depends on organizational structures and team integration, e.g. composition of the team, business goals, and technological solutions. Regarding technology choice it is essential to understand which solutions support which tasks of virtual teamwork and should be chosen accordingly. The choice of the right technology can be supported by using the media synchronicity theory and carrying out a structured analysis of the embedding of the virtual teams in the organization, considering all organizational elements, e.g. context, governance and corporate culture. Therefore, we analyze technology along the media synchronicity theory for the integration of virtual teams, taking a holistic view on these organizational elements. The results contribute to the state of the art by proposing a matching of technology on organizational elements and provide insights for further research interests as well as for technology choice in practice. The matching shows that the functional, cultural, governance related and contextual elements of an organization and most corresponding tasks are especially supported by asynchronous technology.

Keywords: Virtual teams · Media synchronicity theory
Information and communication technology · Organizational transformation

1 Introduction

Virtual teamwork is already well established in practice [1], even presenting the fundamental mode of working in some organizations (e.g., Basecamp, Fire Engine RED, 10up, and Zapier). Scientific research addresses and explains phenomena associated with virtual teamwork step by step. Comprehensive insights regarding, e.g., management [2] and collaboration [3] in virtual teams (VTs) are already being offered. But open questions remain and continue to emerge in this highly dynamic field. Challenges from a scientific point of view, e.g., concern technology choice for virtual teamwork. Related insights are contradictory and incomplete despite extensive research (e.g. [4, 5]). With virtual teamwork already in place, conflicts arise due to non-functioning work processes [6]. Thus, the need for guidance regarding a successful integration of virtual

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 67–80, 2018.
https://doi.org/10.1007/978-3-319-99951-7_5

teamwork persists. We address this research gap by providing scientific insights on technology choice for VTs, also offering references for organizations in practice.

Digitalization as a socio-technological phenomenon manifests itself in the mutual interaction of lifestyles and technology. The way people are socialized and decide to live affects the development of innovative technologies that, in turn, enable new ways of life. This interrelation also manifests in megatrends, such as globalization, urbanization, and flexibilization. The demand for and use of technological solutions for tasks that can be performed in a virtual way, and the innovations of software and hardware can thus be regarded as intertwined [7]. This reflects in a growing virtualization of many areas of life. Social interaction is increasingly taking place via virtual platforms. Errands, like event planning and banking, are prominent examples of everyday tasks for which virtual solutions are already established.

Digitalization and the associated virtualization also affect how and by what means people work in organizations. Virtualization facilitates remote and asynchronous teamwork through information and communication technology (ICT), e.g. regarding communication and database access. In order to achieve competitive advantages by integrating virtual teamwork, existing organizations require adaptation and new organizations will be planned along the requirements of a digitalized society and its economy. Virtual teamwork in an organizational context comes in many shapes. Crowdsourcing and prosuming, e.g., include value created by independent customers and enthusiasts. But VTs are also established as regular workforce and thus core resource within an organization. The proceeding megatrends indicate on the one hand that people seek to lead their working lives according to their socialization and thus lead to a demand for technological solutions regarding work processes and work infrastructure. On the other hand, organizations use virtual work to find and retain qualified employees and to lower costs, e.g. for real estate and travels. Organizations implementing virtual collaboration across globally distributed teams are thus one of the many manifestations of current technology-driven and innovation-driving developments in society as well as in organizational design.

The strategic ideas, business processes, ICT use, corporate culture as well as the governance of organizations require adaptation [8], when employees engaged in VTs represent the core resource for value creation. This is the case in many knowledge-driven organizations presenting an ongoing trend towards digital products and services, automated processes, and knowledge work. Current and future generations grow up in a digitalized world and professional as well as academic education aims at developing knowledge work expertise [9]. Yet, the current competition over professional knowledge workers pushes organizations to meet the needs and requirements of potential employees. Today, these employees are more likely to work for an organization that is regarded as open and innovative digital leader [10], which exerts pressure for organizations to transform. A second driver for organizations to increase virtual teamwork is the leveraging effect that virtuality has on team performance. VTs are highly competitive regarding creativity and overall performance. Studies even show that VTs working together via asynchronous media exceed teams that work face-to-face [11]. As these organizational changes are dynamic and interrelated a comprehensive perspective on organizations is taken for analyzing the technology use by VTs in the following chapters.

No standards exist, which can be used to advise organizations which ICT to use for which organizational element, and the ICT use in reality is highly diverse. An ongoing scientific discussion in various disciplines seeks to provide insights and guidance on ICT support for VTs and organizational performance [12–14]. The performance of VTs in an organization mainly depends on three factors: the structure and composition of the team, the business goals, and the support by ICT. While the first two factors are well studied, it is essential to understand for the third factor which ICT are the "right" ones to support performance and which ones should be chosen accordingly. ICT are rapidly emerging and widely used. Moreover, innovations can be promptly overthrown or are neglected contrary to popular predictions (e.g. digital signatures). Furthermore, software solutions as part of ICT are not static regarding their features and capabilities but evolve and are changed over time, following user requirements and technological progress. It is thus reasonable to analyze ICT regarding their capabilities [15] instead of studying bundled solutions such as wikis or video-calls as objects of analysis. Assigning capabilities to ICT could support a better adjustment of organizations to internal or external change triggers. This flexibility strengthens organizational abilities to adapt and thus increase its resilience [16, 17].

Several theories propose insights concerning what ICT to use for which processes in a given context. Research on the application of these theories still leads to contradictive outcomes to some extent [15, 18]. These contradictions are addressed by the media synchronicity theory (MST) which allows including the dynamically evolving context of teamwork and the different goals within teamwork. Furthermore, MST is confirmed to explain media choice and performance in work teams which has been conducted in several case studies (e.g. [19, 20]). Therefore, we deploy MST for the analysis of organizational technology integration in this study. The following study includes the analysis of ICT use by building on existing literature but also considers the technology use in the whole organization and its context [8], adding to the current state of knowledge.

The goal of this study is thus to provide a comprehensive organizational view on technology use when integrating VTs and by that transforming the organization to a certain degree. The insights will provide links for further research and guidance for planning and analyzing a productive ICT use for VTs in practice. This leads to the following research questions:

RQ1: What insights are provided by research for technology choice in virtual teamwork through the lens of MST?
RQ2: Which technological capabilities are required for facilitating specific organizational tasks towards effective virtual team integration?

In order to provide a consistent understanding throughout this study, concepts for VTs and ICT are proposed in the following chapter. MST and a comprehensive view on organizational structures are introduced as grounding and structuring instruments. A literature review regarding analyses of technology use by VTs along MST is presented. The insights are combined and the results for the organizational elements derived. Finally, the main insights, limitations and links for future research are discussed.

2 Conceptualization and State of the Art

2.1 Conceptual Foundation

In order to analyze the required technological capabilities for effective virtual teamwork and to provide a comprehensive overview of the organizational implementation, the background of VTs, ICT, MST and organizational structures is outlined and discussed. The state of the art combining these concepts is derived subsequently through a structured literature review.

The concepts of VTs differ in research regarding the focus of analysis. If the focus is on cultural differences, VTs are defined to consist of globally dispersed employees. If the focus is on the use of ICT, VTs are defined as groups with intense technology use, even if they are co-located in the same region. VTs can also be defined as independent communities or as groups working on mutual tasks. As we aim to derive a holistic view on the integration of VTs in organizations, a rather minimal definition of VTs based on [21] is appropriate for our approach: VTs are groups of people working towards a mutual goal, relying on ICT due to working asynchronously regarding time or due to geographic dispersion. The goals can be both strategic and operational.

ICT that VTs rely on for communication and collaboration include all facets of technological solutions. The term 'media' represent these technological solutions with various features for our study, including hard- and software as well as how people use and interact with them. Therefore, no further constraints are applied regarding the considered ICT for this study.

The MST allows a structured analysis of this vast sphere concerning ICT by focusing on certain technological capabilities and hence suits the goal of this paper. It updates former theories of media use. The media richness theory has been an anchor for analyzing teamwork and its media use [22]. At the end of the 1990s, MST emerged from the short comings of the media richness theory [22], e.g. the focus on tasks instead of results, taking into account the outcome of teamwork and the communication processes involved. According to MST, media with low synchronicity support conveying information, defined as "conveyance" [23] whereas the communication process for "convergence" is supported by media with high synchronicity [15]. Conveyance is required for any teamwork, technologically supported especially for VTs. Convergence leads to a shared mental model in teams and is regarded as fundamental for a mutual goal specification and interpretation [24]. As both conveyance and convergence are required in teamwork, MST underpins that both asynchronous and synchronous communication processes need to be supported by the deployed media, thus leading to a switching between media or features according to the task being performed [23]. Five capabilities for analyzing ICT are proposed by MST [15, 18]. These capabilities are presented in Fig. 1 and described in the following. They are used for the analysis of organizational elements in the next chapter.

Technology that provides mainly asynchronous features, e.g. storage or email, is categorized as asynchronous media while synchronous media, e.g. phone calls or chats, are regarded as opposite. Both characteristics are supported by certain capabilities. These capabilities are crucial for allowing conveyance and convergence processes and related tasks. Communication via media with high transmission velocity provides

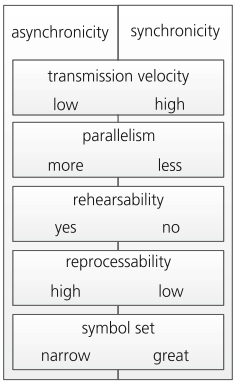


Fig. 1. Capabilities of information and communication technology [15, 18]

messages without delay and thus supports synchronous understanding. A parallel transmission of messages instead of mere serial transmission can impair synchronous understanding as it bears the risk to lower a shared focus. E.g. phone or video calls allow less parallelism than asynchronous media such as forums or wikis. Rehearsability provides the opportunity to edit messages before sending them. This might lead to delays and is thus more appropriate for asynchronous communication. Media that support reprocessability of content, e.g. by reviewing prior messages or stored information, are suitable for asynchronous communication processes, e.g. the conveyance of extensive and complex information. The so called symbol set describes the ways that are available for information transmission, e.g. gestures, pictures and written text. Synchronicity in communication is increased, the better the symbol set fits the information. Further details and examples for these capabilities are described by [15].

Technology use by VTs has been analyzed with varying focuses. In order to work towards a holistic view, the interrelated elements within and around organizations serve as structure for the following analysis. Adaption of the entire organization for integrating virtual teamwork needs to serve strategical as well as operational goals, e.g. recruiting new employees and optimizing team communication. A successful integration leads to effective teamwork and thus value creation. The relevant concepts are summarized in Fig. 2.

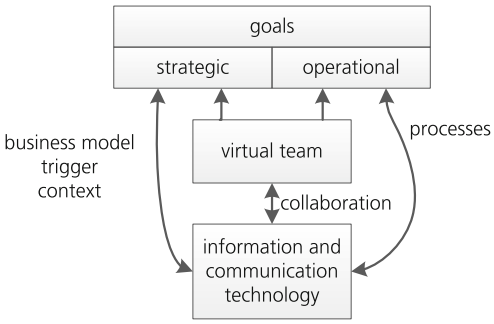


Fig. 2. Virtual teams as organizational core resources

2.2 Literature Review Process

The choice of the right ICT based on technological capabilities can be supported by using MST and by carrying out a structured analysis of the embedding of the VTs in the organization. In order to synthesize existing scientific insights and reveal research gaps, we reviewed literature analyzing technology use by VTs based on MST. The review followed the characteristics as shown in Table 1.

The databases were searched using the full text options. All finding where checked whether they match the scope of VTs and the discussion of MST. Doubles where eliminated and the resulting articles checked for accessibility. All but three articles where accessible via the available licenses. A backward search was performed using the articles reference lists. The resulting sample consists of 14 peer-reviewed scientific journal and conference publications as shown in Table 2.

Table 1. Characteristics of the literature search

Search terms	Media synchronicity theory AND virtual team*
Time span	Until 2018
Databases	EBSCO (Business Source Ultimate, eBook Collection), IEEE Xplore Digital Library, SCOPUS

Table 2. Findings and final sample

Database	Findings
EBSCO	9
IEEE Xplore Digital Library	3
SCOPUS	15
After check for doubles and accessibility	12
After backward search	**14**

The articles were matched to the following criteria: the functional, cultural, governance related and contextual elements (Table 3). The criteria are based on a holistic view on organizations regarding transformation processes and derived from following organizational elements [8, 25]. The functional elements of organizations include strategic initiatives and business models, business processes and the implemented ICT [8]. The cultural elements are embedded in the governance and context of the organization and include structures and related processes regarding company culture, leadership and behavior, power structures and social competencies [3, 26]. The governance includes legal regulations as well as management control systems. Trends and norms of a region or a certain industry sector are elements that are considered when

Table 3. Organizational elements addressed by the reviewed literature

	[28]	[15]	[29]	[30]	[31]	[32]	[33]	[24]	[19]	[34]	[35]	[20]	[14]	[5]
strategy														
processes	■	■	■		■		■	■	■			■		■
ICT			■				■	■	■	■			■	
company culture				■			■		■					■
leadership and behavior	■	■		■	■		■		■					■
power structures		■									■			
social competencies	■		■	■			■					■		
governance														
context														■
trigger														

analyzing the context. The context thus includes parameters presenting the opportunities as well as restrictions for the organizational change. The context and triggers are external factors that cannot be controlled by the organization itself [25]. But their influences require internal reactions and control. Triggers introduce pressure for organizational change of structures [27] and include, e.g., the market situation, megatrends and innovations in ICT. In our case, the endeavor to deploy VTs and technological innovations are triggers of the changes of the organizational elements. The results are described in the following chapter. An overview of the 14 sources and what elements of an organization they address are displayed in Table 3.

In order to derive insights on the technology-fit for all organizational elements, the results are synthesized in the following chapter for all criteria and missing topics are discussed subsequently.

3 Technology Use for Virtual Team Integration

3.1 Current Insights on Technology Use

Tasks are performed in all organizational elements, e.g. defining legal regulations for governance or performing market research for strategic endeavors. All tasks require conveyance and convergence to a different extent [15] as described above and all

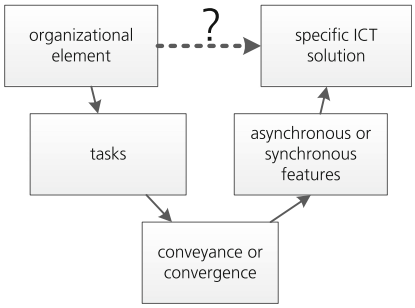


Fig. 3. Process of matching adequate ICT on organizational element

presented elements include several tasks. In order to provide further guidance regarding which ICT to use for which elements of an organization, a mere task-based perspective proved to be too superficial [15]. Therefore, the tasks need to be analyzed concerning their goals, if reaching these goals requires conveyance and convergence to what extent and in which step. This will allow deriving ICT with features that support conveyance and convergence as specifically required for each element. Thus, the matching of ICT to organizational elements needs to be intermediated by a high granular analysis as shown in Fig. 3.

The findings of the literature review are described in the following regarding the business processes and ICT as functional elements of an organization. Then the cultural elements are analyzed regarding company culture, leadership and behavior, power structures and social competencies, and the insights on ICT use for the context of an organization are described. The literature review provided no findings for ICT regarding the strategy, triggers and governance, as shown in Table 3. Therefore, additional ideas are presented for these elements subsequently (Sect. 3.2).

ICT are used to support and improve *processes* [23]. Tasks performed by VTs aim at executing business processes, comparable to everyday work by tradition teams. The view on communication and surrounding business processes when analyzing the use of ICT by VTs is the main perspective of most analyzed sources. Asynchronous media are used for displaying discussions and present results [23]. Shared mental models of the tasks and processes as well as the team structure itself positively influence the team's performance and vice versa [24]. Shared mental models support, e.g., the understanding of goals, help compensating language barriers and thus allow teams to work efficiently and effectively on their tasks. ICT can facilitate shared mental model building by, e.g., documenting goals and milestones and providing various symbol sets for cross-language and intercultural communication. The development of shared mental models can be structured into stages, requiring both conveyance and convergence processes. When forming new VTs or seeking to improve the performance of existing VTs, the mental models require evaluation and their level of sharing by the team members can be enhanced by selecting asynchronous and synchronous media for the respective processes [24]. The execution of most business processes requires decision making. Conference tools that are low in parallelism are found to support decision

making by VTs better than ICT with high parallelism, e.g. virtual environments [34] even though both ICT are suitable for convergence. This is in line with the findings that the media rich communication, especially face-to-face, is not always the best choice [15] and also requires further investigation [14, 36]. The results also show that parallelism requires to be analyzed with respect to the tasks that are performed. An insight contradicting the assumption of MST regarding parallelism and synchronicity was discovered: When tasks are performed via ICT in a parallel manner, e.g. via multiple text-messaging chats, synchronicity can even increase through parallelism, when the tasks are of minor complexity [20].

The use of *ICT* with asynchronous features allows producing higher input than face-to-face communication [23]. Face-to-face communication cannot always be considered to be the richest medium [15], thus less synchronous media can support convergence to even a higher degree than face-to-face communication in some cases. The effectiveness of ICT use can be improved regarding all organizational elements, when VTs are provided with training on the implemented ICT in order to not only know how to use them but to understand their matching as well as risks for specific communication tasks [35].

Corporate *culture* is an integral component of any organization. It has strong influences on successful implementation of processes, ICT use, and behavior and thus on the overall performance of organizations [37]. The use of asynchronous media can be intensified and the use of synchronous media can be reduced over time, when the VTs have established an effective working mode [15, 23]. This is explained by the development of trust and norms and the resulting ability to manage conflicts [33]. It is also related to contradictive findings that for the kick-off of projects, synchronous face-to-face meetings increase team's productivity [23] or that teams working more or solely asynchronously are more productive [11, 32]. As the evolutions of team's ICT use are complex and individually shaped, forecasts are difficult, if not impossible, as of now [31]. Therefore, an ongoing monitoring of the team's ICT use is still necessary and essential for measuring work performance [19]. The overall use of ICT regarding individual mindsets need to be considered when forming VTs. Even though the teams develop over time as described above, individual traits, such as anxiety towards media use can remain stable and compromise team performance [29]. Synchronous media, e.g. video conferencing, can be used for regular work meetings that also include a time slot for informal conversation. These regular conversations in VTs can support the understanding of the other team members' lives and associated impacts on their work behavior [19]. VTs often include employees with different native languages and who are situated in dispersed regions. These aspects also influence communication and media choice, e.g. the preference of asynchronous media due to language barriers [5]. Thus, languages as well as additional influencing factors, e.g. technological socialization and time zone, impact the choice of asynchronous or synchronous media for communication related processes in VTs. Therefore, results differ to the propositions by MST when a variety of team members is taken into account.

Leadership, behavior and power structures are also core elements of the cultural elements of an organization and their manifestations thus influence the overall performance. If single team members are not working according to rules and their goals

deviate from the team's goals, there is a risk that asynchronous media are misused to lower the team's performance [30]. Even if a consistent corporate culture is aspired, the actual use of planned ICT cannot be enforced, especially regarding leadership [23]. Information entails power in social as well as organizational settings. Power dynamics are also bound to information sharing or restrains in teamwork. This also influences the team's performance as teams make better informed decisions regarding mutual goals, when all information is shared [35]. Changes in power dynamics are an important step when implementing new organizational structures. As this does not only affect the corporate culture but also individual professional situations, it is a sensitive and delicate issue which requires to be approached with care.

Social competencies of team members and leaders are important alongside professional skills for a team's performance. Highly synchronous media, even face-to-face meetings, are viable for sensitive topics such as personal issues [23]. This is explained by the ICT's capability to provide immediate feedback, e.g. in video conferences [33]. Yet, low synchronicity appears to be able to balance out individual differences. E.g. individual shyness can be compensated by using asynchronous media for stating arguments and for decision making [32]. Social as well as technological competencies influence media use. Not only the quantity of use differs between individuals regarding their familiarity with ICT as one would expect, but also the quality of use regarding task-oriented and social messaging are influenced by ICT competencies [29].

The *context* which organizations are embedded in consists of trends and norms, corresponding to the megatrends introduced at the beginning. This element was only tackled by one source of our review. Tenzer and Pudelko [5] approach this element by discussing different manifestations of technology choice and team performance regarding MST. In general, contextual information is stored in the team members' intrinsic knowledge and thus communicated mainly indirectly, e.g. through speech behavior and the ways language is used. This is in line with the findings by [5] on the one hand, as these norms differ among people that are e.g. socialized in distant regions. Digitalization, on the other hand, virtually shrinks these distances and thus impacts the extent to which socializations still differ. In order to meet the requirement to embed an organization in its context when adapting to virtual teamwork, characteristics of the organization's context could be documented and transferred asynchronously by monitoring trends and norms and communicating them to the teams. The communication can also be supported by synchronous media, e.g. in intercultural trainings.

3.2 Holistic View on Organizational Technology Use

Most studies focus on communication and collaboration of VTs addressing operational goals and thus the functional element regarding business processes, as shown in Table 3. Strategy, triggers and governance are organizational and contextual elements which are not addressed by the reviewed research papers. But these elements are also affected by the deployment of VTs and their adaption strongly influences the work performance of VTs. In order to also provide guidance regarding appropriate ICT use for these elements, they are analyzed along MST and its five capabilities in the following. All results are summarized in Fig. 4 below.

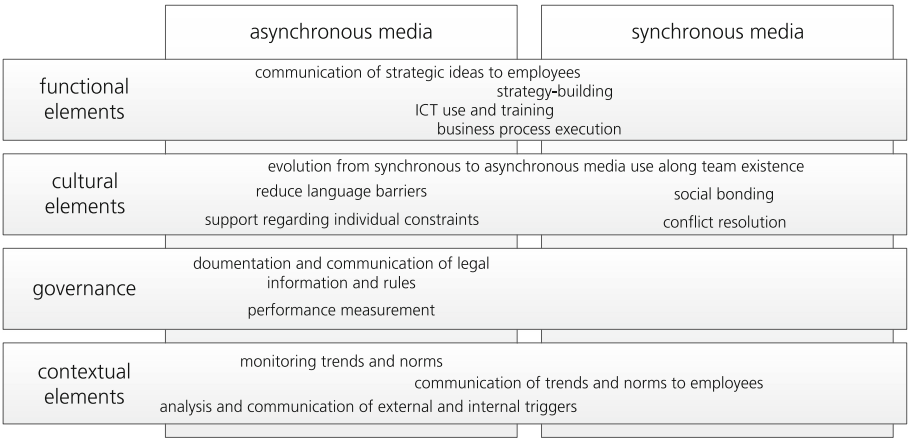


Fig. 4. Technology-oriented matching of tasks and organizational elements

The *business strategy* is shaped by the deployment of VTs regarding strategic planning, e.g. the focus set on customers [38] or product, the positioning in market and competitive network [25] and further internal and external effects of virtuality. Therefore, the ICT deployed for strategic planning and strategy communication needs to be reviewed. Conveyance processes of strategic planning and strategy communication are supported mainly by asynchronous media, as they require severe rehearsability. A rich symbol set supports strategy conveyance, with that addressing every employee, present, virtual and remote. Convergence processes for building or updating an organizational strategy require mutual mental models which are supported by synchronous communication, based on immediate feedback through high transmission velocity, e.g. by phone and video calls.

The *triggers* are not directly influenced by the organization as described above and thus do not require convergence. But the triggers' analysis and communication for and within organizations are relevant tasks as they concern all structures and employees. ICT supporting conveyance processes can support the analysis and communication of external and internal triggers so resulting organizational changes are transparent for all employees. The documentation can be updated and monitored and thus requires high reprocessability. This supports the need for media with asynchronous features.

The insights on triggers also apply to the organizational *governance*. Low synchronicity allows conveying richer information and revising messages for completeness and correctness. This is highly relevant for governance tasks. If legal information is stored and presented, asynchronous media are more qualified due to high rehearsability and reprocessability. This follows traditions of written legal documents and documented intra-organizational rules. Also the governance related tasks of measuring outcomes and providing reports are apt for asynchronous media due to equal reasons. This follows the process analysis of MST, as the governance tasks are all substantially conveyance related processes.

4 Conclusion

The results show that MST supports the analysis and choice of ICT in the context of business transformation regarding the deployment of VTs. The specific requirements of communication and collaboration tasks for strategic and operational goals can be unraveled using MST and a technology choice and implementation be supported (RQ1). The literature review shows that MST is established for analyzing ICT use by VTs. The insights confirm most aspects proposed by MST, but some aspects require refinement when analyzing VTs. Future research could include, e.g., propositions based on the adaptive structuration theory (AST) [31] in order to explain how team behavior evolves over time also leading to alterations in technology use. An initial face-to-face meeting is still handled as standard in forming VTs. As research outcomes are contradictive, further analyses are required for revealing whether this procedure is actually suitable. This also applies for decision making in VTs where results on media choice are contradictive (e.g. [14, 34, 35]). Studies confirm that purely asynchronously working VTs are more creative and successful regarding decision making processes [11, 31], which is contradicting the propositions by MST. Some organizational elements were not analyzed along MST yet, presenting a research gap that was approached by deriving complementary propositions above (RQ2).

Limitations occur due to the focus on insights based on MST for this study. This could be overcome by including more studies that also focus on technological capabilities but not directly refer to MST or to VTs. The results could then be interpreted following MST. First insights are provided above by transferring the findings from the studies to the uncovered elements. All organizational elements were briefly analyzed in order to provide a comprehensive view, preparing for more in depth research for chosen elements in future research.

Our findings show that a comprehensive perspective should be taken, even if the affecting changes are assumed to only tackle parts of the organization. Therefore, adaptive ICT does not only meet technological requirements, but supports overall business resilience [16, 17] by allowing flexible reactions on triggers. The insights presented above add to an ongoing scientific discussion with focus on VTs and support the practical monitoring and improvement of organizations in times of digitalization and virtuality as one of the manifestations.

References

1. Gilson, L., Maynard, M., Young, N., Vartiainen, M., Hakonen, M.: Virtual teams research: 10 years, 10 themes, and 10 opportunities. J. Manag. **41**(5), 1313–1337 (2015)
2. Dunn, S., Grannan, C., Raisinghani, M., Stalling, H.: Communication strategies for successful virtual teams. In: 48th Hawaii International Conference on System Sciences, Hawaii, USA, pp. 364–373 (2015)
3. Salminen-Karlsson, M.: Enabling virtual communities of practice: a case-study of Swedish-Indian collaboration in IT development. Electron. J. Inf. Syst. Eval. **17**(1), 60–70 (2014)
4. Ahmad, R., Lutters, W.: Perceived faultline in virtual teams: the impact of norms of technology use. In: Pacific Asia Conference on Information Systems, p. 133 (2015)

5. Tenzer, H., Pudelko, M.: Media choice in multilingual virtual teams. J. Int. Bus. Stud. **47**(4), 427–452 (2016)
6. Bathini, D., Kandathil, G.: Work from home: a boon or a bane? The missing piece of employee cost. Indian J. Ind. Relat. **50**(4), 568–574 (2015)
7. Adner, R., Levinthal, D.: Demand heterogeneity and technology evolution: implications for product and process innovation. Manag. Sci. **47**(5), 611–628 (2011)
8. Österle, H., Blessing, D.: Business engineering modell. In: Österle, H., Winter, R. (eds.) Business Engineering. BE, pp. 65–85. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-642-19003-2_4
9. Pinzaru, F., et al.: Millennials at work: investigating the specificity of generation Y versus other generations. Manag. Dyn. Knowl. Econ. **4**(2), 173–192 (2016)
10. Kane, G., Palmer, D., Phillips, A.N., Kiron, D., Buckley, N.: Strategy, not technology, drives digital transformation - becoming a digitally mature enterprise. Research Report, MIT (2015)
11. Ocker, R., Fjermestad, J.: Communication differences in virtual design teams: findings from a multi-method analysis of high and low performing experimental teams. DATABASE Adv. Inf. Syst. **39**(1), 51–67 (2008)
12. Minas, R., Dennis, A., Massey, A.: Opening the mind: designing 3D virtual environments to enhance team creativity. In: 49th Hawaii International Conference on System Sciences, Hawaii, USA, pp. 247–256 (2016)
13. Schott, T.: Adopting Workgroup Collaboration Tools in 3D Virtual Worlds. ProQuest LLC, Ann Arbor (2014)
14. Schouten, A., Hooff, B., Feldberg, F.: Virtual team work: group decision making in 3D virtual environments. Commun. Res. **43**(2), 180–210 (2016)
15. Dennis, A., Fuller, R., Valacich, J.: Media, tasks, and communication processes: a theory of media synchronicity. MIS Q. **32**(3), 575–600 (2008)
16. Bustinza, O., Vendrell-Herrero, F., Perez-Arostegui, M., Parry, G.: Technological capabilities, resilience capabilities and organizational effectiveness. Int. J. Hum. Resour. Manag. 1–23 (2016)
17. Linnenluecke, M.: Resilience in business and management research: a review of influential publications and a research agenda. Int. J. Manag. Rev. **19**(1), 4–30 (2017)
18. Dennis, A., Valacich, J.: Rethinking media richness: towards a theory of media synchronicity. In: Proceedings of the 32nd Hawaii International Conference on System Sciences, Hawaii, USA, p. 10p (1999)
19. Niinimäki, T., Piri, A., Lassenius, C., Paasivaara, M.: Reflecting the choice and usage of communication tools in GSD projects with media synchronicity theory. In: International Conference on Global Software Engineering, pp. 3–12 (2010)
20. Randall, D.: Ten red balloons: virtual teams and online communities - a test of media synchronicity theory. iSChannel **6**(2), 13–21 (2013)
21. Schweitzer, L., Duxbury, L.: Conceptualizing and measuring the virtuality of teams. Inf. Syst. J. **20**(3), 267–295 (2010)
22. Dennis, A., Valacich, J., Speier, C., Morris, M.: Beyond media richness: an empirical test of media synchronicity theory. In: Proceedings of the Thirty-First Hawaii International Conference on System Sciences, Hawaii, USA, pp. 48–57 (1998)
23. DeLuca, D., Valacich, J.: Outcomes from conduct of virtual teams at two sites: support for media synchronicity theory. In: Proceedings of the 38th Hawaii International Conference on System Sciences, Hawaii, USA, p. 50b (2005)
24. Maynard, M., Gilson, L.: The role of shared mental model development in understanding virtual team effectiveness. Group Org. Manag. **39**(1), 3–32 (2014)
25. Baumöl, U.: Change Management in Organisationen - Situative Methodenkonstruktion für flexible Veränderungsprozesse. Gabler, Wiesbaden (2008)

26. Österle, H., Winter, R.: Business engineering. In: Österle, H., Winter, R. (eds.) Business Engineering. BE, pp. 3–20. Springer, Heidelberg (2000). https://doi.org/10.1007/978-3-642-98097-8_1
27. Friedel, D., Back, A.: Determination of enterprise 2.0 development levels with a maturity model. In: Nunes, M., Peng, G., Roth, J., Weghorn, H., (eds.) Proceedings of the IADIS International Conferences, Lisbon, pp. 3–9 (2012)
28. DeLuca, D., Valacich, J.: Virtual teams in and out of synchronicity. Information **19**(4), 323–344 (2006)
29. Fuller, R., Vician, C., Brown, S.: Longitudinal effects of computer-mediated communication anxiety on interaction in virtual teams. IEEE Trans. Prof. Commun. **59**(3), 166–185 (2016)
30. Giordano, G., George, J.: The effects of task complexity and group member experience on computer-mediated groups facing deception. IEEE Trans. Prof. Commun. **56**(3), 210–225 (2013)
31. Han, H.-J., Starr, R., Fjermestad, J., Wang, Y.: Does medium matter? A comparison of initial meeting modes for virtual teams. IEEE Trans. Prof. Commun. **54**(4), 376–391 (2011)
32. Hassell, M., Limayem, M.: A portfolio of media: effects of media synchronicity on communication performance. In: Thirty Second International Conference on Information Systems, Shanghai, China, p. 13 (2011)
33. Maruping, L., Agarwal, R.: Managing team interpersonal processes through technology: a task-technology fit perspective. J. Appl. Psychol. **89**(6), 975–990 (2004)
34. Nordbäck, E.: The influence of emergent technologies on decision-making process in virtual teams. In: Proceedings of the 19th Americas Conference on Information Systems, Chicago, USA, pp. 3417–3424 (2013)
35. O'Neill, T., Hancock, S., Zivkov, K., Larson, N., Law, S.: Team decision making in virtual and face-to-face environments. Group Decis. Negot. **25**(5), 995–1020 (2016)
36. Schouten, A., Hooff, B., Feldberg, F.: Real decisions in virtual worlds: team collaboration and decision making in 3D virtual worlds. In: 31st International Conference on Information Systems, Saint Louis, USA, p. 18 (2010)
37. Walker, B., Soule, S.: Changing company culture requires a movement, not a mandate. Harv. Bus. Rev. 2–6 (2017)
38. Shah, D., Rust, R., Parasuraman, A., Staelin, R., Day, G.: The path to customer centricity. J. Serv. Res. **9**(2), 113–124 (2006)

Defining Transformational Patterns for Business Model Innovation

Ilia Bider(✉) and Erik Perjons

DSV, Stockholm University, Stockholm, Sweden
{ilia, perjons}@dsv.su.se

Abstract. The pace of changes in the business environment in which a modern enterprise operates requires the enterprise to constantly review its business models in order to survive and prosper in the dynamic world. This exploratory study investigates how to help the enterprise to innovate their business models based on the concepts of fractal enterprise model and transformational patterns. The paper suggests an approach to Business Model Innovation (BMI) where the focus is on transformational patterns. It discusses the structure of such patterns, and based on examples, it presents an approach on how such patterns can be derived from cases of completed business transformations.

Keywords: Business model · Innovation · Strategy · Business transformation Pattern · Enterprise modelling · Fractals

1 Motivation

In the dynamic world of today, enterprises need to be innovative not only in the line of products and services they offer, but also in who they are and what they do, i.e. under which Business Models they operate. This is needed in order to survive in the turbulent, technology driven business environment and/or exploit the opportunities for growth, emerging due to the changes in the environment. For example, in the future, a traditional manufacturing company may not be able to continue its business as usual, i.e. designing and manufacturing their own products, due to the emergence of mature 3-D printing. Instead, the company may need to change its business model, for example, becoming a designer while letting the customer 'print' the design in the place convenient for the customer, or becoming a manufacturer, providing the customer with a service to 'print' somebody else's design (having both alternative could do as well). This change could be more radical than adding a new product or service to the company's offerings.

In light of the increasing pace of changes in the business environment, it is not a surprise that the topic of Business Model Innovation (BMI) got attention from both practitioners and researchers. This interest is visible in numerous research publications, including books [1] and special issues of journals [2]. According to the classification suggested in [3], there are three ways of innovating a Business Model (BM):

1. *Industry model* innovation - which amounts to change the position in value change, entering new markets, and/or other type of radical changes.

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 81–95, 2018.
https://doi.org/10.1007/978-3-319-99951-7_6

2. *Revenue model innovation* - which results in changes in how a company generates revenues, e.g. reconfiguring offerings and/or introducing new pricing models.
3. *Enterprise model* innovation - which involves innovating the structure of an enterprise, such as enterprise goals, business processes, products and/or services.

In this paper, we focus exclusively on the first type of BMI, i.e. industry model innovation. Note, however, the three types are not independent, but can interweave. This especially concerns the industry model innovation, as, for example, entering a new market may require changes in the existing revenue model as well as the enterprise model. From the strategic perspective, the industry model innovation corresponds to changing the doctrine [4], i.e. who we are, which is the highest level of strategic decision making.

From the field of engineering invention [5], it is well known that over 90% of inventions do not propose anything new, but suggest new combinations of known ideas, or exploiting known ideas in new domains. Based on this fact, Altshuller created a method called TRIZ [5] that "industrializes" invention in the field of engineering by making use of recurrent problems and solutions across industries. In the same way, an enterprise that needs or wants to change its BM does not need to invent a new model but can follow a proven example invented by somebody else.

A number of research papers is devoted to designing tools that could help the decision makers to successfully innovate their BMs. Some of these works suggest standardized procedures for innovation, for example, [6] suggest a procedure to use analogies when innovating a business model. Other works [7, 8] suggest using patterns as help in designing a new business model. Patterns can be on the highest level as in [7], or on the level of elements of the business model, as in [8].

When designing a new business model for an already existing enterprise, it makes sense to take into account and to utilize already existing enterprise structures and elements. Designing a new BM completely unconnected to the existing enterprise does not make sense, as it might be easier to create a new company in this case instead of transforming the old one.

An enterprise has several strategic layers [4]; any of them or a combination can be used as a starting point for innovating a business model. According to [4], the highest strategic layer is the doctrine, i.e. who we are (mentioned above), followed by the layer of *capability/infrastructure*. The other two strategic layers are *grand strategy*, e.g. choice of a sector or strategic alliances, and *strategy as such*, which is defined as a structural coupling to the elements of environment, e.g. competitors, partners, customers, markets, etc. In this paper, we will mostly consider the layer of capability/infrastructure as a starting point for innovating a business model of an existing enterprise. In this case, the innovation consists of reconfiguring capabilities, including the infrastructure, for a new strategic direction. This corresponds to the idea of destruction (of an old model) and creation (of a new one) by reconfiguring the existing elements, as suggested in [9].

The research reported in this paper aims at creating a procedure that facilitates business model innovation based on transformational patterns. The usage of concept of pattern in our approach is different from other works that exploit patterns in BMI, e.g. [7, 8]. In the latter works, patterns are used to design a new business model. Our

patterns are used to find a way to transform an existing business model. As it has been mentioned above, we mostly consider transformations that are based on existing capabilities and infrastructure. This type of transformation requires depicting existing capabilities/infrastructure in some form.

Currently, a business model is depicted using some kind of a canvas, e.g. as suggested in [7]. Such a model may not reveal all capabilities existing in the enterprise, and thus it is not very useful for the sake of transformation. Therefore, we are not using a canvas type of enterprise model, but a more powerful enterprise model that can reveal, if not all, but the essential capabilities of the enterprise in question.

The enterprise model we use for transformational purpose is called a Fractal Enterprise Model (FEM). It was first introduced at PoEM 2012 [10], and then extended and improved in [11]. FEM has a form of a directed graph with two types of nodes *Processes* and *Assets*, where the arrows (edges) from assets to processes show which assets are utilized by which processes and arrows from processes to assets show which processes help to have specific assets in healthy and working order. The arrows are labeled with meta-tags that show in what way a given asset is utilized, e.g. as *workforce*, *reputation*, *infrastructure*, etc., or in what way a given process helps to have the given assets "in order", i.e. *acquire*, *maintain* or *retire* the assets.

A FEM model is built recursively by using a so called unfolding procedure and two types of archetypes: *process-assets archetypes* that show which kind of assets might be needed for running a process, and an *asset-processes* archetype that shows which processes are needed to maintain an asset in order. Unfolding starts with a primary process, a process that delivers value to a customer/beneficiary, by applying process-assets archetypes and alternating them with the asset-processes archetype.

Both assets and processes represent capabilities that exists in the enterprise and can be used as a basis for a transformation. A major organizational change would result in changing FEM. A change that corresponds to an industry model innovation would result in appearing nodes for new primary processes that substitute or complement the old ones. A part of a FEM graph that corresponds to a new primary process would use assets and processes that already exist in the old FEM, though some modification can be made in them. This reuse of nodes from the old FEM in a new one represents employment of the existing capabilities during industrial business model innovation.

The idea of using FEM for business model transformation was first presented in [12], but without any technical details on the transformational procedure and transformational patterns. A more recent publication [13] sets a research agenda for converting the idea from [12] into a practically useful procedure. The agenda consists of several items, including defining transformational patterns (archetypes), and providing computerized tools support. The current paper aims at making a step in fulfilling the research agenda by defining a possible structure of a transformation pattern, and presenting examples of such patterns.

The rest of the paper follows the following structure. In Sect. 2, we give an overview of the related literature. In Sect. 3, we present a short description of FEM. In Sect. 4, we discuss our research approach. In Sect. 5, we demonstrate a way for analysis of examples of BMI. In Sect. 6, we discuss the structure of transformational patterns. In Sect. 7, we present another transformational pattern, also built based on a BMI example. Section 8 contains concluding remarks and plans for the future.

2 An Additional Literature Overview

There is a sizable body of literature devoted to BMI in addition to what has already been reviewed in Sect. 1. The book by [1] provides a systematic review of this literature. The works that are related to the current research belong to the area *Tools and Processes for BPI*. According to [1], "overall, research on the BMI process is still in its infancy and more work is needed in all respects". Still, there are a number of researchers working in this area. Below, we review some works related to our research that represent the main directions in the area of tool and processes for BMI.

The idea of experimentation starting from the existing BM and trying to reconfigure it is expressed in [14]. This work proposes using the business model canvas [7] as an enterprise model to start experimentation. Moreover, it does not suggest any systematic way for experimentation, therefore, the experimentation remains on the ad-hoc level. We are also promoting the idea of experimentation. However, we suggest the use of a richer enterprise model as well as a more systematic way for experimentation by applying transformational patters to the existing model.

The ideas of using the experience of others, so-called best practices, is promoted by a number of researchers. For example, [6] suggest a procedure (process) called *BMI by analogy* that consists of finding cases of BMI and reinterpreting them so that the ideas can be applied in a certain business. In our work, we exploit the idea of analogy implicitly; namely, we suggest (1) having a library of cases instead of scanning the environment for them each time, (2) pre-interpreting the cases by increasing the level of abstraction in the case description, thus converting them into transformational patterns.

Another way of using best practices is by using a library of patterns to be used in the BMI process. Two directions of using patterns can be found in the literature. The first direction is having a pattern of overall composition of a business model, as in [7] where five general BM patterns are defined. The second direction is introducing patterns for individual components of a BM, as suggested in [8]. In both directions, patterns are used for designing a new BM. Our aim of using patterns is different, we suggest using transformational patterns to get an idea for changing/complementing an existing BM. After the new idea has been chosen, the non-transformational patterns can be applied.

Summarizing the short literature overview above, we can conclude:

- The research area of tools and processes for BMI is a relatively new field which requires attention from the researchers.
- Our approach uses some of the already known ideas, like experimentation, analogy and patterns. However, it implements and combines these ideas in a different way, i.e. use a richer business model as well as transformational patterns.

3 Fractal Enterprise Model

The Fractal Enterprise Model (FEM) includes three types of elements: business processes (more exactly, business process types), assets, and relationships between them, see Fig. 1 in which a fragment of a model is presented. The fragment is related to a

hypothetic company that sells books over the Internet. Graphically, a process is represented by an oval, an asset is represented by a rectangle (box), while a relationship between a process and an asset is represented by an arrow. We differentiate two types of relationships in the fractal model. One type represents a relationship of a process "using" an asset; in this case, the arrow points from the asset to the process and has a solid line. The other type represents a relationship of a process changing the asset; in this case, the arrow points from the process to the asset and has a dashed line. These two types of relationships allow tying up processes and assets in a directed graph.

In FEM, a label inside an oval names the given process, and a label inside a rectangle names the given asset. Arrows are also labeled to show the type of relationships between the processes and assets. A label on an arrow pointing from an asset to a process identifies the role the given asset plays in the process, for example, *workforce*, *infrastructure*, etc. A label on an arrow pointing from a process to an asset identifies the way in which the process affects (i.e. changes) the asset. In FEM, an asset is considered as a pool of entities capable of playing a given roles in a given processes. Labels leading into assets from supporting processes reflect the way the pool is affected, for example, a label *acquire* identifies that the process can/should increase the pool size.

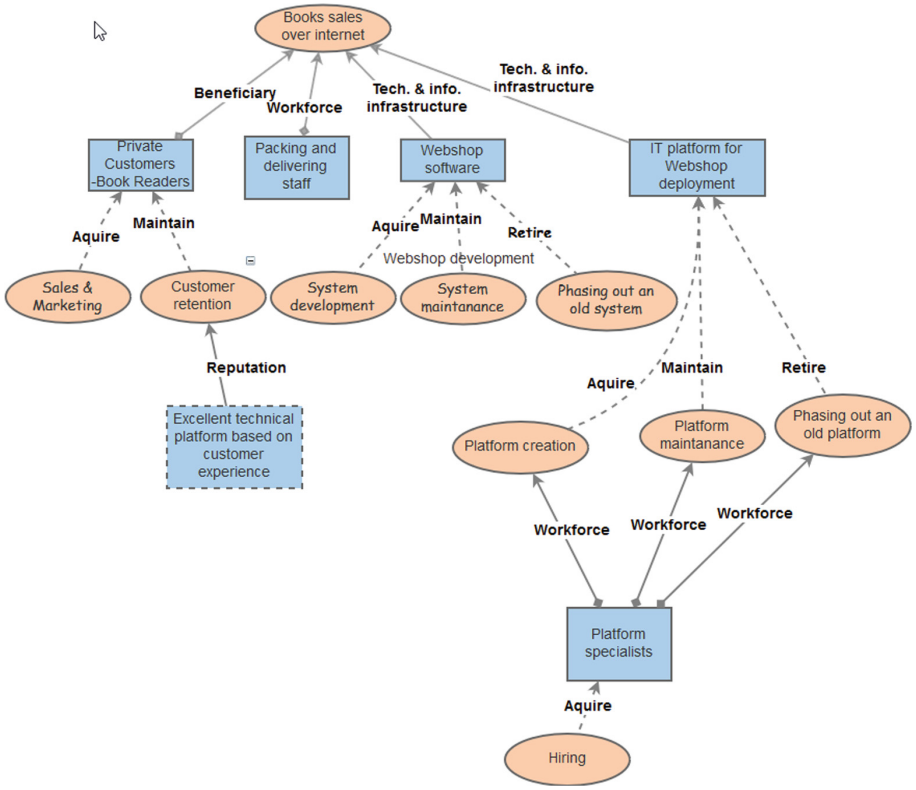


Fig. 1. A fragment of a FEM from [13]

Note that the same asset can be used in two different processes playing the same or different roles in them, which is reflected by labels on the corresponding arrows. It is also possible that the same asset can be used for more than one role in the same process; in this case, there can be more than one arrow between the asset and the process, but with different labels. Similarly, the same process could affect different assets, each in the same or in different ways, which is represented by the corresponding labels on the arrows. Moreover, it is possible that the same process affects the same asset in different ways, which is represented by having two or more arrows from the process to the asset, each with its own label.

In FEM, different styles can be used for shapes to group together different kinds of processes, assets, and/or relationships between them. Such styles can include using dashed or double lines, or lines of different thickness, or colored lines and/or shapes. For example, a dashed border of an asset (see the asset "Excellent technical platform based on customer experience" in Fig. 1) points to the asset being intangible opinion of stakeholders. A diamond start of an arrow from an asset to a process means that the asset is a stakeholder of the process (see the arrows "Workforce" in Fig. 1).

Labels inside ovals, which represent processes, and rectangles, which represent assets, are not standardized. They can be set according to the terminology accepted in the given domain, or be specific for a given organization. Labels on arrows, which represent the relationships between processes and assets, however, can be standardized. This is done by using a relatively abstract set of relationships, like, *workforce*, *acquire*, etc., which are clarified by the domain- and context-specific labels inside ovals and rectangles. Standardization improves the understandability of the models.

To make the work of building a fractal model more systematic, we introduced archetypes (or patterns) for fragments from which a particular model can be built. An archetype is a template defined as a fragment of a model where labels inside ovals (processes) and rectangles (assets) are omitted, but arrows are marked. Instantiating an archetype, means putting the fragment inside the model and labeling ovals and rectangles; it is also possible to add elements absent in the archetype, or omit some elements that are present in the archetype.

We introduce two types of archetypes, process-assets archetypes and asset-processes archetypes. A process-assets archetype represents which kind of assets that can be used in a given category of processes. An asset-processes archetype shows which kinds of processes are aimed at changing the given category of assets.

Note that the fractal model does not represent direct relationships between business processes, such as generalization or composition. On the level of abstraction accepted for FEM, a process with all its possible sub-processes is considered as one process.

4 The Research Approach

The research presented in this paper belongs to the Design Science (DS) paradigm, [15–17], while we use an approach suggested in [18]. Using DS is natural for this research, as the objective is to develop a way of depicting and using transformational patterns for BMI, as well as creating a library of such patterns. The problem the research addresses is that, currently, BMI (at least, industry model innovation) is done

in an ad-hoc manner based on experience and intuition, which may result in a new BM that is too costly, difficult, or impossible to implement. The proposed solution, or artifact in terminology of [15], is transformational patterns that could be applied to a FEM-based enterprise model.

The development of a solution for the problem defined above requires both (a) defining a structure of a transformational pattern, and (b) creating a sizable library of such patterns. An appropriate approach for both tasks is analyzing examples/cases of transformations of the industry model innovation type, successful and unsuccessful, from the international enterprise practice. Information of such transformations can be found in different sources, like books, Internet, and our own practice. The analysis and building of a pattern are envisioned as consisting of the following steps:

1. Build a relevant fragment of FEM before the transformation
2. Build a relevant fragment of FEM after the transformation
3. Relate elements of these two models showing which elements of the original model have been used in the transformed model and how they have been change during the transformation
4. Abstract from the details of the given case creating a transformational pattern
5. Finding other examples of transformation that fit the constructed pattern.

The analysis of a specific transformational example/case can be done in different ways, dependent on the amount of information available about the example. In the best case, information can be available on the rationale behind the transformation, and details on how it has been decided upon and completed. This will help to better understand the logic of transformation to be presented in a transformational pattern. In a less favorable case, only the business activities before and after transformation are known, but not how the decision has been made. In such a case, we can just make logical analysis when comparing FEM before and after transformation, deriving the rationale behind the transformation based on the analysis and imagination. Whether this rationale corresponds to the decision making in the case or not is less important, as long as our analysis can result in defining a transformation pattern and finding other examples of its application. In the following sections, we will analyze both types of transformation cases.

5 Reverse Engineering of a Transformational Case

In this section we will use an example from [13] to demonstrate our approach for analyzing examples/cases. The example corresponds to the transformation completed by Amazon when it created a new business – Amazon Web Services (AWS) - in 2006. The new business was created [19] to complement the main business of selling books over Internet. As we do not know the exact rationale behind this transformation, we just create the relevant fragments of FEM before and after transformation and analyze the relationships between the elements of these two FEMs.

The two FEMs and the relationships are presented in Fig. 2. The FEM fragment that corresponds to book sells over the Internet is represented on the left-hand side of Fig. 2; actually, it repeats the fragment depicted in Fig. 1. The right-hand site of Fig. 2

depicts a FEM fragment related to the new business – AWS. The (green) dashed arrows between the elements of these two FEM's show the relationship between these elements, the labels on the arrows indicating the difference between these elements.

The main idea of the transformation is to lease the internal IT platform to external customers. Thus, the cornerstone of the transformation is the asset *IT platform for workshop deployment*. This asset represents also a capability of maintaining such a platform, including updates, scaling up the power, etc.; this capability is represented by the whole graph that "hangs" on this asset, see Fig. 2. This asset plays a role of *Technical & informational infrastructu*re in the *Books sales* process, which is indicated by the label on the arrow going from this asset to the process. In the new FEM it also plays the role of infrastructure, but the nature of the main process is completely different – platform as a service. Though we do not know the actual rationale of the transformation, we can imagine one as follows:

> *We are maintaining a powerful IT platform for our own needs, and the market for IT platform as a service is growing. We could use our platform capability to enter this market.*

The FEM on the right hand side of Fig. 2 is built based on this rationale. To ensure that the transformation is feasible, we need to compare the new FEM with the old one to see what assets and capabilities from the old business could be employed in the new one. The analysis follows below:

- *Beneficiary* (customers). Customers for the new business are different from the old one. However, the decision makers of the new kind of customers, can be assumed to be book readers (at least some of them), which increases the chance that they have used the original main process for buying books. In this case, they have the first-hand experience of using Webshop software and the IT platform behind it. Therefore, a reputation of having software that is reliable and efficient can be used for acquiring new customers. This thinking is depicted in Fig. 2 by using the reputation asset from the left-hand side of Fig. 2 to underpin the sales process in the new business.
- *Technical & informational infrastructure - the platform itself*. For the book sales business, the infrastructure needs to contain only the components, e.g. database engines, webservers, etc., that are used by the Workshop software. For the AWS business the platform should be more general, e.g. include different database engines. This is depicted as a label that connects the infrastructure assets from two FEMs in Fig. 2.
- *Managing processes for IT platform*. The processes for maintaining the *IT platform* asset - *Acquire*, *Maintain* and *Retire* - also have different characteristics. For the internal use, acquiring a new platform and dismantling the old one can be planned in advanced and performed at regular speed. For the new business, this should be done quickly, i.e. as soon as the customer defines what platform is needed, it should be immediately created. When the customer no longer needs the platform, it should be dismantled to free the resources. This is depicted in labels that connect the managing processes in two FEMs.

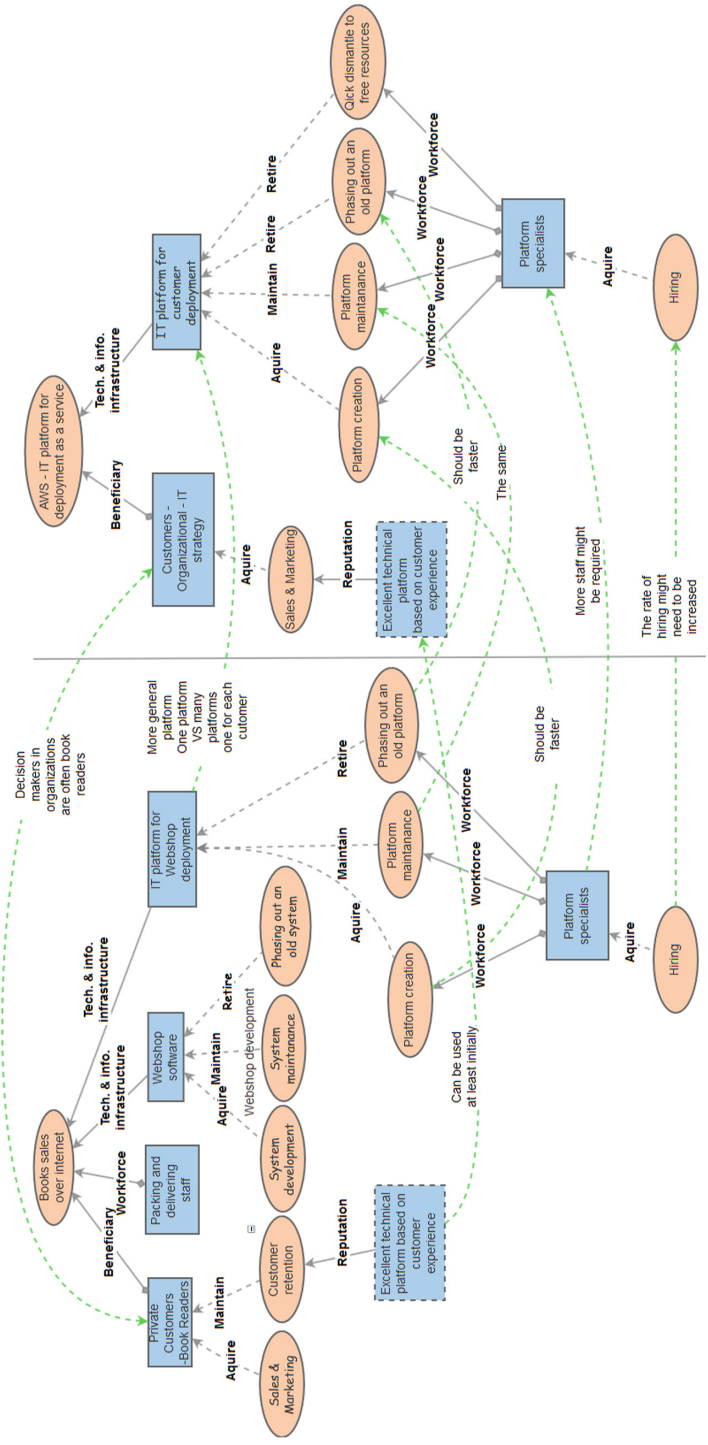


Fig. 2. Analysis of AWS transformation (Color figure online)

- *Assets needed for managing platform* – workforce. The discussion on the differences between two FEMs could continue further by unfolding the FEMs downwards. For example, at the next level the capacities of hiring platform specialists could be evaluated with the result that this capacity might need to be increased for the new FEM, which is depicted in Fig. 2.

6 The Structure of Transformational Patterns

In this section we will use the example from the previous section to show how a transformational pattern can be derived from a business case of BMI, and discuss the structure of transformational patterns. The basic idea behind BMI in Fig. 2 is based on the following reasoning:

- We have an *infrastructure* that supports our main process that could be of use for other processes if made more general
- Part of our current customers might be interesting to use our infrastructure if the latter is generalized
- These customers have positive experience of our infrastructure, though indirect.

A transformation pattern that corresponds to this thinking could be presented as Fig. 3. It consists of two FEM fragments, (green) labeled arrows between the elements of these fragments, and green hexagons with label inside that are connected to the elements of the left-hand FEM fragment. We can differentiate two parts in a pattern - formal and informal:

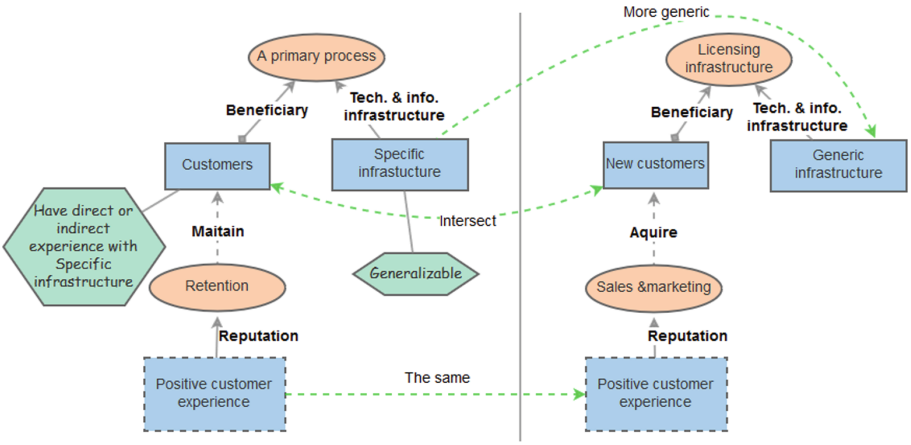


Fig. 3. An example of transformational pattern (Color figure online)

- The *formal* part is a part that can be applied by an algorithm to an existing FEM without any human participation. It includes (1) shapes for processes and assets in both FEM fragments, but not the labels in them, (2) links between these shapes,

including labels, inside each fragment, and (3) green links between the elements of the two FEMs, but not their labels.

- The *informal* part is a part that needs human beings for interpretation. It includes: (1) hexagons connected to the elements of the left-hand FEM that express some conditions on these elements (2) labels inside elements of FEMs that express the semantics of the assets, and (3) labels on the (green) arrows that connects the elements of two FEMs that express the difference between the corresponding elements.

Application of a transformation pattern to an existing FEM is as follows:

1. A place in a FEM graph is found which matches the formal part of the left-hand FEM fragment of the pattern (labels inside shapes excluding)
2. The informal conditions expressed by (green) hexagons and semantic labels inside the pattern's shapes are checked by experts
3. If the conditions are considered as being satisfied, a new FEM is generated and attached to the old one with (green) arrows
4. Conditions included in the labels on the (green) arrows between the new and old FEMs are checked by experts
5. If conditions considered as being satisfied, the process of adding details to the new FEM continues.

Figure 2 can be used as an example of application of this procedure to the FEM in Fig. 1. Note that the pattern from Fig. 3 can be applied not only to the place where *IT-platform* matches *Specific infrastructure*, but also to the place where *Webshop software* matches *Specific infrastructure*. In this case, the new FEM would represent licensing the Webshop to other book sellers. In fact this transformation has also been completed by Amazon.

7 Another Example of a Transformation Pattern

The example of transformational pattern presented in Sect. 6 is more or less straightforward, as the idea behind it is simple – having excellent infrastructure that supports a main processes, the company can decide on leasing this infrastructure to others. This is why the pattern could be constructed based only on facts, without knowledge on the details of decision making. In this section, we present another example, which is not that straightforward and which has been built based on the information on the decision making process of the company that has completed this type of transformation.

The company in question is an American consulting company called Prolifics, for which the first author worked for a short period of time in the middle of 1990th. At this time, the company was a vendor of a high-level software development tool, also called Prolifics, which could be counted as a kind of 4GL tools. The main business was selling licenses of the tool and providing product support. Beside this, there was a consulting business providing experts in Prolifics for various software development projects that used this tool. By the end of 1990th the market for 4GL went down due to the switch to the web and low-level programming (e.g. Java), and in the beginning of 2000th, the company needed to radically change its business in order to survive.

Such a change was successfully completed, and the company reemerged as a consulting business providing expertise in IBM's Websphere. Understanding the rationale behind such a transformation requires some insights in the internal decision making. These were provided by Prolifics' CEO, who explained that the internal changes were minimal. The sales and marketing of the new consulting service was done according to the same routines that were established for consulting arranged around the company's tool (Prolifics). The consulting service itself also used the already established routines. The main difference was that instead of expertise related to the tool Prolifics, the company started to provide expertise related to Websphere. At the same time, activities related to the tool Prolifics where drastically cut off. The support for the tool, and some consulting related to it, remained, but without any marketing and sales activities. The product development went down to bug fixing and releases connected to changing operational environment. The developers where redirected to be Websphere consultants.

The main factor that allowed to successfully complete the transformation above was that Prolifics had interface (API) to a number of third parties products, including Websphere. Therefore, some of their consultants and product developers already had expertise in Websphere. Moreover, a number of their customers had both Prolifics and IBM products working together, as well as some of the consultants providing expertise in both. In addition, the company had a partnership agreement with IBM.

Based on the analysis of the information provided by CEO, we created a transformational pattern in Fig. 4 by abstracting from details specific for Prolifics. In several places in Fig. 4, a connection between an asset and a process labelled ExT is used. ExT stays for Executable Template, and it is used for linking to the process any asset that affects the behavior of process instances that belong to this process. In the first place, ExTs have a form of description of working routines, process maps, policy documents, etc. However, an EXT can be any other asset that affects how the process is executed. In particular, the asset *Own software product* is used as ExT in both main processes in the left-hand FEM in Fig. 4. In *Product Licensing* an installation package is used for copying and sending to the customer, or downloading. In *Consulting*, the product affects the area of the customer project that engages the consultants.

Note also that the two main process in the left-hand FEM are not independent, the consulting process serves for retention of customers that license the product; also the set of customers for consulting is a subset of the customers who have bought a license.

On the whole, the pattern in Fig. 4 represents the main idea of Prolifics's transformation – amending the consulting business by substituting its own product to a third party product. The pattern expresses also the conditions when such transformation is feasible. For example, to have a solid customer base for this kind of transformations, a substantial part of the existing customer base should not only have the company's own product installed, but also the third party product. Moreover, a substantial part of their consultants needs to be experts in both products, and their expertise in the third party products should be revealed to the customers that have both products.

Note that Prolifics, as a company, still exists, but at the end of 2000th, it was sold to an Indian concern, and might have changed the line of its business. The example of the transformation refers to the events of the early 2000th. The authors are grateful to Robert Ismach, its President at the time, for providing the internal insights.

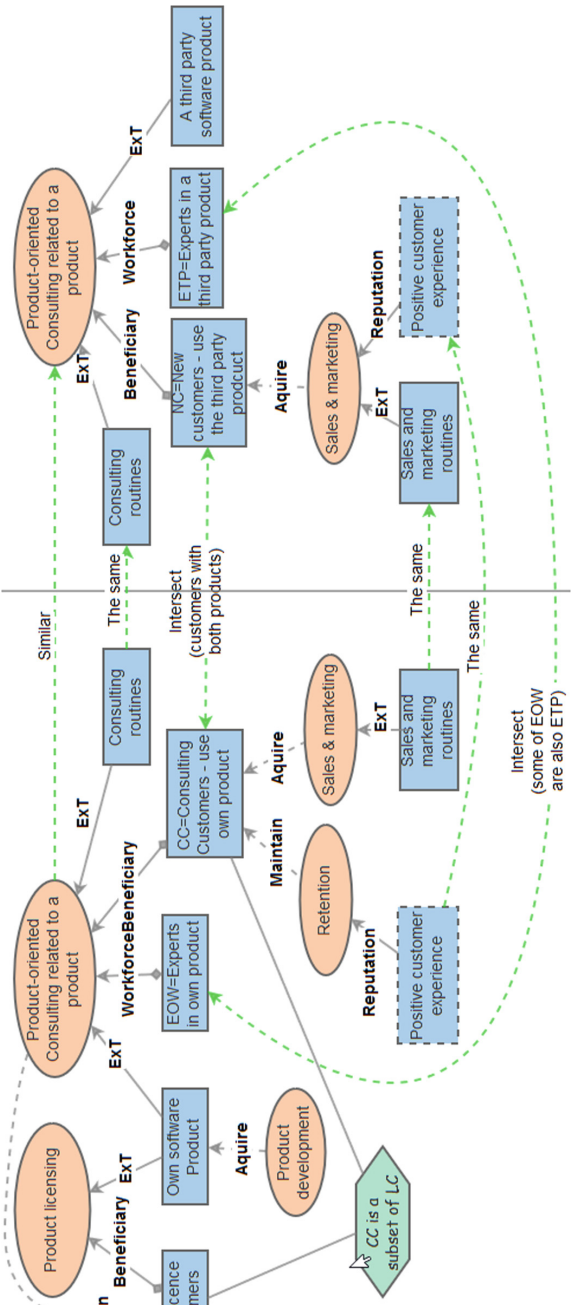


Fig. 4. From a tool vendor to providing expertise in a third party product (Color figure online)

8 Concluding Remarks

According to the literature reviewed in Sects. 1 and 2, the needs for the enterprises to innovate their business models is expected to grow, particularly in connection to advances in technology, and especially in regards to digital transformation. In particular, in this environment, the needs for industrial model innovation is expected to grow, as the companies might be forced to rethink their line of business. As the same literature shows, there is no systematic way for an existing business to conduct this kind of BMI. Currently, it is being done on the intuitive, tacit level with running a risk of failure or great costs, or missing a better opportunity. Management consultants, with whom we discussed our ideas, also confirmed the lack of tools for systematic industrial model innovation.

This exploratory study represents a major step on the way of filling the above gap in theory and practice by providing an innovative approach for BMI that takes into consideration existing capabilities of the enterprise (in form of both processes and assets). Though this approach is based on the known principles - experimentation, analogy (best practices) and patterns - it combines these principles in a new way. The contributions of this paper include:

- Clearly identifying the gap in practice and research that needs filling
- Suggesting an approach to fill the gap by a library of FEM-based transformational patterns
- Suggesting a structure for transformational patterns
- Suggesting and demonstrating an approach for creating new patterns.

Our immediate plans regarding transformational patterns are related to creating and disseminating a library of patterns. Creating a transformational pattern includes finding a good example, abstracting from details and creating a transformational pattern based on this example, and finding other examples that correspond to the constructed pattern. Dissemination can be arrange by creating a website where such patterns are published, giving access to it to any researcher or practitioner who wants to add a new pattern, thus making the library a collaborative work.

More long-term plans include creating a computerized tool that facilitate experimentation by automatically finding places in a FEM model which can serve as a basis for a pattern-based BMI. We also plan to extend our framework so that it can help not only in finding a promising transformation using the pattern library, but also in planning the completion of such a transformation. This requires the development of a set of qualitative and quantitative properties that could be assigned to the nodes of FEM [13]. Planning then can be done based on comparing characteristics of elements of the FEM before transformation with the corresponding elements of the FEM after transformation. We are also planning to investigate attaching other frameworks and method for developing a more detailed plan. Here in the first place, we are looking at pattern of strategy suggested in [4] trying to find a synergy between two types of patterns. The connection between them seems to exist when considering specific examples, e.g. Prolifics' pattern of strategy was changed from coupling to a market to coupling to a bigger partner.

References

1. Andreini, D., Bettinelli, C.: Business Model Innovation. ISAMS. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53351-3
2. Mangematin, V., Ravarini, A., Sharkey Scott, P.: Special issue: business model innovation. J. Bus. Strategy **38**(2) (2017)
3. Giesen, E., Berman, S.J., Bell, R., Blitz, A.: Three ways to successfully innovate your business model. Strategy Leadersh. **35**(6), 27–33 (2007)
4. Hoverstadt, P., Loh, L.: Patterns of Strategy. Taylor & Francis, London (2017)
5. Altshuller, G.: The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity. Technical Innovation Center, Hagerstown (1999)
6. Rumble, R., Minto, N.A.A.: How to use analogies for creative business modelling. J. Bus. Strategy **38**(2), 76–82 (2017)
7. Osterwalder, A., Pigneur, Y.: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Wiley, Hoboken (2014)
8. Echterfeld, J., Gausemeier, J.: How to use business model patterns for exploiting disruptive technologies. In: 24th International Conference on Management of Technology, pp. 2294–2313 (2015)
9. Boyd, J.R.: Destruction and creation. Lecture presented to the U.S. Army Command and General Staff College (1976)
10. Bider, I., Perjons, E., Elias, M.: Untangling the dynamic structure of an enterprise by applying a fractal approach to business processes. In: Sandkuhl, K., Seigerroth, U., Stirna, J. (eds.) PoEM 2012. LNBIP, vol. 134, pp. 61–75. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34549-4_5
11. Bider, I., Perjons, E., Elias, M., Johannesson, P.: A fractal enterprise model and its application for business development. Softw. Syst. Model. **16**, 663–689 (2016)
12. Henkel, M., Bider, I., Perjons, E.: Capability-based business model transformation. In: Iliadis, L., Papazoglou, M., Pohl, K. (eds.) CAiSE 2014. LNBIP, vol. 178, pp. 88–99. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07869-4_8
13. Bider, I., Perjons, E.: Using a fractal enterprise model for business model innovation. In: BPMDS 2017 RADAR, CEUR, vol. 1859, pp. 20–29 (2017)
14. Chesbrough, H.: Business model innovation: opportunities and barriers. Long Range Plan. **43**(2), 354–363 (2010)
15. Hevner, A., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
16. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–78 (2007)
17. Baskerville, R.L., Pries-Heje, J., Venable, J.: Soft design science methodology. In: DERIST 2009, pp. 1–11 (2009)
18. Bider, I., Johannesson, P., Perjons, E.: Design science research as movement between individual and generic situation-problem–solution spaces. In: Baskerville, R., De Marco, M., Spagnoletti, P. (eds.) Designing Organizational Systems. LNISO, vol. 1, pp. 35–61. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-33371-2_3
19. Amazon: Amazon Web Services (AWS). https://aws.amazon.com/

Digital Transformation

Understanding the Transformation Towards Industry 4.0

Kristin Vogelsang[1], Sven Packmohr[2(✉)], Kirsten Liere-Netheler[1], and Uwe Hoppe[1]

[1] Institute of Information Management and Information Systems Engineering (IMU), Osnabrück University, Osnabrück, Germany
{kristin.vogelsang, kirsten.liere-netheler, uwe.hoppe}@uni-osnabrueck.de

[2] Department of Computer Science and Media Technology (DVMT), Malmö University, Malmö, Sweden
sven.packmohr@mau.se

Abstract. The ongoing process of digital transformation in manufacturing – known as Industry 4.0 - hauls fundamental change. The whole value chain of enterprises is affected. As the digital transformation of businesses is still ongoing, many enterprises struggle with the challenges arising. This paper aims to show these struggles but also to contribute by analyzing how enterprises are transforming. We take a phenomenological view of the ongoing transformation. To get in-depth insights, we conducted and analyzed 18 interviews with 10 companies. For most companies, the digital transformation starts in operations with the vision of building a smart factory. Other primary and support activities also need to transform. These essential changes lead to restructuring and extensions of the strategy of manufacturing companies. Following these changes, companies will not need to choose either cost advantage or differentiation as a strategy but instead can do both.

Keywords: Industry 4.0 · Value chain · Qualitative research · Porter Phenomenology

1 Introduction

A keyword conquers the field of the working world: The so-called future project Industry 4.0 was launched by the German Federal Ministry of Education and Research in 2011 [1]. One implication of Industry 4.0 is the overall digital transformation and cross-linkage of the value creation process [2] as information and communication technologies merge with production processes. The digital transformation (DT) in manufacturing affects the whole business. Research regarding the entire enterprise and implications for the complete value creation is still rare [1, 3]. A deeper understanding of how the whole value chain is affected by Industry 4.0 is missing [4]. The transformation towards industry 4.0 is risky and is related to high investments in technologies that are still regarded as new and unpredictable [5]. Many enterprises struggle with this transformation [6]. This paper aims to contribute to the research about the

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 99–112, 2018.
https://doi.org/10.1007/978-3-319-99951-7_7

effects of the digital change in manufacturing. We take an approach from a phenomenological [7] perspective on the ongoing changes. This approach allows us to better focus on the phenomenon of DT from different views of a variety of experts working in the field. As DT is implemented in different usage scenarios, we expect this approach to be useful to understand Industry 4.0. For this reason, we use a qualitative research approach to give in-depth insights into how Industry 4.0 changes the value creation processes. Moreover, we show associated strategies and recent goals of the examined companies.

We will introduce the term Industry 4.0 and show actual research on digital transformation in the next chapter. The well-known value chain by Porter [8] is presented in chapter 2.2. We use Porter's work to outline the impact of Industry 4.0 on the different activities within companies. On the base of 18 in-depth interviews, we describe the changes that occur due to DT. The implications for each activity according to the underlying model will be presented and discussed. We aim at detecting the effects of Industry 4.0 and align the findings to the basic model given by Porter. Finally, we describe limitations and provide an outlook of possible additional future research about the effects of Industry 4.0 on the value chain. We consider this work to be of value to researchers as well as practitioners. The first group can deduce need for further research. The second group may use the insights to assume action plans for the implementation of digital transformation in their businesses.

2 Background

2.1 Industry 4.0

Industry 4.0 is regarded as a kind of vision, a new kind of industrialization. It is characterized by three main dimensions [9]: "(1) horizontal integration across the entire value creation network, (2) end-to-end engineering across the entire product life cycle, as well as (3) vertical integration and networked manufacturing systems". An ideal value creation process in Industry 4.0 is marked by digitalized, cross-linked processes between companies and end-users from the raw material to the final product across the whole product lifecycle. "The intelligent cross-linking and digitalization covers the application of an end-to-end solution using information and communication technologies which are embedded in a cloud"[9].

Basic innovations that support Industry 4.0 are cyber-physical production systems (CPPS) and the Internet of Things (IoT) [10]. Exchange of data is achieved by "seamless, ubiquitous sensing, data analytics, and information representation with cloud computing as the unifying framework" [11]. The modern (and future) production will take place in so-called "smart factories" [9]. A smart factory is defined as a "factory that context-aware assists people and machines in the execution of their tasks" [12]. Smart factories apply various technologies of ubiquitous computing and are autonomously controlled [13].

Industry 4.0 is a keyword for a goal, which manufacturing companies are working on. Technological solutions regarding the realization of digitalizing production processes are diverse. These solutions are often studied in engineering disciplines which

hardly give a holistic view on the topic. However, Industry 4.0 is said to transform whole value chains [1] which leads to a paradigm shift in business and focuses on the organization, rather than technological challenges [14].

In the context of organizational changes triggered by Industry 4.0, we identified major research streams that deal with the changing processes and their implications. The field of digital business strategy [5] and the embedding of the IT development is of particular interest. During the 21st century, digital strategies shifted from a decentralized hierarchical functional structure to an IT-enabled global network structure [15]. Technologies, like big data, [16] and the more competitive environment [17] change the requirements for firms and thus their digital strategies. Moreover, the new role of customers also influences the strategic design [18]. A change of the whole working environment is expected [9, 19]. While Brynjolfsson [20] predicts a pessimistic view with high losses of jobs, moderate approaches [21, 22] also exist. Furthermore, there is much research about the changing role of services [23–25]. Many researchers concentrate on distinct technologies or departments to describe the changes. As Industry 4.0 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. There is more research about the future potential of Industry 4.0 [14, 26, 27] rather than results from enterprises trying to cope with that challenge [6]. This situation leads to a critical research gap as organizational settings and strategies need to adapt to new ways of working.

2.2 Porter's Value Chain

To map value-creating activities Porter's value chain [8] is a widely-used concept. It was developed in the 1980ties and is still of great value to outline the changes in value generation [28]. It combines the activities (processes) within an enterprise with the strategy the enterprise follows [29]. Within the concept nine generic categories of activities are isolated. These activities are technologically and strategically distinct from each other. The aim is to identify activities which could lead to competitive advantage and thereby support the strategy of a company [30]. Porter distinguishes different generic strategies such as *Cost Leadership* and *Cost Focus* as well as *Differentiation* and *Focused Differentiation*. For the first pair, value creation activities have to focus on cost advantages. For the second pair, they have to focus on offering a distinction to customers [8].

Porter identified five primary activities which are: **Inbound logistics** – activities to receive and handle resources internally; **Operations** – transformation activities to convert resources into products or services for customers; **Outbound logistics** – activities to deliver and handle products or services going out to customers; **Marketing and sales** – processes to persuade customers to purchase products or services; **Service** – after-sales activities to keep up the value of products or services purchased.

Secondary activities span horizontally over the primary activities to support each of the primary activities. They are: **Procurement –** activities to secure the supply with resources; **Human resource management –** activities to recruit, train, motivate and bind employees; **Technological development –** activities to stay up-to-date with technologies used; often related to IT technologies; **Infrastructure –** administration

activities to secure daily operations such as accounting, legal, administrative, and general management.

Even though Porter describes only one value chain, activities take place over several independent but connected value chains [8]. Other researchers and practitioners felt the value chain is missing a detailed customer as well as time perspective which were introduced later [31]. Porter was the first to describe a comprehensive model on company-level to specify cost and differentiation strategies based on an activity-based view [32].

One of the advantages of digital transformation is the possibility of value creation in business networks or ecosystems. As such, significant research streams arise in this area [33–35]. One topic of interest is the growing importance of the integration of customers and suppliers in value creating processes. The focus moves from one single company to whole networks [36] where the use of digital services is critical to creating value [37]. This possibility is not explicitly mentioned in Porter's value chain as it represents a description of a single company [38]. However, as we aim at a better understanding of Industry 4.0 by highlighting changes in single processes of companies, we rate it as a framework applicable for further analysis. We will not analyze value creation in-depth in our work. Instead, take a closer look at actual practical changes within different activities.

3 Research Design and Introduction of the Sample

To examine changes in the value creation process resulting from Industry 4.0, we used a qualitative research approach. Though quantitative methods in IS research are prevailing [39, 40], a qualitative research process is regarded as suitable for the examination of complex research fields [39, 41]. Our research was mainly influenced by phenomenology [42]. As we aim at gaining a deeper understanding of the essence of the changes that occur in the context of Industry 4.0 [43]. According to the phenomenological approach, we show how the digital transformation in the context of industry 4.0 is understood and experienced by different people. This approach originates from the social sciences in the 1950s and helps in deducting and interpreting results [44]. Although it is rarely used in IS research, we consider it as a very useful view on complex constructs. The merger of differing views gives a broad picture of the transformation. "The results reflect a careful description of precisely the features of the experienced phenomenon as they present themselves to the consciousness of the researcher" [45]. Also, we borrowed from Grounded Theory, as it gives a cluster of guidelines and techniques for the research in social reality [46, 47]. We did not develop a theoretical framework from the data collected [47], which is the aim of Grounded Theory. Instead, we used advice for sampling and coding from Grounded Theory. For our analysis, we used techniques provided by Mayring [48]. The text analysis according to Mayring can be regarded as a valuable method to receive explanations of a complex phenomenon [49]. We do not count or rate the statements. We collect them and try to align them with the context of the study. Thus, we contribute to the ongoing debate on the meaning, role, and implications of Industry 4.0.

Table 1. Description of the interviewees

No	*Description of the company*	*Interviewees*
01	*Manufacturing:* More than 7000 employees in 38 countries; customer demands for traceability gave the initial trigger for Industry 4.0. The DT started in operations, a complete vertical integration followed	01-01 Head of Production Intelligence 01-02 Development Engineer
02	*Manufacturing:* Machine supplier; in over 130 countries, more than 250 employees; DT since 2007/2008; started with projects to integrate the customer for more service	02-01 Head of Production 02-02 Head of Service Department
03	*Manufacturing:* Electrical engineering, 9 countries, family company, SME; DT since 2012, started at operations with connective products, a cloud-based platform for customers followed; additionally, connection with the supplier for automated delivery of purchased parts	03-01 Chief Technical Officer
04	*Manufacturing:* More than 1000 employees; DT came up with machine supplier, offer of digital connections, work on network engineering to support smart production processes, implementing a fully automated (general) production line to produce more variances	04-01 Process Optimization and Automation Department 04-02 Control Programmer
05	*Agricultural engineering:* More than 500 distribution facilities worldwide; started in automated production, major goal is the increment of product variances. Customers and suppliers are interconnected, enhancing online services via a digital platform	05-01 Head of Quality Management
06	*Agricultural engineering:* With sites all over the world. Online platforms are used to communicate with suppliers and offer additional digital services for customers. The enterprise works continuously on smart products for the customer	06-01 Product Development
07	*Automotive:* Supplier, more than 4000 employees; DT as a central goal of operations, started by customers wanting more variances in smaller charges, aim is extension of self-adaptive systems.	07-01 Head of R&D
08	*Automotive:* Leading car manufacturer with employees worldwide; DT started from 2015; major goal: machine networks and big data aggregation for improved sales	08-01 Teamleader Purchasing 08-02 Industry 4.0 Department
09	*Automotive:* Leading car manufacturer with employees worldwide; development of smart production in 2010, mainly in self-adapting systems and robot support	09-01 Electronical Engineering Department
10	*Automotive:* Leading car manufacturer with employees worldwide; DT started for process consistency to link processes in operations, network-production, and self-adaptive robot support, strategy: security increment, rollout of integrated smart production; improving of HCI	10-01 Automation Planning Department 10-02 Smart Production Manager

According to [50] data, collection and analysis are interrelated processes. For this reason, we developed a sampling strategy to receive answers from many different perspectives and to iteratively generate the database. The interviewees were identified mainly in social network groups dealing with DT. Some interviewees directly contacted the researchers because of former work and the project website. We pre-checked the suitability of the interviewees for this study by asking for information about current digitalization projects in the companies. If the projects had a significant impact on the value creation process of the firm, the interviews were conducted. Table 1 presents the companies and interviewees of the study. We aimed at a heterogeneous sample with interviewees of many different positions and with a variety of experiences to gain broad insights.

The interviewees answered in free speech, without being biased by a framework. We asked the interviewees to describe the experienced transformation process in detail to ensure that DT specific changes have been undergone. The interviews were conducted in German and afterward translated by the authors close to the original content. For the data collection, we conducted 18 semi-structured interviews in 10 German enterprises. The interviews were held to gain information about the enterprises observed and the process of change triggered by Industry 4.0. We chose a sample of a combination of three industries: mechanical engineering (no. 1–4), agriculture (no. 5–6) and automotive (no. 7–10). They belong to the dominating industries in Germany [19]. Moreover, automotive and the mechanical engineering industry is regarded as the most driving ones' for Industry 4.0 [51].

The underlying data were revised several times to identify patterns and similarities as well as deviations. In a first step, we followed the idea of "open coding" [47] using a qualitative content analysis according to Mayring [48] to identify the relevant changes. We adjusted the research for a simultaneous collection of data and analysis following a grounded theory approach. This led to a process of iterative revisions until no more new ideas arose [46]. As a second, selective step, we used Porter's model [8] as a frame to align the activities, to unify the findings and to deduce detailed descriptions of the departments and strategies. This allows us to allocate the statements of the interviews to the activities according to Porter.

4 Results

The digital change affects the primary as well as the supporting activities of the whole value chain. Table 2 shows where the interviewees experience the influence of Industry 4.0 and how the change is organized. Operations seem to be the principal department for the integration of Industry 4.0. Many interviewees describe it as the initial point for the digital change (marked in Table 2 by *). On the one hand, this can be attributed to the topic and the sample of the study. On the other hand, it also presents the central role of operations where customer needs and technical capabilities meet. The change in production is often initiated by customer demands. The ability to create more variances and more quality data are of central importance for the digital change. *"[…] we print identification labels; you find the product information which is on the products there […].* These are recorded at management level and are passed down to the shop floor

Table 2. Description of the Industry 4.0 activities

Activity	Value chain	Phenomenon from the cases	Cases
Primary	Inbound logistics	Automated order and delivery of purchased parts	1, 2, 3, 4*, 8, 10
	Operations	Smart production lines enable more variances, lower failure rates and real-time execution	1*, 2, 3, 4*, 5 *, 7*, 8*, 9*, 10*
	Outbound logistics	Load processes can be monitored in real-time	7, 10
	Marketing and sales	Customer platforms with direct interfaces for product selection and order management	2, 3, 4, 5, 7, 8, 10
	Service	New or enhanced digital platforms to manage custom data and services	2*, 3*, 4, 6, 7, 8, 10
Support	Firm infrastructure	Industry 4.0 became part of the strategic orientation. Digital integration eases quality management and leads to higher demands	1, 2, 9, 10
	HR management	Agile techniques and more training in IT competencies	4, 7, 9, 10
	Technology development	Development of IT architecture to support smart factory (integrated systems). Central integration of IT and data management. Development of interfaces	2, 3, 4, 6, 7, 8, 9, 10
	Procurement	Digital accounting and purchasing opportunities	1, 2, 4, 5, 7, 8, 9

level witho*ut anyone involved. So, this is totally vertical integrated" [01-01].* Only one enterprise reports a supplier-initiated change of production, as the supplier offered smart production equipment (enterprise 4). Often, the digital transformation was triggered within pilot projects. Many interviewees outline the realized gains of the pilots. In the long view, many enterprises work on extensions of the linked networks and smart production. Restructuring is often regarded as a challenge because of more consolidated companies that have to be linked inter-organizationally. The second key department for the digital change is the service department. Digital products and customer platforms trigger change towards industry 4.0.

In the following we will describe the changes in the value creation process deduced from the results of our qualitative research:

Inbound Logistics. The supplier-relation is a major issue for the investigated companies. In one enterprise this was the initial point of digital production processes. Data should be exchanged automatically to improve work, optimize inbound processes and lower stocks. Thus, the interviewees see a lot of opportunities here and have already implemented these into the factories. *"The shelf itself just reads the code of the empty bin, and the information is sent via a mobile connection, this is sent directly to the supplier. So, he receives this message that our shelving unit is empty and organizes the refilling autonomously" [03-01].* Problems of inefficient communication and high

stocks can be solved. Inventory postings happen at real-time. The transparency at the beginning of the value chain has increased. *"So, I know exactly how many parts he has in stock and how fast he can deliver" [10-02].* Connections also influence operations activities positively. *"[…] the electronic delivery notes can more easily be integrated into the ME. The electronic import works without somebody interacting. This saves time and lowers the number of errors" [01-02].*

Operations. Operations is often regarded as the main changing activity of Industry 4.0. In 7 of the 10 enterprises, it represents the initial starting point. The main vision is to come to smart factories which need a connection at all levels. *"Thus, connecting our single components with each other. Then specifically collecting data which we could collect. We have a lot of thoughts about how to evaluate and use our data" [08-01].* Many different technological solutions, as well as benefits, exist. *"We have different actions taken in operations regarding the connection of units, for the documentation units, and for production processes. The ERP system with appropriate connectivity, development of shelves connected by internet technologies and similar things" [03-01].* The automated integration of units eases the traceability of intermediates and goods. The natural border between operations and outbound logistics activities can be overcome by system communication. Furthermore, automated inventory entries and withdrawal postings directly connect operations with inbound logistics.

Outbound Logistics. Digital tags enable the measurement of loading times. The customer receives real-time responses on loading progress and residence times of the products: *"I see the benefit especially for the customer where whole logistics processes will be automated and simplified by Industry 4.0" [04-02].* This traceability can also be part of the contracts, as the suppliers guarantee firms (short) charging times.

Marketing, Sales, and Service. The interviewees describe new ways of customer contacts and interfaces with customers. The customer is more integrated into the production process, as products can be directly ordered in diverse variances. *"We want to give the customer the opportunity to influence the product, to individualize" [10-02].* Additional information about the products is delivered, so that the customer has more opportunities to trace the goods. Marketing and sales can go beyond the pure selling of a product by staying connected and offering value-added services. Hybrid products are becoming more important. *"Other business models besides selling cars need to be considered" [10-01].* During after sales a more active approach in maintenance is possible due to the connection of the delivered product. They are often regarded as additional business areas. In particular, the cases from the agriculture sector show how the customer relationship can be shaped by the digital linkage. *"They sell the service and afterward another digital benefit and again the service" [05-01].*

Firm Infrastructure. A characterizing feature of Industry 4.0 is the impact on the whole enterprise. Nearly all interviewees describe that the digital change was led by strategic decisions. Thus, the firm infrastructure is affected. Most interviewees mention improvements regarding quality management, as more data about components are available. "*It is a part of the re-traceability. We now know what materials and components are used for each product. We can break it down into any part of the end product using the tags" [01-02].* Furthermore, automated process interlocking prevents

failures and rejections. The interviewees detect a change towards a more technology-friendly business strategy within the enterprises.

Human Resources Management. HR has to adopt to the skills and working habits needed in a software company instead of a purely engineering company [29]. More IT competencies will be necessary to be successful. *"If you are looking for new maintenance stuff, you need to train the personnel. This is also a shift in the training profession" [09-01]. "We need more high-qualified staff" [10-02].* Due to a need for more prototype working, *"agile working is becoming more important" [10-01].* This does not only affect factories but also employees in offices. *"I also need connected devices in the offices" [09-01].* Industry 4.0 will not only change the need for technology skills but will also lead to a change in the education of workers. The ability to use and integrate technologies will be in the focus of future education.

Technology Development. Technology development plays a crucial role in times of Industry 4.0. The IT architecture needs to become "ready" for smart factories. This means advanced production systems are necessary and IT systems need to be integrated. *"The big challenge will be to summarize everything, to bundle and to make it consistent and this is a lot of work for the IT department" [10-03].* The IT department is an enabler for Industry 4.0. *"We discuss a lot about different protocols, data, ways of connectivity, interfaces to different cloud providers and also internet data" [08-03].* They need to develop new ideas and work in a testing environment. *"There are many prototypes and proof of concepts which are partially already in the line" [10-02].* The interviewees emphasize a clear upgrade of the technology development. The application and successful integration of technology will be a matter of future sustainability for many enterprises.

Procurement. Procurement is also influenced by Industry 4.0. Via digital interfaces, some parts of the procurement are strongly related to the inbound logistic process. Purchased parts can be automatically delivered and also settled. This process is consistently linked with the order processing. Trust is an important characteristic of this process. *"We want to connect our suppliers to our components to provide them data. Why? Because they most often know better how to evaluate the components than we do" [08–03].*

5 Discussion

Our results show that the activities marketing and sales, as well as service, move to the fore of all processes which is following results from value generation by networks [34]. Industry 4.0 enhances the possibilities of gaining advantages through cooperation and can such contribute to value co-creation of companies. Our study shows that the system implementation was in many enterprises done due to customer demands. Moreover, customers are more and more integrated into the activities of the firm. This also leads to a permanent exchange of information between marketing and sales as well as service. They become closely linked [12]. Therefore, we suggest rethinking the value chain of Porter by putting these activities in focus as central activities shown in Fig. 1. The

activities of Porter's value chain cannot be considered in the given order. Data are transferred consequently during production processes and the activities mutually influence each other. This leads to the customer at the centre of all activities. The change of the value creation process is often pushed by customers. It is important to note that the semi-digitalized structure of the core products enables the integration of the customer into the production process.

These essential changes strongly influence the strategic focus of the enterprises. Almost all interviewees mention the key strategic objectives once formulated by Porter: cost and differentiation strategies. Porter's approach is implicitly building upon a dichotomy between these two. Information and communication technology (ICT) might help to bridge this dichotomy. Today a focus on costs doesn't necessarily mean that differentiation cannot be achieved or vice versa [52]. Differentiation strategies mainly aim at the capability to produce more variances in smaller quantities. *"We will be able to produce more variances of complex products in less time" [01-01].* Smart production lines enable the self-adapted steering of machines to produce small charges. As the products and the machines permanently deliver data in real-time, quality management can be improved. These enhancements allow following the strategy of cost reduction. *"We will have different products from the factory in a messy sequence with the same costs as with the actual mass production" [01-01].* Companies

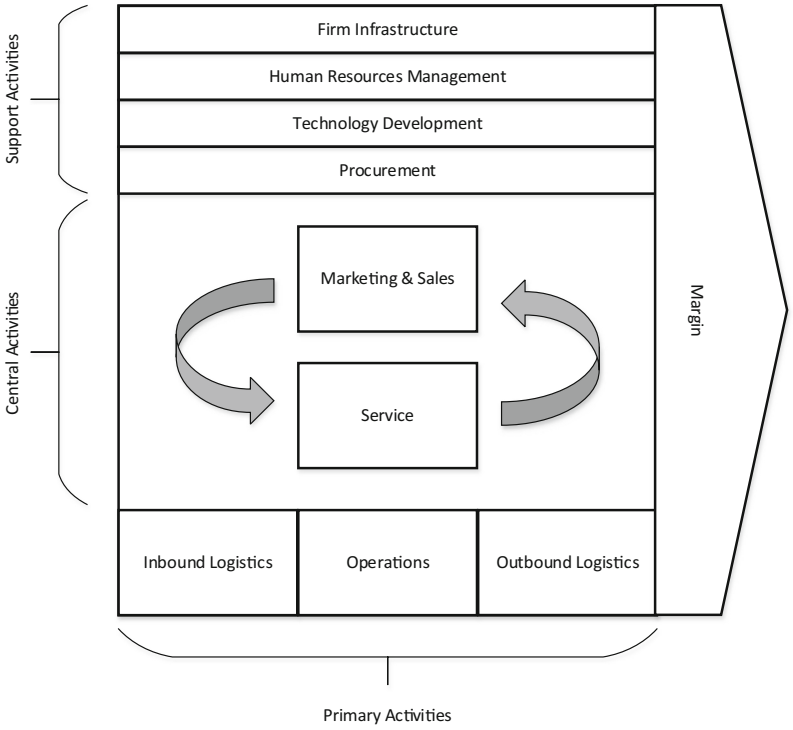


Fig. 1. Adjusted Porter's value chain

will more often be able to use mass customization. With the reduction of rejection rate, faster processes, and automated retooling, both of the major strategies can be pursued. Furthermore, the strong integration of the customer and enhanced services give the opportunity for an additional service strategy. *"The machine suppliers know that. They want to be permanently connected to our machines. They advertise the predictive maintenance. This is why they first sell us the machines and later, they sell the services" [05-01].*

The major phenomenon in the value creation process caused by Industry 4.0 are:

- **Decentral structures** – self-adapting machines and data processing in real-time will need decentral data structures; all departments will need access.
- **Department borders** – the virtual linkage will lead to a reduction of department separations; development and production, as well as service and IT departments, will have to cooperate more with more time constraints.
- **Interfaces to suppliers and customers** – the interfaces with suppliers and customers will be expanded; improvement regarding data security and access permissions is necessary.
- **New employee capabilities** – IT abilities and the readiness for human-computer interaction become more and more important.
- **Integration of the IT strategy** – decisions for or against certain IT will more than ever affect the whole enterprise; therefore, the development of an IT strategy has to be part of an enterprise's long-term strategy.

6 Conclusion and Limitations

This paper shows the impacts provoked by Industry 4.0 in detail for each activity according to Porter [8]. Due to the use of a phenomenological research approach, we gain an in-depth view. The phenomenological perspective allows us to gain insights on a complex phenomenon from different interviewee's perspectives. Most interviewees report operations as the initial starting point. Moreover, they state that the strategic decisions play a crucial role. The distinction between a cost leadership strategy and a differentiation strategy is becoming less important because of new possibilities in Industry 4.0 factories to combine cost savings with distinction. The role of the customers is getting more important by integrating them into the value creation process. This leads us to rethink the activity groups in Porter´s value chain to highlight this central role.

Even though we carefully preceded the research for the recent paper, we have to admit some limitations concerning the underlying research process. The process of Industry 4.0 is still on-going. A broader discussion of findings with additional case studies is needed to develop a common Industry 4.0 value creation process description. Though we carefully chose a sample that covers a variety of characteristics and perceptions about Industry 4.0 to come to a holistic view, we cannot claim overall generalizability of the results. Most of the examined companies were large companies from Germany. Even though many of them are internationally oriented, we assume it as interesting to have a comparison with companies from other countries and cultures.

Moreover, the topic is also important for SMEs which is why comparisons between company sizes also lead to further insights.

References

1. Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. acatech – National Academy of Science and Engineering (2013)
2. Bertschek, I., et al.: Industrie 4.0: Digitale Wirtschaft - Herausforderung und Chance für Unternehmen und Arbeitswelt. Ifo Schnelld. **68**, 3–18 (2015)
3. Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M.: How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 perspective. Int. J. Mech. Ind. Sci. Eng. **8**, 37–44 (2014)
4. Pflaum, A., Bodendorf, F., Prockl, G., Chen, H.: Introduction to the digital supply chain of the future: technologies, applications and business models minitrack. In: Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS) (2017)
5. Holotiuk, F., Beimborn, D.: Critical success factors of digital business strategy. In: Proceedings of Wirtschaftsinformatik (WI), St. Gallen, pp. 991–1005 (2017)
6. Hess, T., Matt, C., Benlian, A., Wiesböck, F.: Options for formulating a digital transformation strategy. MIS Q. Exec. **15**, 123–139 (2016)
7. Husserl, E.: Cartesian Meditations: An Introduction to Phenomenology. Springer, Berlin (2013)
8. Porter, M.E.: Competitive Advantage: Creating and Sustaining Superior Performance. Free Press, New York (1985)
9. Stock, T., Seliger, G.: Opportunities of sustainable manufacturing in Industry 4.0. Proc. CIRP **40**, 536–541 (2016)
10. Ashton, K.: That 'internet of things' thing. RFiD J. **22**, 97–114 (2009)
11. 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)
12. Lucke, D., Constantinescu, C., Westkämper, E.: Smart factory - a step towards the next generation of manufacturing. In: Mitsuishi, M., Ueda, K., Kimura, F. (eds.) Manufacturing Systems and Technologies for the New Frontier, pp. 115–118. Springer, London (2008). https://doi.org/10.1007/978-1-84800-267-8_23
13. Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., Hoffmann, M.: Industry 4.0. Bus. Inf. Syst. Eng. **6**, 239–242 (2014)
14. 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 (2016)
15. Nolan, R.L.: Ubiquitous IT: the case of the Boeing 787 and implications for strategic IT research. J. Strateg. Inf. Syst. **21**, 91–102 (2012)
16. Constantiou, I.D., Kallinikos, J.: New games, new rules: Big data and the changing context of strategy. J. Inf. Technol. **30**, 44–57 (2015)
17. Mithas, S., Tafti, A., Mitchell, W.: How a firm's competitive environment and digital strategy posture influence digital business strategy. Manage. Inf. Syst. Q. **37**, 511–536 (2013)
18. Woodard, C.J., Ramasubbu, N., Tschang, F., Sambamurthy, V.: Design capital and design moves: the logic of digital business strategy. Manage. Inf. Syst. Q. **37**, 537–564 (2013)

19. Kagermann, H.: Change through digitization—value creation in the age of Industry 4.0. In: Albach, H., Meffert, H., Pinkwart, A., Reichwald, R. (eds.) Management of permanent change, pp. 23–45. Springer, Wiesbaden (2015). https://doi.org/10.1007/978-3-658-05014-6_2
20. Brynjolfsson, E., McAfee, A.: The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company, New York (2014)
21. Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., Schlund, S.: Produktionsarbeit der Zukunft-Industrie 4.0. Fraunhofer Verlag, Stuttgart (2013)
22. Dregger, J., Niehaus, J., Ittermann, P., Hirsch-Kreinsen, H., ten Hompel, M.: The digitization of manufacturing and its societal challenges: a framework for the future of industrial labor. In: Proceedings of the IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS) (2016)
23. Baird, A., Raghu, T.S.: Associating consumer perceived value with business models for digital services. Eur. J. Inf. Syst. **24**, 4–22 (2015)
24. Herterich, M.M., Uebernickel, F., Brenner, W.: The impact of cyber-physical systems on industrial services in manufacturing. Proc. CIRP **30**, 323–328 (2015)
25. Lee, J., Kao, H.-A., Yang, S.: Service innovation and smart analytics for Industry 4.0 and big data environment. Procedia CIRP. **16**, 3–8 (2014)
26. Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N.V.: Digital business strategy: toward a next generation of insights. MIS Q. **37**, 471–482 (2013)
27. 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, Cyprus, pp. 1–17 (2016)
28. Teece, D.J., Pisano, G., Shuen, A.: Dynamic capabilities and strategic management. Strateg. Manage. J. **18**, 509–533 (1997)
29. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming companies. Harv. Bus. Rev. **93**, 96–114 (2015)
30. Huggins, R., Izushi, H.: Competition, Competitive Advantage, and Clusters: The Ideas of Michael Porter. Oxford University Press, Oxford (2011)
31. Ghemawat, P.: Competition and business strategy in historical perspective. Bus. Hist. Rev. **76**, 37–74 (2002)
32. Sheehan, N.T., Foss, N.J.: Enhancing the prescriptiveness of the resource-based view through porterian activity analysis. Manage. Decis. **45**, 450–461 (2007)
33. Galvagno, M., Dalli, D.: Theory of value co-creation: a systematic literature review. Manage. Serv. Qual. Int. J. **24**, 643–683 (2014)
34. Mandrella, M., Zander, S., Kolbe, L.M.: IT-based value co-creation: a literature review and directions for future research. In: Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS) (2016)
35. Mustak, M., Jaakkola, E., Halinen, A.: Customer participation and value creation: a systematic review and research implications. Manage. Serv. Qual. Int. J. **23**, 341–359 (2013)
36. Barringer, B.: Walking a tightrope: creating value through interorganizational relationships. J. Manag. **26**, 367–403 (2000)
37. Grönroos, C.: Service logic revisited: who creates value? And who co-creates? Eur. Bus. Rev. **20**, 298–314 (2008)
38. Mustafa, R.: Business model innovation: pervasiveness of mobile banking ecosystem and activity system – an illustrative case of Telenor Easypaisa. J. Strategy Manag. **8**, 342–367 (2015)
39. Sarker, S., Xiao, X., Beaulieu, T.: Qualitative studies in information systems: a critical review and some guiding principles. Manage. Inf. Syst. Q. **37**, iii–xviii (2013)

40. 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)
41. Dubé, L., Paré, G.: Rigor in information systems positivist case research: current practices, trends, and recommendations. MIS Q. **27**, 597–636 (2003)
42. Creswell, J.W.: Qualitative Inquiry and Research Design: Choosing Among Five Approaches. SAGE Publications, Thousand Oaks (2013)
43. Ströker, E., Husserl, E.: Die Krisis der europaeischen Wissenschaften und die transzendentale Phänomenologie, 3rd edn. Meiner, Hamburg (2007)
44. Garfinkel, H.: Conditions of successful degradation ceremonies. Am. J. Sociol. **61**, 420–424 (1956)
45. Giorgi, A.: The Descriptive Phenomenological Method in Psychology: A Modified Husserlian Approach. Duquesne University Press, Pittsburgh (2009)
46. Strauss, A., Corbin, J.: Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage, Newbury Park (1990)
47. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publ. Co., Chicago (1975)
48. Mayring, P.: Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. SSOAR, Klagenfurt (2014)
49. Vogelsang, K., Steinhüser, M., Hoppe, U.: A qualitative approach to examine technology acceptance. In: Proceedings of the International Conference on Information Systems (ICIS), Milan (2013)
50. Corbin, J., Strauss, A.: Grounded theory research: procedures, canons and evaluative criteria. Z. Für Soziol. **19**, 418–427 (1990)
51. Roth, A.: Einführung und Umsetzung von Industrie 40. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-48505-7
52. Fritz, W.: Markteintrittsstrategien in der Internet-Ökonomie. Arbeitspapier, Technische Universität Braunschweig, Institut für Marketing, Braunschweig (1999)

Exploring the Determinants of IoT Adoption: Findings from a Systematic Literature Review

Marian Carcary, Giovanni Maccani(✉), Eileen Doherty, and Gerard Conway

Innovation Value Institute, Maynooth University, Co. Kildare, Ireland
{marian.carcary, giovanni.maccani, eileen.doherty, gerard.conway}@mu.ie

Abstract. The Internet of Things (IoT) heralds a new era of disruptive technologies that provide organizations with both benefits and challenges. However, organizational adoption of IoT is not yet widespread and greater understanding of the phenomenon is required. This study examines the existing literature on the key determinants (drivers, benefits, barriers, and challenges) that influence the adoption of IoT by organizations. Therefore, this paper presents findings from a Systematic Literature Review (SLR) and concept matrix approach to identify these IoT adoption determinants at the organizational level. The key constructs of the Unified Theory of Acceptance and Use of Technology (UTAUT) were examined in relation to the determinants identified to understand applicability of this theory in the IoT context. Future research will complement these findings through an empirical investigation. Therefore, the overall aim of this research is (1) to generate a model that outlines the determinants influencing organizational IoT adoption and (2) to ascertain the applicability of UTAUT in understanding IoT adoption and to further enrich UTAUT by contextualizing its constructs to the IoT phenomenon.

Keywords: Internet of Things (IoT) · IoT adoption determinants
IoT drivers · IoT benefits · IoT barriers · IoT challenges
Systematic literature review · Concept matrix
Unified Theory of Acceptance and Use of Technology (UTAUT)

1 Introduction

The Internet of Things (IoT) paradigm represents a nascent concept that heralds a new era of disruptive technologies, enabling ubiquitous, unbounded connectivity of different types of devices and pervasive computing scenarios [12, 14]. It is the next revolutionary concept in the transformation of the Internet into a fully integrated network, enabling a harmonious interaction between societies, individuals, and smart things [29]. IoT is defined as a network of intelligent heterogeneous physical devices or objects that contain embedded technology and are seamlessly interconnected [2, 43]. These devices are augmented with sensing, computing, and communication capabilities, enabling them to communicate and sense or interact with their internal states or external environments [14]. They are connected to form a network, across which data

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 113–125, 2018.
https://doi.org/10.1007/978-3-319-99951-7_8

can be transferred without human-to-computer or human-to-human interactions [71]. By leveraging insights from the collective analyses of this data, organizations can solve real-world problems and deliver reality-augmented services and functionalities that were not heretofore imagined [7, 53]. These functionalities have the potential of disrupting multiple application areas across both the private and public sectors including supply chain management [49], manufacturing [30], healthcare [42], and traffic management in cities [39]. This study focuses on IoT from an end-to-end perspective where IoT can be conceptualized as a complex 'system of systems' that is comprised of three high level layers. These include: a 'Devices' layer where objects are augmented with sensing and data identification and capture technologies; a 'Network' layer across which data is transferred to local processing technologies where it is collected, filtered, and aggregated; and a 'Data Storage and Analytics' layer, where the data can be leveraged to generate new analysis and insights that deliver business value [36].

IoT is anticipated to have a greater impact than the Internet itself [63], as by 2020 it is projected that the number of Internet-enabled connected devices will reach 50 billion [23]. The economic impact of IoT has been forecast to have a total value of between $3.9 and $11.1 trillion per year up to 2025 [41]. However, organizational adoption of IoT is far from being universal. In a 2014 survey of 1400 C-suite executives, 73% had not made any concrete IoT investments, and only 7% had developed a comprehensive IoT strategy [1]. A 2015 survey of more than 200 IT and business leaders revealed that 42% were only starting to consider the potential impact of IoT on their business [69]. In a further 2016 survey of over 500 business executives, 90% of organizations remained in the IoT planning or proof of concept stages, with only 20% expecting to implement IoT solutions at scale by 2020 [11].

While it is anticipated that the momentum to adopt IoT will rapidly grow due to competitive pressures on organizations to transform [40], the lack of IoT adoption by organizations to date may be partly explained by the fact that it is very different from the adoption of other technologies [9]. This is because IoT requires the adoption of an ensemble of hardware (e.g. sensors, IoT gateways, cloud infrastructure), software (across the various IoT layers), and networking technologies that are supplied by multiple vendors. As a consequence, the IoT landscape reflects a rich variety of protocols, technologies, and devices [51] that have different requirements, boundary conditions, traffic characteristics, and involve multiple stakeholders [7, 21]. In addition, there is a scarcity of studies in relation to the social, behavioral, economic, and managerial issues associated with IoT, thereby hampering an organization's ability to make informed decisions about IoT adoption and use [40]. Given the assertion that IoT adoption is different from that of other technologies, it raises the question as to whether existing IS theories of technology adoption are sufficient to provide the basis for an effective understanding of the IoT phenomenon. IS research acknowledges the need to investigate "How will theories of IT adoption and diffusion develop to take account of IoT?" (see call for papers for the Journal of Strategic Information Systems Special Issue on "IT Governance on the Internet of Things", October 2017 - https://www.journals.elsevier.com/the-journal-of-strategic-information-systems/call-for-papers). Consequently, the research question explored in this study is 'what are the determinants that influence the adoption of IoT by organizations'? This question is addressed in two sequential phases. Phase 1 (discussed in this paper) identifies the current understanding in the literature of the key determinants of IoT

adoption by undertaking a systematic literature review of IoT adoption drivers, benefits, barriers, and challenges, and a content analysis of the relevant literature. Through developing a concept matrix [76], the identified themes are categorized within the constructs of the Unified Theory of Acceptance and Use of Technology (UTAUT) [73]. Thus, this paper contributes through a systematic analysis of the extant literature on IoT adoption through the lens of the UTAUT theory. These findings allow reflecting on the current understanding of IoT adoption and representing the foundation for future empirical research avenues. Hence, Phase 2 of this study (discussed in a future paper) will complement these findings through an empirical investigation. Therefore, the overall aim of this research is to (1) generate a model that outlines the determinants influencing IoT adoption by organizations and (2) ascertain the applicability of UTAUT in understanding IoT adoption and to further enrich UTAUT by contextualizing its constructs to the IoT phenomenon.

2 Theoretical Background

Research in the area of adoption, diffusion, and acceptance of technologies is argued to be a mature field in the IS literature [34, 44, 78]. Several theoretical frameworks have been formulated across disciplines to study these phenomena. Venkatesh et al. (2003) propose a "Unified Theory of Acceptance and Use of Technology" (UTAUT) that integrates eight widely used theories: (1) Theory of Reasoned Action [25], (2) Technology Acceptance Model (TAM) [19] and its extension [72], (3) Motivational Model [20], (4) Theory of Planned Behavior (TPB) [3], (5) Combined TAM and TPB model [68], (6) Model of PC Utilization [70], (7) Diffusion of Innovations Theory [55], and (8) Social Cognitive Theory [17]. For this study, UTAUT was selected as the underpinning theoretical framework to map and interpret the determinants of IoT adoption for the following reasons. Firstly, UTAUT synthesizes many alternative views on IS adoption in the literature and provides a common structure to harmonize the constructs of different theories, many of which were similar in nature. More importantly, UTAUT addresses a fundamental limitation of the models that underpin it, in that they are focused on "individual-oriented information technologies as opposed to more complex and sophisticated" [73, p. 427] ones. IoT clearly represents a complex and sophisticated technology that falls consistently within the scope of UTAUT. Since its inception, UTAUT has been widely employed as a theoretical lens for studies associated with individual and organizational IS adoption and use, with reference made to a wide range of technologies, including those that are more complex and specialized in nature. Examples of such sophisticated technologies within organizations include biometrics authentication systems [5], customer relationship management systems [48], e-government services [4], electronic HRM [32], electronic medical record systems [15], electronic procurement systems [62], and enterprise resource planning systems [35].

This paper considers the four constructs defined within UTAUT as determinants of IS adoption: Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions [73]. An overview of these constructs is provided below, and they are used in subsequent sections as the basis for theoretically framing the IoT adoption determinants systematically identified in the literature. *Performance*

Expectancy is defined as the "degree to which an individual believes that using the system will help him or her to attain gains in job performance" [73, p. 447]. The underlying assumption is that behaviors are influenced by the value outcomes anticipated from using a technology. *Effort Expectancy* is the "degree of ease associated with the use of the system" [73, p. 450]. This construct infers that when institutionalizing a new technology, behaviors are influenced by the degree of effort required to process complex new stimuli. *Social Influence* is "the degree to which an individual perceives that 'important others' believe he or she should use the new system" [73, p. 451]. This implies that behaviors are influenced by how people believe others will view them as a result of using a technology. *Facilitating Conditions* are described as "the degree to which an individual believes that organizational and technical infrastructure exists to support use of the system" [73, p. 453]. The underlying position is that aspects of the technological and organizational environment can lower the barriers for using a technology.

3 Systematic Literature Review (SLR): The Approach

In order to better understand the IoT phenomenon, the approach of this study involved a concept-centric examination of drivers, benefits, barriers, and challenges to IoT adoption in organizations. These key determinants were identified through understanding the requirements for effectively conducting a literature review [50, 58] and following a systematic literature review (SLR) approach [47], and were analyzed using a concept matrix [76]. The SLR adhered to the 8 steps proposed by Okoli [47]. These included:

(1) Purpose of the literature review: The authors sought to identify relevant themes in the stream of research centered on the drivers, benefits, barriers, and challenges to organizational IoT adoption.
(2) Protocol and training: A focused, systematic literature search was undertaken to identify all drivers, benefits, barriers, and challenges to organizational IoT adoption. Analysis of the identified literature was supported through the development of a concept matrix [76], which provided a simple, visual diagrammatic representation of the key determinants identified and their frequency, categorized according to the four UTAUT constructs.
(3) Searching for the literature: The requirement for an article to be considered for analysis was the presence of the following terms within the paper title, abstract, or keywords: 'Internet of Things OR IoT' AND 'adoption OR drivers OR benefits OR barriers OR challenges'. The literature collection sources were focused on the Association of Information Systems "Basket of 8" and the "Business Source Complete" bibliographic collection, and sought to identify scholarly peer reviewed articles published 2010 and 2017 included. The authors identified 253 papers.
(4) Practical screen: Resulting from this step was a list of the literature to be considered for review. An initial screen of the title, abstract, and keywords of the 253 identified papers resulted in the exclusion of 207 papers due, for example, to

language or a focus on issues that did not illuminate specific adoption determinants. All remaining papers were read to verify their relevance. In parallel, analysis of the references of those initially selected papers was also undertaken. Consequently, 46 papers were systematically ordered and selected to achieve the SLR's objective.

(5) Quality appraisal: The 46 eligible papers were also evaluated in terms of quality. Given the search focus for scholarly peer reviewed articles, no quality-related issues were found.

(6) Data extraction: Within each of the shortlisted papers the relevant material to answer the review question was isolated, and all drivers, benefits, barriers, and challenges to IoT adoption were extracted. Hence, this step provided a list of all relevant concepts to synthesize the study.

(7) Synthesis of studies: A content analysis of the material extracted from the literature was undertaken to establish the most common concepts. The authors created a high-level categorization comprising the four constructs of UTAUT, according to which the drivers, benefits, barriers, and challenges were classified based on their fit according to the UTAUT constructs (Table 1). The authors followed the concept matrix method [76] - the matrix rows provide the paper references from which the concepts were extracted, while frequency of occurrence of a particular theme is indicated by the number of 'Xs' in the table columns.

(8) Writing the review: The final step involved a write up of the review. In the next section, the identified determinants are described based on their categorization within the UTAUT constructs.

4 SLR: Key Findings from the Literature

This section provides the key literature findings in relation to organizational IoT adoption that were identified through the SLR. These are discussed separately below for each of the UTAUT constructs.

Performance Expectancy: As outlined in Table 1, six of the twelve determinants identified were found to be related to Performance Expectancy. These primarily emerged as the expected benefits and usefulness of IoT for organizations. A description and discussion of each item is provided below.

Improved Integration and Connectivity Potential: With the advent of IoT, networked connections have become more valuable. Through technologies embedded in sensor and actuator networks, IoT has the potential to enable the constant and ubiquitous connectivity of different devices [14, 22]. A key development trend is that of IoT's integration with existing network systems [79] and support for inter-organizational integration [24].

Real-Time Data Visibility and Sharing: IoT is one of the key sources of big data [12]. The connectivity of IoT devices enables the dynamic generation of vast volumes of real-time or near real- time granular intelligence data [45, 63]. The practice of ongoing sensing increases data richness to include details such as the status of items or

Table 1. Concept matrix of IoT adoption drivers, benefits, barriers, challenges.

	Performance Expectancy						Facilitating Conditions				Social Influence	
	Improved integration/ connectivity	Real-time data visibility/sharing	Improved business analytics/decisions	Monitoring and control	Enhanced efficiencies	Enhanced productivity	Data management challenges	Lack of support/ awareness	Lack of unified inter-connection standard	Scalability concerns	Security concerns	Privacy concerns
Airehrour et al (2016)									X		X	
Al-Turjman et al (2013)									X			
Amadeo et al (2016)									X	X	X	X
Boos et al (2013)		X		X	X							
Borgia (2014)		X	X	X	X	X	X		X	X	X	X
Botta et al (2016)		X							X	X	X	X
Bradley et al (2014)						X	X		X		X	X
Caron et al (2016)	X	X	X	X	X						X	X
Cho et al (2013)											X	X
Corujo et al (2012)									X			
DiMarco et al (2016)									X			
Dominici et al (2016)	X	X	X								X	X
Dutton (2014)		X		X	X			X	X	X	X	X
Ferretti and Schiavone (2016)	X	X	X	X	X	X		X	X		X	X
Fleisch (2010)		X	X	X	X				X			
Georgakopoulos and Jayaraman (2016)				X			X				X	X
Grieco et al (2014)		X		X							X	
Hakiri et al (2015)									X	X	X	X
Henze et al (2016)											X	X
Jacobsson et al (2016)											X	
Jiang et al (2015)							X			X	X	
Lee and Kyoochun (2015)	X	X	X	X	X	X	X		X		X	X
Murray et al (2016)		X	X		X	X						
Neisse et al (2015)		X							X	X		X
Ng (2014)			X		X				X		X	
Patel and Cassou (2015)									X			
Pye (2014)			X	X	X		X		X		X	
Qin et al (2016)		X	X		X		X		X		X	X
Qiu et al (2015)		X	X		X				X		X	X
Roman et al (2013)							X		X	X	X	X
Sahraoui and Bilami (2015)											X	X
Saied et al (2014)									X		X	
Saint (2014)		X	X	X	X	X	X				X	
Shrouf and Miragliotta (2015)		X	X	X	X							
Sicari et al (2015)									X	X	X	X
Sofronijević et al (2014)									X			
Stojkoska and Trivodaliev (2017)		X					X		X			
Storey (2014)								X	X		X	X
Valmohammadi (2016)		X	X		X			X	X		X	X
Weber (2010)											X	X
Weber (2015)									X		X	X
Weinberg et al (2015)		X	X		X	X	X		X		X	X
Xu et al (2016)	X								X		X	X
Yan et al (2014)											X	X
Yun et al (2015)									X	X	X	X
Zhou et al (2017)					X						X	X
	5	19	15	12	17	7	11	4	30	10	35	28

their surroundings, and gives rise to more "trusted" data that is difficult to influence due to its continuous collection [26]. Sharing of this data across interconnected devices improves situational awareness and provides opportunities for smarter services [40] including connected marketing, and reality-augmented customer services [14, 40].

Improved Business Analytics and Decision-Making: The embedding of business analytic tools in IoT devices enables dynamic drill down, querying, and analysis of intelligence data [10, 52], diagnostics and complex problem-solving [46, 54], and enables real-time decision-making to take place at the data source [63, 77]. This often occurs autonomously without human intervention [14]. The analytics enabled by IoT technologies provide actionable insights on the organization's strengths and weaknesses [24], and enables it to react to events promptly [53].

Monitoring and Control: IoT-enabled data availability facilitates real or near real-time performance monitoring and tracking (e.g. of entire product lifecycles, supply chain status, equipment performance, energy consumption) [14, 63]. Organizations can leverage the benefits of remote control capabilities [22, 77] to, for example, enable timely reaction to critical parameter measurements [28] and implement simple preventative maintenance concepts [52].

Enhanced Organizational Efficiencies: Operating efficiencies enabled by IoT can result from improved systems and procedures [40], such as the use of proximity or sensor triggers that can automatically initiate an activity [26]. IoT-enabled redesign and streamlining or automation of business processes can minimize the intervention required by humans [26] or improve coordination between people, products, and procedures [24], and result in greater business agility [52, 77], and procedural innovation [43]. Operating efficiencies can materialize in the form of improved transaction speed and accuracy [43], reduced labor, transaction, and process failure costs [24, 53], reduced waste [10, 23], improved energy efficiency [77], and supply chain responsiveness [71].

Enhanced Organizational Productivity: Through the effective coordination of IoT devices, systems can be harnessed to increase organizational productivity [13, 61]. This includes, for example, improved value chain productivity across product design, inventory management, manufacturing, logistics, and customer service [22, 43, 77]. Similarly, employee productivity can be enhanced, with the organization leveraging its employees' skills to their full potential [24, 40], and over time increasing the value of intellectual capital [43].

Facilitating Conditions: With respect to Facilitating Conditions, four determinants were identified from the literature (Table 1). Of these, *lack of organizational support and awareness* refers to the availability of a specific person or group "for assistance with system difficulties" as defined as part of the original construct [73, p. 460]. The remaining four relate to internal and external constraints encompassing resource facilitating conditions and technology facilitating conditions [3, 68, 73].

Lack of a Unified Interconnection Standard: A unified, global interconnection standard to support seamless integration across IoT devices does not exist [67, 71]. This gives rise to the implementation complexities of fragmented solutions and interoperability challenges for organizations [6, 81], and closed IoT ecosystems [23, 81], thereby reducing the value proposition for IoT investment [46, 71]. Several standardization issues exist including those pertaining to data format, data inter- faces, protocols, service platform, and architecture [10]. Establishing a universal standard is challenged by the number of organizations and industry bodies with a vested interest in the IoT ecosystem [67]. The potential exists for large vendors to frustrate the path to

standardization by developing closed proprietary IoT solutions around their own products [13].

Scalability Concerns: The exponential growth of big data generated by connected IoT devices gives rise to considerable scalability and extensibility challenges [7], increasing scalability requirements many orders of magnitude higher than currently available on the Internet [23]. Confidence in the availability of a scalable, cost efficient architecture to address the complexities of the IoT 'data deluge' is important in ensuring faster uptake of IoT [10].

Data Management Challenges: Estimates suggest that IoT data generation will grow from a present volume of 4 zettabytes to over 40 zettabytes in 2020 [77]. This large-scale data is characterized by issues such as velocity, quality, redundancy, inconsistency, incompleteness, and ambiguity [38, 53]. However, current data center architecture is inadequate to cater for both the data volumes and its heterogeneous nature [40]. In a reversal of trends, IoT is driving the need for distributed data center management to improve data processing efficiency and response time [13]. To support effective organizational adoption, new algorithms and technologies are required for data processing and storage [77]. Efficient representation schemes and indexing methods are required to readily locate specific data items [10], as well as sophisticated data mining tools to make sense of the data [61].

Lack of Organizational Support and Awareness: IoT adoption in organizations is impacted by a lack of senior management knowledge, commitment, and support, inadequate employee technological skills [24, 71], and inadequate awareness of security and data privacy concerns [67]. Putting the organizational and institutional infrastructures in place to address the above is difficult due to the required social change, and potential resistance to changes in day-to-day work practices [23].

Social Influence: Significant emphasis was found in the literature regarding the role of Normative Pressure in influencing IoT adoption. In particular, this emerged as concerns pertaining to privacy and security, both of which are currently challenging the complete implementation of IoT systems.

Security Concerns: IoT-specific security threats result from the proliferation of embedded and connected devices and underlying network heterogeneity [59, 79]. Security attacks include, for example, identity fabrication, unauthorized access to intellectual property, sabotage of critical infrastructure, and denial of service [31]. The implications of IoT specific security threats are often profound, offering cyber criminals the opportunity to take control of physical devices in close proximity to individuals [77]. Research highlights that there are several vulnerabilities inherent in IoT devices such as insecure web interfaces, software and firmware vulnerabilities, privacy issues, and lack of transport encryption and authentication/authorization [2, 40]. Several vulnerabilities are also created by human actors in an IoT system, including disgruntled employees [37].

Privacy Concerns: The ubiquity of IoT devices gives rise to significant privacy concerns, including for example, the challenge of uncontrolled data generation and diffusion, inadequate authentication, preservation of anonymity, and risks pertaining to sensitive data [14, 33]. Akin to the concept of Orwellian surveillance, an issue of considerable importance is unauthorized surveillance that results in large-scale data

collection without an individual's consent, leading to individual tracking and behavior inference [75]. Data ownership is also a concern [22], with questions arising in terms of who owns the data in a system where multiple actors co-create and add value [77]. Privacy concerns are exacerbated when data is outsourced to the cloud, as a result of data transfer, storage, and processing across legislative boundaries, and perceived loss of control with regards to data access by third parties [33].

Effort Expectancy: As outlined in Table 1, none of the IoT adoption determinants uncovered through the SLR were found to be related to Effort Expectancy. Given that this construct is concerned with the degree of effort required in using a new technology, the authors argue that this may be an area currently overlooked in IoT adoption literature, particularly given the differences and complexities associated with IoT compared to other technologies. In the next research phase, it is anticipated that aspects such as required "learning to operate the system" [19, 56, 73] will emerge as empirical data is gathered from organizations that have adopted IoT.

5 Conclusions and Future Research

This IoT adoption is different from that of other technologies in that it reflects adoption of a complex "system of systems", comprised of multiple hardware, software, and networking technologies that are often faced with fragmentation and interoperability challenges. To date, it is poorly understood by the IS research community. A paucity of studies in this area is evident, as of the papers systematically reviewed throughout this study's SLR process, none of those pertaining to IoT adoption determinants were published in journals of the AIS "basket of eight". This paper, proposes an initial contribution to the IS challenge outlined in the forthcoming Special Issue of the Journal of Strategic Information Systems on the need to understand "How will theories of IT adoption and diffusion develop to take account of the IoT?" The findings of this research serve three purposes. Firstly, the paper elucidates the current understanding in the extant literature on the determinants for organizational adoption of IoT through undertaking a SLR. It highlights that adoption may be influenced by the fact that IoT can improve integration and connectivity, real-time data visibility, business analytics and decision-making, to provide improved organizational efficiencies and productivity. However, the decision to adopt may also be influenced by a number of constraints, such as lack of a unified interconnection standard, lack of internal awareness and support, coupled with concerns around scalability, security, privacy, and data management. Secondly, by categorizing these determinants according to the four UTAUT constructs, the paper provides initial insights on the applicability of UTAUT in explaining organizational IoT adoption. In its current formulation, UTAUT appears a viable option for investigating IoT. All of the determinants identified through the SLR could be explained by the Performance Expectancy, Social Influence, and Facilitating Conditions constructs of UTAUT. Interestingly, we did not uncover any determinants that could not be explained by the existing UTAUT constructs, which provides further support for its relevance in relation to explaining adoption of emerging and complex systems of technologies. Thirdly, Effort Expectancy did not underpin any determinant identified in the current

literature. We believe this represents an important gap and a research opportunity for IS scholars. Important questions in this way remain substantially unanswered in the IS literature to date. For example: what is the learning curve to effectively adopt and make full use of IoT? What skills are needed at the individual and organizational level to effectively adopt and use IoT? The work presented in this paper will also serve as the basis for an empirical investigation in the next phase of this study. A quantitative research approach, in the form of an online survey [8], will be employed to gain greater insights into this phenomenon, and to further investigate the relevance of the determinants across different contexts. The survey will target individuals holding CIO, CTO, or equivalent roles in organizations, as those individuals are regarded as most likely to have the greatest knowledge of the organization's IoT adoption efforts. The insights gleaned from the data will produce an empirically validated model that outlines the key determinants influencing IoT adoption in an organizational setting. The contribution from this research will address the current paucity of studies, and thereby from a practical perspective, enable an organization to make more informed decisions regarding IoT adoption and use, and provide IoT solution providers with greater insights into the key determinants to organizational adoption of IoT. From a theoretical perspective, insights gained from the theoretical framework will enable a formal enrichment/extension to UTAUT that contextualizes its constructs to IoT adoption determinants.

References

1. Accenture: From Productivity to Outcomes. Using the Internet of Things to Drive Future Business Strategies. Accenture (2015)
2. Airehrour, D., Gutierrez, J., Ray, S.: Secure routing for internet of things: a survey. J. Netw. Comput. Appl. **66**, 198–213 (2016)
3. Ajzen, I.: The theory of planned behavior. Organ. Behav. Hum. Decis. Process. **50**(2), 179–211 (1991)
4. Al Awadhi, S., Morris, A.: The use of the UTAUT model in the adoption of e-government services in Kuwait. In Proceedings of the 41st Hawaii International Conference on System Sciences, pp. 1–11. IEEE, Hawaii (2008)
5. Al-Harby, F., Qahwaji, R., Kamala, M.: Users' acceptance of secure biometrics authentication system: reliability and validation of an extended UTAUT model. Commun. Comput. Inf. Sci. **87**(2), 254–258 (2010)
6. Al-Turjman, F., Al-Fagih, A., Alsalih, W., Hassanein, H.: A delay-tolerant framework for integrated RSNs in IoT. Comput. Commun. **36**, 998–1010 (2013)
7. Amadeo, M., et al.: Information-centric networking for the internet of things: challenges and opportunities. IEEE Netw. **30**, 92–100 (2016)
8. Blumberg, B., Cooper, D.R., Schlinder, P.S.: Business Research Methods. McGraw Hill, Berkshire (2014)
9. Boos, D., Guenter, H., Grote, G., Kinder, K.: Controllable accountabilities: the internet of things and its challenges for organizations. Behav. Inf. Technol. **32**(5), 449–467 (2013)
10. Borgia, E.: The internet of things vision: key features, applications and open issues. Comput. Commun. **54**, 1–31 (2014)
11. Bosche, A., Crawford, D., Jackson, D., Schallehn, M., Smith, P.: How Providers Can Succeed in the Internet of Things. Bain Insights, Munich (2016)

12. Botta, A., de Donato, W., Persico, V., Pescape, A.: Integration of cloud computing and internet of things: a survey. Future Gener. Comput. Syst. **56**, 684–700 (2016)
13. Bradley, T., Thibodeau, P., Ng, V.: The internet of things - threats and challenges. Netw. World Asia **11**, 16–18 (2014)
14. Caron, X., Bosua, R., Maynard, S., Ahmad, A.: The internet of things (IoT) and its impact on individual privacy: an australian perspective. Comput. Law Secur. Rev. **32**, 4–15 (2016)
15. Chisolm, D.J., Purnell, T.S., Cohen, D.M., McAlearney, A.S.: Clinician perceptions of an electronic medical record during the first year of implementation in emergency services. Pediatr. Emerg. Care **26**(2), 107–110 (2010)
16. Cho, K., Lee, B., Lee, K., Lee, D.: Energy-efficient replica detection for resource- limited mobile devices in the internet of things. IET Commun. **7**(18), 2141–2150 (2013)
17. Compeau, D.R., Higgins, C.A.: Computer self-efficacy: development of a measure and initial test. MIS Q. **19**(2), 189–211 (1995)
18. Corujo, D., Lebre, M., Gomes, D., Aguiar, R.: MINDiT: a framework for media independent access to things. Comput. Commun. **35**, 1772–1785 (2012)
19. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
20. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: Extrinsic and intrinsic motivation to use computers in the workplace. J. Appl. Soc. Psychol. **22**(14), 1111–1132 (1992)
21. DiMarco, P., Athanasiou, G., Mekikis, P., Fischione, C.: MAC-aware routing metrics for the internet of things. Comput. Commun. **74**, 77–86 (2016)
22. Dominici, G., Roblek, V., Abbate, T., Tani, M.: Click and drive. Bus. Process Manage. J. **22** (2), 420–434 (2016)
23. Dutton, W.: Putting things to work: social and policy challenges for the internet of things. Info **16**(3), 1–21 (2014)
24. Ferretti, M., Schiavone, F.: Internet of things and business processes redesign in sea- ports: the case of Hamburg. Bus. Process Manage. J. **22**(2), 271–284 (2016)
25. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wiley Publishing Company, Massachusetts (1975)
26. Fleisch, E.: What is the internet of things? An economic perspective. Econ. Manage. Financ. Markets **5**(2), 125–157 (2010)
27. Georgakopoulos, D., Jayaraman, P.P.: Internet of things: from internet scale sensing to smart services. Computing **98**(10), 1041–1058 (2016)
28. Grieco, L., et al.: IoT-aided robotics applications: technological implications, target domains and open issues. Comput. Commun. **54**, 32–47 (2014)
29. Guo, B., Zhang, D., Wang, Z., Yu, Z., Zhou, X.: Opportunistic IoT: exploring the harmonious interaction between humans and the internet of things. J. Netw. Comput. Appl. **36**, 1531–1539 (2013)
30. Haass, R., Dittmer, P., Veigt, M., Lütjen, M.: Reducing food losses and carbon emission by using autonomous control–a simulation study of the intelligent container. Int. J. Prod. Econ. **164**, 400–408 (2015)
31. Hakiri, A., Berthou, P., Gokhale, A., Abdellatif, S.: Publish/subscribe-enabled software defined networking for efficient and scalable IoT communications. IEEE Commun. Mag. **53** (9), 48–54 (2015)
32. Heikkila, J., Smale, A.: The effects of language standardization on the acceptance and use of e-HRM systems in foreign subsidiaries. J. World Bus. **46**(3), 1–9 (2010)
33. Henze, M., Hermerschmidt, L., Kerpen, D., Häußling, R., Rumpe, B., Wehrle, K.: A comprehensive approach to privacy in the cloud-based internet of things. Future Gener. Comput. Syst. **56**, 701–718 (2016)

34. Hu, P.J., Chau, P.Y., Sheng, O.R.L., Tam, K.Y.: Examining the technology acceptance model using physician acceptance of telemedicine technology. J. Manage. Inf. Syst. **16**(2), 91–112 (1999)
35. Huang, S., Wang, X.: Influence of organizational system to end-users' acceptance of ERP system in chinese enterprises. In: Proceedings of the 9th International Conference on Hybrid Intelligent Systems, pp. 160–164. IEEE, Shenyang (2009)
36. Intel: The Intel IoT Platform – Architecture Specification White Paper (2017). https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/iot-platform-reference-architecture-paper.pdf
37. Jacobsson, A., Boldt, M., Carlsson, B.: A risk analysis of a smart home automation system. Future Gener. Comput. Syst. **56**, 719–733 (2016)
38. Jiang, H., Shen, F., Chen, S., Li, K., Jeong, Y.: A secure and scalable storage system for aggregate data in IoT. Future Gener. Comput. Syst. **49**, 133–141 (2015)
39. Lang, W., Jedermann, R., Mrugala, D., Jabbari, A., Krieg-Brückner, B., Schill, K.: The "intelligent container"—a cognitive sensor network for transport management. IEEE Sens. J. **11**(3), 688–698 (2011)
40. Lee, I., Kyoochun, L.: The internet of things (IoT): applications, investments, and challenges for enterprises. Bus. Horiz. **58**, 431–440 (2015)
41. McKinsey: The Internet of Things: Mapping the Value Beyond the Hype (2015)
42. Mishra, D., Gunasekaran, A., Childe, S.J., Papadopoulos, T., Dubey, R., Wamba, S.: Vision, applications and future challenges of internet of things: a bibliometric study of the recent literature. Ind. Manage. Data Syst. **116**(7), 1331–1355 (2016)
43. Murray, A., Papa, A., Cuozzo, B., Russo, G.: Evaluating the innovation of the internet of things. Bus. Process Manage. J. **22**(2), 341–356 (2016)
44. Nakicenovic, N., Grübler, A. (eds.): Diffusion of Technologies and Social Behavior. Springer, Berlin (2013)
45. Neisse, R., Steri, G., Fovino, I., Baldini, G.: SecKit: a model-based security toolkit for the internet of things. Comput. Secur. **54**, 60–76 (2015)
46. Ng, V.: Drivers and obstacles to IoT adoption in Asia Pacific. Netw. World Asia **11**(3), 12–14 (2014)
47. Okoli, C.: A guide to conducting a standalone systematic literature review. Commun. Associ. Inf. Syst. **37**(43), 879–910 (2015)
48. Pai, J.C., Tu, F.M.: The acceptance and use of customer relationship management (CRM) systems: an empirical study of distribution service industry in Taiwan. Expert Syst. Appl. **38**(1), 579–584 (2011)
49. Papert, M., Pflaum, A.: Development of an ecosystem model for the realization of internet of things services in supply chain management. Electr. Mark. **27**(2), 175 (2017)
50. Paré, G., Trudel, M.C., Jaana, M., Kitsiou, S.: Synthesizing information systems knowledge: a typology of literature reviews. Inf. Manag. **52**(2), 183 (2015)
51. Patel, P., Cassou, D.: Enabling high-level application development for the internet of things. J. Syst. Softw. **103**, 62–84 (2015)
52. Pye, A.: The internet of things: connecting the unconnected. Eng. Technol. **9**, 64–70 (2014)
53. Qin, Y., Sheng, Q., Falkner, N., Dustdar, S., Wang, H., Vasilakos, A.: When things matter: a survey on data-centric internet of things. J. Netw. Comput. Appl. **64**, 137–153 (2016)
54. Qiu, X., Luo, H., Xu, G., Zhong, R., Huang, G.: Physical assets and service sharing for iot-enabled supply hub in industrial park (SHIP). Int. J. Prod. Econ. **159**, 4–15 (2015)
55. Rogers, E.M.: Diffusion of Innovations. Free Press, New York (1995)
56. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster, New York (2010)
57. Roman, R., Zhou, J., Lopez, J.: On the features and challenges of security and privacy in distributed internet of things. Comput. Netw. **57**, 2266–2279 (2013)

58. Rowe, F.: What literature review is not: diversity, boundaries and recommendations. Eur. J. Inf. Syst. **23**(3), 241–255 (2014)
59. Sahraoui, S., Bilami, A.: Efficient HIP-based approach to ensure light weight end-to-end security in the internet of things. Comput. Netw. **91**, 26–45 (2015)
60. Saied, Y., Olivereau, A., Zeghlache, D., Laurent, M.: Lightweight collaborative key establishment scheme for the internet of things. Comput. Netw. **64**, 273–295 (2014)
61. Saint, A.: Internet of things: brave new world. Eng. Technol. **9**(9), 80–83 (2014)
62. Sambasivan, M., Wemyss, G.P., Rose, R.C.: User acceptance of a G2B system: a case of electronic procurement system in Malaysia. Internet Res. **20**(2), 169–187 (2010)
63. Shrouf, F., Miragliotta, G.: Energy management based on internet of things: practices and framework for adoption in production management. J. Clean. Prod. **100**, 235–246 (2015)
64. Sicari, S., Rizzardi, A., Grieco, L., Coen-Porisini, A.: Security, privacy and trust in internet of things: the road ahead. Comput. Netw. **76**, 146–164 (2015)
65. Sofronijević, A., Milićević, V., llić, B.: Smart city as a framework for creating competitive advantages in international business management. Management **71**, 5–15 (2014)
66. Stojkoska, B.L.R., Trivodaliev, K.V.: A review of internet of things for smart home: challenges and solutions. J. Clean. Prod. **140**, 1454–1464 (2017)
67. Storey, A.: There's nothing 'smart' about insecure connected devices. Netw. Secur. **7**, 9–12 (2014)
68. Taylor, S., Todd, P.A.: Understanding information technology usage: a test of competing models. Inf. Syst. Res. **6**(2), 144–176 (1995)
69. TEKSystems: The State of the Internet of Things: Executive Summary (2016)
70. Thompson, R.L., Higgins, C.A., Howell, J.M.: Personal computing: toward a conceptual model of utilization. MIS Q. **15**(1), 124–143 (1991)
71. Valmohammadi, C.: Examining the perception of iranian organizations on internet of things solutions and applications. Ind. Commer. Train. **48**(2), 104–108 (2016)
72. Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance mod- el: four longitudinal field studies. Manage. Sci. **46**(2), 186–204 (2000)
73. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
74. Weber, R.: Internet of things – new security and privacy challenges. Comput. Law Secur. Rev. **26**, 23–30 (2010)
75. Weber, R.: Internet of things: privacy issues revisited. Comput. Law Secur. Rev. **31**, 618–627 (2015)
76. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **26**(2), 13–23 (2002)
77. Weinberg, B., Milne, G., Andonova, Y., Hajjat, F.: Internet of things: convenience vs. privacy and secrecy. Bus. Horiz. **58**, 615–624 (2015)
78. Williams, M.D., Rana, N.P., Dwivedi, Y.K.: The unified theory of acceptance and use of technology (UTAUT): a literature review. J. Enterp. Inf. Manage. **28**(3), 443–488 (2015)
79. Xu, K., Qu, Y., Yang, K.: A tutorial on the internet of things: from a heterogeneous network integration perspective. IEEE Netw. **30**, 102–108 (2016)
80. Yan, Z., Zhang, P., Vasilakos, A.: A survey on trust management for internet of things. J. Netw. Comput. Appl. **42**, 120–134 (2014)
81. Yun, J., Ahn, I., Sung, N., Kim, J.: A device software platform for consumer electronics based on the internet of things. IEEE Trans. Consum. Electr. **61**(4), 564–571 (2015)
82. Zhou, J., Cao, Z., Dong, X., Vasilakos, A.V.: Security and privacy for cloud-based IoT: challenges. IEEE Commun. Mag. **55**(1), 26–33 (2017)

Digital Innovation Based on Digital Signage: Method, Categories and Examples

Kurt Sandkuhl[1,2(✉)], Matthias Wißotzki[4], Alexander Smirnov[2], and Nikolay Shilov[3]

[1] Rostock University, Albert-Einstein-Str. 22, 18059 Rostock, Germany
kurt.sandkuhl@uni-rostock.de
[2] ITMO University, Kronverkskiy pr. 49, 197101 St. Petersburg, Russia
smir@iias.spb.su
[3] SPIIRAS, 14 Line 39, 199178 St. Petersburg, Russia
nick@iias.spb.su
[4] Wismar University, Philipp-Müller-Str. 14, 23966 Wismar, Germany
matthias.wissotzki@uni-rostock.de

Abstract. Delivery of demand-oriented and accurate information for a person's interest has been subject of research since many years. In the last years, focus of attention shifted from just-in-time provision of information to context-oriented personalized information supply. Digital signage is one way of context-oriented information supply and summarizes information presentation to a wide audience on large screens. Work presented in this paper brings together technology-oriented work in digital signage and application-oriented work in the housing industry. More concrete, we have developed new digital signage solutions and investigated their potential for providing innovative services to tenants in apartment buildings, i.e. in the housing industry. The main contribution of this work is (1) an overview to innovative digital signage solutions connected to elevators, (2) the categorization of assistance services in the context of digital signage, and (3) application examples from housing industry.

Keywords: Digital signage · Digital innovation · Assistance service
Housing industry · Elevator · Personalization

1 Introduction

Delivery of demand-oriented and accurate information for a person's interest has been subject of research since many years in various areas in the field of business information systems research, such as decision support systems [1, 2], ubiquitous computing or information logistics [13]. In recent years, the focus of attention shifted from just-in-time provision of information to context-oriented personalized information supply, i.e. to capture and interpret the context of a person and refine the information supply accordingly. However, what context-orientation is and how it can be implemented depends on the application field and the technical means of information supply.

The information presentation to a wide audience on large screens (Digital Signage [3, 4]) is one way of information supply quite popular both in publicly accessible

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 126–139, 2018.
https://doi.org/10.1007/978-3-319-99951-7_9

places (fairs, shopping centres) and places available to a clearly defined and limited group of persons (industrial facilities, office buildings). Digital signage can be used for different kinds of information, like, e.g., commercials, general information about the location for orientation purposes, safety-related information or information directed to individuals or groups of people. In the case of providing targeted information not to one particular person, but to a group, several problems arise. Information might be confidential and cannot be displayed without exactly knowing the audience, it might be difficult to identify the joint interest of a group of persons, or the presentation of certain information might require authentication of persons – to name only some examples.

From a technical perspective, there are no solutions for personalized contextual information supply with digital signage available, although the market potential of such "targeted" digital signage is promising, which is illustrated in an industrial case in this paper. From a research perspective, more work is needed both on technical aspects and on methodical issues regarding business models and contributions to digital innovation.

Work presented in this paper brings together technology-oriented work in digital signage and application-oriented work in the housing industry. More concrete, we developed new digital signage solutions connected to elevators and investigate their potential for providing innovative services to tenants in apartment buildings, i.e. in the housing industry. Due to this combination, we are able to study the process of both, technology development and digital innovation and business model development. From a methodology perspective, the digital innovation process is performed in accordance to the Digital Innovation and Transformation Process (DITP), a method proposed in previous work [14].

The research focus of the paper is on categorizing digital signage solutions and on evaluating and improving DITP. The main contribution of this work is (1) an overview of innovative digital signage solutions connected to elevators, (2) the categorization of assistance services in the context of digital signage, and (3) application examples from housing industry.

The paper is structured as follows: Sect. 2 describes the background for our work from digital signage and digital transformation and innovation. Section 3 briefly summarizes the research method used for the work. In Sect. 4, the industrial case and the digital signage solutions for elevators are presented. Section 5 describes the categorization and discusses lessons learned. Section 6 summarizes the paper and presents future work.

2 Background

2.1 Digital Signage

The term digital signage denotes "a system that transmits information using electronic display devices connected to the network at all places, such as outdoor, shop front, public space, transportation facilities" [6]. Related terms used with similar meaning are pervasive display or digital-out-of-home (DOOH). Taniguchi [7] provides a comprehensive survey of literature in the field of digital signage. Important work streams in this area include:

- Application fields of digital signage and their characteristics, such as corporate communication [8],
- Technology developments and technology applications, such as touchless tactile display [9],
- The effects on user attention and behaviour [10],
- Back-end services and their architectures [11], or
- The impact of digital signage use in different application fields, such as retail [12].

The analysis of existing work in the field shows that technologies such as large screens or touch monitors are frequently used in digital signage solutions but – with exception of [5] – there is no published work on such solutions connected to elevator infrastructures, which encourages our work in such devices and solutions presented in Sect. 4. Furthermore, the specific situation of individuals or groups waiting in front of an elevator door and the applications resulting from this situation so far have not been investigated in research from the perspective of innovation potential.

2.2 Digital Innovation and Transformation Process

In previous work, we proposed the DITP in combination with the qualification profile of "digital business architect" (DBA) as an acknowledgement of the importance to integrate business model innovations into the established enterprise architecture of the company under consideration [14]. In this paper, the DITP will be applied and briefly introduced in the following, whereas the DBA is not discussed.

The DITP is supposed to support enterprises in digital transformation and innovation processes. DITP can typically be applied in the following scenarios:

1. An existing enterprise is facing new product-related, customer-related and/or competitive challenges caused by digitalization and needs to react promptly.
2. A start-up company aims to implement a new business concept.
3. An established company has no urgent need of action regarding a digital transformation but is interested in exploiting its digital potentials and thus in improving or expanding business accordingly.

The overall DITP process triggered by these scenarios is illustrated in Fig. 1 and aims at tackling business-model and architecture-related problems. In the following description of DITP, we will focus on scenario 3 only, as this is the most adequate for the industrial case considered in Sect. 4.

The phase ***"Find Digital Potentials"*** focusses on the analysis of digital trends and potential building blocks for implementing innovations and transformations in the business. This phase is divided into the activities *DITP2.1 Research of digital trends/building blocks (for example by using Gartner Hype Cycle* [18] or *Technology Radar* [19]) and *DITP2.2 Evaluation of digital trends/building blocks.* Examples for building blocks are *automation, digital accesses* (social networks, apps, etc.), *interconnectivity and exchange* (broadband, IoT, Cloud), *artificial intelligence or data* (big data, blockchain).

The next phase includes the modelling of the business model ***"Design New Business Model"***. Different approaches can be used for modelling according to

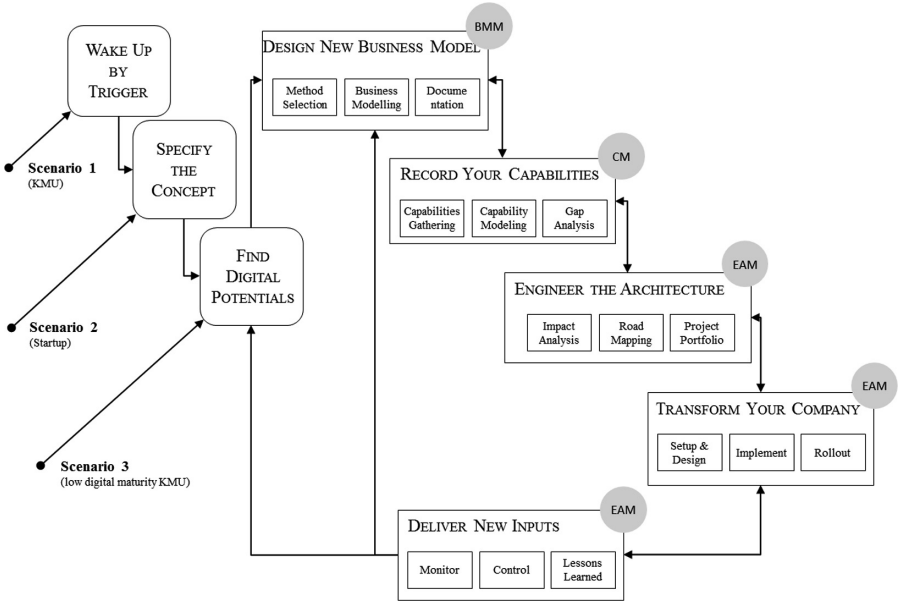


Fig. 1. Digital innovation and transformation process (DITP) adopted from [14].

personal preferences or the requirements of the project [16, 17, 21, 22]. In general, three activities have to be performed: *DITP3.1 Choice of business modelling approach, DITP3.2 Modelling of the business model, DITP3.3 Visualization of the business model & documentation*.

The phase of ***"Record Your Capabilities"*** has to focus on the question *"What capabilities are needed by the company to implement the developed business model?"*. Capabilities [20] are considered as missing link between business units and IT and therefore support the business IT alignment. The Capability Management Guide supports the identification, the structuring and the management of capabilities [15]. Three activities have to be carried out: *DITP4.1 Collection of required and existing capabilities, DITP4.2 Capability modeling, DITP4.3 Gap analysis and adaptations*.

In the phase **"Engineer the Architecture"**, the future enterprise architecture model is defined by means of the detected capabilities and the present architecture. The aim of this phase is to draft an enterprise architecture model including the required architecture objects and their dependencies. In this context, we propose the following activities: *DITP5.1 Choice of an approach for the impact analysis, DITP5.2 Development of the enterprise architecture including roadmap for integration, DITP5.3 Project portfolio for the implementation*.

The phase ***"Transform Your Company"*** focusses the activities *DITP6.1 Project set up, DITP6.2 Implementation and DITP6.3 Roll out* of the enterprise architecture. The planning of strategic initiatives and/or single projects provides the framework for the following change measures and the implementation of the planned enterprise architecture.

One of the most important functions of the final phase ***"Deliver New Inputs"*** is to safeguard the digitalization objectives. In order to monitor the implementation of digitalization objectives within the entire enterprise and the changes related to it, enterprises have to determine suitable indicators to be sustained by a corresponding data base [22].

3 Research Method

The primary research method used in our work is qualitative case study research. Qualitative case study is an approach to research that facilitates exploration of a phenomenon within its context using a variety of data sources. This ensures that the subject under consideration is not explored from only one perspective, but rather from a variety of perspectives which allows for multiple facets of the phenomenon to be revealed and understood. Yin differentiates various kinds of case studies [23]: explanatory, exploratory and descriptive. The case study presented and discussed in Sect. 4 can be classified as exploratory, i.e. we explore digital transformation process in its organizational context. As focus of the case study, we decided to address the innovation potential of digital signage solutions and the DITP use. More concrete, the case study explores the following questions addressing the perspectives:

RQ1: What digital innovation potential in housing industry exists from the perspective of data-driven services for tenants?

The primary focus of the case study is on understanding the innovation potential. In this context, the term "data-driven" services denotes in our work that the value proposition offered to a target group is to a substantial part based on collecting, analyzing and applying data.

RQ2: How can digital signage support this innovation potential?

The focus on digital signage is based on the intention of the case study company to evaluate the options created by this technology.

RQ3: How can the DITP support the digital innovation in the industrial case and what implications does the have for the DITP?

The secondary focus of the case study is on evaluating the DITP.

In the case study, we so far analyzed documents of the case study company, formed a focus group and performed interviews. More details are given in Sect. 4 when introducing the case study company and discussing the results.

4 Industrial Cases

4.1 Housing Industry: EDS Solutions "Doorshow" and Black-Board

The industrial case is a housing company owning and operating more than 200 residential buildings with 36.000 flats. Although many technological innovations were

proposed for housing industries, a lot of innovation potential remains in this area. Smart buildings, ambient assisted living and smart living are only some of many examples connected to introducing digital technology and IT into apartment buildings and flats. The area not yet sufficiently explored by research and industry is the possibility to integrate the business processes of housing companies with the processes of their suppliers and partners (e.g., for maintenance of buildings or facilities) and to offer additional services for the tenants as part of the buildings' infrastructure (e.g., individual assistance services or logistics services). Such services are part of the digital innovation in housing industries and subject of research in our project.

The current focus of the project is to explore and specify what services will be of value for the tenants and at the same time will be promising for the housing company. For selected services, we build prototypes and conduct acceptance tests. The main objective of the company is to improve comfort for the tenants and attractiveness of the buildings without charging extra for the services. Digital signage solutions, such as the elevator related products presented in the following, are a potential technical basis for the new services.

The technical solutions used in the industrial case resulted from cooperation with one of the market leaders in producing and operating elevators and escalators. This company with more than 100 years of history opened a new "digital" business line including elevator digital signage (EDS) solutions. The EDS builds upon a technical infrastructure available in all modern elevators which is independent of EDS: in the elevator shaft there is a data communication facility connecting every elevator cabin with a communication device on top of the shaft to the maintenance unit in the back-office. The data transmitted to the back office includes real-time sensor information and statistical data which are used for preventive maintenance and fault diagnosis purposes.

With this communication facility installed, also the EDS can be connected to the back office. The elevator manufacturer has detailed knowledge who operates the elevators and what kinds of users (i.e., target groups) are frequently using the elevator. This makes the EDS an interesting option for marketing campaigns or presenting information relevant for the organizations or individuals in the building. From a technical perspective, there are three main EDS devices:

- The "doorshow", a short-distance data projector mounted above the elevator door from the outside. For an individual or a group waiting for the elevator, this projector can display content on the elevator door.
- The interactive doorshow, which adds gesture recognition to the conventional doorshow. Persons waiting for the elevator can navigate through the displayed content by using gestures (see Fig. 2).
- The black-board which is a shock-proof touch screen mounted on the side of the elevator door (again from the outside) (see Fig. 3). The black-board is equipped with various sensors (RFID, NFC, camera, etc.) which potentially could be used to identify the person waiting or to collect context information.

Figure 2 shows the EDS in use providing the digital service of advertisements on elevator doors (demonstrated on Hanover Industrial Fair in April 2017). The elevator with the advertisement is visible on the right of the Figure with the short distance projector mounted above the door. The person in the centre of the picture (arm raised

Fig. 2. EDS device "interactive doorshow" in use: advertisements on elevator doors

Fig. 3. The black-board as demonstrated at Hanover Industrial Fair in 2018.

above his head) demonstrates the gesture recognition possibilities of the service to the visitors on the stand.

Figure 3 shows a snapshot of the black-board EDS solution. The blackboard software so far primarily supports two "roles" meant to represent stakeholder groups:

tenant and facility manager. For supporting individual processes and demands of tenants, in the first step an import of the tenant's profile information is foreseen.

4.2 Personalized Information on in-Company Screen

Another solution independent from the actual elevator digital signage case is based on the usage of information screen installed in company offices and aimed at utilizing possibilities related to presence of multiple people next to a digital signage screen to provide for personalized information based on the interests and preferences of these people. The complete developed framework can be split into three major components: viewer detection system, annotated content storage and content management system (Fig. 4).

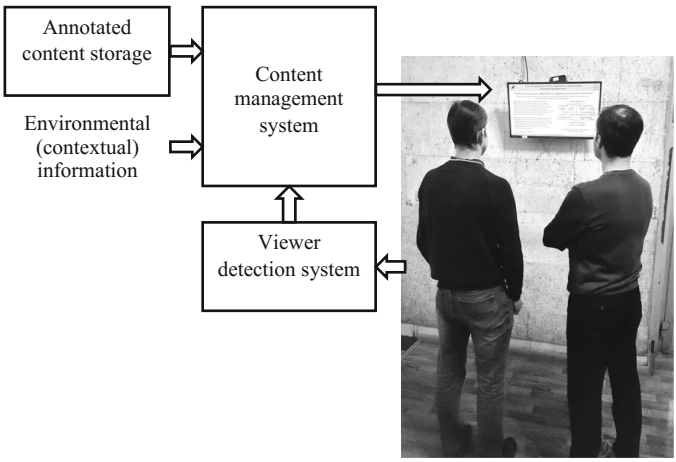


Fig. 4. The framework of the targeted digital signage system.

The viewer detection is based on the application of the WiFi technology. It is currently integrated into most of portable electronic devices. Though technologies of in-door positioning via WiFi are not precise enough without usage of additional complementing technologies, it is still possible to identify if the signal source is really close (in the range of couple of meters) to the receiver what meets the requirements of the case. Unlike WiFi, Bluetooth can only be used if it is either switched on in a discoverable mode (which is unlikely), or the personal device is paired with the scanning device (also not very likely). Identification of viewers via face recognition through a camera photographing the space in front of the digital signage is a complex task that requires additional research and it was not considered in the frame of the presented work.

The annotated content storage is a database storing the information pieces to be shown with tags that relate it to particular persons (e.g., Nikolay Shilov is asked to communicate to the service office), roles (e.g., an announcement to all parents of

children under seven), and environmental conditions (e.g., announcements related to upcoming week).

The content management system is responsible for integration of the environmental information (the context) including time, day of the week, weather, etc., the information about digital signage viewers and available content from the content storage.

When information is presented to a group of people the confidentiality issues become important. However, in a limited community (residents of one floor or company employees), some shared information related to the building floor, company or department can be distributed. An example of such information can be an upcoming events or meetings with their times and locations, announcements of new or changing facilities, advertisements, and other.

5 Case Work and Results

In the industrial case presented in Sect. 4.1, we studied the work on identifying digital innovation potential using EDS. This section first describes the DITP use for this purpose and afterwards investigates lessons learned regarding the research questions defined in Sect. 3.

5.1 DITP Use and Categorization of Services

In the housing company a focus group with managers of residential buildings and service process owners was established. This was the starting point for the "find digital potential" phase of DITP. Topic of the focus group was the digitization and innovation potential of tenant services. The focus group started by developing a list of stakeholder groups of relevance for digitization and of potential relevance for implementing innovations. Examples for stakeholder groups identified by the focus group are tenants (grouped according to different phases of their work life, e.g., students, works in shifts, retired, …), the housing company (again in different groups, like facility managers, administration, …), logistic service providers (mail, transport, delivery), personal service providers (hair dresser, medical personnel, …), relatives and friends to tenants.

With the stakeholder groups defined, an interaction matrix was created showing all stakeholder groups on one axis as potential recipient of value offerings and on the other axis as providers of value creation. In the cells of this matrix, potential value exchange was positioned. Example: the housing company might offer an additional service to tenants to allow logistics partners to deliver parcels to the tenants at a safe facility in the building. This service could be supported using the EDS by displaying the information to the tenant as soon as he/she returns to the building that a parcel arrived and how the access is possible.

More rewarding than the interaction matrix was the discussion of individual processes for the stakeholder groups and how they could be supported, i.e., individual process assistance. A simple example would be a tenant on his/her way to work could receive an information while waiting for the elevator which bus line close to the building would arrive next and to what bus stop to go. This information is only possible

if it is known that the "process" of the tenant is "go to work", where the work is and that riding the elevator is part of the way to work.

The next phase of DITP is on designing the new business model. As a starting point, based on the interaction matrix and the individual process analysis, a list of potentially interesting services, required information and expected user functionality was derived. Furthermore, one of the results of the focus group work was the differentiation of services for tenants regarding (a) information or process assistance and (b) individual or organizational assistance. As a precondition for describing these assistance categories, we have to define the terms process, context and situation:

- A process is a sequence of activities aiming at achieving a defined purpose.
- Context consists of context elements; each context element captures information about defined aspects or factors required for providing assistance.
- Situation is the status of all context elements at a certain point in time.

The assistance function categories are as follows:

- Individual assistance is directing to providing information relevant for the situation of an individual,
- Organizational assistance is directed to providing information to an individual relevant for the situation of an organization,
- Process assistance is providing information in time for supporting the activity at hand.
- Information assistance is providing information relevant for the actual situation (of an individual or an organisation).

Table 1 shows examples for individual/organizational and information/process assistance.

Table 1. Categories of EDS services with examples

	Information assistance	Process assistance
Organisational assistance	Information to employees about the closest restaurant with free capacity for lunch	Guiding technicians to the next device defined in a maintenance process
Individual assistance	Information to person about the fastest public transport option to the target location	Guiding a person to the office in the building where the next step in applying for a permit can be done

From a research perspective, our conclusion is that the value propositions originating from the analysis of individual and organisational processes are of particular interest for business service design.

Another part of the "design new business model" phase is to evaluate the identified technical building blocks for implementing innovative services. In the case study, the EDS devices presented in Sect. 4 are of particular interest. For individual assistance, both information and process assistance, an identification of the individual tenant is required. Without such an identification, the information for an individual's process or

demand profile cannot be provided. Only the black-board and the in-company screen offer the technical means for identification, for example by an RFID in the key ring of the tenant/employee or by face recognition and comparison to the personal profile in the back-office of the (housing) company (if the tenant/employee agrees to these methods).

Organisational assistance could be implemented by all three devices, if we assume that the situation to be supported is identified by evaluating context information from information systems in the organisation or by installed sensors. The interactive doorshow can be expected to have advantages over the conventional doorshow as the person to be supported at least could acknowledge that she/he received the information by using a defined gesture. With the black-board the highest level of organizational support can be expected, as the identification of individual employees would allow for documentation possibilities and for refining organizational assistance.

5.2 Discussion of Conclusions for Research Questions and Experiences

The experiences from the blackboard development and the previous EDS and interactive EDS development are discussed in this section from the perspective of the research questions defined in Sect. 3. Furthermore, the experiences can be used to derive lessons learned for future work.

RQ1: What digital innovation potential in housing industry exists from the perspective of data-driven services for tenants?

The focus group work in the industrial case returned a long list of potential innovations (see Sect. 5.1 for examples). The most promising ones are in the field of information and process assistance for individuals and organizations as illustrated in Table 1. User acceptance and economic viability of the services will have to be shown in future work by implementation of the services. The phases of DITP which were not yet performed in the industrial case describe a way for achieving this implementation.

RQ2: How can digital signage support this innovation potential?

The different EDS devices (see Sect. 4) offer significant support for implementing innovative services, in particular for individual assistance of tenants and for organizational assistance. Most valuable seems to be the blackboard due to its suitability for all identified services. The interactive doorshow and the black-board both were part of our earlier work where we produced proof-of-concept (PoC) implementations. This work is covered in other publications [5].

RQ3: How can the DITP support the digital innovation in the industrial case and what implications does the case have for the DITP?

The DITP proved to be a useful guide through the first phases of identifying digital innovation potential. In the phase “find digital potential”, the activity of analyzing stakeholder demands proved to be not only useful but a mandatory and valuable step in combination with analyzing trends and digital building blocks, i.e. this activity might be useful to integrate into the DITP. The “develop business model” phase of DITP was not performed completely, but only regarding the value creation process and the

identification of target groups. Here, the close inter-relation to the previous phase "find digital potential" and the next phase "integration into the enterprise architecture (EA) became clear. Regarding the EA, the necessity of an analysis of the existing EA was confirmed. This will be part of the future work.

The importance of the different dimensions of the DITP approach (i.e. business model, enterprise architecture, user experience, service design) was analysed based on the time consumed working on the dimensions and based on the perceived importance from the perspective of the engineers involved. Table 2 shows the result of this investigation for the PoC of the interactive elevator door solution compared to the blackboard solution. "1" indicates the highest and "4" the lowest perceived importance. The time consumption is presented in terms of the share of the overall development efforts for the PoC. The table clearly illustrates the greater weight of user experiences in the interactive elevator door show, which is not providing personalized information, (i.e., no "targeted digital signage") as compared to the blackboard which can be classified as targeted digital signage.

Table 2. Comparison of time consumption and perceived importance

	Interactive elevator door solution		Blackboard	
	Time consumption	Perceived importance	Time consumption	Perceived importance
User experience	32%	1	22%	3
Business model	22%	2	16%	1
Architecture	18%	4	33%	2
Service design	28%	3	29%	4

The comparison of the development processes, as illustrated in Fig. 5, confirms this impression. The figure shows in what sequence the different dimensions were tackled, beginning from the "S" (start) and finishing with the "F" (finish).

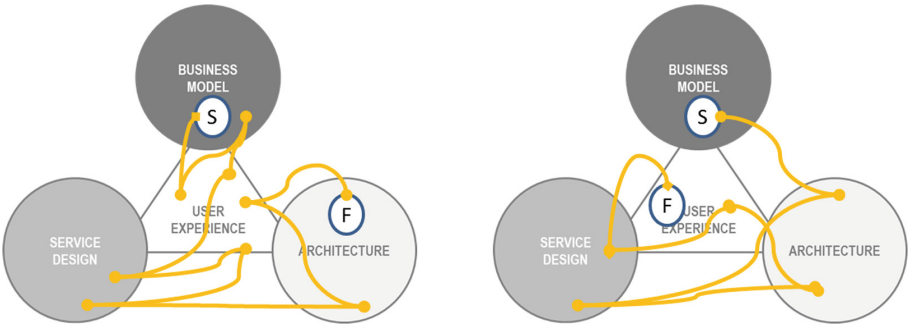


Fig. 5. Development paths for elevator door signage (left) and blackboard (right).

6 Summary and Future Work

This paper focuses on the field of elevator digital signage motivated application needs from housing industry. We identified categories of assistance service in digital signage as individual or organizational process or information assistance.

Future work will have to focus on technical and method aspects. The ideal digital signage solution would be able to provide individualized content for the audience consuming the information at any time: if only one person is present, the content would be personalized for this person, if a group of persons is present the demand of all should be met. This would require continuous recognition and identification of the current audience for the digital signage, their preferences and the matching content, which requires more work in developing our approach for intersecting preferences and integration with content aggregation, filtering and matching as well as user identification and profiling techniques.

The initial EDS solution developed in our previous work improved due to the new business requirements from the industrial partner from housing industry. The demonstration at the fair in Hanover created the impression that gesture recognition and display on elevator doors were more attractive for audiences seeking entertainment and leisure than for target groups with a certain personal plan or purpose on their minds. However, this impression was not seriously investigated at the fair. Thus, the third DSR cycle at the same time was supposed to address a differentiation of digital signage solution in door show, interactive door show and classical blackboards.

From method perspective, the technical work described above will contribute to completing the ongoing DITP application. Furthermore, the implementation of EDS solutions in housing industry will have to be integrated into the EA.

The biggest limitation of our work so far obviously is that we conducted only one case study in one application field. More case study work or experiments tackling usability, user acceptance, benefits and improvement potential are needed to reach generalizable conclusions. However, the focus of the work presented in this paper was on improving the understanding about potentials and limitations of EDS and digital signage. To this aim the paper provided a number of contributions.

Acknowledgements. The research was supported partly by projects funded by grants # 18-07-01201 and 18-07-01272 of the Russian Foundation for Basic Research, by the State Research no. 0073-2018-0002, and by Government of Russian Federation, Grant 08-08.

References

1. Gallacher, S., Papadopoulou, E., Abu-Shaaban, Y., Taylor, N.K., Williams, M.H.: Dynamic context-aware personalisation in a pervasive environment. Pervasive Mob. Comput. **10**, 120–137 (2014)
2. Anagnostopoulos, A., Broder, A.Z., Gabrilovich, E., Josifovski, V., Riedel, L.: Just-in-time contextual advertising. In: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, 272 p, pp. 331–340, November (2007)

3. Schaeffler, J.: Digital Signage: Software, Networks, Advertising, and Displays: A Primer for Understanding the Business, p. 212. CRC Press, Boca Raton (2012)
4. Want, R., Schilit, B.N.: Interactive digital signage. Computer **45**(5), 21–24 (2012)
5. Wißotzki, M., Sandkuhl, K., Smirnov, A., Kashevnik, A., Shilov, N.: Digital signage and targeted advertisement based on personal preferences and digital business models. In: Proceedings of the 21st Conference of Open Innovations Association FRUCT, 6–10 November 2017, Helsinki, Finland, pp. 375–381. ITMO University, St. Petersburg (2017)
6. Digital Signage Consortium, Digital Signage Consotium. http://www.digital-signage.jp/about/
7. Taniguchi, Y.: Content scheduling and adaptation for networked and context-aware digital signage: a literature survey. ITE Trans. Med. Technol. Appl. **6**(1), 18–29 (2018)
8. Bäckström, T.: Large digital screen as a corporate communication channel: enhancing the dissemination of information, administration and interaction. diva-portal.org, DIVA (2018)
9. Corenthy, L., et al.: Touchless tactile displays for digital signage: mid-air haptics meets large screens. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, p. D103. ACM, April 2018
10. Burke, R.R.: Behavioral effects of digital signage. J. Adv. Res. **49**(2), 180–185 (2009)
11. Bauer, C., Dohmen, P., Strauss, C.: A conceptual framework for backend services of contextual digital signage. J. Serv. Sci. Res. **4**(2), 271–297 (2012)
12. Dennis, C., Michon, R., Brakus, J.J., Newman, A., Alamanos, E.: New insights into the impact of digital signage as a retail atmospheric tool. J. Consum. Behav. **11**(6), 454–466 (2012)
13. Sandkuhl, K.: Information logistics in networked organizations: selected concepts and applications. In: Filipe, J., Cordeiro, J., Cardoso, J. (eds.) ICEIS 2007. LNBIP, vol. 12, pp. 43–54. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88710-2_4
14. Wißotzki, M., Sandkuhl, K.: The digital business architect – towards method support for digital innovation and transformation. In: Poels, G., Gailly, F., Serral Asensio, E., Snoeck, M. (eds.) PoEM 2017. LNBIP, vol. 305, pp. 352–362. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70241-4_24
15. Wißotzki, M.: Capability Management Guide, 1st edn. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-658-19233-4
16. Schallmo, D.: Geschäftsmodelle Erfolgreich entwickeln und implementieren: Mit Aufgaben und Kontrollfragen, 1st edn. Springer, Berlin (2013). https://doi.org/10.1007/978-3-662-57605-2
17. Osterwalder, A., Pigneur, Y.: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers, p. 2013. Wiley, Hoboken (2013)
18. Gartner, Gartner Hype Cycle 2016. http://www.gartner.com. Accessed 19 June 2017
19. Thoughtworks, Technology Radar. https://www.thoughtworks.com/de/radar. Accessed on: Jun. 19 2017
20. Offerman, T., Stettina, C.J., Plaat, A.: Business capabilities: a systematic literature review and a research agenda (2017)
21. Hoffmeister, C.: Digital Business Modelling: Digitale Geschäftsmodelle entwickeln und strategisch verankern. München: Hanser (2015)
22. Wirtz, B.W.: Electronic Business, 5th edn. Springer, Wiesbaden (2016)
23. Yin, R.K.: Case Study Research: Design and Methods. SAGE Publications, Thousand Oaks (2013)

Information Technology Management

Efficient and Sustainable Risk Management in Large Project Portfolios

Mats Danielson[1,2] and Love Ekenberg[1,2(✉)]

[1] Department of Computer and Systems Sciences, Stockholm University, PO Box 7003, 164 07 Kista, Sweden
mats.danielson@su.se
[2] International Institute for Applied Systems Analysis, IIASA, Schlossplatz 1, 2361 Laxenburg, Austria
ekenberg@iiasa.ac.at

Abstract. In an ever-increasing competitive business environment, it has become increasingly important to be able to obtain efficient and sustainable business operations not only by efficient core procedures but also by being able to minimise losses incurred by risk taking. The latter by handling both operational risks and financial risks in a unified model. This is important not least in businesses that handle sets of simultaneous large projects, which is the topic of risk handling in project portfolios. In this paper, we present a novel method for business risk handling for project portfolios under strong uncertainty. The method is based on event trees representing each adverse consequence modelled, together with mitigation costs and effects. The aggregation of all consequences for all projects together constitutes the risk portfolio for the business. This method is used in one of Sweden's largest manufacturing enterprises having a vast portfolio of projects in the form of ongoing tenders for orders.

Keywords: Efficient risk analysis · Probability theory · Belief distribution
Sustainable risk management

1 Introduction

While risk taking is integral to every business, risk management is paramount in taking advantage of it. Since risk is unavoidable and ever present, what is required of a successful business is being substantially better at managing risk than its competitors. Thus, the ultimate criterion of a risk managing strategy is its performance relative to competitors in the same business sectors. This paper describes a risk handling method, based on advanced decision theoretic and statistical methods, which has been implemented in a computer tool that currently runs in large corporate settings.

In this article, we present a novel method for business risk handling for project portfolios under strong uncertainty. This method is, for instance, currently used in one of Sweden's largest manufacturing enterprises having a vast portfolio of projects in the form of ongoing tenders for orders. Each tender, and then each successful tender becoming an order, is one entry in the project portfolio. The annual turnover of the enterprise is in excess of 20 billion Euro, and the estimated reduction in losses

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 143–157, 2018.
https://doi.org/10.1007/978-3-319-99951-7_10

attributed to efficient risk handling in the portfolio is between 1–2% of the total order value as estimated by internal auditing within the enterprise. Like many other methods, we utilise event trees for the modelling of the problems, but refrain from using precise numbers. The various estimates could instead be made by using intervals and qualitative assessments. It has, however, turned out that this is not sufficient in many cases as we have discussed in a number of articles regarding decision analysis of event trees, cf., e.g., [8, 23]. From one point of view, the pure interval approaches have some attractive features because we need no assumption regarding the underlying distributions, and the evaluation of decision alternatives can be handled by contraction of the intervals combined with several complementary evaluation rules. Thus, the advantage with the distribution-free approaches is the generality and freedom from assumptions that it allows. However, the interpretation of the results might then be complicated to interpret and there might also be problems in discriminating between the various mitigation measures and their efficiency. In order to alleviate that problem, a new evaluation method is introduced in this article. This is based on a belief mass interpretation of the output intervals, without introducing any further complicating modelling aspects into the problems. This interpretation allows the user (and thus the company) to see how much belief there is in gaining or losing value from each and every mitigation effort considered for the set of adverse consequences that have been specified for a particular project. Without explicitly modelling belief in various risks and their mitigations, it would not be possible to tell which mitigations to apply or not. In the next section, we briefly describe risk analytical approaches to the risk management part of handling portfolio risks in business processes. Thereafter, we provide the conceptual model for our method and explain the input data format. Section 5 enlarges on the evaluation of the value of a portfolio and how this is related to belief modelling, and in Sect. 6, we illustrate the ideas with a small example. Section 7 concludes the paper.

2 Risk Analytical Approaches

A quite usual technique in many risk analytical methods is the use of event trees. Probabilities and costs are assigned to the tree nodes and usually some aggregation of the values is used to get an estimation of, e.g., an expected cost. Most often however, when evaluating the expected cost of a simple incident, the probability and cost variables are provided as numerically precise estimates, putting quite extreme demands on the input capability of the risk analyst. Another problem is that most risks evaluate to a medium risk level with no indication of how to rank the risks within that level. Already a mediocre risk analyst is usually capable of differentiating between disastrous, unacceptable, and acceptable risks without the aid of analysis tools at all, so an important problem remains, i.e., to decide the reduction order of the different unacceptable risks. Hence, when the risk situation is obvious, there is little need for a model, and when it is not, the models are offering none or little help.

Qualitative methods in risk analysis may be used as screening methods in the preliminary risk analysis. A problem here is, however, that it is difficult to really discriminate between various risks regarding their severity, in particular in the mid

segment. Quantitative methods could be better in this respect and are usually divided into deterministic and probabilistic analyses. The deterministic analysis describes the hazards in terms of the severity of the consequences. The probabilistic approach determines the risk in terms of frequency and consequences. Various methods for risk analysis have been developed over the years, such as what if/checklist, hazard and operability (HAZOP) analysis [16], failure modes and effects analysis (FMEA) [18], event tree analysis (ETA) and fault tree analysis (FTA) [5], cause-consequence analysis [19], management oversight risk tree (MORT) [13], and safety management organization review technique (SMORT) [15]. During the last 70 years, various methods for relaxing the strong requirements of precise assignments have been proposed, which have also influenced risk analytical methods. These are, among others, based on sets of probability measures, upper and lower probabilities [3] as well as interval probabilities and utilities [12], fuzzy measures [1, 23] and representations based on the theories of evidence (belief functions) and possibility theory, cf., e.g., [9, 10, 22]. However, these are used in real business process management to a seemingly limited extent. In many cases, this seems to depend on that these more realistic models are perceived as too difficult to use for common practitioners. Furthermore, the complexity often introduced by extended methods can be problematic from a computational viewpoint.

In contrast to feature modelling and feature-oriented domain analysis (which are diagrams of hierarchies of features pertinent to a set of objects, for example in a product portfolio), our method models risks associated with projects and not the features of the products themselves. Further, the method uses decision theoretic event trees to model possible consequences, while for example the DMN (Decision Model and Notation) standard for decision modelling uses decision tables. In this way, the model in this paper adheres more to traditional decision analysis employing event trees, albeit of a particular structure, thus being risk-specific. The event trees have exactly two consequences, with one of them carrying the exact value of zero. Further, the consequences are valued on a ratio scale, not an interval scale as is more common. Those properties lead to more feasible computational loads while not incurring any modelling penalties in the sense of restrictions. Input to the model is taken from other parts of the business processing system that handles all tenders and orders.

3 Risk Modelling

Many larger businesses have a large number of ongoing activities and tenders which could be considered a portfolio of projects. The management of the risks in the portfolio benefit from taking an enterprise-wide approach to risk measurement and mitigation. In this model, each input and output interval consist of a lower bound, an upper bound, and a focal point as utilised already in, e.g., [6, 7], but now with another significance. These three points are interpreted as parameters for belief distributions for probabilities, values and (when multiple criteria are considered) criteria weights. The belief distributions are then combined using combination rules (addition and multiplication) for random variables.

Furthermore, in all reasonably well-managed portfolio, all projects above some size or complexity threshold is considered a separate portfolio entry. Therefore, for each

portfolio entry, a risk project is created to hold the risk information about the project. We first formalize this in some more details.

Definition: A *risk project* $P = \langle R, T \rangle$ consists of a consequence part and a structural part. The consequence part is a risk set R and the structural part is a binary risk tree T.

The risk project contains a set of risks which can either occur or not occur (binary risk with two outcomes). It is desirable that the risk does not occur since the value of a risk occurring is negative, i.e. it incurs a loss for the business.

Definition: A *risk set* R is a set $\{R_j\}$ of binary risks. A *binary risk* R_i (henceforth *risk*) consists of a set $\{C_i, \overline{C_i}\}$ of two consequences and a risk descriptor r_i. The two consequences entail that the risk occurs (C_i) or it does not ($\overline{C_i}$). The consequence set $\{C_i, \overline{C_i}\}$ is exclusive and exhausting, i.e. exactly one of the consequences occur.

To formally describe the risk, we need to estimate the probability that it will occur and the size of the loss (the negative value) of the incidence. This is collected together in a risk descriptor which contains the information required for a formal analysis.

Definition: A *risk descriptor* $r_i = \{P(C_i), V(C_i)\}$ contains the description of a probability distribution and of a cost distribution for the risk. Each risk consists of a consequence pair in which a risk does or does not occur with probabilities $P(C_i)$ and $P(\overline{C_i}) = 1 - P(C_i)$ respectively. The cost of occurring is $V(C_i)$ and the cost of not occurring is zero ($V(\overline{C_i}) = 0$), i.e. the valuation operates on a ratio scale.

The risk descriptor must contain information on how probable it is for the risk occurring, i.e. how the probability is distributed. It is very seldom the case that a probability is known with certainty. Rather, there is a belief distribution over a set of probabilities, usually over an interval. This belief is encoded in a probability distribution for each risk [20].

In the model, different distributions are utilised for probabilities and values. This is necessary because of the normalisation constraint for probabilities, requiring them to add to one for a set of exhaustive and mutually exclusive events in the model. This leads to the use of Dirichlet distributions for probabilities and two- or three-point distributions for values or costs.

The Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a vector $\boldsymbol{\alpha}$ of positive numbers. It is a multivariate generalization of the beta distribution. The Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_K > 0$ has a probability density function given by a function of those parameters that depend on a beta function and the product of the parameters x_i. Thus, $x_1, \ldots, x_K$ belongs to a standard simplex. The normalizing constant in the density function is the multivariate beta function (which in turn can be expressed in terms of the gamma function) [14].

The marginal distributions of Dirichlet are beta distributions. The beta distribution is a family of continuous probability distributions defined on [0, 1] and parameterized by two shape parameters, α and β, that control the shape of the distribution. The probability density function of the beta distribution, for $0 \leq x \leq 1$ and shape parameters $\alpha, \beta > 0$, is a power function of the variable x and its reflection $(1 - x)$.

Since both the Dirichlet and beta distributions operate on the entire [0, 1] range, they are unsuitable to use unaltered in this risk management model. We instead employ

a *bounded* Dirichlet distribution that operates on a user specified [a_i, b_i] range instead. It then gives rise to likewise bounded beta distributions, the so called four-parameter beta distributions, which are defined only on the user specified range.

Definition: A *probability distribution* is a three-point distribution $P(C_i) = f_3(a_i, c_i\ b_i)$. The Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a vector of positive numbers. It is a multivariate generalization of the beta distribution. A *bounded* Dirichlet distribution is used in this context, which means that the outer boundaries are not the traditional [0, 1] but a narrower interval specified by the user. This is expressed as a function $f_3(a_i, c_i\ b_i)$ where c_i is the most likely probability and a_i and b_i are the boundaries of the belief with $a_i < c_i < b_i$ [16]. This is discussed further in the next section.

Further, the risk descriptor must contain information on how big the loss would be if the risk occurs, i.e. how the value is distributed. It is again very seldom the case that the value of a loss is known with certainty. Rather, there is a belief distribution over a set of value, usually over an interval. This belief is encoded in a cost distribution for each risk.

Definition: A *cost distribution* is either a two-point distribution $V(C_i) = h_2(a_i, b_i)$ or a three-point distribution $V(C_i) = h_3(a_i, c_i, b_i)$. The uniform, trapezoid and triangular distributions are used in this context as representing value distributions. This is also discussed further in the next section.

There would be little use for a risk model unless it could also take mitigations into account. The idea is to be able to express the (negative) expected value, i.e. expected cost, of the portfolio prior to a mitigation effort (or a set of mitigation efforts) and compare to the post-mitigation value taking the mitigation effects and the cost of the mitigation itself into account. This enables the decision-makers to evaluate and decide on which mitigations should be applied to each risk in the portfolio.

Definition: A *mitigation* M_i is a modification to one or more risks in which a set $D_i = \{r_1', \ldots, r_s'\}$ of new risk descriptors (new probability distributions and/or new cost distributions) are activated. There is not any modification of the structure or the consequences of any risk. Activating a mitigation incurs a cost m_i. In return for this cost, the risk descriptors r_j are replaced with new descriptors r_j' from the set D_i.

Each risk is then expressed as a risk tree, where each event has two outcomes: the risk occurs or it does not. In the unmitigated scenario, the cost for the risk occurring is specified, while the risk for it not occurring is zero. The same is true for the mitigated scenario with the mitigation cost added. Note that in the figure, costs are positive, i.e. should be minimised.

Definition: A *risk tree* can be seen as a decision tree in which there is only one alternative. Binary risk trees are a subset of risk trees in which every event node has two consequences, the risk occurs or it does not.

As an illustration, consider in Fig. 1 a building contractor deciding on securing a construction site to protect it from floods. If unmitigated, a flood could damage the project severely. When mitigated, a flood can still occur, but most effects will be mitigated by the levees and if an overtop occurs, the consequences will be much less dire to the project. In the figure, the cost of mitigation is 10 and it lowers both the probability of occurring and the severity if an overtop occurs as a result of a levee breaking.

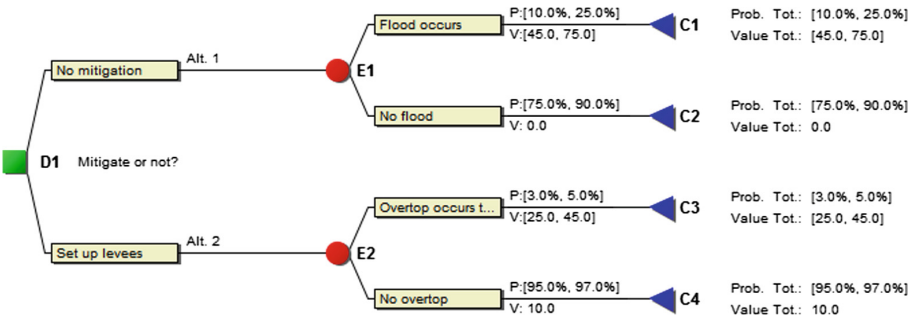


Fig. 1. Binary risk tree for decision to protect against flood or not

A risk tree is thus a decision tree in which there is only two alternatives, to mitigate or not. Binary risk trees are a subset of risk trees in which every event node has exactly two outcomes. These are exclusive and exhaustive; the risk occurs or it does not with probabilities p and $1 - p$ respectively. The value of not occurring is zero, i.e. the evaluation operates on a ratio scale.

Definition: A *binary risk tree* T is a subset of risk trees in which each event node (E_i) has two consequences; the risk occurs (C_i) or it does not ($\overline{C_i}$). The consequence set $\{C_i, \overline{C_i}\}$ is exclusive and exhausting, i.e. exactly one of the consequences occur. A risk R_i is associated to the binary risk event. Thus, binary risk trees consist of consequence pairs in which an event does or does not occur with probabilities $P(C_i)$ and $P(\overline{C_i}) = 1 - P(C_i)$ respectively. The event E_i is associated to a risk descriptor r_i with a probability distribution $P(\cdot)$ and a cost distribution $V(\cdot)$.

4 Input Information

The model allows for two kinds of specification of risks and costs. When little is known about the underlying distribution of a value or cost, it is reasonable to use a two-point distribution for modelling requiring only upper and lower bounds. In the model, the uniform and trapezoid distributions are employed as the preferred two-point distributions. But if the modal outcome is also known or could be reasonably estimated, then the probability of the outcome can be better represented by a three-point distribution. The triangular distribution, along with the Beta and Erlang distributions, is therefore widely used in project management models (such as PERT) to model events which take place within an interval defined by a minimum and maximum value. The triangular distribution yields results similar to Beta-PERT in general, and to Erlang-PERT in particular.

Thus, the triangular distribution is used as a representative for a class of distributions (three-point distributions) commonly used in business management. It is used as a subjective description of a population for which there is only limited sample data, and especially in cases where the relationship between variables is known but data is scarce (possibly because of high cost of collection). It is based on best estimates of the minimum and maximum as well as the modal value. The PERT distribution is a classic, originating from uncertainty in the development of the Polaris missile in the early

1960s. It has been popular with planning sequence models but less so in risk management due to its emphasizing the centre of the distribution. For example, Golenko-Ginzburg [11] studies a large number of PERT networks and draw the conclusion that there is a problem with the most likely estimate which shapes the distribution. Compared with other common three-point distributions, the triangular distribution considers the entire interval $[a_i, b_i]$ to a larger extent and is thus a better and more consistent companion to the two-point uniform distribution in the method. The statements by the decision-maker are interpreted in a more similar way which adds consistency. To compare them, the mean value of several three-point distributions is $\mu(\lambda) = (a_i + b_i + \lambda c_i)/(\lambda + 2)$ with the triangular distribution having $\lambda = 1$. Beta-PERT usually has $\lambda = 4$ and Erlang-PERT has $\lambda = 3$ but various other parameter values including non-integers have been suggested to represent the user input.[1] In practise, a higher λ value tends to underestimate the uncertainty. The triangular distribution is less centre-weighted and thus less prone to underestimation. Thus, there are no compelling reasons to use any other three-point distribution than triangular in business modelling. Consequently, the triangular distribution can be seen as a very good representative of the class of three-point distributions. Together with uniform and trapezoid distributions, they cover a very wide range of management risk modelling needs.

5 Evaluation

A risk portfolio is evaluated using the ordinary expected value. Thus, for a portfolio with m risks (mitigated or not), the expected value is the sum over the m risks

$$\mathrm{EV} = \sum(\mathrm{p_i} \cdot \mathrm{v_i} + \overline{\mathrm{p}}_\mathrm{i} \cdot \overline{\mathrm{v}}_\mathrm{i}), \text{where } \overline{\mathrm{p}}_\mathrm{i} = (1 - \mathrm{p_i}) \text{ and } \overline{\mathrm{v}}_\mathrm{i} = 0 \text{ according to the risk descriptor.}$$

To this is added the mitigation costs $\mathrm{m_i}$, yielding the final portfolio expected value [7]

$$\mathrm{PEV} = \sum(\mathrm{p_i} \cdot \mathrm{v_i} + \overline{\mathrm{p}}_\mathrm{i} \cdot \overline{\mathrm{v}}_\mathrm{i}) + \sum(\mathrm{m_i}), \text{ where again } \overline{\mathrm{p}}_\mathrm{i} = (1 - \mathrm{p_i}) \text{and } \overline{\mathrm{v}}_\mathrm{i} = 0.$$

Since this is a business model, we use the principle of going concern. In this case, it means that all assets and expectations should be valued relative to the assumption that the business will operate for an indefinite time into the future. This assumption leads to the occurrence of a large number of portfolio events which can then be approximated with a normal distribution. But there are two complications that must be taken into account when evaluating the model according to the going concern principle:

1. It will be normally distributed only if the input distributions are symmetric. But neither beta nor triangular distributions are symmetric, thus giving rise to skew-normal distributions instead.

[1] The two-point uniform and trapezoid distributions have $\lambda = 0$.

2. It will be normally distributed only if the input distributions and their aggregations allow for long tails, i.e. not being bounded by lower and upper limits. But the formulation of bounded Dirichlet as well as the uniform and triangular distributions all contain lower and upper bounds, thus giving rise to truncated normal distributions instead.

We will discuss those complications in turn [4]. The skew-normal distribution is a continuous probability distribution that generalises the normal distribution to allow for non-zero skewness by introducing α as a shape parameter. When $\alpha = 0$ the result is the standard normal distribution. When $\alpha = 1$ it models the distribution of the maximum of two independent standard normal variates. When the absolute value of the shape parameter increases, the skewness of the distribution increases. The limit as $|\alpha| \to \infty$ results in a folded normal distribution or half-normal distribution. The distribution is right-skewed if $\alpha > 0$ and left-skewed if $\alpha < 0$. When α changes its sign, the density is reflected about $x = 0$. The cumulative distribution function is expressed as a difference of two terms, where the latter is Owen's T-function $T(h, a)$ [21]. It gives the probability of the event ($X > h$ and $0 < Y < a \cdot X$) where X and Y are independent standard normal random variables. By utilising the skew-normal distribution instead of a normal, asymmetric input can be catered for.

The truncated (skew) normal distribution is the probability distribution of a (skew) normally distributed random variable whose value is either bounded from below, from above, or both. Without loss of generality, suppose that a distribution X has a normal distribution and lies within the interval (a, b). Then X conditional on $a < X < b$ has a truncated normal distribution. Its probability density function, f, for $a \leq x \leq b$, is given by a four-parameter expression that tends to normality as the intervals are widened.[2]

The binary risk tree is evaluated using the B-normal distribution method. The project risk cost distribution is derived from the individual risks in the project. The evaluation result of binary risk trees in business use does not approach a normal distribution even as the number of events increases, not even as a limiting case for an infinite number of events. The B-normal distribution method is based on a modified normal distribution adapted to business use. It expresses the resulting distributions of probabilities and costs of events in the context of business operations. The joint probability and the resulting cost of all modelled consequence pairs is approximately B-normally distributed with the approximation rapidly improving as the number of consequence pairs increases. As the limiting case for an infinite number of events, the risk cost distribution is exactly B-normal. The adaptions of B-normality to binary risk trees for business use consist of

- interpolated truncation toward the orthogonal hull
- control of approximation by successive limiting
- interpretation of binary risk trees as going concern.

In the use of the B-normal distribution, the going concern principle of business administration leads to the B-normal approximation of the total (joint) multi-variate

[2] Refer to any textbook on statistics for particulars, such as [17].

cost/probability distribution for an entire business operation. In the approximation, the tails are cut off for values outside the orthogonal hull of the random variables and the cumulative distribution function for the tails is interpolated between the cut-off points and the tangents of the B-normal CDF curve instead of a steep truncation as in a truncated normal distribution.

The cumulative distribution function (CDF) describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to x. For every real number x, the cumulative distribution function of a real-valued random variable X is given by $F_X(x) = P(X \leq x)$ where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x. The probability that X lies in the interval $(a, b]$, where $a < b$, is therefore $P(a < X \leq b) = F_X(b) - F_X(a)$ where $(a, b]$ is a semi-closed interval. The CDF of a continuous random variable X can be defined as an integral of its probability density function f. The function f is equal to the derivative of F almost everywhere, and is the probability density function of the distribution of X.

In the use of the B-normal distribution, the going concern principle of business administration leads to the B-normal approximation of the total (joint) multi-variate distribution for an entire business operation. In the approximation, the tails are cut off for values outside the orthogonal hull of the combined random variables and the cumulative distribution function for the tails is interpolated between the cut-off points and the tangents of the B-normal CDF curve instead of a steep truncation as in a truncated skew-normal distribution.

6 Example

In the following, we will illustrate the model by an example. Before commencing, we will define some terminology that will be useful in interpreting the example:

Expected risk exposure. The expected value of the risk exposure posed.

Risk exposure limit. The highest value (at the confidence level, in the example 95%) of the risk exposure posed by the project.

S-curve. The cumulative density function (CDF) of the resulting risk cost distribution. Shows the probability of incurring at most a specific cost from project risks. It is popularly called an S-curve due to its shape.

Mitigation strategy net gain. The gain from employing the entire set of mitigations for a project. The gain is defined as below. A gain of 0 means that the expected return is the same as the mitigation cost, leading to a zero net gain. A gain of 1 means that the expected return is the twice the mitigation cost, leading to a net gain in expected return of the same amount as the mitigation cost.[3]

Single mitigation net gain. The gain from employing a specific mitigation. Mitigation gains are not additive; thus, the sum of the mitigations' net gain is not the strategy mitigation net gain.

A construction company is under contract to erect a large building at a foreign site. Apart from the usual project risks of (**A**) substantial cost overrun and (**B**) substantial time overrun, they have identified four additional major risks for this project: (**C**) fluctuations in currency exchange rates above a threshold; (**D**) risk of buyer defaulting; and risk of adverse weather conditions leading to either (**E**) transport roads closed or (**F**) a flooded site. The mitigations for the risks are considered by management to be respectively: (**A**) tighter subcontractor supervision; (**B**) reinforced local project management; **C**) acquiring futures or options for part of the order value; (**D**) prepayments (lower than post-payments) and insurance; (**E**) transporting material ahead of rain season; and (**F**) starting construction earlier and have the completed building standing empty until the delivery date. For each risk, information is collected regarding the probability of it occurring, the cost if it occurs, and the mitigation costs along with the mitigation benefits in terms of reduced probabilities and costs. This leads to the risk situation in Table 1 for the unmitigated risks and Table 2 for the mitigated ones. All probabilities in percent and costs in MEUR (Fig. 2). The model allows for one mitigation to influence more than one risk, but this feature is not explored in the example[3].

Table 1. Unmitigated probabilities and costs.

Unmitigated risk	Probabilities			Costs		
	Minimum	Most likely	Maximum	Minimum	Most likely	Maximum
A	30	45	65	1.2	1.8	3.0
B	25	40	55	1.6	2.2	2.8
C	20	30	60	1.0	1.3	1.8
D	2	3	5	2.0	2.5	5.0
E	5	10	20	2.2	3.0	4.4
F	5	10	15	0.8	1.2	2.2

For each risk, there are two alternative courses of action: either A_1: leave it unmitigated or A_2: mitigate it, leading to reduced vulnerability but at a mitigation cost. A schematic set of trees for risks A through C are shown in Fig. 3. The total risk is the sum of all individual risks.

The risk management process proceeds with evaluation combinations of sets of mitigations, ranging from none to all. The business software supporting the tender/order handling process presents the mitigations from most to least efficient for

[3] Let $y = cdf(x, \mathrm{R})$ be the cumulative probability for the cost x for the project risk R. Then the inverse $x = cdf^{-1}(y, \mathrm{R})$ yields the risk exposure limit for the risk level y. Let $mean(\mathrm{R})$ be the mean cost for the project risk R. Then the following is a risk measure for $y > 0.75$:

$t(y, r) = (cdf^{-1}(y, \mathrm{R}) + mean(\mathrm{R}))/2$

The *net gain* measure for a mitigation M (with mitigation cost $m(\mathrm{M})$) compared to an unmitigated R at the risk level y is $gain(y, \mathrm{R}, \mathrm{M}) = (t(y, \mathrm{R}) - t(y, \mathrm{M}))/m(\mathrm{M}) - 1$ if $m(\mathrm{M}) > 0$.

Table 2. Mitigated probabilities and costs.

Mitigated risk	Mitigation cost	Mitigated probabilities			Mitigated costs		
		Minimum	Most likely	Maximum	Minimum	Most likely	Maximum
A	0.4	5	10	20	Same	Same	Same
B	0.6	10	15	30	Same	Same	Same
C	0.25	10	15	30	Same	Same	Same
D	0.3	0.5	1.0	1.5	0.7	0.9	1.2
E	0.3	2	2.5	3	1.2	1.3	1.5
F	0.2	3	5	8	0.2	0.4	0.5

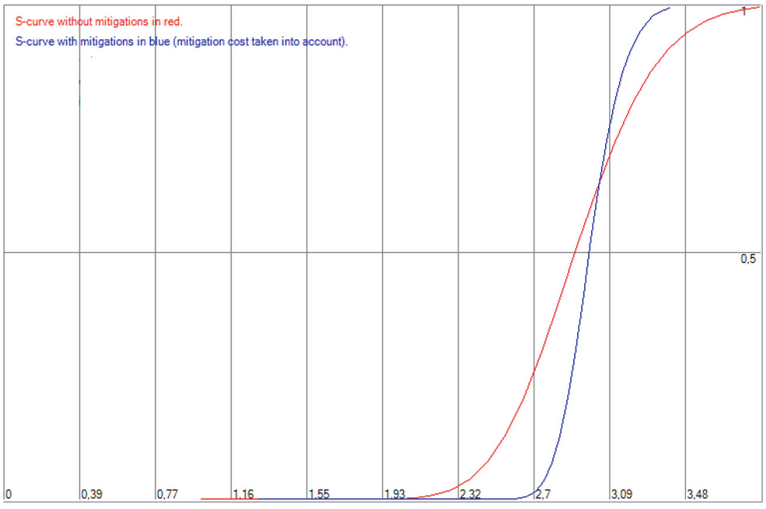


Fig. 2. S-curves of project risk unmitigated and fully mitigated (Color figure online)

the user to select which ones to include among the mitigations chosen for the particular project being handled.

The effects can be seen of employing all mitigations compared to not mitigating at all. The net effect of all mitigations when mitigation costs have been taken into account is that the mitigations bring the risks much more into control (the s-curve is steeper) but the gain in performance is not clear-cut. The expected cost[4] is slightly higher for the mitigated case, but the variance is lower, yielding a positive net gain according to the definition in the beginning of the section emphasizing lower variance, i.e. higher predictability. The summary of the risk analysis is shown in Table 3 using the terminology introduced in the beginning of the section.

[4] The expected cost is not the crossover at 50% CDF since it is asymmetric.

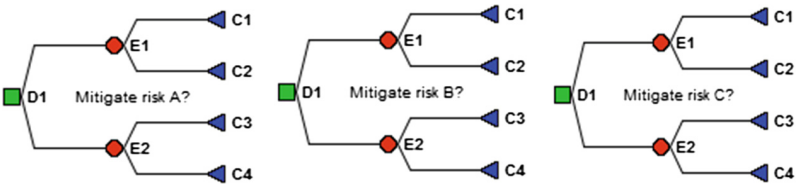


Fig. 3. Alternative courses of action for risks A–C

Table 3. Summary of efficiency of mitigation strategy.

Expected risk exposure	2.93
Risk exposure limit (at 95% confidence)	3.51
Mitigated expected risk exposure	3.00
Mitigated risk exposure limit (at 95% conf.)	3.25
Mitigation strategy efficiency	1.06
Mitigation strategy net gain	0.06

Next, a subset of the mitigations needs to be considered. The tool is able to list the mitigations in decreasing efficiency order, see Table 4.

Table 4. Mitigations in decreasing efficiency order.

Mitigation of	Efficiency
Risk A	2.13
Risk E	1.50
Risk C	1.27
Risk B	0.90
Risk F	0.78
Risk D	0.41

From this analysis, it is evident that only mitigations to risks A, C, and E are efficient in the sense that they contribute more than they cost. Thus, only these are considered further. The other three risks are left unmitigated. This leads to another subsequent risk evaluation displayed in Fig. 4.

The resulting set of mitigations is now efficient (see Fig. 4 and Table 5), leading to a sustainable mitigation strategy for this particular project in the portfolio. The efficiency of the mitigation strategy is 1.47, meaning that for each Euro put into the mitigation efforts of the strategy, an additional 0.47 Euro is gained.

Each project in the business is analysed using the same methods and criteria, leading to a business-wide policy implementation yielding an efficient risk management practice.

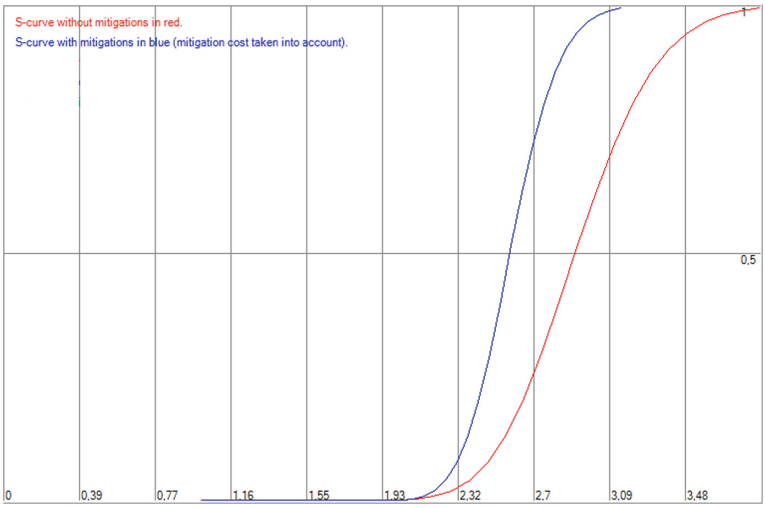


Fig. 4. S-curves of project risk unmitigated and partly mitigated (Color figure online)

Table 5. Summary of efficiency of new mitigation strategy.

Expected risk exposure	2.93
Risk exposure limit (at 95% confidence)	3.51
Mitigated expected risk exposure	2.59
Mitigated risk exposure limit (at 95% conf.)	2.94
Mitigation strategy efficiency	1.47
Mitigation strategy net gain	0.47

7 Concluding Remarks

There is a need to handle both operational risks and financial risks in a unified model, which is important not least in businesses that handle sets of simultaneous large projects and we have presented a novel method for business risk handling for project portfolios under strong uncertainty. We utilize event trees for the modelling of the problems, where the various estimates can be done by intervals and qualitative assessments, utilizing the advantages with distribution-free, while adding an evaluation method based on a belief mass interpretation of the output intervals, without introducing any complicating modelling aspects into the problems under consideration. The underlying idea is to maximize the efficiency of considerate mitigation measures where the planned response will not violate a tolerable level of effectiveness measured by means of several different metrics. Hence the approach conforms to the as-low-as-reasonably-practicable-principle (ALARP) [2]. The issues here are that development of plans and maintenance of response capacity are always bounded by budget constraints, and assessments of response effectiveness is subject to imprecision due to that there is no way to accurately measure an emergency response prior a catastrophic event. But at

the same time, this uncertainty can, and must, be taken into consideration and managed by rational means. The method is in use in one of Sweden's largest manufacturing businesses, having led to substantial reductions in risk costs.

Acknowledgements. This research was partly funded by strategic grants from the Swedish government within ICT – The Next Generation.

References

1. Abdo, H., Flaus, J.M.: Uncertainty quantification in dynamic system risk assessment: a new approach with randomness and fuzzy theory. Int. J. Prod. Res. **54**(19), 5862–5885 (2016)
2. ALARP Guidance Note, Commonwealth Offshore Petroleum and Greenhouse Gas Storage (Safety) Regulations 2009. NOPSEMA (2015)
3. Aven, T., Zio, E.: Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliab. Eng. Syst. Saf. **96**, 64–74 (2011)
4. Azzalini, A.: A class of distributions which includes the normal ones. Scand. J. Stat. **12**, 171–178 (1985)
5. Clemens, P.L., Simmons, R.J.: System safety and risk management. NIOSH instructional module. In: A guide for Engineering Educators. National Institute for Occupational Safety and Health, Cincinnati, Ohio, IX–3 – IX–7 (1998)
6. Danielson, M., Ekenberg, L., Larsson, A.: Distribution of expected utility in decision trees. Int. J. Approx. Reason. **46**(2), 387–407 (2007)
7. Danielson, M., Ekenberg, L.: A Method for Decision and Risk Analysis in Probabilistic and Multiple Criteria Situations. U.S. Patent No. 11/145229 (2004)
8. Danielson, M., Ekenberg, L., Larsson, A., Riabacke, M.: Weighting under ambiguous preferences and imprecise differences in a cardinal rank ordering process. Int. J. Comput. Intell. Syst. **7**(1), 105–112 (2014)
9. Dubois, D.: Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information. Risk Anal. **30**(3), 361–368 (2010)
10. Dutta, P.: Human health risk assessment under uncertain environment and its SWOT analysis. Open Public Health J. **11**, 72–92 (2018)
11. Golenko-Ginzburg, D.: Controlled alternative activity networks in project management. Eur. J. Oper. Res. **37**, 336–346 (1988)
12. Huang, L.: The failure-tree analysis based on imprecise probability and its application on tunnel project. Chem. Eng. Trans. **59**, 463–468 (2017)
13. Johnson, W.G.: MORT – The Management Oversight and Risk Tree. SAN 821-2, U.S. Atomic Energy Commission (1973)
14. Kendall, M.G., Stuart, A.: The Advanced Theory of Statistics. Volume 1: Distribution Theory, 3rd edn. Griffin (1969)
15. Kjellén, U., Tinmannsvik, R.K.: SMORT: Säkerhetsanalys av industriell organisation. Stockholm:Arbetarskyddsnämnden, in Swedish (1989)
16. Kletz, T.: Hazop and Hazan, Notes on the Identification and Assessment of Hazards. Institution of Chemical Engineers, Rugby (1974)
17. Loeve, M.: Probability Theory. Graduate Texts in Mathematics, vol. 45, 4th edn. Springer, Heidelberg (1977). https://doi.org/10.1007/978-1-4684-9464-8
18. Neal, R.A.: Modes of Failure Analysis Summary for the Nerva B-2 Reactor. Westinghouse Electric Corporation Astronuclear Laboratory. WANL–TNR-042 (1962)

19. Nielsen, D.S.: The cause consequence diagram method as a basis for quantitative accident analysis. Report Risö-M-1374. Risö National Laboratory, Denmark (1971)
20. O'Hagan, A., Leonhard, T.: Bayes estimation subject to uncertainty about parameter constraints. Biometrika **63**, 201–202 (1976)
21. Owen, D.B.: Tables for computing bivariate normal probabilities. Ann. Math. Stat. **27**, 1075–1090 (1956)
22. Rohmer, J., Baudrit, C.: The use of the possibility theory to investigate the epistemic uncertainties within scenario-based earthquake risk assessments. Nat. Hazards **56**, 613–632 (2010)
23. Shapiro, A.F., Koissi, M.C.: Risk assessment applications of fuzzy logic. Casualty actuarial society, Canadian Institute of Actuaries, Society of Actuarie (2015)

Viable Systems Model in Triple-Agile Context

Marite Kirikova[1(✉)], Ligita Businska[1], Andrejs Dubrovskis[2], and Edgars Salna[2]

[1] Department of Artificial Intelligence and Systems Engineering, Riga Technical University, Riga, Latvia
{Marite.Kirikova, Ligita.Businska}@rtu.lv
[2] Datorzinibu Centrs, Riga, Latvia
{Andrejs.Dubrovskis, Edgars.Salna}@dzc.lv

Abstract. This paper reports on the results of the research regarding the Triple-Agile concept that has been introduced for supporting agile SMEs with cloud services agilely. The Triple-Agile concept implies three aspects of agility: agility of SME processes, agility of transition to cloud services, and agility of cloud service provision. The paper proposes using a viable system model, in general, and Viable Enterprise Bus design principles, in particular, to reflect services in small and medium sized enterprise ecosystems where both service providers and service consumers are SMEs. The use of a viable system model offers an opportunity to better understand relationships between services and thus may help to improve agility of service provision. The viable system model is applied to a triple-agile ecosystem from two viewpoints: the viewpoint of service and business relationship maturity and the viewpoint of service functions.

Keywords: Agile · SaaS · Service management · SME · Business processes
Viable system model · Fractal systems

1 Introduction

Software as a service (SaaS) has become a popular way in which the IT resources are used by large and small companies [1, 2]. When providing services to small and medium sized companies (SMEs), it is essential to take into account that these companies must be very flexible and fast in order to survive in a turbulent business environment. In other words, they have to be able to act agilely. To support the agility of SMEs, service providers have also to act agilely. To characterize such an ecosystem of companies that are acting agilely, a Triple-Agile concept has been introduced [3]. The Triple-Agile concept implies agile processes of SMEs, agile cooperation between SMEs and cloud providers, and agile cloud services provision.

In this paper, the research question under consideration is as follows: "Can the use of a viable systems model (VSM) help to improve service provision agility in a triple-agile service ecosystem?" The research mainly concerns such service provider-SMEs which have a restricted number of client-SMEs.

To answer the above stated research question, the following research approach was taken. First, the gaps between theoretical issues and practices of "agile" were analyzed.

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 158–171, 2018.
https://doi.org/10.1007/978-3-319-99951-7_11

Then, on the basis of this analysis, a hypothesis, that the Viable Enterprise Bus concept might be useful in agile service provision was applied and tested from two viewpoints. In this paper we will discuss lessons learned from that research.

The paper is structured as follows. The gaps between agile theory and practice are analyzed in Sect. 2. Here the concept of Viable Enterprise Bus is also discussed. The ecosystem of agile SMEs (service providers and service consumers) is reflected with the help of the Triple-Agile concept in Sect. 3. In Sect. 4, the service maturity and business relationship maturity view of the Viable Enterprise Bus concept application is presented and discussed. In Sect. 5, the VSM is applied from the function perspective of service provision. Conclusions and possible directions of further research are provided in Sect. 6.

2 Gaps Between Research and Practice of "Agile"

The word "agile" is used in many different contexts. In practice it is most often mentioned in the context of software development where it refers to an incremental and iterative way of software development. However, agility can be discussed more broadly and deeply as a theory looking from such perspectives as flexibility, speed, quality, knowledge management, etc. A comprehensive analysis of agility in information systems development is provided by Conboy in [4]. He summarizes the results of analysis in the following definition of "agile": "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". The Triple-Agile concept [3] used in this paper complies with this definition. However, it addresses a broader scope of issues looking beyond just the information systems development. This will be discussed in more detail in Sect. 3 when referring to the concept "agile ecosystem".

In this paper, when analyzing gaps between research and practice of "agile", we will refer to the works of Hobbs (2008) [5] and Hobbs and Scheepers (2009) [6] which show how a VSM [7] is a useful meta-level theory that captures the theoretical elements of information systems literature. These researchers use the VSM to identify the features of the IT functions that ensure their agility. From the practical side, the authors describe the agility, through the objective of COBIT best practices, to create an IT agility that is linked to the following management goals:

- Control over the IT process "Define Information Architecture" that provides the ability to respond to user requirements.
- Control over the process "Define IT processes, organization and linkages" that allows responsible reaction to a business strategy.
- Control over the process "Manage IT human resources".
- Control over the process "Acquire and maintain technology infrastructure".

The authors of [5, 6] point to a number of gaps between the theory and practice of "agile":

1. In practice, agility is defined more narrowly than in theory. In practice, it is more related to changes in requirements, while in theory it is more related to changes in the external environment.
2. Theory requires real-time identification and evaluation of future needs directly from the external environment, while practice does not emphasize such a necessity.
3. Theory demands the possibility of gradual development of the system according to the possibilities (initial investments) which are further developed depending on the situation (i.e. there are various "digital variants"), but in practice it is not considered.
4. Theory requires real-time identification and evaluation of the use of the information system, but practice does not emphasize this need.

In order to eliminate these gaps, the authors of [5, 6] recommend the use of a VSM, which pays special attention to issues of control, intelligence, and portfolio management. Figure 1 shows a simplified model that allows the key elements that are important in breaking the gap between the theory and practice of "agile" to be taken in account. The numbers shown in the image correspond to the gap numbers in the list of the gaps provided above.

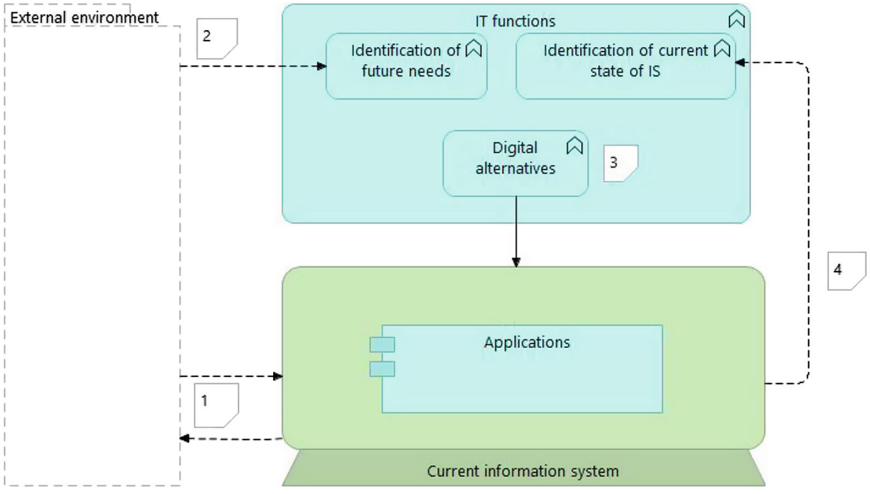


Fig. 1. Gaps between theory and practice of "agile."

The VSM, according to the authors of [5, 6], allows consideration of all four gaps between theory and practice that are shown in Fig. 1.

The VSM [7–9] has been used in various industries. Most applications are in business management [10, 11], business process management [12, 13], service systems [14], software development architectures [15–17], in the development of information

systems [18, 19] and other areas. Taking into account the fact that the VSM has been recognized as an appropriate meta-structure for connecting the theoretical aspects of the theory to the practice of "agile", and that the model has been successfully used both in business and in the context of information technology, we can conclude that the model might be useful in the Triple-Agile context that considers agility of customer processes, service provision and cooperation with customers, that includes and goes beyond the agility of just information systems development.

At the same time, it should be noted that the VSM is not a simple design. It includes 5 systems (operational – System 1, coordination – System 2, control – System 3, intelligence – System 4, and strategic management – System 5) as well as various information and management links between these systems and the external environment. The model also includes fractal decomposition (a simultaneous aggregation and specialization [20]) in the sense that each operating system itself can be expressed by a VSM at an unlimited number of decomposition levels. The relative complexity of the model might be considered as a challenge with respect to simplicity which is one of the features of agility.

An interesting work with respect to modeling cloud provision platforms has been developed by Jafarov [21]. The work suggests accommodating in the model 15 design principles based on cybernetic concepts of Beer's VSM [7]. The design principles are as follows:

- Variety – this principle basically refers to service interoperability.
- Viability, which refers to standardized integration of services for the sake of the interoperability.
- Value creation: the principle, which mainly refers to service reusability and standardization.
- Value preservation, that points to the interoperability between existing, newly uploaded or alternative services.
- Black box, which imposes service contracts to contain only the essential information about the services.
- Channels: this refers to the messaging system.
- Service recursion: the principle implies service abstraction, reusability, statelessness, discoverability, and composability.
- Service autonomy; the principle refers to high level of control for services over their environment.
- Service deviation – refers to good coordination of operations to avoid deviations in subsystems.
- Service bargain; the principle requires regulation of resource consumption.
- Service performance; the principle implies monitoring of service performance.
- Service audit: high variety and intra-operational checks on the services.
- Service intelligence: service awareness of their environment.
- Service policy – compliance with legislative and regulatory obligations.
- Service alert, the principle implies that the system must be endowed with a mechanism that would alert it in situations that require immediate actions.

The principles proposed in [21] are in line with findings in [13] and practically cover almost all issues of modeling of agile business process, cooperation processes,

and service development and management processes. These principles also correspond to the essential non-functional cloud service capabilities discussed in [22], namely: elasticity, reliability, quality of service, agility and adaptability, and availability. Therefore these principles and the frame of the model suggested in [22], namely Viable Enterprise Bus will be applied in Sect. 4 to illustrate the VSM application in a Triple-Agile ecosystem from the point of view of service maturity and business relationship maturity.

3 Triple-Agile Ecosystem

The concept of "agile ecosystem" is rarely used in literature; and, if used, it is usually related to development of information technology solutions; for instance, as in [23]. In this study, the triple-agile ecosystem is considered in conjunction with the Triple-Agile concept discussed in [3]. The concept concerns the following three types of agility, the expected synergistic effect of which is the agility of a whole ecosystem where service consuming and service providing SMEs are involved:

1. SME agility.
2. Agility of IT solutions. This agility type includes agile engineering (requirements-design-development), maintenance, and management of services.
3. Agility of cooperation between SMEs and the IT solution providers.

A simplified model of a triple-agile ecosystem is represented in Fig. 2 using enterprise architecture modeling language ArchiMate [24].

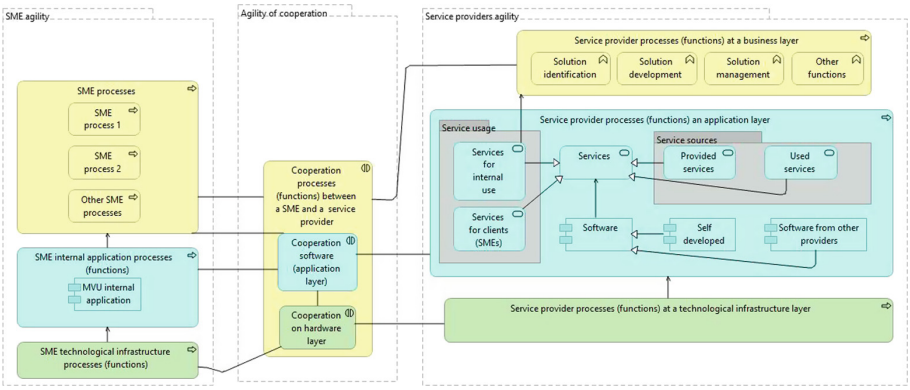


Fig. 2. Triple-agile ecosystem. (Color figure online)

Each business ecosystem's subsystem is shown in its block of agility and refers to a particular agility type. Each block has all three layers of business architecture: the business layer (depicted in yellow), the application layer (blue), and the layer of technological infrastructure (hardware, system software) – reflected in green. The collaborative

processes of the SME and IT solution (cloud services, e.g. SaaS) provider can take place at only the business layer, at the business and applications layers, and at all three layers simultaneously. Similarly, SME business processes can take place in only the business layer or in two or three layers at the same time. Service providers must identify and develop solutions appropriate for agile SMEs. For this purpose, a set of services can be used, which is provided by software applications. With regard to services, two service groups of service sources can be distinguished: (1) provided services and (2) used services (services provided by other service providers, e.g. outsourced testing environment services). On the other hand, with regard to the usage of services, the services can be divided into services used directly by SMEs (customer services) and services that are not directly used by SMEs (internal services).

While the agile ecosystem is reflected in Fig. 2 at a high level of abstraction and does not include many details, it still helps to set the context of application of a VSM that will be discussed in the next two sections.

4 Application of VSM: Maturity Viewpoint

The issues relevant for the Triple-Agile concept and triple-agile ecosystem are maturity levels of SaaS and maturity levels of cooperation between SMEs and SaaS providers [25]. There are several maturity models proposed ranging from four (e.g., Microsoft SaaS maturity model [26]) to six levels of maturity (e.g Forester SaaS maturity model [27]). Regardless of exact stages of maturity, the configurability is required at the second level of maturity, configurability and multi-tenancy are required at the third level of maturity; and configurability, multi-tenancy and scalability are required at the higher levels of maturity [25].

Similarly to SaaS maturity levels, the maturity levels of business relationship in SaaS are proposed in [28], namely:

1. Ad-hoc transaction implying static service delivery.
2. Repeatable transaction which requires service delivery with stable capability.
3. Configurable transaction which requires service delivery with configurable capability.
4. Long term relationship, where customers perform continuous evaluation of service quality and service providers ensure integrated delivery with customer extension.
5. Strategic partnership with governance of service (e.g. risk prevention) by customers and dynamic delivery with change management by service providers.

These levels well describe different cooperation options between SMEs and cloud service providers. It is important to see that a particular level of cooperation maturity does not imply the same level of SaaS maturity.

The levels of SaaS maturity and business relationship in SaaS maturity are taken into consideration when applying the idea of Viable Enterprise Bus (based on VSM) to the Triple-Agile context. The model implies that an SaaS provider can operate at different levels of maturity with respect to different types of services and different types of relationships with service consumers. Not all SMEs are eager to use high maturity level SaaS. This can be seen also from the business relationship in SaaS maturity model [28].

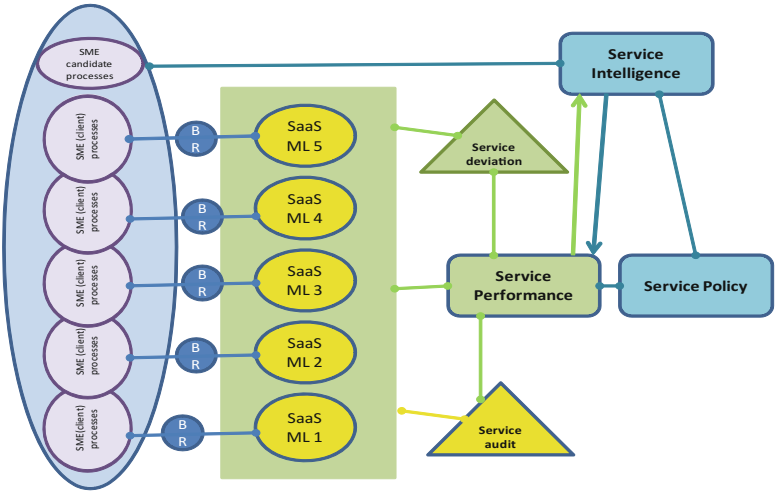


Fig. 3. The Viable Enterprise Bus idea applied to triple-agile ecosystem: BR stands for "business relationships in SaaS", ML stands for the "maturity level". The colors in this figure are not related to colors in figures presenting the models developed in ArchiMate graphical language.

The model that is based on a Viable Enterprise Bus idea, and represents an integrated view of application of VSM in triple-agile ecosystem from the point of view of service and business relationship maturity, is illustrated in Fig. 3.

The model presented in Fig. 3 accommodates all 15 design principles described in [21]. The difference from the model proposed in [21] is one of organizing the services according to their maturity level. The Forester maturity levels are used here, excluding the level 0 [27]. While, for the sake of simplicity, the model does not show any relationships between the service levels, some of them can be represented in the model. For instance, the user, i.e. SME, can require changing the SaaS maturity level for specific services from the third maturity level to the first one. While it may look like a simple change, in reality, from the service provider point of view, a service development project will be necessary to get the service from the third to the first level of maturity.

Each element of the model can be described at the lower level of detail. In this study we have used two forms of description: service dependency models and business process models in BPMN. It is essential to mention here that satisfactory richness of representation was achievable only by using both forms of representation together. In all modeling sessions the model types were proposed by researchers, and then the modeling process was done together with the practitioners. The service dependency models were easy to construct when detailed information about information exchange between services was not required, but not thereafter. The lower level representations of the elements of the model presented in Fig. 3 are not included in this paper. The challenges of representation of information flows will be discussed in the next section of the paper with respect to another viewpoint of representation.

Application of VSM gave an opportunity to clearly distinguish between maturity levels of services. Actually, it helped to structure the services by their maturity levels and served as a fractal registry of services. As mentioned above, this way of VSM application did not help to represent information flows between the services. It also did not help to consider issues of functional grouping of services and to analyze, in detail, functions of cooperation between service providers and service consumers that are essential in a triple-agile ecosystem.

5 Application of VSM: Function Viewpoint

In the previous section we used the form of representation of a Viable Enterprise Bus that is close to the one offered by Jafarov [21], which, in turn, is close to the original representation of the VSM [7]. The main difference was the introduction of business relationship elements between the external environment and the SaaS at different levels of maturity.

In this section we use ArchiMate architecture modeling language [24] as a means of VSM based model representation. In VSM, several links can be interpreted as actions. There are two main blocks of elements; namely, the one belonging to the external environment and one belonging to the service (see Fig. 3). Thus the business relationship functions are moved into the service block. A number of assumptions/requirements concerning the use of a VSM have been taken into account when applying the Viable Enterprise Bus concept to the triple-agile ecosystem from the functional perspective, namely:

1. Each IT tool/service provided to SMEs can be considered as a viable system.
2. In the context of enterprise architecture, actions and links of the original VSM model can be replaced by functions and corresponding information links in all five systems of VSM and in the external environment.
3. The structure of functions and their executable objects may vary in different layers of enterprise architecture for one and the same service.
4. With the purpose of closing the gaps between “agile” theory and practice (see Sect. 2) the minimum requirements for a service provider’s enterprise bus is the ability to ensure the consistent application of the theory and practice of “agile” (see Fig. 1).
5. Taking into the account the previous point, the context (external environment) of each VSM operating system (service) must be modeled so that the following types of external environment can be distinguished:
   a. an external environment to be served,
   b. an extended external environment (this could also be called “to be investigated external environment” – it includes the environment being served and the surrounding area),
   c. an internal collaboration environment (other operational systems (services, functions) at different fractal levels of the model).

When discussing the model, first, we will consider a template of an operation level services representation with VSM according to the design principles of Viable

Enterprise Bus. An operational system (service) representation template for an independent service is reflected in Fig. 4. The independent service template is shown here for simplification of description and visualization. The ArchiMate layering (business, application software, technology) is not reflected so as not to overcomplicate the representation. The case for the model of two functionally different types of services (e.g. one of them could be a document management service and another one could be an e-learning service) is shown in Fig. 5. It is essential that the model template for both services does not differ, i.e. it is possible to use one and the same template for handling all provided services.

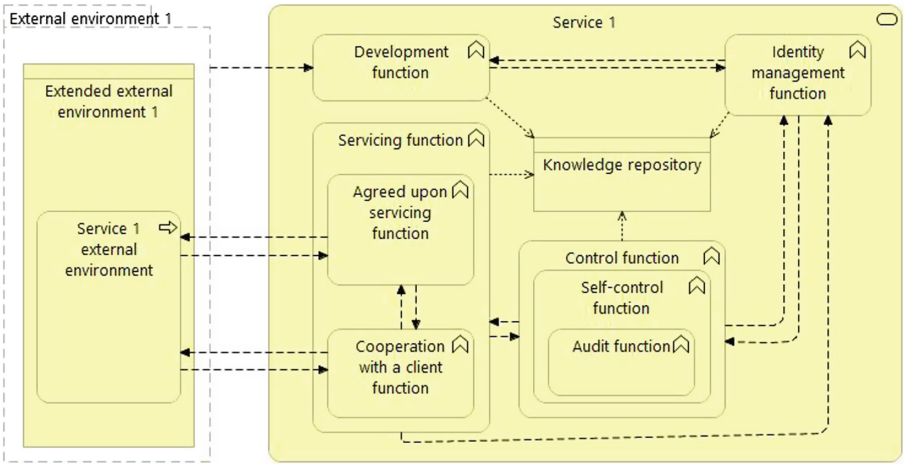


Fig. 4. Functions and information flows in an independent service (template).

The independent service template includes the external environment and the following functions:

- Servicing function that consists of two sub-functions (Agreed upon servicing function and Cooperation with a client servicing function) which can exchange information between themselves. The Servicing function corresponds to System 1 in the VSM.
- Control function that includes Self-control function, which can perform the Auditing function. The Control function corresponds to System 3 in the VSM.
- Development function that corresponds to System 4 in the VSM.
- Identity management function that corresponds to System 5 in the VSM.

The above listed functions fully correspond to the original VSM. However, two essential differences can be seen in the template. First, the template does not show the System 2 (coordination) at the given level of granularity of the representation. It could be reflected as a third sub-function of the Servicing function; however, its linkages to the other sub-functions of Servicing depend on the structure of sub-functions. Therefore the coordination function can be introduced on a lower level of a granularity of

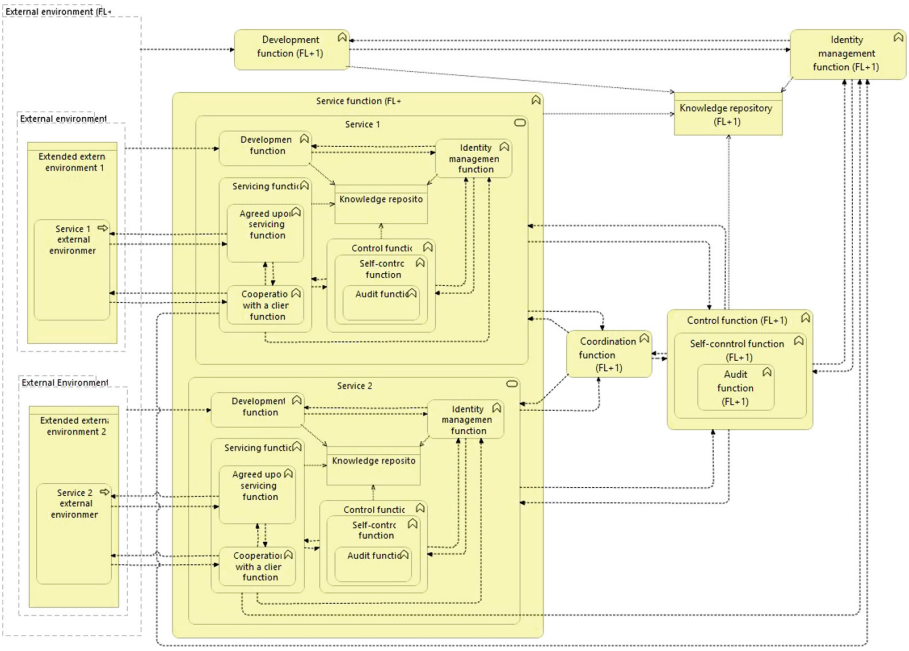


Fig. 5. Fractal architecture one level up (functions with (FL+1) refer to the upper fractal level).

Servicing function if it is necessary. On the other hand, coordination function is explicitly represented at the next fractal level, where more than one service is considered (see Fig. 5).

Second, the template includes a knowledge repository which amalgamates all types of knowledge needed for service provision and development. It can include digital variants of service provision, description of new technologies, documentation of existing services and other knowledge items. Such knowledge repository is not explicitly shown in original representations of the VSM and Viable Enterprise Bus. However, during the project, the necessity of such repository became obvious and a considerable effort was put in development of its structure and initial content. The VSM as such implies usage of organizational knowledge, but it is mainly utilized by the management rather than operational functions. In triple-agile ecosystem the knowledge is needed at an operational level as well. Therefore the Knowledge repository is linked to all other functions so that each function can use and update its contents.

It is important for each service to have the functions shown in the Fig. 4. As was mentioned above, these functions can be split between, or interact with, different business architecture layers (business, application software, and technology). The following information links are important for managing services to be provided to the

client processes (processes of SMEs located in the external environment of the service provider):

- A link from the Service external environment (SME business processes) to the service Servicing function to ensure that the service provision process is in accordance with the service level agreement.
- A link from the Servicing function to the service external environment (SME business processes) to ensure that the service usage process is in accordance with the service level agreement.
- A link from the Service external environment (SME business processes) to the Cooperation with a client function to ensure the agility of the service usage process.
- A link from the Cooperation with a client function to the Service external environment (SME business processes) to ensure the service provision agility.
- A link from the Extended external environment to the Development function that helps to identify potential new IT solutions. Development function here is to be understood at an enterprise level, not just as a software development function.
- A link from the Service function to the Control function, such as the data needed for the audit.
- A link from the Control function to the Service function, such as a feedback on the audit results.
- A link from the Development function to the Identity management function, such as an initiative to develop new services.
- A link from the Identity management function to the Development function, such as the decision to design or use a new IT solution (or technology).
- A link from the Control function to the Identity management function, such as information about the resources required for the service operation and its current performance quality.
- A link from the Identity management function to the Control function, such as switching to other technological resources or application software.
- An escalation link from the Service function to Identity management function when unexpected service activity is required (for example, the prevention of unforeseen security risks).

The functions and information links shown in Fig. 4 are sufficient for only one service. If another service operates at the same fractal level, there is an additional Coordination function between these services and the functions of the coordination, development and identity management of the next fractal level (FL+1), and related information links. The knowledge repository will also appear at the new fractal level, as it applies to both the individual services in the system and their possible synergistic combinations. The system of several services (in this case – two services) at the next fractal level is shown in Fig. 5.

In Fig. 5, for simplicity, only two services are represented at the lowest fractal level. In general, the number of these services is not limited and they are all tied to the Coordination function (FL+1). They are also related to the higher level (FL+1) Control function, Development function, and Identity management function that are represented in the same way (with the same type of information links) as for the template shown in the Fig. 4.

In terms of fractal levels, the model shown in Fig. 5 is itself a service, and it, together with other services, can form the next fractal level in the same way as the services shown in Fig. 5. In this case, there will also be a need for a new fractional knowledge repository. Thus, it is also important to note here that Coordination functions, Control functions, Development functions, Identity management functions, and Knowledge repositories, from the point of view of the system as a whole, can form a fractal structure.

In the triple-agile ecosystem, the use of the VSM helped to distinguish between different types of services from a functional point of view, relate them to the knowledge repository, and establish control over service change management. The planned support for SME processes was well mapped to the model represented in Fig. 4, however, only functionality wise. In experiments it appeared that most of the information links prescribed by the models reflected in Figs. 4 and 5 were challenging when they came to be documented and maintained. While having these links recorded was seen as a factor that can positively influence service delivery, maintenance, and development; no decisions have been made regarding the method for their permanent implementation. Still, it was clear that the recording should be human effortless and should be supported with multiple control mechanisms and interfaces.

6 Conclusions

In this paper the viable systems model, in general, and a Viable Enterprise Bus concept, in particular were applied in the context of triple-agile ecosystem composed of cloud service providers and service consumers-SMEs. The VSM was applied from two perspectives, namely, (1) from the software as a service maturity and service provider and consumer business relationship maturity level perspective; and (2) service function perspective. Both perspectives gave an opportunity to better structure the services and showed fractal relationships between them. Fractal structuring helped to transparently represent the services and thus contributed to service provision agility.

The paper contributes the following knowledge with respect to the use of VSM in the Triple-Agile context:

- It illustrated that the VSM is applicable in service modeling in the Triple-Agile context;
- It presents a fractal, VSM based, service provisioning model that helps to close four gaps between “agile” research and practice;
- It identifies a set of information links to be taken into account in agile service provision;
- It introduces a new element in the VSM based model, namely, a knowledge repository that actually opens new opportunities in knowledge management in fractal VSM based systems.

It is a matter for further research into how to ensure the recording and maintenance of the set of identified information links; and what are the best methods of knowledge management in the Triple-Agile context.

In this paper, the VSM based representations of the maturity and function perspectives were reflected in separate models using different modeling notations. Also, as mentioned in the introduction, the research was conducted in the SME ecosystem where the number of service consumers and consumed services was restricted. One more direction of further research, thus, is to merge both perspectives into a common multi-fractal model of service handling in the Triple-Agile ecosystem that is not restricted by the size of enterprises and the number of services.

Acknowledgment. The research has been supported in part by the funding from the research project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/16/A/007 signed between IT Competence Centre and Central Finance and Contracting Agency, project No. 1.13.

References

1. Schäfer, G., Schulze, M., Yusuf, Y., Musa, A.: The Benefits of SaaS-Based Enterprise Systems for SMEs - A Literature Review. In: Hernández, Jorge E., Liu, S., Delibašić, B., Zaraté, P., Dargam, F., Ribeiro, R. (eds.) EWG-DSS 2012. LNBIP, vol. 164, pp. 34–45. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41077-2_3
2. Benlian, A., Hess, T., Buxmann, P.: Drivers of SaaS-adoption – an empirical study of different application types. Bus. Inf. Syst. Eng. **1**(357), 357–369 (2009)
3. Kirikova, M., Salna, E.: Triple-agile: cloud solutions for SMEs. In: Johansson, B., Møller, C., Chaudhuri, A., Sudzina, F. (eds.) BIR 2017. LNBIP, vol. 295, pp. 260–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64930-6_19
4. Conboy, K.: Agility from first principles: reconstructing the concept of agility in information systems development. Inf. Syst. Res. **20**(3), 329–354 (2009)
5. Hobbs, G.: Forming a practice perspective for agility in information systems. In: Proceedings of ACIS 2008, pp. 391–400 (2008)
6. Hobbs, G., Scheepers, R.: Identifying capabilities for the IT function to create agility in information systems. In: Proceedings of PACIS 2009, pp. 1–12 (2009)
7. Beer, S.: Diagnosing the Systems for Organizations. Wiley, Chichester (1985)
8. Beer, S.: Platform for Change. Wiley, Chichester (1975)
9. Beer, S.: The Viable System Model: Its Provenance, Development, Methodology and Pathology. Cwarel Isaf Institute (1989)
10. Espejo, R., Reyes, A.: Organizational Systems. Managing Complexity with the Viable System Model. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19109-1
11. Rios, J.P.: Design and Diagnosis for Sustainable Organization. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-22318-1
12. Azadeh, A., Darivandi, K., Fathi, E.: Diagnosing, simulating and improving business process using cybernetic laws and viable systems model: the case of purchasing process. Syst. Res. Behav. Sci. **29**, 66–86 (2012)
13. Bider, I., Bellinger, G., Perjons, E.: Modeling an agile enterprise: reconciling systems and process thinking. In: Johannesson, P., Krogstie, J., Opdahl, Andreas L. (eds.) PoEM 2011. LNBIP, vol. 92, pp. 238–252. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24849-8_18
14. Golnam, A., Regev, G., Wegmann, A.: On viable service systems: developing modeling framework for analysis of viability in service systems. In: Snene, M., Ralyte, J., Morin, J.-H. (eds.) IESS 2011, LNBIP 82, pp. 30–41 (2011)

15. Herring, C., Kaplan, S.: Viable systems: the control paradigm for software architecture revisited, 0-7695-0631-3/00, 9 p. IEEE (2000)
16. Lamberth-Cocca, S., Meiren, T.: Towards a reference model for agile new service development using the example of e-mobility service systems. Proc. CIRP **64**, 259–264 (2017)
17. Stoyanov, E., Dalakakis, St., Roller, D., and Wischy, M.: Supporting the rapid product development by viable software architecture. In: Samuel, A., Lewis, W. (eds.) Engineering Design and Global Economy, 13 p. Institute of Engineers, Australia (2005)
18. Preece, G., Shaw, D., Hayashi, H.: Using the viable system model (VSM) to structure information processing complexity in disaster response. Eur. J. Oper. Res. **224**, 209–218 (2013)
19. Richter, J., Basten, D.: Applications of the viable systems model in is research – a comprehensive overview and analysis. In: Proceedings of HICSS 2014, pp. 4589–4598. IEEE (2014)
20. Kirikova, M.: Work systems paradigm and frames for fractal architecture of information systems. In: Nurcan, S., Pimenidis, E. (eds.) CAiSE Forum 2014. LNBIP, vol. 204, pp. 165–180. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19270-3_11
21. Jafarov, N.: Viable enterprise service bus model: a model for designing a viable service integration platform, thesis. http://www.unsworks.unsw.edu.au/primo_library/libweb/action/dlDisplay.do?vid=UNSWORKS&docId=unsworks_42290&fromSitemap=1&afterPDS=true
22. Carroll, N., Helfert, M., Lynn, T.: Towards the development of a cloud service capability assessment framework. In: Mahmood, Z. (ed.) Continued Rise of the Cloud. CCN, pp. 289–336. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6452-4_12
23. Medmilla, A.: Agile Kaizen. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54991-5
24. The ArchiMate® Enterprise Architecture Modeling Language. http://www.opengroup.org/subjectareas/enterprise/archimate-overview
25. Sorenson, P., Chen, X.: Towards SaaS (Software as a Service) Evaluation model. http://ssrg.cs.ualberta.ca/images/3/35/CASCON2008-Sorenson.pdf
26. SaaS Maturity Model by Microsoft (2006). http://www.theresearchpedia.com/research-articles/saas-maturity-model-by-microsoft
27. SaaS Maturity Model According to Forester (2008). https://blogs.msdn.microsoft.com/architectsrule/2008/08/18/saas-maturity-model-according-to-forrester/
28. Yang, Y.: A SaaS evaluation model based on competitive advantages. J. Residuals Sci. Technol. **13**(6), 40.1–40.4 (2016)

Enterprise Systems and Data Analytics

Merging Situation Analytics and Model-Based User Interface Development for Building Runtime-Adaptive Business Applications

Christian Märtin[1(✉)], Felix Kampfer[1], Christian Herdin[1,2], and Lionnelle Biawan Yameni[1]

[1] Faculty of Computer Science, Augsburg University of Applied Sciences, An der Hochschule 1, 86161 Augsburg, Germany
{Christian.Maertin, Felix.Kampfer, Christian.Herdin, LionnelleBiawan.Yameni}@hs-augsburg.de
[2] Institute of Computer Science, University of Rostock, Albert-Einstein-Straße 22, 18051 Rostock, Germany

Abstract. Situation analytics can be used to recognize the changing behavior, emotional state and environmental context of a user during complex task processing. This paper discusses the SitAdapt 2.0 architecture that combines a situation analytics platform with pattern- and model-based user interface construction tools in order to build runtime-adaptive interactive business applications with enhanced user experience and task-accomplishment characteristics. The paper focuses on the software components and tools for observing and tracking the user, modeling situations, recognizing situations, and modeling and generating structural changes and actions for generating the dynamic adaptations. The adaptive functionality of the system is demonstrated for a web-application for long-distance travel booking.

Keywords: Situation analytics · Adaptive user interfaces
Model-based user interface development environments (MBUID)
HCI-patterns · Emotion recognition · User tracking · E-business applications

1 Introduction

Situation- and context-aware computing has an established tradition in the area of distributed mobile computing [21] and complex interactive systems [8]. The availability of easy-to-use soft- and hardware tools for gathering visual or bio-physical user data and for observing user interactions and tracking activities has created the desire to systematically exploit the collected data for engineering dynamically adaptive software systems. Application domains range from sport and fitness apps, driver assistance systems to web shops and other business applications, e.g., support for digital marketing activities. The term situation analytics [3] was coined in order to define the requirements for a new branch of software engineering that deals with the construction of user-centered applications that are able to dynamically adapt to the situations of users, and to guide them to meeting their objectives and successfully finishing their tasks.

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 175–189, 2018.
https://doi.org/10.1007/978-3-319-99951-7_12

1.1 SitAdapt

The SitAdapt interpreter is a structured HCI-focused approach for building situation aware interactive software with a current focus on web applications. The SitAdapt interpreter is integrated into the PaMGIS model-based user interface development environment (MBUIDE) [4] and has access to its pattern- and model based tools and resources. It uses a situation analytics component for observing the interactions and emotions of users at work and for finding and generating the adaptive changes needed for supporting users during work sessions. The SitAdapt architecture provides a new and flexible approach for constructing high-quality interactive applications by recognizing and interpreting individual user situations in real-time. It enables rapid dynamic changes of the appearance and the behavior of interactive systems and raises the overall user experience level when working with such applications.

An observer component records and synchronizes eye-tracker-signals, recognized visual emotions, bio-physical wristband data, as well as user tracking and application meta-data. These data are interpreted by the situation analytics component. An adaptation component then decides whether a dynamic adaptation is necessary or not and controls the generation of an appropriate modification of the target software at runtime.

Using an MBUID environment [16, 17] for constructing interactive applications generally offers many advantages before, during, and after target system construction. For instance, models at different abstraction levels can even be accessed after the target system was implemented or generated. To enable real-time adaptations we use an enhanced version of the PaMGIS development environment that was developed in accordance with the CAMELEON reference framework (CRF) [2].

1.2 PaMGIS

PaMGIS [4, 5] offers pattern and model repositories that can be used for the development of interactive applications. The resources available in the repositories, e.g. HCI-patterns, user interface templates, or glue code for linking the user interface to business objects play an important role during the dynamic adaptation process.

The framework also supports automated generation of model refinements and final user interface code. To specify, organize and apply patterns and models, several software tools were integrated into the framework. An interface for improving the resource base with usability evaluation results is also provided by the framework. Another main contribution of PaMGIS is the PPSL (PaMGIS Pattern Specification Language) [6], an extended superset of all major HCI-pattern modeling languages that can also be used to define the resources for pattern-based code generation and to embed patterns into the overall modeling context.

1.3 Focus of This Paper

After we have demonstrated with a prototype, how the observation, decision-making, and adaptation components of a situation interpreter are collaborating [13], our recent research activities have architected the integration of the SitAdapt interpreter into the PaMGIS framework and specified the adaptation process [14, 15] in detail. The present

paper focuses on the operation of the components of the SitAdapt interpreter in its current version SitAdapt 2.0 and the interplay between SitAdapt and the PaMGIS models and pattern-resources. It also discusses the administrator tools for defining new situation rules and adaptations and demonstrates, how situations are recognized and adaptations are triggered by situation rules. It is also shown that Sit-Adapt can be used for adding adaptive behavior to applications that were developed without the PaMGIS framework.

The rest of this paper is structured as follows: Sect. 2 gives an overview of the field of context- and situation-aware computing and discusses the related work. Section 3 examines the SitAdapt 2.0 implementation and its components and tools. Section 4 discusses the concepts for modeling adaptability within the environment and the interplay between the situation analytics component, the target applications, and the pattern- and model-based resources of the PaMGIS framework. Section 5 demonstrates the system at work, e.g., with scenarios from a long-distance travel booking web-application. Section 6 concludes the paper and gives an outlook to the planned future work.

2 Related Work

The CRF [2, 17], a de-facto standard architecture for the model-driven construction of interactive systems, contains some structural guidelines for adapting the target software in pre-defined ways, mainly for responding to the requirements of different platforms for dynamically migrating applications from device to device.

However, in order to design interactive systems that are able to adapt dynamically to situational and contextual changes in a way completely tailored to the needs of the individual user, a new approach had to be engineered that has to support two major goals:

- *Improved task accomplishment and quality of work*. By observing the user on her way to reaching a goal or possibly failing to reach this goal, the system can compare the actual path taken by the user with the workflow proposed by the task model. If necessary, the system can offer help or dynamically restructure the user interface or the task workflow to support the successful completion of the task.
- *Better user experience*. By observing the current emotional and physical state of the user, the system can propose or generate modifications of the user interface or even the business layer in order to improve the individual user's sentiments and the overall user experience.

2.1 Context- and Situation-Awareness

Central to this new approach is the concept of situation-awareness that is strongly related with context-aware computing. The concept of context-aware computing was first proposed for distributed mobile computing in [20]. In addition to software and communication challenges to be solved when dynamically migrating an application to various devices and locations within a distributed environment, the definition of

context then also included environmental and social aspects (e.g. lighting and sound environment, persons in the room, roles of the persons, etc.).

Early definitions of the term situation-awareness appearing in psychology and the cognitive sciences were aimed at supporting human operators in complex situations, e.g., pilots during the landing phase, by defining situation-dependent requirements for allowing a smooth and correct task accomplishment [7, 8].

With a rising level of machine intelligence the borderline between a human operator, who easily recognizes a situation, because the software system delivers helpful visualizations and data, or a technical system that recognizes the situation and acts on its own behalf, is flexible. Situation-awareness has to cover both concepts.

Context, however, is a complex and not easily tangible concept. In [22] it is shown that context can either be seen in a reifying way by defining the semantics of every relevant object surrounding the entities engaged in interactions. Or, context can play a deeper role in interactions and must be considered, when users are forming meanings over time by also interpreting pragmatic contextual information contained in utterances of all types. For incorporating the latter view on context into automated recognition systems, deep knowledge of the domain and common-sense reasoning are required.

To capture the individual requirements of a situation, Chang [3] has suggested that a situation specification must cover the user's operational environment *E*, the user's social behavior *B,* by interpreting his or her actions, and a hidden context *M* that includes the users' mental states and emotions. A situation *Sit* at a given time *t* can thus be defined as $Sit = <M, B, E>_t$. A user's intention for using a specific software service for reaching a goal can then be formulated as temporal sequence $<Sit_1, Sit_2, ..., Sit_n>$, where Sit_1 is the situation that triggers the usage of a service and Sit_n is the goal-satisfying situation. For recognizing and evaluating situations adaptive runtime environments have to provide the cognitive and analytic capabilities to interpret the multitude of available signals and meta-data and to infer a goal-oriented set of adaptations.

SitAdapt was inspired by [3]. Our aim is to design a high-quality and practicable software engineering environment for building situation-aware target systems. Therefore, we follow a combined approach by linking the PaMGIS MBUID-framework and its domain and user interface models with a user-centric situation-aware adaptation component. For integrating the necessary reasoning capabilities, we introduce situation rules that have access to both, the observed raw situations and the modeling resources of the MBUID.

2.2 Observing the User

For implementing the emotion recognition functionality SitAdapt captures both visual and biometric data signals. In its current version user monitoring within several e-business scenarios (e.g. travel booking, online banking, finding and ordering beauty products in a web shop) is implemented in an advanced usability lab environment. The user is observed already before starting to interact with the application, during interaction, and until after the session is closed.

In [9] we discuss the interplay of the various recognition methods used in the SitAdapt system. Our work was influenced by several current research approaches for visual and bio-physical emotion recognition, e.g., [18, 19, 21]. We have also studied

the practical aspects of integrating runtime analytics, e.g. [12], and the consequences for sharing such information and privacy issues, e.g. [23].

An interesting related approach for tracking non-verbal visual utterances like smiling, nodding, and head shaking for assessing the users' social behavior is provided by the *MACH* social skills training environment [11].

A study for recognizing negative emotions with a high-level of accuracy by only exploiting mouse movement and cursor speed data is presented in [10]. Such methods can help to build more unobtrusive affective systems. As SitAdapt also records mouse and keyboard interaction data, a study for analyzing correlations between visually recognized emotions and mouse and keyboard behavior is currently set up at our lab.

The evaluation of observed user emotions while interacting with the target system and executing different tasks can lead to a precise recognition of users' intentions and their current mental states. To exploit this information, e.g., for individualized adaptations of commercial applications in order to change customer behavior does not only include UX related aspects, but also poses some ethical questions. In order to arrive at a better understanding of user reactions to automatically generated adaptations, e.g., when are they welcomed, when are they taken as manipulations, when aren't they recognized at all by the target applications' users, we will include these aspects into the mentioned lab study, both for a travel-booking application and a beauty products web-shop.

3 SitAdapt Implementation and Operation

SitAdapt 2.0 is an integrated software system for enabling situation-aware real-time adaptations for web and mobile applications. A recording component stores and synchronizes the various input signals as well as application meta-data in real-time (Fig. 1). These data are interpreted by the situation analytics component that is responsible for deciding whether a dynamic adaptation is necessary or not and controls the generation of an appropriate modification of the target software at runtime.

When we designed the architecture and discussed the functional requirements for SitAdapt 2.0 we had in mind that the field of affective computing is currently evolving rapidly both from the soft- and hardware point of view. We therefore opted for building a completely new observer and analytics platform that can easily integrate future recording devices and synchronize the interfaces of previously unrelated input systems. We decided to use a rule-based approach for interpreting the various input data and for triggering real-time adaptations, because of the easy way of accessing the collected situation analytics data and the resources of the coupled PaMGIS system. Future implementations could include a learning component that improves the quality of adaptations by altering and optimizing situation rules based on the assessment of the observed user reactions and level of task accomplishment.

In order to profit from earlier results in the field of MBUID systems, we planned to integrate the SitAdapt interpreter into the PaMGIS (Pattern-Based Modeling and Generation of Interactive Systems) development framework from the beginning since we created our first prototype [13]. With the integration now fully operational this defines at least three different use cases for the resulting development environment:

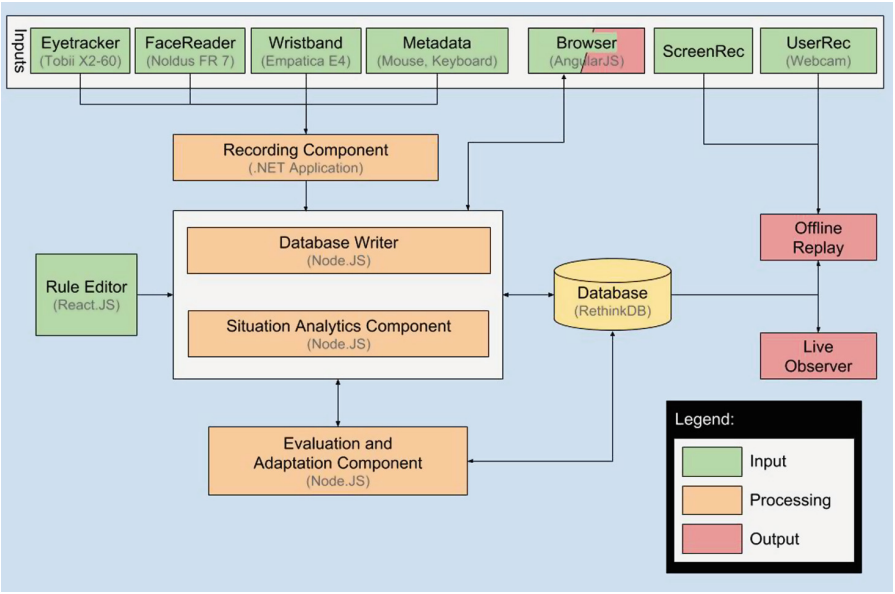


Fig. 1. Structure and compoments of the SitAdapt Interpreter in its current implementation SitAdapt 2.0.

1. SitAdapt 2.0 can be used as a situation-rule-based stand-alone system for adapting all sorts of web applications at runtime. In this case the situation rules trigger adaptations via a direct interface to the application. The application has to provide the necessary resources or settings for modifying the user interface or the accessed content.
2. With the SitAdapt interpreter integrated into the PaMGIS framework, adaptations due to context and situational changes can be modeled at design time. The user interface or even the accessed business objects will be adapted at runtime, if a pre-modeled situation or context change arises. The resources for adaptation are kept in the pattern and model repositories of the PaMGIS system.
3. The integrated framework can also be used for enhancing web applications with support for modeling and adaptation, e.g., user interface and task models can be added to an existing app without a priori models. The new models can link an app to model and code fragments that are used for implementing possible adaptations. When a predefined situation occurs, situation rules trigger the generation of the adaptations typically in one or more additional windows or by generating separate interaction objects or actions.

The SitAdapt 2.0 system consists of the following parts:

- The *data interfaces* from the different devices (eye-tracker, wristband, facial expression recognition, metadata from the application). Our prototypical adaptive applications were built with Angular JS and are executed in the browser. The communication between the application and the PaMGIS models, if available, is controlled by the situation analytics and the adaptation components.

- The *recording component* that synchronizes the different input records with a timestamp. The component also records the eye- and gaze-tracking signal of the user and observes his or her emotional video facial expression with the Noldus Face-Reader software as a combination of the six basic emotions (happy, sad, scared, disgusted, surprised, and angry). Other recorded data about the user are, e.g., age-range and gender. The stress-level and other bio-physical data are recorded by reading the data from a wristband or other signal-emitting devices in real-time. In addition, mouse movements and keyboard logs are protocolled.

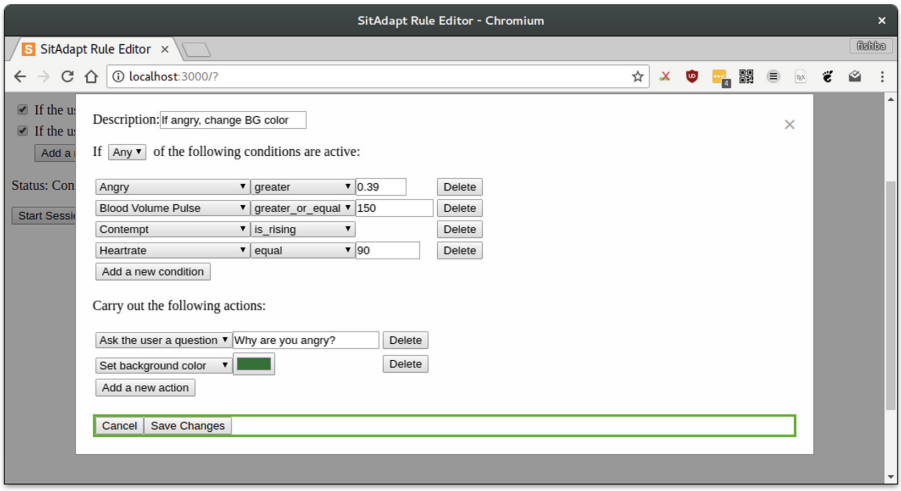


Fig. 2. SitAdapt 2.0 rule editor

- The *database writer* stores the data from the recording component and from the browser in the database, where the raw situations and situation profiles are managed, and also controls communication with the rule editor.
- The *rule editor* allows the definition and modification of situation rules, e.g. for specifying the different user states (e.g., if an angry state is observed, it will only become relevant, if the state lasts more than five seconds and the grade of the emotion surpasses a certain activation level) and the resulting actions. The rule editor is very flexible and can use all input types, values and temporal changes for formulating rule conditions. Figure 2 shows the creation of a simple situation rule with four conditions and two simple actions. Figure 3 shows, how existing situation rules are activated at design time. At runtime they are fired by the situation analytics component for adapting the user interface, if the conditions of one or more rules apply. In this case none of the PaMGIS models is affected and only a dialog with the user or a modified presentation attribute is created. However, situation rules can also activate HCI-patterns in the pattern repository. These patterns come with different levels of abstractions. The may contain concrete templates for generating interaction objects or for modifying low-level user interface attributes. A more abstract HCI-pattern can be exploited by the PaMGIS framework at runtime to generate user

interface adaptations from predefined UI-, task-, or domain-model fragments directly referenced by the HCI-pattern.

- The *situation analytics component* analyzes and assesses situations by exploiting the observed data. Situation rules are fired by the situation analytics component when the rule conditions are satisfied. Situation rules interact with the situation profiles stored in the SitAdapt database. The rule actions either directly trigger simple adaptations or interact with the PaMGIS resources as described above.
- The *evaluation and decision component* uses the data that are provided by the situation analytics component to decide whether an adaptation of the user interface is currently meaningful and necessary. For this purpose the component evaluates one or more applicable situation rules and has to solve possible conflicts between the rules. Whether an adaptation is meaningful depends on the predefined purpose of the situation-aware target application. Goals to meet can range from successful marketing activities in e-business, e.g. having the user buying an item from the e-shop or letting her or him browse through the latest special offers, to improved user experience levels, or to meeting user desires defined by the hidden mental states of the user. Such goals can be detected, if one or more situations in the situation profile trigger an application dependent or domain independent situation rule. Situation rules are related to patterns. They define behavioral and context-related situational patterns. If the decision component decides that a complex adaptation is necessary, it has to provide the artifacts from the PaMGIS pattern and model repositories to allow for the modification of the target application by the adaptation component. The situation rules provide the links and control information for accessing and composing HCI-patterns and model fragments, necessary for constructing the modifications.

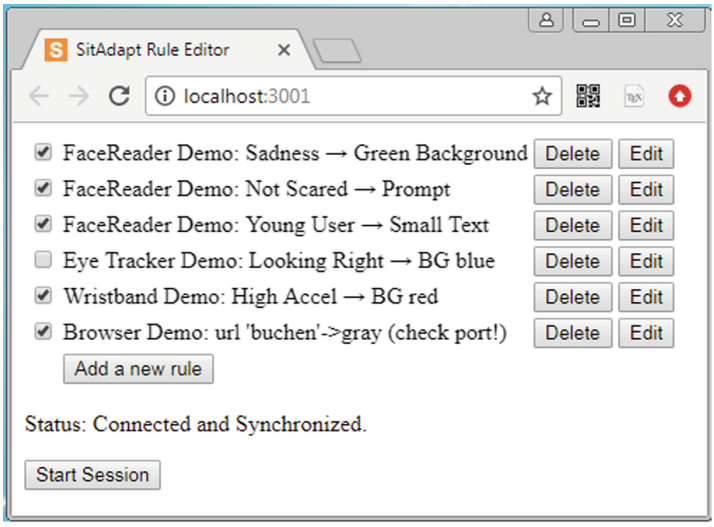


Fig. 3. Preselected situation rules that are active at application runtime

- The *adaptation component* finally generates the necessary modifications of the interactive target application.

These architectural components provided by the SitAdapt Interpreter are necessary for enabling the PaMGIS framework to support automated adaptive user interfaces.

4 Modeling Adaptations

The SitAdapt interpreter is integrated into the PaMGIS-framework (Fig. 4), a pattern- and model-based MBUID environment that uses the ontological domain- and context models as proposed by the CAMELEON reference framework (CRF) [2].

The SitAdapt interpreter has full access to all sub-models of PaMGIS and directly interacts with the user interface model. It supports (semi-)automated dynamic adaptation before and during runtime [1, 24]. Within CRF-conforming systems the abstract user interface model (AUI) is generated from the information contained in the domain

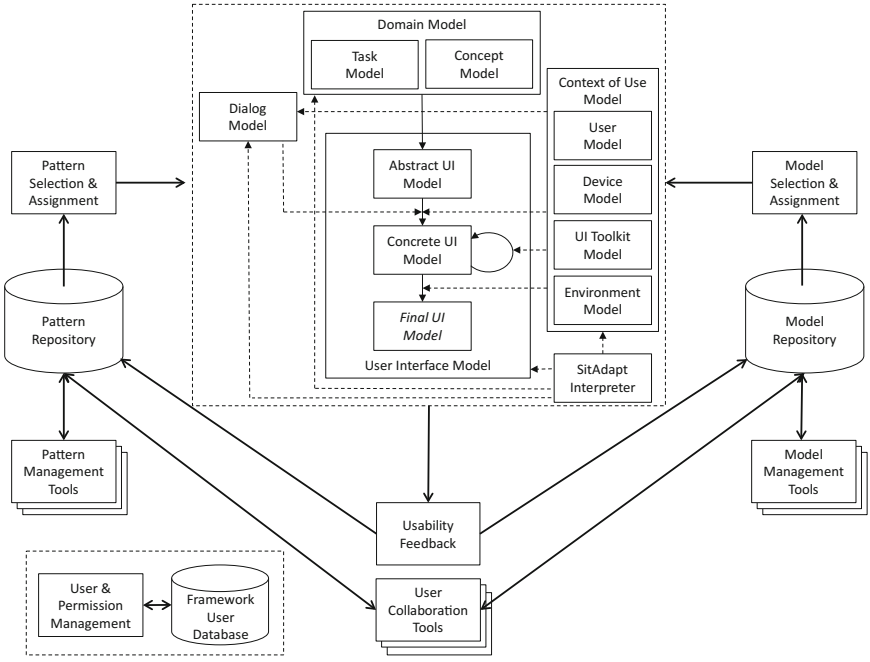


Fig. 4. PaMGIS MBUID framework with SitAdapt Interpreter and communication flow for CRF-model interaction [14]. The situation analytics component resides in the SitAdapt Interpreter. Adaptations triggered by the situation analytics component may involve the pattern and model resources of the PaMGIS framework.

model of the application that includes a task model and a concept model and defines abstract user interface objects that are still independent of the context of use.

The AUI model can then be transformed into a concrete user interface model (CUI) that already exploits the context model and the dialog model, which is responsible for the dynamic user interface behavior. The next step generates the final user interface model (FUI) by parsing the CUI model.

If an adaptation requires predictable structural changes of the user interface or different task workflows, an adaptation variant can be modeled by providing alternative models for all affected modeling levels. The adaptation process is discussed in more detail in [14].

For the situation-aware adaptation process, the task and context of use models are the most relevant of the PaMGIS sub-models. The *user model* holds both static information about the current user, and dynamic data describing the emotional state as well as the biometric state. It is structured as follows [14]:

```
<UserCharacteristics>
        <UserIdentData>
        <UserAbilities>
                <USRUA_Visual>
                <USRUA_Acoustic>
                <USRUA_Motor>
                <USRUA_Mental>
        <UserExperiences>
                <USRUE_Domain>
                <USRUE_Handling>
        <UserDistraction>
        <UserLegalCapacity>
        <UserEmotionalState>
        <UserPsychologicalState>
        <UserBiometricState>
                <USRBS_Pulse>
                <USRBS_Attention>
                <USRBS_PulseChangeRate>
                …
```

Not all of these attributes need to be filled with concrete data values. The dynamic values concerning emotional state and biometric state are taken from the *situation profile* that resides in the SitAdapt database, whenever an adaptation decision has to be made. Note, that the situation profile that is generated by the SitAdapt Interpreter contains the whole sequence of emotional and biometric user states from session start until session termination. The temporal granularity of the recorded raw situations is variable. An example of a situation profile with a granularity of one second for a ticket booking application is the following:

```
<SituationProfile>
<TargetApplication>
<User>
...
<Situation_booking>
       <SituationTime> 29 s
       ...
       <Eye_Tracking> eye_t(29 s)
       ...
</Situation_booking>
<Situation_booking>
       <SituationTime> 30 s
       <FUI_link> Wizard_Part_1
       <AUI_link> model_AUI_1
       <CUI_link> model_CUI_1
       <Dialog_link> order_ticket
       <Task_link> order_ticket
       <Concept_link> concept_1
       <Eye_Tracking> eye_t(30 s)
       <EmotionalState> neutral
       <UserPsychologicalState> tired
       <BiometricState>
              <Pulse> normal
              <StressLevel> orange
</Situation_booking>
...
</SituationProfile>
```

Along with the recorded data, the situations are also linked with the PaMGIS models of the currently used interactive components. The links can be modified whenever an adaptation occurs. Via the FUI_link information about the currently used interaction object as well as the mouse and keyboard input data is immediately available.

Before starting a target application, a *UI configuration file* is generated that contains a *ContextOfUse* tag field for each modeled task. It has sub-tags that may serve as context variables holding information relevant for controlling the UI configuration and later, at runtime, the adaptation process.

For instance, one of the sub-tags may contain information that a task “ticket sale” is only authorized for users from age 16. When the situation analytics component at runtime discovers that the current user is less than 16 years old, a hint is given in the final user interface (FUI) model that she or he is not authorized to buy a ticket.

5 SitAdapt at Work

Designed alongside SitAdapt 2.0, we developed a prototypical travel-booking web application to highlight and evaluate some of the system's capabilities. The application features elements typical for e-commerce applications, including the ability to enter query parameters, viewing and selecting query results, viewing product details, registering as a new user, logging in, modifying a selected product, and viewing a summary of all entered data before making a purchase.

Having the ability to collect information about a user's physical, physiological and emotional properties allows application designers to offer products and product enrichment opportunities ("extras") that center on the user's immediate and/or long-term needs, benefitting producers and consumers. In the field of air travel, for instance, knowledge of a user's demographic and his or her current disposition can influence the kinds of seat and airline upgrade opportunities offered to the user.

In our first example (Fig. 5), the system can recognize fear, anger, or a higher pulse rate, while a user is in the process of choosing a flight. In response, it offers an effective remedy to combat the customer's fear of flying and/or other negative flight-associated emotions.

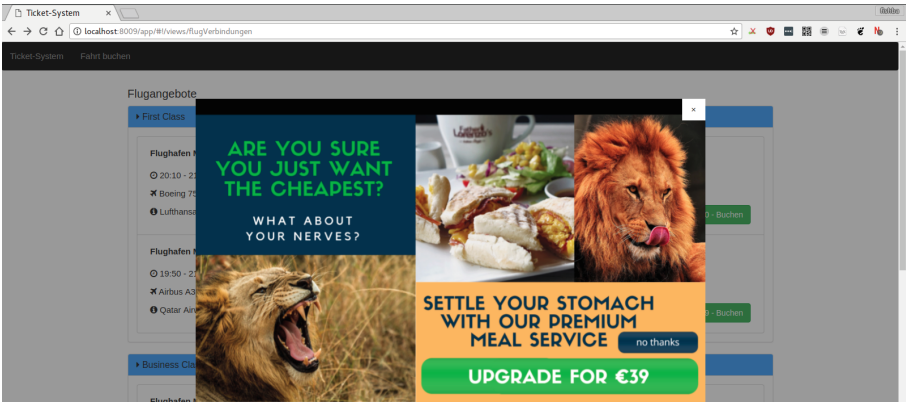


Fig. 5. Offering an additional meal service

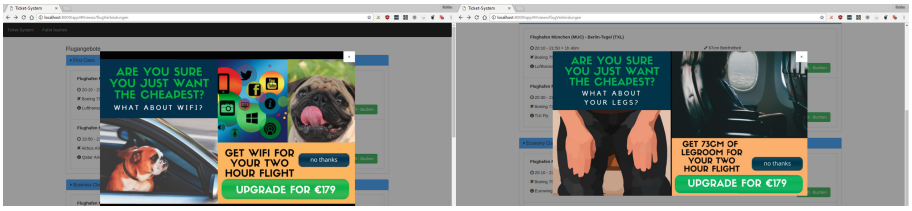


Fig. 6. (a) Offering Wi-Fi upgrade for young passengers. (b) Offering more legroom for senior citizens.

In the second example (Fig. 6a and b), the system uses a demographic-focused approach to determine positive flight-related desires and presents an option to fulfill these desires to the user. A user between the ages of 15 and 40 may be presented with an option to purchase Wi-Fi access, while for users with an age over 50 a seat upgrade opportunity may be presented to increase their level of comfort throughout the flight.

Additional sensor information can be used to make predictions about the likelihood of users taking advantage of these product enrichment opportunities. Eye tracking and browser-based mouse-tracking data can offer information about customers' financial flexibility as they consider or reconsider their budgetary restrictions when looking at the different opportunities. Such data can be exploited by specific situation rules.

The SitAdapt interpreter's capabilities used by the prototypical applications are not limited to targeted advertising. Application developers can also use our system to change color schemes to appeal to the user's current mood, change layouts to suit demographic-based technological prowess, change styles to suit possible accessibility needs, record specific diagnostic information and offer help (or ask for specific feedback) when a user becomes frustrated with a certain feature.

In a final example we specify a situation rule that can be used in a web shop environment, e.g. for beauty products, where SitAdapt recognizes the user's interest in a certain product. After three minutes a text is displayed that notifies the user that in case of the purchase of product (Id) within a certain time-span, a voucher is granted for the user's next purchase. A link to the voucher-processing task in the task model is activated:

```
<SituationRule> OfferingVoucher
          FOR N <Situation1> IN 180s
          <Eye_Tracking> Field Product Product(Id)
          <Gaze_Tracking> Contains Field Product
                 Product(Id) (>5)
          <Pulse> (85-100)
          <PulseRate> rising
          <Emotion> excited
          <StressLevel> orange
          <Action> SHOW AT 180s VoucherText1FUI
          <Action> WAIT VoucherText1FUIInput
          <Action> LINK VoucherText1FUIInput
                 TaskModel VoucherProcessingTask
```

6 Conclusion

In this paper we have presented the SitAdapt 2.0 system that supports situation-aware dynamic adaptations of the user interface at design and runtime. The system is currently being evaluated for several business applications in a use-lab environment.

The evaluation process includes user studies for both the travel-booking application and the beauty products portal and will assess the quality of the real-time adaptations during a session, their impact on task accomplishment, on usability, as well as on the

way users perceive and rate such adaptations with respect to user experience, privacy, and possible manipulative effects.

We use the results for different purposes:

- Defining and testing situational patterns and possible user interface adaptations
- Testing ideas for digital marketing activities in the later stages of the customer journey
- Identifying domain-dependent and domain-independent user personas
- Analyzing recorded user and meta-data to mine correlations between visual, bio-physical and mouse - and keyboard tracking data in order to support anonymous or at least less intrusive support for emotion-based adaptations in the future.

We expect a rapid evolution of visual and sensor-based emotion recognition technology and its hard- and software integration into all device types from desktop computers to smart phones and other personal devices. If the recorded data can be securely kept privately on the devices, the benefits generated from individualized adaptations will be experienced by the users without compromising their privacy. Better user experience and task accomplishment can thus also drive the success of business models built around situation-aware software technology.

References

1. Akiki, P.A., et al.: Integrating adaptive user interface capabilities in enterprise applications. In: Proceedings of the 36th International Conference on Software Engineering (ICSE 2014), pp. 712–723. ACM (2014)
2. Calvary, G., Coutaz, J., Bouillon, L., et al.: The CAMELEON Reference Framework (2002). http://giove.isti.cnr.it/projects/cameleon/pdf/CAMELEON%20D1.1RefFramework.pdf. Accessed 25 Aug 2016
3. Chang, C.K.: Situation analytics: a foundation for a new software engineering paradigm. IEEE Comput. **49**, 24–33 (2016)
4. Engel, J., Märtin, C., Forbrig, P.: A concerted model-driven and pattern-based framework for developing user interfaces of interactive ubiquitous applications. In: Proceedings of First International Workshop on Large-Scale and Model-based Interactive Systems, Duisburg, pp. 35–41 (2015)
5. Engel, J., Märtin, C., Forbrig, P.: Practical aspects of pattern-supported model-driven user interface generation. In: Kurosu, M. (ed.) HCI 2017. LNCS, vol. 10271, pp. 397–414. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58071-5_30
6. Engel, J., Märtin, C., Forbrig, P.: A unified pattern specification formalism to support user interface generation. In: Kurosu, M. (ed.) HCI 2016. LNCS, vol. 9731, pp. 445–456. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39510-4_41
7. Flach, J.M.: Situation awareness: the emperor's new clothes. In: Mouloua, M., Parasuaman, R. (eds.): Human Performance in Automated Systems: Current Research and Trends, pp. 241–248. Erlbaum (1994)
8. Flach, J.M., Mulder, M, Van Paassen, M.M.: The concept of the situation in psychology. In: Banbury, S., Tremblay, S. (eds) A Cognitive Approach to Situation Awareness: Theory and Applications, pp. 42–60. Ashgate Publisching, Oxon (2004)

9. Herdin, C., Märtin, C., Forbrig, P.: SitAdapt: an architecture for situation-aware runtime adaptation of interactive systems. In: Kurosu, M. (ed.) HCI 2017. LNCS, vol. 10271, pp. 447–455. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58071-5_33
10. Hibbeln, M., Jenkins, J.L., Schneider, C., Valacich, J.S., Weinmann, M.: How is your user feeling? Inferring emotion through human-computer interaction devices. MIS Q. **41**(1), 1–21 (2017)
11. Hoque, M., Picard, R.W.: Rich nonverbal sensing technology for automated social skills training. Computer **47**, 28–35 (2014)
12. Lee, Y., Balan, R.K.: The case for human-centric personal analytics. In: Proceedings of WPA 2014, pp. 25–29. ACM (2014)
13. Märtin, C., Rashid, S., Herdin, C.: Designing responsive interactive applications by emotion-tracking and pattern-based dynamic user interface adaptation. In: Kurosu, M. (ed.) HCI 2016. LNCS, vol. 9733, pp. 28–36. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39513-5_3
14. Märtin, C., Herdin, C., Engel, J.: Model-based user-interface adaptation by exploiting situations, emotions and software patterns. In: Proceedings of CHIRA 2017, Funchal, Madeira, 31 October–2 November. SCITEPRESS (2017)
15. Märtin, C., Herdin, C.: Enabling decision-making for situation-aware adaptations of interactive systems. In: Proceedings of 10th Forum Media Technology, FMT 2017, 29–30 November, St. Pölten, Austria (2017)
16. Meixner, G., Calvary, G., Coutaz, J.: Introduction to model-based user interfaces. W3C Working Group Note, 07 January 2014. http://www.w3.org/TR/mbui-intero/. Accessed 27 May 2015
17. Melchior, J., Vanderdonckt, J., Van Roy, P.: A model-based approach for distributed user interfaces. In: Proceedings of EICS 2011, pp. 11–20. ACM (2011)
18. Picard, R.: Recognizing stress, engagement, and positive emotion. In: Proceedings of IUI 2015, 29 March–1 April 2015, Atlanta, GA, USA, pp. 3–4 (2015)
19. Qu, F., Wang, S.-J., Yan, W.-J., Fu, X.: CAS(ME)^2: a database of spontaneous macro-expressions and micro-expressions. In: Kurosu, M. (ed.) HCI 2016. LNCS, vol. 9733, pp. 48–59. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39513-5_5
20. Schilit, B.N., Theimer, M.M.: Disseminating active map information to mobile hosts. IEEE Netw. **8**(5), 22–32 (1994)
21. Schmidt, A.: Biosignals in human-computer interaction. Interactions **2016**, 76–79 (2016)
22. Vieweg, S., Hodges, A.: Rethinking context: leveraging human and machine computation in disaster response. Computer **47**, 22–27 (2014)
23. Warshaw, J. et al.: Can an Algorithm Know the "Real You"? Understanding people's reactions to hyper-personal analytics systems. In: Proceedings of CHI 2015, pp. 797–806. ACM (2015)
24. Yigitbas, E., Sauer, S., Engels, G.: A model-based framework for multi-adaptive migratory user interfaces. In: Kurosu, M. (ed.) HCI 2015. LNCS, vol. 9170, pp. 563–572. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20916-6_52

Agile Software Engineering Practices and ERP Implementation with Focus on SAP Activate Methodology

Adnan Kraljić(✉) and Tarik Kraljić

Ghent University, Ghent, Belgium
{adnan.kraljic,tarik.kraljic}@ugent.be

Abstract. The Enterprise Resource Planning (ERP) implementation is a complex and active process, one that involves a mixture of technological and organizational interactions. Often it is the largest IT project that an organization has ever launched and requires a mutual fit of system and organization. Concept of an ERP implementation supporting business processes across different departments in organization is not a generic, rigid and uniform process - it is a vivid one and depends on number of different factors. As a result, the issues addressing the ERP implementation process have been one of the major concerns in industry. Therefore, ERP implementation process receives profound attention from practitioners and scholars in academic or industry papers. However, research on ERP systems so far has been mainly focused on diffusion, use and impact issues. Less attention has been given to the methods/methodologies used during the configuration and the implementation of ERP systems; even though they are commonly used in practice, they still remain largely undocumented in Information Systems research domain. This paper aims to provide insight from practice (SAP ERP implementation team up with 20 SAP consultants including authors of this paper) regarding the agile engineering practices in ERP implementation process. One of stubbornly persists belief was that ERP systems cannot be part of agile development due to their complexity and nature. However, it is becoming obvious that agile engineering practices will not be anymore exclusively linked to software development as SAP (biggest world ERP vendor) recently introduced its first agile ERP implementation methodology named SAP Activate Methodology.

Keywords: ERP · ERP implementation · Agile implementation methodology Sprints · Phases

1 Introduction

Implementing an ERP system is a major project demanding a significant level of resources, commitment and adjustments throughout the organization.

Problem Description. Often the ERP implementation project is the single biggest project that an organization has ever launched. As a result, the issues surrounding the implementation process have been one of the major concerns in industry. And it further worsens because of numerous failed cases including a few “fatal” disasters which lead

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 190–201, 2018.
https://doi.org/10.1007/978-3-319-99951-7_13

to the demise of some companies. In previous studies can be found that almost 70% of ERP implementations fail to achieve their estimated benefits [1]. Although ERP can provide many benefits for organization, goals are often changed to getting the system operational instead of realizing the goals. Reflecting such a level of importance, the largest number of articles in literature belongs to this theme. It comprises more than 40% of the entire articles. Less attention has been given to the methods/methodologies used during the configuration and the implementation of ERP systems; even though they are commonly used in practice, they still remain largely unexplored and undocumented in Information Systems research domain. However, practically relevant research that addresses industry sectors that have to apply more than just agile (e.g., due to the development of safety critical systems or legal regulations) is rare. Back in 2003, Boehm and Turner [1] described a first approach how to combine agile and traditional software development for defining a balanced software development strategy, and Diebold and Zehler described how agile can be integrated into rich processes. Yet, as also argued in, systematic construction procedures are missing as most available research documents ad-hoc and user specific approaches to construct organization- and project-specific development approaches.

Objective. The overall goal of the research presented in this paper is to provide an insight from industry regarding the ERP implementation methodologies trend (focusing on world's biggest ERP provider SAP). The research presented aims to show the key components of SAP ERP implementation methodologies focusing on differences between agile and waterfall elements in each of them (and their evolution).

Contribution. In this paper, we present the ongoing research on agile software engineering practices applied on ERP implementation methodologies in practice.

2 ERP Implementation Methodologies in Literature

Several models of ERP implementation methodologies are provided in literature and they vary according to e.g. the number of phases. The phases in ERP implementation frameworks are often counted as between three and six, according to Somers and Nelson. However, the model of includes 11 phases and it gives practical checklist-type guidance for an ERP implementation. On the other hand, the models of Markus and Tanis or Parr and Shanks are very general, and are merely used for analyzing ERP implementation projects [2]. The models are useful in studying, analyzing and planning ERP implementation. The selection of ERP implementation method mentioned in paper is based on the degree of "institutionalization" in the scientific community. Livari and Hirschheim described six criteria to determine institutionalization: including (1) the existence of scientific journals, (2) scientific conferences, (3) textbooks, (4) professional associations, (5) informational and formal communication networks, and (6) citations. There are number of different ERP implementation methodologies mentioned and described in literature. However, there is an issue with methodology scope, context and its ambiguity. For example, some methodologies treat the phases before the acquisition of an ERP system (and are focused on it), while some methodologies put stress on phases after the ERP system has started to be used (production phase) [3].

Next table summarize list of proposed implementation methodologies followed by the degree of institutionalization in scientific community [4] (Table 1).

Table 1. ERP implementation methodologies in literature

Author(s)	ERP implementation model
Bancroft et al. (1998)	(1) Focus, (2) Creating As – Is picture, (3) Creating of the To-Be design, (4) Construction and testing and (5) Actual Implementation
Kuruppuarachchi et al. (2000)	(1) Initiation, (2) Requirement definition, (3) Acquisition/development, (4) Implementation, and (5) Termination
Markus and Tanis (2000)	(1) Project chartering, (2) The project, (3) Shakedown, and (4) Onward and upward
Makipaa (2003)	(1) Initiative, (2) Evaluation, (3) Selection, (4) Modification, Business process Reengineering, and Conversion of Data, (5) Training, (6) Go – Live, (7) Termination, and (8) Exploitation and Development
Parr and Shanks (2000a)	(1) Planning, (2) Project: a. setup, b. reengineer, c. design, d. configuration and testing, e. installation (3) Enhancement
Ross (1999)	(1) Design, (2) Implementation, (3) Stabilization, (4) Continues improvement and (5) Transformation
Shields (2001)	Rapid implementation model of three phases and 12 major activates
Umble et al. (2003)	(1) Review the pre-implementation process to date, (2) Install and test any new hardware, (3) Install the software and perform the computer room pilot, (4) Attend system training, (5) Train on the conference room pilot, (6) Established security and necessary permissions, (7) Ensure that all data bridges are sufficiently robust and the data are sufficiently accurate, (8) Document policies and procedures, (9) Bring the entire organization on – line, either in a total cutover or in a phased approach, (10) Celebrate, and (11) Improve continually
Verviell and Halingten	(1) Planning, (2) Information search, (3) Selection, (4) Evaluations, and (5) Negotiation

It is evident that there is no ground based ERP implementation methodology - widely accepted and tested. Even though they are commonly used in practice (ERP implementation methodologies) they still remain largely unexplored and undocumented in Information Systems research domain. Additionally, academic literature does not provide any relevant material regarding the influence of agile software engineering practices on ERP implementation methodology. In next paragraph we will describe newly introduced SAP ERP implementation methodology (SAP Activate Methodology).

3 SAP Activate Methodology (Agile Based Methodology)

In business informatics, software project methodologies define implementation and development guidelines for “out-of-box” software implementation as SAP. Historically, SAP had its own project methodology called ASAP, which stands for

Accelerated SAP (and was completely based on waterfall project management principles). Few years ago, SAP introduced SAP Launch project methodology for its Cloud product portfolio which was also based on waterfall principles. Launch methodology has been transformed into SAP Activate Implementation methodology (currently in use for SAP ERP products) and presents the first agile based ERP implementation methodology introduced by SAP. The SAP Activate methodology comprises of six phases as highlighted in Fig. 1, which is a disciplined approach to managing complex projects, organizational change management, solution management, & industry specific implementations [5]. In next paragraph, we will briefly describe SAP Activate ERP implementation phases.

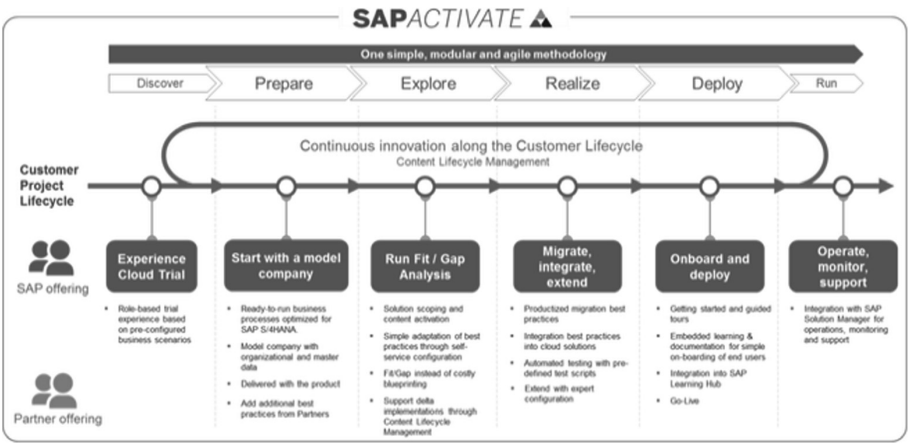


Fig. 1. SAP activate methodology

Prepare - This phase encompasses the entire project preparation and planning activities with infrastructure, hardware/network sizing requirements completed. It involves setting up the infrastructure, team, project goals, charter, and agree upon schedule, budget, risk baseline, proof-of-concept planning if applicable with implementation sequence. The project manager on the ground will discuss with the customer project manager to identify risks early on with a mitigation plan. The PM will be responsible for drafting a high-level project plan with all milestones with a detailed task level plan chalked out with critical dependencies. Each phase deliverable should be agreed between both parties. Finally, a project organization, steering committee is organized with assigned resources [6].

Explore - This is the most crucial phase of the project for a project manager as he just about to steer the ship, like a captain. The objective of this phase is to be on a common platform on how the company plans to run SAP for their business operations. Thus, a PM is responsible for analyzing the project goals and objectives and revise the overall project schedule if required. In simple terms, it is the critical requirements gathering phase, A PM might use appropriate tools to collect requirements with required traceability. The

result is the Business Blueprint, which is a detailed flow of business process AS-IS, how they run the business operations with a TO-BE mapped in SAP, on how these business operations will run in SAP. Depending on the implementation complexities, number of business process, Blueprint workshops might span for a few days or weeks or even months, in a complex environment. The output of this phase is the baseline configuration in SAP with detailed custom code requirements analysis done [7].

Realize - In simple terms, realization is the actual development phase of the project, where you'd configure, develop custom code and conduct required testing. It involves coding-unit testing-integration testing-User acceptance testing (UAT). As per the business blueprint and mapping the SAP system as agreed with business, all the business process requirements will be implemented. In reality, there are two major work packages: (a) Baseline (major scope); and (b) Final configuration (remaining scope). The success of any implementation project relies on how closely you're able to develop custom code, test and release it to the UAT phase, in order to support adequate testing by the users. Also, the challenge is to adopt changes as indicated during the UAT. This phase is resource intensive and the team is at peak team size to ensure all deliverables are met and sign-off. Often times Integration fail due to lack of test data, and testing in a "PRD" like environment to be able to test all critical business scenarios. A good practice is to copy a "PRD"-like environment and start testing if the system already exists. If it is GreenField environment, ensure adequate test data is available to test it rigorously [8].

Deploy - Final preparations before cutover to production ensure that that the system, users, and data are ready for transition to productive use. The transition to operations includes setting up and launching support, then handing off operations to the organization managing the environment.

In next few paragraphs we will provide industry report from major medical institution in Bosnia and Herzegovina.

4 Report from the Industry

This industry report provides insight into an implementation of SAP ERP solution in a major medical institution in Bosnia and Herzegovina, with several thousands of employees. The project was initiated by management of the hospital with purpose to eliminate the ineffectiveness of the current information system. Analysis of the current financial system and the list of new system requirements have been prepared by external consulting company. This was prerequisite for announcing a public tender for selection of ERP software solution integrator. After several months of tender procedure and assessment of the best vendor (price was eliminatory criteria, in accordance with the law), software integrator was selected. At the end, SAP All-in One solution was preferred one. In next few paragraphs are briefly provided some quick facts regarding the project. As recommended by external consultants the main tasks of the project were:

- To centralize the information system;
- To increase data integrity and consistency;

- To focus on accounting and financial department processes;
- To improve drug warehouse management and billing system;
- To provide comprehensive and accurate reports for top management.

The project incorporated five SAP modules: FI (Finance), CO (Controlling), MM (Material management), SD (Sales and Distribution) and HR (Human resources). SAP integrator offered a team of 20 SAP solution consultants, including one SAP system administrator. Additionally, two consultants (ABAP programmers) were teamed up for specific ABAP developments. The implemented system was standard SAP ERP All-in-One. Since it was not specialized health care solution, additional industry-specific functionalities needed to be developed to fulfill basic needs. They were mostly related to the processes of Materials Management and Sales and Distribution.

4.1 Remarks from Participations

Agile SAP implementation methodology named SAP Activate was presented to team members at the kick off meeting. Some of the remarks presented in next paragraphs are also mentioned in industry whitepaper named Whitepaper "Scrum and ERP – do they go together" from Mr. Boris Dloger. Some of the most repeated remarks from participants were:

- "What is the intention for using agile practices for standard ERP implementations? Typically ERP implementation does not involve any kind of development, just customizing of software to the customer's needs."
- "Agile terminology is ambiguous for us – it is terminology of software development process, not ERP implementation process!"
- "Team structure is ambiguous and not possible in ERP implementation process – lack of the rigid hierarchy will create chaos"
- "In order to customize ERP and adjust the diverse and complex ERP components we need dedicated experts. Work in cross functional teams is complex!"

We will describe those remarks in detail in next chapters providing the experience of team members.

4.2 Agile Terminology in SAP ERP Implementation

Getting the terminology right is often a complex task. In order to have a better insight in this topic author used industry experience and knowledge of Mr. Anton Karnaukhov, SAP expert, available on his blog. Agile project management brings a number of new terms that are critical to understand in order to be able to comprehend how agile can help you implement SAP solutions in a better way. Projects, especially large and complex, typically should have both waterfall and agile characteristics and elements to them. A combination of elements from both approaches usually works best for any given project, regardless of context, complexity, size or any other factor. The key roles that should be integrated in an organization that mixes traditional waterfall and agile practices in are simplified in Table 2 [9].

Table 2. Traditional waterfall vs. Agile practices

Waterfall role	Meet half way	Agile role
Sponsor	Same	Sponsor
Product manager	Product manager would typically be accountable for the whole product or product line, and oversee potentially many product owners, or act as a product owner herself	Product owner
Project manager	Solution architect would be elevated to oversee programs and dependencies of projects, leaving teams to manage more autonomously	Solution architect
Project methodology "champion"	Same idea here, if either practices are applied, care should be taken to apply with true understanding of why they are applied to not dilute the intent and effect	Scrum master (Scrum being by far the most widely adopted agile framework, hence this role used here)
End users	Active involvement from the beginning to the end. Traditional stage gates with wider presentation, key roles from core stakeholder groups involved continually	End users
Other stakeholders	Address and respect the product owner	Other stakeholders

Furthermore, those are "items" of terminology used regarding agile implementation process, shown in Table 3 [10].

4.3 Team Structure of Project

In the past, different variations of project team composition was suggested as standardized, but the fact is that there is no canned project team setup that will work on every project [9].

At the very beginning of any SAP implementation project there is a process of defining a baseline of features that make up the general scope of a particular product. A release is typically associated with a go-live - for example going live with Material Management (MM) and Finance (FI) modules of SAP ERP, or a particular business unit going live with an SAP Extended Warehouse Management (EWM) system. One important thing to note is that the initial list of features in a release is not "frozen for changes". Most of the time, if not all of the time, new features will be added to the list as the project progresses and some of the old features will be either moved to a later release or even removed entirely from the project. This is perfectly normal and even encouraged. In fact, part of embracing change is constantly re-evaluating the prioritization of project features and user stories and focusing on the ones that bring the most business value.

Table 3. Agile terminology

Agile terminology	Description (analogy to SAP ERP implementation terminology)
Release	Each release is associated with some type of go-live where a number of features are moved to production. For example, a "big bang" SAP implementation project could have just a single large release. A more phased approach could lead to many releases within a single SAP implementation project - SAP ERP Finance and Human Resources in one release and Logistics modules taking place in a separate release shortly after
Feature	These are sometimes called Epic Stories or Epics. Features represent large sets of functionality, for example - sales order processing, warehouse management picking, month-end close, etc.
User story	Each user story describes a particular business requirement and is assigned to a single feature. In many ways a user story is the next level down in terms of detail after a feature
Sprint	A sprint is typically a pre-determined period of time (2 weeks, 4 weeks, 6 weeks, etc.) within which a set of identified user stories needs to complete
Kanban board	A visual board where columns represent a state that a user story can be in, for example - planned, blueprinting, realization, testing, done. Stories are arranged on the Kanban board and moved from one column to another as progress is being made. Many teams build physical Kanban boards by using tape and post-it notes
Retrospective	A retrospective is a focused session where your team looks back at how the current agile approach is working and which areas can be improved. Many agile teams conduct retrospectives at set intervals of time (every 8 weeks or at the end of every sprint)

A good starting point for building a list of features for an SAP implementation project is SAP's Best Practices library. Once a baseline or industry-specific package is chosen customer is able to see its requirements presented with a list of SAP best practice scenarios, each designated by a unique number. For example, under the baseline package you will be able to find scenario 112 called "Sales Quotation". The end result is a list of features that outlines the first pass at defining the overall scope of what we think we are going to accomplish in a particular project [11].

Once the initial feature list is put together we can start building our project team structure. Here are some general guidelines that we try to follow.

4.4 Leadership

In the context of an agile SAP implementations these roles are much less about management and are much more about leadership, which is often a source of tension in many organizations that are trying to do agile for the first time [9]. What is meant by this is that many traditional managers are used to working in an environment where they create layers of superfluous entities (status reports, WBS structures, functional and technical specification documents, detailed tracking of estimated vs actual time) all typically under the banner of "proper management" and for purposes of historical

reference, and then depend on "data" from those entities to reveal red flags in order to take some "corrective action". With agile the use of such activities is greatly discouraged and many managers who are not successful leaders often struggle to find their niche within an agile project.

4.5 Self-directed Teams

In general, each team is responsible for managing all of the work (features, user stories, tasks, etc.) from start to finish. They are responsible for gathering user stories and tasks (blueprinting), prioritize the stories; performing the configuration and development work (realization), testing, providing documentation, and even of training end-users. All teams are expected to create their own user stories (and tasks) and keep their Kanban boards up-to-date. Teams are expected to engage in continuous dialog with business users and prioritize their own workload [9]. One of the most crucial techniques for promoting self-directed teams is the use of daily stand-ups. These are short and focused meetings, typically at the beginning of each day, where each team member answers the following 3 questions:

- What have I accomplished yesterday?
- What am I planning to accomplish today?
- Is there anything in my way that is preventing me from making progress?

Product Owner

Four product owners where assigned to project team structure. They are important part in providing help with prioritization of user stories, but given the volume of stories in an average SAP implementation (from 1000 to 5000), their involvement in prioritization at the individual story level is limited [12]. We can summarize product owner's primary objectives as:

- Identification and prioritization of features (in tandem with project manager aka Scrum Master)
- Review and acceptance of delivered solutions through a live demo (usually a batch of 15–30 user stories)

Solution Architect

One solution architects was assigned to project team structure whose role was to ensure that the overall SAP implementation across sub-teams is synchronized. Solution architect is professional with extensive technical experience across many SAP systems and modules. And even though cross-functional sub-teams often rely on solution architects to get direction in complex integration scenarios, the teams themselves still retain the responsibility of aiming all of their user stories and tasks forward [13]. Solution architects are typically responsible for:

- Cohesiveness and robustness of the overall delivered solution for the entire project
- Integration design and testing across teams, systems and modules

Cross – Functional Teams

This is often a source of confusion in SAP project as the term cross-functional has historically referred to cross-module (for example SD, FI, MM, etc.) in the SAP world.

However, in the context of an agile team structure for an SAP project cross-functional refers to the various functions performed by the project team - requirement gathering, configuration, development, documentation, training, etc. [12]. Typically a number of such cross-functional sub-teams is build, each focused on a small number of SAP systems/modules. For example:

- Sales and Distribution team - 4 BPOs, 3 analysts, 1 developer, 1 trainer/instructional designer
- Materials Management team - 5 BPOs, 2 analyst, 1 developer, 1 trainer/instructional designer

4.6 Impact on Project Managers

In Agile engineering practices, a traditional project manager role may not be required to manage an implementation, as the agile teams are self-sufficient (common Project Management roles in an agile project are product owners, scrum masters and scrum team).

- Scrum master coaches the development team to use scrum principles.
- A product owner will generally come from the customer side and will own the requirements and will be part of the scrum team.
- Product owner will be responsible for documentation of requirements which are normally written as user stories. He will also be in charge of the prioritization of the requirements [10].

4.7 Impact on Project Stakeholders and Sponsors

SAP Activate Methodology ensures consistent involvement of project stakeholders and sponsors. Project stakeholders are involved in project planning and retrospective meetings every 2–4 weeks. In a traditional/Waterfall project management scenarios, the client gets involved at a much later stage which results in a mismatch of expectations and project delivery [12].

5 Conclusion

It is evident that agile practices will not be exclusively linked to software development anymore. Some of the remarks that arise from team members were stated in previous chapter. We will restate them again, but this time, providing an insight as a result of "hands on" experience in agile SAP ERP implementation project.

> *"Why should agile methods be used for standard ERP implementations? This does not normally involve any kind of development, just adapting software to the existing processes."*

ERP implementations intervene in the way many employees work, whether standard or in-house development: changes give rise to uncertainty. In agile engineering practices, great scope is given to communication with the user by means of regular interviews. The users test the product increments and give feedback as to what works well and what does not.

"We don't understand agile terminology – it is terminology of software development process, not ERP implementation process!"

Getting the terminology right is very important part adopting agile engineering practices truly [14]. Agile project management brings with it a number of new terms that are critical to understand in order to be able to comprehend any of the detailed discussions on how agile can help you implement SAP solutions in a better way. Listed in this paper, in the form of table, we showed that it is possible to map agile and waterfall terminology "one – to – one"

"Team structure is ambiguous and not possible in ERP implementation process – lack of the rigid hierarchy will create chaos"

In general, each team is charged with owning and managing all of their logical units of work (features, user stories, tasks, etc.) from start to finish. That means they are responsible for gathering user stories and tasks (blueprinting), working with the business to prioritize those stories, performing the configuration and development work (realization), testing, documentation, demoing solutions and even training end-users [6].

"To be able to configure and adapt the diverse and complex ERP components we need specialized experts. That makes work in cross-functional teams more difficult."

It has also proved effective in ERP projects to unite different skills in one team: ERP consultants work with ERP developers, CRM experts with MM experts, while business analysts or system architects also enhance such teams. In this way the requirements are viewed from different perspectives and the exchange of ideas within the team brings aspects to light that the individual alone would not have detected – this again ensures that the "right" product is delivered [12].

In upcoming years we will see more and more research papers and case studies about the influence of agile practices on ERP implementation process. It is expected that all major ERP providers present theirs unique agile driven ERP implementation methodologies [14]. In upcoming papers (and as a part of PhD project) we will focus on providing hybrid Agile Waterfall ERP implementation methodology designed on design science postulates.

References

1. Abrahamsson, P., Salo, O., Ronkainen, J., Warsta, J.: Agile Software Development (2002)
2. Aiken, L.S., West, S.G.: Multiple Regression: Testing and Interpreting Interactions. Sage, Newbury Park (1992)
3. Akgün, A.E., Byrne, J., Keskin, H.: Knowledge networks in new product development projects: a transactive memory perspective. Inf. Manag. **42**, 1105–1120 (2005)
4. Akgün, A.E., Keskin, H., Byrne, J., Imamoglu, S.Z.: Antecedents and consequences of team potency in software development projects. Inf. Manag. **44**, 646–656 (2007)
5. Al-Fatish, F., Roemer, M., Fassunge, M., Reinstorf, T., Staader, J.: ASE: Immer busser mit starken Teams! Agile Software Engineering bei der SAP. Object Spectr. **1** (2011)
6. Anderson, D.J.: Agile Management for Software Engineering: Applying the Theory of Constraints for Business Results. The Coad Series. Prentice Hall, Upper Saddle River (2004)
7. Andreessen, M.: Why software is eating the world. Wall Street J. (2011)

8. Argote, L.: Input uncertainty and organizational coordination in hospital emergency units. Adm. Sci. Q. **27**, 420–434 (1982)
9. http://www.xpertminds.net/blog/111
10. Baccarini, D.: The concept of project complexity: a review. Int. J. Project Manag. **14**, 201–204 (1996)
11. Balijepally, V.G., Mahapatra, R.K., Nerur, S., Price, K.H.: Are two heads better than one for software development? The productivity paradox of pair programming. Manag. Inf. Syst. Q. **33**, 91–118 (2009)
12. https://go.forrester.com/consulting/
13. Jalali, S., Wohlin, C.: Global software engineering and agile practices: a systematic review. J. Softw.: Evol. Process **24**, 643–659 (2012)
14. Glogler, B.: Scrum and ERP: do they go together? Industry Whitepaper (2014). https://borisgloger.com/wpcontent/uploads/2014/07/Whitepaper_Scrum_and_ERP_en.pdf

Interviews Aided with Machine Learning

Robin Ambrosius[1(✉)], Morgan Ericsson[2], Welf Löwe[2], and Anna Wingkvist[2]

[1] Dezember IT GmbH, Edenkoben, Germany
ambrosius@robin.dezember.co
[2] Department of Computer Science, Linnaeus University, Växjö, Sweden
{morgan.ericsson,welf.lowe,anna.wingkvist}@lnu.se

Abstract. We have designed and implemented a Computer Aided Personal Interview (CAPI) system that learns from expert interviews and can support less experienced interviewers by for example suggesting questions to ask or skip. We were particularly interested to streamline the due diligence process when estimating the value for software startups. For our design we evaluated some machine learning algorithms and their trade-offs, and in a small case study we evaluates their implementation and performance. We find that while there is room for improvement, the system can learn and recommend questions. The CAPI system can in principle be applied to any domain in which long interview sessions should be shortened without sacrificing the quality of the assessment.

Keywords: Due diligence · Artificial intelligence · Interview software

1 Introduction

The corporate world has changed and there are more startups than ever before. Startups often depend on funding from investors or business angels, especially when they want to expand their business beyond the "minimal viable product". As the number of startup increases, so does the need for funding, so startup investments have increased in recent years [1]. An essential part of investing in a startup is estimating the value in a so-called due diligence process. The process is particularly challenging if the startup deals with software and software services, or information, since these have limited physical value and are difficult to appraise. For example, how to appraise the value of software that is not ready to be launched and thus have no customers and does not generate any revenue?

The company DueDive[1] has developed a due diligence process for software startups that assesses the capabilities of their IT organization. All needed information is known or can be estimated. The facts of the evaluation are retrieved in manual interviews, and the findings from those are used to determine the value. An interviewer from DueDive and key people of the assessed company take part in the interview. The interview is a semi-structured conversation, and a question

[1] https://duedive.com.

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 202–216, 2018.
https://doi.org/10.1007/978-3-319-99951-7_14

can be included or skipped based on interviewer's experience. The goal is to gain the highest amount of information with the least amount of questions. However, since the conversation is the basis for the value estimation, it is vital that the right questions are asked so that the required information can be combined and assessed carefully. A lot of experience is needed for carrying out such an analysis and arriving at a realistic estimate.

Give the importance of the interviews and the required expertise to ask the rights questions, only experts in due diligence of software startups can perform the interviews and come to realistic estimates. This makes it hard to scale since it takes time to find or train experts. Currently, DueDive can perform approximately one interview per week. In complicated cases, with multiple interviews and multiple interviewers, it can also be challenging to find all the correlations between answers on the spot and be able to ask the right questions in an interview. So, there is a high potential for improvement.

We suggest a Computer Aided Personal Interview (CAPI) system that uses machine learning to help interviewers ask the right questions, no matter their experience level. The system should help control the flow of the interview by suggesting which question to ask next to maximize the information gain and which questions can be skipped because the answer is always known or because the question is not applicable in this case. It should also determine whether there are any additional questions, not in the script, that should be asked based on provided answers. Expert interviewers do this, and the machine learning algorithms should learn from conducted sessions and provide support for less experienced interviewers when they conduct an interview.

There exist multiple systems that use natural language processing and machine learning to interact with humans e.g., chat bots and automated support systems, but we were not able to find any system that focuses on computer aided interviews. Some systems focus on fully automated interviews, but these are often focused on speech and text recognition and generation, and not on the flow of questions.

There exist also systems for simplifying interviews in specific application domains. For instance, Wall et al. suggest shortening a questionnaire that diagnoses autism using machine learning [2]. The question count was reduced from 93 to 7 while keeping the accuracy of the test high. This proves that machine learning has the potential to shorten interviews that cover a complex topic that usually involved a large questionnaire. We generalize this and other approaches and reduce the number of significant questions regardless of the application domain.

Therefore, design science research guidelines presented by Hevner et al. [3] will provide the basic methodology for our CAPI system. We outline the required components of the system and evaluate potential machine learning algorithms for the flow continuation, flow shortening, and the prediction of additional questions. Based on this design, we implement and evaluate the system in a small case study. Our main contributions are the design of such a system, with the algorithms and an evaluation of their performance and trade-offs, as well as the prototype

that was implemented in collaboration with DueDive. The prototype was well received, and they think it can potentially increase the number of interviews they can perform in a week.

2 Background

Some algorithms related to machine learning will be evaluated and used to design and implement the CAPI system. We provide a brief introduction to the essential algorithms used in or considered for our system below. We also provide a brief discussion of natural language processing and the approaches that will be used in the design and implementation.

Decision trees are tree structures that enable a decision by differentiating between various attributes [4]. In this context, an attribute is an object e.g., an interview question or an answer. Attributes can be of different data types e.g., integers or Boolean variables. Decision trees are classified as supervised learning and are a clear and easy to understand classification technique.

Markov chains are stochastic models that try to predict a future state based on the current state [5]. Markov chains are often used in reinforcement learning and resemble state machines, with the difference that each state has a probability of transferring into another state spontaneously. In first-order Markov chains, the transfer probabilities only depend on the current state, while in higher order chains they also depend on previous states [6]. Hence, in a second-order chain, the transfer probability depends on the current state and the state that the machine was in before the current one. Some advanced models are based on Markov chains, such as Markov chains Monte Carlo [7] that use previous trails to predict the future state and the autoregressive model [8] that also bases the output on previous states.

Imputation is a supervised learning approach that recreates missing values based on existing values. It is often used in statistics, as many algorithms do not work with incomplete data [9]. However, the result needs to be assessed with care, as the imputed values might have skewed the data set. The quality of the imputation depends on the used method. Standard methods are Mean substitutions or k-Nearest-Neighbors (KNN).

The mean substitution is one of the simplest and fastest ways to impute a value. It verifies all values of the attribute that are not missing and uses the mean of those to predict the missing value. Other statistical methods, such as the median or the mode can also be used. This simple approach is attractive if the attributes do not correlate.

For use cases where correlations between attributes exist, valuable information can be lost. KNN instead evaluates the current state of an entry that is missing value and compares this state to other entries with values. These other entries are the nearest neighbors. The value is then imputed from the k-nearest entries using, for example the mean.

Basket analysis is a supervised learning algorithm that can be used to, e.g., find items that are commonly bought together in shops. The approach analyzes baskets that were checkout and group items together that occur regularly. The generated rules indicate which item leads to another one. The rules can often be improved by considering the support, lift, and confidence [10]:

- Support is the relative frequency of a rule in a data set and indicates how often a rule occurs. A high support value indicates a useful rule.
- Lift indicates how independent two events are. A value around 1 indicates an independent rule.
- Confidence measures how reliable a rule is, i.e., how often the consequence occurs when the cause is given.

Basket analysis does not support ordinal values, the values in a matrix need to be binary [10]. *Cross tabulation* allows ordinal values but at a high runtime complexity [11]. Though, a large sample size is necessary to get reliable calculations, and the cross tabulation fails if data points are missing.

Natural Language Processing. (NLP) includes a number of different approaches e.g., speech recognition, text understanding, and text generation [12]. We focus on understanding written text, mainly on finding similar texts.

The first possible method to find similar texts is a so-called sequence matcher. This approach analyzes a text by its tokens and calculates the similarity of the texts by comparing the distance of the tokens. Tokens are often n-Grams, sequences of n items from a given element set. These element sets can be numbers, letters or any other finite set of elements, but for natural language processing, they are often words. n-grams are often used to predict the next word in a sequence e.g., in a sentence. First, a set of n-grams is extracted from an existing set of sentences. These n-grams are then compared to an incomplete word to find possible missing endings of the word.

Another approach that is often used in recommender systems is the analysis of *Term Frequency-Inverse Document Frequency* (TF-IDF) [13]. It analyses how common a word is compared to all other words in the text, which can be used to find similar texts.

Another commonly approach used in this context is *entity extraction*. Instead of analysing common words this method searches for entities that are mentioned in the text [14]. These entities can be grouped into different categories and later be compared to find similar texts by checking which ones involve the same entities.

3 Design

We begin by describing the interview process and extract entities and components from the description. Since the learning component is the most complex part, it requires the most design. We divide its functionality into three parts; interview administration, interview process and learning component; and discuss solutions to each.

3.1 The DueDive Due Diligence Interview Process

We focus on the interview part of the due diligence process at DueDive, which currently relies on an expert interviewer. Figure 1 outlines the main steps of this process. Every time a new due diligence request arrives (step 1), the company performs interviews with key people. In cooperation with the company that is assessed, a representative of DueDive goes through a catalog of questions (Q)(step 2). Not all of the questions are asked in every session, as some might not be applicable based on the current circumstances. The answers (A)(step 3) aim to give the interviewer an idea of the value, based on their experience. The problem with this process is that the experience is only in the mind of the interviewer (step 4). The information is only collected in a single document, which does not allow linking information between different interview sessions.

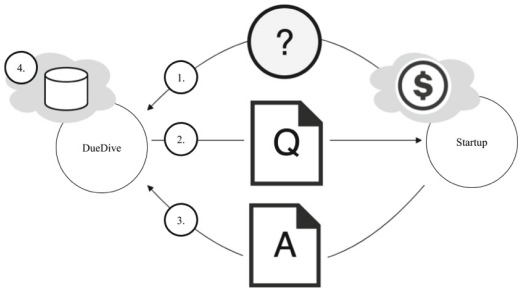


Fig. 1. The due diligence process.

3.2 Entities of the Solution

We extract a set of entities from the description of the interview process. These will form the basis for our solution:

Interviewer carries out the interview process by asking questions and interact with the CAPI system to add all necessary information.
Interviewee answers the questions but does not interact with the CAPI system.
Interview consists of a set of interview sections.
Interview Section groups questions that belong to the same topic to give more structure and help the interviewer cover all relevant information.
Question asked by the interviewer to get information from the interviewee.
Answer is a text representation of the response to the question given by the interviewee. The interviewer classifies the answers or marks them not applicable if they do not add any new information.
Additional Question comes to the mind of the interviewer after a specific answer. Additional questions are important because they are fitted to the interview session and the current circumstances.
Interview Session is an particular instance of an interview that could cover at most all questions contained in that interview.

3.3 Components of the Solution

The description of the interview process focuses on a conversation, which can be supported by a CAPI system showing questions and allowing the interviewer to record the answers e.g., by typing. The questions need to be defined and arranged beforehand, and the answers need to be accessible afterwards. Also there is a need for a learning component that can help the interviewer with for example the order of the questions. Our solution requires the following three components:

Interview Administration. At the beginning of each interview, the structure needs to be defined, which includes the interview categories that each contain multiple questions. For each question, it is possible to define what a high or low classification of a question is.

Interview Process. Once the interview is set up, the interview process itself can begin. A new interview session is created for each interviewee. During an interview session, all the questions are listed and can be picked in any order. The answers are recorded and classified by the interviewer. In case any additional questions come up as part of a question, they can be added to the answer.

Learning Component. When an interviewer performs multiple interview sessions, patterns in additional questions emerge. The learning component should predict whether previous additional questions are applicable in the current session.

3.4 Designing the Learning Component

There are some factors that influence the due diligence interviews, e.g., company size, investment type, and technology stack. If these factors are similar between two different companies, the interview flow should also be similar. An expert due diligence assessor can, based on these factors, identify the right questions to ask and in what order. The learning component should emulate this behavior, and based on the answers, suggest a question flow to a less experienced interviewer.

We want to be able to change the interview process e.g., adding a new category of questions or new questions, without breaking the learning component. The learning component should, in addition to being flexible, bring value to the interview process as quickly as possible and continue to improve as more sessions are available to learn from. We divide the learning component into the following features and discuss design alternatives for each of them:

Flow Continuation. enables the interviewer to make an informed decision based on the same data an expert interviewer would use to pick the question that yields the most information next.

Flow Shortening. A set of questions that are predetermined and might contain questions that do not apply to a specific interview, so to improve efficiency, the learning component should decide whether to skip any questions in this particular session.

Prediction of Additional Questions. When an interviewer performs multiple interview sessions, patterns in additional questions emerge. The learning component should predict whether previous additional questions are applicable in the current session.

Design Options for Flow Continuation. The flow continuation should mimic the way an expert interviewer uses a mental model of how questions are connected to guide the interview. Based on the classifications of particular answers, the interviewer will decide the next question. The learning component should learn this behavior by predicting specific attributes of the questions. These attributes include how likely the question is to be answered, if the question is applicable, and which classification the answer might get. This knowledge is derived from previously completed interview sessions of the same interview.

The learning component should take into account that questions can be answered in any order and that not all questions need to be asked during every interview. Also, empty attributes need to be handled properly. The prediction should be based on the current state of the session, i.e., the set of questions asked and answers given. The following criteria are used to evaluate which of the methods is the most suitable:

Multi Attribute Prediction. To predict multiple attributes, such as classification and the assessment whether a question retrieved new information at once. Can predict other attributes besides the order of questions.
Order Independent. Question order does not matter.
Session Completion. Can predict incomplete sessions.
Matching Prediction. Predicts accurately given a 100 % match with previous sessions.
Closest Matching Prediction. Predicts quite accurately given a less than 100 % match with previous sessions.
Flexible Interview Structure. Supports change of structure after initial setup.
Result Count Independence. Calculates results for both few and many sessions completed.

Decision Tree. Assume that questions in the interview are states and that there is a certain probability to go from one state to another, based on the current position in the interview session. In this case, the interview session forms a tree where questions (nodes) are connected to each other, and the state of the current session can predict missing nodes. This is similar to a decision tree.

However, decision trees predict a single attribute. In an interview, we need to predict several attributes at once, such as the classification of answers, the probability of getting an answer, and whether the question is applicable. This would require that every attribute that should be predicted forms a tree, which might result in performance problems as the number of attributes grows. Additionally, every question that is not answered and therefore needs a prediction needs the same number of trees again.

As the order of the questions is not fixed, the trees cannot be pre-calculated; there are too many sequences in which the questions can be answered. So, the trees must be calculated at runtime, which can cause performance problem.

Markov Chains. Markov chains can predict future states based on the current state. Higher order Markov chains are for example used to predict the next words based on previously typed words on mobile phone. The questions in an interview can be considered as words that are typed, so similarly to predict the next word it can be used to predict the next question. The higher the order of Markov chains, the longer the streaks of questions are considered.

Considering questions as words result in that only the order of the questions is encoded. This leaves out many attributes that are gathered by the previous session and makes it impossible to differentiate between a high or low classification of answers. So, the quality of the prediction might be reduced.

Sentence Prediction (n-grams). Similarly to Markov chains, n-grams can be used to predict next letters when typing. For an interview, a question would be a single letter in a word, and the previous sessions would be complete words. Markov chains were not able to complete a currently unfinished interview without a brute force search.

A problem with Markov chains is that they can predict an entirely new session, but they cannot complete a unfinished session. An alternative would be to predict new sessions until a partial match with the current session is found, but this brute force search is again a highly inefficient procedure. Compared to Markov chains, n-grams can complete currently unfinished interview sessions, so no brute force is necessary to retrieve a completion of the current session.

Even though n-grams solve one problem of Markov chains, some of the other issues persist. For example, n-grams would still only predict the order of questions but not the state of the question. Even though the current incomplete session can be completed, it does not take similar matches of previous words into account. In practice, when predicting words, the predictions take into account the letters that are close to each other to make up for typing errors. This could be used to check for similar sessions, but we do not know which questions that are close to each other.

Imputation. One way of predicting a new attribute is to take the mean or median of the same attribute in the completed sessions and use this to predict all relevant attributes of all questions. This can be pre-computed and does not change during the interview. The problem is that the attributes that are predicted are not based on the current state but only on the previously completed sessions. This is illustrated in Table 1. In sessions 5 and 6, questions C and D are predicted even though both sessions have different states in questions A and B since the imputations are the same.

If we increase the weight of completed sessions that are similar to the current session and compare completed sessions that have a completely different flow and classification of questions, we can improve how well the approach performs. We

Table 1. Example of imputation to predict flow, using Mean and KNN. 1 indicates answered and 0 indicates not answered. ? indicates not completed yet and the number indicates the probability that the question should be asked in this session.

Session	Mean				KNN			
	Question				Question			
	A	B	C	D	A	B	C	D
1	1	0	0	1	1	0	0	1
2	1	0	0	0	1	0	0	0
3	1	0	0	1	1	0	0	1
4	1	1	1	0	1	1	1	0
5	1	0	? → 1/4	? → 1/2	1	0	? → 0	? → 2/3
6	1	1	? → 1/4	? → 1/2	1	1	? → 1	? → 0

calculate a distance measure of both sessions using, e.g., KNN, as illustrated in Table 1. Session 5 is now matched with its nearest neighbors sessions 1, 2, and 3. The prediction is now only based on those sessions and leaves out sessions that are too different in the current state. The state of question C in all three neighbors is the same so the prediction is also not to answer C. On the other hand, question D has different states in the neighbor sessions, and the mean will be used to predict the value. This suggests that the question D should be answered with a probability of 2/3 in session 4.

Since the predictions will then be the same as the neighboring session, it is vital to complete enough interviews to have enough data to train the learning component.

Design Options for Flow Shortening. Since the time and attention span of both the interviewer and the interviewee are limited, the interview should be as short as possible. Furthermore, not all questions are always applicable, depending on the circumstances derived previously in the interview session. Sometimes, questions correlate and do not result in any information gain, so asking both questions would only extend the length of the interview. Questions to double-check if the interviewee is paying attention and trustworthy are neglected here, as the expert interviewer will judge this. Therefore, it can be efficient not to ask all the questions.

Compared to flow continuation, flow shortening will only be used after completing a session and not continuously during the process, mainly because the session needs to be completed to be useful for analysis. Secondly, a domain expert needs to handpick the rules, as they are not added automatically.

After the expert selects the applicable rules, these can be used in the interview process to shorten it. Specific questions can be marked as not applicable based on a flow-shortening rule even if the flow continuation states that this question should be answered.

Table 2. Basket analysis results.

#	Antecedents	Consequents	Support	Lift	Confidence
1	$Q1_{low}$	$Q3_{high}$	0.4	1.7	1.0
2	$Q2_{high}$	$Q3_{low}$	0.3	1.4	0.9
3	$Q1_{low}$	$Q2_{low}$	0.2	1.5	1.0

Basket Analysis. Basket analysis can find questions that are often answered similarly between sessions. When we apply basket analysis to the interview domain, the questions are columns, and the sessions are rows of the matrix.

Since a basket analysis usually cannot handle ordinal values, we need to map these to nominal values. We do this by mapping a classification of 4 or 5 to a high attribute classification and 1 or 2 to a low attribute classification. Table 2 shows the result of basket analysis. The rules are found in the columns Antecedents and Consequents. Once the rules are generated, support, lift, and confidence of the basket analysis together with expert knowledge is used to evaluate the rules and determine which are most applicable.

Cross Tabulation. The ordinal values, especially those in the classification of the answer, can be of value to the analysis. Since the ordinal values are partly lost in the basket analysis, cross tabulation should be considered to keep this information encoded in order of the classification.

The benefit of keeping the ordinal values is a granular differentiation between the classifications and the order that the classification initially contained. This improvement can be compared to the imputation for the flow continuation, as the close matches could be considered with a higher relevance than matches that differ a lot from the classification of another session.

But cross tabulation makes many underlying assumptions that are problematic, especially in the early stages of an interview. Cross tabulation requires large sample sizes for proper analysis. So it will take time to collect a large enough sample. Missing values are another issue, as the analysis cannot handle them. Since many questions are not mandatory and some questions can be skipped, empty values can break the algorithm. Skipping empty values for the analysis or filling the missing value with a neutral value solve this problem.

Similar Question Ranking. There are correlations in between questions of the interview. This aids the interviewer in picking from the additional questions that were previously asked. Since there might be several additional questions from previous sessions, it would be helpful to the interviewer to rank these in order of applicability.

We assume that an additional question can be imputed to the flow continuation. However, this could only lead to the statement of whether or not there should be an additional question. Continually expanding the data structure with all the additional questions is not sufficient since the cause for an additional question is often a previous answer.

To decide which additional questions are most likely in the current situation, we developed a similarity measure that compares questions that the additional questions are based on. If they are flagged as not applicable, the similarity is 1, i.e., the highest possible similarity. If both questions are flagged as applicable, the similarity is determined as $1 - |current\_class - other\_class|/\#steps$. This value can then be used to rank the additional questions.

3.5 Prediction Enhancements Using Natural Language Processing

If we assume that an answer text that is similar would lead to the same additional question, it would be useful to compare answer texts. Up until now the answer text is not evaluated and only recorded for manual analysis by the expert. Though, the similarity of answer texts could be evaluated using e.g., TF-IDF or entity extraction.

To evaluate the feasibility of these algorithms, we used book reviews from Amazon and tried to automatically assess which reviews belong to which book. We found that using the sequence matcher and TF-IDF appeared to have significant success rate in choosing the correct books. Using entity extraction on its own is not enough, because the entities still need to be compared and associated, and henceforth this is done with TF-IDF.

4 Implementation

We describe the prototype implementation and architecture in a real environment. Figure 2 depicts the architecture of the CAPI system. The system is implemented as a web application with Django and VueJS. PostgreSQL is used to store the questions and answers. Since the system should work offline, all components need to be installed on the computer that is used during the interviews. To make the installation process more manageable, the system is packaged as a Docker container. This makes it easy to install on Windows and macOS computers. Interview data can be shared between computers and sessions by backing up and restoring the database.

The learning component was implemented as a REST-API. The primary challenge was to implement flow shortening using basket analysis. The learning component is supposed to generate a suggestion as soon as possible, so the algorithm needs to operate on as few interview sessions as possible. This resulted in long calculation times for the a priori algorithm, due to the sheer amount of relations. To fix this, we introduced a parameter that allows us to specify a threshold to manage how many rules that are generated. As a result, the same view can be used in the early stages of an interview with few completed interview sessions and in the later stages with many completed sessions. Depending on the progress, the threshold can be adjusted to generate more or fewer rules.

Before implementing the user interface, a mock-up was developed for the interview-process component. This mock-up helped to gather input and feedback from DueDive. Figure 3 shows the list of questions of an interview. Next

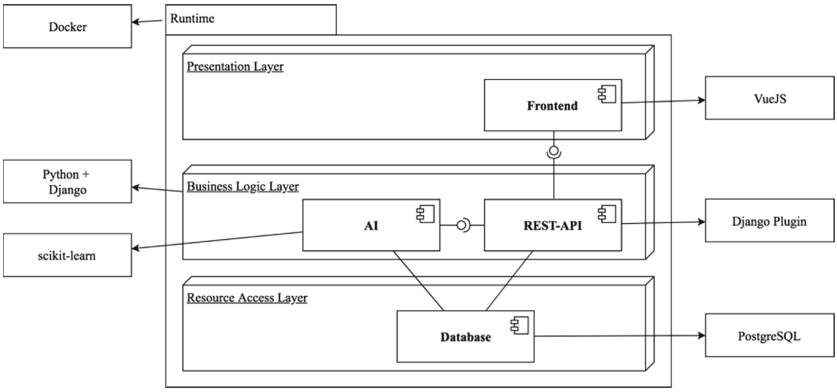


Fig. 2. The architecture of the CAPI system.

to the questions, some question attributes and details from the learning component are displayed. After selecting a question, the question detail view opens, as seen in Fig. 4. It shows all information about the question, including additional questions.

5 Evaluation

We evaluated the prototype CAPI system to determine how well it supports the interviewer and the interview process. We specifically wanted to answer: does the user interface support conducting the interview, is the learning component helpful during the process, and can the predictions potentially result in a relative short but informative interview?

To get a sufficient dataset and avoid performing a large number of interviews, we initially wanted to do a retrospective case study. There was not enough data available from DueDive, so we investigated available datasets online, e.g., from Kaggle and other machine learning dataset repositories. We were unable to find a perfect dataset so the best candidate was the Young People Survey[2], which contains 150 questions with mostly ordinal values. The problem with this dataset is that many attributes that are essential to the prediction algorithms are missing.

Since we were unable to find a suitable dataset, we performed a prospective case study instead, and developed a set of questions and selected and subsequently interviewed a group of people with relevant but varying personalities and traits concerning the questions. Based on seven interviews, we found the user interface comfortable; it provided a good overview of the current state of the interview. The learning components integrated seamlessly and showed the most critical information.

[2] https://www.kaggle.com/miroslavsabo/young-people-survey.

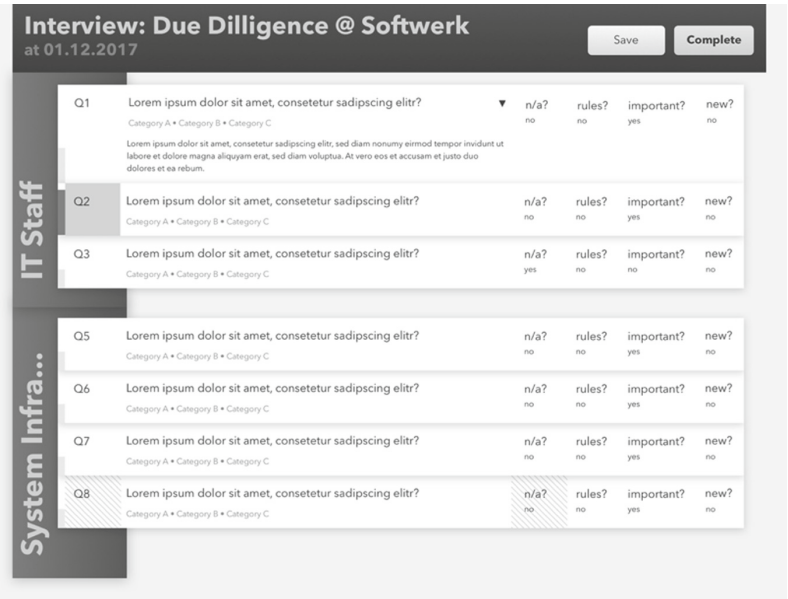


Fig. 3. The interview session overview. The first question is already answered and predictions for the other questions are shown.

While assessing the imputations made by the flow continuation component, it became evident that correlations between one and another question are difficult to discover. This is because the imputation assesses the whole interview at once and other questions introduce too much noise into the imputation. However, this can be neglected as the flow-shortening component covers those correlations.

Other than that, it was apparent that a small amount of sessions leads to highly biased recommendations by the learning components. The bias improves with the number of sessions completed but interviewer should keep in mind to avoid being misled by the prediction in the early stages of an interview. It might be better if all questions are asked during the training phase, and once enough data is collected, the flow shortening can be enabled. This will improve the stability of the predictions and prevent skipping some questions entirely. Questions that are only answered once will have a substantial impact and should be assessed with care, even after the training phase. At the same time, answering all questions eventually results in the learning component predicting the answer for many other questions. One way to deal with this can be to adjust the threshold that is used to predict that a question should be asked. This should be done after a few real-life sessions are carried out.

We find that the user interface supports the interview and that the learning component could potentially do so as well, but it will need more than seven sessions and require some adjustments of various thresholds and parameters based on the interviews, questions, etc.

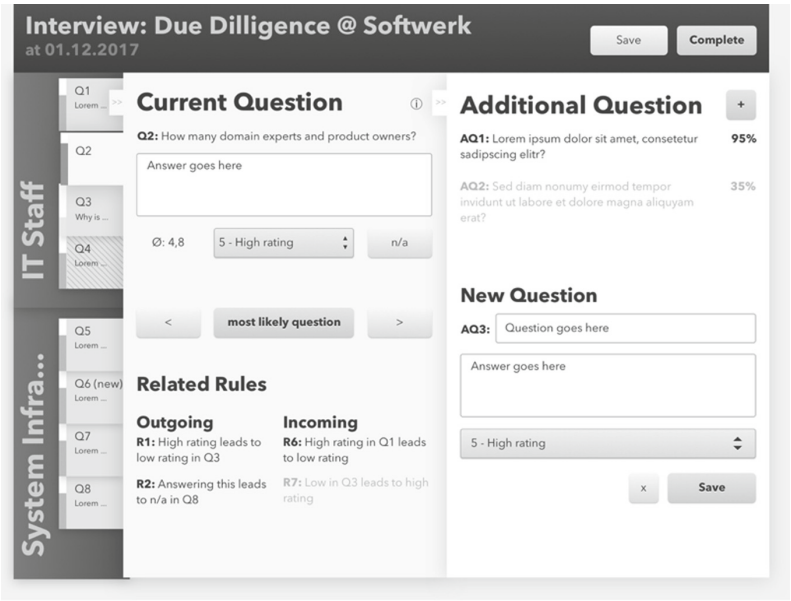


Fig. 4. The question detail view shows information about a question, including additional questions.

6 Conclusions and Future Work

We designed and implemented a CAPI system where machine learning is used to help guide non-experts through due diligence interviews. An experienced interviewer can change the flow of an interview by skipping or changing the order of questions. They can also find the need for additional questions based on the answers to previous questions. In our system, machine learning is used to change or shorten the flow of an interview and to suggest relevant additional questions. Based on our evaluation, we found that the CAPI system was easy to use, but that the machine learning components need more data and additional tuning to perform well. So as part of future work the prototype need to be used more to investigate how the machine learning components can be improved.

In our small case study, we found that is important that an expert who knows which questions to ask conducts the initial interviews. As soon as there are enough sessions in the database, less experienced interviewers can start to carry out interviews. The CAPI system should be extended with a mechanism to signal once enough data is available. We included a general questions section since we assumed this would improve the flow shortening by gathering basic data that could be used to predict answers to later questions. We did not observe this, so more interviews are needed to see if this holds. We only used the KNN algorithm for imputation and flow continuation. We used it because it is easy to use and assess, but other algorithms and their performance could be tested,

such as Matrix Factorization and Nuclear Norm Minimization. There are also alternatives to imputation that can be considered; Markov chains Monte Carlo or autoregressive models might be able to overcome some of the problems with Markov chains or n-grams for flow continuation.

In the end, we did not use cross tabulation, since the conversion from ordinal to nominal is sufficient. But this could change once we have more data and we need to adopt. Finally, there are multiple ways that natural language processing could be used to improve the process. We could for example use it to detect contradictions in the answers.

References

1. Richter, F.: Startup Funding Shows Signs of New Tech Bubble, September 2014. https://www.statista.com/chart/2732/venture-capital-investments-in-the-us/. Accessed 15 May 2018
2. Wall, D., Dally, R., Riannon, L., Jung, J., DeLuca, T.: Use of artificial intelligence to shorten the behavioral diagnosis of autism. PLOS One **7**(8), e43855 (2012). 1–8
3. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
4. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
5. Sarukkai, R.R.: Link prediction and path analysis using Markov chains. Comput. Netw. **33**(1–6), 377–386 (2000)
6. Ching, W.K., Huang, X., Ng, M.K., Siu, T.K.: Markov Chains: Models, Algorithms and Applications. Springer, Heidelberg (2013). https://doi.org/10.1007/0-387-29337-X
7. Diaconis, P.: The Markov chain Monte Carlo revolution. Am. Math. Soci. Bull. New Ser. **46**(2), 179–205 (2009)
8. Akaike, H.: Fitting autoregressive models for prediction. Ann. Inst. Stat. Math. **21**(1), 243–247 (1969)
9. Gelman, A., Hill, J.: Data Analysis Using Regression and Multilevel/Hierarchical Models. Analytical Methods for Social Research. Cambridge University Press, Cambridge (2006)
10. Linoff, G.S., Berry, M.J.A.: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3rd edn. Wiley, Hoboken (2011)
11. Karel, F.: Quantitative and ordinal association rules mining (QAR mining). In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) KES 2006. LNCS (LNAI), vol. 4251, pp. 195–202. Springer, Heidelberg (2006). https://doi.org/10.1007/11892960_24
12. Bates, M.: Models of natural language understanding. Natl. Acad. Sci. **92**(22), 9977–9982 (1995)
13. Rajaraman, A., Ullman, J.D.: Mining of Massive Datasets. Cambridge University Press, Cambridge (2011)
14. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)

Business and Information Systems Development

A Projection-Based Approach for Development of Domain-Specific Languages

Boris Ulitin(✉), Eduard Babkin, and Tatiana Babkina

National Research University Higher School of Economics,
Nizhny Novgorod, Russia
{bulitin, eababkin, tbabkina}@hse.ru

Abstract. The article is related to the problem of sustainable flexibility of a domain-specific language (DSL) in response to evolution of the application domain and users' capabilities. We offer a solution of that problem based on a particular model-driven approach. We propose to create a DSL structure from the domain-semantic-model (DSM) through the so-called semantic projection mechanism. The semantic projection is an operation, which is conducted over DSM. The result of the projection is a fragment of DSM, which describes the semantic model of a particular DSL dialect. We suggest to apply a group of model-to-model (M2M) transformations for practical implementation of semantic projections and producing corresponding DSL artefacts. We demonstrate the application of the proposed approach by the example in railway allocation domain.

Keywords: Domain-specific language · Domain-semantic model
Projection · Model-driven development · Model-to-model transformations
Evolution · Railway transportation

1 Introduction

Currently, domain-specific languages (DSL) become more and more widespread. Such popularity can be explained by the fact, that DSL is a fairly simple and convenient way of organizing work in a certain target domain. DSLs contain only the required set of terms of the domain, representing some kind of its conceptualization, because every DSL uses as its basis some model of the application domain, which is named as Domain Semantic Model (DSM) [1]. Using the concept of DSM allows the users to achieve addressing resilience of businesses information systems because of integration of DSM and general model-oriented approach in development of DSL and its consequent evolution process.

The DSL life cycle consists of five development phases: decision, analysis, design, implementation and deployment [2]. In our work we concentrate on the first three phases, leaving behind a framework for considering the specific application of the created DSLs. The first two phases are connected with the target domain, for which DSL is developing, and result in the definition of some domain model and gathering domain knowledge. Only after that the DSL design is initiated. Traditionally, the DSL design is organized in two orthogonal ways: to develop DSL from scratch or to base it on the existing one. However, both cases do not take into account any formal results,

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 219–234, 2018.
https://doi.org/10.1007/978-3-319-99951-7_15

which can be obtained during the domain analysis phase. Although DSM contains all terms and relationships between them, needed to be included into DSL created. As a result, the designer actually repeats the provided analysis phase while designing DSL and, actually, creates one more DSM for DSL created instead of transferring once defined DSM into DSL structure.

This is especially useful, considering that the DSL structure contains semantic and syntactic levels. The syntactic level is responsible for the opportunity to define some context, while the semantic level reflects this context on the concepts of the target domain. As a result, the semantic level absolutely depends on the model of the target domain, used during DSL creation. Taking into account, that DSL is based on a specific DSM, we can argue that the semantic model of the DSL can be obtained by projecting the corresponding DSM.

At the same time, the syntactic level of DSL is a result of combination of the general domain concepts with the problem-specific tasks of the users. This means, the syntactic part of DSL is a result of adaptation of the DSL semantic model to information and operational needs, experience and skills of every specific user. As a result, in parallel to the development of the skills and knowledge of the user, the set of DSL terms, that he/she operates with, can also change. The syntactic model of DSL is also the result of projection of the DSL semantic model on the user's needs and can be based on the principles of projecting DSM into the DSL semantic model. This co-directional development of both levels of DSL leads to the necessity of sustainable maintaining a growing arbitrary tree of syntactic and semantic dialects of the original DSL. Some dialects may be based on the one semantic DSL model, while other dialects use different DSL semantic models on the one DSM.

Certain attempts to facilitate such sustainable maintenance of DSL dialects were done by several researchers. For example, Cleenewerck in [3] tries to identify the tree of DSL dialects using graph-transformations for transition between different DSL dialects. However, every DSL dialect in this approach is a combination of the syntactic and semantic parts. Such a mechanism, when the environment stores not just different components of DSL, but the complete structure of different DSL dialects and its handlers, is called versification [2]. Obviously, this approach does not allow us to organize a flexible and rapid transition from one version of DSL to another, because it causes the semantic transition between different DSLs.

A similar model-oriented approach is described by Haav et al. in [12], but they concentrate only on the opportunity to transform the DSL semantic level into the syntactic level, without reference to DSM. Furthermore, the structure of DSL in this case is considered stable over the time. In order to support co-directional development of both DSL levels, we propose to use the notion of semantic projections. Semantic projection is a procedure which produces a subset of the original DSM. This subset fully describes the DSL semantic model of a particular DSL dialect. In our case the group of semantic projections formally specifies transitions between all known DSL models.

For practical implementation of our approach we suggest to consider the object-oriented paradigm, because in that case DSL becomes a combination of objects and operations on these objects. Consequently, the semantic part of DSL can be represented as an object-oriented model, which is a result of DSM semantic projection.

Furthermore, the syntactic part can also be formalized as an object-oriented model, which components are projections of the semantic model. From such point of view, we can organize the DSL development as a unified process of sequential projection of some models into others. In order to provide such projections, the mechanism of model-to-model (M2M) transformations [10] can be used. In order to achieve these goals, we adopt the graph-oriented transformations, described in our previous article [15], for projection of DSM into DSL semantic model and expand it for organizing projections of the DSL semantic model to the syntactic level. From the point of implementation, we propose to use such model-driven instruments as Eclipse GMF platform for models' definition and QVT language for realization of M2M transformations among them.

Our projection-based principle of sequential creation of DSL from DSM to a specific semantic model, differs substantially from existing approaches [2–4], because the proposed projection-based procedure of DSL development allows not only to unify all stages of DSL creation, but also modify different parts of DSL independently. Also, our approach eliminates the need to recreate the DSL in the case of making any changes to it without the opportunity to modify different DSL parts separately.

As a result, the proposed approach allows not only to construct one specific DSL, but also to modify it in accordance with changes in DSM. Thus, it is possible to determine the different syntactic dialects of DSL over one DSL semantic model as well as to define several semantic models of the DSL over one DSM model. All these models (DSM, the DSL semantic and syntactic models) are dynamic and can be updated consistently over the time in accordance with the changes that are taking place in the domain.

The proposed approach can be the most effective in the domains with a high need for adaptation to end users. For example, in a case of definition of DSLs for elderly people, which are limited not only because of the domain's restrictions, but and physical and psychological limitations of the actual users. Another case of application are rapidly evolving domains, where new terms are appeared and have to be included in DSL developed. We show the effectiveness of the proposed approach by the example of DSL development for the railway allocation domain: starting from the DSM development and the further definition of two different consistent DSL dialects, indicating cyber-resilience of the established system.

The article describes our proposed approach as follows. In Sect. 2 we give some facts from the theory of design of domain-specific languages, define the DSL structure in model-oriented manner and analyze various tools and notations, appropriate for DSL model-oriented development. Section 3 describes a high-level design and technologies of the proposed approach to DSL development. Section 4 demonstrates an example of application of the proposed approach to the railway allocation domain. We conclude the article with analysis of the results and specification of the future researches.

2 Background

2.1 Definition of Domain Semantic Model (DSM)

DSM offers a flexible and agile representation of domain knowledge. DSM can be constituted by either just small pieces of a domain knowledge (e.g. small taxonomies equipped with few rules) or rich and complex ontologies [13] (obtained, for example, by translating existing ontologies). That gives respectively weak or rich and detailed representation of a domain [14]. More formally DSM is a seven-tuple of the form:

$$DSM = (\mathcal{H}_C, \mathcal{H}_R, O, R, A, M, D)$$

where

- $\mathcal{H}_C$ and $\mathcal{H}_R$ are sets of classes and relations schemas. Each schema is constituted by a set of attributes, the type of each attribute is a class. In both $\mathcal{H}_C$ and $\mathcal{H}_R$ are defined partial orders allowing the representation of concepts and relation taxonomies;
- O and R are sets of class and relation instances also called objects and tuples;
- A is a set of axioms represented by special rules expressing constraints about the represented knowledge;
- M is a set of reasoning modules that are logic programs constituted by a set of (disjunctive) rules that allows to reason about the represented and stored knowledge, so new knowledge not explicitly declared can be inferred;
- D is a set of descriptors (i.e. production rules in a two-dimensional object-oriented attribute grammar) enabling the recognition of class (concept) instances contained in O, so their annotation, extraction and storing is possible.

In our research, we focus only on the sets O and R. Consideration of other parts of DSM is beyond the scope of our study, since it determines more meta-characteristics of DSM itself, rather than the objects and connections between them, which are the most interesting for the further development of the DSL semantic model.

It is also important to note, that DSM usually has a dynamic structure, demonstrates a tendency to changes over the time (evolution, in other words). In accordance with the evolution of the domain, the evolution of its DSM also occurs [2]. As a result, any DSL, based on the corresponding DSM, should be adopted according to the changes. Consequently, the structure of the DSL metamodel should be as close as possible to the structure of DSM in order to guarantee the coherence between DSL and the target domain. So, that is reasonable to select a common meta-meta model which will be used both for definition of a DSL metamodel and DSM. We believe that a widely accepted object-oriented meta-meta model can be suitable for our purposes. The following manifestation of DSL as a special kind of the object-oriented model proves that believe.

2.2 Manifestation of DSL in Terms of Object-Oriented Models

A domain-specific language (DSL) is a computer language specialized to a particular application domain. This is in contrast to a general-purpose language, which is broadly applicable across domains, and lacks specialized features for a particular domain [1]. In [5] two parts of the DSL are identified: (1) a syntactic part, which defines the constructions of DSL; and (2) a semantic part, which manifests itself in the semantic model. The first part allows defining the context for working with the second one, which defines meaning of DSL commands in terms of the target domain.

From the formal point of view, a semantic part of DSL is derived from the set of objects of the target domain and operations on them. The structure of every model is a combination (E, R) of some entities (each entity is a set of its attributes $e_i = \{attr_{i_1}, attr_{i_2}, \ldots, attr_{i_M}\}, M \in \mathbb{N},\ i = 1, N$) and relations between them. Therefore, we can formalize the semantic DSL level in a model-oriented manner as a combination (O, R) of some objects of the target domain and relations between them, where each object is a set of its attributes and operations $o_i = (Attr_i, Opp_i) = (\{attr_{i_1}, attr_{i_2}, \ldots, attr_{i_M}\}, \{opp_{i_1}, opp_{i_2}, \ldots, opp_{i_K}\}), M, K \in \mathbb{N},\ i = 1, N$).

In these terms, the syntactic part of DSL can be represented as a subset of the semantic level, needed for representation of a certain problem situation. The very one difference is, that the syntactic part may not absolutely reflect the semantic constructions but identify its own definitions (pseudonyms) for the semantic constructions, according to the user's needs.

In addition, a syntactic part of DSL can also be separated into two levels: the level of objects and the level of functions. The object-level is equivalent to the set of objects of the semantic level. The functional level contains operations, which allow to specify the operational context for the objects.

As follows, the structure of the syntactic level can be formalized as a triple $\left(O_{syntax}, R_{syntax}, Alias_{syntax}\right)$, where $O_{syntax} \subseteq O$ and $R_{syntax} \subseteq R$ are the subsets of objects and relations between them of the semantic DSL level respectively, and $Alias_{syntax}$ is a set of pseudonyms for objects' components (attributes and operations).

Under these circumstances, we can tell about the complete model-oriented representation of the DSL structure. The structure allows us not only to describe both levels of DSL in structured and unified manner but optimize the process of DSL development and evolution by introducing several syntactic DSL dialects on one fixed semantic level. Furthermore, the versification of DSL can be provided in a similar way on the semantic level as well as on the syntactic level, without need to re-create the whole DSL structure every time, when the changes are required.

3 Proposed Approach

3.1 A Semantic Hierarchy of Model-Oriented DSL Development

Combining the object-oriented model of the DSL structure with the formal definition of DSM on the basis of a single meta-meta model, we can specialize a well-known semantic hierarchy of meta-models for our approach to model-oriented development and evolution of DSL (Fig. 1).

In our case this hierarchy is separated into four layers, according to the stages of the DSL development. Each lower level is based on the model artefacts of the upper level.

A single M3 meta-meta-model determines common grounds for all meta- and models of the lower levels. This meta-level defines also notations in which concrete models will be defined and what rules for their transformations will be used.

The structure of the semantic hierarchy determines the corresponding process of DSL creation. It starts with the definition of DSM, containing all important entities of the target domain and relationships between them. The process of DSM creation is beyond the scope of our current research, we propose its consistency for DSL development. For more details on DSM definition and checking its correctness, see [13].

When DSM is created, we can build the DSL semantic model by the operation of semantic projection. Any semantic projection performs a certain M2M transformation of DSM to some its fragment. Thus, semantic projection fully determines the semantic model of a particular dialect of DSL. In this case the semantic model becomes an object-temporal structure, because it should be adopted according changes in DSM over the time, thereby defining a new object filling of the DSM.

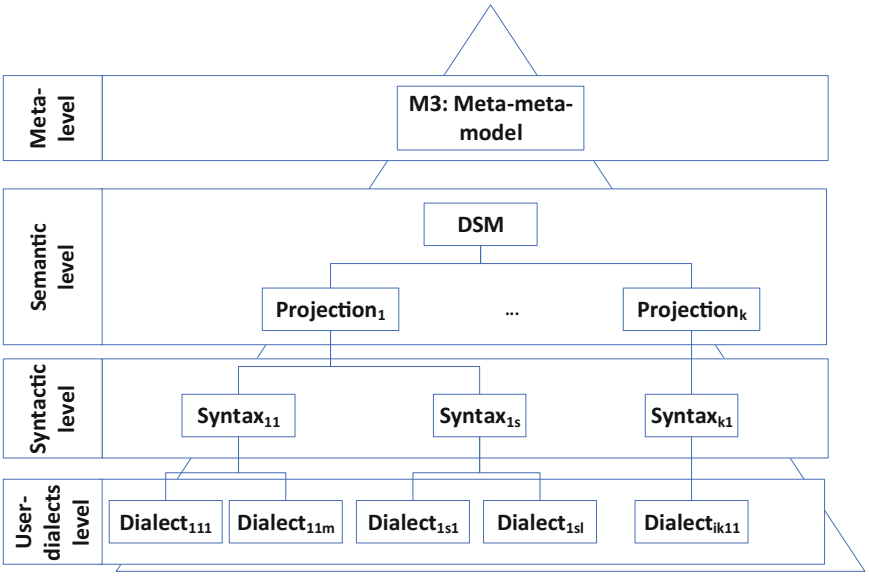


Fig. 1. The semantic hierarchy of projection-based DSL development

After the semantic projection was performed, the syntactic level of DSL can be developed by a M2M transformation of the result of the corresponding projection. What is important, these DSL syntactic models are independent of each other and are determined by end-users in accordance with the adaptation of the semantic projection to their own tasks.

Finally, created syntaxes are used by the end-users of DSL, who determine the set of DSL dialects within the single specific syntactic model.

Figure 2 shows differences between traditional approaches and our proposals. Traditional approaches start with the manual definition of the DSL concrete syntax which is followed by the translation of the syntax in terms of grammars. Consequently, every change in the target domain leads to the need to redefine the DSL concrete syntax and re-create the corresponding grammar. A similar process repeats in a case, when changes in DSL are caused by the end-users. As a result, outcomes of traditional approaches contain inconsistent dialects of DSL, which cannot be mapped among themselves due to differences in all levels of the DSL structure.

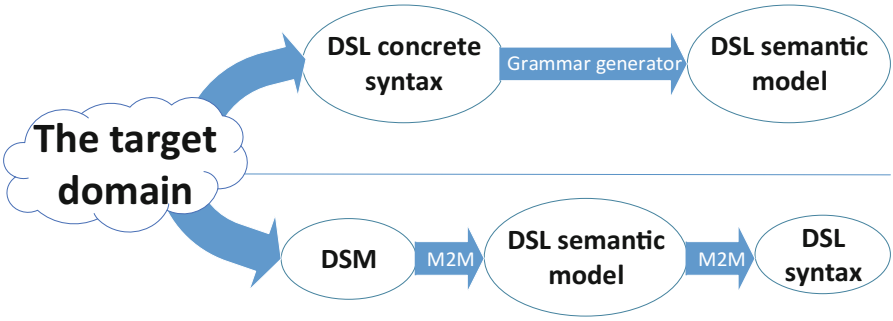


Fig. 2. The scheme of differences between traditional (top) and projection-based (bottom) DSL development approaches

In comparison to traditional approaches, the proposed projection-based scheme (Fig. 2top) of DSL development is organized in the strong correspondence of the target domain. Such correspondence is provided by the consequent projections among different models in a semi-automatic way through M2M transformations: from DSM into a DSL semantic model and then into a syntactic model of the specific DSL dialect. As a result, we can define several DSL syntactic dialects over one specific DSL semantic model, which will be consistent and can be transformed between themselves without the redefinition of the DSL semantic models.

3.2 Model-Driven Tools for DSL Development

After the definition of the projection-based approach for DSL development, we need to analyze available tools, which allow us to practically develop DSL according to the projection-based principles.

This seems to be important, because each tool supports a limited set of notations for the model-oriented DSL representation. That, in turn, any tool imposes restrictions on the feasibility of practical implementation of M2M transformations.

Our analysis shows that there are several tools, which allow us to define or transform DSL in a model-oriented manner (see Table 1).

Table 1. Available tools for DSL model-oriented development

Tool	Supported notations	Transformation support	Supported transformation formats
MetaEdit+	GOPRR	None	None
Eclipse GMP + XText	MetaGMP, ECORE	Complete	QVT, ATL
MPS	Own textual notation	None	None
MetaLanguage	Own textual notation	Partial (whole DSL structure versions, without transformation on separate levels)	ATL

As we can see, only two of four tools provide an opportunity not only to create DSL, but also to make changes in it: Eclipse GMP [7] and MetaLanguage (http://pespmc1.vub.ac.be/METALARE.html). Both tools support organization of transformations of the once created DSL structure, using the ATL language. However, MetaLanguage is based on the following inappropriate principle: any DSL transformation is equivalent to the DSL re-creation, that means the need to redefine the whole DSL structure instead of definition of the transformation of its separated components of the semantic and/or syntactic levels. Such a limited toolkit paired with the need to use MetaLanguage specific notation to determine the DSL model, makes it impossible to conduct flexible transformations of the DSL in this environment.

```
modeltype modelTypeName;
transformation ruleName (in i : originalModel,
                         out o : targetModel);
mapping originalModelType::originalModelElement() : targetMod-
elElement
{
  actions
}
```

Fig. 3. The scheme for definition the transformation rules in QVT

The only possible option for carrying out flexible DSL modifications is the Eclipse GMP. This platform provides unified and standardized notations to define the model-oriented structure of DSL, and also allows users to organize the transformations of various DSL components without redefining the previously created structure. In order to define the formal specification of transformations a set of specialized graph-transformation languages can be used such as ATL Transformation Language (http://www.eclipse.org/atl/), QVT (Query/View/Transformation) [6], etc.

In our research, we propose to use QVT, because this language allows us to describe the transformation rules from any original model into any target model, conducting a transformation at the level of meta-models. In QVT the rule definition includes five components: the name of the model type and the rule, the type of the element of the original model, the type of the equivalent element of the target model and the block for definition of the additional actions with transforming and corresponding components of the original and the target model (Fig. 3). In addition, QVT supports not only a textual form for rule definition, but also a model-driven rule architecture, which permits automatic organization of the transformation according to the definition of the original and target models.

In our research we have studied appropriateness of the Eclipse GMP platform for design and further modification of DSL in the particular domain of resources allocation in the railways. However, this does not limit the possibility of using this tool to create DSLs in other domains. The only thing that will be required for this is to determine the structure of DSL in terms of the ECORE notation and the rules for its further modification.

ECORE is an EMF's UML-based object-oriented modelling language. Inheriting from the object-oriented UML notation, the ECORE notation also represents a model in the form of a collection of entities and relationships between them. Main elements of the ECORE notation include: *EClass* in order to define a specific class/entity, *EAttribute* and *EDataType* in order to specify the entity's attributes and their types, *EReference* in order to define relationships between different entities.

ECORE notation provides the same universal expressive abilities to represent both levels of DSL, that greatly simplifies the process of its development and allows to organize the creation of the DSL syntactic level as a transformation of the DSL semantic level, defined in the ECORE notation.

In addition, the ECORE model is not strongly connected with one fixed meta-model, thereby allowing the user to enter own components into the language and reuse previously created DSL models and the possibility of their flexible and coherent integration, that is the main goal of our research.

4 Application of the Proposed Approach to the Railway Domain

4.1 DSM of the Railway Allocation Domain

The railway allocation problem identifies the case, where the effectiveness and performance of the user strongly depends on the convenience and adaptability of DSL. The more flexible and native DSL syntax leads to the faster process of real time resource allocation. DSL allows, in particular, to allocate trains among railways, specify what services have to be provided to trains, what brigades perform these services, etc.

However, the concrete syntax can be defined only on the general semantic model of DSL. As a result, in order to increase the flexibility and create opportunities for customization of the DSL syntax for each specific user, we need to define a semantic model of the language as such accurate as possible. Consequently, it's vital to identify all the types of resources and operations on them in this domain.

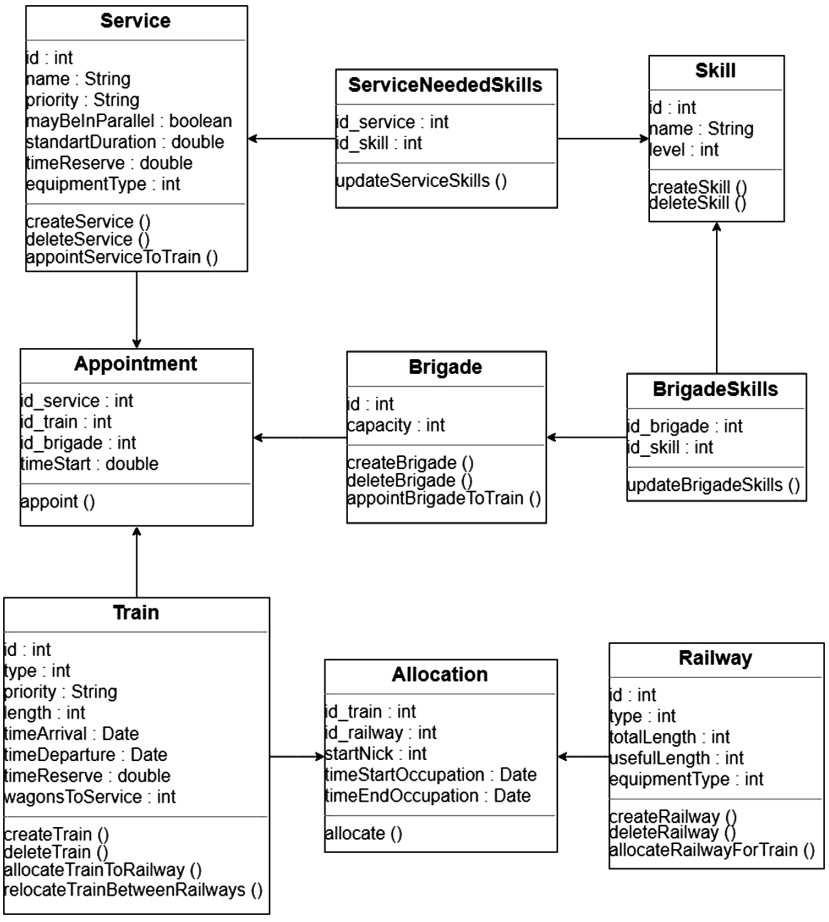


Fig. 4. The UML representation of DSM for the railway allocation domain

Developed DSM of the railway allocation domain is more complete in comparison with that considered in our previous work [15], since it contains the specification of the requirements (Skills) both for the Services and for the Brigades providing them. There are three general resources with own specific attributes for any railway station (Fig. 4): ***Trains***, ***Railways*** and ***Servicing brigades***. The more fundamental description and analysis of these resources is represented in our previous work [15], here we are interested only in the object-oriented structure of them.

In order to represent the relations between classes, two additional entities are defined: ***Allocation*** (to store the relations between the Train and the Railway) and ***Appointment*** (to store the relations between the Trains, the Service, needed for this Train, and the Brigade, which provide needed Service for the Train).

The behavior of all classes is defined using the methods of the corresponding classes. For example, the user can create a new Train, Brigade, Railway; appoint the Brigade for the created Train and allocate it on the created Railway. Other sceneries of the behavior can be realized through the corresponding methods.

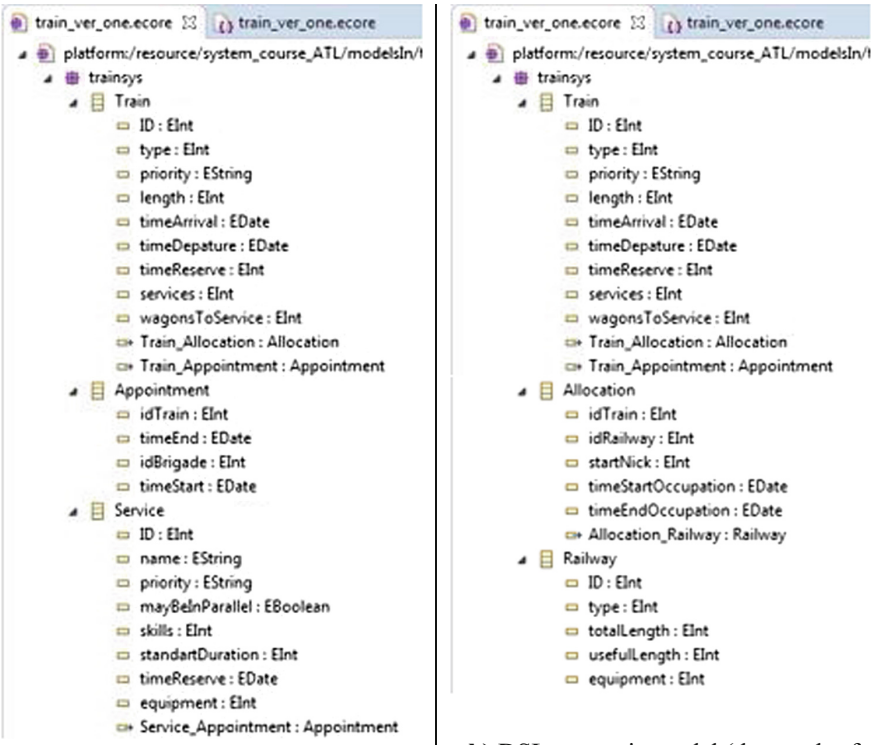
4.2 The DSL Semantic Level Definition

After the definition of DSM of the target domain, we can identify the semantic level of DSL. For this we need to reformulate the above created scheme in the UML notation on terms of the ECORE notation. In order to simplify this process, we use the graphical tool of Eclipse GMP, which translates the UML-description of DSM, into the ECORE model. As a result, the following ECORE representation of DSM is generated (the fragment is represented in Fig. 5a).

After this stage, our approach allows the users maintaining of DSL dialects based on the projections of DSM into DSL semantic models. In our particular case we need to create two DSLs dialects in response to the users' demands: the first one has to provide the opportunity to allocate Trains among Railways, while the second one - to appoint Brigades for Service delivery among Trains. For producing both DSL dialects, we should project only required set of DSM elements in order to achieve strong correspondence between the DSL structure and its purpose: Train, Railway and Allocation for the first DSL, and all DSM entities for the second DSL.

In order to implement corresponding semantic projections and realize needed DSL semantic models, the system of QVT transformation rules have to be defined. In more details the set of QVT and ATL rules, needed for different types of model transformations, is described in [15]. For the first DSL the rules (Fig. 6) remove all unnecessary entities from DSM, leaving only those necessary for the first DSL. As an output we have an ECORE-representation of the DSL semantic model (Fig. 5b). As we can see, this model contains a subset of entities from DSM necessary for specifying the corresponding DSL dialect: Train, Railway and Allocation.

In order to realize the projection for the second DSL semantic model, we have to use only the final mapping rule from the previously created (Fig. 6), because the DSL semantic model completely coincides with the DSM in this case (Fig. 5a).



a) DSM

b) DSL semantic model (the result of projection)

Fig. 5. The fragments of the key modeling artefacts (ECORE-representation)

```
modeltype ECORE;
transformation SemanticToSintactic (
   in i : "semanticDSLlvl.ecore", out o : "syntacticDSLlvl.ecore" );
rule killRedundEntities {
   from
      s : Ecore!ENamedElement (
        s.name <> 'Train' and s.name <> 'Railway' and
        s.name <> 'Allocation')
   to drop }
rule killRedundMethods {
   from
      s : Ecore!ENamedElement (
        s.rootObject().name <> 'Train' and
        s.rootObject()name <> 'Railway' and
        s.rootObject()name <> 'Allocation')
   to drop }
mapping EPackage :: copyPackage() : EPackage {
   name :=self.name; eClassifiers += self.eClassifiers[EClass]; }
```

Fig. 6. QVT M2M transformation rules

4.3 The DSL Syntactic Level Definition

After we have got the ready semantic model of DSL, we can define its syntactic level. In fact, this means that we need to describe a number of aliases for accessing specific functions on the semantic level of DSL. In our case end-user DSLs have a graphical form. Therefore, the aliases have a form of graphical pictograms. Each pictogram represents the corresponding entity of the DSL semantic model. The line connectors between the pair of pictograms represent the semantic connection between corresponding domain entities.

In order to achieve our goal, the Eclipse XTend plugin (http://www.eclipse.org/xtend/) can be used. This plugin allows us to identify the pictograms above ECORE-model entities and provides the opportunity to automatically create the form for the definition of the corresponding entity attributes. Such automated form creation is available only for the attributes, which type is internal for the ECORE-metamodel. Custom types are not supported and require manual implementation of form fields. In our case all attributes are standard and are implemented in ECORE. That means that automatic generation of the DSL syntax is available for both DSL dialects. The examples of the created interface for both DSLs are represented in Figs. 7 and 8 respectively.

Created interfaces are absolutely coherent with the corresponding DSL semantic models and allow to organize the corresponding DSL dialects scenarios: allocate Trains among Railways in the first dialect and appoint Brigades for providing Services for Trains in the second dialect.

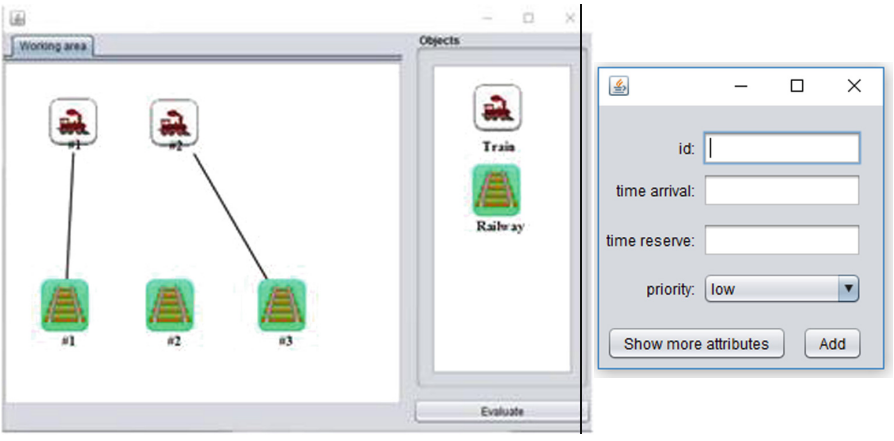


Fig. 7. The interface and forms prototypes for the first DSL dialect

What is more important, both interfaces are based on the same more general DSM. That results in the opportunity to update and extend them without the need to re-create the whole DSL scheme. Only a modification of the QVT rule and the interface concretization are needed for this operation and can be provided by the end-users.

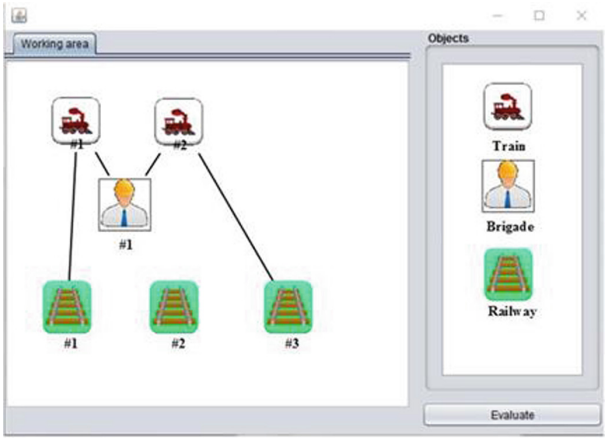


Fig. 8. The interface prototype for the second DSL dialect

Comparing the proposed approach of DSL model-oriented development with other approaches in terms of time-consuming, we can state that our approach significantly accelerates the process of DSL development. For example, in our previous work [15] a full cycle of DSL development took about two months, while using the proposed approach this time was reduced to one week. In fact, during this time frame, two different DSLs were developed against one in the previous case. The main part of time was spent on building the original full DSM, while all subsequent operations took several hours, because they were organized in (semi)automatic manner. Such a significant reduction in the time to develop DSL shows the effectiveness of the proposed approach and the possibility of its use in the industrial development of customizable DSLs.

5 Conclusion

In our research we explored a model-oriented and projection-based approach for DSL development. Proposed approach is based on the idea, that every DSL contains two parts: semantic and the syntactic. Both can be represented as a set of interconnected objects, characterized by specific attributes. Such object-oriented description of DSL levels results in the opportunity to define the DSL as a model-oriented structure, where a certain entity is assigned to each element.

What is the most important, the DSL structure is created automatically from the specific DSM using the so-called semantic projection mechanism. The semantic projection is an operation, which is conducted over the DSM and the result of which is also a semantic model that is obtained by transforming the original DSM. The result of projection describes the DSL semantic model.

In contrast to the superset modelling language [8], DSM in this case is not only the formal representation for the domain's terms and relationships between them, but also reflects users' knowledge and semantical scheme of the domain, allowing to take it into account during the DSL creation.

In comparison with the existing approaches to DSL development (like [3, 9, 10]), which use a traditional cycle of DSL development, starting from the definition of the DSL concrete syntax, our approach starts with the generation of DSM, which is a dynamic, time-varying structure. Under these circumstances, the DSL semantic model can be obtained as a projection of such DSM through M2M transformations. Furthermore, the DSL syntactic model is also the result of the projection of the DSL semantic model onto users' requirements and needs.

As a result, the proposed DSL development process is conducted in full accordance with the conceptual scheme of the target domain, thereby ensuring the participation of end-users in the process of its creation. In addition, the projection-based principle of the DSL development allows the users to achieve the resilience of the DSL created both with the target domain (represented by DSM) and users' requirements. Created DSL can be transformed on the semantic and syntactic levels separately, using M2M transformations for projections realizations. At the same time, the consistency of the created DSL dialects is preserved.

Among advantages of the approach proposed its reusability and end-user orientation should be mentioned. The approach can be transferred to any domain for which the DSM is defined. The model-oriented structure of DSL is also an understandable and convenient for end-user. This format of DSL representation results in the opportunity to make changes to DSL without special programming skills. The defining a specific syntax for different versions of DSL is maximally automated and is carried out at the level of models, rather than in the program code. Such automation significantly reduces the time required to complete the DSL development cycle, that was demonstrated by the examples.

Planning further research, we consider to organize the whole model-driven DSL life-cycle with opportunity to define the more complex behavior for the user through the combination of different syntactic DSL elements. One more possible improvement can be the implementation of versification mechanism not only at the level of syntax, but also at the level of semantics. For achieving this goal an approach, described by Bell, can be used. In [11] Bell tries to implement the coherent changes in several DSL meta-models, using the automated transformations, that can allow to organize the consistent changes in the semantic and syntactic DSL levels models within our approach.

References

1. Fowler, M.: Domain Specific Languages. Addison Wesley, Boston (2010)
2. Laird, P., Barrett, S.: Towards dynamic evolution of domain specific languages. In: van den Brand, M., Gašević, D., Gray, J. (eds.) SLE 2009. LNCS, vol. 5969, pp. 144–153. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12107-4_11

3. Cleenewerck, T.: Component-based DSL development. In: Pfenning, F., Smaragdakis, Y. (eds.) GPCE 2003. LNCS, vol. 2830, pp. 245–264. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-39815-8_15
4. Challenger, M., Demirkol, S., Getir, S., Mernik, M., Kardas, G., Kosar, T.: On the use of a domain-specific modeling language in the development of multiagent systems. In: Engineering Applications of Artificial Intelligence, pp. 111–141 (2014)
5. Parr, T.: Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages. Pragmatic Bookshelf (2012)
6. QVT (Query/View/Transformation). https://projects.eclipse.org/projects/modeling.mmt.qvt-oml
7. Elipse Graphical Modeling Project (GMP). http://www.eclipse.org/modeling/gmp/
8. Lucassen, G., Robeer, M., Dalpiaz, F., Werf, G.M., Brinkkemper, S.: Extracting conceptual models from user stories with Visual Narrator. Requir. Eng. **22**(3), 339–358 (2017)
9. Sprinkle, J.: A domain-specific visual language for domain model evolution. J. Vis. Lang. Comput. **15**, 291–307 (2004)
10. Kosar, T., Martınez Lopez, P., Barrientos, P., Mernik, M.: A preliminary study on various implementation approaches of domain-specific language. In: Information and Software Technology, pp. 390–405. Elsevier (2008)
11. Bell, P.: Automated transformation of statements within evolving domain specific languages. In: Computer Science and Information System Reports, pp. 172–177 (2007)
12. Haav, H.-M., Ojamaa, A., Grigorenko, P., Kotkas, V.: Ontology-based integration of software artefacts for DSL development. In: Ciuciu, I., et al. (eds.) OTM 2015. LNCS, vol. 9416, pp. 309–318. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26138-6_34
13. Guizzardi, G.: Ontological foundations for structural conceptual models. Telematica Instituut Fundamental Research Series, The Netherlands, no. 15 (2005). ISBN 90-75176-81-3
14. Ruffolo, M., Sidhu, I., Guadagno, L.: Semantic enterprise technologies. In: Proceedings of the First International Conference on Industrial Results of Semantic Technologies, vol. 293, pp. 70–84 (2007)
15. Ulitin, B., Babkin, E.: Ontology and DSL co-evolution using graph transformations methods. In: Johansson, B., Møller, C., Chaudhuri, A., Sudzina, F. (eds.) BIR 2017. LNBIP, vol. 295, pp. 233–247. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-64930-6_17

Utilizing Customer Feedback for Business Process Performance Analysis

Amina Mustansir, Khurram Shahzad(✉), Syed Irtaza Muzaffar, and Kamran Malik

Punjab University College of Information Technology, University of the Punjab, Lahore, Pakistan
{amina.mustansir, khurram, syed.irtaza, kamran.malik}
@pucit.edu.pk

Abstract. There are numerous organizations that limit themselves to the modeling of business processes and do not automate and execute business processes using a Business Process Management System (BPMS). Consequently, these organizations cannot exploit the potential of process execution logs for the performance analysis of their business processes. In this study, we contend that customer feedback in natural language is a rich source of information that can provide valuable indications about process performance. However, as the number of comments increases, the time and effort required to synthesize them and elicit indications about process performance increases exponentially. To that end, in this paper we propose an approach that takes customer feedback as input and provides valuable indications about process performance. Specifically, the proposed approach is composed of three types of classification that leverage from a supervised learning technique to automatically classify customer feedback. The classified comments are subsequently used to generate answers to performance related questions. Evaluation of the proposed approach is conducted by using a real-life case study of an academic institute. For the case study we collected 4366 student comments about the admission process of the institute and manually classified the comments for each type of classification. Subsequently, these classified comments are used as a benchmark for the evaluation of our proposed approach. The results demonstrate the effectiveness of the proposed approach.

Keywords: Business Process Management · Process performance analysis
Devil's Quadrangle · Supervised learning techniques · Customer feedback

1 Introduction

Business Process Management (BPM) lifecycle comprises of modeling business processes, implementing the modeled processes, executing them, and monitoring their performance [1]. During the last decade, growing number of organizations have embraced all the phases of business process management lifecycle [2]. However, presently there are numerous organizations that limit themselves to business process modeling i.e. these organizations do not implement and execute their business processes. The organizations that embrace all phases of the BPM lifecycle have vast

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 235–249, 2018.
https://doi.org/10.1007/978-3-319-99951-7_16

opportunities for performance analysis of their processes using Activity Monitoring [3, 4], Process Warehouse [5], and Process Mining [6], etc. On the contrary, the organizations that limit themselves to business process modeling do not have access to the execution logs of their processes. Consequently, these organizations cannot exploit the potential of execution logs, generated during process execution, for the performance analysis of business processes [7]. These deprived organizations desire to get a sense of their process performance by utilizing the continuously growing amount of business data, such as customer feedback.

Customer feedback on a business process is a rich source of information that can provide valuable insights about process performance. Typically, customer feedback in natural language text includes the feeling and/or opinion of users regarding an idea, product, business or service [10]. It is widely established that the effective utilization of customer feedback may yield several benefits for organizations, which include introducing new products or services, improving the quality of existing products or services, avoiding unwanted features, emphasizing the trending features, and increasing sales [11–15]. Furthermore, Kujala et al. [16], has conducted a more extensive study to establish that the effective utilization of customer feedback can increase sales by 25%. The study has also discovered that the cost-benefit ratio of utilizing customer feedback is 1:100 and in the context of automation the effective utilization of customer feedback can lead up to 90% change in a software.

In this study, we propose that organizations can utilize customer feedback in natural language to generate valuable indications about the process performance. However, a central problem to such an approach is that the presence of a large number of comments requires substantial time and effort to synthesize the comments and elicit indications about process performance. More concretely, it is not practical for a process owner to read a couple of thousand comments and elicit some indications about process performance. To address that problem, in this paper, we propose an approach that takes customer feedback in natural language as input and returns valuable indications about process performance. These indicates are useful in several ways, such as getting a sense of process performance, identifying the performance dimensions that should essentially be considered during the subsequent process redesign attempt, and choosing snippets from customer feedback for identifying areas of improvement.

In essence, the approach is composed of three types of classification (Performance Relevance, Performance Dimension, and Sentiment Analysis) and aggregation. The aim of the first type of classification is to segregate process performance comments for further use. The second type of classification ascertains the process performance dimension (time, cost, quality and flexibility) of each comment. The third type of classification determines the polarity (positive, negative or neutral) of each comment. These three types of classifications leverage from supervised learning techniques to automatically classify the comments. Finally, aggregate functions, such as count and average, are applied on top of the classified comments to generate indications about process performance. Below, we provide an example that is used to illustrate the proposed approach.

Figure 1 shows an excerpt version of a university admission process. It can be observed from the figure that the process starts when a student visits the institute to collect an application form. The institute generates an application form for each student

that includes a randomly generated unique number, formally called application number. Each student is required to complete the application form, deposit the fees to a bank account, and submit the form. The physical presence of a student is not mandatory for submitting the completed application form. That is, the completed application form can also be sent by postal mail.

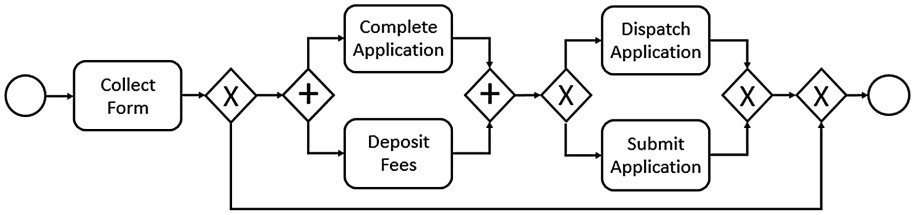


Fig. 1. University admission process, an excerpt version.

Table 1 contains an excerpt of student feedback about the admission process. The comments listed in the table are extracted from the actual feedback provided by the students during the case study presented in Sect. 3. Utilizing the given feedback to generate indications about process performance requires synthesis of the comments presented in the table. For instance, it can be observed from the table that the comment C1 and C2 are related to the behavior of the admission staff, C3 and C5 are related to the time consumed in performing a task, and C4 and C6 are related to the availability of an alternate for a task. The comments can also be synthesized based on the feelings of the customers. In this case, C1 and C6 indicates satisfaction of students, whereas C2, C3, and C7 indicates dissatisfaction of students. However, the dissatisfaction in C7 is related to the limited knowledge of taxi driver, i.e. not related to the admission process of the university.

Table 1. Example comments about the university admission process.

ID	Example comments
C1.	The form was easily available and the staff was very friendly
C2.	The staff did not guide me properly about completing the form
C3.	I had to stand for hours in hot weather to deposit fees
C4.	In this digital era, the form was not available online
C5.	It took me an hour to fill the application form
C6.	Glad that there was an option to send the application form by courier
C7.	The taxi driver did not know where the campus was
C8.	The application form has seven pages

The time and effort required to synthesize these small number of comments and elicit indications about process performance are considerable due to the diversity of comments and the synthesis possibilities discussed above. The diversity in the

comments as well as the synthesis possibilities may increase further as the number of comments increase. Consequently, it is not practical for a process owner to read some thousand comments and elicit some indications about process performance. To that end, in this paper we propose an approach that takes input customer feedback in natural language and elicit indications about process performance.

The rest of the paper is organized as follows: Sect. 2 presents our proposed approach for utilizing customer feedback to elicit performance indications about the performance of a process. Section 3 introduces the real-world case that we have used for the evaluation of the proposed approach. Section 4 presents the evaluation of the proposed approach. Finally, Sect. 5 concludes the paper.

2 The Proposed Approach

In this section, we present our proposed approach for utilizing customer feedback to get indications about process performance. In essence, the approach is composed of three types of classification and subsequent aggregation of the classified feedback, to generate valuable indications about process performance. For the automatic classification we rely on Support Vector Machine (SVM) – an established machine learning algorithm. The choice of the algorithm stems from *literature* as well as from our *experimentation*. Regarding *literature*, a recent study has highlighted the usefulness of SVM in business process management domain, due to its ability to deal with small datasets as well as low risk of overfitting and scalability [18]. Regarding *experimentation*, during the evaluation of the proposed approach we performed experiments using SVM, Naïve Bayes and Random Forest. The results indicated that SVM is the most suitable choice.

2.1 Overview of the Approach

The overall approach is shown in Fig. 2. It can be seen in the figure that the first three steps classify the comments while the fourth step aggregates the classified comments to provide answers to performance questions. A brief overview of the proposed approach is as follows:

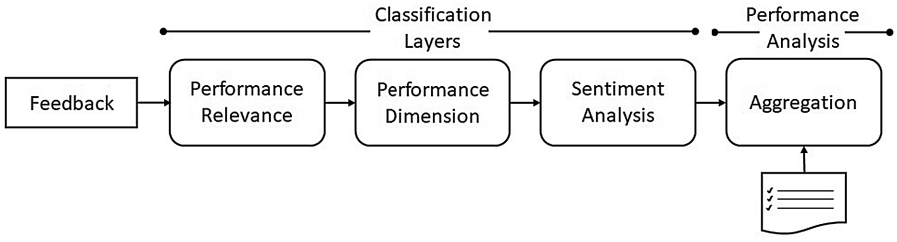


Fig. 2. An overview of the proposed approach

Performance Relevance. The purpose of the first type of classification is to distinguish between the comments that are about the execution of the process, from the

comments that are not about the execution of the process. It is necessary to distinguish between relevant and irrelevant comments because any indications obtained by processing the entire collection of comments, without omitting irrelevant comments, may be misleading [7]. For instance, in the illustrative example, C5 (it took me an hour to fill the application form) is clearly about the process performance because it refers to the time consumed to fill the form. On the contrary, C7 (the taxi driver did not know where the campus was) is related to the location of campus rather than execution of the admission process. The further utilization of C7 comment can skew the aggregation score and therefore it should be excluded at this staged. Similarly, C8 (the application form has seven pages) may or may not be considered as a performance related comment. That is, in case the interpretation of the comment is that, the application form was lengthy and it consumed substantial human effort to fill the application form, the comment is related to process performance. On the contrary, the comment can also be considered as a comment about application form rather than the act of filling the form. In this case, the comment can also be omitted at this step. These examples indicate that prior to any further processing of customer feedback, it is desirable to first recognize the performance related comments and omit the comments that are not related to the process performance.

Performance Dimension. The aim of the second type of classification is to determine what performance aspect of the process is discussed in a comment. For that, we rely on Devil's Quadrangle Framework (DQF), a useful conceptual framework for the process performance analysis [8, 9]. The reasons for choosing DQF are two-folds, it is widely acknowledged as an *ideal* framework for the performance analysis of processes [7, 8], and the framework offers a pertinent list of the performance dimensions that should be considered for process performance analysis [19]. The framework distinguishes between four main performance dimensions, Time, Cost, Quality and Flexibility [9]. Time dimension represents the robustness of process execution. The lower score of this dimension is desired because a higher value of this dimension represents undesired delay in process execution. Cost dimension represents the human effort required for the execution of the process or revenue consumed during the execution. Both from the organizational as well as the customer perspective, the lower value of this dimension is desired because a higher value of cost represents expensive execution of the process. Quality dimension from the customer perspective refers to the satisfaction with the specification or characteristics of the process. Higher value of this dimension is desired because it represents customers' satisfaction. Typically, flexibility is defined from organizational perspective, and it refers to the ability of a process to react to changes induced by the organization. In this study, we define flexibility from customer perspective, and it refers to the availability of alternates to the customer for performing a task, such as multiple ways of depositing fees and alternates to submit application form. A higher value of flexibility may be desired from customers' perspective to facilitate customers.

To illustrate this step of the approach, Table 2 contains six comments that are related to process performance and the performance dimension of each comment. For instance, C1 is about customers satisfaction with the service delivered by the staff and

therefore should be classified as quality related comment. C3 is regarding the waiting time it is thus related to the time dimension.

Table 2. Classification and sentiment analysis of relevant comments

ID	Performance relevant comments	Dimension	Sentiment
C1.	The form was easily available and the staff was very friendly	Quality	Positive
C2.	The staff did not guide me properly about completing the form	Quality	Negative
C3.	I had to stand for hours in hot weather to deposit fees	Time	Negative
C4.	In this digital era, the form was not available online	Flexibility	Negative
C5.	It took me an hour to fill the application form	Time	Neutral
C6.	Glad that there was an option to send the application form by courier	Flexibility	Positive

Sentiment Analysis. The aim of the third type of classification is to classify the customer feedback based on the feelings expressed in the comment. For that, we rely on the widely used four types of sentiments, Positive, Negative, Neutral, and Both. A comment that indicates satisfaction of customer with the execution of the process is a positive comment. For instance, C1 (The form was easily available and the staff was very friendly) is a positive comment because it indicates the satisfaction of customers about the friendly behavior of staff. On the contrary, a comment that indicates dissatisfaction of customer with the execution of the process is a negative comment. For instance, C2 (The staff did not guide me properly about completing the form) is a negative comment because it indicates the dissatisfaction of customer with the guidance provided by the university staff. Another possibility is that the comment neither expresses satisfaction nor dissatisfaction with process execution. Such a comment is a neutral comment. C5 (It took me an hour to fill the application form.) is an example of a neutral comment because the expression in the comment does not explicitly represent a positive or a negative feeling. Another possibility is that a single comment may contain both positive and negative expression. For instance, consider the comment 'The application form was expensive but the staff was very cooperative'. The comment contains a positive sentiment about the staff and a negative comment about the cost of the application form.

Aggregation. The last step of the approach relies on aggregate functions, such as count, or average, to quantify the classified comments and generate indications about process performance. For that, we have proposed three categories of questions, that can provide valuable indications about process performance. Furthermore, for each category, four concrete questions are presented in Table 3. We contend that the list offers a critical starting point for performance analysis and it can be extended on demand. The three categories of questions are *basic*, *advance* and *profound*. The *basic* questions provide fundamental information about the performance of a process, based on the customer feedback. The questions like, how often customers comment about the

robustness of the process, are included in this category. The *advance* questions provide contextual information about the process to precisely interpret the answers of the basic questions. For instance, given that there is no restriction on the number of comments by a customer, it would be valuable to know how many customers comment about the robustness of the process, in addition to the number of comments. Lastly, the *profound* questions provide enriched information about the basic questions. The questions like, how many comments are positive or how many comments are negative, are included in this category.

Table 3. Categories and concrete questions about process performance.

ID	Basic questions (fundamental information)
Q1.	How often customers comment about robustness of the process?
Q2.	How often customers comment about flexibility of the process?
Q3.	How often customers comment about quality of the process?
Q4.	How often customers comment about the cost incurred in the process execution?
	Advance questions (contextual information)
Q5.	How many customers comment about robustness of the process?
Q6.	How many customers comment about flexibility of the process?
Q7.	How many customers comment about quality of the process?
Q8.	How many customers comment about the cost incurred in the process execution?
	Profound questions (enriched information)
Q9.	Does customers talk positively or negatively about robustness?
Q10.	Does customers talk positively or negatively about flexibility?
Q11.	Does customers talk positively or negatively about quality?
Q12.	Does customers talk positively or negatively about execution cost?

For the illustrative example, the answer to Q1 can be generated by counting the number of comments about time dimension in Table 2, i.e. there are two comments about robustness of the process. Similarly, there are two comments about quality as well as about flexibility. However, there is no comment about execution cost of the process. Similarly, answer to the advanced questions can be generated by counting the number of students. Furthermore, the number of positive and negative comments can be computed from Table 2. These numbers indicate that students have not expressed any concern about the cost, but they are concerned about time. Therefore, the reduction in waiting time should be kept in focus during process redesign.

3 The Case of an Academic Institute

In this section, we introduce the real-world case study that we have used for the evaluation of the proposed approach. Specifically, Sect. 3.1 introduces the case of an academic institute as well as the business process about which the feedback is collected. Section 3.2 explains the procedure that we have used to collect the feedback and the specification of the collected feedback is presented in Sect. 3.3.

3.1 Description of the Case

A public-sector institute offers several degree programs in which admissions are offered once a year. Due to the reputation of the institute, over 10,000 students apply for admission to its various degree programs. All these students are required to visit the campus during a short span of two weeks to collect the admission form for the entry test, and return the completed forms. In addition to completing the form, the applicants are required to deposit fee for the entry test to a designated bank. The institute conducts an admission test for the applicants and based on the results of the admission test the best performing students are selected.

Every year the administration receives multiple complaints and suggestions to improve the admission process. Responding to the requests, the administration desires to develop a mechanism for assessing the performance of its admission process from the students' perspective. For that, we formed a team of researchers and BPM experts to model the admission process model. During the last year, the team interviewed several stakeholders of the admission process, including members of admission committee and institutional staff, to design the admission process. This exercise was necessary to understand the exact scope and specification of the admission process. However, due to space limitation the complete model is not presented in the paper.

3.2 Student Feedback Collection

As a first step, we conducted a pilot study with 60 volunteer students of the college. For the study, a brief description was provided to the students to explain the objectives of the study and they were asked to provide their feedback about the admission process. Subsequently, we analyzed the feedback and used it to refine the study design. The feedback collection form included a brief introduction to the project, instruction related to the use of English language in feedback, and the motivation for collecting the data. In addition to that, we added few feedback provoking questions/statements as well as example answers. The examples of these are as follows: Share your experience about admission process. If you are not satisfied with any task related to the admission process, please specify it. You can help us improve each step of the admission process by sharing your feedback. This decision of including examples stems from the problems faced during the pilot study. For collecting the data, a Google form was shared and students were given six weeks to collect their thoughts and share their experiences of students. To generate a large collection of comments, students were encouraged to be exhaustive in providing feedback.

3.3 Specification of the Feedback

Table 4 shows the specification of the two studies, the pilot study and the complete study. From the table it can be seen, 60 students participated in the pilot study and more than 100 comments were collected. However, these comments were not included in the final collection of comments to avoid any bias in the data.

It can also be seen from the table that in the complete study 117 students provided feedback on the admission process. For this study, each sentence in the student

Table 4. Specification of the collected feedback.

	Pilot study	Complete study
Number of students	60	117
Data collection timeline	Spring 2017	Fall 2017
Study duration	1 day	6 weeks
Number of comments collected	>100	4366

feedback is treated as a separate comment. The decision of considering each sentence as a separate comment stems from the manual analysis of randomly selected feedback of 30 students. During the analysis it was observed that, (a) students expressed their feeling about a certain part of the process in one sentence, (b) a large majority of the co-located sentences express their feeling about different parts of a process, and (c) there are sentences that contain more than one expressions. Examples of the three types of comments are presented in Table 5. In total, 4366 comments were collected from 117 students making the average number of comments per student 37. This indicates that the students followed the guidelines and they were thorough in their feedback.

Table 5. Examples of the three types of comments.

Type of comment	Comments
Single sentence	I had to stand for hours in hot weather to deposit fees
Co-located	It took me an hour to complete the form. The staff was not willing to help. Why do I have to deposit the fee in THAT bank
Multi-expression	The form was easily available and the staff was very friendly

4 Evaluation of the Proposed Approach

In this section, we use the collected feedback to evaluate the effectiveness of our proposed approach. Specifically, we present the systematic procedure that is used to prepare the dataset for training and testing purpose, for each type of classification. Subsequently, the details of the *Performance Relevance* and *Performance Dimension* and *Sentiment Analysis* experimentations are presented.

4.1 Data Preparation for Automatic Classification

To prepare the data for SVM, we first developed three criteria for the classification of comments. For generating the criteria, we randomly collected a pool of 100 comments and asked two researchers to review the comments independently. Subsequently, the researchers were asked to generate initial thoughts regarding *Performance Relevance*, *Performance Dimension* and *Sentiment Analysis* criteria. For that, researchers were asked to carefully review the definitions of performance dimensions and polarity of sentiment in literature, including books [8, 9], and research articles [17]. This was

followed by a synthesis of criteria as well as multiple discussion sessions to generate mutually agreed candidate criteria.

To further refine the candidate criteria, two researchers independently classified a quarter of randomly selected comments, i.e. nearly 1000 comments. The generated results were compared to identify the discrepancies. The classification results revealed higher agreement between the two researchers (greater than 86%). This higher agreement indicates that the criteria contains sufficient information for classification of comments at all three levels. Keeping in view the higher percentage of agreement, the remaining 3366 comments were classified by a single researcher. The process used to develop the classification criteria is shown in Fig. 3.

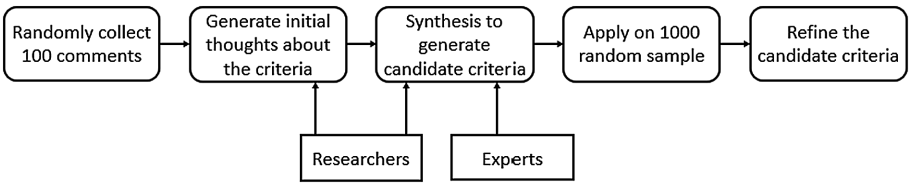


Fig. 3. The process of generating the criteria.

Generating the Benchmark Dataset. As discussed above, one researcher parsed all the remaining comments to identify performance related comments and further classifying them into performance dimensions. Accordingly, the generated dataset includes 3450 comments about process performance and the 916 comments that are not related to process performance. Out of the 3450 performance related comments, 571 comments are about time, 619 comments are about flexibility, 2097 comments about quality, and 249 comments about cost. For the experimentation, we use 3536 comments because some multi-expression comments were classified into more than one classes, so their representation in both classes was created by duplicating the comments. Furthermore, the dataset includes, 1421 Positive comments, 1269 Negative comments, 45 comments that includes both (positive and negative) expressions, and 801 comments that does not include a positive or negative expression. The detailed specification of the collected feedback is presented below in Table 6. These classified comments are used as input for training and testing of SVM.

Table 6. Specification of the collected feedback.

	Relevant	Positive	Negative	Both	Others
Time	571	119	365	21	66
Flexibility	619	152	76	5	386
Quality	249	1089	707	16	285
Cost	2097	61	121	3	64
Total	3536	1421	1269	45	801

4.2 Conducting the Classification Experiments

To evaluate the effectiveness of supervised learning techniques, we performed three experiments using SVM. The input to each classification task is a 1-gram (hereafter, unigram) feature matrix in which the columns represent a set of (unique) words in all the comments and rows represent IDs of the comments. Thus, each cell of the table contains a binary value 1 or 0, where 1 represents that a word in the column is present in the comment and 0 represents that the word is not present in the comment. For each technique, we have used training and testing ratio of 65:35 and we have performed 10-fold cross-validation to reduce the bias that may arise due to sampling. The results presented in the subsequent section are average scores of the 10 folds. Furthermore, we have repeated the experiments using all possible combinations of preprocessing (punctuation removal, stop words removal and stemming), and the best results are presented in the subsequent section. Similarly, we have performed experiments using 2-gram (hereafter, bigram) feature matrix and 3-gram (hereafter, trigram) feature matrix.

4.3 Results of the Classification Experimentations

Table 7 shows the average precision, recall and F_1 scores of each classification experimentation. From the table it can be observed that our proposed approach shows promising results for all the classification tasks i.e. 0.89 for performance relevance, 0.70 for performance dimension, and 0.70 for sentiment analysis. However, due to space limitations, the discussion of the results is limited to two key observations, effective feature for learning and performance variation between tasks.

Effective Feature for Learning: From Table 7 it can be observed that unigram is the most discriminating feature for recognition as well as for classification. This observation is true for all the levels of classification. It can also be observed from the figure that as the value of n in word n gram increases from 1 to 3, the performance of SVM for all classification tasks decreases. It is because, in case of unigram, the number of words used as features are significantly less than that of bigram and trigram. As a result, the size of the unigram feature matrix is significantly smaller than the size of bigram and trigram feature matrices. The large size of matrix limits the learning capability of the supervised learning techniques and hence impedes its performance.

Performance Variation Between Tasks: From the table it can also be observed that SVM is more effective for identifying performance related comments than for the performance dimension identification and sentiment analysis. A key reason is that the size of the training dataset used for the recognition task is significantly larger than the training dataset used for the other two tasks. For instance, for the performance recognition task, 2319 (65% comments) performance relevant comments and 519 (65% comments) comments not relevant to performance, are used. In contrast, for the performance dimension task, 396, 161, 1,359 and 403 comments about time, cost, quality and flexibility, respectively, are used for training. This smaller size of the classification training datasets, limits the learning abilities of the supervised learning technique, and hence impedes the performance of the technique.

Table 7. Potential questions about process performance that can be answered.

Technique	Features	Precision	Recall	F1 Score
Performance relevance	Unigram	0.89	0.89	0.89
	Bigram	0.85	0.77	0.79
	Trigram	0.71	0.71	0.71
Performance classification	Unigram	0.70	0.70	0.70
	Bigram	0.65	0.66	0.65
	Trigram	0.56	0.60	0.55
Sentiment analysis	Unigram	0.70	0.70	0.70
	Bigram	0.68	0.68	0.68
	Trigram	0.63	0.57	0.54

4.4 Results of Performance Analysis

The fourth step of the proposed approach concerns the analysis of the classified comments to generate indications about process performance. To that end, we proposed three categories of questions (basic, advanced and profound) that can provide indications about the process performance. Furthermore, for each category of questions we presented four concrete questions that needs to be answered for generating indications about process performance. In this section, we present the answers of the concrete questions that are generated from classified comments by using aggregate functions.

Table 8 presents answers of the basic and advanced questions for our case study. From the table it can be observed that a large majority of the relevant comments (2,097 out of 3,536) are about the quality of the process and merely 249 comments are about the cost incurred during the admission process. These numbers may give the impression that the applicants are mostly concerned about the execution quality and rarely about the cost of the process. Furthermore, there is no significant difference between the number of comments about the flexibility and the robustness of the process, that are 619 and 571, respectively.

The answers of the advanced questions, presented in Table 9, offers valuable contextual information to interpret the answers of the basic questions. For instance, it can be observed from the table that the number of applicants commenting about quality are comparable with the ones commenting about robustness and flexibility of the process ($113 \approx 107 \approx 109$). These results are contrary to the impression generated by the results of the basic questions. That is, undoubtedly, the number of comments about quality are significantly greater than that of robustness, cost and flexibility. However, it does not indicate that the number of applicants interested in robustness and flexibility of the process are less than that of quality.

To further obtain indications about the process performance, we have also generated the answers of the profound questions (see Table 9). In the table, each question has three corresponding values representing the number of positive comments, negative comments and other comments – neither positive nor negative. From the table, it can be observed that the number of negative comments about robustness and cost of the process are significantly higher than positive comments ($386 > 140$ and $124 > 64$).

In contrast to that, the number of positive comments about flexibility and quality are higher than the negative comments (157 > 81 and 1105 > 723).

Table 8. Answers of process performance analysis, basis and advanced questions.

Basic questions (fundamental information)	Count
How often applicants comment about the robustness of the process?	571
How often applicants comment about the flexibility of the process?	619
How often applicants comment about the quality of the process?	2097
How often applicants comment about the cost incurred by the process?	249
Advanced questions (contextual information)	
How many applicants comment about the robustness of the process?	107
How many applicants comment about the flexibility of the process?	109
How many applicants comment about the quality of the process?	113
How many applicants comment about the cost incurred by the process?	89

Table 9. Answers of process performance analysis, profound questions.

Profound questions (enriched information)	Positive	Negative	Others
Does applicants talk positively or negatively about the robustness of the process?	140	386	66
Does applicants think we are rigid in our process?	157	81	386
Does applicants talk positively or negatively about the quality of the process?	1105	723	285
Does applicants talk positively or negatively about the execution cost?	64	124	64

To conclude, for the admission process, a significant percentage of comments are about the execution quality of the process, and nearly every applicant commented about quality. Furthermore, majority of these quality comments are positive. In contrast to that, there are fewer comments about the cost, and a significant number of applicants comment about cost. However, most of the cost related comments are negative. These changing impressions across the three types of questions indicate that for an accurate interpretation of results, the answers of all three types of questions are necessary. Hence, endorsing the suitability of our approach.

5 Conclusions and Limitations

Numerous organizations limit themselves to modeling their processes, that is they do not automate and implement their processes in BPMS. These organizations desire to gain indications about their process performance, however no work has been done to the benefit of these organizations. To that end, in this work we have proposed an approach that takes input customers feedback in natural language and provides valuable

indications about process performance. The proposed approach employs a supervised learning technique, Support Vector Machine, to classify customer feedback at three levels, Performance Relevance, Performance Dimension, and Sentiment Analysis. The classified comments are aggregated to generate answers to several questions, hence, providing key indications about process performance. For the evaluation of the propose approach, we have conducted a real-life case study of an academic institute by firstly collected 4366 comments about the admission process. Secondly, we have manually classified these comments at three levels to generate a human benchmark. Thirdly, we have used Support Vector Machine to automatically classify the comments at three levels. Finally, we demonstrate that the classified comments can answer basic, advanced as well as profound questions about process performance. In the future, we plan to increase the size of the dataset and to make it publicly available for advancement of research in this novel direction.

We contend that the proposed approach has highlighted the need for developing tools and technologies to equip the organizations that do not automate their processes using BPMS. The underlying mechanism of the proposed approach can be enhanced in several ways. For instance, instead of conducting the survey, the social media comments can be scrapped, cleaned, and subsequently used for process performance analysis. Furthermore, the three categories of questions as well as the list of questions can be extended to fulfill the requirements of process performance analysis. Finally, we recognize that a more comprehensive evaluation of the proposed approach is required to evaluate the usefulness of the proposed solution. Furthermore, several questions, such as what value does the proposed approach create for business, and how effective is the approach, are yet to be answered. In the future, we plan to answer these questions and use the results of the analysis for business process redesign.

References

1. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.: Fundamentals of Business Process Management, 1st edn. Springer, Holland (2013). https://doi.org/10.1007/978-3-662-56509-4
2. Chang, J.F.: Business Process Management Systems: Strategy and Implementation. CRC Press, Boca Raton (2016)
3. Jalali, A., Johannesson, P.: Multi-perspective business process monitoring. In: Nurcan, S., et al. (eds.) BPMDS/EMMSAD 2013. LNBIP, vol. 147, pp. 199–213. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38484-4_15
4. Vera-Baquero, A., Colomo-Palacios, R., Molloy, O.: Real-time business activity monitoring and analysis of process performance on big-data domains. Telemat. Inform. **3**(33), 793–807 (2016)
5. Casati, F., Castellanos, M., Dayal, U., Salazar, N.: A generic solution for warehousing business process data. In: Proceedings of the 33rd International Conference on Very Large Data Bases, pp. 1128–1137 (2007)
6. van der Aalst, W.M.P.: Process Mining: Data Science in Action, 2nd edn. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4
7. Ahmad, S., Muzaffar, S.I., Shahzad, K., Malik, K.: Using BPM frameworks for identifying customer feedback about process performance. In: Matulevičius, R., Dijkman, R. (eds.) CAiSE 2018. LNBIP, vol. 316, pp. 55–69. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92898-2_5

8. Dumas, M., van der Aalst, W.M.P., Hofstede, A.H.T.: Process-Aware Information Systems: Bridging People and Software Through Process Technology, 1st edn. Wiley, New York (2005)
9. Kis, I., Bachhofner, S., Di Ciccio, C., Mendling, J.: Towards a data-driven framework for measuring process performance. In: Reinhartz-Berger, I., Gulden, J., Nurcan, S., Guédria, W., Bera, P. (eds.) BPMDS/EMMSAD 2017. LNBIP, vol. 287, pp. 3–18. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59466-8_1
10. Naumann, K.: https://www.huffingtonpost.com/karen-naumann/5-reasons-why-feedback-is_b_8728332.html. Accessed 28 Feb 2018
11. Thomke, S., von Hippel, E.: Customers as innovators: a new way to create value. Harv. Bus. Rev. **80**(4), 74–85 (2002)
12. Hauser, J.R.: How Puritan-Bennett used the house of quality. Sloan Manag. Rev. **34**(3), 61–71 (1993)
13. Gibbert, M., Leibold, M., Probst, G.: Five styles of customer knowledge management, and how smart companies use them to create value. Eur. Manag. J. **5**(20), 459–469 (2002)
14. Baroudi, J.J., Olson, M.H., Ives, B.: An empirical study of the impact of user involvement on system usage and information satisfaction. Commun. ACM **29**(3), 232–238 (1986)
15. Woodruff, R.B.: Customer value: the next source for competitive advantage. J. Acad. Mark. Sci. **25**(2), 139–153 (1997)
16. Kujala, S.: User involvement: a review of the benefits and challenges. Behav. Inf. Technol. **1** (22), 1–16 (2003)
17. van der Aalst, W.M.P., Rosa, M., Santoro, F.M.: Business process management: do not forget to improve the process. Bus. Info. Syst. Eng. **58**(1), 1–6 (2016)
18. Leopold, H., van der Aa, H., Reijers, H.A.: Identifying candidate tasks for robotic process automation in textual process descriptions. In: Gulden, J., Reinhartz-Berger, I., Schmidt, R., Guerreiro, S., Guédria, W., Bera, P. (eds.) BPMDS/EMMSAD 2018. LNBIP, vol. 318, pp. 67–81. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91704-7_5
19. Jansen-Vullers, M.H., Loosschilder, M.W.N.C., Kleingeld, P.A.M., Reijers, H.A.: Performance measures to evaluate the impact of best practices. In: Proceedings of the BPMDS 2007, Trondheim, pp. 359–368 (2007)

Teaching Business Process Development Through Experience-Based Learning and Agile Principle

Amin Jalali(✉)

Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden
aj@dsv.su.se

Abstract. Business process development is a challenging and important task in organizations that enables performing operations more effectively and efficiently. Traditionally, business processes have been developed through waterfall approaches which are described by different business process management life-cycles. Such development requires a lot of time that is one of the barriers to applying business process management in practice. Teaching business process development in the traditional way is also a challenge in education due to the long-running cycle to develop a business process. In this paper, we propose a new approach to teach business process development through experience-based learning and agile principle. The approach is implemented in an advanced course on business process management where students developed a real public business process in an agile way. The study shows that agile business process development not only enables students to automate the business process but also enhance their understanding of business process management.

Keywords: Business process management · Agile principle
Pedagogy · Teaching · Experience-based learning

1 Introduction

Business Process Management (BPM) is an area of research that aims to enable efficient and effective operation of activities in organizations to achieve business goals [10]. Business process models play a very important role in BPM. They not only provide a ground for negotiation but also support analysis of business processes that enable redesigning business processes. The redesign of business processes enables changing the as-is process models to to-be models. The models need to be implemented and configured to support enactment of business processes with the help of IT systems. The result of enactment also enables analyzing business processes based on data, which reveals more information to improve the business process by repeating this cycle. Thus, the business processes can be developed over time to fulfill the requirements.

J. Zdravkovic et al. (Eds.): BIR 2018, LNBIP 330, pp. 250–265, 2018.
https://doi.org/10.1007/978-3-319-99951-7_17

Business process development is a challenging and important task in organizations. There are different BPM life-cycles that propose different approaches to develop business processes, e.g., [4,9,10]. Despite differences among these life-cycles, they follow the same trend that proposes to design, analyze, implement, enact and monitor business processes. The development of business processes is usually considered as an ongoing process that requires a lot of iterations, and each iteration is performed on a long-running cycle. In each cycle, the implicit knowledge of people about the business process is transformed into explicit knowledge in terms of process models and their implementation. Such explicit knowledge can support knowledge transformation as described by Nonaka's theory of knowledge transformation [7], as described in [2]. These approaches are waterfall in nature, and they require a lot of cost and time to deliver the result.

The traditional business process development requires a lot of time that is one of the barriers to applying business process management in practice. Therefore, the idea of Agile Business Process Development (Agile BPD) has received more attention recently to overcome this problem. Agile BPD promises to develop business processes by splitting the project into several small process deliverables. Each deliverable can be performed in much shorter time. The approach can be very promising not only for developing business processes but also to teach business process development. Indeed, it can enable students to learn different phases of business process development by experiencing the life-cycle.

Therefore, we propose a new approach to teach business process development using agile business process development [2] through experience-based learning [5]. The approach is implemented at an advanced level course on business process management where students developed a real public business process in an agile way. The study shows that agile business process development not only enables students to automate the business process but also enhance their understanding of business process management. The shorter cycles enable students to understand and implement the business process in an effective and efficient way.

The remainder of this paper is organized as follows. Section 2 describes ideas of both agile business process development and experience-based learning. Section 3 introduces the approach that is used for supporting agile business process development. Section 4 reports the experiment of applying agile business process development as an experience-based learning case. Section 5 reports the results. Section 6 discusses limitations, future directions and related literature. Finally, Sect. 7 concludes the paper and introduces future works.

2 Background

This section describes Experience-Based Learning and Agile Business Process Development briefly.

2.1 Experience-Based Learning

Teaching is an important topic that enables learners to obtain a new or refine their existing knowledge. There are different learning methods which are

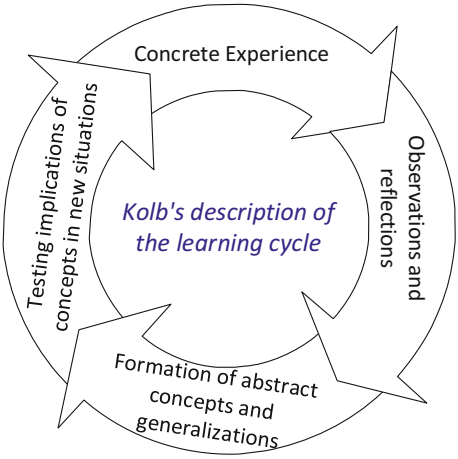


Fig. 1. The learning cycle proposed by Kolb et al. [5]

proposed by different researchers. Experience-Based Learning (EBL) is a method that considers experience as a core part of the learning process [5]. The aim is to enable learners to translate experience into concepts, which play the role of guideline for new experiences [3]. The transformation of experience to the concept is performed in a learning process that is described by Kolb et al. as a learning cycle (see Fig. 1).

This transformation is a process in which the learners obtain tacit knowledge from their experience, and it has four stages, i.e., (i) Concrete experience, (ii) Observations and reflections, (iii) Formation of abstract concepts and generalizations, and (iv) Testing implications of concepts in new situations.

As it can be seen the *experience* is the heart of this learning process, which should be defined by the instructors to enable learners to face new challenges.

The *observation and reflection* are also very important, which should be followed after experiencing the challenges. As one can look at a mirror to perceive more about herself or himself, the reflection deal with observing and elaborating on the own experience. It provides a room for thinking and grasping the experience that has performed. There are many different techniques that are used for reflection like informal group discussion, Comparing notes, roundtable discussions, etc [3].

Formation of abstract concepts and generalizations refer to humans capability to inductively make hypothesis based on experience. Such a hypothesis is a form of assumption that needs to be tested by new experiences later. However, it plays an important role in building tacit knowledge based on an experience that has been conducted.

Testing implications of concepts in new situations indicate the use of the obtained tacit knowledge in another setting. It means that we can decide if an obtained knowledge can be applied in another similar situation or not. The actual use of the knowledge in another setting means repeating the cycle that can enable refining the knowledge further.

2.2 Agile Business Process Development

The idea of agile business process development has been received more attention recently in both academic and industry. It proposes to develop a business process through iterations by dividing the project into manageable small deliverables. The main focus is on how the business process knowledge can be supported using business process management systems. The traditional knowledge transformation lifecycle in the BPM area is adopted from the knowledge transformation lifecycle proposed by Nonaka [7], and it has four phases, i.e., Externalization, Combination, Embedment, and Adoption (ECEA) [2]. The main focus is on how the tacit knowledge can be transferred into explicit knowledge.

To support agile business process development, the traditional lifecycle does not function very effective due to its long-running cycles. Thus, a new lifecycle is defined (see Fig. 2). The main focus in agile business process development is on how the tacit knowledge can be transferred from the business environment into embedded knowledge, and in turn how to transfer the embedded knowledge into tacit knowledge of users through adpotation [2]. The embedded knowledge is the semantics which is implemented in systems. The knowledge transformation cycle is different from traditional knowledge transformation lifecycle in business process development. The model is abbreviated to SEA, which stays for Socialization, Embedment, and Adoption.

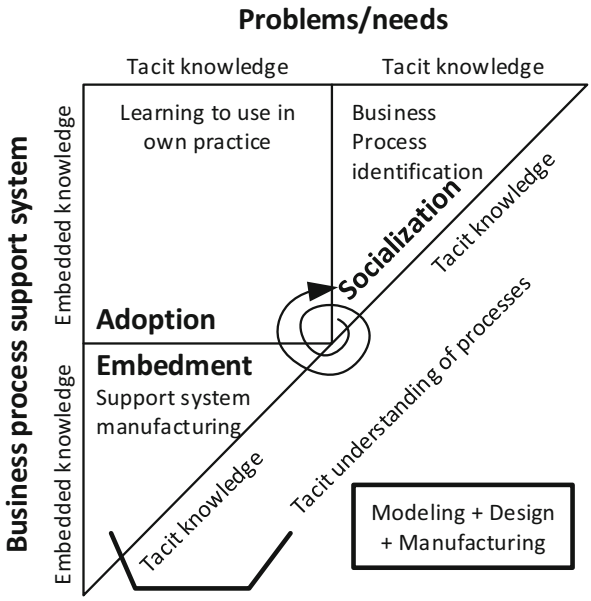


Fig. 2. Knowledge transformation lifecycle in Agile Business Process Development [2]

- *Socialization*: The cycle starts with Socialization, where tacit knowledge is transferred from the heads of one group of people to others via informal means, e.g., conversations during the coffee breaks, meetings, observations, working together, etc.
- *Embedment*: In this cycle, process modeling, system design, and manufacturing are merged into one phase. Indeed, one big cycle in previous knowledge transformation model is substituted by many smaller and shorter ones in this model. The system is built iteratively starting with the basic functionality that does not limit the flexibility of process participants to experiment with the new process. During the usage of the basic system, a better understanding of the needs is acquired, which is converted into adding details to the system in the next cycle.
- *Adoption*: In this cycle, the users of the system learn how to use it in own practice. It consists of transforming embedded knowledge (which is in the system) into the tacit knowledge of how to use it.

3 Approach

This section explains a new approach that proposes how agile business process development can support the learning process in teaching business process development. The approach is defined by combining the agile business process development lifecycle [2] and the learning cycle proposed by Kolb et al. [5]. Figure 3 shows the overall picture of the approach.

The approach has three main phases, called (i) Plan, (ii) Collaborate, and (iii) Deliver.

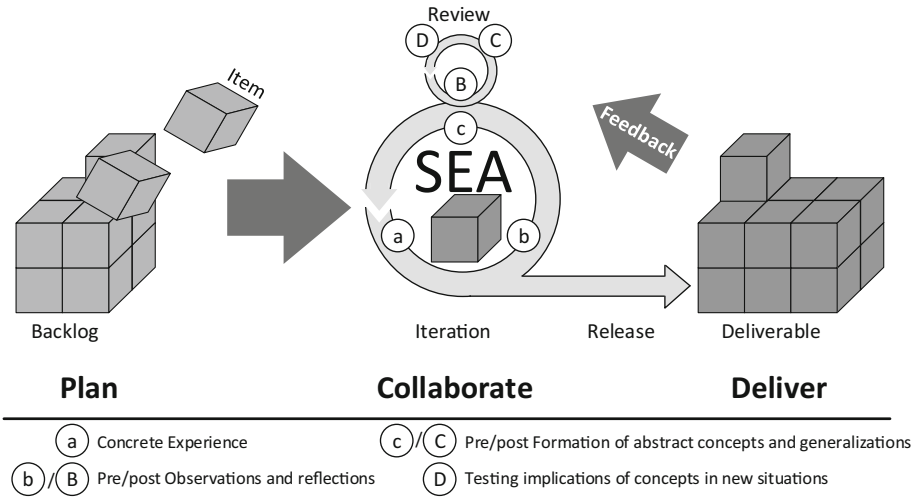


Fig. 3. Agile Business Process Development method for teaching, inspired from SCRUM method

- *In Plan phase*, the requirements of a business process needs to be prioritized based on their importance, feasibility, and other criteria. The set of prioritized requirements are called backlog - to comply with agile definitions in project management. The backlog contains several items each of which describes one requirement of the business process and its relation to business goals. These items will be implemented in iterations in the next phase.
- *In Collaborate phase*, the development team will implement items iteratively. In each iteration, they develop one requirement that has higher priority. The development includes (re-)designing, analyzing, implementing and testing the business process according to the new requirement. The result is a module that needs to be delivered.
- *In Deliver phase*, the developed module will be released to the user. In this phase, the feedback from users is very important. This feedback initiates another cycle in the collaboration phase that is called review. In the review cycle, the development teams need to consider and apply the feedback that they have received.

As it can be seen from the figure, the collaboration phases include two cycles, i.e., *Iteration* and *Review*. The development team experiences different stages of the learning cycle which is described earlier.

In the Iteration cycle, process developers develop the requirements following the SEA (Socialization, Embedment, and Adoption) cycle. In Socialization, they identify the scope of the business process and understand the needs to implement the selected item. In this way, they collaborate and obtain the tacit knowledge about the context that they want to implement. In Embedment, they implement the models into business process management systems in terms of embedded knowledge. In Adoption, they test the implemented process, which might give them new tacit knowledge since it may broaden their understanding of the business process. The rest of the adoption will happen in the Delivery phase when users adopt their work according to the implemented module.

In terms of the learning stages, the development team follows three stages of the learning cycle, which are annotated with a, b and c in Fig. 3. First, they have a concrete experience by developing the business process. Second, they will have pre-observation and reflections as they test the model. Finally, they have pre-formation of abstract concepts and generalizations based on their development experience.

Note that this tacit knowledge can be obtained for example by reading and understanding the business requirements, developing the as-is process, analysing the as-is process, refining the model as a to-be process, and developing the to-be process. Although the users may not be involved in this cycle, the development team can obtain some tacit knowledge from different sources. It is worth mentioning that if users are involved in requirement gathering, more knowledge can be obtained by development team.

In the Review cycle, the development team follows three stages of the learning cycle, which are annotated with A, B, and C in Fig. 3. First, they have a post-observation and reflections as they adopt their understanding based on the feedback that they had received. Second, they have post- formation of abstract

concepts and generalizations based on the received feedback. Finally, they need to consider if some parts were not accepted when testing implications of concepts in the new situation (as the situation is different if they understood it wrongly and had a deviation with the understanding of the user).

The approach is performed iteratively, and it contains short cycles due to the small scope of each item. Thus, it can help the development team to experience and have a good understanding of different phases of business process development life-cycle. In the next section, we will explain how the approach is evaluated.

4 Experimental Study

To evaluate our approach, we simulate the development process by conducting an experimental study. The study is performed on an advanced-level course in Business Process Management at Stockholm University in 2016 and 2017. In 2016 and 2017, 37 and 58 students have actively participated in the course respectively. The students formed their groups to develop a business process as their assignment. In total, there were 30 groups of students who participated in this study.

Each group received a business case that specifies different requirements. The business case is borrowed from a real project that had been conducted at Stockholm University in collaboration with one municipality in Sweden. The aim of this project is to develop a process to support social services that are open to the public and support citizens in their contacts with the municipalities. The municipality has 65,000 citizens, among which 1500 citizens are elderly, and 850 are disabled. One part of the project is to develop a service that supports these citizens when they need to apply for an emergency phone. The student received the business case and requirements for this process.

Each group needs to develop the business process in an agile way that is described in the previous section. They have used YAWL as a tool to develop business processes. Students had no previous knowledge about the tool, but some of them had some very basic understanding of business process design on a very general level. The author played the role of the stakeholder to whom the project should be delivered. To enable students to deliver iteratively, several timeslots are defined during the course in which students can present their development and receive feedback. These timeslots were optional, so it was up to students on how to define their backlogs and number of iterations that they are interested to go through, i.e., either many small iterations or fewer big iterations. Each group needed to deliver their project by the same deadline at the end of the course.

The two events of the course started and ended in these periods: 2016-03-21 to 2016-06-04, and 2017-03-21 to 2017-06-03, and the project can be considered as a long-running project in regards to a simulation setting. Each group submitted their project including all versions of the developed business process finally. The assignment is graded from 0 to 8, depends on the details that they have implemented and the maturity of the process to cover different perspectives and scenarios.

4.1 Summary of the Business Case

The business case based on which students developed the business process consists of several pages and documents. Here, we give a summary of the process, so it can help readers to follow the experimental study. This summary is borrowed from the course material.

The process in this project starts by receiving an application. The application is first registered. Based on some predefined criteria, the application is either routed to the customer relationship officer (CRO) for a manual handling or a positive decision automatically made after which the citizen is informed about the decision, and at the same time, the process progressed to an emergency phone installation phase.

During the emergency phone installation phase, an alarming number is obtained from the emergency service center (ESC). Then, installation order is sent to the installers and allocated to the first one to accept it. This installer is hence responsible for the installation of the emergency telephone equipment in the citizen's home. The installer calls the citizen and schedules the installation. A confirmation of the time is sent to the citizen. 24 hours before installation a reminder is sent, and the process progressed to the installation task.

Sometimes problems appear during the installation of an emergency phone. For instance, an IP phone may require a special configuration, or keys that are provided by the customer may turn out not to be working. In some cases, when the problem is technical, the installer helps the citizen to resolve it. In other cases, the citizen needs to resolve the problem. When the problem is resolved, a new installation time is booked, and the installation will proceed. After three unsuccessful installation attempts, however, the work is interrupted, the participants are informed about the cancelation, and the case is filed.

Note that the customer also has the right to cancel the process if her or his case is handling through the system. The customer can cancel the process all times before the installation has been performed successfully. The reasons for failing of installation is archived in a log. This log accumulates data from all installation attempts as well as the solutions that are reported for resolving the problems.

Before the installation, the installer configures the extra phone device with the alarming number obtained from the ESC. After the installation, the installer performs a test to ensure that the emergency phone works and checks out keys from the citizen's home. (S)He also enters some installation and keys data into the system and hands over the keys to the alarm group (AG), who are responsible for handling the keys in a secure way.

The alarm group allocates a storage place to the keys and stores them. Henceforth, the citizen can start using their emergency phone. The citizen is added to the accounting system, and the start time for the service fee payment is calculated. These two steps are today done manually, by an assistant. Finally, the citizen is informed about the start service and payment time, and the case is filed. The CRO is notified of the completion of the installation and the filing of the case.

The installer logs all attempts (s)he has made to book installation time with a citizen. This is done for reporting purposes. The alarm group looks at this information when it has taken a long time to come in contact with a citizen for booking an installation. After a citizen has been informed that (s)he has been granted the service, (s)he can withdraw his/her application. This can be done until the installation has taken place. For instance, it sometimes happens if elderly people move to special housing soon after they had launched an application.

If a citizen wishes to withdraw his/her application after the installation has taken place, a different process is used, namely the process "withdrawer of an emergency phone service".

5 Result

This section reports the result of simulating the business process development using our proposed approach. In total, 30 groups have been involved in this study. Each group includes 2, 3 or 4 members.

5.1 The Overall Picture

Figure 4 includes three charts (two box-and-whisker plots and one scattered plot) that describe overall statistical information. *The left box-and-whisker plot* shows that a quarter of groups received 7 or 8 points, which indicates a complete model in final delivery. It also shows that only one-quarter of groups received 4 or less for the final grade. *The right box-and-whisker plot* shows that a quarter of groups performed the project in less than 27 iterations, while half of the group performed the project in between 27 and 103 iterations. It is also interesting that a quarter of groups has performed the project in more than 103 iterations! *The scattered plot* shows the relation between the number of iterations and the final grade. In general, a group receives a better grade as the number of iterations increased except one group that performed the project in 117 iterations but received a low grade (3 out of 8). The iteration variable depicts the number of iterations that had happened in both iteration and review cycles.

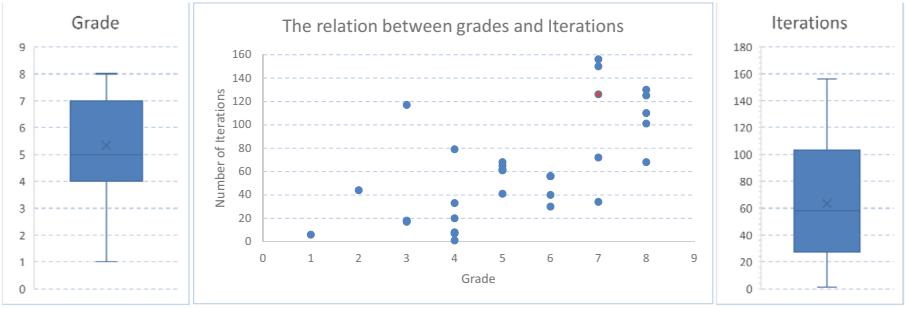


Fig. 4. The overall picture of result

We also plotted two scattered plots based on the number of group members and final grade and number of group members and the number of iterations. The result indicates that there is no significantly meaningful relationship between these variables.

5.2 An Observation of Following Agile BPD

The agile business process development approach that is proposed in this paper is followed by each group. In this section, we selected one case to show how this approach is applied in practice by students. A group from 2016 is selected. This group had two members and developed the process in 126 iterations, and they received 7 as the final grade. This group is marked with a red dot in Fig. 4. Studying how the process is developed in iterations gives more insight into how the approach is applied. We selected a few versions to show how the process is developed over time.

Figure 5 shows one of the early version of the process model. As it can be seen, it only has four activities. First, the application is registered. The requirements indicate that the application can be handled manually or be supported by the implemented version of the process. The *prepare and send installation order* indicates the flow in which the process should be supported by the IT system. The other flow is related to the manual handling. This version does not comply with the requirement, and as we can see in the next versions, the development team revised the process model after receiving the review.

Figure 6 shows one of the version after several iterations. As it can be seen, the models are refined to incorporate more activities to describe the normal flow of the process. At this stage, we can see the process model is enriched with resource perspective information as well. There are still some problems in the model regarding syntactical correctness and completeness. For example, the process contains a deadlock. The deadlock can appear if the application is registered, and the process continues in any flow except manual handling. The deadlock happens before the *install emergency phone* task because it is decorated

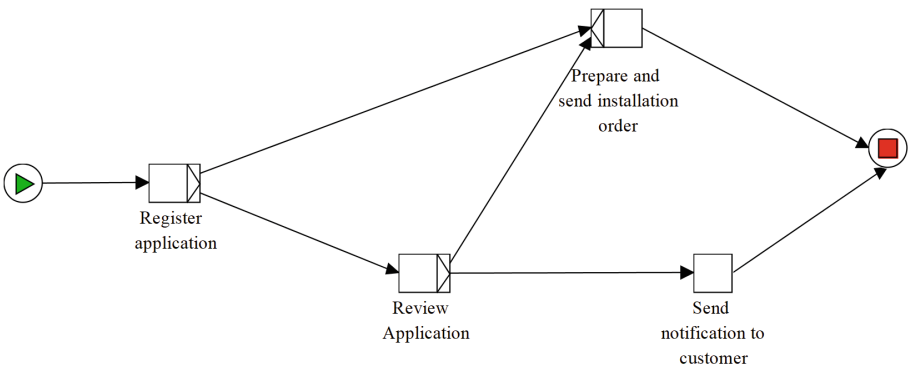


Fig. 5. The second version of the process model

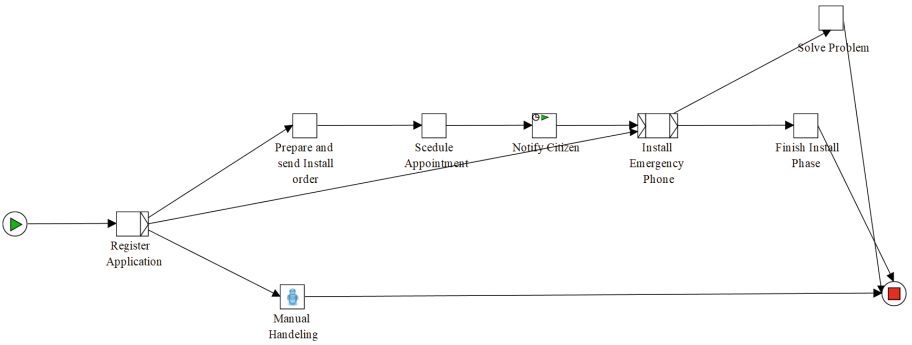


Fig. 6. The version in which the normal flow is emerged

with an AND join. This join will never happen since one of the incoming flows can happen, and the AND join semantics require all flows to be fulfilled. This shows that the development team had not written all possible test case scenarios, and they would have been realized it after receiving the feedback!

Figure 7 shows some iterations in the middle of the business process development. As it can be seen, the model is evolved and gets more mature, and

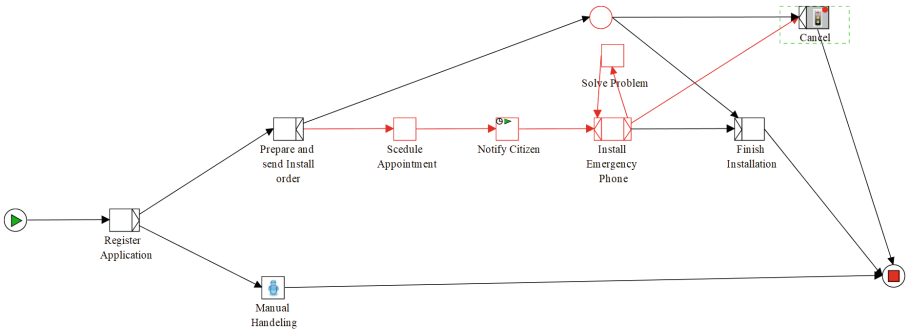


Fig. 7. The version in which the modeling of exceptions are started

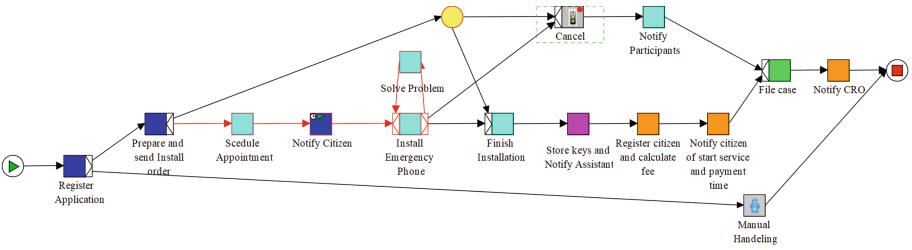


Fig. 8. The version that includes colour to manage the process model complexity

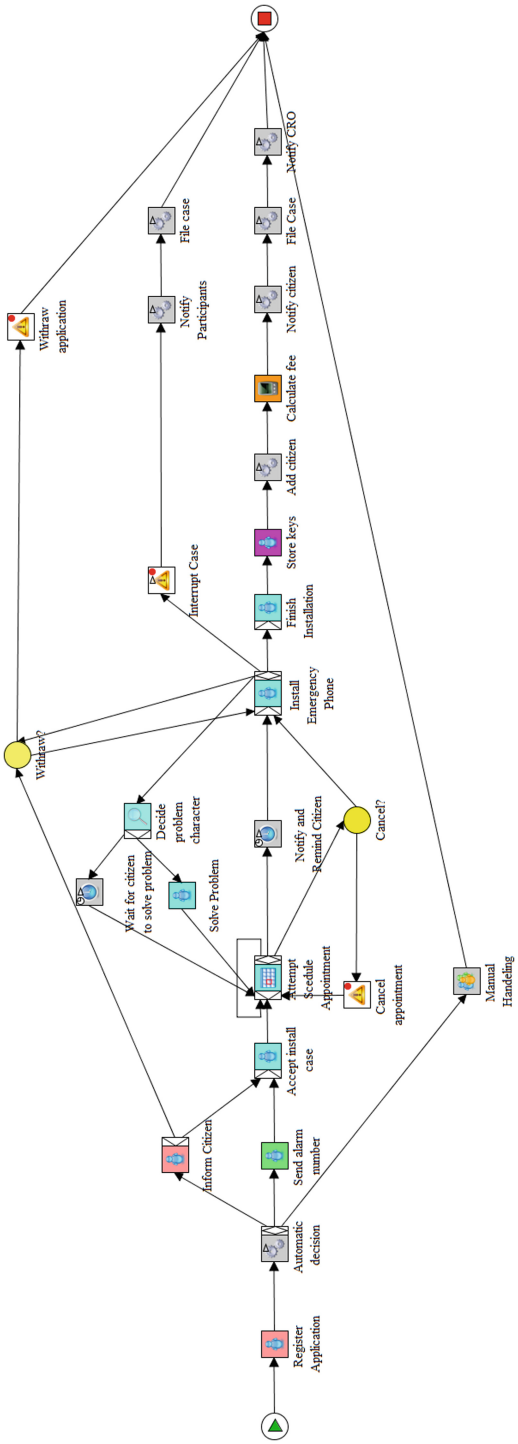


Fig. 9. The final version of the project

it does not have any syntactical error. This shows that the knowledge of the team is also developed regarding syntactical and semantical errors in business process models. This model also supports a cancellation feature which is a kind of exception in the business process. It shows that the development teams have started to extend the exceptions into the normal path. The cancellation feature in this version enables citizens to cancel their request after preparation activity and before the installation is finished.

Figure 8 shows more matured version of the model. The development team added more activities to the model. They have also used some decoration technique to increase the readability of the model. They colored different activities based on different resources and roles who can perform the tasks. This shows that they started to think about applying different design techniques to overcome the complexity of the model as it increases. The process is still not complete. It is correct regarding syntactical and semantical rules.

Figure 9 shows the final version of the model. As it can be seen, the process is now more matured. It covers most of the requirements, and it is correct regarding syntactical and semantical rules. It still can be improved, e.g., the AND join before Finish Installation task can be removed since there is only one incoming flow. It does not make model incorrect, but it is redundant. Also, the development team can still apply different heuristic process redesign techniques to improve the process. For example, Install Emergency Phone and Finish Installation can be merged into one task because they are performed by one resource, and the states can be managed by the system rather than be manually entered by installers.

5.3 Students' Reflections

Each group is asked to submit their reflection in addition to all versions of the developed process. Here, we summarize some of their reflections in regards to agile business process development.

- "... We first modeled our case on paper and later implemented it in the YAWL Editor. We thought that it would be easiest to model the whole workflow first and then implement data perspective and recourses. This was our initial hypothesis. This proved more difficult than we expected. ...
  ... agile approach in part contradicted our first approach which was very traditional in terms of IT-development. In this assignment we learnt that although its essential to know the direction in which your model should be heading, it's good to work in iterations. The iterations help one (us) to deal with problems one at a time and thereby create a model that is susceptible to changing e.g. circumstances. ..."
- "... We have learned to work in an iterative way with testing after each implemented element, the importance of testing variables and parameters became very clear to us to know what was a success, and what was not."
- "... The best part with the assignment was to see how the model is executable ..."

6 Discussion and Related Literature

This section discusses limitations, future directions, and related literature.

6.1 Limitations and Future Directions

As one limitation, it should be noted that this study did not evaluate the effectiveness of the proposed approach in comparison to traditional BPM development techniques. Indeed, it is only focused to show a case in which the approach succeed to be used as an effective teaching method in action.

Moreover, the iteration numbers refer to both iteration and review cycles. It would be nice if the number of iterations can be distinguished in future study since it can reveal more important information about this method. It also should be noted that the small iterations might seems too small for professional process developers, yet we need to consider that students are new to these concepts, and they are in their early learning process. It is also interesting to see if the iterations have moved from a smaller one into a big one over time, which needs another study that considers taking this variable into account. Also, it would be interesting to investigate how the total number of iterations and the success of the project depend on the number of customer-developer reviews.

6.2 Related Literature

There are few works that proposed a new approach or investigated the effectiveness of some approaches in teaching Business Process Management.

Bandara et al. [1] investigated the Status and challenges in teaching Business Process Management in academia. They performed an extensive study in five different universities in four different countries. Based on their investigation, they identified six sorts of challenges for educating business process management subject. *First*, they mention that the lack of qualified instructors to teach BPM and pedagogical resources are very challenging. *Second*, they identified that BPM courses are offered by different schools which could have different expectations and even different sort of students, which is challenging in defining educational resources that fit all. *Third*, the variety of technologies and phases in BPM is considered as another challenge in teaching this topic. It is also mentioned that the delivery of the BPM courses is challenging due to the heavy technology load which exists in this area. Also, licenses and upgrades in software is another challenge for educators. *Forth*, they mentioned that the university might be interested to lower the enrolment requirements to build interest on students to take the course, which can make the teaching of this subject very challenging. *Fifth*, the lack of pedagogical research related to BPM education is identified as another challenge. Such research might assist others in identifying best practices and pedagogical strategies that can be applied in other courses. *Sixth*, the unclear career pathway and student position in this area are mentioned as another challenge. This roots in the inconsistent use of terms by organizations that seek to develop or improve their business processes.

The approach that is proposed in this paper may be able to help educators to deal with the third and fifth challenges that are identified in [1]. The short iterations can enable a better understanding of one technology since the students can focus on one technology at a time and evaluate their understanding by implementing the artifacts. Also, this paper can help others to employ this technique to evaluate how this approach can help them to teach business process development.

There are also some studies about educating other non-workflow business process modeling notations. For example, Koutsopoulos et al. [6] investigated teaching and learning State-Oriented Business Process Modeling (StoBPM). They identified challenges that are related to StoBPM as they mentioned that this technique is substantially differs from the mainstream modeling techniques-including the one that is applied in this paper. Routis et al. [8] evaluated how Case Management Model and Notation (CMMN) may be used by modelers. This notation mainly is used in knowledge intensive processes. Teaching this notation might be challenging at the university level since modeling these processes might need more experience and knowledge about the business domain.

7 Conclusion

This paper introduced a new approach to teach business process development based on experience-based learning and agile principle. The approach is grounded based on existing theories behind experience-based learning and agile business process development, and it is simulated in an advanced course on business process management at Stockholm University in two different years. In this simulation, the students received the requirements for developing support systems for a business process in public domain, which was borrowed from a real project. The role of the end user is played by the teacher. The students handed over all versions that they have developed, and the analysis of all versions shows that the teaching method was successful to enable students to learn business process development. The study shows that agile business process development not only enables students to automate the business process but also enhance their understanding of business process management. The approach can enable teaching business process development in the absence of business users in educational context.

As future directions, it is interesting to apply this technique in real process development projects. It is also interesting to do a test study and compare the use of agile business process development and traditional business process development on student learning.

References

1. Bandara, W., et al.: Business process management education in academia: status, challenges, and recommendations. CAIS **27**, 743–776 (2010)
2. Bider, I., Jalali, A.: Agile business process development: why, how and when - applying Nonaka's theory of knowledge transformation to business process development. Inf. Syst. e-Bus. Manag. **14**(4), 693–731 (2016)

3. Boud, D., Keogh, R., Walker, D.: Reflection: Turning Experience into Learning. Routledge, Abingdon (2013)
4. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management, 2nd edn. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-56509-4
5. Kolb, D.A., Rubin, I.M., McIntyre, J.M.: Organizational Psychology: An Experiential Aproach. Prentice-Hall, Upper Saddle River (1974)
6. Koutsopoulos, G., Bider, I.: Teaching and learning state-oriented business process modeling. experience report. In: Reinhartz-Berger, I., Gulden, J., Nurcan, S., Guédria, W., Bera, P. (eds.) BPMDS/EMMSAD 2017. LNBIP, vol. 287, pp. 171–185. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59466-8_11
7. Nonaka, I.: A dynamic theory of organizational knowledge creation. In: Knowledge, Groupware and the Internet, pp. 3–42. Elsevier (2000)
8. Routis, I., Nikolaidou, M., Anagnostopoulos, D.: Modeling collaborative processes with CMMN: success or failure? An experience report. In: Gulden, J., Reinhartz-Berger, I., Schmidt, R., Guerreiro, S., Guédria, W., Bera, P. (eds.) BPMDS/EMMSAD 2018. LNBIP, vol. 318, pp. 199–210. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91704-7_13
9. Van der Aalst, W.M.: Business process management: a comprehensive survey. ISRN Softw. Eng. **2013** (2013)
10. Weske, M.: Business process management architectures. In: Weske, M. (ed.) Business Process Management, pp. 333–371. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28616-2_7

Author Index