Tiago Prince Sales · Henderik A. Proper ·
Giancarlo Guizzardi · Marco Montali ·
Fabrizio Maria Maggi · Claudenir M. Fonseca (Eds.)

LNBIP 466

Enterprise Design, Operations, and Computing

EDOC 2022 Workshops

IDAMS, SoEA4EE, TEAR, EDOC Forum,
Demonstrations Track and Doctoral Consortium
Bozen-Bolzano, Italy, October 4–7, 2022
Revised Selected Papers

Springer

Lecture Notes in Business Information Processing 466

Series Editors

Wil van der Aalst, *RWTH Aachen University, Aachen, Germany*
Sudha Ram, *University of Arizona, Tucson, AZ, USA*
Michael Rosemann, *Queensland University of Technology, Brisbane, QLD, Australia*
Clemens Szyperski, *Microsoft Research, Redmond, WA, USA*
Giancarlo Guizzardi, *University of Twente, Enschede, The Netherlands*

LNBIP reports state-of-the-art results in areas related to business information systems and industrial application software development – timely, at a high level, and in both printed and electronic form.
The type of material published includes

- Proceedings (published in time for the respective event)
- Postproceedings (consisting of thoroughly revised and/or extended final papers)
- Other edited monographs (such as, for example, project reports or invited volumes)
- Tutorials (coherently integrated collections of lectures given at advanced courses, seminars, schools, etc.)
- Award-winning or exceptional theses

LNBIP is abstracted/indexed in DBLP, EI and Scopus. LNBIP volumes are also submitted for the inclusion in ISI Proceedings.

Tiago Prince Sales · Henderik A. Proper ·
Giancarlo Guizzardi · Marco Montali ·
Fabrizio Maria Maggi · Claudenir M. Fonseca
Editors

Enterprise Design, Operations, and Computing

EDOC 2022 Workshops

IDAMS, SoEA4EE, TEAR, EDOC Forum,
Demonstrations Track and Doctoral Consortium
Bozen-Bolzano, Italy, October 4–7, 2022
Revised Selected Papers

Springer

Editors
Tiago Prince Sales
University of Twente
Enschede, The Netherlands

Henderik A. Proper
TU Wien
Vienna, Austria

Giancarlo Guizzardi
University of Twente
Enschede, The Netherlands

Marco Montali
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy

Fabrizio Maria Maggi
Free University of Bozen-Bolzano
Bolzano, Italy

Claudenir M. Fonseca
University of Twente
Enschede, The Netherlands

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-031-26885-4 ISBN 978-3-031-26886-1 (eBook)
https://doi.org/10.1007/978-3-031-26886-1

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2023
Chapter "Prosimos: Discovering and Simulating Business Processes with Differentiated Resources" is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). For further details see license information in the chapter.
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

For over twenty-five years the EDOC conference has been the primary annual event for disseminating and discussing the latest developments in enterprise computing. In addition to the main track, EDOC 2022 offered a forum, a demonstration track, and a doctoral consortium. It also hosted three workshops of interest to the community. All of these events were held in-person, together with the main conference in Bozen-Bolzano, Italy.

The forum was introduced this year as a track within the main conference where authors were given a platform to present and discuss early-stage work. In this first edition, we accepted four forum papers for presentation and publication in this proceedings volume.

The demonstration track offered a highly interactive outlet for researchers and practitioners to present prototypes and applications in the context of enterprise computing. This year, we accepted five tools for presentation, each of which was accompanied with a short paper published here.

The doctoral consortium is a track designed to encourage early-stage doctoral candidates to present their research projects and exchange with other researchers in their fields. All accepted projects are reviewed by senior researchers, who then provide valuable feedback to the candidates. This year, we accepted three projects for presentation and publication in this proceedings volume.

Workshops within EDOC cover more focused topics and allow for the presentation and discussion of work that is in the earlier development stages. As such, the workshops provide an excellent venue for discussing enterprise computing topics that can become important research streams in the future, as well as topics that are already important in a smaller and more focused setting. This year, we are proud to have hosted three workshops, namely:

- Workshop on Trends in Enterprise Architecture Research (TEAR)
- Workshop on Service-Oriented Enterprise Architecture for Enterprise Engineering (SoEA4EE)
- Workshop on Intelligent Digital Architecture, Methods, and Services for Industry 4.0 and Society 5.0 (IDAMS)

These workshops were selected based on their fit with EDOC's topics of interest, their format and target audience, and their potential to attract high-quality papers. Each of the workshop programs was established in collaboration with the workshop chairs and their accepted papers have undergone a rigorous reviewing and selection process. As the competition for the main track was high, some submissions were redirected to the relevant workshops, where they were accepted for presentation and publication. All in all, 14 workshop papers are published in this proceedings volume.

A noteworthy change we made this year at EDOC was the adoption of a "post-proceedings" format for this volume. With this change, the authors of papers accepted for one of the workshops, the doctoral consortium, the forum, and the demonstration track were allowed to improve their papers after receiving feedback at the conference. We believe this led to better and more mature publications.

Lastly, we would like to thank everyone involved in the organization of EDOC 2022. Their contribution was key to the success of the conference and its satellite events. We give a special thankyou to our workshop organizers, the chairs of the forum, doctoral consortium, and demonstrations track, as well as to the authors of the papers published here.

October 2022

Tiago Prince Sales
Henderik A. Proper
Giancarlo Guizzardi
Marco Montali
Fabrizio Maria Maggi
Claudenir M. Fonseca

Organization

General Chairs

Giancarlo Guizzardi	University of Twente, The Netherlands
Marco Montali	Free University of Bozen-Bolzano, Italy
Fabrizio Maria Maggi	Free University of Bozen-Bolzano, Italy

Program Committee Chairs

João Paulo A. Almeida	Federal University of Espírito Santo, Brazil
Dimka Karastoyanova	University of Groningen, The Netherlands

Workshop Chairs

Tiago Prince Sales	University of Twente, The Netherlands
Henderik A. Proper	Luxembourg Institute of Science and Technology & University of Luxembourg, Luxembourg

Demonstrations Track Chairs

Massimiliano de Leoni	University of Padua, Italy
Ivan Donadello	Free University of Bozen-Bolzano, Italy
Cristine Griffo	Free University of Bozen-Bolzano, Italy

Doctoral Consortium Chairs

Felix Mannhardt	Eindhoven University of Technology, The Netherlands
Chiara Di Francescomarino	Fondazione Bruno Kessler, Italy

Industrial Chairs

Zoran Milosevic	Deontik/Best Practice Software, Australia
Flavia Santoro	Rio de Janeiro State University, Brazil

Proceedings Chair

Claudenir M. Fonseca	University of Twente, The Netherlands

Financial Chairs

Cristine Griffo	Free University of Bozen-Bolzano, Italy
Glenda Amaral	Free University of Bozen-Bolzano, Italy
Renata S. S. Guizzardi	University of Twente, The Netherlands

Publicity Chairs

Dominik Bork	TU Wien, Austria
Estefanía Serral	KU Leuven, Belgium

Website Chair

Claudenir M. Fonseca	University of Twente, The Netherlands

Local Organization Chairs

Tiago Prince Sales	University of Twente, The Netherlands
Mattia Fumagalli	Free University of Bozen-Bolzano, Italy
Pedro Paulo F. Barcelos	Free University of Bozen-Bolzano, Italy
Claudenir M. Fonseca	University of Twente, The Netherlands

Local Organization Committee

Cristine Griffo	Free University of Bozen-Bolzano, Italy
Elena Romanenko	Free University of Bozen-Bolzano, Italy

Hannes Hell	Free University of Bozen-Bolzano, Italy
Isadora Valle Sousa	Free University of Bozen-Bolzano, Italy
Ítalo Oliveira	Free University of Bozen-Bolzano, Italy

Steering Committee

Selmin Nurcan	University Paris 1 Panthéon-Sorbonne, France
Alan Wee-Chung Liew	Griffith University, Australia
Colin Atkinson	University of Mannheim, Germany
Georg Grossmann	University of South Australia, Australia
João Paulo A. Almeida	Federal University of Espírito Santo, Brazil
Marten van Sinderen	University of Twente, The Netherlands
Remco Dijkman	Eindhoven Univ. of Technology, The Netherlands
Stefanie Rinderle-Ma	University of Vienna, Austria
Sylvain Hallé	Université du Québec à Chicoutimi, Canada
Zoran Milosevic	Deontik & Best Practice Software, Australia

Contents

TEAR 2022

EDOC Forum

Demonstrations Track

Doctoral Consortium

IDAMS 2022

Intelligent Digital Architecture, Methods, and Services for Industry 4.0 and Society 5.0 (IDAMS 2022)

Preface

The digital transformation of global industries and value chains and the associated need for structured research and standardization has given rise to major global and national initiatives. These initiatives address the potential and challenges of digitalization. Enterprises and societies currently face crucial challenges, while Industry 4.0 becomes increasingly important in the global manufacturing industry. Industry 4.0 offers a range of opportunities for companies to increase the flexibility and efficiency of production processes.

The development of new business models can be promoted with digital platforms and architectures for Industry 4.0. Industry 4.0 is dedicated to research and practice for industry and supports the implementation of this vision, especially in manufacturing companies. According to the Japanese government, Society 5.0 is more general and can be defined as a fusion between cyberspace and physical space, addressing economic progress aligned with solving social problems by providing goods and services to meet repeated latent needs regardless of location, age, gender, or language.

Contemporary advances in the field of artificial intelligence have led to a rapidly growing number of intelligent systems that can operate entirely independently of human intervention or enable interactions of unprecedented complexity with humans. Data plays a central role in intelligent digital architecture and allows automation of decisions impacting all stakeholders.

The use of artificial intelligence techniques enables decisions that were previously reserved for humans to be made autonomously. Intelligent systems augment processes by creating automated interfaces to human beings and replacing human decision making with machine-based decision-making. Intelligent digital architectures support requests for, configuration of, and fulfillment of services.

Digitalization promotes the creation of intelligent systems and services with an intelligent digital architecture. Products based on intelligent digital architectures become aware of their environment, act upon it, are able to interact with human beings, and can change their functionality during their lifetime. Products and services based on intelligent digital architectures have local autonomous capabilities and extend them dynamically by accessing external services.

Platforms become feasible by matching the supply and demand of services, resources, and products. Intelligent digital architectures also enable and enhance business models by integrating resources and leveraging decision making in unprecedented ways, for instance, by applying a Digital Enterprise Architecture Framework such as the Adaptive Integrated Digital Architecture Framework (AIDAF). Public discourse on "autonomous" algorithms which work on "passively" collected data contributes to this view.

The EDOC Workshop - Intelligent Digital Architecture, Methods, and Services for Industry 4.0 and Society 5.0 - covers fundamental and practical aspects to support digital transformation. This disruptive change interacts with all information processes and systems, which for years has been important business enablers for the digital transformation. Intelligent digital architectures enable intense interaction with customers and products. The customer is closely integrated with business processes and interacts like a co-worker by using implicit touchpoints, which are provided by mobility and wearable systems and the Internet of Things. In this way, customer experience is fostered with disruptive transformation and continuous improvement.

The IDAMS 2022 workshop was a half-day workshop in conjunction with EDOC 2022. All submissions have been peer-reviewed by at least three members of the EDOC 2022 and IDAMS international program committee.

We wish to thank all authors for having shared their work with us, as well as the members of the IDAMS 2022 Program Committee and the organizers of EDOC 2022 for their help with the organization of the workshop.

October 2022

Alfred Zimmermann
Yohimasa Masuda
Rainer Schmidt

Organization

Workshop Chairs

Alfred Zimmermann	Reutlingen University, Germany
Yohimasa Masuda	Keio University, Japan & Carnegie Mellon University, USA
Rainer Schmidt	Munich University of Applied Sciences, Germany

Program Committee

Abdellah Chehri	Université du Québec à Chicoutimi, Canada
Andreas Speck	Christian-Albrechts-University Kiel, Germany
Asif Gill	University of Technology Sydney, Australia
Christian Schweda	Reutlingen University, Germany
Dierk Jugel	Reutlingen University, Germany
Dimka Karastoyanova	University of Groningen, The Netherlands
Hironori Takeuchi	Musashi University, Japan
Janis Stirna	Stockholm University, Sweden
John Gøtze	IT University of Copenhagen, Denmark
Karlheinz Blank	T-Systems International, Germany
Kurt Sandkuhl	University of Rostock, Germany
Marco Aiello	University of Stuttgart, Germany
Matthias Wissotzki	Wismar University of Applied Sciences, Germany
Michael Möhring	Munich University of Applied Sciences, Germany
Milan Simic	RMIT University, Australia
Oliver Bossert	McKinsey & Company, Germany
Shuichiro Yamamoto	Nagoya University, Japan
Ulrike Steffens	Hamburg University of Applied Sciences, Germany

Digital Architectures Under Society 5.0: An Enterprise Architecture Perspective

Jean Paul Sebastian Piest[1](✉), Yoshimasa Masuda[2,3], and Maria Eugenia Iacob[1]

[1] University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
j.p.s.piest@utwente.nl
[2] Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[3] Tokyo University of Science, 1 Chome-3 Kagurazaka, Shinjuku City 162-8601, Tokyo, Japan

Abstract. As Japan's vision regarding Society 5.0 is unfolding, several digital architectures are (being) designed and constructed. Although Society 5.0 is extensively researched and its application is illustrated in several domains, little research is conducted regarding digital architectures under Society 5.0. The aim of this paper is to identify and analyze digital architectures that are (being) designed and constructed under Society 5.0 from an enterprise architecture perspective. Extending the results of related studies, a systematic literature review is conducted to find out which enterprise architecture frameworks and principles are utilized to create digital architectures under Society 5.0. The initial search in three scientific databases resulted in a sample of 857 documents. After removing duplicates and screening based on abstracts, 116 documents have been analyzed. This revealed a variety of 37 digital architectures in different domains. The main finding is that the analyzed digital architectures mainly use conceptual and system representations, make little use of established enterprise architecture frameworks, are not based on explicit principles, and do not refer to a reference architecture for Society 5.0. This paper presents an enterprise architecture perspective on Society 5.0 and is considered of value for scholars and enterprise architects that are interested in digital architectures under Society 5.0. This study is limited to scientific databases and selected published articles. Empirical research is required to validate the findings. As the interest in Society 5.0 is increasing, it is recommended to develop reference architectures and guiding principles to support digital transformation processes. Future research can contribute to develop prescriptive design knowledge and supporting methods for designing, constructing, and evaluating digital architectures under Society 5.0.

Keywords: Society 5.0 · Digital architecture · Enterprise architecture · Frameworks · Principles · Systematic literature review · VOSviewer

1 Introduction

Since its introduction in Japan's 5th Science and Technology Basic Plan 2016–2021, the topic Society 5.0 (hereafter S5.0) attracted the interest of scholarly and policy makers

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 5–24, 2023.
https://doi.org/10.1007/978-3-031-26886-1_1

around the globe. S5.0 is defined as "A human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space" [1]. Originally developed in Japan, S5.0 is being adopted in many other countries and closely related to Industry 4.0 (hereafter I4.0) and the Sustainable Development Goals (SDGs) [2–4]. I4.0 refers to four distinct industrial revolutions and the national strategic initiative in Germany that focuses on realizing the smart factory based on Cyber-Physical Systems (CPS), Internet of Things (IoT), and Artificial Intelligence (AI) [3]. The SDGs are part of the United Nations Member States's 2030 agenda for sustainable development and provides a shared blueprint for peace and prosperity for people and the planet, for now and towards the future [5]. The S5.0 vision and related development programs in Japan are compared, amongst other countries, to I4.0 in Europe, advanced manufacturing in the US, and Made in China 2025, and more recently linked to the emergence of I5.0 [3].

With the completion of the 5th Science and Technology Basic Plan various digital architectures are (being) constructed under S5.0. Surprisingly, S5.0 attracted fairly limited attention in the Enterprise Architecture (EA) discipline. In order to guide the large-scale implementation of I4.0, reference architectures and standards have been developed, including the Reference Architecture Model for Industry 4.0 (RAMI4.0). On the contrary, no publicly available reference architecture for S5.0 seems to exist.

The main aim of this paper is to identify and analyze digital architectures that are (being) constructed under S5.0. The following research questions are formulated:

- Which digital architectures are developed under S5.0?
- Which enterprise architecture frameworks and/or principles are used to design and construct S5.0 applications?
- Is there a reference architecture for S5.0? If not, is there a need for developing a reference architecture?

Extending the scope of earlier conducted Systematic Literature Reviews (SLR), an updated SLR is conducted using the PRISMA 2020 guidelines (https://prisma-statement.org/). The envisioned contribution of this research is to provide a common ground to start developing EA perspectives regarding S5.0. and supporting digital transformation processes. More specifically, this paper provides scholars and enterprise architects an overview of digital architectures that are (being) designed and constructed under S5.0. in relation to EA frameworks and principles.

This paper is structured as follows. Section 2 lays down the theoretical foundation by summarizing related work regarding S5.0, EA frameworks and design approaches. Section 3 is concerned with the methodological approach of the SLR. Section 4 presents the results and findings of the SLR regarding digital architectures under S5.0, emphasizing the use of EA frameworks and principles. Section 5 discusses the SLR outcomes. Section 6 concludes and positions opportunities for future research.

2 Related Work

This section summarizes relevant related work regarding S5.0 and EA frameworks. Section 2.1 discusses the topic S5.0. and related SLRs. Section 2.2 highlights established

EA frameworks and principles. Section 2.3 presents related work concerning the design of S5.0 solutions.

2.1 S5.0

The Hitachi-UTokyo Laboratory published a book regarding S5.0 encompassing the people centric, super smart society [4]. The authors contextualized the five societies in terms of productive approach, material, transport, settlement form, and ideals. S5.0 is categorized as the super smart society in which cyberspace and the physical space are merged, that is built with smart materials, provides transport based on autonomous driving, and is organized in autonomous decentralized cities. The city ideal is humanity. This is emphasized in the statement that S5.0 is leveraging technology to balance economic development with the resolution of social problems, taking into account the demographic developments, and bring wealth, comfort, and quality of life to people. S5.0 leverages technology to establish a knowledge-intensive and data-driven society. Here, digital platforms play an important role. Various examples are presented to illustrate the S5.0 vision, including personal health robots, autonomous transport, smart energy consumption, and data-driven governmental services. Establishing smart cities requires a massive transformation to integrate IT and sensors into urban planning, public infrastructure, energy systems, and individual households. This raises the important question and/or concern how to balance the needs of society and individuals. The habitat innovation framework presents a set of Key Performance Indicators (KPIs) and formulas to balance structural transformation with technological innovation and quality of life. The KPIs support measuring residents quality of life, evaluating social issues, and predicting drivers for social issues. From an EA perspective, the so-called urban datarization will raise challenges regarding integrating data and systems. The trends and development of smart cities in Japan are compared by the authors with developments in the US and the EU. The book provides an extensive overview of S5.0. Additionally to the book, some related review studies and SLRs are identified regarding S5.0.

Shahidan et al. [2] conducted a SLR regarding S5.0 to map the emergence of S5.0 and its core concepts. Their SLR is based on the PRISMA guidelines (version not reported), included 142 documents from the Scopus database (initial sample not reported), and a bibliometric analysis using VOSviewer version 1.6.16. Their SLR shows that S5.0 is a relatively young research field and is predominantly studied in engineering disciplines, but has interdisciplinary traits. The research indicates a strong link between S5.0 and I4.0. and its related technologies to build smart cities. Furthermore, the study shows that S5.0 is studied in 41 countries. The authors conclude that S5.0 is an under researched topic that requires further research. More specifically, the authors recommend that additional databases (e.g., Web of Science) should be included and research can be done based on analysis of co-citations and bibliographic couplings.

Roblek et al. [3] conducted a similar SLR regarding S5.0 and the relation to I4.0 and I5.0. Their SLR is based on a three-step approach, includes 37 documents from an initial sample of 916 documents from the Web of Science database, and a bibliometric analysis using VOSviewer version 1.6.16. The authors examined S5.0 from a historical, economical, and technological perspective. Furthermore, the author emphasize the transformation challenges to transition from S4.0 to S5.0. The authors recommend to develop a set of good practices to foster implementation (in other countries) and pay more attention to the risks related to the digital society.

2.2 EA: Frameworks and Principles

In the past decades, the discipline of EA evolved to a well-known practice of business and IT alignment. A review illustrates how the discipline first focused on understanding and modelling EA, and gradually switched focus to managing EA [6].

EA frameworks have various approaches, levels of granularity, and abstraction. A comparison study [7] identified several established EA frameworks, varying from generic EA frameworks to specific EA frameworks. Generic EA frameworks include the Zachman framework for EA and The Open Group Architectural Framework (TOGAF). Domain specific EA frameworks include the Department of Defense Architecture Framework (DoDAF) and Federal Enterprise Architecture Framework (FEAF). A similar review study complements previous studies with more recent EA Frameworks, including, amongst others, the Adaptive Integrated Digital Architecture Framework (AIDAF) [8]. AIDAF is based on modern development paradigms and emphasizes the need to support digital transformation processes. Furthermore, another review study identified the main EA implementation methodologies, which are concerned with the practices to model, develop, and maintain EA [9]. Reference architectures contribute to effective implementation of EA. A review study revealed a variety of 162 reference architectures, including RAMI4.0 and smart cities [10]. However, to the best of our knowledge, a reference architecture for S5.0 is not publicly available.

Principles are important elements of EA frameworks. A review regarding design principles indicated the lack of a general accepted definition and conceptual framework for enterprise architecture principles [11]. The author illustrates that principles exist on different levels (e.g., business, IT, data) and architectures essentially can be seen as networks of interrelated principles. The author presents a conceptual foundation for enterprise architecture principles in which principles are described as fundamental propositions related to the design, guiding both the construction and evaluation, and/or the representation, guiding the description and modelling, of architectures. Furthermore, the author differentiates generic and enterprise-specific principles.

2.3 Designing S5.0 Solutions

Bartoloni et al. [12] present a design approach for S5.0 solutions based on design thinking and the quadruple helix innovation framework. The proposed conceptual model is constructed around circular, multilateral S5.0 logic and applied to design and implement the SMARTAGE platform based on action design research in a single case study. The authors acknowledge the presence of design principles for I4.0, but emphasize that S5.0 solutions require a holistic approach and design principles for S5.0 are lacking. Furthermore, the authors stress that S5.0 is lacking empirical studies that provide understanding in the design and development processes and interactions between stakeholders to realize solutions that leverage I4.0 technologies to solve societal problems to benefit the entire society. The authors argue that the application of S5.0 logic is based on human-centric design, goes beyond the political-ideological concept, and should be based on an interdisciplinary approach to ensure that stakeholder needs are incorporated in the design and implementation of S5.0 solutions.

3 Methodology

This research is based on a SLR using the PRISMA 2020 guidelines [13]. Section 3.1 highlights the use of the PRISMA 2020 guidelines. Section 3.2 describes the topic identification process and scope of the SLR. Section 3.3 summarizes the screening process within the SLR. Section 3.4 describes the SLR in- and exclusion process.

3.1 Overview

The PRISMA 2020 guidelines are utilized to identify the topic and scope of the SLR, screen scientific databases for literature, and create a sample by extracting documents based on in- and exclusion criteria, as shown in Fig. 1. The PRISMA 2020 checklist is used to assess 27 items. Next, the use of PRISMA 2020 is visualized and described based on the guidelines and checklist.

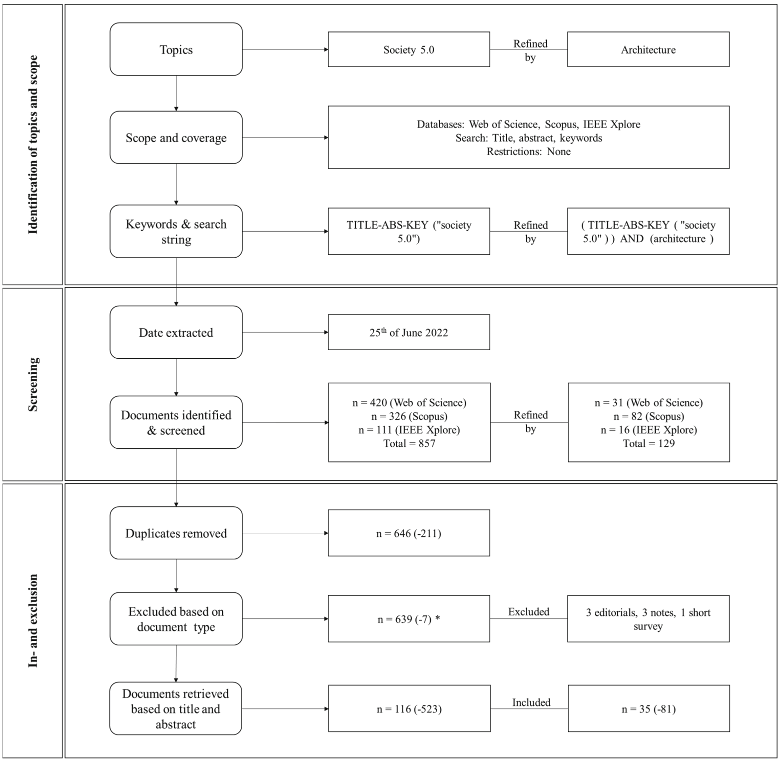


Fig. 1. Overview of the SLR regarding digital architectures under S5.0 (adapted based on [13]).

This study is identified as SLR. The abstract is checked using the checklist. The rationale of the SLR is described in the introduction together with the objective and research questions. In- and exclusion criteria are specified. The information sources and date of search are specified, including the search strategy and filters. The selection and data processing process are conducted independently by the first author without the use of automation tools. The study selection process and individual selected studies are described. The data items are listed and defined in this paper. The synthesis methods are described, visualized, and tabulated. The second author assessed the results and methodology. The risks of bias in studies are assessed for each individual study. The results of the syntheses are assessed to identify biases. Specific attention is given to the certainty of evidence. The risk of bias and certainty assessment are reported. The third author assessed the risks of bias. The research is conducted independently and received no funding or financial support. The authors declare no conflict of interest. All details are published in this paper and were subject to peer-review by multiple anonymous reviewers. The review is not registered.

3.2 Identification of Topic and Scope

The main topic of the SLR is S5.0 in relation to architecture to extend related studies [2, 3, 6] and book regarding S5.0 [4]. Three scientific databases were selected: Web of Science (https://www.webofscience.com/), Scopus (https://www.scopus.com/), and IEEE Xplore (https://ieeexplore.ieee.org/). No restrictions were set given the fact that S5.0 is a relatively new topic. The TITLE-ABS-KEY ("society 5.0") was used to query the databases for an initial search and TITLE-ABS-KEY ("society 5.0" AND "architecture") was incorporated to refine the results for the purpose of this study.

3.3 Screening

The three databases are queried on the 25th of June 2022 and resulted in a sample of 857 documents. The refined search yielded a sample of 129 documents (15% of the sample). Due to the relatively limited sample size all 857 documents were screened.

3.4 In- and Exclusion

First, 211 duplicates were removed from the sample. Next, 7 documents were excluded based on document type. This concerned 3 editorials, 3 notes, and 1 short survey. Then, 523 documents were excluded based on analysis of the title and abstract. Here, the main criteria were the relatedness to the main topic S5.0. and presence of a system or application. Documents that did not contain a clear relation to S.5.0 or lack a system or application in the title or abstract were excluded. This reduced the sample to 116 documents, which were obtained and analyzed in detail for the presence of either a digital and/or an enterprise architecture. Documents that lacked content regarding digital and/or enterprise architectures were excluded. Based on the analysis of the 116 documents, a total of 35 documents are included in this SLR.

4 Results and Findings

In this section, the results and findings of the SLR are presented to answer the research questions. Section 4.1 summarizes the results of the bibliometric analysis of the research topic S5.0. Section 4.2 lists and categorizes the identified digital architecture under S5.0. Section 4.3 analyzes the identified digital architectures under S5.0 in relation to EA frameworks and principles.

4.1 Bibliometric Analysis

VOSviewer (https://www.vosviewer.com/) version 1.6.18 is used to visualize, explore, and analyze bibliometric networks. The bibliometric database files from Web of Science and Scopus can be imported in VOSviewer. VOSviewer does not support the import of multiple databases. Therefore, the database with the most documents is used first for bibliometric analysis. The procedure is repeated for the other database.

Based on 420 documents from the Web of Science database, a keyword co-occurrence analysis is conducted to determine the relatedness of items based on the number of documents in which they occur. The sample includes 1857 keywords. Based on the minimum occurrence of 5, a total of 49 keywords meets the threshold. VOSviewer calculates the strength of the co-occurrence link with other keywords for each of the 49 keywords. The keywords that have the greatest link strength will be selected. The following keywords are manually removed "0", "Society 5", "Industry 4", and "Industry 5" to remove noise. The keyword "enterprise architecture" occurs 2 times and "architecture" is part of 12 keywords.

Figure 2 shows the visualization of keyword co-occurrence based on the bibliometric data. The scale is set to 2.0 to emphasize links with other keywords.

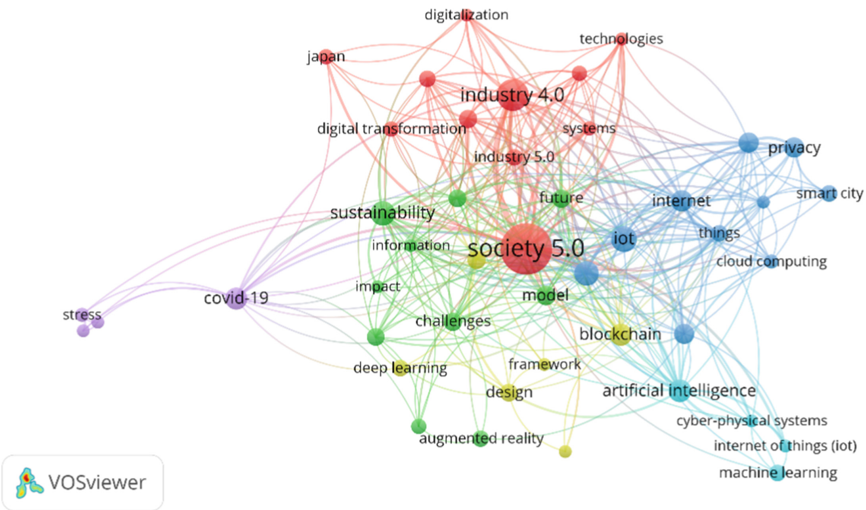


Fig. 2. Keyword co-occurrence >5 regarding S5.0 in Web of Science (image from VOSviewer).

Next to the visualization, VOSviewer identified six clusters based on the 45 keywords. Table 1 presents the discovered six clusters and keywords from Web of Science.

Table 1. Clusters and their keywords in the Web of Science database.

Cluster	Number of keywords	Keywords
1 (red)	11	Digital transformation, digitization, industry 4.0, industry 5.0, innovation, japan, smart society, society 5.0, systems, technologies, technology
2 (green)	10	Augmented reality, challenges, education, future, impact, information, management, model, performance, sustainability
3 (blue)	10	Big data, cloud computing, internet, internet of things, iot, privacy, security, smart cities, smart city, things
4 (yellow)	6	Blockchain, deep learning, design, framework, implementation, system
5 (purple)	4	Anxiety, covid-19, online learning, stress
6 (sky blue)	4	Artificial intelligence, cyber-physical systems, internet of things (iot), machine learning

A similar bibliometric analysis is conducted based on the 327 documents from the Scopus database, which revealed significant differences in terms of keywords (2328) and keyword co-occurrence of >5 (58 keywords), as shown in Fig. 3. The keyword “enterprise architecture” occurs 3 times and “architecture” is part of 12 keywords.

VOSviewer identified six clusters based on the 58 keywords. Table 2 presents the discovered six clusters and keywords in the Scopus.

Comparing both figures shows that using a single database introduces a risk of bias. The difference is also reflected in the relatively limited number of duplicates (-211). Therefore, the results must be cross-validated by (manually) merging both databases.

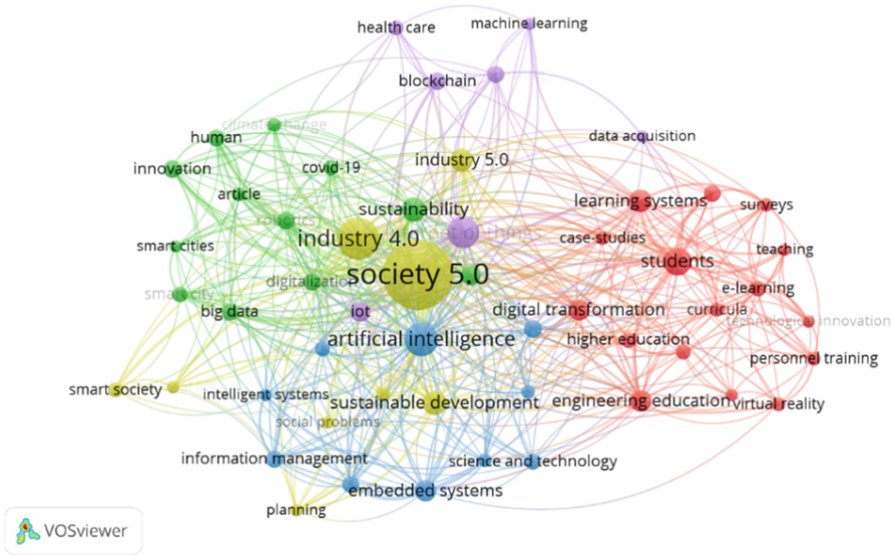


Fig. 3. Keyword co-occurrence >5 regarding S5.0 in Scopus (image from VOSviewer).

Table 2. Clusters and their keywords in the Scopus database.

Cluster	Number of keywords	Keywords
1 (red)	18	Biospherics, case-studies, curricula, data acquisition, digital transformation, e-learning, education computing, engineering education, higher education, indonesia, industrial revolutions, learning systems, personnel training, students, surveysm teaching, technological innovation, virtual reality
2 (green)	15	Artificial intelligence, cyber physical system, cyber physicals, cyber-physical systems, embedded systems, information management, internet of things, internet of things (iot), iot, philosophical aspects, science and technology, smart cities, smart city, social systems
3 (blue)	10	Article, big data, climate change, covid-19, digitalization, human, innovation, japan, robotics, sustainability
4 (yellow)	9	Digital technologies, industry 4.0, industry 5.0, planning, quality of life, smart society, social problems, society 5.0, sustainable development
5 (purple)	5	Block-chain, blockchain, deep learning, health care, machine learning
6 (sky blue)	1	Information and communication technologies

4.2 Digital Architectures Under S5.0

Based on the SLR and the selected 116 documents, a total of 37 digital architectures are identified. Table 3 lists and categorizes the digital architectures by domain, year and source(s). Figure 4 shows the publications containing digital architectures under S5.0 by year. Figure 5 categorizes digital architectures under S5.0 by domain.

Table 3. Identified digital architectures under S5.0.

ID	Digital architecture	Domain	Year	Source(s)
1	Group management system design	Generic	2017	[14]
2	Future energy system	Energy	2017	[15]
3	Cone-X-ion system	Retail	2018	[16]
4	Gurarda framework for smart cities	Smart cities	2018	[17]
5	Civil registration and population data system	Government	2018	[18]
6	Visual analytics framework for condition monitoring in cyber-physical systems	Manufacturing	2019	[19]
7	Digital healthcare platform	Healthcare	2020	[20, 21]
8	Cyber-physical system setup for predicting tool wear in machining	Manufacturing	2020	[22]
9	Facemask and physical distancing detection with alarm systems	Security	2020	[23]
10	Enhanced living environments with ambient assisted living	Living	2020	[24]
11	Societal security concept for society 5.0	Generic	2020	[25]
12	Human-machine system for finishing mill process	Manufacturing	2020	[26]
13	Learning health system	Healthcare	2020	[26]
14	Marketplace of services	Generic	2020	[27]
15	Human-centred management system for scheduling in operations and maintenance	Manufacturing	2020	[28]
16	Smart state net scheme	Smart cities	2020	[29]
17	Teleoperation system	Generic	2021	[30]
18	Secure hash algorithm (SHA)-256 accelerator	Generic	2021	[31]
19	AI-based communication-as-a-service	Generic	2021	[32]
20	AI-based heart monitoring system	Healthcare	2021	[33]
21	Intelligent health service	Healthcare	2021	[34]
22	AIoT based picking algorithm	eCommerce	2021	[35]
23	Robotics and digital platform (IoT)	Healthcare	2021	[21]
24	Reference architecture for framework for MNCRP	Generic	2021	[36]
25	530-mW multicore blockchain accelerator	Generic	2021	[37]

(*continued*)

Table 3. (*continued*)

ID	Digital architecture	Domain	Year	Source(s)
26	Anticipatory smart energy system	Energy	2021	[38]
27	Flexible/modular manufacturing system architecture	Manufacturing	2021	[39]
28	Digital twin for smart cities	Smart cities	2021	[40]
29	A generalized framework for tactile internet	Generic	2021	[41]
30	Shiojiri environmental data compilation platform and IoT network	Generic	2021	[42]
31	SMARTAGE platform	Healthcare	2022	[12]
32	Delay-tolerant public blockchain network in healthcare systems	Healthcare	2022	[43]
33	Cipher policy-attribute-based encryption for smart healthcare systems	Healthcare	2022	[44]
34	Dialog and recommender system	Retail	2022	[45]
35	A platform for integrated/connected government	Government	2022	[46]
36	Long-range real-time monitoring for precision irrigation and rural farming	Smart cities	2022	[47]
37	Cyber-physical network architecture for data stream provisioning in complex ecosystems	Smart cities	2022	[48]

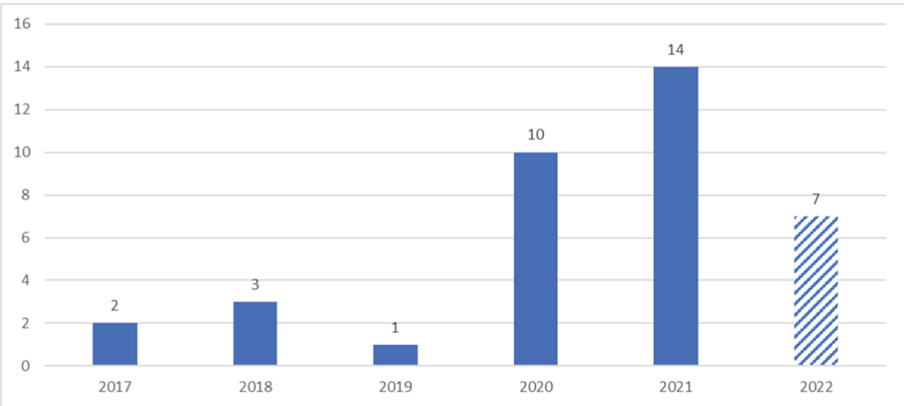


Fig. 4. Publications containing digital architectures under S5.0.

The list is numbered and sorted by year of publication, but is limited to published articles in the selected databases and the sample based on the in- and excluding criteria. Therefore, the list is non-exhaustive and should be seen as a starting point to identify digital architectures under S5.0. The sample presents digital architectures from

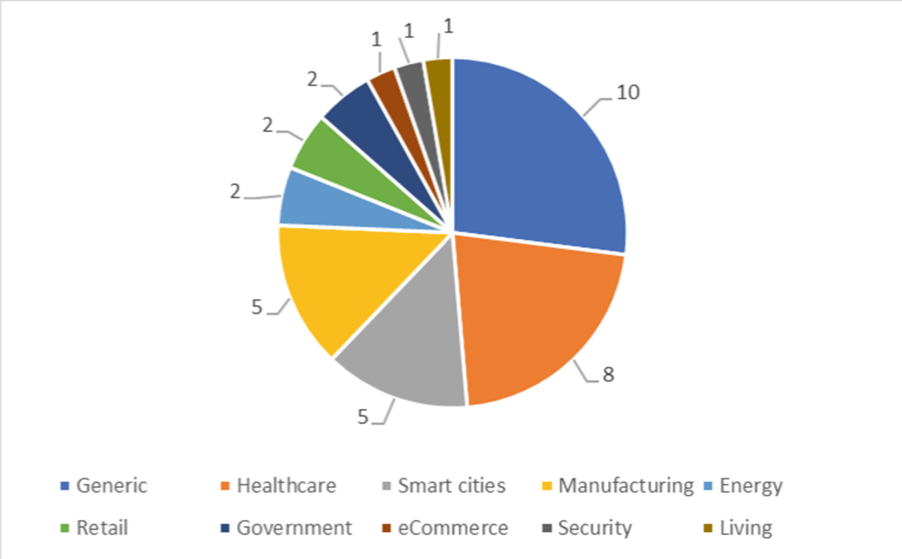


Fig. 5. Digital architectures under S5.0 by domain.

various domains. However, not all domains are represented (e.g., education, transportation, finance). Moreover, the certainty that the identified architectures are designed and constructed under S5.0 can not properly and solely be assessed based a SLR.

4.3 EA Frameworks and Principles

The identified digital architectures are analyzed using an EA perspective. Table 4 presents an additional overview of the representation(s) of the digital architectures under S5.0 and the use of EA framework(s) and principle(s). Figure 6 visualizes the use of representations of selected digital architectures under S5.0.

The majority of the digital architectures is based on conceptual and system representations. Quite some digital architectures contain a layered viewpoint and some present a framework. Furthermore, a variety of detailed representations are used, including specific layers, application components or algorithms. Only 2 publications do not contain any representation.

In terms of EA frameworks, 25 publications do not state the use of any EA framework. Most of the digital architectures are based on a domain specific approach, e.g. AIDAF in Healthcare [7, 23]. Interesting EA developments are reported in Indonesia using TOGAF [35]. None of the publications refer to a reference architecture for S5.0.

Table 4. Representation(s), use of EA framework(s), and principle(s) in digital architectures.

ID	Representation(s)	EA framework(s)	Principle(s)
1	System	Not stated	Not explicit
2	Conceptual; Layered; System; Reference	Smart Grid Architecture Model	Not explicit
3	Conceptual	ZEF framework	Not explicit
4	Conceptual; Framework; Living Lab	ASEAN Smart Cities Network; Garuda Smart City Framework (own contribution)	Living lab; Scientific method
5	System	Not stated	Not explicit
6	Framework; Conceptual; System; Lab	CPS framework (own contribution)	Not explicit
7	Conceptual; Framework; System	AIDAF; RAMI4.0	AIDAF; Agile; Design Thinking; GDPR
8	Conceptual; Prototype	Not stated	Not explicit
9	Algorithm	Not stated	Not explicit
10	Conceptual; Framework; Domain	Not stated	ELE-Industry 4.0; Human habitat innovation
11	Conceptual	Not stated	Crowd
12	System; Functional	Not stated	Not explicit
13	Conceptual	Not stated	Not explicit
14	Layered	Not stated	Not explicit
15	Conceptual; System; Functional	Not stated	Not explicit
16	Conceptual	Not stated	Society 5.0
17	Conceptual; Network; Infrastructure; Prototype	Not stated	Not explicit
18	System; Component; Data flow	Not stated	Not explicit
19	Layered; Conceptual; Network	Not stated	Not explicit
20	System; Algorithm	Not stated	Not explicit
21	Conceptual; Layered	Not stated	Not explicit
22	System; Information flow; Data; Algorithm	Not stated	AIoT
23	-	AIDAF; RAMI4.0	AIDAF

(*continued*)

Table 4. (*continued*)

ID	Representation(s)	EA framework(s)	Principle(s)
24	Conceptual; Layered; Reference	ArchiMate; Unified Architecture Framework	Modeling by Angelov; General architecture
25	Conceptual; System; Component; Data flow	Not stated	Not explicit
26	Conceptual; Layered; System; Data	Not stated	Society 5.0
27	System	Not stated	Digital Twin
28	System	Not stated	Gemini; Circular
29	Conceptual; System; Network; Reference	International mobile telecommunications framework; Tactile internet framework (own contribution)	Not explicit
30	System	Not stated	SDG
31	–	Quintuple helix framework	Human-centric design; Design thinking; Quintuple helix
32	Conceptual; Layered; Information flow	Not stated	Not explicit
33	Conceptual; Layered; Data flow	Not stated	Not explicit
34	System; Interaction	Not stated	Not explicit
35	Framework	TOGAF; Indonesian EA framework (own contribution)	IEA Development Method
36	Conceptual; System	Absolute innovation management framework	Principles of workplace; Design principles I4.0; Industry 5.0
37	Conceptual; Layered; System; Network; Algorithm	iFog framework (own contribution)	Clustering spatial-temporal data system

Similar to EA frameworks the use of principles is not stated in 21 publications. 2 publications [16, 26] refer to S5.0 principles, but it remains unclear how these principles are used for the design, construction and evaluation of S5.0 applications. 2 publications [7, 23] report the use of AIDAF principles in combination with design thinking and agile development principles and GDPR principles.

The synthesis is based on the analysis of selected published articles. There is a risk that the article does not (fully) report the use of EA frameworks, principles, and digital transformation processes. Therefore, the certainty can not be assessed properly and solely based on the SLR. Empirical research is required to validate the findings.

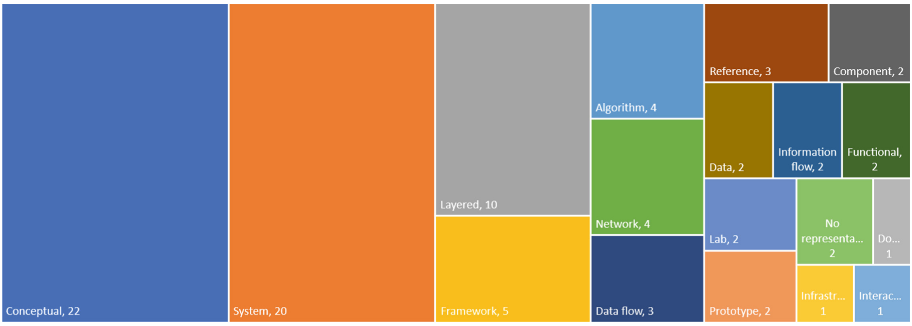


Fig. 6. Representations of digital architectures under S5.0. by type.

5 Discussion

Extending the scope of related SLRs, this research adds an EA perspective regarding S5.0 to the knowledge base by identifying 37 digital architectures. These digital architectures are analyzed based on their representation(s), the use of EA framework(s), and principle(s). The majority of digital architectures are based on a conceptual and/or system representation. Others are based on a layered architecture representation or represented as a framework. Some digital architectures focus on specific layers (e.g., network, process). EA frameworks (e.g., TOGAF, AIDAF) are used fairly limited and lack explicit principles. None of the publications contains, nor refers to, a reference architecture for S5.0.

The SLR results regarding S5.0 indicate a strong link with I4.0 and related technologies. This is in line with related work. On the contrary, the link to the SDGs is not clearly visible. However sustainability and sustainable development are among the keywords with a co-occurrence >5, none of the identified and analyzed digital architectures specifically addresses SDGs in relation to goal realization.

The bibliometric analysis in VOSviewer shows that the keyword "enterprise architecture" only occurs 2 times in the sample from Web of Science and 3 times in the sample from Scopus. The keyword "architecture" occurs in 12 keywords in both databases. There seems no direct link between the ideas and vision of S5.0 and the design, construction and evaluation of digital architectures. This supports our initial assumption that S5.0 and digital architectures are loosely coupled. This is where reference architectures and principles can be of value.

Taken together, the SLR and related work show that the interest in S5.0 is increasing and tends to evolve to an interdisciplinary topic. This makes S5.0 a relevant topic for the EA discipline. However, given the increasing number of publications, it is remarkable that S5.0 attracted little attention in the EA discipline. This can be explained by the fact that S5.0 is a relatively new research topic. Next to that, S5.0 extends the scope of individual enterprises and domains. This raises the question whether EA frameworks are suitable for S5.0. There are several reference architectures that guide the design of smart cities. Furthermore, interesting related work illustrates how TOGAF can be applied to create a governmental EA perspective regarding S5.0. The identified digital architectures can serve as a starting point to develop EA perspectives, but S5.0 needs to be further examined using disciplinary knowledge, exceeding the current scope regarding EA frameworks and principles. The design approach, discussed in Sect. 2.3, and habitat innovation framework, introduced in Sect. 2.1, can offer a starting point together with existing EA frameworks, principles and reference architectures, referred to in Sect. 2.2. Alternatively, existing domain specific approaches can be extended with guiding S5.0 principles, e.g. AIDAF for Healthcare under S5.0.

Based on the assessment of risks of bias and certainty, empirical research is needed to validate the findings of the SLR. Moreover, the current SLR revealed a bias in bibliographic analysis using a single database (which may effect related SLRs). Therefore, the results of this study must be cross-validated.

6 Conclusion

The main aim of this paper was to identify and analyze digital architectures that are (being) designed and constructed under S5.0. The SLR illustrates the increasing knowledge base regarding S5.0 and presence of a variety of digital architectures in different domains. More specifically, this study links S5.0 to the EA discipline and adds an EA perspective regarding S5.0 by identifying, listing, categorizing, and analyzing 37 digital architectures in different domains in terms of representation(s), use of EA framework(s), and principle(s). The main finding is that the digital architectures are based on conceptual and system representations and make fairly limited use of established EA frameworks and principles. Although some relevant and related reference architectures are present, a dedicated reference architecture for S5.0 is not found in this study based on the SLR in the three selected scientific databases.

From an EA perspective, scholars and practitioners can leverage this study and available disciplinary knowledge, including existing EA frameworks, methods, and tools, to support the design, representation, construction, implementation, and management of digital architectures under S5.0. EA frameworks can support the development of a holistic and systematic approach to support the lifecycle of digital architectures under S5.0. Given the broad scope of S5.0, a reference architecture may support the realization. Existing principles can guide the design and representation, but specific principles are required to ensure incorporation of the core ideas of S5.0 (in other countries).

Despite the systematic approach of this study, there are several limitations to address. The main limitation is that the current study is based on scientific literature in three selected databases in the English language. Extending the scope of related SLRs, this study focused on developing an EA perspective, but its scope is limited to EA frameworks and principles. The list of digital architectures provides an overview of digital architectures under S5.0, but is a non-exhausting list. Therefore, the list should be treated as a first starting point to develop EA perspectives regarding S5.0. Empirical research is required to validate the results and findings of the SLR. The bibliometric analysis of Web of Science and Scopus revealed a risk of bias in (existing) single database studies. Therefore, the results must be cross-validated by (manually) merging multiple databases. Due to practical limitations this cross-validation has not taken place. For this reason, the identified future work based on bibliographic coupling and co-citation analysis are not included in this SLR.

Additional research is required to develop a theoretical foundation for EA research regarding S5.0. First, the results of this SLR should be cross-validated by (manually) merging multiple databases. Then, depending on the outcomes, further analysis can take place regarding bibliographic coupling and co-citation. Based on related work and this SLR, existing guiding principles and reference models can be evaluated, or if needed developed or extended, to support the design, construction and evaluation of digital architectures under S5.0 and supporting digital transformation processes. For the broader adoption of S5.0 (in other countries), it is recommended to develop both general principles for S5.0 and specific principles (e.g., per country, domain and/or application). Furthermore, the development of a risk and security perspective is considered to be an interesting avenue for future research.

References

1. Cabinet Office: Society 5.0. [online]. https://www8.cao.go.jp/cstp/english/society5_0/index.html (2022). Last accessed: 16 Jul 2022
2. Shahidan, N.H., Latiff, A.S.A., Wahab, S.A.: Moving towards society 5.0: a bibliometric and visualization analysis. In: Gerber, A., Hinkelmann, K. (eds.) Society 5.0 2021. CCIS, vol. 1477, pp. 93–104. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86761-4_8
3. Roblek, V., Meško, M., Podbregar, I.: Mapping of the emergence of society 5.0: a bibliometric analysis. Organizacija **54**, 293–305 (2021). https://doi.org/10.2478/orga-2021-0020
4. Deguchi, A., et al.: Society 5.0 A People-centric Super-smart Society. Hitachi-UTokyo Laboratory (H-UTokyo Lab.) The University of Tokyo Bunkyo-ku, Tokyo, Japan. Springer open (2020). https://doi.org/10.1007/978-981-15-2989-4
5. United Nations: Transforming our world: implementing the 2030 agenda through sustainable development goal indicators [online]. Available from: https://sdgs.un.org/2030agenda (2016). Last accessed: 16 Jul 2022
6. Gampfer, F., Jürgens, A., Müller, M., Buchkremer, R.: Past, current and future trends in enterprise architecture—a view beyond the horizon. Comput. Ind. **100**, 70–84 (2018). https://doi.org/10.1016/j.compind.2018.03.006
7. Urbaczewski, L., Mrdalj, S.: A comparison of enterprise architecture frameworks. Issues in Inform. Syst. **VII**, 18–23 (2006). https://doi.org/10.48009/2_iis_2006_18-23
8. Júnior, S.H.D.L., Silva, F.Í.C., Albuquerque, G.S.G., de Medeiros, F.P.A., Lira, H.B.: Enterprise Architecture in Healthcare Systems: A systematic literature review. arXiv preprint arXiv:2007.06767 (2020). https://doi.org/10.48550/arXiv.2007.06767
9. Rouhani, B.D., Mahrin, M.N., Nikpay, F., Binti Ahmad, R., Nikfard, P.: A systematic literature review on enterprise architecture implementation methodologies. Inform. Softw. Technol. **62**, 1–20 (2015). https://doi.org/10.1016/j.infsof.2015.01.012
10. Garcés, L., et al.: Three decades of software reference architectures: a systematic mapping study. J. Syst. Softw. **179**, 111004 (2021). https://doi.org/10.1016/j.jss.2021.111004
11. Stelzer, D.: Enterprise architecture principles: literature review and research directions. In: Dan, A., Gittler, F., Toumani, F. (eds.) ICSOC/ServiceWave -2009. LNCS, vol. 6275, pp. 12–21. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16132-2_2
12. Bartoloni, S., et al.: Towards designing society 5.0 solutions: the new quintuple helix - design thinking approach to technology. Technovation **113**, 102413 (2022). https://doi.org/10.1016/j.technovation.2021.102413
13. Page, M.J., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ **372**, n71 (2021). https://doi.org/10.1136/bmj.n71

14. Prasetyo, Y.A., Arman, A.A.: Group management system design for supporting society 5.0 in smart society platform. In: 2017 International Conference on Information Technology Systems and Innovation (ICITSI), pp. 398–404 (2017). https://doi.org/10.1109/ICITSI.2017.8267977
15. Izui, Y., Koyama, M.: Future energy and electric power systems and smart technologies. IEEJ Trans. Electr. Electron. Eng. **12**(4), 453–464 (2017). https://doi.org/10.1002/tee.22436
16. Rahmandita, A., Fajar, A.N., Shofi, I.M., Girsang, A.S.: Analysis and design of XYZ integrated system based on service oriented architecture. In: 2018 International Conference on ICT for Smart Society (ICISS), pp. 1–5 (2018). https://doi.org/10.1109/ICTSS.2018.8549942
17. Tay, K., Supangkat, S.H., Cornelius, G., Arman, A.A.: The SMART initiative and the garuda smart city framework for the development of smart cities. In: 2018 International Conference on ICT for Smart Society (ICISS), pp. 1–10 (2018). https://doi.org/10.1109/ICTSS.2018.8549961
18. Kerlooza, Y., Setiawan, A., Asrianto, R.: Towards smart society: a study on multi-channel and public participation-based system architecture for civil registration and population data in Indonesia. In: 2018 International Conference on ICT for Smart Society (ICISS), pp. 1–5 (2018). https://doi.org/10.1109/ICTSS.2018.8549956
19. Villalonga, A., Castaño, F., Beruvides, G., Haber, R., Strzelczak, S., Kossakowska. J.: Visual analytics framework for condition monitoring in cyber-physical systems. In: 2019 23rd International Conference on System Theory, Control and Computing (ICSTCC), pp. 55–60 (2019). https://doi.org/10.1109/ICSTCC.2019.8885611
20. Masuda, Y., Zimmermann, A., Shepard, D.S., Schmidt, R., Shirasaka, S.: An adaptive enterprise architecture design for a digital healthcare platform : toward digitized society – industry 4.0, society 5.0. In: 2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW), pp. 138–146 (2021). https://doi.org/10.1109/EDOCW52865.2021.00043
21. Masuda, Y., Zimmermann, A., Sandkuhl, K., Schmidt, R., Nakamura, O., Toma, T.: Applying AIDAF for enabling industry 4.0 in open healthcare platform 2030. In: Zimmermann, A., Howlett, R.J., Jain, L.C., Schmidt, R. (eds.) KES-HCIS 2021. SIST, vol. 244, pp. 211–221. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-3264-8_20
22. Kondo, Y., Yamaguchi, M., Sakamoto, S., Yamaguchi, K.: A study on cyber-physical system architecture to predict cutting tool condition in machining. Int. J. Mech. Eng. Robot. Res. **9**(4), 565–569 (2020). https://doi.org/10.18178/ijmerr.9.4.565-569
23. Militante, S.V., Dionisio, N.V.: Deep learning implementation of facemask and physical distancing detection with alarm systems. In: 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE), pp. 1–5 (2020). https://doi.org/10.1109/ICVEE50212.2020.9243183
24. Caro Anzola, E.W., Mendoza Moreno, M.Á.: Enhanced living environments (ELE): a paradigm based on integration of industry 4.0 and society 5.0 contexts with ambient assisted living (AAL). In: García-Alonso, J., Fonseca, C. (eds.) IWoG 2020. LNB, pp. 121–132. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72567-9_12
25. Aldabbas, M., Xie, X., Teufel, B., Teufel, S.: Future security challenges for smart societies: overview from technical and societal perspectives. In: 2020 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), pp. 103–111 (2020). https://doi.org/10.1109/ICSGCE49177.2020.9275630
26. Sawaragi, T., Horiguchi, Y., Hirose, T.: Design of productive socio-technical systems by human-system co-creation for super-smart society. IFAC-PapersOnLine **53**(2), 10101–10108 (2020). https://doi.org/10.1016/j.ifacol.2020.12.2734
27. Olariu, S.: Smart communities: from sensors to internet of things and to a marketplace of services. In: SENSORNETS, pp. 7–18 (2020). https://doi.org/10.5220/0009430700070018

28. Foresti, R., Rossi, S., Magnani, M., Bianco, C.G.L., Delmonte, N.: Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. Engineering **6**(7), 835–846 (2020). https://doi.org/10.1016/j.eng.2019.11.014
29. Gurjanov, A.V., Zakoldaev, D.A., Shukalov, A.V., Zharinov, I.O.: The smart city technology in the super-intellectual Society 5.0. J. Phys.: Conf. Ser. **1679**(3), 032029 (2020). https://doi.org/10.1088/1742-6596/1679/3/032029
30. Maier, M.: 6G as if people mattered: from industry 4.0 toward society 5.0: (Invited Paper). In: 2021 International Conference on Computer Communications and Networks (ICCCN), pp. 1–10 (2021). https://doi.org/10.1109/ICCCN52240.2021.9522181
31. Tran, T.H., Pham, H.L., Nakashima, Y.: A high-performance multimem SHA-256 accelerator for society 5.0. IEEE Access **9**, 39182–39192 (2021). https://doi.org/10.1109/ACCESS.2021.3063485
32. Ghosh, T., Saha, R., Roy, A., Misra, S., Raghuwanshi, N.S.: AI-based communication-as-a-service for network management in society 5.0. IEEE Trans. Netw. Serv. Manage. **18**(4), 4030–4041 (2021). https://doi.org/10.1109/TNSM.2021.3119531
33. Dampage, U., Balasuriya, C., Thilakarathna, S., Rathnayaka, D., Kalubowila, L.: AI-based heart monitoring system. In: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1–6 (2021). https://doi.org/10.1109/GUCON50781.2021.9573888
34. Al Mamun, S., Kaiser, M.S., Mahmud, M.: An artificial intelligence based approach towards inclusive healthcare provisioning in society 5.0: a perspective on brain disorder. In: Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q., Zhong, N. (eds.) BI 2021. LNCS (LNAI), vol. 12960, pp. 157–169. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86993-9_15
35. Muslikhin, M., Horng, J.-R., Yang, S.-Y., Wang, M.-S., Awaluddin, B.-A.: An artificial intelligence of things-based picking algorithm for online shop in the society 5.0's context. Sensors **21**(8), 2813 (2021). https://doi.org/10.3390/s21082813
36. Siriweera, A., Naruse, K.: Survey on cloud robotics architecture and model-driven reference architecture for decentralized multicloud heterogeneous-robotics platform. IEEE Access **9**, 40521–40539 (2021). https://doi.org/10.1109/ACCESS.2021.3064192
37. Tran, T.H., Pham, H.L., Phan, T.D., Nakashima, Y.: BCA: a 530-mW multicore blockchain accelerator for power-constrained devices in securing decentralized networks. IEEE Trans. Circuits Syst. I Regul. Pap. **68**(10), 4245–4258 (2021). https://doi.org/10.1109/TCSI.2021.3102618
38. Darani, Z.H., Taheri Demne, M., Zanjirani, D.M., Zackery, A.: Conceptualization of a new generation of smart energy systems and the transition toward them using anticipatory systems. Eur. J. Futures Res. **9**(1), 1–17 (2021). https://doi.org/10.1186/s40309-021-00184-1
39. Kalogeras, G., Anagnostopoulos, C., Alexakos, C., Kalogeras, A., Mylonas, G.: Cyber physical systems for smarter society: a use case in the manufacturing sector. In: 2021 IEEE International Conference on Smart Internet of Things (SmartIoT), pp. 371–376 (2021). https://doi.org/10.1109/SmartIoT52359.2021.00069
40. Mylonas, G., Kalogeras, A., Kalogeras, G., Anagnostopoulos, C., Alexakos, C., Muñoz, L.: Digital twins from smart manufacturing to smart cities: a survey. IEEE Access **9**, 143222–143249 (2021). https://doi.org/10.1109/ACCESS.2021.3120843
41. Mourtzis, D., Angelopoulos, J., Panopoulos, N.: Smart manufacturing and tactile internet based on 5g in industry 4.0: challenges, applications and new trends. Electronics **10**(24), 3175 (2021). https://doi.org/10.3390/electronics10243175
42. Narvaez Rojas, C., Adolfo, G., Peñafiel, A., Loaiza Buitrago, D.F., Tavera Romero, C.A.: Society 5.0: a japanese concept for a superintelligent society. Sustainability **13**(12), 6567 (2021). https://doi.org/10.3390/su13126567

43. Ghosh, T., Roy, A., Misra, S.: B2H: enabling delay-tolerant blockchain network in healthcare for Society 5.0. Comput. Netw. **210**, 108860 (2022). https://doi.org/10.1016/j.comnet.2022.108860
44. Ghosh, T., Roy, A., Misra, S., Raghuwanshi, N.S.: CASE: a context-aware security scheme for preserving data privacy in IoT-enabled society 5.0. IEEE Internet of Things J. **9**(4), 2497–2504 (2022). https://doi.org/10.1109/JIOT.2021.3101115
45. Sakai, K., Nakamura, Y., Yoshikawa, Y., Ishiguro, H.: Effect of robot embodiment on satisfaction with recommendations in shopping malls. IEEE Robot. Autom. Lett. **7**(1), 366–372 (2022). https://doi.org/10.1109/LRA.2021.3128233
46. Saiya, A.A., Arman, A.A.: Indonesian enterprise architecture framework: a platform for integrated and connected government. In: 2018 International Conference ISS, 2018, pp. 1–6 (2022). https://doi.org/10.1109/ICTSS.2018.8549990
47. Singh, D.K., Sobti, R.: Long-range real-time monitoring strategy for Precision Irrigation in urban and rural farming in society 5.0. Comput. Ind. Eng. **167**, 107997 (2022). https://doi.org/10.1016/j.cie.2022.107997
48. Okafor, K.C., Ndinechi, M.C., Misra, S.: Cyber-physical network architecture for data stream provisioning in complex ecosystems. Trans. Emerg. Telecommun. Technol. **33**(4), e4407 (2022). https://doi.org/10.1002/ett.4407

Towards a Framework for Context Awareness Based on Textual Process Data: Case Study Insights

Aleksandra Revina[1,2(✉)], Nina Rizun[3], and Ünal Aksu[4]

[1] Technical University of Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
[2] Technische Hochschule Brandenburg, Magdeburger Str. 50, 14770 Brandenburg an der Havel, Germany
revina@th-brandenburg.de
[3] Gdansk University of Technology, G. Narutowicza 11/12, 80-233 Gdansk, Poland
nina.rizun@pg.edu.pl
[4] Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands
u.aksu@uu.nl

Abstract. Context awareness is critical for the successful execution of processes. In the abundance of business process management (BPM) research, frameworks exclusively devoted to extracting context from textual process data are scarce. With the deluge of textual data and its increasing value for organizations, it becomes essential to employ relevant text analytics techniques to increase the awareness of process workers, which is important for process execution. The present paper addresses this demand by developing a framework for context awareness based on process executions-related textual data using a well-established layered BPM context model. This framework combines and maps various text analytics techniques to the layers of the context model, aiming to increase the context awareness of process workers and facilitate informed decision-making. The framework is applied in an IT ticket processing case study. The findings show that contextual information obtained using our framework enriches the awareness of process workers regarding the process instance urgency, complexity, and upcoming tasks and assists in making decisions in terms of these aspects.

Keywords: Context awareness · Textual data · Text analytics · Business process management

1 Introduction

Business processes are one of the key assets of organizations through which they provide products and services to their customers. To lead to an outcome that is valuable for customers, activities, events, and decisions are instrumented within business processes by process workers [1]. Moreover, how such process elements will be instrumented is influenced by the context of the processes, i.e., the environmental properties [2]. For example, a typical order processing is likely influenced by various situational factors

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 25–41, 2023.
https://doi.org/10.1007/978-3-031-26886-1_2

necessary for process execution, such as weather, terms and conditions, customer and order type. Hence, business processes are bound to their context. In fact, many Business Process Management (BPM) projects are known to fail as they base on standard guidelines and schemes neglecting the context [3]. In empirical research, context is frequently viewed as an outside threat that needs to be controlled or removed [4]. There are multiple attempts in information systems research to develop solutions to control the context, such as promoting generalizability [7], increasing causality, and improving robustness [8]. However, one of the most profound and prominent theoretical findings in information systems research is that context matters [4].

In line with the importance of context in business processes, to be able to execute processes appropriately, workers need to comprehend their context. Specifically, decision-making of humans while performing processes requires awareness and processing of all possible information about the process context. However, humans are constrained in their ability to effectively comprehend and process large amounts of data. As the majority of the information on process context is textual data and unstructured, how to facilitate contextualization of such data is considered rather challenging [5, 6].

BPM literature introduces a number of methods to deal with context. For example, in the recent work [9], building on a special type of directed graph, the authors present an approach incorporating contextual information in the analysis and visualization of the process execution data, i.e., event log. Other research focuses on the development of holistic frameworks [10], ontologies [11], taxonomies, and specific event log- [12] and business process model-based [6] methods. Despite the abundance of these approaches, the potential of textual data for context awareness and decision-making support of process workers in process execution has not been comprehensively explored [13]. Human workers have limited ability to process large amounts of data [14]. Hence, their awareness of certain process-relevant information, such as customer-related or expected process complexity, naturally supports efficient process execution. Additionally, most studies lack addressing the practical implementation of context awareness approaches and focus on conceptual work [9]. However, providing or using publicly available tools and techniques as well as specifying the implementation details are important aspects of the work reproducibility and value for research and practice.

In this paper, we propose a framework for context awareness, explicitly focusing on textual data *related to process executions*. In particular, we focus on linguistic features of textual data (such as syntactic structure, meaning, style, text parts, word choice, and order) that can provide process context and assist process workers in executing processes in a proper way. We take an established BPM context model as a basis [2] and adopt it for leveraging textual data on context awareness. Specifically, we enhance the model, particularly its external, internal, and immediate context layers, with common text analytics techniques. This way, we aim to "practically implement" or operationalize the model, which forms our theoretical and methodological contribution. Practical contributions are demonstrated in an IT ticket processing case study. The awareness regarding the process instance *urgency*, *complexity*, as well as *expected tasks* can help process workers to *prioritize* their work, *facilitate task assignment* and *resource allocation* in the short run. In the long run, it may be beneficial for successful and fast process executions increasing the satisfaction of multiple stakeholders, e.g., managers, process workers, and customers.

Further, as we use a well-established comprehensive context model and common text analytics techniques, we believe that the framework can be applied in several real-life settings and domains with justified effort.

The remainder of the paper is as follows. Section 2 summarizes the related literature on context awareness in BPM. The research methodology is explained in Sect. 3. Section 4 describes how we conceptualize the framework for context awareness based on textual process data. Then, we provide the case study-based insights, evaluate the framework, and discuss the findings and limitations in Sect. 5. Finally, in Sect. 6, we draw conclusions and present ideas for future work.

2 Related Work

The context of processes and information about such context received much attention in BPM [15, 16]. At the same time, detecting and incorporating contextual factors in processes is considered rather challenging [17]. These factors may be encountered at different levels, and their number and range can be diverse [2]. Moreover, they can be close to the process itself, for example, a minimal time required to execute a process or far beyond, like country import regulations. Therefore, developing context-aware solutions requires a profound knowledge of internal and external factors impacting the processes [18, 19]. In this regard, various context awareness approaches in BPM are suggested for detecting contextual factors. We analyzed state-of-the-art studies about context awareness in BPM aligned with our goal. With this, we aim to identify a comprehensive BPM context model serving as a basis for our framework. We elaborate on the studies[1] that are relevant to our framework development.

The research on contextual BPM is still in its early stages [15] and demands more context awareness inclusion in BPM method design and exploration [20]. However, we could identify both rather extensive, like [15], and very specific, such as [12], approaches to context awareness. In recent context awareness studies, various topics are covered, for example, process modeling [21], decision-making [22], process mining [23], IoT [6], and cloud computing [24].

At the same time, automated analyses of the language and text characteristics are valuable in various settings [25]. Text generated as communication via emails, chats, social media, and documents can naturally imply rich information on different contextual factors. Hereby, big data analytics in general [26] and text analytics in particular [27] have become popular techniques to extract contextual information from large amounts of textual data. In BPM, we observe that context awareness approaches inherently consider textual data. However, this consideration is prevailingly limited to textual data employed in process models [28], event logs [12], and ontologies [6]. Moreover, no studies suggest the linking of various semantic aspects inherent in textual data to context types. For example, one of the latest studies on BPM context [20] proposes linking various BPM contexts to appropriate BPM methods leaving the contextual aspects and textual data focus out of scope.

[1] The overview of the studies and literature search process can be found on the Github page.

Hence, our study aims to address these shortcomings while developing the framework. To base our framework development, we searched for the groundwork in the related literature, i.e., a BPM context model that could incorporate varying multilayered knowledge inherent in textual data [29]. In this regard, the context model by [2] is one of the earliest efforts in BPM context awareness, providing a foundation for future, more focused research [30, 31]. It represents a comprehensive universal taxonomy as an initial reference for process contextualization in organizations. Unlike other approaches, the model is characterized by a layered structure, each layer (immediate, internal, external, environmental) having a broader coverage. For example, whereas the immediate layer covers the aspects influencing process execution, like data and resources, the environmental layer addresses the aspects outside the business network of an organization, such as national policies [2]. These layers are intertwined so that the innermost layer gathers data from all the outer layers. Such a structure provides the necessary flexibility and semantics to integrate various knowledge extracted from textual data.

Nonetheless, the selected BPM context model [2] reveals certain limitations. For example, it provides only a high-level construct that lacks a strategy on how to practically implement the model in the sense of necessary data, analysis methods, and techniques, i.e., make it operational for an organization. Further, the demonstration of how a process worker can benefit from the model is also missing. The next section describes the methodology used to address these limitations while developing the framework.

3 Methodology

Taking Design Science Research Methodology (DSRM) as a basis [32], this study aims to build a framework for context awareness based on textual data. In the first phase of DSRM, we define the problem we focus on as follows: (1) process workers must be aware of the process context to be able to carry out the business process properly, (2) textual data inherently contains such contextual information, (3) however, workers have limited capability to comprehend textual data, especially in large amounts.

In the second phase, we set the objectives of the envisioned framework. Relying on solid prior research in this context [2] and dealing with its limitations, the envisioned framework aims to analyze available textual data relevant for process execution using various text analytics methods. With this, process-related insights can be extracted at the immediate, internal, and external contextual layers [2] and provided to process workers in a comprehensive manner.

In the third phase, we employ conceptual modeling. It serves the purpose to abstractly represent specific aspects of a domain with the help of graphical representations [33]. In this paper, we use conceptual modeling to develop our framework based on the context of a selected organization. As the context itself may imply an endless amount of information, the decision was made to select an organization interested in such a solution and involve its experts in the framework development. Hence, the framework has been conceptualized in an industrial IT Service Management (ITSM) Change Management (CHM) [34] process from a large international telecom provider. The CHM department of the company started a project to get insights into its CHM operations using the textual data of IT tickets issued to perform changes in IT products and services offered by the company. The complexity of these change requests varies from clearly formulated simple requests such as "*please grant the user XYZ on application Y admin rights*" to rather complex ones like "*install the hotfixes on XXX production environment: YYY, ZZZ, HHH, XXX MUST NOT BE RUNNING during the fix installation, expected time up to ten hours!!!!!! NOTE pre and post installation tasks*". All the requests must be entered into a sophisticated IT ticket processing system. Hereby, estimation of complexity, priority, risks, breakdown to tasks, identification of teams, and assignment of tasks take place. To achieve data-driven service delivery and provide real-time decision-making support for the workers, the CHM department declared an interest in creating context awareness solutions using textual data.

As this setting is highly relevant to our framework, we conceptualize it in the CHM department. Hereby, we worked with 12 domain experts (CHM managers and coordinators) who handle IT tickets daily and are well-versed in this process. In a one-day workshop, the experts provided us with the necessary knowledge of the problems related to IT ticket processing. Moreover, they explained the information contained in textual data, which is important for and related to the process context.

In the fourth phase, we use the case study to evaluate our framework and demonstrate its applicability. We use the real-world IT ticket textual data and related process goals to showcase the value of the framework for process worker decision-making support.

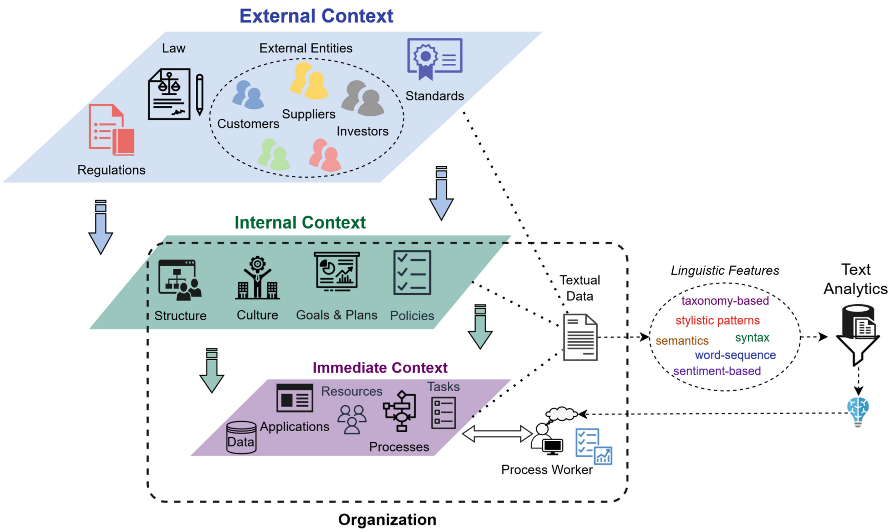


Fig. 1. Design of a framework for context awareness

4 Conceptual Framework Development

As presented in the related work section, Rosemann et al. introduce four layers of the model serving as a basis for our framework: immediate, internal, external, and environmental [2]. We take the first three context layers and enrich them with text analytics and process worker support in decision-making. We exclude the environmental layer, i.e., society, nature, and technological developments, as being far from process operations. According to [2], this layer addresses weather, strikes, policies, and work norms. Although this information can be considered in our framework, impactful cases involving such information are rare compared to others. Hence, we propose considering the environmental layer as a part of our future work.

The framework design reveals the exemplary elements in each layer based on [2] and enriching elements, i.e., text analytics, linguistic features, and process worker (see Fig. 1). Below, we elaborate on the conceptual framework layer by layer, starting with the external context layer. While doing that, we base the elaborations on our previous work [13, 35, 36]. The main reason is that, in this work, we showed the textual data potential in decision-making and awareness of process workers.

As shown in Fig. 1, the external layer includes concepts such as law, regulations, standards, and external entities, i.e., customers, suppliers, and investors. According to the experts from the CHM department of the organization where the framework is conceptualized, the most significant external layer information is related to the customers, i.e., the authors of the tickets. Further, at the internal layer addressing the organization or department-specific information, such as goals, plans, and policies, ticket processing implicit guidelines would play an important role. Finally, the immediate layer deals with the process execution and involves applications, resources, tasks, and process instances. While the experts consider all this information important, we focus primarily on the tasks necessary to process an IT ticket due to the organizational privacy policy.

4.1 Framework Layers

External Context. As presented in Fig. 1, the external context represents the outer layer of our framework comprising regulations, laws, standards, and external entities. Referring to a definition of a process [1], the ultimate goal of any process is bringing value to its customers. Knowledge about customers, essential for building sustainable businesses [37], naturally belongs to the external context layer and has also been identified as the most relevant for the organization by the experts. Hereby, sentiment analysis allows extracting subjective, i.e., customer-related, information such as attitude, opinion, and emotions from the text [38]. This can help process workers better understand the customer's pain points and react accordingly while executing the processes. Hence, we suggest sentiment analysis as an operational and easily applicable approach to obtain external information regarding customers.

In brief, at the external context layer, with the help of sentiment analysis, the process worker receives important latent information regarding the author of the IT ticket, i.e., the customer anxiety level. This knowledge extends the awareness with the customer context element. Hence, the worker can assess the urgency and significance of the request, which facilitates the planning. For example, if the number of urgent tickets gets high, a delegation of work or involvement of other colleagues can occur.

Internal Context. The internal context denotes the middle layer of the framework. It includes information about the internal organizational environment affecting process execution, such as a company or department structure, goals, and work plan (see Fig. 1). In BPM, knowledge, i.e., awareness of the organization- or department-specific rules, can be regarded as the next level of decision-making support closer to direct process operations. The extraction of this information is known to be achieved with the help of text analytics related to organizational knowledge management, like lexicons, thesauri, taxonomies, and ontologies [39]. Such knowledge representation can not only help experienced workers to manage the processes efficiently and faster but also enable novice workers to execute any process instance. Thus, we propose using taxonomies to include organizational knowledge in the decision-making support of process workers.

In short, at the internal context layer, the process worker gets an understanding of the process cognition in the organizational unit context based on the mutually agreed meaning of keywords formalized in the form of taxonomy. Then, using this knowledge enriched by personal, contextual experience, the worker can estimate the complexity of the effort and time needed to execute the process and its tasks.

Immediate Context. The immediate context constitutes the innermost layer of our framework. According to [2] and Fig. 1, the immediate context includes those aspects directly assisting in process execution, for example, information on required data, organizational resources, activities, IT, and applications. In its meaning, the immediate context is somewhat similar to the internal context. The distinction is its proximity to the direct process execution. Hence, we draw on the same considerations of the internal context in our choice of text analytics technique. Hereby, we highlight the difference related to the thematic specificity of the text analytics commonly applied for knowledge management, i.e., lexicons, thesauri, and taxonomies. In the immediate context, this thematic specificity should reflect tasks of processes that should be realized in the taxonomic approach.

Thus, at the immediate context layer, the process worker gets an understanding of those tasks and their content based on the keywords grouped in the form of taxonomy. Using this knowledge enriched by personal, contextual experience, the worker can plan specific tasks and resources needed to execute them.

4.2 Decision-Making Support Enabled by the Framework

After getting a meaningful representation of the three context layers, we elaborate on the solution for decision-making support of a process worker: (1) assessment of the process instance based on the textual data related to the process execution in terms of (i) the customer-related information (sentiment, external layer) and (ii) unit-specific knowledge (taxonomies, internal and immediate layers) related to the processing of the similar process instances, (2) based on (1), recommendation of possible actions. Together with the experts, we searched for a simple technique enabling us to integrate all the knowledge from three context layers within one text analytics approach. Due to their common usage and implementation simplicity, association rules represent an appropriate technique to enable such integration.

To sum up, based on the information collected in all the layers with the help of association rules, the process worker gets (i) an immediate comprehension of the current situation, i.e., process instance urgency, complexity in terms of time and resources, and (ii) a recommendation of tasks relevant for the process execution. Using this recommendation, the worker is able to adequately assess the situation and plan the implementation of the ticket.

5 Case Study

We used both ticket and corresponding task textual descriptions from the same organization to evaluate and demonstrate the applicability of the framework in a case study. After cleaning and preprocessing, the final dataset is comprised of 4623 entries. The time of dataset covers prevailingly the first half of 2019. In the case study, experiments were performed using Python 3.6, and the association rules were implemented in R, using the arules package. As further important textual data sources, we used ITIL handbooks and process descriptions in the company. These data have been used to develop case-study-specific vocabularies and taxonomies necessary in each context layer.

For illustration, we provide an anonymized IT ticket example from our case study. A CHM worker receives the following customer request via email: "*Dear colleagues, we need a service pack installation on production SAP HANA. Online installation is possible. Kind regards, XXX*". This request is entered into the IT ticketing system, as a rule, in its original form. However, it might be slightly modified or extended by the CHM workers. Afterward, while planning the ticket execution, the CHM worker breaks down the ticket into separate tasks. In the example, the following three tasks were planned: "*execute SAP HANA service pack installation on YYY*", "*4EP*", "*QA task, QA task will include healthcheck validation by ZZZ, check of logs, check of application*". Below, we explain the application of the proposed framework layer by layer and illustrate it using the example.

5.1 External Layer

In line with the consideration of the context at the external layer (see Sect. 4.1), we deploy a lexicon-based context-specific Business Sentiment (BS) to measure the "emotional" component, or level of anxiety, implied by the customer in the request description. As a rule, standard lexicons do not function well in domain-specific applications [40]. This prompted us to create a domain-specific BS lexicon utilizing the state-of-the-art VADER [41] and Latent Dirichlet Allocation (LDA) algorithm [42]. For details, we refer to [36] and the Github page. Using VADER, we, *first*, identify the emotionally loaded keywords and expressions and create the BS lexicon. Afterward, we enrich the BS lexicon with the keywords obtained from the LDA implementation. To do so, two sources are used: (1) IT ticket texts and (2) CHM process descriptions from the ITIL handbook. Each keyword is associated with a positive, negative, or neutral sentiment. Keywords with valence scores greater than 0 are considered positive, whereas those with less than 0 are marked as negative. Other keywords are denoted as having a neutral sentiment. *Second*, we compute the normalized total score of BS keywords with the pre-assigned valence and unique significance markers (syntactic and semantic intensifiers) for each ticket text in the dataset. *Third*, using the CHM workers' feedback, threshold rules are developed. *Fourth*, based on the normalized score and threshold rules, the BS is measured as the customer anxiety level for each ticket on the qualitative scale of normal, moderate, and severe anxiety [45].

In the motivating example, we observe the BS lexicon keywords with neutral valence (*dear, kind regards*) and no syntactic and semantic intensifiers, i.e., normal anxiety.

5.2 Internal Layer

Based on the contextual information considerations at the internal layer in Sect. 4.1, we apply the taxonomic approach. Its realization is accomplished in two steps. *First*, we build a hierarchical taxonomy to determine a process cognition level. Thus, we call it Decision-Making Logic (DML) taxonomy. With this, we aim to discover the decision-making character of activities inherent in processes. We differentiate between the following three DML levels: routine, semi-cognitive, and cognitive. The most significant keywords in the IT ticket texts are identified using LDA. Combined with the CHM workers' feedback, we organize these keywords into one of the three DML levels. For details of DML taxonomy building, we refer to [35] and the Github page. Accordingly, the internal context is realized by assigning the contextually associated keywords to the DML levels based on the knowledge intensity and process complexity [43].

In our motivating example, the keyword *online* indicates a routine, i.e., simple, activity as no service outage needs to be planned for this ticket. The keyword *colleagues* refers to the semi-cognitive level as one needs to have a contextual understanding and know the history of similar requests to determine who will process the ticket. *Second*, we calculate the relative occurrence of detected keywords in the IT ticket text and determine the DML level using the context-specific threshold rules defined by the case study employees. Our motivating example reveals four routine *online, install, pack, need* ($4/7 = 0.57$) and three semi-cognitive ($3/7 = 0.43$) *colleague, service, production* keywords. Thus, the overall DML level of such a ticket is routine.

5.3 Immediate Layer

According to the conceptualization of the immediate layer in Sect. 4.1, to get a meaningful representation of task sets, we develop a typology of ticket types and task types and subtypes in the form of a hierarchical taxonomy [44]. Under ticket and task type, we infer an actual activity that the customer requests. In the motivating example, the type of ticket is *install*. *First*, to build a ticket and task types taxonomy, we extract topics from the ticket and task descriptions using LDA. Afterward, we fine-tune the topics with the involved CHM workers. In the case of tasks, we used the IT ticketing system manual, where three types of tasks are distinguished: (1) projected service outage (PSO), (2) implementation of the ticket itself, and (3) quality assurance (QA). The PSO task means the service disconnection for performing requested changes. The implementation task directly references activities such as installation, update, and migration. QA is necessary for ensuring service level quality, for example, the four-eyes principle (4EP). As a result, ticket types and task types (including subtypes) are organized into one hierarchical taxonomy. *Second*, we label tickets and tasks in our dataset with the types and subtypes using the developed taxonomy. For details on typology building, we refer to the Github page.

In the motivating example, the ticket type *installation* reveals one implementation task (subtype *installation*), two QA tasks (four-eyes principle, i.e., *4EP*, and *healthcheck* subtypes), and no PSO task since the installation can be executed online.

5.4 Decision-Making Support Enabled by the Framework

As suggested in Sect. 4.1 and shown in Table 1, we enhance the association rules with the information obtained at the internal and external layers, i.e., DML/process cognition and BS/customer anxiety values. In particular, the association rules are applied to determine possible sets of task types and subtypes (rule body, i.e., consequent) based on the ticket type (rule head, i.e., condition). For details of the implementation and more examples, we refer to the Github page.

In Table 1, using our motivating example, row 1 represents association rules based on ticket types and task types and subtypes. Our ticket of type *installation* is predicted to have three tasks with the support of 7%: *implementation* (subtype *installation*) and two *QA* tasks (subtype *4EP* and *healthcheck*). It means that such a pattern occurs in 7% of the cases, i.e., rows in the dataset. The confidence of 55% shows how frequently the rule head *{Ticket: Installation}* appears among all other rows containing the rule body *{Implementation: Installation, QA: 4EP, QA: Healthcheck}*. The rows below indicate the contextual enhancements of the same ticket, i.e., ticket type (rule head) is enhanced with DML/process cognition and BS/customer anxiety values. Performing this experiment, we could make the following main observations: (1) on average, when adding the contextual information, the support and confidence values increase, (2) using the support and confidence values, the most frequent patterns can be identified, (3) the influence of specific contextual information in the rule head, i.e., DML/process cognition and BS/customer anxiety, on the task sets in the rule body can be identified. Based on this information, the process worker can assess the urgency of the customer request (ticket) and how to better deal with/contact that customer, estimate the ticket complexity, and get a recommendation of task sets needed to process the ticket.

Table 1. Association rules based on motivating example

No	Association rule	Support (%)	Confidence (%)
1	{Ticket: Installation} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	7	55
2	{Ticket: Installation (routine)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	13	88
3	{Ticket: Installation (routine, normal)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	12	89
4	{Ticket: Installation (routine, moderate)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck}	10	83
5	{Ticket: Installation (semi-cognitive)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup}	10	71
6	{Ticket: Installation (semi-cognitive, normal)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup}	9	70
7	{Ticket: Installation (semi-cognitive, moderate)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup}	11	72
8	{Ticket: Installation (cognitive, normal)} => {Implementation: Installation, QA: 4EP, QA: Healthcheck, QA: Backup, QA: Test}	8	66

In Table 2, we summarize all the steps and inputs necessary to apply our framework, from data collection and preprocessing to developing decision-making support.

Table 2. Framework for context awareness based on textual data. Overview of steps

1. Data collection and preprocessing
Input: textual data serving as input to a process (i.e., containing requests), tools: standard NLP processing software, e.g., Python and NLTK library *Processing*: 1) special preprocessing (retaining capitalization, exclamation and question marks, specific symbols), 2) standard preprocessing (removal of numbers, special symbols, punctuation, converting to lowercase, stemming), *Output*: files 1) and 2) with preprocessed textual data
2. External context knowledge extraction
Input: file 1), BS lexicon, threshold rules for BS level assignment, tools: Python, NLTK *Processing:* identification of BS keywords and their valence, intensifiers, calculation of the normalized total score, BS level assignment *Output:* file 3) with BS total scores, normalized total scores, assigned BS level for each textual entry
3. Internal context knowledge extraction
Input: file 2), DML taxonomy, threshold rules for DML levels assignment, tools: Python, NLTK *Processing*: identification of DML keywords, calculation of the relative occurrence of the keywords of each category, DML level assignment *Output:* file 4) with DML keywords and assigned DML for each textual entry

(*continued*)

Table 2. (*continued*)

4. Immediate context knowledge extraction
Input: file 2) enriched with textual data related to process execution (i.e., containing tasks or activities), task typology taxonomy, tools: Python, NLTK *Processing:* identification of ticket types, task types and subtypes based on the principle of the maximum relative distribution *Output:* file 5) with ticket type, task type and subtype keywords/expressions and their number, assigned ticket type, task type and subtype for each task text
5. Decision-making support
Input: files 3), 4), 5), tools: R (arules package) *Processing:* application of association rules *Output*: file 6) with the association rules (see Table 2)

Below, we provide the evaluation details and findings with validity and limitations.

5.5 Evaluation and Discussion

When performing the evaluation, we set out to test how far the association rules can predict a potentially relevant task set for an incoming ticket. The prediction considers the following: contextual information describing the ticket content, i.e., ticket type, DML/process cognition, and BS/customer anxiety. While assessing the prediction quality, we rely on two indicators: (1) support and confidence values and (2) a proprietary rule-based algorithm based on [45] involving the following steps:

- Training and test dataset generation (70%:30%) to assess the prediction quality.
- Deriving prediction rules by transforming the association rules obtained at the mining stage from the training dataset into *condition – consequent* pairs. *Condition* is the head of the association rule in one of the three formats: (1) ticket type, (2) ticket type and DML/process cognition, (3) ticket type, DML/process cognition, BS/customer anxiety. *Consequent* is the rule body, i.e., a set of tasks associated with a given *condition.*
- Forming a rule-based algorithm for the assessment of prediction quality consisting of the following. (a) We look for the *condition* matching the *condition* in the test dataset (in one of the three formats) to perform prediction and use the top-3 *consequents* as predictions of a task set ordered by the support score. (b) We evaluate the quality of our prediction by the number of attempts to find an exact or partial match between the top-3 *consequents* and the corresponding task sets from the test dataset.
- Finally, we evaluate the *consequents* and compare the results for three different types of *conditions*. Hereby, we determine how often the next task set can be correctly predicted in the first three prediction attempts (top-3). We suppose that the three suggestions are a reasonable number to display to a process worker as a recommendation.

Thus, we correctly predicted the task set, i.e., rule body or *consequent*, based on the ticket type *condition* in 43% of cases using one attempt. With increasing attempts, this number has grown to 50% (three attempts). Ticket type and DML/process cognition in the rule head (*condition*) demonstrated an evident increase in prediction quality, i.e., 59% of cases (one attempt) and 65% (three attempts). Finally, adding the BS/customer anxiety contextual information in the rule head showed no substantial influence: 57% (one attempt) and 63% (three attempts).

The described two-fold evaluation of the prediction quality allows the process worker to choose the best option from the top-3 recommended task sets. The enrichment of the rule head (*condition*) with the DML/process cognition contextual information positively influences the support and confidence values, in contrast to BS/customer anxiety. Similarly, DML/process cognition apparently impacts the task sets from the content viewpoint, i.e., task types and subtypes. The higher the DML/process cognition is, the higher the amount of QA and PSO tasks are. On the contrary, BS/customer anxiety has a minimal impact on the task sets. However, this information can be used by the process workers to make a correct prioritization in the process execution.

Due to the nature of context information [46], our study reveals several threats to validity and limitations, which we have either partially addressed or plan as a part of future work:

- *Context information has a range of temporal characteristics.* Currently, a process worker gets a recommendation in the form of a task set and needs to adjust the correct execution order based on the individual experience. Using the task execution time information, we plan to perform task mining at the immediate context layer to provide recommendations on the task execution sequence.
- *Context information is imperfect.* It might contain issues in reflecting reality. In our study, the decision support has a recommendation character and consists of a set of possible options. It is transparent regarding the deployed text analytics and recommended task sets.
- *Context information is highly interrelated.* We have already managed to identify several relationships between the DML/process cognition context information in the rule head and certain task types in the rule body (or their absence in the case of BS/customer anxiety). While including the recommendation on the task execution order, we plan to examine possible relationships in this regard.

A further limitation is related to the textual process data itself. As one can conclude from the framework conceptualization section, two important requirements on the data are the following: (1) these contain a request from a customer, and (2) textual data records on the tasks or activities necessary to process the requests are available. The ITSM case study has been selected as it fulfills both requirements. Although ITSM has become rather popular, especially in large organizations, we suggest other cases where our framework can be applied. In particular, many cases exist where textual data used as input to a process impacts its execution. In healthcare, for example, patients' complaints arriving via email or over a form on the website usually determine the required diagnostics and, hence, activities necessary to address the reported health problem. Similarly, in public administration, citizens' requests directly influence the required activities to handle the request. In other words, our framework can be applied in those settings where a process execution relies on textual data input. An additional limitation related to the data is the focus on BPM. To address it, we envision considering the challenges of context modeling and analysis for an improved system design.

There is another limitation related to obtaining linguistic features: same as context and textual data, linguistic features can also potentially bring cognitive overload and subconsciously misinform process workers. To avoid the potential cognitive overload, in future work, we aim to develop a dashboard with two focus points: (1) color-coded representation of DML/process cognition and BS/ customer anxiety and (2) recommendation of the tasks or activities and their order to solve the incoming request. For the second limitation on possible errors in our recommendations, we highlight that our approach and implemented text analytics techniques enable the transparency of those keywords and phrases, which might lead to misinformation of the process worker. Further, our approach has a recommendation character, i.e., process workers can decide to rely on their own experience or other sources of information.

6 Conclusion and Future Work

Under conditions of increased remote communication, the amount and importance of textual data in organizations cannot be underestimated. As humans are limited in processing large amounts of data, decision-making support solutions extracting important context-related aspects are in demand. In this study, we aimed to develop a framework for context awareness based on textual process data. Specifically, we extended a prominent BPM context model where we bring forward the human aspect in BPM through the textual data perspective. Hereby, the framework, our main theoretical and methodological contribution, emphasizes the value of non-technical artifacts in BPM.

We developed the framework in the IT ticket processing case study that allowed us to demonstrate its applicability and practical value. Specifically, we highlight the context layers, linguistics features, and text analytics to extract the latter. As a result, process worker context awareness is significantly improved, enabling specific decision-making support on predicting process cognition and customer anxiety level, i.e., complexity and urgency, implied in the text. Further, the process worker is supported with a list of recommended task sets for handling IT tickets. For the recommendation, the framework obtains additional contextual information from the textual data contained in IT tickets. Our findings showed that considering the contextual information enabled by our framework increases the quality of decision-making support. As a part of future work, from the performance viewpoint, we aim to enhance our framework by employing predictive process cognition and customer anxiety as well as develop a dashboard providing recommendations. From the context viewpoint, we will consider cases to enrich the framework with the environmental layer. Further, from the rigor viewpoint, we will test our framework in other settings like healthcare or public administration.

References

1. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-33143-5
2. Rosemann, M., Recker, J., Flender, C.: Contextualization of business processes. Int. J. Bus. Process. Integr. Manag. **3**, 47–60 (2008)
3. vom Brocke, J., Schmiedel, T., Recker, J., Trkman, P., Mertens, W., Viaene, S.: Ten principles of good business process management. Bus. Process. Manag. J. **20**, 530–548 (2014)
4. Avgerou, C.: Contextual explanation: alternative approaches and persistent challenges. MIS Q. **43**, 977–1006 (2019)
5. Sundermann, C.V., de Pádua, R., Tonon, V.R., Domingues, M.A., Rezende, S.O.: A context-aware recommender method based on text mining. In: Moura Oliveira, P., Novais, P., Reis, L.P. (eds.) EPIA 2019. LNCS (LNAI), vol. 11805, pp. 385–396. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30244-3_32
6. Song, R., Vanthienen, J., Cui, W., Wang, Y., Huang, L.: Context-aware BPM using IoT-integrated context ontologies and IoT-enhanced decision models. In: IEEE Conference on Business Informatics. pp. 541–550. IEEE (2019)
7. Whetten, D.A.: An examination of the interface between context and theory applied to the study of chinese organizations. Manag. Organ. Rev. **5**, 29–55 (2009)
8. Johns, G.: The essential impact of context on organizational behavior. Acad. Manag. Rev. **31**, 386–408 (2006)
9. Pentland, B.T., Recker, J., Wolf, J.R., Wyner, G.: Bringing context inside process research with digital trace data. J. Assoc. Inform. Syst. **21**(5), 1214–1236 (2020)
10. Müller, O., Junglas, I., Debortoli, S., vom Brocke, J.: Using text analytics to derive customer service management benefits from unstructured data. MIS Q. Exec. **15**, 243–258 (2016)
11. Hoang, H.H., Jung, J.J.: An ontological framework for context-aware collaborative business process formulation. Comput. Informatics. **33**, 553–569 (2014)
12. Hompes, B.F.A., Buijs, J.C.A.M., van der Aalst, W.M.P.: A generic framework for context-aware process performance analysis. In: Debruyne, C., et al. (eds.) OTM 2016. LNCS, vol. 10033, pp. 300–317. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48472-3_17
13. Rizun, N., Revina, A., Meister, V.G.: Assessing business process complexity based on textual data: Evidence from ITIL IT ticket processing. Bus. Process Manag. J. **27**(7), 1966–1998 (2021). https://doi.org/10.1108/BPMJ-04-2021-0217
14. Paas, F., Sweller, J., Paas, F., Sweller, J.: An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. Educ. Psychol. Rev. **24**, 27–45 (2011)
15. vom Brocke, J., Zelt, S., Schmiedel, T.: On the role of context in business process management. Int. J. Inf. Manage. **36**, 486–495 (2016)
16. Zelt, S., Recker, J., Schmiedel, T., vom Brocke, J.: A theory of contingent business process management. Bus. Process. Manag. J. **25**, 1291–1316 (2019)
17. Weber, M., Grisold, T., vom Brocke, J., Kamm, M.: Context-aware business process modeling: empirical insights from a project with a globally operating company. In: European Conference on Information Systems. AIS, Marrakesh, Morocco (2021)
18. Rosemann, M., Recker, J.: Context-aware process design: exploring the extrinsic drivers for process flexibility. In: Workshop on Business Process Modeling, Development, and Support at CAiSE, pp. 149–158. CEUR, Luxembourg (2006)
19. Zelt, S., Recker, J., Schmiedel, T., vom Brocke, J.: Development and validation of an instrument to measure and manage organizational process variety. PLoS ONE **13**, e0206198 (2018)

20. vom Brocke, J., Baier, M.-S., Schmiedel, T., Stelzl, K., Röglinger, M., Wehking, C.: Context-aware business process management. Bus. Inf. Syst. Eng. **63**(5), 533–550 (2021). https://doi.org/10.1007/s12599-021-00685-0
21. Boukadi, K., Chaabane, A., Vincent, L.: Context-aware business processes modelling: concepts, issues and framework. IFAC Proc. Volumes **42**(4), 1376–1381 (2009). https://doi.org/10.3182/20090603-3-RU-2001.0291
22. Enrique, H.V., De Maio, C., Fenza, G., Loia, V., Orciuoli, F.: A context-aware fuzzy linguistic consensus model supporting innovation processes. In: IEEE International Conference on Fuzzy Systems. pp. 1685–1692. IEEE (2016)
23. Mounira, Z., Mahmoud, B.: Context-aware process mining framework for business process flexibility. In: International Conference on Information Integration and Web-Based Applications and Services, pp. 421–426 (2010)
24. Hidri, W., M'tir, R.H., Bellamine, N., Saoud, B., Ghedira-Guegan, C.: A meta-model for context-aware adaptive business process as a service in collaborative cloud environment. Procedia Comput. Sci. **164**, 177–186 (2019). https://doi.org/10.1016/j.procs.2019.12.170
25. Graesser, A.C., McNamara, D.S., Kulikowich, J.M.: Coh-metrix: providing multilevel analyses of text characteristics. Educ. Res. **40**, 223–234 (2011)
26. Dinh, L.T.N., Karmakar, G., Kamruzzaman, J.: A survey on context awareness in big data analytics for business applications. Knowl. Inf. Syst. **62**(9), 3387–3415 (2020). https://doi.org/10.1007/s10115-020-01462-3
27. Purnomo, F., Heryadi, Y., Gaol, F.L., Ricky, M.Y.: Smart city's context awareness using social media. In: International Conference on ICT for Smart Society, pp. 119–123. IEEE (2016)
28. Cartelli, V., Di Modica, G., Tomarchio, O.: A cost-centric model for context-aware simulations of business processes. In: International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, pp. 303–314. SciTePress (2015)
29. Daelemans, W.: Explanation in computational stylometry. In: Gelbukh, A. (ed.) CICLing 2013. LNCS, vol. 7817, pp. 451–462. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37256-8_37
30. Anastassiu, M., Santoro, F.M., Recker, J., Rosemann, M.: The quest for organizational flexibility: driving changes in business processes through the identification of relevant context. Bus. Process. Manag. J. **22**, 763–790 (2016)
31. Ploesser, K., Recker, J., Rosemann, M.: Building a methodology for context-aware business processes: insights from an exploratory case study. In: European Conference on Information Systems IT to Empower, pp. 1–12. University of Pretoria, South Africa (2010)
32. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**, 45–77 (2007)
33. Wand, Y., Weber, R.: Research commentary: information systems and conceptual modeling – a research agenda. Inf. Syst. Res. **13**, 363–376 (2002)
34. Axelos: ITIL® Service Transition. TSO, London (2011)
35. Rizun, N., Revina, A., Meister, V.: Method of decision-making logic discovery in the business process textual data. In: Abramowicz, W., Corchuelo, R. (eds.) BIS 2019. LNBIP, vol. 353, pp. 70–84. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20485-3_6
36. Rizun, N., Revina, A.: Business sentiment analysis. concept and method for perceived anticipated effort identification. In: Information Systems Development: Information Systems Beyond 2020, pp. 1–12. AIS eLibrary (2019)
37. Grossnickle, J., Raskin, O.: The Handbook of Online Marketing Research: Knowing Your Customer Using the Net. McGraw-Hill Education (2000)
38. Beigi, G., Hu, X., Maciejewski, R., Liu, H.: An overview of sentiment analysis in social media and its applications in disaster relief. In: Pedrycz, W., Chen, S.-M. (eds.) Sentiment Analysis and Ontology Engineering. SCI, vol. 639, pp. 313–340. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30319-2_13

39. Medelyan, O., Witten, I.H., Divoli, A., Broekstra, J.: Automatic construction of lexicons, taxonomies, ontologies, and other knowledge structures. Wiley Interdisc. Rev.: Data Min. Know. Discovery **3**, 257–279 (2013)
40. Hammer, H., Yazidi, A., Bai, A., Engelstad, P.: Building domain specific sentiment lexicons combining information from many sentiment lexicons and a domain specific corpus. In: Amine, A., Bellatreche, L., Elberrichi, Z., Neuhold, E.J., Wrembel, R. (eds.) CIIA 2015. IAICT, vol. 456, pp. 205–216. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19578-0_17
41. Hutto, C., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc. Int. AAAI Conf. Web Soc. Media **8**(1), 216–225 (2014). https://doi.org/10.1609/icwsm.v8i1.14550
42. Blei, D.: Probabilistic topic models. Commun. ACM **55**, 77–84 (2012)
43. Eppler, M.J., Seifried, P., Röpnack, A.: Improving knowledge intensive processes through an enterprise knowledge medium. In: Conference on Managing Organizational Knowledge for Strategic Advantage: The Key Role of Information Technology and Personnel. pp. 222–230. Gabler (1999)
44. Rizun, N., Revina, A., Meister, V.G.: Analyzing content of tasks in Business Process Management. Blending task execution and organization perspectives. Comput. Ind. **130**, 103463 (2021)
45. Wright, A.P., Wright, A.T., McCoy, A.B., Sittig, D.F.: The use of sequential pattern mining to predict next prescribed medications. J. Biomed. Inform. **53**, 73–80 (2015)
46. Henricksen, K., Indulska, J., Rakotonirainy, A.: Modeling context information in pervasive computing systems. In: Mattern, F., Naghshineh, M. (eds.) pervasive computing, pp. 167–180. Springer Berlin Heidelberg, Berlin, Heidelberg (2002). https://doi.org/10.1007/3-540-45866-2_14

Check for updates

Digital Technologies Supporting Digitalization: A Maturity Model to Manage Their Usage Risks

Lamiae Benhayoun[1](✉) and Imed Boughzala[2]

[1] UIR Rabat Business School, Rabat, Morocco
lamiae.benhayoun@uir.ac.ma
[2] Institut Mines Telecom Business School, Evry, France

Abstract. Digital technologies in support of digitalization allows organizations to improve their strategic and operational performance, but also harbors risks of security, oversizing or loss of control, among others. While the management of these risks is essential to promote the success of digital transformation, no research offers an integrative framework to help monitor them. This study adopts a design science approach to conceive a maturity model evaluating the risks of using digital technologies. This framework is the initial step towards the supervision of these digital risks to succeed in digital transformation. We relied on an in-depth literature review and an empirical study using a Delphi approach and a focus group with 19 practitioners. Accordingly, we identify three dimensions of risks related to data, stakeholders and technology governance and distinguish them according to each digital technology in the spectrum of SMAC and DARQ technologies. We additionally define a maturity scale to assess these risks and a protocol to implement the maturity model. The paper concludes with its theoretical and practical implications as well as a research agenda.

Keywords: Digital technologies · Digital transformation · Risk management · Maturity · SMAC · DARQ

1 Introduction

To reap the benefits of the recent digital technologies, companies in all sectors are increasingly embracing deep digital transformation projects [1], for which expenditure will reach two trillion dollars by the end of 2022 [2]. Digital transformation or digitalization refers to changes in working methods, roles and behaviors of individuals, and commercial offers, that are induced by the intensive use of digital technologies in the organization and in its operational environment [3]. To accomplish a digital transformation process, companies rely on digitization by dematerializing information [1] and transforming their products, existing services and processes in digital variants [4]. This wave of transformation was driven by the advent of SMAC (Social, Mobile, Analytics, Cloud) technologies that enabled companies to improve their operational performance through the reduction of costs and execution times [3]. These technologies have also revolutionized the business models through which the company delivers value to customers,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 42–62, 2023.
https://doi.org/10.1007/978-3-031-26886-1_3

while providing them with an innovative user experience [5]. Today, being digital-first for a company is no longer an innovation or a competitive advantage, but a minimum condition for surviving in a constantly changing market. Successful companies are those that genuinely combine SMAC technologies with the new generation of DARQ technologies (Distributed ledger, Artificial intelligence, extended Reality, Quantum calculation) marking the post-digital era [6].

Over the past five years, digital transformation has aroused growing interest among practitioners and researchers in the Information Systems stream. Some scholars focused on identifying opportunities to improve operational and strategic performance offered by the use of digital technologies [7]. Others explored the changes in organizational practices that accompany the adoption and acceptance of these technologies [8], as they require different postures and capacities from previous technological waves [9]. Finally, few researchers took an interest in identifying the risks induced by these technologies, in particular security risks [10], compliance with standards and regulations [11], relationships with third parties [7] and employee governance [12]. The exploration of such risks concerned specific sectoral contexts, often the military [13] and health [14], or specific technologies namely social networks [15] and artificial intelligence [16]. However, no study offers a model to monitor the risks related to the use of digital technologies as a whole, regardless of the organizational context. Designing such an integrative framework to manage these risks simultaneously rather than in isolation, would allow researchers to better understand the interconnected nature of digital technologies and the complexity that their risks add to existing operational problems. Also, this model would help firms implement preventive practices to better benefit from investments and efforts in deploying digital technologies.

In this respect, maturity models are often used to evaluate the organization's abilities, identify the most critical issues and initiate improvement activities [17]. They suggest that an enhanced action will lead to a better outcome [18]. Accordingly, we raise the following research question: *How to design a maturity model assessing the risks of using digital technologies in support to digital transformation*? To answer this question, we follow a Design Science approach [19, 20] for maturity model development [21]. This approach relies on a thorough analysis of the literature, a Delphi method, and a focus group with 19 practitioners. The resulting maturity model constitutes the initial step in the successful monitoring of risks associated with the use of digital technologies. The paper is structured as follows. Section 2 presents our conceptual foundations, namely the technologies at the heart of digital transformation and the risks they entail. Section 3 explains the methodology adopted for this study. The results, consisting of the proposals made by the professionals with regards to the maturity model, are presented and discussed in Sect. 4. Finally, the conclusion highlights the study limitations and discusses the main directions of future research.

2 Literature Review

2.1 Digital Transformation: A Change Supported by Digital Technologies

To respond to a digitally disrupted environment, many companies joined the wave of digital transformation or digitalization over the past decade [22]. This phenomenon reflects a profound and intentional restructuring of their capacities, resources, and value creation pathways to benefit from the advantages offered by digital technologies [7, 23, 24]. This alteration aims to seize revolutionary opportunities in three major areas [9]. First, it seeks to improve the user experience by creating customized products and services and establishing a transparent and personalized digital relationship with clients [5]. Second, digital transformation offers opportunities for streamlining business processes that improve agility and responsiveness [3]. Finally, it allows the creation of new business models to increase the strategic benefits of using digital technologies [25].

Digital transformation is supported by digital technologies [26] encompassing all systems, tools, devices, and electronic resources used to generate, store and process data [27]. It was induced by the advent of SMAC technologies referring to Social, Mobile, Analytics and Cloud [1], and continues to intensify with the emergence of DARQ technologies (Distributed ledger, Artificial intelligence, extended Reality, Quantum computing) which are moving companies towards a post-digital era [6]. These technologies can quickly and severely alter the competitive dynamics of industries, that is why digitalization is now a priority for many organizations [7]. If digitalization represents undeniable advantages for companies, it nonetheless conceals major challenges. Scholars pointed out that digital transformation carries risks related to IS adoption [28], data governance [29], well-being at work [30], skills development [31], strategic alignment with IT [32], etc.

The term "transformation" expresses the entirety of actions to be taken when organizations are faced with a disruption. It goes beyond functional thinking and addresses the opportunities, but also the risks associated with change [33]. Hence, identifying and managing the risks inherent in a transformation process is a prerequisite for its success [34]. However, unlike the grey literature such as reports from consulting firms and white papers from companies [e.g., 6, 35, 36], very few IS researchers characterized the risks associated with digital transformation [e.g., 7, 37]. Even fewer proposed approaches to monitor these risks [38]. This creates a real discrepancy: companies are increasingly implementing initiatives to manage the risks of digital transformation, while the academic literature does not scientifically explore these practices. Such research would promote the cross-fertilization of academic and practitioner knowledge.

2.2 Risks Related to the Use of Digital Technologies

Digital transformation involves a multitude of risks related, among others, to employee well-being [30], strategic alignment [32] and skills development [31]. The characterization and assessment of these risks constitutes a research gap that needs to be addressed in order to provide a better understanding and support of digital learning pathways and digital governance at individual, organizational and governmental levels. In this paper, we chose to focus on the risks that accompany a key dimension of digital transformation: the use of digital technologies. Their implementation forms the foundation of the digitalization process [39] and poses particular risks, as digital technologies require postures and capabilities different from the previous technological waves [9]. Indeed, due to the alterations in communication, work and more generally intellectual style induced by these digital technologies, it is crucial to understand them, know how to interact with them [40], and to manipulate the data they contain [41].

We therefore conducted a literature review to assess the state of scientific knowledge about the risks of using digital technologies. We relied on the reference platform "Web of Science" to explore articles in peer-reviewed journals and renowned conferences in the Information Systems' field. We limited our search to articles written in English and published since 2005, as this year marks the first scientific work on the digital transformation of companies [42]. To define our keywords, we were guided by the definition of digital technologies proposed by [27] and presented in the previous section, and that of the term "risk". According to [43], risk corresponds to something that can be lost and the probability of actually losing it. This concept reflects a hazard or a potential malfunction, more or less foreseeable, and which can cause damage [44]. Thus, we combined the terms "Digital transformation" or "Digitalization", with the keywords "Risk", "Threat" or "Danger", and with "Technology", "Digital technology" or one of the eight digital technologies (SMAC, DARQ) involved in digital transformation. In total, we selected 61 articles, the vast majority of which were published since 2016. Most of this research is based on case studies or exploratory interviews and none offers a classification of the risks associated with the use of digital technologies. To relate the risks identified in this state of the art, we relied on the 5W approach (Who, What, When, Where, Why) recommended by [45]. It delivers an exhaustive characterization of a phenomenon by classifying it according to its different dimensions. Table 1 explains these five classification criteria in the context of our study and provides examples of risks for each criterion resulting from our state of the art. These risks are detailed in the following paragraphs.

Who: The risks of using digital technologies have been analyzed at the levels of employees, companies and the government. Employee-specific risks mainly involve the lack of skills, leadership, creativity and of entrepreneurial spirit related to digital technologies [46]. They also include resistance to the change induced by these technologies [47] and the hampering of learning efforts following the misuse of artificial intelligence [48]. At the organizational level, the most critical risks refer to the non-alignment of IT and business strategies [32], the absence of a digital culture [49], and a low level of digital maturity in the case of SMEs [50]. Finally, the risks identified for government authorities concern the change in the structure of the labor market [51], the strengthening of social inequalities [52] and the difficulty of interacting with citizens reliably and efficiently across the multitude of social networks [15].

What: Although the existing theoretical corpus has addressed the risks associated with different aspects of digital transformation, most studies focused on the operational process aspect. Some authors pointed out that digitalization generates risks of substantial production and infrastructure development costs [53]. Others specified processual issues of data confidentiality [14], technology sustainability [13] and vulnerability [16], and lack of competent digital actors [54]. The literature also addressed the risks associated with the digital transformation of the user experience, namely miscontrol over their communicated data [55], citizen security in smart cities [56] and the difficulty of establishing an effective relationship with consumers through mobile marketing [57]. Finally, few studies underlined the risks of business model digitalization, particularly the difficulty to renew the strategies in light of new advanced technologies [58].

Table 1. Analysis of the state of the art

Classification criterion	Explanation of criterion	Examples of results
Who	Categorizes the risks according to the level of analysis considered	• Individual: Degradation of skills, reluctance to change • Organization: Lack of alignment between IT and Business strategies, lack of digital culture • Government: Alteration of the job market, reinforcement of social inequalities
What	Classifies the risks according to the impacted transformation component	• Operational processes: Development costs, infrastructure instability, lack of competent actors • User experience: inappropriate use of technologies, confidentiality of data communicated • Business model: disruption of business activity, the institution's role in the socio-economic landscape
Where	Distinguishes the risks according to the sector of activity	• Bank-insurance: Fraud, huge flow of data that is hard to process • Military: Durability of technologies • Health: Data Privacy
Why	Qualifies the risks according to the technologies that cause them	• Artificial Intelligence: Ethical issues, limits of algorithms • Data Analytics: Data Privacy, Cybersecurity • Mobile: Network issues, device obsolescence • Blockchain: Lack of standards, lack of competent actors
When	Provides a reading of the risks according to the moment of their occurrence in the digitalization process	• Development of technology: Integration of Privacy by design, Inadequacy of certain digital project methods • Day-to-day use: Cybersecurity, data leakage

Where: Several authors emphasized process transformation risks in specific sectors. The health sector highlights the problem of data confidentiality [14], which results in the difficulty of establishing effective collaboration between healthcare providers and developers of specialized technologies [59]. The military sector is characterized by the concern of technology sustainability [13], while the agricultural sector presents risks of vulnerability and security of the connected objects used [16]. The digitalization of processes in the field of logistics is accompanied by significant development costs and a lack of experts with digital skills [54]. The financial and telecommunication sectors are mostly concerned by business model risks, namely the need to review the role of certain financial institutions [58] and telecommunications operators in the digital era [60].

Why: The 'Why' in our study qualifies the risks according to the originating technology among the SMAC/DARQ spectrum. Most studies focus on the risks generated by artificial intelligence, such as the ethical issues of unemployment [61], and the limits of algorithms and their vulnerability [16]. A multitude of researchers also investigated the issues of mobile technologies in terms of data confidentiality [62], network vulnerability [63], and rapid obsolescence of devices [64]. The risks associated with Data analytics technologies are also very present, in particular fraud and cybersecurity [65], the difficulty of processing the volume of data which continues to increase [66], and data privacy [67]. This privacy issue also represents a significant risk when implementing Cloud technologies [14, 67] and social media [65]. These media additionally involve risks of inappropriate use to design an effective user experience [15] aligned with his expectations [57]. Numerous authors highlighted the risks of adopting Distributed ledger technology (e.g., Blockchain), in particular the immaturity of the technology, the absence of common standards [68], the scarcity of competent actors in this field, and the inefficiency of the institutional environment [69]. Finally, to our knowledge, no research deals with the risks of using extended reality and quantum technologies. Besides, some studies underlined the risks associated with digital technologies as a whole [e.g., 51, 70]. We discuss these risks in depth in Table 3.

When: Regarding the moment of risk occurrence within the transformation process, our analysis of the literature showed that most of the risks are present during the development of technologies in support of digital transformation, for example the low quality of the Cloud developed in agile mode [70], the sustainability of digital engineering outputs in the military [13], the cost of developing digital technologies for warehousing [54], and the consideration of privacy issues in technology design [67]. A large part of the risks also takes place during the day-to-day use of digital technologies, for example resistance to change [47], skill degradation [48], alteration of social structures [52], leaks of sensitive data [62], and cybersecurity [55].

In short, no study offers an integrating framework or an overall classification of the risks associated with the use of digital technologies. To cover this gap and enrich the risks identified in our state of the art, we followed a design science approach to conceive a maturity model assessing these risks' management.

3 Research Methodology

3.1 Data Collection

Maturity models have been extensively investigated in several domains as an instrument for continuous improvement [17, 72]. They suppose that when activities are defined, managed, and executed effectively, they lead to better performance [18]. In this research, we follow the seminal design science methodology proposed by [21] to conceive a maturity model. Additional to the design stage, the authors suggest that the models should also undergo field applications and frequent updates for maintenance to comfort their continuous validity.

The construction of a maturity model requires determining the key process areas (KPAs), i.e., themes that are mutually exclusive and collectively exhaustive to describe the evaluated object [21]. Each KPA is defined through associated practices, implemented collectively to satisfy the goals of the area. These KPAs are described at a number of levels of performance [73, 74]. The highest maturity level is where the KPA's practices are efficiently applied and culturally rooted [17, 75].

To determinate the KPAs and the maturity scale, we relied on an in-depth literature review combined with a Delphi type approach. The Delphi method is generally used in Information Systems' research when the generation of ideas is necessary to convey a consensual opinion on a well-determined subject [75]. This approach is based on the remote surveying technique [76]. During our study, to collect a field reflection (bottom up) on the risks associated with the use of digital technologies, we set up an electronic survey system with 19 practitioners from 9 organizations. Table 2 summarizes the key characteristics of the participants and their organizations. The latter were chosen because of their in-depth knowledge of the subject and the richness of points of view that they can provide based on their complementary profiles.

The survey was conducted in one-to-one sessions with participants and the entire process was recorded to circumvect any missing point. Each session was initiated by the introduction of the study objectives, namely to propose a structure of the maturity model and an assessment protocol. We also provided examples of risks associated with the use of digital technologies resulting from our 5W literature analysis. Subsequently, the participants reacted to questions structured according to three interaction times:

First, each participant had to refer to his digital transformation experience in order to propose risk management capabilities that are related to the use of digital technologies as a whole, and explain their components or items. These capabilities would ultimately represent the KPAs of our model.

Secondly, we asked them to provide us with a more detailed vision of these risks according to each of the SMAC/DARQ technologies in support of digital transformation.

Finally, we invited each participant to share with us the maturity assessment approaches he was knowledgeable of, and which could be suitable to the risks of using digital technologies. We mainly requested them to suggest a definition and criteria of maturity to be used in an evaluation scale.

Table 2. Characteristics of the participants and their organizations

Characteristics of the participants			Characteristics of the organizations				
Number of participants	19		Org	Sector	Type	Size	Disciplines
Youngest	31		Org 1	IT consulting	Private	80	Analytics and AI expert
Oldest	58		Org 2	Bank	Private	>50000	Analytics and AI expert, financial analyst
Average age	46		Org 3	Insurance	Private	≈10000	Analytics and AI expert, financial consultant
Male	14		Org 4	University	Public	1000	Cybersecurity expert
Female	5		Org 5	University	Public	1000	R&D managers
Education level	7 PhD, 10 Master, 2 Bachelor		Org 6	Electricity	Private	300	CEO, R&D manager, Logistics manager
Longest experience	25 years		Org 7	Pharmaceutical	Private	500	CEO, Biotech researcher, R&D manager, Purchasing manager
Shortest experience	9 years		Org 8	Automotive	Private	> 50000	Purchasing manager, logistics manager, assembly manager
Average experience	15 years		Org 9	Automotive	Private	> 50000	R&D manager, logistics manager

3.2 Data Analysis

To analyze the data collected from this panel, we proceeded according to the three themes addressed during the survey, namely the capabilities related to the risks of using digital technologies, the refinement of these risks according to the SMAC/DARQ spectrum, and finally the proposal of a maturity evaluation scale. To analyze each theme, we carried out a double purification as recommended by [77] which we explain below.

The first purification took place throughout the investigation process by relying on the system in support of our Delphi approach. Indeed, the latter makes it possible to collect feedback from participants following a structured and proven process. In addition, it facilitates the development of a summary document synthesizing their responses. As we implemented our data collection approach, we proceeded to clarify, reduce, and organize participant feedback.

The second purification took place at the end of all the sessions. The participants were invited to a full day focus group structured in two sessions of three hours each, moderated by the researchers. A focus group is a discussion of a particular topic under the direction of a moderator who promotes group participation and communication and manages the discussion through a series of interactions [78]. During the first half-day, we presented the results of the Delphi survey and refined them. This refinement involved the reformulation, when necessary, of the answers, their consolidation, interpretation, and the verification of the proposals. In this step, we were guided by two criteria: the non-redundancy of ideas in the classification of participants' feedback, and the relatedness of each response to the topic of risks to use digital technologies.

During the second half-day, the developed maturity model was applied to a use case with the participation of the group members. An electricity firm taking part in the focus group was interested in implementing the model to assess the organization's maturity for managing the risks related to the use of digital technologies. This use case comforts the validity of the model as recommended by [21]. We present and discuss the field application in the findings' section.

4 Findings

4.1 The Maturity Model's KPAs

The maturity model developed in this research aims to assess the overall ability to manage the risks related to the use of digital technologies. It informs the firm on its strengths and weaknesses, that can serve on the definition of an action plan. To define the KPAs, we relied on an in-depth literature review combined with the insights of our Delphi approach. We asked the participants to refer to their experience of digital transformation to describe the main risks that characterize the use of digital technologies in general, and then each of the eight SMAC/DARQ technologies. This helped enrich our state of the art with new risks emerging from our study.

Following the analysis of the collected data and the validation during the focus group, we distinguished three classes of risks that are exhaustive and exclusive to represent the KPAs of our maturity model. These risks are listed in Table 3, which mentions whether each has already been highlighted in the literature or whether it emerged solely from our empirical study. In the next paragraphs, we provide verbatim on the risks that were heavily discussed during the focus group.

Data Sensitivity: Data sensitivity concerns information that should be protected from unauthorized access or disclosure due to its delicate nature. Participants underlined that "*digital technologies by nature generate data without the knowledge of the individual. The sharing of this data entails a risk in itself*". This is even more critical when it comes to "*personal data such as medical, financial or other sensitive data that could impact the reputation of the person*".

Relationships with Third Parties: According to the participants, "*digital technologies enable opening up to other stakeholders (customer, supplier, partner, competitor), and subsequently to more risks in terms of managing these relationships*". In particular, such technologies "*can cause a lot of formality in relationships, which harms the efficiency of partnerships*". Furthermore, some group members highlighted that "*the use of digital technologies makes the organization more exposed to the outside world, which can quickly put it at a disadvantage in the case of bad recommendations by customers*".

Governance of Digital Technologies: The group members insisted that "*the recent and constantly evolving nature of digital technologies poses challenges to the management of information systems within the company*". Indeed, "*these technologies require time to be mastered by the IT department and can quickly become obsolete*". The rapid mastery of these technologies is even more necessary to "*meet the instantaneous needs of the businesses in the organization. The digital age is precisely characterized by the culture of 'immediately'*". To meet their needs, businesses can "*resort to Shadow IT, which can pose a real governance problem for the IT Department. These technologies are increasingly accessible with the advent of digital. Apps and technologies developed yesterday are within everyone's reach!*". Furthermore, the IT Department often encounters the problem of "*data overflow in case of poor management of the data generation and distribution channels*".

Table 3. KPAs and their associated risks of using digital technologies

KPA	Nature of the digital technology	Risks
Data sensitivity	Digital technologies in general	Data generated without the knowledge of individuals
	Digital technologies in general	Access to unauthorized personal data (medical, financial, etc.)
	Digital technologies in general	Access to data that could damage the reputation of individuals
	Social	Exposure (fishing, harassment, grooming)
	Mobile	Access to private data
	Mobile	Cyberattacks
	Mobile	Identity theft
	Mobile	Data theft, Data alteration

(*continued*)

Table 3. (*continued*)

KPA	Nature of the digital technology	Risks
	Analytics	Prohibited manipulation of data (fingerprints, genomes, etc.)
	Cloud	Lack of protection of private data
	Cloud	Security/ data leakage
	Distributed ledger (e.g., Blockchain)	Fraud
Relationships with third parties	Digital technologies in general	*Difficulty to manage digital relationships with a multitude of stakeholders**
	Digital technologies in general	*Rigid professional relations**
	Digital technologies in general	Risky outdoor exposure
	Social	Bad reputation and negative publicity resulting from social amplification
	Social	Professionalization of the means to harm the person
	Cloud	Pressure from the suppliers to implement inadequate Cloud
	Artificial Intelligence	Financial opportunism at the expense of other social classes
Governance of digital technologies	Digital technologies in general	Difficulty to align with the evolving nature of digital
	Digital technologies in general	Investments to master technologies that can quickly become obsolete
	Digital technologies in general	*Difficulty to meet the immediate needs of businesses**
	Digital technologies in general	Excessive use of Shadow IT
	Digital technologies in general	Overflow by large data flows
	Mobile	Poor technology design
	Analytics	Wrong measurement

(*continued*)

Table 3. (*continued*)

KPA	Nature of the digital technology	Risks
	Analytics	Non-compliance with GDPR
	Cloud	Non-performance of technology
	Cloud	Data and country sovereignty
	Cloud	Loss of assets and their location
	Distributed ledger (e.g., Blockchain)	*Oversizing**
	Distributed ledger (e.g., Blockchain)	*Deployment by mimicry**
	Distributed ledger (e.g., Blockchain)	*Unnecessary investments**
	Artificial Intelligence	Algorithms not mastered
	Artificial Intelligence	Irresponsible innovation
	Artificial Intelligence	Inappropriate use
	Artificial Intelligence	*Poor choice of use cases**
	Artificial Intelligence	Misleading use
	Extended Reality	*Unnecessary investments to align with a fad**
	Quantum computing	*Overcapacity, oversizing to respond to simple problems**

* *Risk emerging from the empirical study.*

Regarding the risks associated to the particular use of a technology within the SMAC/DARQ spectrum, we note that a large part of the risks related to SMAC and Artificial Intelligence technologies have previously been identified in the literature, while the risks of using Distributed ledger, extended Reality and Quantum computing technologies have almost entirely emerged from our empirical approach.

4.2 The Maturity Scale Adopted in Our Model

Most studies that designed maturity models rely on the definition of maturity and the scale proposed in the CMMI (Capability Maturity Model Integration), a seminal model developed by [79]. CMMI defines maturity as the degree to which processes are formally organized and executed to produce the desired results [73]. Most of the participants were aware of the CMMI and, consequently, responded in the Delphi questionnaire by referring to its definition of maturity and its scale composed of four levels namely ad hoc, exploring, managing, and optimizing.

However, when discussing with the group participants the adequacy of this scale to assess the risk management of using digital technologies, we concluded that it does not meet the peculiarities of this evaluated topic. Indeed, digital risk management does not correspond to institutionalized processes as is the case of CMMI-based models and additionally integrates a behavioral dimension of awareness. Thus, based on the practitioners' feedback, we propose to evaluate the maturity in our model not only in terms of the organization's ability to operationally implement practices dealing with the risks in Table 3, i.e. '*Capable to do'*, but also its propensity toward these practices, i.e. '*Willing to do'*.

To evaluate the organization's maturity according to these criteria, we adopted a hybrid descriptive approach as recommended by [21] for topics that were never operationalized from a maturity perspective. It consists of asking a question conveying the highest level of maturity for each evaluated item. In our case, for each risk, we ask two questions associated with the two maturity criteria. For the capability criterion, we evaluate if the organization is able to implement the methods and tools to prevent this risk and control it when it occurs. For the willingness criterion, we assess whether the organization perceives the relevance of managing the risk and is willing to implement the necessary practices in order to monitor it. After discussing with the focus group participants, we defined 4 levels of maturity associated with these two maturity criteria. A level of maturity is granted to the organization depending on its answers to the two questions on a scale of 1 to 4.

For capability, if the organization states that it perfectly masters the methods and tools to monitor the risk, it is considered (4)*expert*. If it feels capable but needs more formalization, the maturity level is (3)*capable*. If the organization has some ideas but does not know how to proceed to mitigate the risk, it is (2)*novice*. If it does not have any idea, method or tool to proceed with the risk, it is (1*)not capable.*

As for willingness, the organization is a (4)*firm believer* if it is fully convinced with the benefits of managing the risk and is inclined to implement all the necessary practices. It is (3)*culturally rooted* if it perceives the interest of managing the risk and would eventually agree to implement some practices in this respect. The organization is (2)*potentially receptive* if it is not against managing the risk but is not convinced with the interest of this action. Finally, it is (1)*culturally resistant* if it does not find any relevance in monitoring the risk.

4.3 Feedback from the Field Application

4.3.1 Presentation of the Case Application

We performed a field application during the second half-day of the focus group to improve the designed maturity model and evaluate its usability, usefulness, and completeness [73]. The CEO of the company "Org 6", expressed his interest to assess the maturity of his organization with respect to the management of risks related to the use of digital technologies. This SME operates in the electricity sector. It proposed solutions for electrical mobility and energy storage that respect the principle of planned sustainability, in contrast to planned obsolescence.

This principle aims to face the uncertainties related to health, financial and ecological crises, by adopting a design and manufacturing mode that fosters sharing, reuse and cooperation. Sustainability is at the heart of this organization's activities: from the design, production, maintenance of its products and the logistical choices, to the management of its employees. Org 6 relies on digital technologies to implement this planned sustainability principle. Therefore, it found extreme relevance in evaluating its maturity in terms of digital use, considering the importance of this dimension for its activities and performance.

4.3.2 The Maturity Evaluation Protocol

During the second half-day of the focus group, the 3 participating members of Org 6 (CEO, R&D manager, Logistics manager) applied our maturity model. The model was implemented in a VBA tool prior to the focus group and was adjusted as we discussed with the practitioners during the first half-day. The evaluation was split into four phases. After each one, a debriefing was performed with these 3 actors as well as with the rest of the focus group members.

The first phase consisted in an introduction to the maturity tool's structure and functionalities. The participants were invited to browse a one-page documentation which recalls the objectives of the model and the instructions to guide its implementation. Following the debriefing, the content of this introductory page was improved according to the participants' comments to make it a frame of reference enabling future firms to use the model without the presence of the researcher as an external moderator.

During the second phase, the 3 firm actors proceeded to self-assessment by completing a questionnaire adapted from the Delphi survey to assess the organization's maturity according to the established maturity scale. The discussion with the participants following this stage enabled to clarify some maturity items by reformulating them or illustrating them with examples.

In the third phase, participants scrolled, together with the moderator, the assessment results that are automatically generated in a summary report. It depicts an aggregated score for each KPA accompanied with detailed results for each risk according to the two maturity criteria. Participants unwrapped their results from the least to the most performant KPA and discussed the reasons for the maturity gap of each KPA's items. Accordingly, recommendations for improvement were formulated with the help of the focus group members to enable cross fertilization of best practices.

The fourth phase aimed at verifying the success related to the usage of the model by asking all the focus group members to answer a set of questions assessing its usability, completeness, and utility [73]. Also, the participants could fill blank spaces to provide any further feedback that was not delineated through the proposed questions. Besides collecting their comments and propositions in writing, the researchers made sure to record the oral, gestural, and visual reactions of the participants throughout the evaluation session which had served to improve the content of the model.

4.3.3 Results of the Maturity Assessment

The assessment showed that Org 6 was very mature regarding the relationships with third parties. It was using digital technologies efficiently with its suppliers, particularly the Cloud and AI, while controlling for their risks. Org 6 applies these technologies to monitor the supply and demand and enable more flexibility on its partnerships. It also harnesses the power of social media to observe the trends and collect the users' feedback on its products. Regarding this topic, we noted a difference of understanding among the three evaluating members. While the CEO was aware of the digitalization projects within the company aiming at fully benefitting from the potential of social media, the R&D and logistics' manager thought that Org 6 is still immature regarding these aspects. After deep discussions, they agreed that the organization is mindful that social media could eventually harm its reputation and was implementing an AI based scanning to spot any negative tweets or comments and resolve the related issue.

This firm depicted several maturity problems regarding the KPA of Data sensitivity. In fact, it was in the process of implementing GDPR principles in its organization, such as the assignment of a chief digital officer and the collection of consent when personal data is eventually extracted. This firm was also lacking cybersecurity protocols and protective actions for its Cloud. Nevertheless, all these issues concerned the capability maturity criterion and not the willingness. Org 6 was convinced with the necessity to act on these risks and had indeed started an internal reflection on the actions to put in place.

4.3.4 Evaluation of the Model's Utility, Completeness, and Usability

In the fourth phase of the model's application underlined in the evaluation protocol (Sect. 4.3.2), we asked the focus group members to answer a set of questions regarding the model's completeness, usability, and utility. A debriefing was then performed to understand the pros and cons of our operationalized model and enhance it accordingly.

Regarding the model's **completeness**, we asked the participants to compare the accuracy of its content with classifications or frameworks of risks associated with the use of digital technologies that they were aware of. None of the participants was knowledgeable of such risks' classification. Nevertheless, they were informed of some tools to assess digital maturity, for instance the DQ framework by the DQ institute [80], the European Digital Competence Framework for Citizens by the European Commission [81], and the Digital Quotient by McKinsey & company [82]. A participant from the insurance sector stated the existence of a risk management process in a concurrent organization, entirely designed in-house, and built according to the life cycle of data in the insurance world. However, it only deals with the risks related to data. Compared to all these tools, the participants stressed that the maturity model developed in this study was "*more accurate in comparison to well-known generic tools. It follows a rigorous scientific approach far from any speculation and is co-constructed with practitioners to align with their frame of reference*". Also, the fact that its design involved organizations from different sectors made it an integrative framework that "*would promote mutual learning between organizations from a variety of sectors, especially as it is difficult for each company to have individual access to constructive feedback about these emerging technologies*".

For the model's **usability**, the participants appreciated the content of the introductory page which methodically explains how to use the model at each step. They then insisted that "*a collective evaluation involving key actors of the organization and representatives of its trades is crucial to capture the firm maturity profile and identify trouble spots*", and "*discern differences of understanding within the same firm that could hinder the implementation of any change management initiative*". Such interactivity would even foster "*actions to correct behavioral issues regarding digital technologies*". Indeed, "*a* *collective feedback based on the implicit recommendations of the model would guide the firm toward a better distribution of responsibilities with respect to the governance and use of digital technologies*".

As for the **utility** of the model, two key findings of the focus group discussion and the field application should be underlined. The first one was the participants' ability to clearly distinguish digital technologies from other types of Information Technologies. In fact, the group now understood that "*digital technologies are more 'data' oriented and involve a new cognitive framework*". These technologies "*alter our ways of communicating and working and require capacities and postures different from IT*". The respondents explained that "*classical IT is more closed, internal to the company*" while "*digital technologies involve the generation and exchange of data between several internal and external actors, whose volume and speed are important*". In addition, "*digital extends information and communication technologies and leads to new practices such as Do It Yourself and Bring Your Own Device*".

The second utility finding was particularly highlighted by Org 6. The CEO stressed that the model "*clearly ascertains the risks to be managed. It helped us take a step back on our abilities and pointed out our weaknesses and strengths of which we were not necessarily aware. This makes it easier to initiate targeted improvement actions to manage the alarming risks*". However, a participant underlined that "*this risk management should extend to all dimensions of digital transformation and not only to the use of technologies*". In this regard, another group member explained that "*digitalization involves risks of different kinds: legal risk, image risk, financial risk, ethical risk, time to market risk, usage risk, security risk, risk of non-availability, risk of monitoring, etc.*". Finally, the group participants argued that an adjustment phase of the model prior to its use "*can make the results more significant and concise by focusing the evaluation on the risks that are most relevant to the firm*". This scoping step could address "*all the key individuals within the organization and external experts to revise the model according to the firm's strategic and operational concerns and then apply it*".

5 Conclusion

The advent of digital technologies has induced a wave of transformation within organizations. The latter aims to redefine operational processes, business models, and user experiences to capitalize on the potential offered by digital technologies. However, this digitalization also harbors risks, among others, ethical, technical, and intellectual, of which companies must be aware in order to manage them throughout the transformation process. In this sense, this paper focused on the risks associated with a particular dimension of digital transformation, the use of digital technologies, as it constitutes the foundation of the digitalization process.

Our results combine both the insights of the literature and of an empirical study with practitioners following a Delphi approach and a focus group. From a theoretical standpoint, they contribute to defining the concept of digital risk capability by characterizing its dimensions and components. In this respect, they underline the existence of three classes of risks for using digital technologies, namely risks related to data sensitivity, relationships with third parties, and the governance of digital technologies. In addition, this research brings out new risks specific to the SMAC/DARQ spectrum and accordingly enhances our understanding of these technologies' common and different properties. Furthermore, our findings provide a definition of maturity specific to digital risks that is aligned with the conceptual nature of these items and the practitioners' cognition. From a managerial standpoint, the present study answers a contemporary practical need of firms who embrace digital transformation projects to prosper in this post-digital era. It raises managers' awareness of the risks they should prevent and monitor and provides them with a user-friendly tool to accompany the implementation of digital technologies within their organizations and initiate the necessary improvement actions to capture digital benefits in a secure manner.

This study, however, has some limitations that represent promising research perspectives. First, this paper focused only on the dimension of using digital technologies. It would be relevant to address in depth the risks related to the other aspects of the digitalization such as the alignment between the business strategy and the digital one or the development of employee skills. This would enable a better understanding of the interconnection between digital technologies, but also their relation to the different aspects of organizational performance. Second, the panel of respondents was limited to some specific sectors and, thus, was not statistically representative of the various industries that possess disparate levels of digital maturity. Further case studies would enrich our results and evaluate their applicability in other organizational contexts. Such studies can also inform the digital learning pathways and governance at the individual, organizational and governmental levels. Finally, this research was limited to the design of a maturity model for managing the risks associated to the use of digital technologies, without addressing the most appropriate way to manage them through concrete actions. Future field applications across different sectors could help build a framework of actions to be undertaken with respect to the maturity deficiency of each risk. Also, a quantitative study can help build a predictive model to prioritize the risks according to their probability of occurrence in an organization considering its contextual peculiarities.

References

1. Legner, C., et al.: Digitalization: opportunity and challenge for the business and information systems engineering community. Bus. Inf. Syst. Eng. **59**(4), 301–308 (2017)
2. IDC: Worldwide Digital Transformation Spending Guide. Retrieved from https://www.idc.com/getdoc.jsp?containerId=IDC_P32575 (2018)
3. Parviainen, P., Tihinen, M., Kääriäinen, J., Teppola, S.: Tackling the digitalization challenge: how to benefit from digitalization in practice. Int. J. Inf. Syst. Proj. Manag. **5**(1), 63–77 (2017)
4. Gassmann, O., Frankenberger, K., Csik, M.: The business model navigator: 55 models that will revolutionize your business. Pearson, UK (2014)
5. Reis, J., Amorim, M., Melão, N., Matos, P.: Digital transformation: a literature review and guidelines for future research. In: Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S. (eds.) WorldCIST'18 2018. AISC, vol. 745, pp. 411–421. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77703-0_41
6. Accenture: The Post-Digital Era is Upon Us: ARE YOU READY FOR WHAT'S NEXT?. Retrieved from https://www.accenture.com/_acnmedia/pdf-94/accenture-techvision-2019-tech-trends-report.pdf (2019)

7. Hess, T., Matt, C., Benlian, A., Wiesböck, F.: Options for formulating a digital transformation strategy. MIS Q. Executive **15**(2), 6 (2016)
8. Majchrzak, A., Markus, M.L., Wareham, J.: Designing for digital transformation: lessons for information systems research from the study of ICT and societal challenges. MIS Q. **40**(2), 267–277 (2016)
9. Singh, A., Hess, T.: How chief digital officers promote the digital transformation of their companies. MIS Q. Executive **16**(1), 101890 (2017)
10. Ardolino, M., Rapaccini, M., Saccani, N., Gaiardelli, P., Crespi, G., Ruggeri, C.: The role of digital technologies for the service transformation of industrial companies. Int. J. Prod. Res. **56**(6), 2116–2132 (2018)
11. Preece, R.: The GDPR accountability principle and the use of scenario workshops in the digital age. J. Data Prot. Priv. **2**(1), 34–40 (2018)
12. Garmann-Johnsen, N.F., Helmersen, M., Eikebrokk, T.R.: Digital transformation in healthcare: enabling employee co-creation through web 2.0. In: Americas Conference on Information Systems (2018)
13. Zimmerman, P., Gilbert, T., Salvatore, F.: Digital engineering transformation across the Department of Defense. The J. Defense Model. Simul. **16**, 325–328 (2019)
14. Yang, L., Zheng, Q., Fan, X.: RSPP: A reliable, searchable and privacy-preserving e-healthcare system for cloud-assisted body area networks. In: IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pp. 1–9. IEEE (2017)
15. Dneprovskaya, N., Bayaskalanova, T., Shevtsova, I., Urintsov, A.: Digital transformation of communication between government authorities and citizens. In: Proceedings of the International Conference on Research Paradigms Transformation in Social Sciences (2018)
16. Barreto, L., Amaral, A.: Smart farming: cyber security challenges. In: 2018 International Conference on Intelligent Systems (IS), pp. 870–876. IEEE (2018)
17. Van Looy, A., Poels, G., Snoeck, M.: Evaluating business process maturity models. J. Assoc. Inf. Syst. **18**(6), 1 (2017)
18. Dooley, K., Subra, A., Anderson, J.: Maturity and its impact on new product development project performance. Res. Eng. Design **13**(1), 23–29 (2001). https://doi.org/10.1007/s001630100003
19. Bichler, M.: Design science in information systems research. Wirtschaftsinformatik **48**(2), 133–135 (2006). https://doi.org/10.1007/s11576-006-0028-8
20. Hevner, A., Gregor, S.: Envisioning entrepreneurship and digital innovation through a design science research lens: a matrix approach. Inform. Manag. **59**, 103350 (2020)
21. Maier, A.M., Moultrie, J., Clarkson, P.J.: Assessing organizational capabilities: reviewing and guiding the development of maturity grids. IEEE Trans. Eng. Manage. **59**(1), 138–159 (2012)
22. Vial, G.: Understanding digital transformation: a review and a research agenda. The J. Strateg. Inform. Syst. **28**(2), 118–144 (2019). https://doi.org/10.1016/j.jsis.2019.01.003
23. Sebastian, I.M., Ross, J.W., Beath, C., Mocker, M., Moloney, K.G., Fonstad, N.O.: How big old companies navigate digital transformation. MIS Q, Executive **16**, 197–213 (2017)
24. Svahn, F., Mathiassen, L., Lindgren, R., Kane, G.C.: Mastering the digital innovation challenge. MIT Sloan Manag. Rev. **58**(3), 14 (2017)
25. Lepore, D., Nambisan, S., Tucci, C.L., Zahra, S.A.: Digital transformation & firms' innovative strategies: capabilities, ecosystems, and business models. In: Academy of Management Proceedings, vol. 2019, No. 1, p. 14623. Academy of Management, Briarcliff Manor, NY 10510 (2019)
26. Stolterman, E., Fors, A Croon: Information technology and the good life. In: Kaplan, B., Truex, D. P., Wastell, D., Wood-Harper, A.T., DeGross, J.I. (eds.) Information systems research. IIFIP, vol. 143, pp. 687–692. Springer, Boston, MA (2004). https://doi.org/10.1007/1-4020-8095-6_45

27. Strachan, R., Aljabali, S.: Investigation into Undergraduate International Students' Use of Digital Technology and Their Application in Formal and Informal Settings. International Association for Development of the Information Society (2015)
28. Gombault, A., Allal-Chérif, O., Décamps, A.: ICT adoption in heritage organizations: crossing the chasm. J. Bus. Res. **69**(11), 5135–5140 (2016)
29. Earley, S., Maislin, S.: Data governance and digital transformation: using the customer journey to define a framework. Appl. Mark. Analytics **2**(1), 25–40 (2016)
30. Koffer, S.: Designing the digital workplace of the future–what scholars recommend to practitioners. In: International Conference on Interaction Sciences (2015)
31. Muller, E., Hopf, H.: Competence center for the digital transformation in small and medium-sized enterprises. Procedia Manuf. **11**, 1495–1500 (2017)
32. Masuda, Y., Shirasaka, S., Yamamoto, S., Hardjono, T.: Architecture board practices in adaptive enterprise architecture with digital platform: a case of global healthcare enterprise. Int. J. Enterp. Inform. Syst. (IJEIS) **14**(1), 1–20 (2018)
33. Karimi, J., Walter, Z.: The role of dynamic capabilities in responding to digital disruption: a factor-based study of the newspaper industry. J. Manag. Inf. Syst. **32**(1), 39–81 (2015)
34. McDaniels, T., Small, M. (eds.): Risk Analysis and Society: An Interdisciplinary Characterization of the Field. Cambridge University Press (2004)
35. Capgemini: Digital Transformation Review: 12th edn. Retrieved from https://www.capgemini.com/fr-fr/etudes/digital-transformation-review-12/ (2019)
36. Deloitte: Managing Risk in Digital Transformation. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/in-ra-managing-risk-in-digital-transformation-1-noexp.pdf (2018)
37. Sambamurthy, V., Zmud, R.W.: Guiding the digital transformation of organizations. Lightness Digital Press (2012)
38. Carcary, M., Doherty, E.: 'The digital wild west': managing the risks of digital disruption. In: The European Conference on Information Systems Management, p. 29. Academic Conferences International Limited (2016)
39. Matt, C., Hess, T., Benlian, A.: Digital transformation strategies. Bus. Inf. Syst. Eng. **57**(5), 339–343 (2015)
40. Kineshanko, M.K., Jugdev, K.: Enhancing digital intelligence through communities of learning. In: Khare, A., Hurst, D. (eds.) On the Line, pp. 111–125. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-62776-2_10
41. Adams, N.B.: Digital intelligence fostered by technology. J. Technol. Stud. **30**(2), 93–97 (2004)
42. Zhu, K., Dong, S., Xu, S.X., Kraemer, K.L.: Innovation diffusion in global contexts: determinants of post-adoption digital transformation of European companies. Eur. J. Inf. Syst. **15**(6), 601–616 (2006)
43. Merriam Webster: Merriam-Webster's collegiate dictionary. Advocate [Internet][cited 31 Jan 2013]. Available from: www.merriam-webster.com/dictionary (2001)
44. Mili, A., Bassetto, S., Siadat, A., Tollenaere, M.: Dynamic risk management unveil productivity improvements. J. Loss Prev. Process Ind. **22**(1), 25–34 (2009)
45. Szostak, R.: Classifying science, pp. 1–22. Springer, Netherlands (2004)
46. Tekic, Z., Koroteev, D.: From disruptively digital to proudly analog: a holistic typology of digital transformation strategies. Bus. Horiz. **62**(6), 683–693 (2019)
47. Dasi, A., Elter, F., Gooderham, P.N., Pedersen, T.: New business models in-the-making in extant MNCs: digital transformation in a telco. In: Breaking up the Global Value Chain: Opportunities and Consequences, pp. 29–53. EmeraldPublishing Limited (2017)
48. Sułkowski, Ł, Kaczorowska-Spychalska, D.: Internet of things - new paradigm of learning. Challenges for business. In: Ayaz, H., Mazur, L. (eds.) AHFE 2018. AISC, vol. 775, pp. 307–318. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-94866-9_31

49. Romero, D., Flores, M., Herrera, M., Resendez, H.: Five management pillars for digital transformation integrating the lean thinking philosophy. In: 2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–8. IEEE (2019)
50. Riera, C., Iijima, J.: The role of IT and organizational capabilities on digital business value. Pacific Asia J. Assoc. Inform. Syst. **11**(2), 67–95 (2019)
51. Sushko, V.A., Dekhanova, N.G., Kholodenko, U.A.: Employment policy in the conditions of the digital revolution. Dilemmas Contemporáneos: Education, Política y Valores **7**(1), 1–18 (2019)
52. Lopes, N., Rao, H.R., McKenna, S.A., Yang, S., Estevez, E., Nielsen, M.: Panel: digital transformation impact on society. In: 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG), pp. 19–21. IEEE (2019)
53. Kushzhanov, N.V., Mahammadli, Dashqin: The digital transformation of the oil and gas sector in Kazakhstan: priorities and problems. NEWS Natl. Acad. Sci. Repub. Kaz. **3**(435), 203–212 (2019). https://doi.org/10.32014/2019.2518-170X.86
54. Borisova, V., Taymashanov, K., Tasueva, T.: Digital warehousing as a leading logistics potential. In: Strielkowski, W. (ed.) Sustainable Leadership for Entrepreneurs and Academics. SPBE, pp. 279–287. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15495-0_29
55. Kushzhanov, N.V., Aliyev, U.Z.: Digital space: changes in society and security awareness. Bull. Nat. Acad. Sci. Repub. Kaz. **1**, 94–101 (2018)
56. Medapati, P.K., Tejo Murthy, P.H.S., Sridhar, K.P.: LAMSTAR: for IoT-based face recognition system to manage the safety factor in smart cities. Trans. Emerg. Telecommun. Technol. **31**, e3843 (2019)
57. Kaczorowska-Spychalska, D.: Shaping consumer behavior in fashion industry by the interactive communication forms. Fibres Text. Eastern Europe **26**(4(130)), 13–19 (2018)
58. Lindman, J., Tuunainen, V.K., Rossi, M.: Opportunities and risks of Blockchain Technologies–a research agenda. In: Proceedings of the 50th Hawaii International Conference on System Sciences (2017)
59. Dugstad, J., Eide, T., Nilsen, E.R., Eide, H.: Towards successful digital transformation through co-creation: a longitudinal study of a four-year implementation of digital monitoring technology in residential care for persons with dementia. BMC Health Serv. Res. **19**(1), 366 (2019)
60. Cave, M.: How disruptive is 5G? Telecommun. Policy **42**(8), 653–658 (2018)
61. Arntz, M., Gregory, T., Zierahn, U.: Revisiting the risk of automation. Econ. Lett. **159**, 157–160 (2017)
62. Harshanath, S.B.: Detection and protection related to data sharing technologies. In: TENCON 2018–2018 IEEE Region 10 Conference, pp. 0156–0161. IEEE (2018)
63. Kartit, Z., Diouri, O.: Security extension for routing protocols in ad hoc mobile networks: a comparative study. In: Proceedings of the 2nd International Conference on Networking, Information Systems & Security, p. 69. ACM (2019)
64. Prioteasa, AL., Chicu, N., Ciocoiu, C.N.: Implications of digitization on risk management in romanian companies. In: Proceedings of the 31st IBIMA Conference, Milan, Italy (2018)
65. Shah, S., et al.: Compromised user credentials detection in a digital enterprise using behavioral analytics. Futur. Gener. Comput. Syst. **93**, 407–417 (2019)
66. Mitra, A., Munir, K.: Influence of big data in managing cyber assets. Built Env. Project Asset Manag. **9**, 503–514 (2019). https://doi.org/10.1108/BEPAM-07-2018-0098
67. Preuveneers, D., Joosen, W., Ilie-Zudor, E.: Data protection compliance regulations and implications for smart factories of the future. In: 2016 12th International Conference on Intelligent Environments (IE), pp. 40–47. IEEE (2016)
68. Malyavkina, L.I., Savina, A.G., Parshutina, I.G.: Blockchain technology as the basis for digital transformation of the supply chain management system: benefits and implementation

challenges. In: 1st International Scientific Conference “Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth” (MTDE 2019). AtlantisPress (2019)

69. Grigoryeva,, E.E., Sentizova, N.R.: Features of the Russian raw and cut diamonds business digitalization. In: 1st International Scientific Conference “Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth” (MTDE 2019). AtlantisPress (2019)
70. Muntés-Mulero, V., et al.: Agile risk management for multi-cloud software development. IET Softw. **13**(3), 172–181 (2018)
71. Wendler, R.: The maturity of maturity model research: a systematic mapping study. Inf. Softw. Technol. **54**(12), 1317–1339 (2012)
72. Akhlaghpour, S., Lapointe, L.: From placebo to panacea: studying the diffusion of IT management techniques with ambiguous efficiencies: the case of capability maturity model. J. Assoc. Inf. Syst. **19**(6), 441–502 (2018)
73. Fraser, P., Farrukh, C., Gregory, M.: Managing product development collaborations—a process maturity approach. Proc. Inst. Mech. Eng., J. Eng. Manuf. **217**(11), 1499–1519 (2003)
74. Moultrie, J., Clarkson, P.J., Probert, D.: Development of a design audit tool for technology SMEs. J. Prod. Innov. Manag. **24**(4), 335–368 (2007)
75. Keil, M., Tiwana, A., Bush, A.: Reconciling user and project manager perceptions of IT project risk: a Delphi study 1. Inf. Syst. J. **12**(2), 103–119 (2002)
76. Schmidt, R., Lyytinen, K., Keil, M., Cule, P.: Identifying software project risks: an international Delphi study. J. Manag. Inf. Syst. **17**(4), 5–36 (2001)
77. Gioia, D.A., Corley, K.G., Hamilton, A.L.: Seeking qualitative rigor in inductive research: notes on the Gioia methodology. Organ. Res. Methods **16**(1), 15–31 (2013)
78. McDonald, W.J.: Provider perceptions of focus group research use: a multicountry perspective. J. Acad. Mark. Sci. **22**(3), 265–273 (1994)
79. SEI: Software for development, Version 1.2. Software Engineering Institute (2006)
80. Digital Intelligence (DQ): A Conceptual Framework & Methodology for Teaching and Measuring Digital Citizenship. https://www.dqinstitute.org/wp-content/uploads/2017/08/DQ-Framework-White-Paper-Ver1-31Aug17.pdf (2017)
81. European Commission: How European Education Keeps up Nowadays. https://eavi.eu/how-europeaneducation-keeps-up-nowadays-e-learning-and-e-education (2018)
82. McKinsey & Company: Raising your Digital Quotient. https://www.mckinsey.com/businessfunctions/strategy-and-corporate-finance/our-insights/raising-your-digital-quotient (2015)

DTMN a Modelling Notation for Digital Twins

Flavio Corradini, Arianna Fedeli, Fabrizio Fornari(✉), Andrea Polini, and Barbara Re

Computer Science Division, University of Camerino, Camerino, Italy
{flavio.corradini,arianna.fedeli,fabrizio.fornari, andrea.polini,barbara.re}@unicam.it

Abstract. Modelling and developing digital twin solutions is a growing and promising trend followed by enterprises with the ambition to improve decision-making and accelerate risk assessment and production time. However, as a current emerging trend, there is no recognised standard nor a unique solution that provides support for all the characteristics of a digital twin. This article builds upon the result of a literature review that we conducted to extract the main characteristics attributed to Digital Twins. The identified characteristics guided the proposal of a *Digital Twin Modelling Notation* (DTMN). In this work we present the DTMN meta-model supported by a graphical modelling notation. This modelling notation can be used as a starting point to design and reason about Digital Twin solutions.

Keywords: Digital twin · Meta-model · Modelling notation

1 Introduction

The recent advancement in technological fields such as the Internet of Things, Artificial Intelligence, Cloud Computing, Virtual and Augmented Reality, Robotics, Embedded Systems, 3D Modelling, and Simulation are contributing to the rise of Digital Twins that benefit from the mentioned technologies, and that have become more affordable and promise to drive the future of complex systems [29,35]. A digital twin is generally defined as the virtual replica of a real-world entity or process that can be used to run simulations and, therefore, see the effect of changes and choices before deploying them on the actual entity [3]. Gartner's report on Emerging Technologies predicts that the digital twin market will cross the chasm in 2026 to reach $183 billion in revenue by 2031, with composite digital twins presenting the most significant opportunity [48]. Using digital twins can help organisations improve quality and productivity by optimising their decision-making process, reducing costs, and flavouring individual stakeholder demands while optimising resource consumption [23,36]. For this reasons, Digital Twins are recognised as one of the key pioneers of the current industrial revolution [40]. Digital twins enhance the value produced by traditional information systems, allowing the automatic control and information exchange between all the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 63–78, 2023.
https://doi.org/10.1007/978-3-031-26886-1_4

entities involved, such as services, products, clients, transactions, suppliers and many more. This information must be managed optimally to allow continuous adaptations and improvements so to enhance the value of the whole system, which is not feasible in traditional information systems [10].

Although digital twins constitute a growing and promising trend, there is no recognised standard definition or a unique solution to implement all the characteristics and functionalities of a digital twin within the research community [5]. Industry pioneers and researchers have interpreted the Digital Twin concept during the last years, each reporting and emphasising different characteristics. Also, standard organisations attempted to provide a definition of the term digital twin (e.g., ISO[1], NIST[2], Digital Twin Consortium[3]) [24]. Several definitions can be found with different focus (e.g., 3D representation of the entity, data retrieval and analysis, simulation and prediction) [6,33,38,44]. One of the open challenges recognised by the literature is the absence of a modelling language for representing digital twins [9]. Therefore, this paper focuses on the definition of a meta-model and graphical notation that form the *Digital Twin Modelling Notation (DTMN)* that can be used to design and reason about digital twin solutions. The designed digital twin solutions could then be implemented using platforms as software instruments that recently started incorporating some digital twin concepts [26,33,35].

Contribution. The contribution of this work consists in: *(i)* an overview of the main characteristics of digital twins; *(ii)* a detailed description of the developed *Digital Twin Modelling Notation (DTMN)*, with its meta-model and graphical notation provided within the ADOxx meta-modelling platform[4]; *(iii)* a comparison between existing platforms that can be used to implement digital twin solutions. The result of our work can support digital twin solutions designers and also guide: *(1)* the development of new digital twin platforms; *(2)* the extension of already available platforms to fully incorporate the main digital twin characteristics; *(3)* the development of cross-platform solutions that can benefit from the digital twin characteristics already supported by multiple platforms.

Outline. The paper is structured as follows. Section 2 provides an analysis of digital twin characteristics retrieved from the literature. Section 3 describes the Digital Twin Modelling Notation. Section 4 reports a discussion on the implementation of digital twins. Section 5 discusses related work on the design of digital twin solutions. Section 6 concludes the paper highlighting current limitations and proposing insights for future work.

2 Digital Twin Characteristics

A standard and well recognised definition of a digital twin is still missing [9, 20,32]. However, several studies shed light on the characteristics of digital twins

[1] https://www.iso.org/standard/81442.html.
[2] https://csrc.nist.gov/publications/detail/nistir/8356/draft.
[3] www.digitaltwinconsortium.org/.
[4] https://www.adoxx.org/live/home.

and the functionalities they can support. In this section, we report on the results of a study that we conducted to extract digital twin characteristics from the available literature [17]. The identified characteristics are reported in Table 1 with references to the research work that mentions them. Focusing on specific applications domain could have led us to include characteristics that are too domain-specific therefore we mainly analysed research works that discuss the general concept of digital twins without narrowly focusing on specific application domains so to be able to generalise the retrieved characteristics to any digital twin.

Table 1. The Digital Twin characteristics derived from the literature.

	Characteristic/Paper	[7]	[31]	[30]	[37]	[42]	[8]	[11]	[38]	[32]	[47]	[34]	[20]	[49]	[43]	[41]	[39]	[45]	[3]	[28]	[46]	Category
1	Graphical / Visual	✓		✓	✓	✓	✓		✓	✓	✓		✓			✓						Representation
2	Material Composition / Physics Dimensions	✓	✓	✓	✓		✓		✓	✓		✓	✓	✓	✓	✓	✓	✓			✓	
3	Multiple Components	✓	✓	✓			✓	✓		✓	✓	✓	✓	✓	✓	✓			✓	✓		
4	State		✓			✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
5	Context	✓	✓					✓		✓			✓	✓	✓		✓	✓	✓		✓	
6	DT - DT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Interaction
7	DT - Information System		✓				✓	✓	✓					✓			✓		✓		✓	
8	DT - Physical World	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
9	Data Processing	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Functionality
10	Data Analytics	✓	✓	✓			✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
11	Simulation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
12	Augmented Functionalities	✓		✓	✓		✓	✓			✓		✓			✓	✓	✓				

The analysis of the literature revealed twelve main characteristics that are attributed to digital twins. For presentation purposes, we grouped them into three categories: *Representation*, *Interaction*, and *Functionality*.

Representation. A digital twin can be described as a virtual representation of a real-world entity and its characteristics. In representing the digital twin, the analysed definitions focus on different aspects of the real entity that can be represented.

1. Graphic/Visual - A graphical or visual representation of the entity and the data it produces through 2D/3D models can help visualise and reason about the entity to better guide decision-making. Graphical representations could also include the structural schema of the real counterpart (e.g., blueprints). This representation could exploit Virtual and Augmented Reality concepts to represent the digital twin's physical counterpart as faithfully as possible.

2. Material Composition/Physics Dimensions - Information regarding the materials of which the physical counterpart and its components are made, together with their physical dimensions and limitations (e.g., the degrees of freedom of the entire entity or of a specific component) is essential for a proper representation of the physical entity and its possible behaviour.

3. Multiple Components - A physical entity is often composed of smaller components (e.g., machinery can have several mechanical parts). Such components can be represented by digital twins, leading to a holonic composition of digital twins.

4. State - Keeping track of all the states in which a physical entity can be (e.g., turned on, turned off, active, idle, or error state) is essential to properly monitor it and to reason about the actions it is performing so to plan further operations.

5. Context - Context information is necessary to uniquely describe, represent, and categorise a digital twin and its relation with the environment and the context in which it operates. This type of information can contribute to provide a more specific representation of the virtual entity and its physical counterpart [27].

Interaction. Digital twins are also attributed with the capability to interact with each other and with traditional information systems and the physical world. This interaction can be bidirectional, as digital twins can receive and send data.

6. DT - DT - The possibility for digital twins to exchange data is seen as an important characteristic necessary for establishing networks or communities of digital twins. Indeed, a digital twin should interact with other digital twins, even outside the same organization. In this way, a digital twin can request and exchange information with one another to learn additional information that could modify or improve its behaviour. The information that an entity could send or receive can report environmental data (e.g., an entity that senses the temperature of a room could inform another regarding the room state) and those that describe the physical entity itself (e.g., its battery percentage and energy consumption so to advice if some problem occurs).

7. DT - Information System - Digital twins should exchange data with information systems to collaborate and provide more elaborated services. This way, a digital twin can send and receive data from/to external sources or services.

8. DT - Physical World - As a digital twin is a digital representation of a physical entity, it is fundamental that a connection between the physical and the virtual replica exists so that the digital twin can receive and send data from/to the physical entity. This interaction is necessary to maintain the synchronisation between the physical and virtual entities in such a way as to reflect changes from the virtual world to the physical one and vice versa. This process is often referred to as *twinning process* [28]. For example, when an action is requested from a digital twin, a command could be sent to the physical counterpart so that it also performs that action. Vice versa, a performed action in the physical world will require updating the virtual replica with data and information about the new reached state.

Functionality. All definitions analysed have inherent features for which a digital twin can provide support.

9. Data Processing - It refers to operations that a digital twin can perform, such as filtering, removing, merging and modifying data. According to all the works analysed, manipulating this data is an important characteristic of understanding the conditions of the real environment and planning actions that the physical counterpart will perform.

10. Data Analytics - It refers to the possibility of analysing historical data or data streams to understand the occurred events and applying machine learning and artificial intelligence to predict subsequent events or outcomes.

11. Simulation - A characteristic of DTs that is fully recognised by all the analysed works is the possibility to conduct a simulation of data and state variations concerning the physical counterpart. This allows for foreseeing the outcomes of some actions the physical counterpart could perform and predicting undesired situations (e.g., reaching an error state).

12. Augmented Functionalities - The possibility of having a virtual replica of the entity that enhances the characteristics of the physical counterpart to perform additional tasks are highlighted in different works. Providing advanced mechanisms increases the value of the analysed data. For example, a digital twin that can autonomously reason about its status and perform possible automatic adjustments is something the physical counterpart could not do by itself.

The identified digital twins characteristics guided the proposal of a modelling notation described in the following sections.

3 Digital Twin Modelling Notation

In this section, we introduce the *Digital Twin Modelling Notation (DTMN)* to design and reason about digital twin solutions. We describe the meta-model and the respective graphical notation showing its usage in a sample scenario. *DTMN* is provided as a library within the ADOxx meta-modelling platform and made available to the public.[5]

3.1 The DTMN Meta-model

The meta-model proposed in Fig. 1, aims to represent all the main characteristics attributed to a digital twin, as described in Sect. 2. Here, we report a description of the meta-model dividing it into different modules, each representing a specific concept related to digital twins: *Entity Management*, *Data Manipulation*, *Visualisation*, *Entity State*, and *Context*. In the following, we provide a detailed explanation of each module, highlighting which characteristics they fulfil.

Entity Management. This module describes all the information concerning a digital twin entity. A digital twin is a virtual representation of a physical *Entity* and it aims to represent it in its entirety. An entity could be also a device that can be catalogued as a sensor, an actuator or a tag based on its functionality. Information regarding the physical entity can be represented, such as the

[5] The proposed library and additional information about *DTMN* are available on: https://pros.unicam.it/dtmn/.

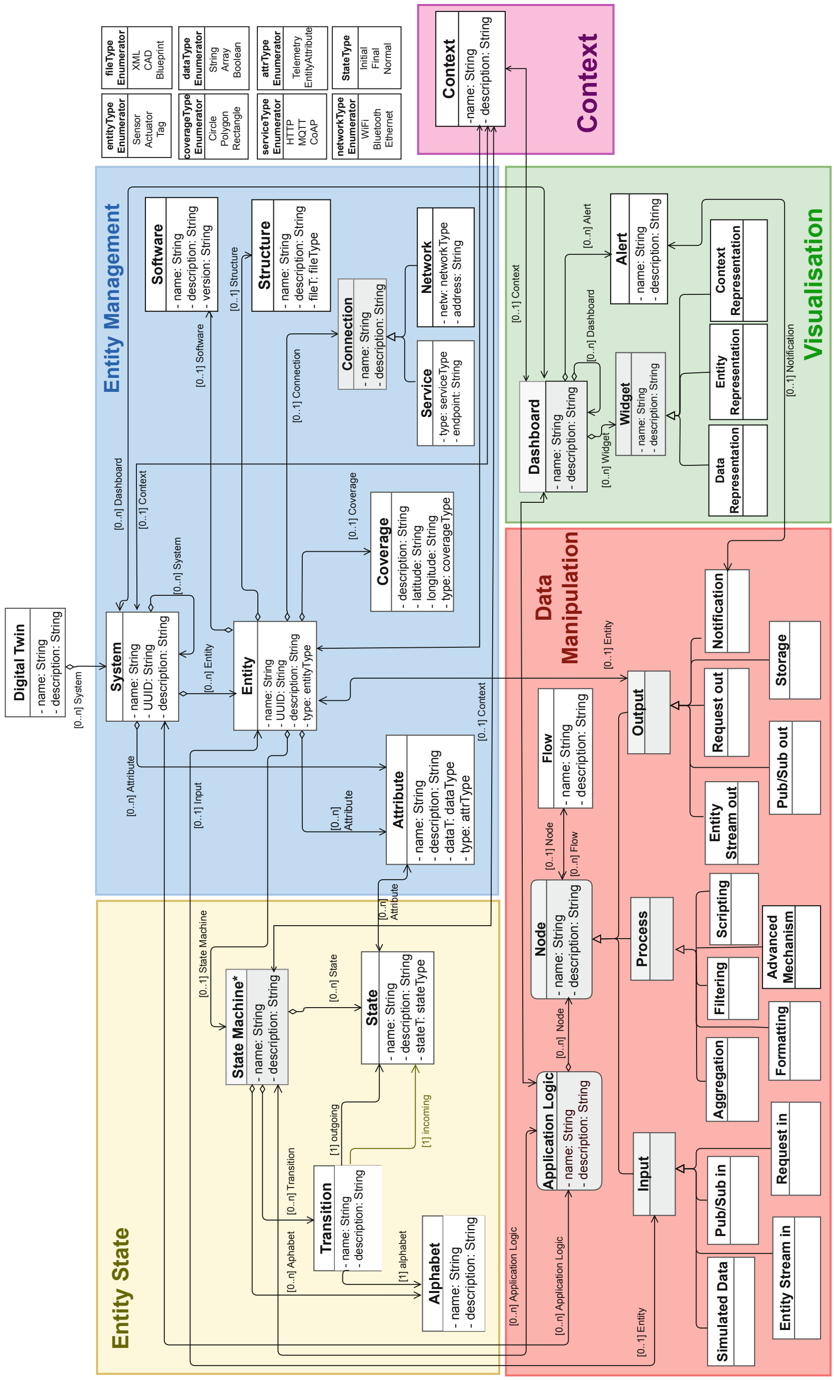


Fig. 1. The proposed digital twin meta-model.

exact *Coverage* location of the physical entity in the geographic space (e.g., latitude and longitude) and the geometric area of the coverage action range of the entity in the actual physical environment. The *Structure* of an entity, together with the composition of its materials and physical dimensions (e.g., the height, width and weight), can be represented by referring to the parts that compose it **(Char. 2)**. Files that report such structural information could also be associated with the digital twin (e.g., CAD, XML, blueprints). Also, details regarding the *Software* associated to the entity can be represented (e.g., operating system, firmware). Information regarding the connection between the virtual and the physical world can be defined. The *Connection* element incorporates the *Network* and the *Service* elements, as they represent the possibility of defining an entry point to interact with the physical world to send and receive data **(Char. 8)**. The *Network* defines how the entity is connected and available to others (e.g., Bluetooth, WiFi, Ethernet or other kinds of networks) and has attributes such as the entity's address (e.g., IP address and MAC address). Instead, the *Service* element describes the protocol that the entity uses to communicate (e.g., MQTT, HTTP and CoAP protocols). Such a connection permits a constant data exchange with the digital twin that can be updated in real-time to properly reflect events in the physical world. The data sent and received from/to the digital twin are represented in the meta-model by the *Attribute* element. Attributes are data collected from the environment and the entity itself. The former is referred to as telemetry data. (e.g., we can refer to the temperature of a room produced on a virtual level by receiving and manipulating the data provided by different temperature sensors). On the other hand, the latter are entity attributes such as battery status, current voltage, and entity amperage. A digital twin can also be described as the composition of one or more digital twins of physical entities [7] that operate together to achieve a complex goal. This is defined in the meta-model as a *System*. At the same time, a digital twin can be seen as a system of digital twins when digital twins represent the parts that compose the physical entity **(Char. 3)**. However, having a digital twin of each component may be an expensive, overkill and not useful approach in some instances.

Context. Digital twins of entities have a strong relation with the environment **(Char. 5)** in which their physical counterpart is deployed [21]. As reported above, data from the environment can be captured by physical entities, which allows for realising part of the connection between the physical and the virtual world. The context can be represented textually, through a detailed description, or through a graphical representation. This allows for a better representation of the physical world and provides valuable solutions. Examples of a context representation can be a description of the digital twin context, a 2D map, or a 3D model of a room, which reports the positioning of each entity.

Entity State. This module deals with the representation of the possible digital twin states. To represent them, we relied on state machine representation that has already been used for similar purposes [1,51]. Therefore, we included in the meta-model the concept of *State*, which allows representing all the states on which an entity can be (e.g., states can be on, off, occupied, idle, etc.) **(Char. 4)**. The passage from one state to another is represented through a *Transition*. Each transition is labelled with the input *Alphabet* symbol that generates that transition. This concept is widely used with digital twins to express entity behaviour by simulating it through attributes (e.g., affect what happens if an entity moves from one state, defined as the current state, to another, defined as the desired state).

Visualisation. This module describes all the information concerning the graphical visualisation of the digital twin data and its physical characteristics (e.g., shape and structure) **(Char. 1)**. Having a clear visualisation of these data allows, also in real-time, performing analyses and forecasts on future events based on current data generating improvements to the entire system. Usually, the visualisation is provided through a graphical *Dashboard* that provides an intuitive view of the system's data from various entities. A dashboard can contain different *Widgets* that exposes data, entity and context in a graphical format (e.g., a numerical temperature value is displayed in a gauge widget, while its historical data is in the tabular form). These widgets can be interactive and enable the end-users to modify the physical environment's parameters remotely (e.g., it is possible to send a signal to the air conditioner system to change its state and then turn it on or off, based on the end user's necessities). Widgets can also display a 2D/3D visualisation of the entity and allow a user to interact with the virtual entity for inspection and simulate a real interaction with the physical entity. The widget can be used to visualise the context in which the physical entity has been deployed in a graphical representation of the surroundings and of the other entities or digital twins that populate the area. The ability to have *Alerts* in the dashboard is also crucial for analysing data and events in real-time, especially for error events (e.g., excess current or unexpected error from an entity).

Data Manipulation. This module represents how a digital twin can process and exchange data. This characteristic involves collecting and processing data streams coming from multiple entities to produce more valuable data **(Char. 9)** or to generate commands that can be sent to physical entities, as well as other digital twins **(Char. 6)**, and external systems **(Char. 7)**. In the metamodel the concepts of *Node* and *Flow* are used to describe the entire *Application Logic*. A node can be seen as an atomic processing unit that receives input data, operates over it and produces a specific output. We categorise nodes based on the support they provide for the main characteristics of an information system: Input, Process, and Output. These characteristics refer to the widely used IPO (Input-Process-Output) model [50], an approach in software engineering to represent the structure of a process or an information processing program. The *Input* element groups the operations that can be performed to retrieve data and define data collections from different sources. The *Process* element groups the operations that can be performed to modify data or elaborate them. The *Output* element groups the operations that can be performed to send the obtained data to a target. In particular, inside the Process element group, the *Advanced Mech-*

anisms element can provide support for *Data Analytics* **(Char. 10)** through such things as machine learning and artificial intelligence. This element can be used for making an upcoming prediction about events or outcomes related to the digital twin or the context of the physical counterpart. The *Simulation* service can also enable the possibility of assigning fictitious values to data that usually are retrieved from the physical entity environment or from the entity itself. This would allow foreseeing effects on the digital twin activities and states, together with the impact on the related digital twin system, the other digital twins, and the entire context **(Char. 11)**. Graphical representations for digital twins and their context would also allow one to experience such a simulation visually; if virtual reality is supported, the simulation could be even more immersive. Being a virtual entity, the digital twin can easily be modified and enhanced to provide *Augmented Functionalities* **(Char. 12)**. The possibility to develop new application logic - new software that can be directly assigned to a digital twin - allows for enhancing the digital twin with capabilities that the physical entity does not present.

3.2 The Proposed Notation

Based on the presented meta-model, we designed and implemented a graphical notation made available as a library for the ADOxx metamodelling platform. The proposed notation is intended to provide first support for digital twin solution designers.

We present in Fig. 2 the usage of *DTMN* to design a digital twin solution for a sample scenario regarding a smart meeting room.

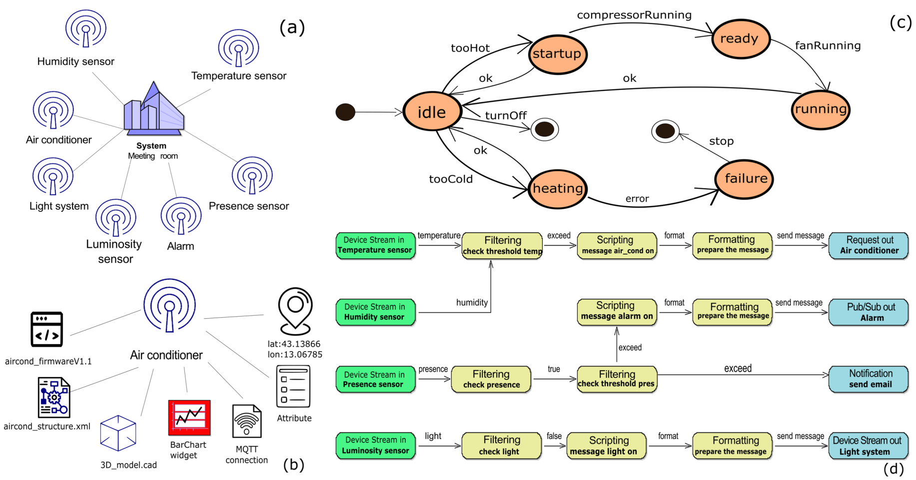


Fig. 2. Models representing (a) all the entities involved in the meeting room system, (b) a detailed visualisation of an air conditioner digital twin, (c) the air conditioner state machine, and (d) the entire meeting room system application logic.

Since a digital twin comprises numerous elements, the notation supports its modelling from several perspectives. In the following, we describe the notation elements highlighting their correspondence with the meta-model's modules.

Figure 2(a), shows the ***Context*** perspective through a model of the meeting room and the various entities involved. It is possible to assign basic information such as the name and description of systems and entities. Then, for each entity, additional specific information can be inserted.

Figure 2(b) shows this ***Entity Management*** perspective, representing an Air conditioner that could be further detailed with additional elements corresponding to digital twin characteristics such as its structure, attributes, location, connection, or corresponding description files, as described in the meta-model. In addition, there is the possibility to provide a graphical **Visualisation** about the entity data, inserting a widget. Representing each entity in a system as a digital twin may be overkill. *DTMN* makes it possible to avoid treating all entities as digital twins. The choice is left to the modeller and system developer based on the needs of that scenario.

It is also possible to model each ***Entity State***. Figure 2(c) refers to the perspective of the Air conditioner state machine. For each digital twin, it is possible to design a state machine that represents the states in which it can be and the transition that can cause a state change.

Finally, Fig. 2(d) shows the ***Data Manipulation*** perspective: a model describing the logic that guides the interactions between entities, digital twins, and external information services for the control management of the smart meeting room. Different nodes represent different and specific functionalities. In this case, nodes and flows allow to: handle the air conditioner based on temperature and humidity, email the manager, activate an alarm in an overcrowding situation, and handle the light system automatically through luminosity sensors.

4 Digital Twin Implementation

The model designed with our *DTMN* notation allows the representation of high-level digital twin concepts without considering the actual implementation of the system. Given all the different digital twin characteristics, Sect. 2, identifying a unified solution that can allow the implementation of a digital twin entirely is a non-trivial task. In fact, at present, no tool or platform can fully support the development of a digital twin [35].

The Internet of Things (IoT) has been recognised as the backbone for developing and using digital twins [25]. In particular, IoT platforms are software instruments generally used to organise and handle: IoT devices and their interactions, the data they capture from the environment, and the manipulation, processing, the visualisation of such data. In addition, they also provide support for the interaction with external services. All of these IoT platforms functionalities allow managing some digital twin characteristics [26,33] and making them become a suitable option for implementing digital twin solutions [6,7,35]. We performed scouting and reported in Table 2 the available IoT platforms that provide support for digital twins.

Table 2. Comparison of Digital Twin characteristics and IoT Platform functionalities. (✓ for supported; * for partially supported; empty for no support)

Platform/ Characteristic	**Virtual Representation**					**Physical and Virtual World Connection**			**Enabled Services**			
	1	2	3	4	5	6	7	8	9	10	11	12
Amazon AWS IoT	*		✓		*	✓	✓	✓	✓	✓	✓	✓
Microsoft Azure	*	*	✓		*	✓	✓	✓	✓	✓	*	✓
IBM Watson	*	*	✓		*	✓	✓	✓	✓	✓	*	✓
Bosch IoT	*		✓		*	✓	✓	✓	✓	✓	✓	✓
Google Cloud IoT Core	*		✓		*	✓	✓	✓	✓	✓	*	✓
Losant	*		✓		*	✓	✓	✓	✓		*	✓
Thingsboard	*		✓		*	✓	✓	✓	✓	*	*	✓
KAA	*		✓			✓		✓	*	*	*	
Particle	*		✓			✓	✓	✓	✓	*	✓	✓
Carriots	*		✓		*	✓	✓	✓	✓		*	✓
SAP Leonardo	*	*	✓		*	✓	✓	✓	✓	✓	*	✓
ThingWorx	*		✓		*	✓	✓	✓	✓	✓	*	✓
MindSphere	*		✓		*	✓	✓	✓	✓	✓		✓
Lumada	*		✓		*	✓	✓	✓	✓			✓
Predix	*		✓		*	✓	✓	✓	✓	*	✓	✓

Most analysed platforms can implement the data exchange from and to physical entities and external services (**Char 6–8**). Also, the possibility to represent an entity as a composition of multiple ones is a peculiarity of IoT platforms which often refer to such a concept with the name *asset* **(Char. 3)**. All the platforms support the collection, aggregation, and manipulation of a vast amount of data coming from different entities and sources **(Char. 9)**. Most platforms also provide support for storing data which enables the possibility to perform data analytics, with the use machine learning and artificial intelligence, to derive insights that could guide decision making **(Char. 10)**. IoT platforms can also be used to develop additional functionalities that the physical counterpart of a digital twin does not present (e.g., sending an email when the entity is in an error state) **(Char. 12)**.

Some IoT platforms provide a form of graphical representation **(Char. 1)**. Digital twin data representation is provided through dashboards and widgets. Few platforms allow storing information regarding the material composition of the physical entity and its physics dimensions **(Char. 2)**. Almost all the platforms support simulations **(Char. 11)**. Still, they do not provide advanced mechanisms of simulation integrating virtual reality.

None of the analysed platforms fully provide a precise way to represent the context in which an entity is deployed **(Char. 5)**. Platforms can only support the description of the digital twin's space and who can interact with it. None of the platforms provides a specific representation or model of the digital twin state space (e.g., state machines) **(Char. 4)**. Most of them rely on the data the entity provides that need to be analysed to understand their state.

Different solutions could be applied to overcome the partial support provided by IoT platforms in supporting digital twins. A first solution would require a high IoT expertise of the actors involved, high-level programming skills and a high amount of time and resources to develop a digital twin platform [22] that incorporates support for all the digital twins' characteristics. Studies discourage the development from scratch of a new platform as it requires too much effort in terms of cost, time and resources. It is estimated that approximately 2.5 years are needed to develop a new platform [2]. To avoid the development of a new platform, a second solution could involve extending one of the already existing IoT platforms adding the missing support for digital twins' characteristics. In this case, advanced knowledge of the chosen IoT platform is required. Therefore the IoT platform company itself, or at its place, an open community of developers (if the platform is open-source), could take care of extending it.

The use of a cross-platform approach that allows the design of a digital twin solution once and the deployment on multiple target platforms can be envisaged as another possible solution [18]. This approach would allow digital twin solution designers to select multiple target platforms based on their specific support for each digital twin characteristic.

5 Related Work

Several characteristics are attributed to digital twins, as also emerged from the study we conducted and reported in Sect. 2. As well as for other domains (e.g., Business Process Management, Internet of Things), modelling notations (e.g., BPMN, UML) can be used to design and reason about complex systems [14–16]. Concerning digital twins, the design of a modelling language has been recognised as a potential means to manage the complexity of digital twin solutions design [4,9,12,13]. Different works provide designing approaches to develop digital twins but refer to specific application domains. As examples in the manufacturing environment, a re-configurable digital twin model has been used to implement a shop-floor configuration [52], as well as a digital twin for an injection moulding system [19]. The authors design digital twins and define cockpits to view, monitor and aggregate digital twin data. In [36], the authors analyse the suitability of enterprise modelling and capability management for developing and managing business-driven digital twins. The models reported in those related works include some of the elements provided in our meta-model, such as those referring to physical entities (physics, geometry, capability information), virtual models representing behaviour and rules, services, digital twins data and connections.

Referring to the implementation of the digital twin solutions, several tools and technologies have been analysed [35]. In particular, the Internet of Things is seen as the key enabler to filling the gap between the physical and virtual worlds, implementing digital twin characteristics [11]. Therefore, it seems reasonable to start implementing digital twin solutions from already available IoT platforms.

6 Conclusions and Future Work

In this work, we provided an overview of the main digital twin characteristics and presented our proposal for a *Digital Twin Modelling Notation (DTMN)*. The first version of *DTMN* is implemented within the ADOxx meta-modelling platform and can be extended to incorporate additional or domain-specific characteristics. Overall, our contribution aims at supporting digital twin solution designers providing them with an instrument to design and reason about digital twin solutions. The result of our work can also be taken as a reference to guide: the development of new digital twin platforms, the extension of already existing ones, or the adoption of a cross-platform approach to support the development of the digital twin across multiple platforms. In such a direction, we discussed IoT platforms that may be used to implement the designed solutions, highlighting the digital twin characteristics they support.

As future work, we consider conducting a validation through real scenarios involving domain experts and developers. We also foresee the possibility of providing support for implementing the designed digital twin solutions.

Acknowledgements. This work has been partially supported by the MIUR project PRIN "Fluidware" (A Novel Approach for Large-Scale IoT Systems, n. 2017KRC7KT) and by Marche Region in implementation of the financial programme POR MARCHE FESR 2014-2020, project "Miracle" (Marche Innovation and Research fAcilities for Connected and sustainable Living Environments), CUP B28I19000330007.

References

1. Alam, K.M., El-Saddik, A.: C2PS: a digital twin architecture reference model for the cloud-based cyber-physical systems. IEEE Access **5**, 2050–2062 (2017)
2. Analytics, I.: IoT platforms: the central backbone of the Internet of Things. In: White Paper (2015)
3. Anto Budiardjo, D.M.: Digital twin system interoperability framework. Digital Twin Consortium (2021)
4. Atkinson, C., Kuhne, T.: Model-driven development: a metamodeling foundation. IEEE Softw. **20**(5), 36–41 (2003)
5. Atkinson, C., Kühne, T.: Taming the complexity of digital twins. IEEE Softw. **39**(2), 27–32 (2022)
6. Ayoobkhan, M.U.A., et al.: Smart connected digital products and IoT platform with the digital twin, pp. 330–350. IGI Global (2021)
7. Bhattacharyya, A., Izgi, E.: Digital twin technologies for high performance manufacturing. IBM White paper (2018)
8. Boje, C., Guerriero, A., Kubicki, S., Rezgui, Y.: Towards a semantic construction digital twin: directions for future research. Autom. Constr. **114**, 103179 (2020)
9. Bordeleau, F., et al.: Towards model-driven digital twin engineering: current opportunities and future challenges. In: International Conference on Systems Modelling and Management, pp. 43–54 (2020)
10. Botkina, D., Hedlind, M., Olsson, B., Henser, J., Lundholm, T.: Digital twin of a cutting tool. Procedia Cirp **72**, 215–218 (2018)

11. Christian, J., et al.: Model-driven digital twin construction: synthesizing the integration of cyber-physical systems with their information systems. In: Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, pp. 90–101 (2020)
12. Clark, T., Kulkarni, V., Whittle, J., Breu, R.: Engineering digital twin-enabled systems. IEEE Softw. **39**(2), 16–19 (2022)
13. Clark, T., Sammut, P., Willans, J.: Applied metamodelling: a foundation for language driven development. arXiv:1505.00149 (2015)
14. Compagnucci, I., Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F.: Modelling notations for IoT-aware business processes: a systematic literature review. In: Del Río Ortega, A., Leopold, H., Santoro, F.M. (eds.) BPM 2020. LNBIP, vol. 397, pp. 108–121. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66498-5_9
15. Corradini, F., Fedeli, A., Fornari, F., Polini, A., Re, B.: FloWare: an approach for IoT support and application development. In: Augusto, A., Gill, A., Nurcan, S., Reinhartz-Berger, I., Schmidt, R., Zdravkovic, J. (eds.) BPMDS/EMMSAD -2021. LNBIP, vol. 421, pp. 350–365. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-79186-5_23
16. Corradini, F., Fedeli, A., Fornari, F., Polini, A., Re, B.: Floware: a model-driven approach fostering reuse and customisation in IoT applications modelling and development. Softw. Syst. Model. 1–28 (2022). https://doi.org/10.1007/s10270-022-01026-9
17. Corradini, F., Fedeli, A., Fornari, F., Polini, A., Re, B.: Towards a digital twin modelling notation. In: DASC/PiCOM/CBDCom/CyberSciTech22. IEEE (2022)
18. Corradini, F., Fedeli, A., Fornari, F., Polini, A., Re, B.: X-IoT: a model-driven approach for cross-platform IoT applications development. In: SAC 2022 pp. 1448–1451 (2022)
19. Dalibor, M., et al.: Towards a model-driven architecture for interactive digital twin cockpits. Concept. Model. **12400**, 377–387 (2020)
20. Demkovich, N., Yablochnikov, E., Abaev, G.: Multiscale modeling and simulation for industrial cyber-physical systems. In: IEEE Industrial Cyber-Physical Systems, pp. 291–296 (2018)
21. Dobrescu, R., Merezeanu, D., Mocanu, S.: Context-aware control and monitoring system with IoT and cloud support. Comput. Electron. Agric. **160**, 91–99 (2019)
22. Fahmideh, M., Zowghi, D.: An exploration of IoT platform development. Inf. Syst. **87**, 1–25 (2020). Article number 101409
23. Fill, H.: Enterprise modeling: from digital transformation to digital ubiquity. In: Federated Conference on Computer Science and Information Systems, vol. 21, pp. 1–4 (2020)
24. Harper, K.E., Ganz, C., Malakuti, S.: Digital twin architecture and standards. IIC J. Innov. **12**, 72–83 (2019)
25. He, Y., Guo, J., Zheng, X.: From surveillance to digital twin: challenges and recent advances of signal processing for industrial internet of things. IEEE Signal Process. **35**(5), 120–129 (2018)
26. Hoffmann, J., Heimes, P., Senel, S.: IoT platforms for the internet of production. IEEE Internet Things J. **6**(3), 4098–4105 (2019)
27. Hribernik, K., Cabri, G., Mandreoli, F., Mentzas, G.: Autonomous, context-aware, adaptive digital twins-state of the art and roadmap. Comput. Ind. **133**, 103508 (2021)
28. Jones, D., Snider, C., Nassehi, A., Yon, J., Hicks, B.: Characterising the digital twin: a systematic literature review. J. Manuf. Sci. Technol. **29**, 36–52 (2020)

29. Kulkarni, V., Clark, T.: Towards adaptive enterprises using digital twins. In: International Conference on Research Challenges in Information Science, pp. 1–5 (2019)
30. Lim, K.Y.H., Zheng, P., Chen, C.: A state-of-the-art survey of digital twin: techniques, engineering product lifecycle management and business innovation perspectives. J. Intell. Manuf. **31**(6), 1313–1337 (2020)
31. Lu, Y., Liu, C., Wang, K.I., Huang, H., Xu, X.: Digital Twin-driven smart manufacturing: connotation, reference model, applications and research issues. Robot. Comput. Integr. Manuf. **61**, 101837 (2020)
32. Madni, A.M., Madni, C.C., Lucero, S.D.: Leveraging digital twin technology in model-based systems engineering. Syst. **7**, 7 (2019)
33. Minerva, R., Lee, G.M., Crespi, N.: Digital twin in the IoT context: a survey on technical features, scenarios, and architectural models. IEEE Proc. **108**(10), 1785–1824 (2020)
34. Negri, E., Fumagalli, L., Macchi, M.: A review of the roles of digital twin in cps-based production systems. Procedia Manuf. **11**, 939–948 (2017)
35. Qi, Q., et al.: Enabling technologies and tools for digital twin. Manuf. Syst. J. **58**, 3–21 (2021)
36. Sandkuhl, K., Stirna, J.: Supporting early phases of digital twin development with enterprise modeling and capability management: requirements from two industrial cases. Enterp. Bus. Process Inf. Syst. Model. **387**, 284–299 (2020)
37. Schroeder, G., et al.: Visualising the digital twin using web services and augmented reality. In: Conference on Industrial Informatics, pp. 522–527 (2016)
38. Schroeder, G.N., Steinmetz, C., Rodrigues, R.N., Henriques, R.V.B., Rettberg, A., Pereira, C.E.: A methodology for digital twin modeling and deployment for industry 4.0. Proc. IEEE **109**(4), 556–567 (2021)
39. Semeraro, C., Lezoche, M., Panetto, H., Dassisti, M.: Digital twin paradigm: a systematic literature review. Comput. Ind. **130**, 103469 (2021)
40. Singh, M., Fuenmayor, E., Hinchy, E.P., Qiao, Y., Murray, N., Devine, D.: Digital twin: origin to future. Appl. Syst. Innov. **4**(2), 1–19 (2021). Article number 36
41. Stark, R., Damerau, T.: Digital twin. In: Chatti, S., Tolio, T. (eds.) CIRP Encyclopedia of Production Engineering. Springer, Berlin, Heidelberg (2019). https://doi.org/10.1007/978-3-642-35950-7_16870-1
42. Steinmetz, C., Rettberg, A., Ribeiro, F.G.C., Schroeder, G., Pereira, C.E.: Internet of Things ontology for digital twin in cyber physical systems. In: Symposium on Computing Systems Engineering, pp. 154–159 (2018)
43. Talkhestani, B.A., Jazdi, N., Schloegl, W., Weyrich, M.: Consistency check to synchronize the digital twin of manufacturing automation based on anchor points. Procedia CIRP **72**, 159–164 (2018)
44. Tao, F., Qi, Q., Wang, L., Nee, A.: Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: correlation and comparison. Engineering **5**(4), 653–661 (2019)
45. Tao, F., Zhang, H., Liu, A., Nee, A.Y.C.: Digital twin in industry: state-of-the-art. IEEE Trans. Ind. Inf. **15**(4), 2405–2415 (2019)
46. Trauer, J.E.A.: What is a digital twin? - definitions and insights from an industrial case study in technical product development. In: DESIGN Conference, vol. 1, pp. 757–766 (2020)
47. VanDerHorn, E., Mahadevan, S.: Digital twin: generalization, characterization and implementation. Decis. Support Syst. **145**, 113524 (2021)
48. Velosa, A., Middleton, P.: Emerging Technologies: Revenue Opportunity Projection of Digital Twins, Gartner, Stamford, Conn. (2022). Accessed June 2022

49. Vrabič, R., Erkoyuncu, J.A., Butala, P., Roy, R.: Digital twins: understanding the added value of integrated models for through-life engineering services. Procedia Manuf. **16**, 139–146 (2018). int. Conf. on Through-life Engineering Services
50. Davis, W.S.: HIPO hierarchy plus input-process-output. the information system consultant's handbook: systems analysis and design, pp. 503–511 (1988)
51. Xiao, R., Wu, Z., Wang, D.: A finite-state-machine model driven service composition architecture for Internet of Things rapid prototyping. Futur. Gener. Comput. Syst. **99**, 473–488 (2019)
52. Zhang, C., Xu, W., Liu, J., Liu, Z., Zhou, Z., Pham, D.T.: A reconfigurable modeling approach for digital twin-based manufacturing system. Procedia Cirp **83**, 118–125 (2019)

Adaptive Management of Cyber-Physical Workflows by Means of Case-Based Reasoning and Automated Planning

Lukas Malburg[1,2](✉), Florian Brand[1], and Ralph Bergmann[1,2]

[1] Artificial Intelligence and Intelligent Information Systems, University of Trier, 54296 Trier, Germany
{malburgl,s4fnbran,bergmann}@uni-trier.de

[2] German Research Center for Artificial Intelligence (DFKI), Branch University of Trier, Behringstraße 21, 54296 Trier, Germany
{lukas.malburg,ralph.bergmann}@dfki.de
http://www.wi2.uni-trier.de

Abstract. Today, it is difficult for companies to react to unforeseen events, e. g., global crises. Highly standardized manufacturing processes are particularly limited in their ability to react flexibly, creating a demand for more advanced workflow management techniques, e. g., extended by artificial intelligence methods. In this paper, we describe how *Case-Based Reasoning (CBR)* can be combined with automated planning to enhance flexibility in cyber-physical production workflows. We present a compositional adaptation method complemented with generative adaptation to resolve unexpected situations during workflow execution. This synergy is advantageous since CBR provides specific knowledge about already experienced situations, whereas planning assists with general knowledge about the domain. In an experimental evaluation, we show that CBR offers a good basis by reusing cases and by adapting them to better suit the current problem. The combination with automated planning further improves these results and, thus, contributes to enhance the flexibility of cyber-physical workflows.

Keywords: Case-based reasoning · Automated planning · Industry 4.0 · Adaptive workflow management · Cyber-Physical Workflows

1 Introduction

Recently, global crises have shown that manufacturing processes and in general supply chains cannot easily be adapted to respond to unforeseen and dynamic events. This is among others because manufacturing processes are often highly standardized and therefore only provide a limited degree of flexibility [12,16]. One of the goals of the *Fourth Industrial Revolution (Industry 4.0)* [12] is to enhance this limited flexibility by applying *Artificial Intelligence (AI)* methods [13] in *Cyber-Physical Production Systems (CPPSs)* [20]. Consequently, failure

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 79–95, 2023.
https://doi.org/10.1007/978-3-031-26886-1_5

situations that occur during production should be resolved in an autonomous, self-adaptive way, strengthening resilience of workflows against failures and unexpected situations during their execution [15,25]. To adapt workflows automatically and, in turn, to ensure their continuing execution in problem situations, current research applies search-intensive approaches such as AI planning (e. g., [7,17,24]) and knowledge-intensive approaches such as *Case-Based Reasoning (CBR)* [1] (e. g., [19,21,28,29]). However, search-intensive techniques require fully observable planning domain descriptions, which are rather rare and difficult to obtain in real-world applications, to generate appropriate solutions [22,30]. In addition, AI planning is sometimes not applicable for large problems due to the high computational complexity [4,5,27]. To remedy these issues, the combination of AI planning and CBR offers significant potential for improvement, leading to research directions such as *Case-Based Planning* [4,5,27,30] in which plans are reused in similar situations. CBR, in this regard, provides specific experience knowledge that can be utilized in similar problem situations, limiting the knowledge acquisition and modeling effort for a comprehensive planning domain. However, today, existing approaches are not examined for production planning or for advanced adaptive workflow management in which currently either pure planning techniques or pure CBR methods are applied (e. g., [17,18,28,29]). For this purpose, this paper presents application scenarios in which *Case-Based Reasoning (CBR)* [1] can contribute to enhance the flexibility of cyber-physical manufacturing workflows. Thereby, we focus on automatic workflow adaptations in the *Reuse* phase to resolve unexpected failure situations that can occur during workflow execution (see [15] for the architectural framework). Compositional adaptation is used with AI planning to overcome the outlined disadvantages and to apply both methods in a synergistic way. A key advantage of our approach is that it limits the knowledge acquisition and modeling effort typically required for creating comprehensive planning domains by incorporating experiential knowledge during problem-solving. To evaluate the approach, we use a physical smart factory model from *Fischertechnik (FT)*, which enables us to conduct laboratory experiments while maintaining real world environmental conditions of production lines. In the following, Sect. 2 presents the used physical smart factory and describes application scenarios in which CBR can contribute to enhance flexibility of production workflows. Our approach for automatic workflow adaptation by combining compositional adaptation and AI planning is presented in Sect. 3. To measure the effectiveness and suitability of the approach, we present an experimental evaluation in Sect. 4. Finally, Sect. 5 summarizes the paper and gives an outlook on future research directions.

2 Foundations and Related Work

We describe the main characteristics of cyber-physical workflows and present the used smart factory model in Sect. 2.1. Afterwards, we discuss in Sect. 2.2 related work using AI-based methods for adaptive workflow management. Finally, Sect. 2.3 introduces *Process-Oriented Case-Based Reasoning (POCBR)* [3] as a special kind of CBR for reusing procedural experiential knowledge. In addition, application scenarios in which POCBR can contribute to enhance flexibility in cyber-physical workflows are described.

2.1 Cyber-Physical Workflows and Physical Smart Factory

Cyber-Physical Workflows (CPWs) [18,25] are a new branch of workflows in which the presence of *Internet of Things (IoT)* technologies influences the execution of workflows in the real world and vice versa. For example, actuators such as machines are used to execute tasks in the environment and sensor data from IoT devices can be used for guiding workflow execution or to detect failures during production. Based on these detected situations, AI techniques can be applied to resolve problems and to continue workflow execution in the physical world [16,17]. These advantages can be achieved for workflow management by exploiting the bidirectional coupling between process-based systems and the smart environment [26] and, thus, by using the variety of IoT sensor data from it. The environment itself benefits by using well-established methods from workflow management research [10]. However, to profit from these advantages, several challenges must be addressed (cf. [10]), which is why the use of advanced AI methods in this area is still in its infancy.

Using IoT environments such as real manufacturing shop floors for research purposes poses many difficulties [16]. Thus, small-scale physical models can be used for executing manufacturing workflows. We use a smart factory model from *Fischertechnik (FT)* to conduct process-oriented research for Industry 4.0 [14, 16,26]. The custom model[1] emulates two independently working production lines consisting of two shop floors that are linked to exchange workpieces (see Fig. 1). Each production line consists of six identical machines. In addition, there are individual machines on each shop floor, i. e., a *Punching Machine (PM)* and a *Human Workstation (HW)* on the first shop floor and a *Drilling Machine (DM)* on the second one. For control purposes, there are several light barriers, switches, and capacitive sensors on each shop floor. The workpieces used for simulating manufacturing are small cylindrical blocks. Each workpiece is equipped with an NFC tag with information about the individual workpiece such as the current production state and the production history with time stamps.

2.2 AI-Based Methods for Adaptive Workflow Management

To enhance workflow flexibility by automatic workflow adaptations, two types of situations [17] that can occur during execution must be handled: *expected* and *unexpected* situations. Whereas expected situations that are known in advance can be handled by appropriate exception handling techniques (cf. [23]), unexpected situations require *ad-hoc changes* during runtime and, thus, more advanced techniques. The *ADEPT* framework by Dadam and Reichert [6] is one

[1] More information about the smart factory model and a video can be found at https://iot.uni-trier.de.

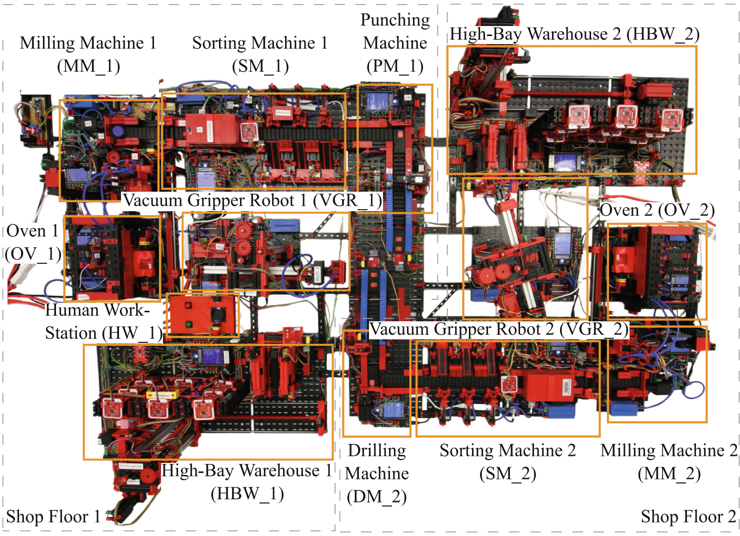


Fig. 1. The physical Fischertechnik smart factory model [16].

of the earliest and most prominent approaches for increasing process flexibility via ad-hoc changes of running processes. Thereby, monitoring for process failures and adapting a process is mainly performed manually. If an error occurs, the user selects a suitable ad-hoc change pattern such as inserting and deleting tasks or changing their execution order. To enable fully automatic ad-hoc changes, AI-based methods, which can be divided into *search-intensive* and *knowledge-intensive* [27], can be applied (e. g., [7,17–19,24,28,30]): Search-intensive techniques aim at finding a solution in the search space, which is usually implemented by using AI planning. In addition to the high search effort that is needed to generate solutions for large problems, a full and comprehensive planning domain description is required for problem-solving. However, such comprehensive planning descriptions are rather rare in real-world application domains. In addition, planning domain descriptions are sometimes sparse and only incomplete domain models with partial knowledge are available for planning [22,30]. Knowledge-intensive methods such as CBR use experiential knowledge, e. g., gained from employees working on the shop floor, to generate solutions. This significantly reduces the effort required to solve problems but also means that sufficient experiential knowledge must be available. This is especially difficult in dynamic IoT environments, where many unexpected problem situations can occur. Consequently, the resulting adapted workflows are sometimes not executable and require additional manual adjustments by users (e. g., [21,29]). Even though the individual methods provide good adaptation results, executability and correctness of adapted cyber-physical workflows are important as improperly configured and adapted workflows can cause considerable damage. To overcome the weaknesses of the methods in this regard, the combination of search-intensive and knowledge-intensive techniques promises valuable advantages [4,5,27,30].

2.3 Process-Oriented Case-Based Reasoning for Cyber-Physical Workflows

In our research, we apply *Process-Oriented Case-Based Reasoning (POCBR)* [3] as a knowledge-intensive method for advanced workflow management. POCBR integrates *Case-Based Reasoning (CBR)* [1] with process-aware information systems [23] such as workflow management systems or enterprise resource planning systems. A case in POCBR expresses procedural experiential knowledge gained in previous problem-solving situations, and the case base consists of best-practice workflows for reuse. To represent procedural experiential knowledge, we use a semantic workflow graph representation called *NEST graph* introduced in [3]: A NEST graph is a quadruple $G = (N, E, S, T)$ where N is a set of nodes and $E \subseteq N \times N$ represents the edges between nodes. Semantic descriptions S can be used for semantic annotations of individual nodes or edges. T specifies the type of the node or edge.

An exemplary cyber-physical manufacturing workflow represented in the *NEST* graph format is depicted in Fig. 2. It represents a sheet metal production process that can be physically executed in the smart factory model[2] (see Sect. 2.1). We use these kinds of manufacturing workflows because they are well suited for the factory layout used, and they are highly customized for a client, which also implies increased flexibility during execution. However, the used sheet metal workflows are placeholders for other arbitrary industrial processes. In the shown workflow, an unprocessed steel slab is unloaded from the high-bay warehouse and transported to the oven. In the oven, the steel slab is burned and rolled into a thick, middle-sized sheet metal. Afterwards, the processed sheet metal is transported to and stored in the high-bay warehouse.

In cyber-physical workflows, task nodes ($TN \subseteq N$) denote the production steps that are executed by actuators in the physical IoT environment. The semantic descriptions of task nodes can be used to further describe the properties of each activity, e. g., the concrete machine parameters. In the illustrated workflow, the *Burn* task is configured by the size and thickness parameters to produce the required sheet metal. In addition to this, the state of the task is captured in the semantic description (*COMPLETED*, *ACTIVE*, *EXECUTABLE*, *FAILED*, or *BLOCKED*). A task is blocked or fails during execution if the IoT resource needed to perform the activity is not functional, e. g., due to a defect. Data nodes ($DN \subseteq N$) can be consumed or produced by task nodes. Data-flow edges ($DE \subseteq E$) represent a consumption of a data node with an ingoing edge to the task node and a production of a data node by an outgoing one. A data node represents the state of the workpiece during this point in the manufacturing workflow. The semantic description of a data node features the properties of the workpiece, e. g., the position of the workpiece and the production properties such as size and thickness. Consequently, state changes of a single workpiece are represented by data nodes in the context of the overall workflow (similar to the usage of data nodes in [29]). Thus, a task node can only be executed after a

[2] More information about the execution of workflows in the Fischertechnik smart factory model can be found in [14,16,26].

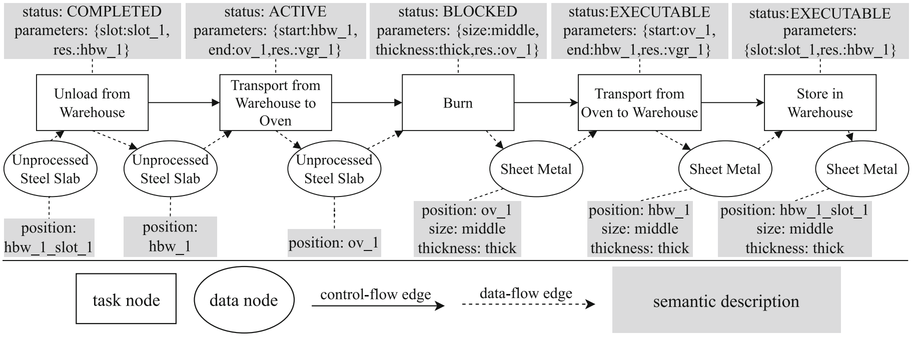


Fig. 2. Graph representation of a cyber-physical workflow.

previous task has produced the required properties. In this context, the control-flow can be derived from the data-flow by using control-flow edges ($CE \subseteq E$) that are specified between task nodes. One important aspect of cyber-physical workflows is the point of execution: As they are workflows executed by actuators in a physical IoT environment, the workflows need to be safely executable. For the smart factory used, it means that the machines need to be active and the activities represented by task nodes need to be configured correctly. Additionally, the state of the workpiece must be a valid input for the executing task. For example, the drilling machine can only drill holes into the workpiece if it is a sheet metal and not an unprocessed steel slab. In addition, the workpiece must be located at the drilling machine.

Application Scenarios for Cyber-Physical Workflows. One application scenario of POCBR for cyber-physical workflows is the retrieval of suitable workflows for execution based on a product specification and the currently available production capacities. For example, a client can specify the properties of the desired product in an order. These requirements are then used as a query to retrieve a suitable production workflow that can satisfy it. Afterwards, the workflow can be executed by a workflow management system. A further application scenario for using POCBR for cyber-physical workflows is the topic of workflow adaptation in the *Reuse* phase (see [15] for a generic architectural framework). Much of the current work in POCBR deals with retrieving similar workflow cases (e. g., [3,9,11]) and only few approaches (e. g., [19,21,28,29]) investigate the complex topic of automatic adaptation of retrieved workflows. Assume that the exemplary manufacturing workflow depicted in Fig. 2 is currently executed in the smart factory (see Sect. 2.1). During the execution, a failure due to a defect of the required oven occurs which leads to the task being blocked and not executable, i. e., the production cannot be continued. The goal of using POCBR in this scenario is to retrieve a case where a similar situation occurred previously, involving similar states of the machines and a similar not executable workflow. This retrieved case builds the basis for solving the current problem. The stored solution in the case is a *Change Plan* [17] that, when applied to the non-executable workflow, recovers the workflow to be further executed and, thus, to produce the final product. However, it is rather unlikely that a retrieved case solution solves the problem completely (see Sect. 2.2). For example, a similar case for the given problem is retrieved in which the oven in the second production line (*OV_2*) is used as alternative. However, the transports from the current position to the second oven and back are not contained. In these cases, the change plan should be adapted to better suit the current problem situation.

3 Adaptive Management of Cyber-Physical Workflows by CBR and Automated Planning

Based on the presented application scenarios of POCBR for cyber-physical workflows, we present our approach for combining experience-based adaptation by POCBR with automated planning for adaptive management of cyber-physical workflows in this section. The approach can be applied in the *Reuse* phase in the generic architectural framework presented in [15]. First, we examine how the compositional adaptation with workflow streams [21] can be used for cyber-physical workflows (see Sect. 3.1). Afterwards, we describe how experience-based adaptation can be enhanced by using AI planning (see Sect. 3.2).

3.1 Compositional Adaptation with Workflow Streams for Cyber-Physical Workflows

Compositional adaptation by using *Workflow Streams* [21] decomposes a workflow into smaller suitable sub-workflows, each of which produces a *partial workflow output* that is essential for achieving the overall workflow goal. More precisely, partial workflow outputs are intermediate steps in the workflow that are combined for achieving the final workflow goal such as the end product. Manufacturing workflows often consist of such intermediate produced subcomponents that are finally combined into an end product. Each partial output and, thus, each workflow stream represents a self-contained part of the workflow that can be replaced by another stream, e. g., with a different task sequence or parameterization, but producing the same output during adaptation. Learned streams from several previously experienced cases can be used for: 1) replacing streams in the case solution, 2) adding new streams to the case solution, or 3) deleting not needed streams in the case solution. The goal of adaptation is to modify the change plan, i. e., the solution in the retrieved case, in such a way that, on the one hand, the workflow goal of the current problem workflow is still reached and, on the other hand, only currently functional machines are used (see Sect. 2.3). In the following, we describe how the compositional adaptation method works in detail for cyber-physical workflows. Thereby, we distinguish between the automatic learning of workflow streams and applying them during adaptation in the *Reuse* phase.

Learning Workflow Streams. Each workflow stream produces a partial output of the workflow, i. e., a data node called *creator data node*. The task node that produces this special data node is called *creator task* and also marks the end of each workflow stream. In contrast to the native approach by Müller and Bergmann [21] in which creator tasks are defined by using the data-flow edges in the graph, this definition cannot be used for cyber-physical workflows. This is because data nodes in cyber-physical workflows are used to represent the state changes of the workpieces (see Sect. 2.3). Definition 1 specifies the modified definition for creating appropriate creator tasks in cyber-physical workflows:

Definition 1. *A task node t is a creator task ct, iff it adds at least one new property to the manufactured product. In the graph representation used, a new property can be determined by a larger number of attributes in the semantic description of the produced data node, i. e., the creator data node, compared to the consumed data node. The set of creator tasks CT is defined as follows:*

$$CT = \{t \in TN \mid \exists d_1, d_2 \in DN : ((d_1, t) \in DE \wedge (t, d_2) \in DE) \wedge |S(d_1)| < |S(d_2)|\}$$

Figure 3 depicts a manufacturing workflow with its marked creator tasks ($\odot$) and corresponding workflow streams. The tasks *Burn*, *Deburr*, and *Drill* are creator task nodes since they add a relevant property to the produced workpiece. Following Definition 1, the data nodes produced by the creator tasks have a larger number of attributes in their semantic descriptions than the previously consumed data nodes. For example, the *Burn* activity adds a concrete size and thickness of the produced sheet metal to the state of the workpiece. After specifying the tasks in the workflow that are creator tasks, the workflow can be partitioned into a set of workflow streams with the restriction that each task node $t \in TN$ is exclusively assigned to one workflow stream. The streams are constructed by applying the following rules [21,29]: A task node $t \in TN \setminus CT$ is assigned to a stream WS, 1) iff t is executed before the creator task $ct \in CT$ in the workflow, 2) iff t is not already contained in another stream, and 3) iff t is directly or transitively data-flow connected (cf. [21]) to the creator task $ct \in CT$. For example, the *Transport from Oven to Milling Machine* task is executed before the creator task *Deburr*, it is not already assigned to a stream, and there exists a data node that connects the *Transport from Oven to Milling Machine* task directly with

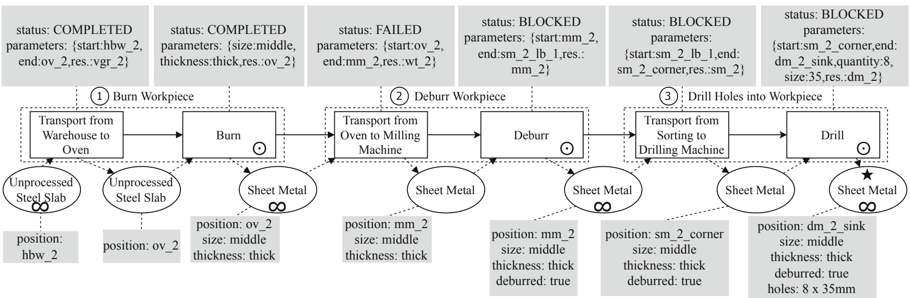
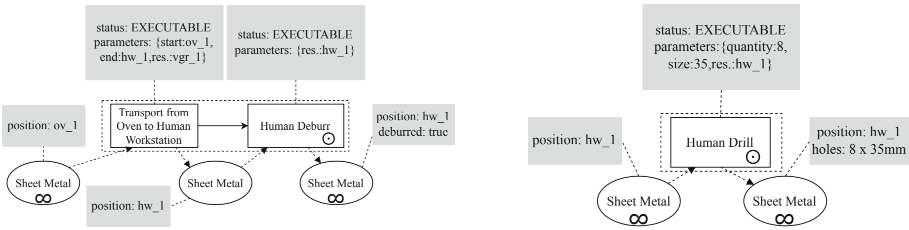


Fig. 3. Problem manufacturing workflow as part of the query.

the *Deburr* creator task (see Fig. 3). However, it could also be the case that several normal tasks are between this task node and the next creator task. In this case, the data-flow connectedness is transitively given.

Applying Workflow Streams. After the partitioning of the workflows in the case base and the construction of workflow streams, the learned streams stored in a stream repository can be applied in the *Reuse* phase. During adaptation, streams in the retrieved case solution are replaced or deleted, or new streams are added. The goal of the adaptation is to modify the case solution in such a way that it resolves the current problem by continuing workflow execution. More precisely, the adapted change plan leads to the fact that the currently not executable workflow can be continued while retaining the workflow goal (see Sect. 2.3). As an example, assume that the illustrated manufacturing workflow in Fig. 3 is currently executed in the smart factory. During production, a failure occurs at the *Transport from Oven to Milling Machine* task since the workstation transport machine in the second shop floor is not functional. After detecting this problem situation, the statuses of all tasks are updated. During this process, it is determined that the following tasks are blocked since the milling machine, the sorting machine, and the drilling machine in the second shop floor also cannot be used. A query consisting of the problem workflow graph and the described factory states is generated and a retrieval for a similar problem situation is performed.



(a) Change Plan As Case Solution From the Retrieved Case.

(b) Learned Workflow Stream.

Fig. 4. Compositional adaptation of cyber-physical workflows.

Figure 4a illustrates the change plan as case solution from the most similar retrieved case. Instead of using the currently defective milling machine, the change plan uses the human workstation, which is currently ready to deburr the workpiece. Although no defective machines are used anymore, inserting this change plan into the currently non-executable workflow and, thus, replacing streams 2 and 3 would not lead to the desired workflow goal marked with ★ in the problem workflow. For this reason, the stream repository is searched for further streams that are executable based on the current machine states and that generate the goal data node of the problem workflow. Figure 4b depicts a workflow stream learned from another case from the case base. Adding this stream to the retrieved change plan fully satisfies the adaptation goal. For this reason, it is added to the change plan and the modified change plan replaces the currently non-executable parts, i. e., streams 2 and 3, in the problem workflow. Adding a stream to the change plan requires maintaining the connection to the other nodes in the change plan. The *creator data node* is the final partial output of a stream. In addition to this output data node, the data node consumed by the first task in each stream is also marked with ∞ since it is required that this input must match with the output of the previous stream during insertion. These data nodes are called *anchor data nodes*. This means that the output of the change plan must be a valid input for the inserted stream. In the example, it is the case since the only condition for processing the workpiece during human drill is that the workpiece must be a sheet metal. Please note that for the validity check, only the workpiece attributes and not the *position* that represents the current physical location of the workpiece are considered. This is because otherwise it would prevent using suitable streams achieving the workflow goal. For example, if the learned stream from the repository (see Fig. 4b) uses the milling machine in the first shop floor for drilling holes, it could never be added to the change plan since the positions did not match. However, it would also achieve the goal data node that is required in the problem workflow.

To ensure that only streams are replaced or added in which no defective machines are used and by considering the goal data node of the problem workflow, a semantic similarity measure based on the measure presented in [3] is used. This similarity measure assesses the similarity between the goal data node of the problem workflow and the partial workflow output of the stream. In addition, the similarity between the current machine states and the machines needed to execute the stream is determined. If the similarity increases after a replacement or an addition of a stream, the modification is suitable for solving the current problem. At the end, we check if unnecessary streams can be deleted from the change plan while still achieving the workflow goal.

3.2 Integrating Automated Planning for Resolving Inconsistencies

Cyber-physical workflows are executed in physical IoT environments by actuators (see Sect. 2.1). For this reason, adapted workflows must be valid for execution since misconfigured workflows could lead to damage to machines or products or could lead to dangerous situations for humans. However, the compositional adaptation method cannot guarantee that adapted workflows are finally semantically correct and, thus, executable [21,29]. In addition, appropriate adaptation knowledge for solving the problem situation may not be available. One possibility to overcome these issues is to present the adapted workflows before execution to users. In this process, users can fix inconsistent parts (similar to [28]). Although this is a viable possibility, it requires a domain expert to perform these adjustments manually (cf. [6]), which can be complex and time-consuming. Another possibility is to use a further adaptation technique to fix inconsistencies in the adapted workflow that cannot be solved by compositional adaptation.

To overcome these problems, we propose to combine POCBR as a knowledge-intensive technique and automated planning as a search-intensive method. The main advantage of this combined approach is that the complete adaptation problem is divided into smaller and, thus, easier to solve sub-problems (*Divide and Conquer*), some of which are solved by the POCBR system and the compositional adaptation and some by AI planning. This combined approach also compensates incomplete planning domain models that are often common in real-world applications [22,30]. In the following, we present how this combination can be used for automatic workflow adaptations.

In the example in Sect. 3.1, a learned stream from the stream repository (see Fig. 4b) is added to the retrieved change plan (see Fig. 4a) to achieve the adaptation goal. However, the modified change plan requires that the workpiece is located at the oven in the first shop floor and not at the oven in the second shop floor. Thus, the adapted workflow is syntactically correct but not semantically and, therefore, not executable in the smart factory. Since the change plan does not include a transport, it must be added to the adapted workflow to ensure executability. In this context, we check after replacing the failed part of the problem workflow with the change plan whether the semantic correctness is given or whether inconsistencies exist. An inconsistency exists, for example, if the workpiece is not located at the correct position or if the workflow goal is not achieved with the change plan used. In these cases, we automatically determine the inconsistencies and generate a corresponding planning problem that consists of the current environmental conditions, i. e., the machine states, the initial state, and the goal state that should be reached by planning. Figure 5 depicts the adapted workflow with marked inconsistencies. To generate the planning problem, we use the output anchor data node of the first stream as the initial state and the input anchor data node of the second stream as the desired goal state of the planning problem. To maintain executability, the current state of the IoT environment, i. e., the machine states, are also defined in the planning problem so that only executable tasks can be used by the planner. Giving the generated planning problem to a state-of-the-art planner, it can easily solve the problem by adding actions/tasks that transport the workpiece from the oven in the first shop floor

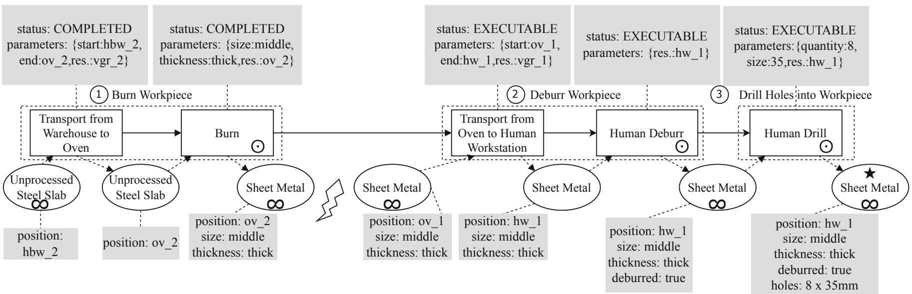


Fig. 5. Manufacturing workflow as result of compositional adaptation.

to the oven in the second shop floor. By adding these tasks, the adapted workflow is finally semantically correct and, thus, executable. The generative adaptation in our approach mainly needs knowledge about possible transportation routes since the knowledge about production steps such as *Burn*, *Drill*, or *Deburr* is already considered by the creator tasks in the workflow streams. In this way, the laborious and error-prone task of creating a complete planning domain as needed to generate appropriate solutions from scratch can be limited [4,17,22,24,27,30]. However, it could be possible that some situations cannot be solved since the required knowledge is not available in the POCBR system and in the (incomplete) planning domain. In these cases, the approach can support users with a pre-adapted workflow that builds the basis for performing final modifications to ensure executability manually (cf. [6,28]).

4 Experimental Evaluation

In this section, we present the experimental evaluation of the proposed approach conducted in our physical smart factory (see Sect. 2.1). For this purpose, we implemented the approach in the open-source POCBR framework ProCAKE[3] [2]. We use Fast Downward [8] with an A* search using the landmark-cut heuristic (*lmcut*) as a planner. Moreover, we create a full planning domain description since it is manageable for our rather small smart factory use case, and it allows us to use AI planning as gold standard in the experiment. The domain is written in PDDL 2.1 by using non-durative actions with action costs and negative preconditions. In total, 256 planning actions with several parameters, 27 relational predicates, and one functional numeric fluent, i. e., the total-cost function for the action costs, are contained in the domain[4]. To evaluate the approach, we measure the fulfillment of the performed adaptations by measuring the semantic similarity and the costs for executing the change plan in the factory. In addition, we check whether the adapted workflows are semantically correct and, thus, executable or not. In the experiment, we investigate the following hypotheses:

- H1 The compositional adaptation (CA) results in equal or better adapted workflows w. r. t. the described criteria than only using the retrieved case without adaptation (w/o).
- H2 Using the combined approach consisting of compositional and generative adaptation (CGA) leads to better results in terms of executability and semantic correctness than the pure compositional adaptation (CA).
- H3 The results generated by the combined approach (CGA) have similar total costs than solving the adaptation problem from scratch by using AI planning (GA).

[3] http://procake.uni-trier.de.

[4] The PDDL 2.1 domain and all planning problems are available at https://gitlab.rlp.net/iot-lab-uni-trier/edoc-2022-idams-workshop.

4.1 Experimental Setup

The Fischertechnik smart factory (see Sect. 2.1) represents two independent production lines and is used for the experiment. The goal of the experiment is to check whether it is possible to modify currently executed workflows after a failure so that they can still be executed, e. g., by using machines from the other production line that can perform the required activities (see application scenario in Sect. 2.3). To obtain real-world problems, a failure during the production is injected in a manufacturing resource of the smart factory. For this purpose, we use a *failure generation engine* that randomly selects machines and switches them to defective. We parameterized the engine to have at most two failures simultaneously, where each individual failure lasts at least 25 and at most 45 s. During the entire runtime of the workflows (approx. 6 to 9 min for each run), several machine resource failures are generated. If a failure occurs in the smart factory, the affected workflow is stopped and its current state is captured to use it for AI planning (GA) and as a query for the POCBR system (w/o, CA, and CGA) [15]. We utilize *four different production workflows* throughout the experiment that are executed in pairs of two. The used workflows deal with sheet metal production, as already introduced before (see Fig. 2 for an example). Thereby, each shop floor executes a single production workflow at a time: W 1.1 and W 1.2 are executed on the first shop floor and use 6 out of 7 different machines; both containing 12 tasks. W 2.1 and W 2.2 are executed on the second shop floor, use all 7 different machines, and contain 16 and 19 tasks. We apply a *Train and Test* scenario in which we first generate 20 random problems (10 for each pair of workflows) that are solved by using AI planning (GA) if failures occur during execution. Four generated problem situations could not be solved, since no other machines are available as alternative. However, these four problems and the 16 adapted workflows are stored in the case base as best-practice solutions. Based on these 16 correctly adapted workflows, we partitioned each workflow into its workflow streams and store them as adaptation knowledge (*Train* phase) in a stream repository. Finally, we generate 10 (5 for each pair of workflows) further problems that are used for evaluation (*Test* phase).

4.2 Experimental Results

The experiment focuses on the executability of adapted cyber-physical workflows for resolving failure situations during runtime. Table 1 depicts the experimental results of 10 random-generated failure situations conducted in the smart factory. We measure the semantic similarity between the solution and the goal to achieve (first row for each method) and the total-costs for executing the change plan (second row for each method). The total-costs reflect the execution time of the adapted workflow in seconds, i. e., the time required in the smart factory to execute the adapted part of the production process. A space in the table indicates that the measured values are equal to the method above it. In addition, we indicate whether the adapted workflows are semantically correct (✓) and, thus, executable or not (✗). As a gold standard, we use generative adaptation (GA) to

Table 1. Results of adapting cyber-physical workflows

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Avg.
w/o	1.0 ✓	0.95 ✗	1.0 ✓	0.98 ✗	1.0 ✗	0.86 ✗	0.97 ✗	1.0 ✓	1.0 ✗	0.96 ✗	0.972
	456	⁒	456	342	451	416	⁒	451	416	⁒	
CA				1.0 ✗		0.89 ✗					0.977
				530		419					
CGA					1.0 ✓	1.0 ✓			1.0 ✓		0.988
					484	817			493		
GA	456	223	456	⁒	416	382	376	451	451	376	

compare the POCBR approach against it. For queries *Q1*, *Q3*, and *Q8*, the failure situation can be solved immediately, since similar problems already occurred during training. In this context, complete solutions are contained in the case base and, thus, no further adaptations are required. For the generated failure situations *Q2*, *Q7*, and *Q10*, the retrieved cases are part of the 4 problems that could not be solved by AI planning in the *Train* phase before. For this reason, the POCBR system can only provide limited support to the user, since no more similar cases exist that can be used to solve the problem. However, the system could provide a complete solution for *Q10* if it had previously included *Q7* as a new experienced case, which shows the potential of the proposed self-learning approach. The workflows adapted in *Q5*, *Q6*, and *Q9* are executable but only after they were modified by the combined approach. The sole use of *CA* only leads to an increase in similarity for *Q4* and *Q6*. *Q4* is a special case since a rather similar case is retrieved and adapted, but no solution could be generated as all required machines for recovery are not functional, i. e., workflow executability cannot be achieved. All in all, the experiment shows high potential for using POCBR for adaptive cyber-physical workflows. Some problems can be solved even without adaptation since the solution has already been experienced during training. Whenever an adaptation has been performed, it always results in the same or a slightly higher similarity in contrast to pure retrieval (see **Avg.** column). Thus, we accept H1. In addition, the combined approach *CGA* leads to a higher number of executable workflows compared to *CA* (see *Q5*, *Q6*, and *Q9*). However, during *CGA* adaptation, the total costs sometimes increase strongly and are significantly higher than the costs for generating the solution from scratch. For this reason, we accept H2 but reject H3. To conclude, the use of the combined adaptation approach leads to adapted workflows that can be further executed by achieving their goals in 6 out of 10 cases in the experiment.

5 Conclusion and Future Work

We present an approach for using *Case-Based Reasoning (CBR)* to enhance the flexibility of cyber-physical workflows. In this context, we focus on combining compositional adaptation with generative AI planning for resolving failures during manufacturing. The proposed approach utilizes procedural experiential knowledge and, thus, limits the typically high knowledge acquisition and modeling effort to create comprehensive planning domains. In this context, complete planning domains are required for planning from scratch but often only incomplete domain models are available for planning in real application domains [22,30]. In contrast, planning in the proposed approach is used to solve smaller sub-problems, requiring general knowledge about transportation routes rather than specific knowledge as stored in cases. However, AI planning is needed in the approach to satisfy the challenging requirement of executability of automatic adapted cyber-physical workflows (see Sect. 2.1 and 2.2). Otherwise, improperly configured and adapted workflows can lead to considerable damage. In an experimental evaluation conducted in a physical smart factory, we showed that the approach can solve most problem situations and can adapt workflows suitably. This is performed either by reusing the case without modifications or by subsequent adaptation with the combined approach. Although the proposed approach has been implemented and validated in the domain of cyber-physical workflows, it can also be applied to other workflow domains (e. g., [21,29]). This is because compared to other domains, cyber-physical workflows have more specific requirements w. r. t. executability and correctness of the adapted workflows.

In the future, more than one case should be used for adaptation and the best adaptation result should be returned. Moreover, using conversational techniques (e. g., [28,29]) may promise better adaptation results while integrating domain experts. Transferring the proposed approach to real production lines with larger production workflows and more possible actions is also an interesting aspect for future work but also requires faster workflow retrieval methods (cf. [9]). In this context, it should be investigated how scalable the approaches are w. r. t. the workflow size and the domain complexity. We expect an improved solution quality since the case base contains more and larger workflows of already solved problems and a significantly better computation time than applying a solely generative approach (cf. [4,5,27]). Finally, it should be examined how much formalized knowledge is inevitably required in the planning domain, i. e., how incomplete the planning domain can be, for the approaches.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun. **7**(1), 39–59 (1994)
2. Bergmann, R., et al.: ProCAKE: a process-oriented case-based reasoning framework. In: 27th ICCBR Workshops, vol. 2567, pp. 156–161. CEUR-WS.org (2019)
3. Bergmann, R., Gil, Y.: Similarity assessment and efficient retrieval of semantic workflows. Inf. Syst. **40**, 115–127 (2014)
4. Bergmann, R., Muñoz-Avila, H., Veloso, M., Melis, E.: CBR applied to planning. In: Lenz, M., Burkhard, H.-D., Bartsch-Spörl, B., Wess, S. (eds.) Case-Based Reasoning Technology. LNCS (LNAI), vol. 1400, pp. 169–199. Springer, Heidelberg (1998). https://doi.org/10.1007/3-540-69351-3_7
5. Borrajo, D., Roubícková, A., Serina, I.: Progress in case-based planning. ACM Comput. Surv. **47**(2), 35:1–35:39 (2014)

6. Dadam, P., Reichert, M.: The ADEPT project: a decade of research and development for robust and flexible process support. Comp. Sci. Res. Dev. **23**(2), 81–97 (2009)
7. Gil, Y., et al.: Wings: intelligent workflow-based design of computational experiments. IEEE Intell. Syst. **26**(1), 62–72 (2011)
8. Helmert, M.: The fast downward planning system. J. Artif. Intell. Res. **26**, 191–246 (2006)
9. Hoffmann, M., Malburg, L., Bach, N., Bergmann, R.: GPU-based graph matching for accelerating similarity assessment in process-oriented case-based reasoning. In: Keane, M.T., Wiratunga, N. (eds.) Case-Based Reasoning Research and Development. ICCBR 2022. LNCS, vol. 13405, pp. 240–255. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-14923-8_16
10. Janiesch, C., et al.: The internet-of-things meets business process management. A manifesto. IEEE Syst. Man Cybern. Mag. **6**(4), 34–44 (2020)
11. Kendall-Morwick, J., Leake, D.: A study of two-phase retrieval for process-oriented case-based reasoning. In: Montani, S., Jain, L. (eds.) Successful Case-based Reasoning Applications-2. Studies in Computational Intelligence, vol. 494, pp. 7–27. Springer, Berlin, Heidelberg (2014). https://doi.org/10.1007/978-3-642-38736-4_2
12. Lasi, H., et al.: Industry 4.0. BISE **6**(4), 239–242 (2014)
13. 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)
14. Malburg, L., et al.: Semantic web services for AI-research with physical factory simulation models in industry 4.0. In: 1st IN4PL, pp. 32–43. ScitePress (2020)
15. Malburg, L., Bergmann, R.: Towards adaptive workflow management by case-based reasoning and automated planning. In: 30th ICCBR Workshops. CEUR-WS.org (2022). Accepted for Publication
16. Malburg, L., Seiger, R., Bergmann, R., Weber, B.: Using physical factory simulation models for business process management research. In: Del Río Ortega, A., Leopold, H., Santoro, F.M. (eds.) BPM 2020. LNBIP, vol. 397, pp. 95–107. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-66498-5_8
17. Marrella, A., Mecella, M., Sardiña, S.: Intelligent process adaptation in the SmartPM system. ACM Trans. Intell. Syst. Technol. **8**(2), 25:1–25:43 (2017)
18. Marrella, A., Mecella, M., Sardiña, S.: Supporting adaptiveness of cyber-physical processes through action-based formalisms. AI Commun. **31**(1), 47–74 (2018)
19. Minor, M., et al.: Case-based adaptation of workflows. Inf. Syst. **40**, 142–152 (2014)
20. Monostori, L.: Cyber-physical production systems: roots, expectations and R&D challenges. Procedia CIRP **17**, 9–13 (2014)
21. Müller, G., Bergmann, R.: Workflow streams: a means for compositional adaptation in process-oriented CBR. In: Lamontagne, L., Plaza, E. (eds.) ICCBR 2014. LNCS (LNAI), vol. 8765, pp. 315–329. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11209-1_23
22. Nguyen, T.A., Sreedharan, S., Kambhampati, S.: Robust planning with incomplete domain models. Artif. Intell. **245**, 134–161 (2017)
23. Reichert, M., Weber, B.: Enabling Flexibility in Process-Aware Information Systems - Challenges, Methods, Technologies. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30409-5
24. Rodríguez-Moreno, M.D., Borrajo, D., Cesta, A., Oddi, A.: Integrating planning and scheduling in workflow domains. Expert Syst. Appl. **33**(2), 389–406 (2007)
25. Seiger, R., et al.: Toward a framework for self-adaptive workflows in cyber-physical systems. Softw. Syst. Model. **18**(2), 1117–1134 (2019)

26. Seiger, R., et al.: Integrating process management and event processing in smart factories: a systems architecture and use cases. J. Manuf. Syst. **63**, 575–592 (2022)
27. Veloso, M.M. (ed.): Planning and Learning by Analogical Reasoning. LNCS, vol. 886. Springer, Heidelberg (1994). https://doi.org/10.1007/3-540-58811-6
28. Weber, B., Wild, W., Breu, R.: CBRFlow: enabling adaptive workflow management through conversational case-based reasoning. In: Funk, P., González Calero, P.A. (eds.) ECCBR 2004. LNCS (LNAI), vol. 3155, pp. 434–448. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28631-8_32
29. Zeyen, C., Malburg, L., Bergmann, R.: Adaptation of scientific workflows by means of process-oriented case-based reasoning. In: Bach, K., Marling, C. (eds.) ICCBR 2019. LNCS (LNAI), vol. 11680, pp. 388–403. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29249-2_26
30. Zhuo, H.H., Nguyen, T.A., Kambhampati, S.: Model-lite case-based planning. In: 27th AAAI. AAAI Press (2013)

Check for updates

An Optimization Ontology for Goal Modelling Frameworks

Krishna Gaur[1], Raghu Raj Sodani[1], Akshat Dobriyal[1], Anuj Mohan Pillai[1], Swasti Khurana[1], Novarun Deb[1](✉), Sajib Mistry[2], and Aditya K. Ghose[3]

[1] Indian Institute of Information Technology, Vadodara, India
{201951083,201951118,201951016,201952205,201852030, novarun_deb}@iiitvadodara.ac.in

[2] Curtin University, Perth, Australia
sajib.mistry@curtin.edu.au

[3] University of Wollongong, Wollongong, NSW, Australia
aditya@uow.edu.au

Abstract. Enterprises are driven by specific business objectives which should be delivered in the best possible manner within constrained business environments. In this paper, we propose a new ontology for goal modeling frameworks that enables enterprises to formally capture the business objectives and constraints associated with business goals as an optimization problem. We associate optimization problems with goals in such a manner that the solutions to the problems at the lower levels of the goal decomposition tree can be combined to obtain the solution for goals higher up in the tree. As a proof-of-concept, we have developed a web-based supply chain configuration management tool, deployed over the Ethereum blockchain, that relies on the Google OR-tools API to find optimal solutions. The interface allows the user to specify business objectives and constraints and then provides the best possible supply chain configuration that optimizes the business objectives.

Keywords: Optimization ontology · Business objectives · Business constraints · Goal models · Goal optimization

1 Introduction

The research community has been enabling requirement engineers with enriched goal ontologies for better analysis and management of requirements. The community has facilitated the advent of goal-oriented frameworks like i^* [17], *KAOS* [8], *NFR* [7] and *Tropos* [5]. These frameworks have further evolved to more enriched frameworks like *Secure Tropos* [15], *Formal Tropos* [9], and many more. Although different directions of enrichment have been explored in the past, a literature survey shows that there has been limited research with respect to optimization ontologies for goal models. Some of the initial works in this direction have been documented in Sect. 2.

The existing enrichment for requirement goal models is insufficient to represent optimization objectives and constraints associated with requirement goals.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 96–108, 2023.
https://doi.org/10.1007/978-3-031-26886-1_6

In this paper, we intend to make a conscious attempt in this direction. We propose to enrich goal model specifications using optimization ontologies. The main research challenge here is to correlate the optimization objectives and constraints associated with a parent goal to its child goals and ensure that they are consistent.

The optimization ontology proposed for goal models makes certain assumptions. We use a canonical representation of optimization problems where the objective function is always represented in *minimization* form and the constraints can be of *inequality* and *equality* types [4]. Inequality constraints are only represented using the $\leq$ operator. We adopt a bottom-up approach and define the optimization problem for the leaf level goals using the proposed optimization ontology. Next, we compose the optimization problem of child goals to derive the optimization problem of the parent goals. This approach ensures that the optimization problems associated with goal decompositions are consistent.

As a proof of concept, we have deployed the optimization ontology on a blockchain-based Supply Chain Management framework. Ethereum is used as the underlying blockchain technology. The use of smart contracts allows immutability and transparency in the movement of product batches across different stakeholders involved in the supply chain. The goals of the supply chain are derived from the demands placed on it while minimizing specific parameters like *cost* (or *time*). The proof-of-concept demonstrates how the optimization ontology can be adopted for the Covid-19 Vaccine Supply Chain.

The main contribution of this paper are as follows -

- Offering a richer ontology for formalising goal optimization problems.
- Offering a formal definition for correlating (or aligning) goal decompositions and goal optimizations.
- A proof-of-concept that aids the above correlation or alignment.

The rest of this paper is organized as follows. Section 2 presents the existing body of literature that exists in this direction. This is followed by Sect. 3 where we present our optimization ontology in detail. In Sect. 4, we demonstrate a deployment of the optimization ontology over a blockchain-based supply chain management web-app. Section 5 discusses some limitations of our proposed work. Section 6 concludes the paper.

2 Related Works

Although an optimization ontology for goal models is fairly new, there exists a body of literature that focus on norms and obligations and how they relate to business objectives. The encoding of norms as objective functions in [11] puts forward the notion of norm compliance as satisfying objective functions. We intend to illustrate a similar relation with goal decompositions. We build on the concept of refinement of an objective function in [10], bringing a new *objective function composition* as detailed in Sect. 3.2.

Optimization aligned with goal fulfillment is discussed in [6], however the objective function here is random. The focus of their model is to find the probability of success in achieving the goal and an expected level of shortfall.

Business contracts essentially have an ultimate goal. In the common setting of RuleML, [12] works toward forming business contracts from a human-oriented form into an executable representation, and [13] enables software agents to create, evaluate, negotiate, and execute business contracts with substantial automation and modularity. Both these papers correlate representation of business contracts while we focus on the broader goal modelling framework.

For a set of agents, [3] discusses the idea of obligations - task assigned to a group of agents based on their roles and responsibilities, and negotiations - distribution of set of obligations among the agents. Obligations and negotiation in [3] is akin to constraints in an objective function and finding an optimal solution, respectively. Moreover, the approach of [3] focuses on multi-agent systems whereas our approach brings in objective functions along with goal modelling.

Two recent works discuss fuzzy goal modelling and preference relations. [2] brings forward a model to balance the trade-off between fuzzy goals and preference relations. [1] discusses that preference relations are provided in terms of linguistic relationships, which brings a certain ambiguity. Although the discussions focus on fuzzy goal programming, it applies to requirements engineering and goal modelling as well. Further bringing it our ontology, optimization problems can be viewed as a kind of preference relations, where we focus on solutions that fit the preferences (or constraints in the case of optimization). We formally define the ontology without the ambiguity of linguistic relations.

The works [16] and [14] are more closely related to the idea of constraints and preferences in goal modelling. [14] addresses preferences in requirements, i.e. some requirements have more importance over others. However, it provides a plan which is a series of tasks to be completed. Our proposition focuses on a solution for the optimization problem in the form of a preferred plan among the possible plans and in agreement with the objective function. [16] proposes a constrained goal model with the main focus being the CGM-tool, whereas our efforts attempt on bringing forward a new ontology that would allow optimization to be a part of the goal modelling framework.

3 A Goal Optimization Ontology

In this section, we introduce the Goal Optimization Modeling Language (GOML) for augmenting goal semantics with optimization structures. The proposed ontology enables requirement engineers to define optimization problems with every goal in the goal decomposition tree of a system.

3.1 General Structure

We propose a *goal optimization problem* to contain the following:

1. **Goal:** A goal defines a state of affairs that an organization intends to achieve in the real world. (e.g., a vaccine manufacturer could have the goal of "*delivering vaccines for immunizing people against Covid-19*").

2. **Objective Function:** An objective function is a popularly used construct in operations research that allows us to define a preference relation on the feasible solutions of an optimization problem. These constructs are typically captured as *minimize f* (or, *maximize f*) where *'f'* is a function defined on the *decision variables 'x'*. For the purpose of this paper, we consider only the *minimize f* construct assuming that every maximization function can be converted to a corresponding minimization function using the negation operator. We represent the objective function defined over a goal as follows -

$$\min_{x} \quad f(x)$$

   Objective functions are used to capture business strategies that involve the measurement of performance metrics (or indicators) that are relevant for that particular enterprise. (e.g., the vaccine manufacturer could have the objective of "*minimizing the costs of producing vaccine vials*").
3. **Constraints:** The feasible solution space - on which an objective function establishes a preference relation - is defined by a set of constraints. Constraints are defined using the set (or subset) of *decision variables 'x'* that were used to define the objective function. The structure of a constraint consists of the following three components:
   - *Function* - defined over the set (or a subset) of the decision variables *'x'*.
   - *Threshold* - that bounds the value which the function may take.
   - *Operator* - which connects the threshold with the function. Operators could be any of $\{=, \leq, \geq\}$.

   For the purposes of this paper, we restrict the constraints to only two types based on the set of *Operators* - namely $\{=, \leq\}$. The underlying assumption is that $\geq$ constraints can be converted to $\leq$ constraints with the help of the negation operator. The two types of constraints are as follows -
   (a) ***Inequality Constraints:*** These are represented for the function $g(x)$ and the threshold C as follows -

$$g(x) \leq C$$
$$\text{or, } g(x) - C \leq 0$$

   An example of inequality constraint could be that "*every manufacturer can contribute a maximum of 40% of the total vaccine requirement*".
   (b) ***Equality Constraints:*** These are represented by the function $h(x)$ and the threshold C as follows -

$$h(x) = C$$
$$\text{or, } h(x) - C = 0$$

   An example of equality constraint could be that "*the total contribution from all manufacturers should be equal to 1.2 million doses*".

Combining the above concepts, we fix the structure of the optimization problem associated with a goal. The general structure of the optimization problem associated with a goal G_A would be as follows -

$$\begin{aligned} \min_{x} \quad & f^A(x) \\ \text{s.t.} \quad & g_i^A(x) \leq 0, \\ & h_j^A(x) = 0 \end{aligned}$$

where $f^A(x)$ is the objective function, $g_i^A(x) \leq 0$ are the 'n' inequality constraints ($1 \leq i \leq n$), and $h_j^A(x) = 0$ are the 'm' equality constraints ($1 \leq j \leq m$) defining the solution space of the optimization problem associated with goal G_A.

3.2 Goal Optimization Composition

Solving the optimization problem (structure shown above) involves assigning values to the set of *decision variables* from the feasible solution space, defined by the *constraints*, such that the *objective function* is minimized. This is equivalent to saying that the objective function specifies a *preference relation* on the feasible solution space, i.e., a solution which gives a lower value for the objective function is given higher preference. We use an ordered set $\{s_i, s_j\}$ to denote that the objective function prefers s_i over s_j, where s_i and s_j are feasible solutions to our optimization problem. In general, the solution at the i-th position is preferred over the solution at the j-th position of the ordered set, if $i < j$.

Next, we define the notion of *consistency* between two objective functions - $f_1(x)$ and $f_2(x)$ - defined over the same set of decision variables x. Let S_1 and S_2 denote the sets of feasible solutions for the two optimization problems, respectively. $\langle f_1, f_2 \rangle$ are said to be *consistent* if the preference relation between every pair of solutions s_i and s_j (that are present in both S_1 and S_2) is *preserved.* Preserving the preference relation between two solutions s_i and s_j implies that both S_1 and S_2 must satisfy the order $\{s_i, s_j\}$ (or $\{s_j, s_i\}$). It should be noted that this notion of consistency is dependent on the solution sets S_1 and S_2. We may have a different set of solutions S_1' and S_2', for the same pair of objective functions, which renders them *inconsistent* with respect to some s_i' and s_j'.

Definition 1. *Objective Function Composition.* Let $\{f_1, f_2, \cdots, f_n\}$ denote a set of objective functions defined over the same set of decision variables x. Also, let S_k represent the solution set for objective function f_k, respectively, for all k. A composition of the set $\{f_1, f_2, \cdots, f_n\}$ into an objective function $F = \min_k(w_k.f_k)$, with a solution set S_F, is a valid composition if and only if the following conditions hold:

(i) All weights w_k are normalised on the range $[-1, 1]$.
(ii) S_F is contained in the union of the solution sets of the individual objective functions, i.e., $S_F \subseteq \bigcup_{k=1}^{n} S_k$.
(iii) S_F is minimal, i.e., there does not exist any $S_F' \subset \bigcup_{k=1}^{n} S_k$ such that $S_F \subseteq S_F'$.

(iv) No preference relation $\{s_i, s_j\}$ in S_F is contradicted in the solution spaces S_k, for any k.
(v) The objective functions $\langle F, f_k \rangle$ are consistent, for all k. ■

The first condition allows us to compose an objective function from a set of given objective functions. In certain situations, if it is observed that $p = q + r$ or $p = q * r$, then we can conclude that *minimize* p implies *minimize* q and *minimize* r. This is equivalent to assigning positive weights in the composition function, i.e., $0 < w_q, w_r \leq 1$. If in certain other situations it is observed that $p = q - r$ or $p = q/r$, then we can conclude that *minimize* p implies *minimize* q and *maximize* r. For the purposes of this paper, we have defined objective functions as being only minimization problems and, hence, *maximize* r is represented as *minimize -(r)*. Thus, in the latter case, we assign positive weight to q $(0 < w_q \leq 1)$ and negative weight to $r(-1 \leq w_r < 0)$. The weight 0 is considered in a special case which we will discuss in the next section.

The second condition ensures that the preference relations captured by the composite objective function are preserved in the preference relations set up by the individual objective functions. The third condition ensures that the composition of objective functions does not introduce redundancy (i.e., no additional and unnecessary preference relations are added as a consequence of composition). The fourth condition states that the composition of objective functions does not set up preference relations which are in direct conflict with the preference relations set forth in any of the individual objective functions. The fifth condition requires the composite objective function to be *consistent* (as defined earlier) with each of the individual objective functions.

Definition 2. *Constraint Composition.* Let $\{f_1, f_2, \cdots, f_p\}$ denote the objective functions of the optimization problems that are to be composed. Let g_i^k and h_j^k represent the 'n' inequality constraints $(0 \leq i \leq n)$ and 'm' equality constraints $(0 \leq j \leq m)$ for the objective function f_k, respectively, for $1 \leq k \leq p$. The composition of the equality and inequality constraints, given by -

$$G = \bigcup_{k=1}^{p} \bigcup_{i=1}^{n} w_k.g_i^k \text{ and } H = \bigcup_{k=1}^{p} \bigcup_{j=1}^{m} w_k.h_j^k,$$

respectively, are valid compositions if and only if the following conditions hold:

(i) The weight w_k is the same defined (in *Definition*-1) for the corresponding objective functions f_k, for all k.
(ii) *Entailment:* $\{\bigcup_{k=1}^{p} g_i^k\} \vDash G$ and $\{\bigcup_{k=1}^{p} h_j^k\} \vDash H$.
(iii) *Consistency:* $\{\bigcup_{k=1}^{p} g_i^k\} \nvDash \perp$ and $\{\bigcup_{k=1}^{p} h_j^k\} \nvDash \perp$.
(iv) *Minimality:* $\nexists G' \subset G$ such that $G' \vDash G$ and $\nexists H' \subset H$ such that $H' \vDash H$. ■

The first condition ensures that the constraints are composed using the same weights as their corresponding objective function. Composition criteria (ii)–(iv) are identical to the goal refinement machinery proposed by Lamswerde in his KAOS framework [reference]. The interpretations of entailment, consistency, and minimality are also on the same lines. This approach helps us to align the newly proposed goal optimization ontology on top of the goal decomposition frameworks existing in the literature.

3.3 Goal Decomposition with Optimization Composition

In this section, we elaborate on how the goal optimization ontology, described in the previous sections, can be correlated with goal decompositions. We specifically look at two simple structures within goal decomposition trees - *AND*-decompositions and *OR*-decompositions - demonstrate how the proposed ontology can be adapted for these structures.

We assume that a goal G_A is decomposed to two child goals G_B and G_C. Let the optimization problem associated with the child goal G_B be as follows -

$$\begin{aligned} M_B: \min_x \quad & f_B(x) \\ \text{s.t.} \quad & g_i^B(x) \leq 0, \\ & h_j^B(x) = 0 \end{aligned}$$

Let the optimization problem associated with the child goal G_C be as follows -

$$\begin{aligned} M_C: \min_x \quad & f_C(x) \\ \text{s.t.} \quad & g_i^C(x) \leq 0, \\ & h_j^C(x) = 0 \end{aligned}$$

The optimization problems of the child goals are based on the ontology structure proposed in Sect. 3.1. According to the optimization composition defined in Sect. 3.2, the generic structure of the optimization problem at the parent goal would be as follows:

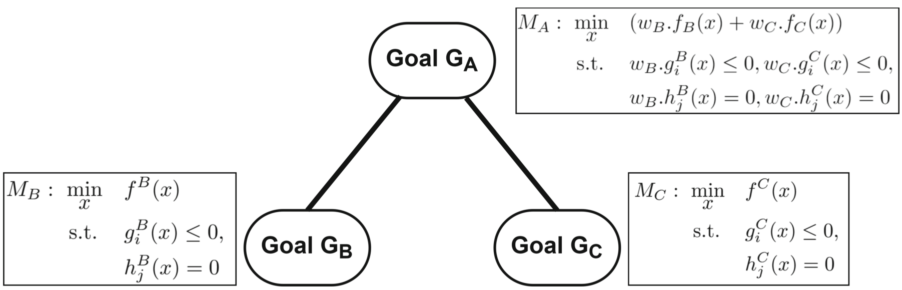


Fig. 1. The optimization composition problem for goal decompositions.

$$\begin{aligned} M_A: \min_x \quad & (w_B.f_B(x) + w_C.f_C(x)) \\ \text{s.t.} \quad & w_B.g_i^B(x) \leq 0, w_C.g_i^C(x) \leq 0, \\ & w_B.h_j^B(x) = 0, w_C.h_j^C(x) = 0 \end{aligned}$$

Figure 1 captures this idea.Based on the type of decomposition that exists between the parent goal G_A and the child goals G_B and G_C, we need to define the weights w_B, w_C. We look at the semantics associated with *AND*-decompositions and *OR*-decompositions and comment on these weights as follows -

OR Decomposition. An *OR*-decomposition semantically implies that goals G_B and G_C represent alternate strategies to achieve the state of affairs represented by goal G_A. The weights w_B, w_C are assigned depending on which strategy is chosen. We can list the possible weight assignments as follows:

(i) If strategy G_B is selected and the optimization problem M_B is a *minimization* problem, then $0 < w_B \leq 1$ and $w_C = 0$.
(ii) If strategy G_B is selected and the optimization problem M_B is a *maximization* problem, then $-1 \leq w_B < 0$ and $w_C = 0$.
(iii) If strategy G_C is selected, then $w_B = 0$ and w_C lies in (0, 1] or [−1, 0) depending on whether M_C is a minimization or maximization problem, respectively.

AND Decomposition. An *AND*-decomposition semantically implies that both goals G_B and G_C are required to achieve the state of affairs represented by G_A. The weights w_B, w_C are assigned as follows:

(i) If M_B, M_C are both *minimization* problems, then $0 < w_B, w_C \leq 1$.
(ii) If M_B, M_C are both *maximization* problems, then $-1 \leq w_B, w_C < 0$.
(iii) If either M_B or M_C is a maximization problem, then the corresponding weight is kept in the negative range [1, 0) and the other weight is kept in the positive range (0, 1], respectively.

It is to be noted that during the composition of optimization problems at the parent level, the weights are treated as input parameters. For example, if we want to minimize the cost of vaccination, then we would like to minimize the cost of manufacturing and distributing the vaccines, separately. However, while composing the optimization problem at the parent level, we may choose to give more precedence to the distribution cost rather than the manufacturing cost (or vice-versa). It is upto the supply chain owner (or manager) to choose the weight parameters.

4 Implementation of PoC

The theoretical concepts pertaining to the optimization ontology explained above have been partially deployed through a web-app. In this section, we describe a blockchain-based supply chain management tool that we developed and use it to build a case study for the optimization ontology. We would like to mention that the novelty of this paper is on the new optimization ontology and not the web- app developed. The deployment only serves as a proof-of-concept (PoC) to strengthen our claims made in the earlier sections.

4.1 PoC Technology Stack

The tool is build on a React based front-end and Django based back-end. We use Ethereum blockchain alongside the back-end for transparency and immutability purposes. The optimization problem is solved using the Google OR-Tools API. A technology viewpoint for our PoC has been created using the Archi Tool and presented in Fig. 2.

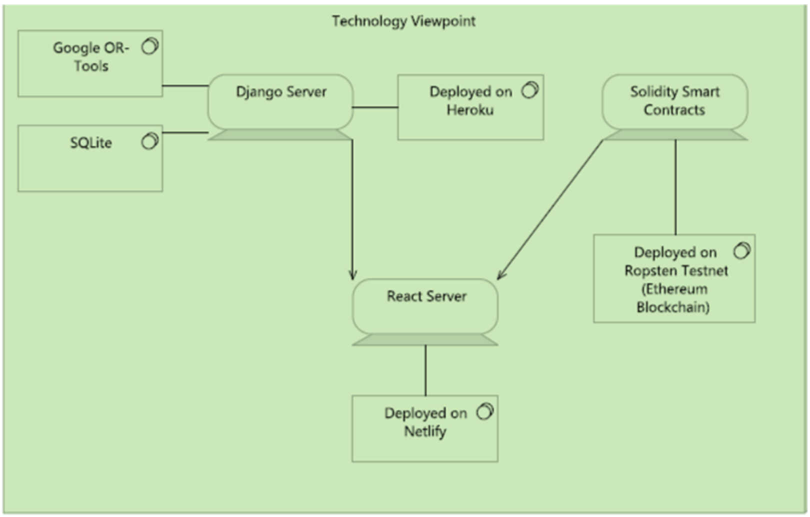


Fig. 2. Technology viewpoint of the deployed PoC.

4.2 Optimization Ontology Deployment

There are two types of users for the tool - the *supply chain managers* and the *participant stakeholders*. Supply chain managers own and handle the multiple participating entities. They also handle the product routes within the supply chain. On the other hand, participant stakeholders have a product (or service) that contributes to the supply chain. They can connect to other entities of the supply chain.

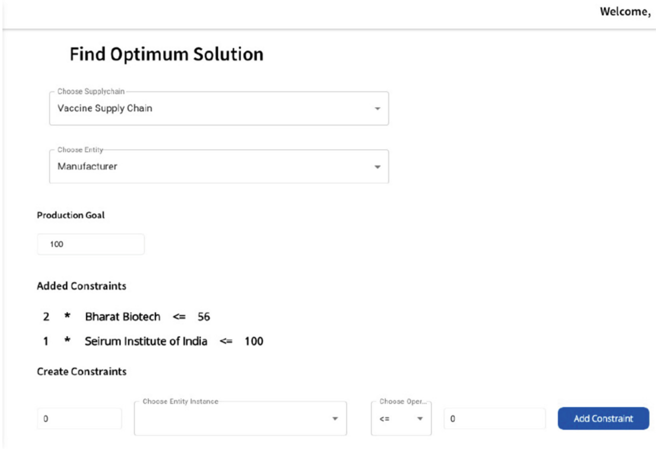


Fig. 3. Optimization and goal input interface.

We consider the Covid-19 Vaccine Supply Chain. It consists of multiple participating entities. These entities could be manufacturers, distributors, transporters, raw material providers, etc. A transportation company can then become a part of the supply chain by becoming an instance of the transporter entity. Connections between entities are specified by supply chain managers. An example would be connections between manufacturers and distributors, as manufacturers would require the vaccine vials to be distributed across different geographical locations.

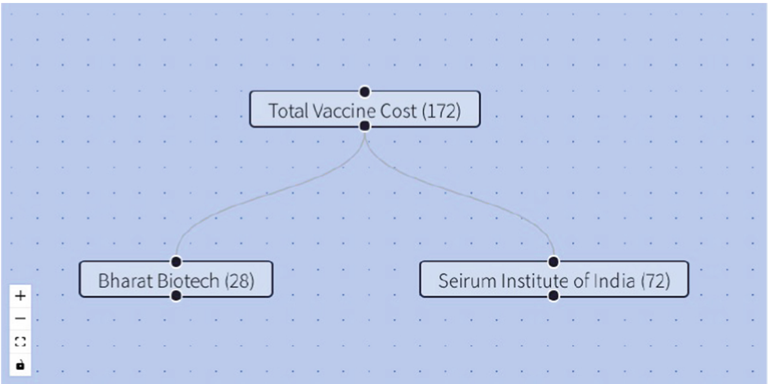


Fig. 4. Optimization solution graph.

In this deployment, we have demonstrated how the optimization problem associated with the goal of manufacturing vaccines can be captured and solved using the ontology proposed in Sect. 3. The objective function is *manufacturing a certain quantity of vaccines with minimum cost*. Each manufacturer provides the cost per unit of manufacturing vaccines while registering into the Vaccine Supply Chain. We take the vaccine quantity to be manufactured as input from the supply chain manager. Figure 3 shows the goal as manufacturing 100 units of vaccine. In *create constraint*, we capture the constraints on the amount of units one instance can produce. For example, in this case we are constraining Bharat Biotech to produce a maximum of 28 vaccine units in its contribution to the goal (captured using the constraint $2 \times BharatBiotech \leq 56$).

After fetching all the front-end data, we feed the objective function and the constraints to our backend which uses Google OR tools to solve our constrained optimization problem. The optimal solution is then pushed to the frontend and displayed on a graph. Figure 4 shows the solution to the optimization problem. The root node shows the total cost of manufacturing the 100 units of vaccine. The child nodes show the number of vaccine units produced by each of the manufacturers. The total vaccine cost (to minimize) is calculated automatically at the backend as a sum of the contribution from each manufacturer multiplied by the respective per unit cost of production. The details of each manufacturer and its respective per unit cost of vaccine production is fetched from our backend.

The chain of activities being performed in our web-app deployment is shown in Fig. 5.

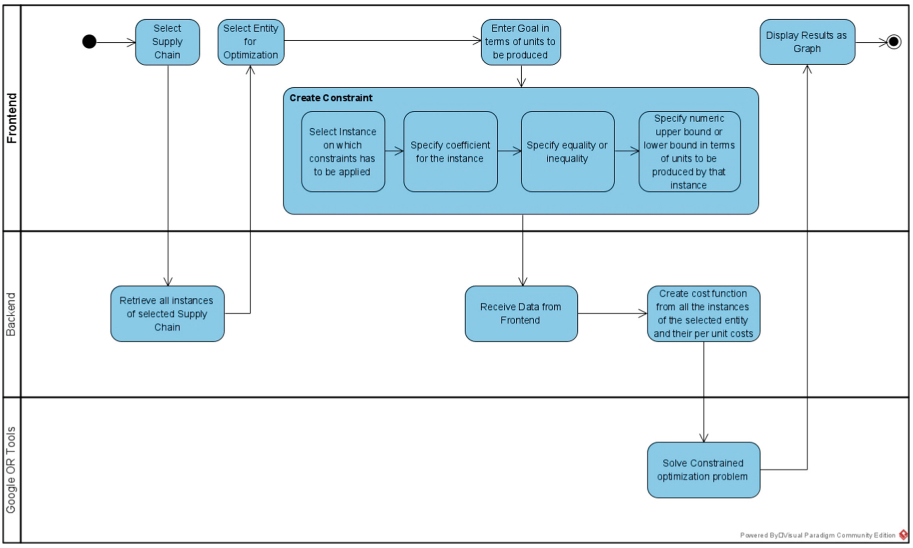


Fig. 5. Control flow for solving optimization problems associated with goals.

5 Limitations

The optimization ontology presented in Sect. 3 lacks a proper framework for integrating with existing goal model frameworks like jUCMNav, Tropos, and KAOS. It is open to the interpretation of the requirement engineers how they would like to use the ontology within their framework. As part of our future work, we intend to develop a plugin that would enable integration of the optimization ontology into existing frameworks. Another limitation is that we have not done a detailed empirical evaluation with other similar works in the literature. We aim to take up this task as soon as a plugin is ready for easy integration and deployment.

6 Conclusion

As part of this work, we propose an optimization ontology for goal model specifications that capture goal decompositions between parent and child goals. The optimization problem, associated with child goals, can be solved with the help of standard operations research tools (for example, the Google OR API framework). Combining the solutions at the child level is expected to generate the solution at the parent level. We have demonstrated how we can deploy this ontology over a blockchain-based supply chain management web app. The deployment is at a PoC level and needs further work to become a fully functional tool.

References

1. Abdul Ghaffar, A.R., Hasan, M., Ashraf, Z., Khan, M.F., et al.: Fuzzy goal programming with an imprecise intuitionistic fuzzy preference relations. Symmetry **12**(9), 1548 (2020)
2. Ahmadini, A.A.H., Ahmad, F.: A novel intuitionistic fuzzy preference relations for multiobjective goal programming problems. J. Intell. Fuzzy Syst. **40**(3), 4761–4777 (2021)
3. Boella, G., van der Torre, L.: Negotiating the distribution of obligations with sanctions among autonomous agents. In: ECAI, vol. 16, p. 13. Citeseer (2004)
4. Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2014). https://doi.org/10.1017/CBO9780511804441, https://web.stanford.edu/%7Eboyd/cvxbook/
5. Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., Mylopoulos, J.: Tropos: an agent-oriented software development methodology. Auton. Agents Multi Agent Syst. **8**(3), 203–236 (2004). https://doi.org/10.1023/B:AGNT.0000018806.20944.ef
6. Chen, W., Sim, M.: Goal-driven optimization. Oper. Res. **57**(2), 342–357 (2009)
7. Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J.: The NFR framework in action. In: Chung, L., Nixon, B.A., Yu, E., Mylopoulos, J. (eds.) Non-Functional Requirements in Software Engineering. International Series in Software Engineering, vol. 5, pp. 15–45. Springer, Boston (2000). https://doi.org/10.1007/978-1-4615-5269-7_2

8. Darimont, R., Delor, E., Massonet, P., van Lamsweerde, A.: GRAIL/KAOS: an environment for goal-driven requirements analysis, integration and layout. In: 3rd IEEE International Symposium on Requirements Engineering (RE 1997), 5–8 January 1997, Annapolis, MD, USA, p. 140. IEEE Computer Society (1997). https://doi.org/10.1109/ISRE.1997.566851
9. Fuxman, A., Pistore, M., Mylopoulos, J., Traverso, P.: Model checking early requirements specifications in tropos. In: Proceedings Fifth IEEE International Symposium on Requirements Engineering, pp. 174–181 (2001). https://doi.org/10.1109/ISRE.2001.948557
10. Ghose, A., Lê, L.-S., Morrison, E.: Correlating services with business objectives in the ServAlign framework. In: Franch, X., Soffer, P. (eds.) CAiSE 2013. LNBIP, vol. 148, pp. 162–167. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38490-5_14
11. Ghose, A., Savarimuthu, T.B.R.: Norms as objectives: revisiting compliance management in multi-agent systems. In: Aldewereld, H., Sichman, J.S. (eds.) COIN 2012. LNCS (LNAI), vol. 7756, pp. 105–122. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37756-3_7
12. Governatori, G.: Representing business contracts in RuleML. Int. J. Coop. Inf. Syst. **14**(02n03), 181–216 (2005)
13. Grosof, B.N., Poon, T.C.: SweetDeal: representing agent contracts with exceptions using XML rules, ontologies, and process descriptions. In: Proceedings of the 12th International Conference on World Wide Web, pp. 340–349 (2003)
14. Liaskos, S., McIlraith, S.A., Sohrabi, S., Mylopoulos, J.: Integrating preferences into goal models for requirements engineering. In: 2010 18th IEEE International Requirements Engineering Conference, pp. 135–144. IEEE (2010)
15. Mellado, D., Mouratidis, H., Fernández-Medina, E.: Secure tropos framework for software product lines requirements engineering. Comput. Stand. Interfaces **36**(4), 711–722 (2014). https://doi.org/10.1016/j.csi.2013.12.006, https://www.sciencedirect.com/science/article/pii/S0920548913001803. Security in Information Systems: Advances and new Challenges
16. Nguyen, C.M., Sebastiani, R., Giorgini, P., Mylopoulos, J.: Multi-objective reasoning with constrained goal models. Requirements Eng. **23**(2), 189–225 (2018)
17. Yu, E.S.K., Mylopoulos, J.: From E-R to “A-R” - modelling strategic actor relationships for business process reengineering. Int. J. Coop. Inf. Syst. **4**(2–3), 125–144 (1995). https://doi.org/10.1142/S0218843095000056

SoEA4EE 2022

14th Workshop on Service-Oriented Enterprise Architecture for Enterprise Engineering (SOEA4EE 2022)

Preface

Since its foundation in 2009, the SoEA4EE workshop complements well-established topics of the EDOC conferences such as service-oriented architectures and enterprise service architectures by addressing the coupling of business processes and services and the alignment of business and IT. The SoEA4EE workshop also includes topics such as Business Process Management, Enterprise Service Architectures, Analytics, Big Data, Networked Enterprise Solutions, and their connections.

Smart companies define how they (will) do business (using an operating model) and design processes and infrastructure critical to their current and future operations using enterprise architecture (EA). Enterprise Engineering (EE) is the application of engineering principles to the design of enterprise architectures. It enables deriving the EA from the enterprise goals and strategy and aligning it with the enterprise resources. EA is used to map the enterprise goal and strategy to the enterprise's resources (actors, assets, IT supports) and to support the evolution of this mapping. There are different paradigms for creating enterprise architectures. In the SoEA4EE series, the one which is considered as the most promising is to encapsulate the functionalities of IT resources as services. By this means, it is possible to clearly describe the contributions of IT both in terms of functionality and quality and to define a service-oriented enterprise architecture (SoEA).

Two special journal issues have been published in the International Journal of Information Systems in the Service Sector, the first with invited papers from SoEA4EE 2010, 2011, and 2012, the second with invited papers from SoEA4EE 2013, 2014 and 2015.

SoEA4EE 2022 workshop was a full-day workshop in conjunction with EDOC 2022. All submissions were peer-reviewed by at least three members of SoEA4EE 2022 international program committee. We wish to thank all authors for having shared their work with us, as well as the members of the SoEA4EE 2022 Program committee and the organizers of EDOC 2022 for their help with the organization of the workshop.

October 2022

Selmin Nurcan
Rainer Schmidt

Organization

Workshop Chairs

Selmin Nurcan	University Paris 1 Panthéon Sorbonne, France
Rainer Schmidt	Munich University of Applied Sciences, Germany

Program Committee

Alfred Zimmermann	Hochschule Reutlingen, Germany
Colin Atkinson	University of Mannheim, Germany
Eric Dubois	Luxembourg Institute of Science and Technology, Luxembourg
Henderik A. Proper	Luxembourg Institute of Science and Technology & University of Luxembourg, Luxembourg
François Charoy	Université de Lorraine - LORIA, France
Jānis Grabis	Riga Technical University, Latvia
Jolita Ralyté	University of Geneva, Switzerland
Julio Cesar Nardi	Federal Institute of Espírito Santo, Brazil
Khalid Benali	LORIA, Nancy, France
Kurt Sandkuhl	University of Rostock, Germany
Rainer Schmidt	Munich University of Applied Sciences, Germany
Ron Kenett	KPA Ltd., Israel
Said Assar	Institut Mines-Telecom Business School, France
Selmin Nurcan	Université Paris 1 Panthéon-Sorbonne, France
Sung-Kook Han	Wonkwang University, South Korea
Ulrike Steffens	Hamburg University of Applied Sciences, Germany

Visual Description of Digital IT Consulting Services Using DITCOS-DN: Proposal and Evaluation of a Graphical Editor

Meikel Bode(✉), Maya Daneva, and Marten J. van Sinderen

Department of Semantics, Cybersecurity and Services, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
{m.bode,m.daneva,m.j.vansinderen}@utwente.nl
https://www.utwente.nl/en/eemcs/scs

Abstract. The digital transformation of the IT consulting domain recently gained momentum due to the Covid-19 pandemic. However, the range of IT consulting services that are fully digital is still very limited. Plus, there are no standardized and established methods for describing digital IT consulting services, nor there is any suitable tooling for digital IT consulting service provisioning. The present work aims to reduce this gap by contributing to establishing a well-defined approach to formally describing digital IT consulting services that could possibly be a candidate for standardization. Building upon (i) the ontology DITCOS-O, which provides the semantic basis for our approach, and (ii) the YAML-based description notation DITCOS-DN, which we leverage to describe digital IT consulting service models, we propose a graphical, web-based editor (called DITCOS-ModEd) to simplify service model maintenance. Following a design science based research process, we developed a prototype and empirically evaluated its applicability with the help of IT consultants. This first evaluation allowed us to identify some limitations and to plan specific improvements, both to the underlying artifacts DITCOS-O and DITCOS-DN, as well as to DITCOS-ModEd itself.

Keywords: Digital transformation · IT consulting · Service ontology · Service description · Virtualization · Consulting platform · Graphical editor · DITCOS-O · DITCOS-DN · DITCOS-ModEd · Design science · YAML

1 Introduction

During the Covid-19 pandemic the digital transformation (DT) of the IT consulting (ITC) domain gained momentum. The core of the ITC business is to advise clients on how to digitally transform their respective processes and businesses models. Gartner defines ITC services to be "...advisory services that help clients assess different technology strategies and, in doing so, align their technology strategies with their business or process strategies. These services support

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 113–128, 2023.
https://doi.org/10.1007/978-3-031-26886-1_7

customers' IT initiatives by providing strategic, architectural, operational and implementation planning" [7, term: 'it consulting'].

In this work we use the terms digitization, digitalization, and digital transformations in line with the definitions of Gartner. Digitization is the conversion of physical resources to digital representations; digitalization is the use of digital resources within IT systems; and digital transformation refers to digitalization in the context of business processes in ITC [7, terms: 'digitization', 'digitalization', 'digital transformation'].

However, digital ITC services and suitable tools for the digital provision of these services are still rare. We use the term *digital ITC service* as defined in [3]: "Digital IT consulting services are technology-based consulting services represented by standardized, modularized, re-combinable, reusable, and customizable service assets that carry specific service commitments and are provided either in an automated, hybrid, or manual mode by human and/or technical agents or in a self-service manner and are instantiated, delivered, monitored, and orchestrated by digital consulting platforms." [3].

In the literature of digital ITC, to the best of our knowledge there are hereonly a small number of examples, such as the *eConsulting Store* provided by Werth et al. (2016) and the customer-tailored web-based self-service project assessment solution provided by Nissen et al. (2019) [10,13]. These artifacts have in common that they are tailored to a special problem context and do not aim on solving issues such as semantic standardization, modularization, or reusability as we identified to be necessities regarding the digital transformation of ITC [3].

To address these gaps in our recent research, we have lately contributed two artifacts [4]: (1) DITCOS-O, an ontology covering relevant concepts of digital ITC, and (2) DITCOS-DN, a formal description notation based on DITCOS-O providing a YAML syntax for service model definition. DITCOS-O constitutes an ontological sound basis to build upon. DITCOS-DN consumes these concepts and acts as a kind of 'programming language' to define ITC service models that are understandable for humans as well as interpretable by technical systems, such as digital ITC platforms.

With the current paper we build upon the published artifacts [4] DITCOS-O and DITCOS-DN and contribute a new artifact, namely the editor DITCOS-ModEd that supports end users to easily describe DITCOS-DN-based service models web-based and graphically. Our editor was empirically evaluated for suitability and usefulness by means of a two-step evaluation study, including perception-based research and experiments.

The remaining paper is structured as follows. Section 2 presents our research goals. Section 3 is on background and related work. Section 4 is on our research process. Section 5 describes our application of design science in order to create and evaluate our prototype editor. Section 6 discusses our results and Sect. 7 concludes.

2 Research Goals

As stated in the introduction, this article leverages our previously published results [4], namely the ontology DITCOS-O and the description notation DITCOS-DN. Both were empirically evaluated [4] with practitioners in a study that investigated the understandability of the YAML-based DITCOS-DN notation. Our evaluation results indicated that manual maintenance of DITCOS-O models using DITCOS-DN is well-supported by existing tools, since we assured automatic syntax checking, code completion and code formatting by means of integrated development environments (IDEs). However, during our evaluation it became apparent that even if good tool support for the textual creation of DITCOS-O service models described in DITCOS-DN exists, this might not be the preferred method from the practitioners' point of view. In fact, it became clear to us that practitioners would much more prefer working with a visual service description editor over using a tool for textual descriptions. Moreover, by means of a graphical editor, we believe we will create an important scientific instrument, which will be extremely useful and necessary for us in the context of our short-term future research activities especially in collaboration with practitioners. With this in mind, we set out the following goals with the present work:

1. To provide a graphical editor for the maintenance of DITCOS-O service models to be described in DITCOS-DN.
2. To apply experimentally the graphical editor in a real-world context with the participation of practitioners and to evaluate the suitability and usefulness of the solution in order to collect feedback for improvement of the underlying artifacts DITCOS-O and DITCOS-DN, as well as the graphical editor DITCOS-ModEd.

3 Background and Related Work

There are two streams of related work that are relevant for this paper: publications on service description approaches and on graphical service modelling. Regarding service description, the existing approaches are of two types: ontology-based and textual approaches using frameworks like IT Infrastructure Library (ITIL). As this research adopts an ontology-based approach, in what follows we provide related work concerning approaches of this type. One example approach is LinkedUSDL [5]. LinkedUSDL (Universal Service Description Language) was designed as an *upper ontology* with the aim to cover all relevant service contexts and concepts [5]. It uses the Resource Description Framework (RDF) to describe concepts based on *triples*, using the structure *subject* → *predicate* → *object*. LinkedUSDL follows the linked data principles [6] in the sense that it requires each element of the triple to be an Uniform Resource Identifier (URI), which, in the optimal case, points to additional content related to the respective RDF element. LinkedUSDL was organized into different sub-ontologies that complement and build upon each other.

Next, regarding graphical service modeling, an example is OBELIX [1], an ontology-based approach that helps to describe real-world services and service bundles based on the flow of resources and generated value (value webs). OBELIX defines *service elements* that have input and output *interfaces*, each of which supports an arbitrary number of *ports*. Output ports connect to input ports, while multiple service elements could form *service bundles*.

While searching for related work for the purpose of this research, we found that even though numerous approaches exist in the literature on technical service or real-world service description, we were not able to find approaches dedicated to the description of digitalized real-world ITC services provided through digital consulting platforms. We noticed that either the approaches aim on being as generic as possible (LinkedUSDL) and require complex RDF based descriptions or mainly focus on value flow, such as OBELIX.

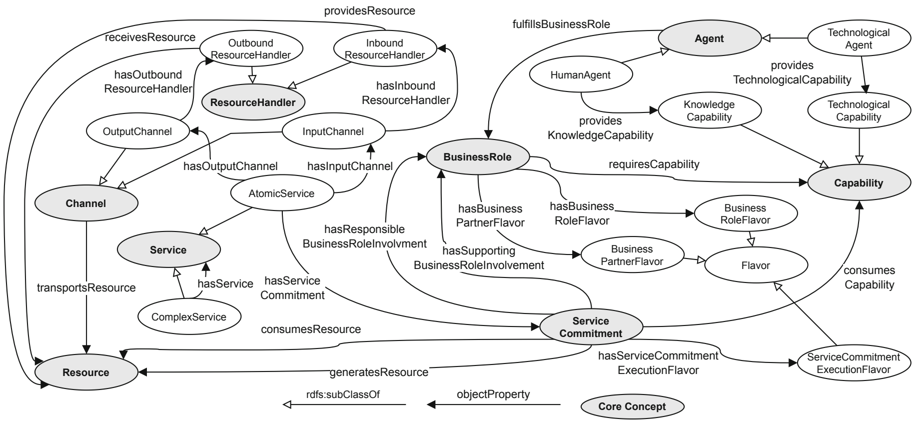


Fig. 1. The DITCOS-O and DITCOS-DN core concepts

Unlike the existing ontology-based approaches, with DITCOS-O and the corresponding YAML-based description notation DITCOS-DN, we want to provide a notation that aims on covering relevant core concepts of the ITC domain (see Fig. 1 how the core concepts relate), being easy to learn, to understand, and to use by practitioners (IT consultants), and is at the same time interpretable by systems (see Listing 1 for an simplified, reduced and invalid example. For a full example see [4]). It is meant to describe ITC services that are digitally provided through digital consulting platforms. The provisioning process might be either (1) only aided by helping agents, e.g., IT consultants, to conduct certain activities, (2) orchestrate incorporated agents, such as consultants, clients, or technical systems, or (3) fully automate the service provisioning process. To the best of our knowledge, none of the approaches in the literature covers this.

```
ditcosModel:
  metadata: [id, name, version, description, author, entryService]
  businessRoles: [businesRoleA, businessRoleB, ...]
  resources: [resourceA, resourceB, ...]
  services: [atomicServiceA, ..., complexServiceA, ...]
```

Listing 1. Simplified Example of a DITCOS-DN Service Model

4 Our Research Process

This section explains and motivates the process used to develop the DITCOS-ModEd artifact. We note that this work is part of a larger research project and builds upon already contributed artifacts [4]. Our research process adopted the design science (DS) research methodology of [11]. We chose it, because DS is recommended to research contexts such as ours where solutions (called ‘artifacts’) are designed to counter industry-relevant problems and issues [11]. Following Peffers et al. [11], our research process consists of the stages: (1) problem identification, (2) definition of objectives, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication. In the following section we report on our execution of these stages, except stage six which is covered by the overall paper implicitly.

5 Designing the DITCOS-ModEd Graphical Editor

5.1 Problem Identification

In our recently published work [4], we already created and evaluated DITCOS-O service models using the YAML-based notation DITCOS-DN. These kind of service models are stored in a *service repository* (SR) to be later consumed by a digital ITC platform [2]. A central part of the platform should be an editor that supports textual (YAML-based) as well as graphical creation of DITCOS-O service models described using DITCOS-DN (see Fig. 2). We emphasize that the SR and the digital ITC platform are subject of our immediate future research.

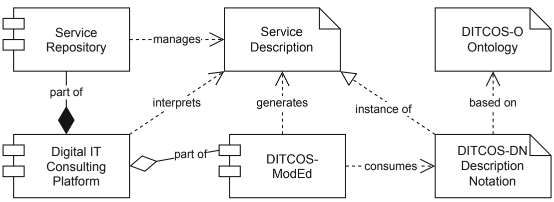


Fig. 2. Interplay of our research artifacts

Even though graphical service model creation is not technically required, as all descriptive power is provided by DITCOS-DN, the practical application of the textual modeling approach remains more complex and technical as it needs to be. Hence, we decided to provide the possibility to graphically create DITCOS-DN-based service models with the help of an appropriate, web-based editor. By providing this graphical editor, we also aim to increase its possible adoption by practitioners who may not be familiar with programming languages and have more of a business focus. It is worthwhile noting that our decision to provide a graphical and web-based editor component is also consistent with the architectural requirements identified in our previous exploratory study focused on practitioners' requirements elicitation [2]. We refer to a subset of these requirements in column 'Source' in Table 1.

5.2 Definition of Design Objectives

Our two goals for this research were stated in Sect. 2. Linked to them, we defined the following objectives in form of functional and non-functional requirements [9]. These requirements listed in Table 1 are then later to be used during the evaluation of our newly proposed editor. We assigned to each requirement in Table 1 an identifier (ID), a name, a description, a type (either functional or non-functional), and the source we collected the requirement from.

Table 1. Functional and non-functional requirements for DITCOS-ModEd

ID	Name	Description	Type	Source
R1	Web-based	Build the editor using web-based technologies	NF	[2]
R2	Component-based	Design the editor to be built from reusable components	NF	
R3	External Lookups	Enhance JSON-Schema based value lookups to support web-services	F	[4]
R4	CRUD Support	Support creation, reading, updating, and deleting service models	F	[2]
R5	Graphical View	Support the graphical rendering of service models and component interconnection	F	
R6	YAML View	Support the YAML code inspection	F	
R7	DITCOS-O Coverage	Support concepts of the DITCOS-O ontology expressible in DITCOS-DN description notation, such as Atomic Service, Complex Service, Business Role, Capability, Resource, Service Commitment	F	

5.3 Design and Development

This section briefly describes the architecture and the design we chose for our proposed editor as well as the tools we used for development. Below we present the elements included in our design and the choices we made in regard to each element.

Editor Architecture. Based on the software design pattern model-view--controller (MVC) [12] we structure the editor artifact into three layers: (1) presentation, (2) business logic, and (3) data access. The data access layer consumes an externally provided data services layer. These layers will be discussed in the next sections in more detail. Overall, the graphical editor will be implemented as a standalone prototype application that is not yet embedded into a larger architecture as we depicted in [2]. The reason for this is, that we mainly want to test the graphical modeling approach of DITCOS-DN and the generation of valid YAML-based service descriptions. The integration of DITCOS-ModEd into the overall architecture will therefore be covered in our future research.

Presentation Layer. Important requirements on the graphical editor are web-based and component-based design. Today, web-applications follow either the backend-rendering, the frontend-rendering, or a hybrid approach. For our implementation we decided to go for the frontend-rendering approach using the REACT framework, because it is a recent framework and the main author already has experience using it. We checked other options, such as Vue or AngularJS but decided against it as this would have meant to learn a new framework.

The loading and execution of a REACT application is initiated after the page that references the REACT JavaScript-based logic was transferred to the browser. The JavaScript logic then renders the entire web-application document object model (DOM) dynamically and manages subsequent page updates and data acquisition by executing HTTP requests on demand. REACT aims on the design and implementation of simple and independent components which can be combined to build more complex applications. The REACT components render out HTML5 compliant code by directly manipulating the DOM. For styling the HTML elements cascading style sheets (CSS) are being used. We used the common and widely-used web component theme 'Material Design' for our REACT components. We decomposed the prototype into 17 independent REACT components that work together. All components are described in Table 2.

Business Logic Layer. Next to the presentation layer and the associated presentation logic, REACT applications contain their required business logic. This is due to the fact, that the web-application is self-contained. All application logic gets loaded during the initialization phase after the stub of the application has been transferred from the web server to the browser. Upon initialization, the REACT App component (see Table 2) gets executed, which itself triggers initialization and execution of all other components in the REACT component hierarchy. All component executions load their required data either individually or shared from remote sources, which are usually provided by the server that also serves the web-application stub and its logic. Additional business logic might be externalized to the server side and exposed by web-based API to consuming web-applications, such as duplicate checks or other kind of data validations before being persisted to a store. This kind of APIs often provides common or shared functionality that can be used by different (web) applications.

Table 2. REACT components of DITCOS-ModEd

Name	Description
App	Represents the central component of an REACT application. It embeds the *ModelManager*, *ModelView*, and the *ModelEditor* top-level components
ModelManager	A top-level component that represents the interface to existing DITCOS-O service models and supports their creation, deletion, and loading
ModelViews	A top-level component that embeds the components *GraphicalView* and *YAMLEditor*
ModelEditor	A top-level component that embeds the components *ModelMedatadata*, *BusinessRoles*, *Resources*, and *Services*. It provides the functionality to persist the currently loaded model
GraphicalView	Provides the graphical representation of the defined DITCOS-O service model currently loaded in form of a directed graph consisting of nodes and edges. It is based on the external REACT component *REACTFlow*. It supports interactions such as zooming in and out, shifting the graphicalized graph around, the selection of nodes and edges. It also provides a *MiniMap* component that represents a minimized version of the whole graph for quick navigation
YAMLEditor	Provides the textual representation of the currently loaded DITCOS-O service model. It is based on the external REACT code editor component *Monaco*. It supports line numbering, folding, formatting, coloring, and indent
ModelMedatadata	Represents input elements for model metadata, such as id, name, description, author, and version
BusinessRoles	Represents all business roles defined by the model and supports their creation, deletion, and modification by linking to *BusinessRoleDetailsDialog* component
BusinessRole-DetailsDialog	Provides input elements to model a business role and link to the capabilities that constitutes it
Resources	Represents all resources defined by the model and supports their creation, deletion, and modification by linking to *ResourceDetailsDialog* component
ResourceDetails-Dialog	Provides input elements to model a resource generated, consumed, or updated by service commitments of defined atomic services
Services	Represents all atomic and complex services defined by the model and supports their creation, deletion, and modification by linking to *AtomicServiceDetailsDialog* and *ComplexServiceDetailsDialog* components
AtomicService-DetailsDialog	Provides input elements to model an atomic service and its service commitments by linking to *ServiceCommitmentDetailsDialog* component
ServiceCommit-mentDetailsDialog	Provides input elements to model a service commitment and to refer to resources by linking to *InvolvedResourcesDetailsDialog* component
InvolvedResource-DetailsDialog	Provides input elements to refer to a certain defined resource and its type of involvement in the related service commitment
ComplexService-DetailsDialog	Provides input elements to model a complex service and to link to its constituting atomic and complex services by linking to *InvolvedServiceDetailsDialog* component
InvolvedService-DetailsDialog	Provides input elements to refer to a certain defined atomic or complex service and to define its dependencies on other services

Data Access Layer. The data access of REACT applications usually is realized by consuming REST-based APIs exposed by the server side. Other HTTP-based communication protocols, such das GraphQL or OData are also common. The data exchange format can vary, but nowadays mostly JSON formatted messages are interchanged with incorporating the data service layer. For our editor we chose a combination of REST-API and JSON as message format.

Data Service Layer. The data services used by our prototype are provided by a JavaScript based server component. All data entities used are represented by corresponding REST-API endpoints which support the CRUD pattern by incorporating different HTTP methods, such as GET, PUT, PATCH, and DELETE. For our prototype we use a simple server component without applying extended validations. This is sufficient, as building such an API would be delegated to a dedicated service repository which would be part of a larger digital ITC platform architecture as proposed in [2].

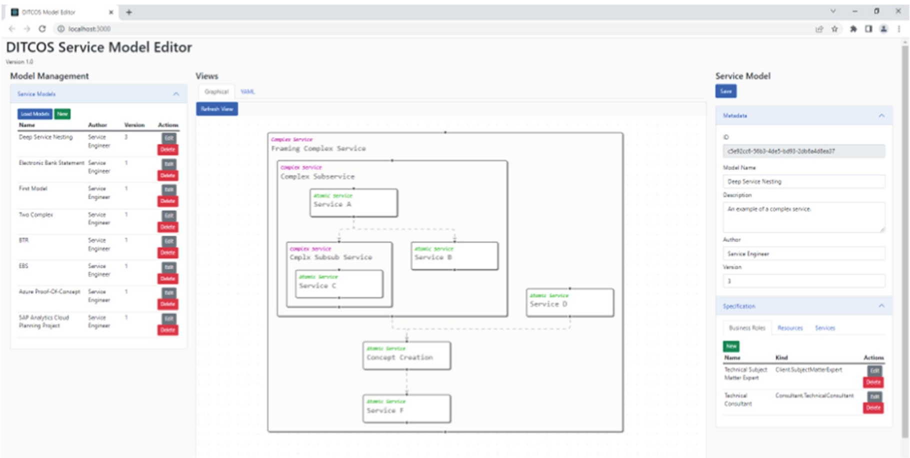


Fig. 3. DITCOS-ModEd showing the available service models and the graphical view as well as the details of the currently loaded service model (left to right).

5.4 Proof-of-Concept Study Demonstration

We set up a proof-of-concept study [14] to explore the practical applicability of our proposed editor from the perspective of IT consultants. The underlying motivation for this was to have a very first prototype demonstration and collection of first feedback from practitioners. We wanted to assure that our prototype design work goes in the right direction and that the editor matches the possible expectations of practitioners. To this end, DITCOS-ModEd was demonstrated by the first author to five practitioners. Each joined a one-on-one 60 min long session with the researcher. The five participants in our proof-of-concept study are all working for a midsize IT consultancy in Germany. Three were technical experts, while two were business process analysis experts with relatively little technical background. The areas in which the five participants were consulting their clients varied from SAP Finance, SAP Business Intelligence, SAP Logistics, Business Strategy, and Microsoft Azure Cloud.

Proof-of-Concept Session Setup. We prepared the editor and run it locally at the computer of the first author using the IDE built-in features. Every session started with an empty and clean editor instance.

Session Execution. We conducted all sessions remotely via Microsoft Teams due to the lockdown work-from-home policies. We prepared a note form which we used to take notes during the each session. For all sessions we executed the following process described in Table 3.

Table 3. Session execution phases

Phase and Description
Phase 1 (~5 min) We welcomed the practitioner and explained what are the aims of the session are and how we will proceed
Phase 2 (~15 min) We introduced DITCOS-O and DITCOS-DN shortly to the practitioner to make him familiar with the concepts of ITC service description. In particular the concepts of a DITCOS-O service model, such as atomic vs. complex service, service commitments, business roles, capabilities, and resources (see Fig. 1)
Phase 3 (~10 min) We introduced the practitioner to the DITCOS-ModEd editor and referred to the concepts presented in phase 2. Furthermore, we explained the user interface and its functionality to the practitioners. Using an example DITCOS-O service model, we showed how the previously introduced concepts can be modeled using the editor and how the graphical modeling reflects to the resulting YAML-based DITCOS-DN textual service description
Phase 4 (~25 min) We asked the practitioner to model a selected ITC service using the editor. This was done with the help the first researcher concerning the use of the editor's interface and navigation. Each expert did the conceptual modeling work himself, while the researcher took over the handling of the interface. This was since the session would have required a much longer training of the practitioner at this early stage of the editor's development. During this phase we continuously discussed the current state of the DITCOS-DN model and inspected the resulting YAML code. During this very short phase of 25 min, the aim was to give the practitioner a feeling for creating DITCOS-O service models using the editor with the goal of gathering his feedback which we would consider for inclusion in our future design cycles of the editor. A complete realization of a DITCOS-O service model would not have been possible within the time frame of 25 min
Phase 5 (~5 min) This phase was to wrap-up and let the practitioner tell us about his experience with the editor. We asked for feedback related to (i) the graphical user interface (GUI), (ii) the usability and concepts, and (iii) general suggestions and comments on possibilities to extend the prototype. The last included possible 'nice to have' integration to other tools that consultants use. In addition to that, we also noted our own (researcher) suggestions and observations during the sessions

5.5 Collected Feedback and Suggestions

All five of the practitioners liked the representation and the Look&Feel of the editor. One had the impression that it supported all the relevant DITCOS-O concepts well (see Fig. 1). Another practitioner stated that the interaction and working process with the GUI was particularly fluent. In addition, we received seven feedback items (see Table 4) to our notation concepts and components of the editor and four suggestions (see Table 5) for improvement of the artifacts.

Table 4. Feedback on usability and concepts

ID	Concept/Component	Name and Description
F1	AtomicServiceKind, BusinessRoleKind, Capability	**Insufficient Coverage** Realistic modeling requires more choices of the named concepts. Add more kinds of each concept for subsequent experiments
F2	BusinessRole	**Agent Occupation Mode** Add an attribute to the business role concept that indicates the 'agent occupation mode' to indicate that a certain role must be fulfilled by a dedicated agent that cannot have other roles assigned within the same service commitment
F3	BusinessRole	**Capability Level** Modeling levels of a capability is not possible in the sense that a certain business role requires 'SAP FI Senior' and 'SAP ABAP Intermediate' skill levels. Actually one can only define the overall level of a business role by assigning the capability 'Knowledge. Experience.(Junior, Intermediate, Senior)'
F4	BusinessRole	**Decision Power** Modeling of 'decision power' of a business role is not possible. This might be relevant if a business role must make decisions during the service provisioning. This flag seems to be relevant in cases necessitating the modeling of client-side roles. *Example*: A client subject matter expert is required to decide to go for a certain customization variant related to the chart of accounts
F5	ServiceCommitment	**Inline Definition of Concepts** The current version of the editor requires that the service modeler defines all resources and business roles before starting with the modeling of service commitments of atomic services. This does not feel 'natural' because, sometimes required resources or business roles only become obvious during the service modeling itself. In the current editor's version, the user would have to abort and step out of the service modeling and create additional resources or business roles and then re-enter the service modeling
F6	ServiceCommitment	**Planned Duration** Add 'planned duration' as attribute to the service commitment concept. Based on this, it would become possible to calculate estimated time consumption for a DITCOS-O service model. (Added by researcher)
F7	GraphicalView	**Drill Down** Add a recursive drill down functionality to the service nodes

5.6 Reflection on How the Editor Reaches Its Goals and Requirements

An important part of the DS research process, according to Peffers et al. [11] includes an analytical reflection on how the designed artifact meets its goals set at the beginning of the research. This section reports our analytically reflection on the results obtained throughout the design of our editor and the proof-of-concept evaluation. The reflection is in two regards: (1) how the editor meets our main research goals defined in Sect. 2, and (2) how it meets the additional design objectives formulated in Sect. 5.2 as requirements.

Table 5. Suggestions and comments regarding nice-to-have extensions

ID	Name and Description
S1	**BPMN Diagram** Add a BPMN diagram to graphicalize a DITCOS-O service model as BPMN process
S2	**Gantt Diagram** Add a Gantt diagram to graphicalize the sequence/parallel dependencies of a DITCOS-O service model
S3	**Integrated Help System** The editor requires an integrated help system to guide the service modeler continuously
S4	**Service Reuse** Avoidance of duplicated names for model entities. This is important when services shall be reused in complex services
S5	**Tags** Allow arbitrary tags for all model concepts to assign non-service-related information. Predefine tags to avoid duplicates

Results Pertaining Our Research Goals. We formulated two main research goals. We present the results in the next paragraphs:

Provide an Editor for DITCOS-O Service Models. We aimed on delivering an editor to graphically create DITCOS-O models defined in DITCOS-DN. We were able to achieve this goal. DITCOS-O services models created using the DITCOS-ModEd editor cover all DITCOS-DN concepts, such as atomic and complex service, service commitment, business role, capability, and resource. We checked the completeness and syntactic correctness for each of the mentioned concepts as well as the interaction of the concepts in the context of a DITCOS-O service model.

Proof-of-Concept Evaluation of the Artifact. We presented the DITCOS-ModEd editor to five practitioners in an experimental setting where each session last around 60 min. Beside an introduction to the DITCOS-O, DITCOS-DN artifacts, as well as the editor, we modeled jointly with the practitioner an ITC service, without the aim of being complete. Our goal was to utilize the features and functionalities of the editor and not to create a fully elaborated DITCOS-O service model. During the modeling process we constantly discussed the activity and collected feedback as well as suggestions from the practitioners.

Results Pertaining the Design Objectives

R1: Web-based. We built DITCOS-ModEd using REACT, one of the latest JavaScript frameworks to build recent web-applications. REACT applications are self-contained in sense of logic and consume web-based APIs, such as REST. With our artifact we were able to fulfill this requirement.

R2: Component-based. Our artifact is based 17 independent and reusable REACT components. All components consume a common data layer, that represents the currently active DITCOS-O service model expressed in DITCOS-DN. All components interact with their respective part of the DITCOS-DN description. This requirement could be fulfilled.

R3: External Lookups. This requirement is twofold. First, we provide certain value completion functionality to the different REACT components. Examples for value helps are AtomicServiceKinds, BusinessRoleKinds, ResourceKinds, or Capabilities (see Fig. 1). If a certain component gets activated it consumes an external REST API, that provides e.g., a constant list of capabilities as a response. This list gets consumed by the REACT component to provide a value help to the user. This first perspective could be fulfilled. Second, another perspective relates to dynamically created instances of concepts, such as atomic or complex services, business roles, or resources. Typically, these instances would have been created by other users and within other DITCOS-ModEd instances. They would have been persisted in a central service repository and provided to DITCOS-ModEd by appropriate REST-APIs. This functionality will only be supported by a later prototype of DITCOS-ModEd. The reason is, that we are at an early stage of the artifact development. Hence, our focus is on the required basic functionality to support the creation of DITCOS-O service models. A later prototype will support reuse of defined concept instances, what was also suggested (S4) by a practitioner (see Sect. 5.5). Overall we were able to fulfill this requirement partly.
R4: CRUD Support. With the current prototype of DITCOS-ModEd we fulfill this requirement for the following DITCOS-O primary concepts: atomic and complex service, service commitment, and business role. For these concepts we created dedicated REACT components, that support creation, update and delete completely (see Table 2). The existing components enable users to create feature-complete DITCOS-O service models. Nevertheless, with the current prototype we decided not to support CRUD for instances of supporting concepts, such as AtomicServiceKind, BusinessRoleKind, Capability, ResourceKind, ResourceInvolvementType MIMEType, ServiceCommitmentKind, and ServiceCommitmentExecutionFlavor (see Fig. 1). Instead, we decided to focus only on the creation of DITCOS-O service models and the required primary concepts and to assume the existence of the named supporting concepts. We simply provide instances of the supporting concepts in form of constant lists, as already described. In other words, for these supporting concepts, we only provide 'read' functionality with the current DITCOS-ModEd prototype. In light of this discussion, we could say that this requirement could be only partly fulfilled.
R5: Graphical View. With the current prototype we where able to provide a sophisticated graphical view on DITCOS-DN based DITCOS-O services models already. The view supports nodes, edges, zooming, change of viewport just to mention some features. Related to DITCOS-O service models anyhow, we only support a subset of the concepts with the current prototype. These are namely atomic and complex services as well as the edges to connect subsequent concepts (see Fig. 1). Nevertheless, the current prototype of the editor can provide a graphical representation of the two core concepts and also supports the required visualization of hierarchical organized complex services, which may contain sub-services which themselves could be complex services again. Based on this reasoning, we think that overall we fulfilled this requirement partly.

R6: YAML View. The current prototype of DITCOS-ModEd supports the complete rendering of DITCOS-DN based DITCOS-O service models in YAML representation. The YAML viewer supports code formatting, code-indent, and coloring. The current implementation only supports one-way YAML rendering based on an in-memory representation of the corresponding DITCOS-DN. Changes to the YAML code therefore, do not trigger changes to the REACT components dynamically. As this was not our aim with this early prototype of DITCOS-ModEd, we fulfilled this requirement completely.
R7: DITCOS-O Coverage. With the current prototype of the DITCOS-ModEd we realized a DITCOS-O coverage regarding to DITCOS-DN based service models a 100% completeness. Therefore, we conclude that we achieved this requirement completely.

6 Discussion of the Findings in Our First Evaluation

Our proof-of-concept study with the ITC practitioners showed that the current prototype of DITCOS-ModEd, even though it is at early stage, already was recognized as a helpful tool that supports the modeling process of DITCOS-O service models compared to the pure textual modeling by manually coding the necessary YAML using DITCOS-DN description notation. The feedback on GUI and the Look&Feel of DITCOS-ModEd is clear (see Sect. 5.5). Anyhow, the feedback on concepts we collected, as well as our own experience with DITCOS-ModEd made clear, that the editor is not yet ready to be transferred to the consulting practice. While this was expected, our proof-of-concept study indicated that the proposed editor did meet our research goals and requirements (to a very large extent).

Our immediate future research that builds upon DITCOS-ModEd must explicitly respond to the feedback (F1) related to insufficient coverage regarding to the concepts of AtomicServiceKind, BusinessRoleKind, and Capability (see Sect. 5.5). These concepts are consumed during the modeling process at different stages. For the current prototype it was sufficient, to only provide a small set of instances of each concept, as we only aimed on supporting the general modeling approach and not to be complete. During follow-up empirical evaluation we will ensure that the set of instances of these concepts are more complete.

Feedback F1 has no direct impact on the underlying artifacts DITCOS-O and DITCOS-DN, as only instances of these concepts are required, and they can be created at runtime as required. However, feedback points F2 to F6 directly affect the concepts BusinessRole and ServiceCommitment. These points require changes to the underlying ontology DITCOS-O and adjustments related to the description notation DITCOS-DN.

Related to the DITCOS-ModEd, we collected one suggestion (F7) namely, to add a drill down functionality to the graphical representation of the underlying DITCOS-DN description. From our point of view this makes absolutely sense. During the design cycles we experimented with different levels of details we added to the view. To ensure clarity, we opted for a rather reduced view of the model concepts. A drill-down to display further relevant details seems very useful here.

The general suggestions we received where also very helpful and interesting. In particular, the improvement ideas to add a BPMN and a Gantt representation to the underlying DITCOS-DN representation of the DITCOS-O service model seem worthy of attention. The question arises, which further requirements and additions to DITCOS-O/DN would be necessary for this?

Even though our aim is to provide an easy-to-use service description notation, the concepts might not always be self-explaining. Adding an integrated help system (S3) to DITCOS-ModEd is a desirable goal.

The practitioners' suggestions to reuse defined services (S4) and to allow arbitrary tags attached to all concepts, are also very good. As one participant indicated, service reuse is considered very important. Anyhow, we suggest 'concept reuse' as a more general improvement. In future prototypes of DITCOS-ModEd we plan to support the reuse of concepts in the sense, that concepts defined by a certain user can be referenced from other concepts created by another user. A prerequisite to support this feature is to implement a service repository as depicted and described in our digital ITC platform architecture proposal [2]. We will evaluate how to tackle the feedback items F2-F7 and suggestions S1-S5 during future research.

Finally, we reflected on the limitations of our empirical findings. We involved five practitioners and clearly it might be the case that if we could have included many more, it would have been possible to collect feedback points different from those that we have now. However, we note that we selected participants with exposure in a variety of consulting areas (ranging from finance, to logistics to cloud). Their perceptions and experiences of the 60 min sessions overlapped, specifically regarding the mater that they liked the editor and thought it indeed filled a gap in the consulting practice. We think that it might well be possible that similar perceptions might come if we possibly include in our sessions other consultants who might share the context in which our participants work. Consultants working in client organizations in the same business sector, implementing the same technology and following the same consulting approach, might well have similar conceptual modeling experiences and similar needs of an editor such as our participants do. As Ghaisas et al. indicate, similar contexts might possibly create similar work experiences [8]. However, as we plan to create more prototypes of the editor, a more empirical and longer-term evaluation is needed to fully assess its applicability, usefulness, and usability in a real-world context. This is our priority in the future.

7 Conclusions and Future Work

With this work we contributed the web-based, graphical editor DITCOS-ModEd to create DITCOS-O service models described using the DITCOS-DN description notation. Following the DS methodology [11], we developed a prototype of the editor and evaluated it in a proof-of-concept experimental study with ITC practitioners. The latter revealed some limitations of DITCOS-ModEd and enabled us to collect valuable and interesting feedback and improvement suggestions to enhance both the underlying artifacts DITCOS-O and DITCOS-DN, as well as the editor DITCOS-ModEd itself. Our reflection on the limitations helped us formulate immediate future research steps towards creating the next and improved prototype of our proposed editor.

References

1. Akkermans, H., Baida, Z., Gordijn, J., Peiia, N., Altuna, A., Laresgoiti, I.: Value webs: using ontologies to bundle real-world services. IEEE Intell. Syst. **19**(4), 57–66 (2004). https://doi.org/10.1109/MIS.2004.35
2. Bode, M., Daneva, M., van Sinderen, M.J.: Digital IT consulting service provisioning – a practice-driven platform architecture proposal. In: 25th IEEE International Enterprise Distributed Object Computing Workshop. The Gold Coast, Australia (2021). https://doi.org/10.1109/EDOCW52865.2021.00056
3. Bode, M., Daneva, M., van Sinderen, M.J.: Characterising the digital transformation of IT consulting services - results from a systematic mapping study. IET Softw. (2022). https://doi.org/10.1049/sfw2.12068
4. Bode, M., Daneva, M., van Sinderen, M.J.: Describing digital IT consulting services: the ditcos ontology proposal and its evaluation. In: 2022 IEEE 24th Conference on Business Informatics (CBI), Amsterdam, The Netherlands (2022). https://doi.org/10.1109/CBI54897.2022.00029
5. Cardoso, J., Pedrinaci, C.: Evolution and overview of linked USDL. In: Nóvoa, H., Drăgoicea, M. (eds.) IESS 2015. LNBIP, vol. 201, pp. 50–64. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14980-6_5
6. Frank, A.G., Mendes, G.H., Ayala, N.F., Ghezzi, A.: Servitization and Industry 4.0 convergence in the digital transformation of product firms: a business model innovation perspective. Technol. Forecast. Soc. Change **141**, 341–351 (2019). https://doi.org/10.1016/j.techfore.2019.01.014
7. Gartner: Gloassary. https://www.gartner.com/en/information-technology/glossary (2022)
8. Ghaisas, S., Rose, P., Daneva, M., Sikkel, K., Wieringa, R.J.: Generalizing by similarity: lessons learnt from industrial case studies. In: 2013 1st International Workshop on Conducting Empirical Studies in Industry (CESI), pp. 37–42. IEEE, San Francisco, CA, USA (2013). https://doi.org/10.1109/CESI.2013.6618468
9. Lauesen, S.: Software Requirements: Styles and Techniques. Addison-Wesley, Boston (2002)
10. Nissen, V. (ed.): Advances in Consulting Research. CMS, Springer, Cham (2019). https://doi.org/10.1007/978-3-319-95999-3
11. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007). https://doi.org/10.2753/MIS0742-1222240302
12. Starke, G.: Effektive Softwarearchitekturen, 7th edn. Hanser, München (2015)
13. Werth, D., Greff, T., Scheer, A.W.: Consulting 4.0 - Die Digitalisierung der Unternehmensberatung. HMD Praxis der Wirtschaftsinformatik **53**(1), 55–70 (2016). https://doi.org/10.1365/s40702-015-0198-1
14. Wieringa, R., Daneva, M.: Six strategies for generalizing software engineering theories. Sci. Comput. Program. **101**, 136–152 (2015). https://doi.org/10.1016/j.scico.2014.11.013

Implementing a Service-Oriented Rural Smartness Platform: Lessons Learned from a Technical Action Research in West Java, Indonesia

Iqbal Yulizar Mukti[1(✉)], Setiaji[2], Indah Dwianti[2], Adina Aldea[1], and Maria E. Iacob[1]

[1] Department of Industrial Engineering and Business Information Systems, University of Twente, Enschede, The Netherlands
i.y.mukti@utwente.nl

[2] Information and Communication Agency, West Java Provincial Government, Bandung, Indonesia

Abstract. Despite the growth of the digital economy in many developing countries, rural communities are still marginalized in the current digital business ecosystems, which are highly urban-oriented. To address this issue, we have proposed in our previous works a reference architecture of a digital platform based on the principles of service orientation. In this paper, we evaluate the real-world applicability of the proposed reference architecture by using a technical action research approach in the West Java province of Indonesia. Our study demonstrates the feasibility of implementing the proposed architecture in a real-world setting, and analyzes its effectiveness in improving the economic climate in rural areas.

Keywords: Smart rural · Service-oriented architecture · Digital business ecosystems

1 Introduction

Agglomeration of economic activities in urban areas, particularly, in many of the developing countries, has stimulated a rapid rural-urban migration [1]. Although this migration has the economic benefits of fueling economic activities in the urban areas, the unmanageable pace of rural-urban migration often results in serious negative consequences. Overpopulation in urban areas puts critical urban infrastructures under a lot of stress which causes various urban problems, such as, traffic congestion, energy crisis, and degradation of environmental quality [1]. Rural areas, on the other hand, lack the talent needed to fuel their economic engine since much of the population at a productive age is pulled towards cities [2]. This, in turn, makes alleviation of poverty, which mostly occurs in rural areas [3], more difficult. Lastly, because a large proportion of rural-urban migrants lack the skills necessary for well-paid jobs in urban areas, they typically end up in low-wage jobs, which traps them in a vicious cycle of poverty [1].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 129–144, 2023.
https://doi.org/10.1007/978-3-031-26886-1_8

To slow down rural-urban migration and reduce its negative consequences, the rural economic climate needs to be improved. A better economic climate in rural areas will allow rural citizens to have a decent livelihood without having to migrate to urban areas. In this regard, recent studies pointed out that the improvement of rural economic climate could be achieved by enabling rural communities to participate in a digital business ecosystem (DBE): an environment of interacting organizations and individuals that co-create value through shared digital platforms [4]. In practice, the leading sectors of a DBE are online commerce, online media, online travel, and online financial services [5].

DBEs are growing rapidly in many developing countries and are seen as a promising vehicle for economic growth. As an example, the internet economy in Southeast Asian countries has tripled from $32 billion in 2015 to $100 billion in 2019 [6] and is projected to exceed 8% of the GDP by 2025, closing the gap with a developed market like the United States where the internet economy accounted for 6.5% of the GDP in 2016 [6].

However, despite its promising economic potential, rural communities are still marginalized in the current DBE which is highly urban-oriented [7]**.** The situation in Indonesia and Vietnam illustrates this phenomenon. In Indonesia, the online sales volume from rural sellers only accounts for 0.5% of the total sales volume in 2017 ($ 0.024 billion compared to $ 5 billion) [8], and in Vietnam, 75% of the total e-commerce sales were generated only from its two big cities: Hanoi and Ho Chi Minh City [7].

Several barriers contribute to the marginalized situation of rural communities in the current DBE. First, rural communities have insufficient IT infrastructures and limited digital literacy [9] that hinder them from using the services provided in a DBE. Second, as the party responsible for rural development, the government has a strict financial budget [10] that limits its capacity to facilitate digital services for rural communities. Lastly, facilitating the rural communities to participate in a DBE is not economically attractive for service providers from the private sector [11]. Reasons behind that argument are the low purchasing power of the rural communities and the investment to improve their digital readiness seen as unprofitable [11].

With the aims to lower the aforementioned barriers and increase the participation of rural communities in the DBE, we have proposed in our previous works a reference architecture of a digital platform based on the principles of service-oriented architecture (SOA) that we refer to as the rural smartness platform [12, 13]. Through the platform, the provisioning of digital business services for rural communities is realized by a collaboration between the government and third-party service providers. We put forward the idea that the platform can work as an efficient mechanism to unlock the economic potential of the rural economy and stimulate the establishment of a rural DBE.

This paper aims to evaluate the implementation feasibility of the proposed reference architecture in the actual situation and identify readiness factors that must be satisfied to ensure a successful implementation. To achieve these aims, we carry out a technical action research (TAR) that will be further explained in the following section.

2 Research Methodology: Technical Action Research

TAR is a research methodology to evaluate a design artefact in a real-world setting by helping real actors with implementing it [14]. It is the process where the researcher is scaling up the evaluation of the design artefact from "laboratory" conditions to the unprotected conditions of practice [14]. In this study, we adapt the nested structure of TAR provided by Wieringa [14] into the systematic steps shown in Fig. 1. The structure provided in [14] is chosen because it comes with very clear steps to evaluate a design artefact in a real situation and it can be aligned to the evaluation phase of the design science research methodology (DSRM) that we used to design the proposed reference architecture in our previous works [12, 13].

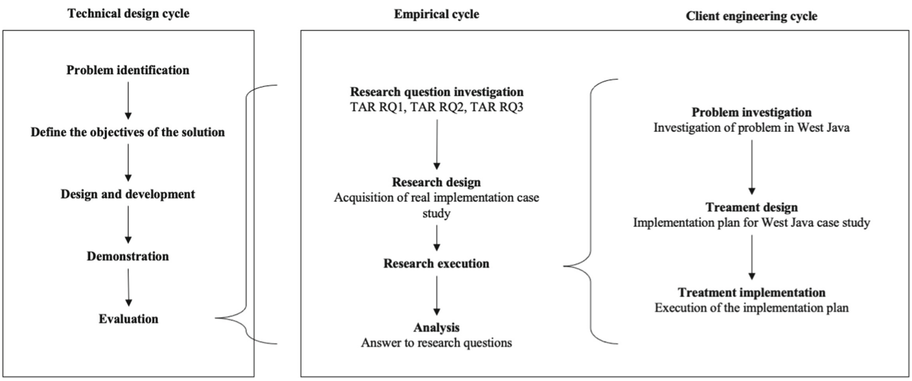


Fig. 1. Structure of TAR for this study (adapted from Wieringa [14])

In its essence, the nested structure of TAR consists of three research cycles: technical design cycle, empirical cycle, and client-engineering cycle [14]. The design cycle (the left column of Fig. 1) is a cycle where the researcher develops a design artefact to address a class of problems. The empirical cycle (the middle column of Fig. 1) is a cycle where the researcher evaluates research questions concerning the applicability of the design artefact in a real setting. Lastly, the client engineering cycle (the right column of Fig. 1) is a cycle where the researcher helps a real-world actor to implement the design artefact.

In our context, the technical design cycle has been carried out through the DSRM research activities in our previous work [12, 13], where we ultimately proposed a reference architecture for the rural smartness platform that is summarized in Sect. 3. The empirical cycle and the client-engineering cycle, which are the primary focus of this paper, are presented in the subsequent sections. We start the empirical cycle in Sect. 4 with the definition of research questions and research design for this TAR. Then, we continue with the client-engineering cycle in Sect. 5. Finally, based on our findings in the client-engineering cycle, in Sect. 6, we go back to the empirical cycle with analysis to answer the research questions.

By carrying out the research activities as prescribed by the above methodology, the study presented in this paper makes two contributions. First, the implementation approach and lessons learned presented in this paper can be used by practitioners as starting point for implementing the reference architecture of the rural smartness platform in their own context. Second, we propose an extension and adaptation of the TAR methodology (see Fig. 1), which can be used in future work as an evaluation method for a design artefact developed with DSRM.

3 Reference Architecture of the Rural Smartness Platform

The rural smartness platform is a digital platform aimed at improving the rural economic climate designed based on the SOA principles [15]. It provides digital services to accelerate the participation of rural businesses in a DBE through a collaboration between the government and the third-party service providers.

We follow the SOA approach as the basis for designing the platform because of the limitations in the previous approaches: top-down, community-driven, and single-partnership approaches [16]. The top-down approach, where the government acts as the main actor that provides end-to-end digital business services for rural communities, requires a large amount of government investment [17] which would be difficult to apply in low-middle income countries. The community-driven approach, in which different actors collectively drive and self-organize themselves to establish a DBE, takes time and does not always evolve into the desired state, especially when the rural areas have low digital readiness [16, 18]. Lastly, in the single partnership approach, the establishment of a rural DBE is done through exclusive cooperation between the government and a major digital service provider [16]. This approach is found to be successful in China, where the DBE is dominated by a single player and has strong legitimate support from the government [19]. However, it could be challenging to apply the same approach in a DBE environment dominated by multiple players (e.g., Southeast Asian countries, India, and Brazil), since a single partnership can threaten the healthy market competition which is essential for long-term economic vitality [20]. Therefore, with the above limitations in mind, we propose a multi-partnership approach based on SOA to establish a rural DBE.

In our previous work [12, 13], we have carried out DSRM research activities to propose a reference architecture of the rural smartness platform that operationalizes our approach. It is depicted with the ArchiMate standard [21] and consists of several viewpoints. A layered viewpoint of the reference architecture that summarizes our proposed approach is presented in Fig. 2.

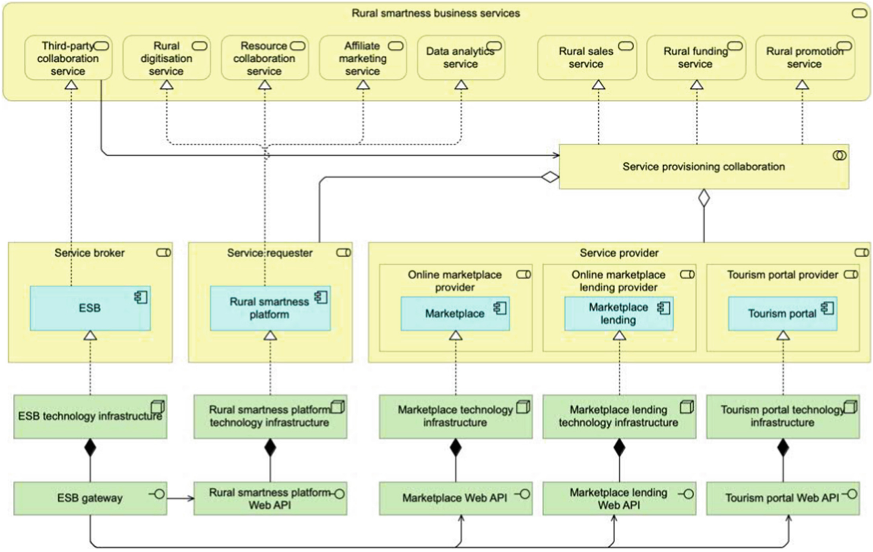


Fig. 2. Layered architecture of the rural smartness platform (adapted from [13])

Table 1. Rural smartness business services

Business service	Description
Rural digitisation service	Service that facilitates rural communities to convert their business-related information (e.g., business profile, offerings, or local attractions) into a digital form
Third-party collaboration service	Service that enables the provision of digital services through collaboration with third-party providers
Rural sales service	Service that facilitates rural businesses to sell their products to a broader market. The sales processes are carried out through a data exchange mechanism where the rural products are registered and managed in the rural smartness platform, and the products' sales fulfillment is facilitated by the third-party online marketplaces
Rural funding service	Service that facilitates rural businesses to obtain funding from external funding sources. The funding processes are carried out through a data exchange mechanism where the funding requests are managed in the rural smartness platform, whereas fulfillment of the funding is facilitated by the third-party marketplace lending providers

(*continued*)

Table 1. (*continued*)

Business service	Description
Rural promotion service	Service that disseminates information on rural touristic attractions. The touristic information is conveyed to broader audiences through a data exchange mechanism with tourism portals that share information on tourism destinations over the internet
Resource collaboration service	Service that facilitates online collaboration among rural businesses for resource procurement
Affiliate marketing service	Service that enables citizens to participate in promoting rural offerings over the internet to support the rural sales service
Data analytics service	Service that facilitates analytics process on the data collected by the platform

The layered viewpoint (Fig. 2) depicts the interaction between the service broker, the service requester, and the service provider. In the context of the proposed platform, the government takes responsibility as the service requester and the service broker, establishing the rural smartness business services since rural development activities are seen as unprofitable by the private sector [11]. On the other hand, third parties from the private sector play the role of the service provider.

In essence, the layered viewpoint explains that the third-party collaboration service provided by the service broker enables the service requester to collaborate with the service provider to realize the rural sales service, the rural funding service, and the rural promotion service. Meanwhile, the service requester provides, by itself, the rural digitization service, the resource collaboration service, the affiliate marketing service, and the data analytics service. As for the technological components, the web API technology facilitates the information exchange between the service requester and the service provider.

With this collaborative approach, the government could concentrate its efforts on ensuring the digital readiness of the rural communities, without large investments in the development of functionality needed for the required services. On the other hand, the third-party service providers are able to offer the rural communities access to their digital services without the burden of ensuring their digital readiness.

The focus of this paper is to evaluate whether our approach presented in the reference architecture is feasible to be implemented in a real setting and is perceived to be effective by the actual users. Therefore, the process of constructing the proposed reference architecture and its detail is not part of the scope of this paper. Those belong to our previous works which can be found in [12, 13].

4 Research Questions and Research Design

The aim of TAR in this study is to evaluate the applicability of the rural smartness platform reference architecture in a real-world setting. Therefore, to achieve that aim, we investigate the following research questions: 1) can the proposed reference architecture of the rural smartness platform be implemented in a real-world setting? (TAR RQ1), 2) is the rural smartness platform, which instantiates the proposed reference architecture, perceived to be effective in improving the rural economic climate by the actual users? (TAR RQ2), and 3) what kind of readiness factors must be satisfied to ensure a feasible instantiation of the proposed reference architecture in a real setting? (TAR RQ3).

To be able to answer the above research questions, we need to acquire a real implementation case study. For that purpose, we collaborated with the ICT Agency of West Java province, Indonesia, which took our reference architecture as a baseline for the actual large-scale implementation of a rural smartness platform in the region.

5 West Java Implementation Case Study

We carry out several research activities to implement the proposed reference architecture in West Java, Indonesia. First, we investigate the problem concerning the rural economic climate in West Java to see whether the implementation of the proposed reference architecture is relevant to the actual context (Subsect. 5.1). Then, we formulate an implementation plan that is suitable for the context in West Java (Subsect. 5.2). Finally, based on the plan, together with the ICT Agency of West Java, we execute the implementation project (Subsect. 5.3).

5.1 Problem Investigation

West Java has a lot of economic potential available in its rural areas. Depending on the geographical location, local resources, and socio-economic situation, this potential can materialise itself in agriculture (e.g., crops and livestock), consumer goods (e.g., culinary products and crafts), or tourism (e.g., natural landscapes and cultural events). However, rural businesses face difficulties to turn this potential into actual economic values, mainly due to limited market access, limited awareness of rural attractions, inefficient supply chain, and limited funding opportunities for rural entrepreneurs.

To address the aforementioned difficulties, the provincial government of West Java has been accelerating the diffusion of IT innovation in its rural areas through the West Java Digital Village program. As part of this program, the government has been implementing basic internet infrastructure throughout rural areas of West Java. As of March 2020, 4,541 out of 5,342 villages (85%) have been equipped with internet access [22]. However, our empirical study from a sample of 202 villages in West Java suggests that the current advancement of internet penetration in rural areas has a low impact on the improvement of the rural economic climate [23]. One of the main reasons behind that finding, as suggested from our survey in [23], is because the dominant usage of the provided internet services is for social communication (e.g., social media and instant messaging) and information access (e.g., internet browsing), thus, mostly for entertainment purposes and not for the online activities that have economic benefits.

To improve the impact of the current IT infrastructure advancement in its rural areas, the provincial government aims to facilitate rural businesses with digital services to improve their market access, supply chain, access to funding sources, and awareness of rural attractions. However, due to limited budget and capabilities, the provincial goverment cannot facilitate and manage the necessary digital business services to fulfil the needs of the whole rural business community timely.

Furthermore, the government cannot rely on third-party providers from the private sector to provide these necessary digital business services. This is because the citizens in the rural areas of West Java have relatively low purchasing power. Additionally, since the digital village program is still at an early stage, they still lag behind in terms of digital literacy. Therefore, although the third-party digital service providers understand the potential of the rural market and are interested in offering their digital services to this market, engagement with rural businesses at this time is not economically attractive. In addition, the government need to get accurate information concerning the economic potential and business activities in rural areas to improve the effectiveness of their strategic intervention for rural economic development. Such information would be difficult to obtain if they fully rely on third-party service providers.

The above challenges in providing the necessary rural digital business services motivate the provincial government of West Java to implement the proposed reference architecture of the rural smartness platform in its region. By implementing the proposed reference architecture, the government could efficiently provide the necessary digital business services since the fulfilment of those services is carried out by third-party service providers. Thus, the government could focus its effort on improving the digital literacy of rural communities. On the other hand, the third-party service providers no longer need to make a high investment to facilitate the access of rural businesses to their services. Instead, they just have to open their APIs to the rural smartness platform.

5.2 Implementation Plan

Together with the ICT Agency of West Java, we made a plan to instantiate the proposed reference architecture for the use of rural businesses in West Java. The goal is to implement all of the rural smartness business services (see Table 1) that are covered by the proposed reference architecture. However, considering the existing capability of the ICT Agency of West Java, it is unrealistic to implement all of those services at once. Therefore, we decided to split the implementation project into two phases: the foundation phase and the enhancement phase.

The foundation phase aims to implement the SOA approach as the backbone of the reference architecture and facilitates sales for rural products since it can trigger acceleration in rural business activities. Therefore, this phase includes the implementation of the rural digitisation services, the third-party collaboration service, the rural sales service, and the data analytics service. The remainder of the rural smartness business services (i.e., the rural funding service, the resource collaboration service, the rural promotion service, and the affiliate marketing service) are planned to be implemented in the enhancement phase.

5.3 Execution of the Implementation Plan

In this sub-section, we give an account on the execution of the implementation plan. However, since the implementation project is still ongoing, we only report the implementation within the foundation phase, which has been carried out from March 2021 until March 2022.

The ICT Agency of West Java started the implementation project by establishing the implementation project team. Initially, the implementation of the actual rural smartness platform was planned to be fully developed by the ICT Agency of West Java. However, due to the covid-19 pandemic, most of the agency's resources were re-directed to handle the pandemic in the region. Therefore, the government eventually formed a partnership with a private company to develop the actual platform. In the partnership, the government's main responsibility is the stakeholder and user engagement, whereas the main responsibility of the private company is the technical development and operation of the platform. Although the platform is developed by the private company, the platform is part of the West Java Digital Village program, and the government has the right to access the data within the platform. Meanwhile, our role during this implementation project is to help the implementation team by explaining the details of the reference architecture, supporting the day-to-day development activities, and assisting them in rolling out the implementation to the actual users.

During the period of the foundation phase, the rural smartness business services within the scope have been implemented. The rural digitization service has been realized by facilitating the registration of rural products (see Fig. 3.a). The third-party collaboration service has been realized by the API integration with two of the largest online marketplace providers in Indonesia: Tokopedia and Shopee (see Fig. 3.b, Fig. 3.c and Fig. 3.d). The rural sales service has been realized through the facilitation of an end-to-end sales process that is done through the data exchange mechanism between the rural smartness platform and the partnered marketplaces (see Fig. 3.e). Finally, the data analytics service has been realized by the data reporting functionality that grants the government access to the data generated in the platform, including products, transactions, and users' demographical data (see Fig. 3.f).

To attract rural businesses to become users of this platform, we help the ICT Agency of West Java to carry out several user engagement activities. To this end, we conducted workshops for selected rural businesses and give them continuous assistance on the usage of the platform after the workshops. However, due to the covid-19 pandemic, we were not able to organize on-site workshops. Therefore, the all workshops took place online.

We have conducted two online workshops with two groups of participants during the foundation phase: rural micro-businesses (i.e., business entities owned by rural entrepreneurs with fewer than 10 employees) and village-owned enterprises (i.e., business entities owned by the village government that has the responsibility to carry out business activities for the welfare of the community within that respective village). During the workshops, we explained the overall concept of the rural smartness platform, demonstrated the use of the actual rural smartness platform, and then guided them to use it. In addition, for the rural smartness business services not covered within the foundation phase, we demonstrated to the participants our prototype of the applications that cover those services. After the workshops, we added the participants to an instant messaging group which was used as the communication channel for continuous assistance on the usage of the platform.

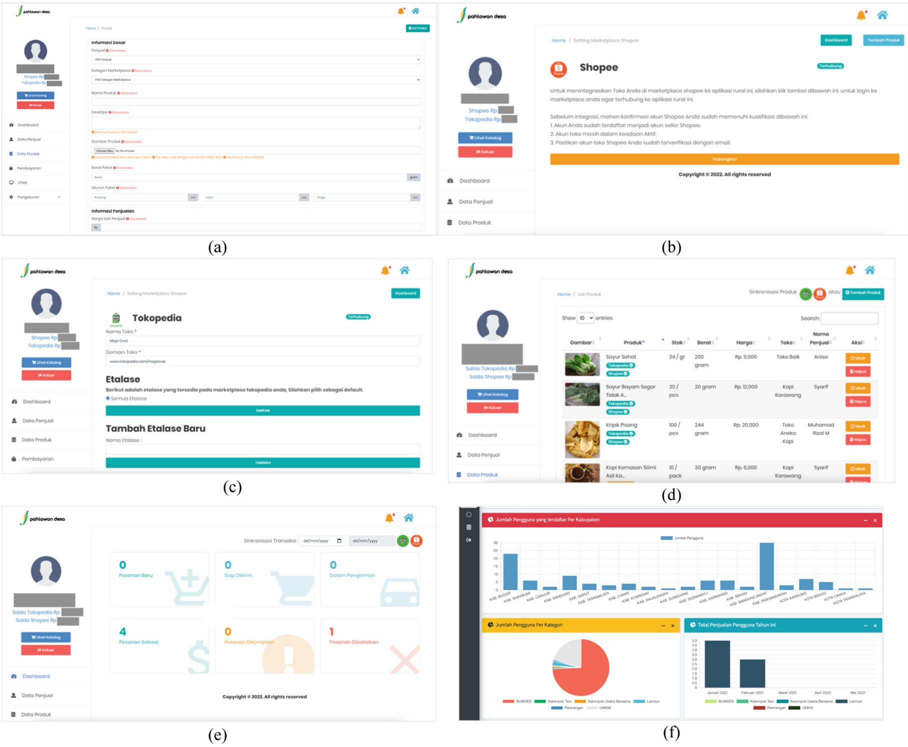


Fig. 3. Screenshots of the actual platform: a) registration of the rural product, b) synchronization setup with Shopee, c) synchronization setup with Tokopedia, d) synchronized products to the partnered online marketplaces, e) dashboard panel to manage incoming orders from the partnered marketplace, f) data reporting dashboard

6 Analysis: Answers to the TAR Research Questions

This section describes the last step of the TAR where we provide answers to the research questions based on our findings during the implementation of the foundation phase. These answers are described below.

TAR RQ1: Can the proposed reference architecture of the rural smartness platform be implemented in a real-world setting?

As described in Subsect. 5.3, during the foundation phase, the rural smartness business services within the scope have been implemented. These business services cover the implementation of the SOA approach as the backbone of the reference architecture and sales facilitation to trigger business activities in rural areas. The SOA approach has been realised by the facilitation of data exchange mechanism between the ICT Agency of West Java (as the platform operator) and the online marketplace companies (i.e., Tokopedia and Shopee) in providing the rural sales business service. In SOA terminology, the West Java ICT Agency plays the role of both service requester and service broker. On the other hand, the online marketplace companies play the role of service providers.

The aforementioned implementation results during the foundation phase confirm the general feasibility of the proposed platform architecture. The main reason for this claim is the fact that our platform architecture is essentially service-oriented, which is a guarantee for the feasibility of plugging in other remaining services (i.e., the rural funding service, the resource collaboration service, the rural promotion service, and the affiliate marketing service) at a later phase.

However, for the successful implementation of the complete architecture, engagement with the third-party service providers is a critical readiness factor that needs to be ensured. This readiness factor concerns both business and technical aspects. For the business aspect, it is essential to formulate a business model that can add value to the service providers. Furthermore, it is also important to define a clear responsibility for each party involved in the collaboration. On the other hand, for the technical aspect, it is critical to ensure interoperability with the service providers. This interoperability readiness concerns with the differences in application processes, data fields, message formats, and APIs used by the multiple service providers participating in the rural smartness platform.

TAR RQ2: Is the rural smartness platform, which instantiates the proposed reference architecture, perceived to be effective in improving the rural economic climate by the actual users?

We answer this question with the help of a survey of the rural businesses that have been registered on the actual platform and participated in the workshop during the foundation phase. The survey aims to assess their perception of whether their use of the rural smartness platform will be more effective in supporting the rural business ecosystem in comparison to their current situation (i.e., doing business without using the platform).

The survey is operationalized with 6 measurement indicators described in Table 2. These indicators are chosen because they represent innovativeness and competitiveness that were proved to be the strong predictors of rural economic welfare improvement, as has been empirically validated in the theoretical model of rural smartness [23].

Table 2. Description of measurement indicators

Measurement indicators	Description
Innovativeness	
Business collaboration	The degree to which the rural smartness platform helps the rural businesses to carry out business collaboration
Value creation	The degree to which the rural smartness platform helps the rural businesses to create new economic value
Entrepreneurship	The degree to which the rural smartness platform supports the growth of entrepreneurship in rural areas
Competitiveness	
Market access	The degree to which the rural smartness platform helps the rural businesses to sell their offerings to a broader market
Business productivity	The degree to which the rural smartness platform helps the rural businesses increase their productivity
Business efficiency	The degree to which the rural smartness platform helps the rural businesses to improve the utilization, sharing, and distribution of their resources

We distributed the survey through an online self-administered questionnaire. The respondents were asked to score their expected performance (on a five-point scale) with respect to the measurement indicators in the two scenarios: if they use the rural smartness platform and if they don't.

There were 47 respondents that provided responses for both scenarios. Based on these responses, we used the multivariate analysis of variance (MANOVA) to test whether the use of the rural smartness platform has a better impact on the performance of the rural business environment (which is measured by innovativeness and competitiveness) in comparison to the existing situation. Table 3 shows the test results. Since the sample size for each scenario was above 30 (i.e., a group size of 47), these results are robust against the violations of the MANOVA assumptions [24].

To identify the difference between the two scenarios (i.e., with and without platform use), we utilised Pillai's Trace statistic. Pillai's Trace was chosen because it is more robust than the other multivariate test criteria [25]. The results of the Pillai's Trace ($F_{6,87} = 13.308$, $p < 0.001$) provide empirical evidence that a significant difference exists between using the rural smartness platform and relying on the current situation. Further assessment of the mean score of each measurement indicator shows that the use of the rural smartness platform is perceived by the respondents to bring a significant improvement to the rural business ecosystem compared to the current situation. Moreover, according to Richardson [26], the value of the partial eta squared (η_p^2) on each measurement indicator, that is greater than 0.14, indicates that the use of the rural smartness platform has large effects on improving the innovativeness and the competitiveness of business ecosystem in rural areas.

Table 3. MANOVA results

Measurement indicator	Using the rural smartness platform		Current situation		p-Value	η_p^2	Multivariate test
	Mean	Std.	Mean	Std.			
Innovativeness							Pillai's Trace: *F = 13.308* *df = 6* *error df = 87* *p < 0.001*
Business collaboration	3.9787	0.92052	2.7234	0.99350	<0.001	0.305	
Value creation	4.0851	0.77543	2.6170	0.87360	<0.001	0.447	
Entrepreneurship	4.0426	0.72103	2.7660	1.04700	<0.001	0.340	
Competitiveness							
Market access	4.1277	0.79720	2.6596	1.14733	<0.001	0.361	
Business productivity	4.0851	0.77543	2.6596	0.96181	<0.001	0.405	
Business efficiency	4.0213	0.84672	2.6383	0.87042	<0.001	0.399	

TAR RQ3: what kind of readiness factors must be satisfied to ensure a feasible implementation of the proposed reference architecture in a real setting?

To answer this question, we analyze our experience in carrying out the foundation phase of the implementation project. We extract factors that must be satisfied to implement the proposed reference architecture in three contexts following the TOE framework [27]: technological, organizational, and environmental. The technological context relates to the required technological elements enabling the implementation of the proposed reference architecture. The organizational context captures the organizational capabilities needed to ensure the implementation of the reference architecture is feasible. Lastly, the environmental context includes the necessary readiness factors in the social structure surrounding the implementation of the proposed reference architecture. This framework is chosen because it has been widely used to study the diffusion of technology innovation in various contexts and has a high explanatory power [28].

Table 4 lists our findings. Our results show that the capabilities of the regional government (readiness factors for the organizational context) hold a crucial role in driving the implementation of the proposed reference architecture as it determines the fulfilment of the requirements in both technological and environmental contexts. From our experience during the implementation project in West Java, the readiness of IT infrastructure and digital literacy in rural areas are strongly reliant on the government because they are not economically attractive and feasible for the private sector. Furthermore, implementation of the proposed reference architecture requires an active collaboration of the government (as the operator of the rural smartness platform) with the third-party service providers. However, this collaboration is unlikely to happen if the government does not initiate it, because rural communities are not the priority of the third-party service providers as they have predominant urban-orientation [7, 8].

Table 4. Readiness factors to implement the proposed reference architecture in a real setting

Context	Readiness factors
Technological	1. Internet access and IT devices for internet connectivity are available in rural areas, at least at the level of the formal rural business units (e.g., village-owned enterprises and rural cooperatives) 2. The third-party digital services (i.e., the online marketplace platforms, the online marketplace lending platforms, and the tourism portals) are possible to be connected with the rural smartness platform (e.g., through the availability of open APIs)
Organisational	1. The regional government has a strong commitment to adopting IT innovation to improve the rural economic climate 2. The regional government drives the initiatives to provide IT infrastructure in rural areas (internet connectivity and IT devices) and to improve the digital literacy of rural citizens 3. The regional government has the necessary resources and capability (owned by the government or provided through collaboration) to develop and implement the rural smartness platform 4. The regional government is capable to collaborate with the related stakeholders, including the third-party service providers, the related government agencies within the region, and the rural communities
Environmental	1. Rural areas in the region have potential offerings that can be turned into economic values 2. Rural businesses are willing to utilise technology in their business activities 3. Rural businesses have sufficient digital literacy to use an internet-based application, at least at the level of the formal rural business units (e.g., village-owned enterprises and rural cooperatives) 4. The third-party service providers are willing to collaborate in the implementation of the rural smartness platform

In particular, we found an important finding in the environmental context. The reliance of the rural businesses' digital readiness on the government has a practical consequence on user engagement, especially at the early stage of the implementation project. Instead of engaging with random rural businesses, the user engagement activities should focus on attracting rural business units that are part of the government's structure, such as village-owned enterprises, to become users of the platform as its early adopters. The main argument for this preference is that the government has already an established relationship with them and can easily support them in achieving digital readiness (IT infrastructure and digital literacy). Our observation during the implementation of the foundation phase justifies this approach. We observed that the village-owned enterprises have a better digital readiness and are more active in using the platform than the rural micro-businesses. In addition, the village-owned enterprises can act as a hub that also sells the rural microbusinesses' products. This way, the rural micro-businesses can focus their effort on creating the products instead of struggling with their digital readiness.

7 Conclusion and Future Work

In this paper, we evaluate whether the reference architecture of the rural smartness platform that we proposed in our previous work [12] is applicable and effective in a real-world setting. For that purpose, we carry out a TAR by working closely with the ICT Agency of West Java provincial government, Indonesia, which took our proposed reference architecture as the baseline for the actual large-scale implementation of the actual platform in their region.

The implementation results of TAR in West Java confirm the general feasibility of the proposed platform architecture. During the foundation phase of the implementation project, we have implemented the SOA approach, which is the backbone of the reference architecture to facilitate the acceleration of rural business activities. Furthermore, the statistical analysis based on the survey data from rural businesses that participated during the foundation phase of the implementation project provides empirical evidence that the use of the rural smartness platform is perceived to have large effects on improving the innovativeness and competitiveness of the business ecosystem in rural areas. In addition, we have extracted from our experience in the TAR, the readiness factors that must be satisfied to ensure a feasible implementation of the proposed reference architecture in a real setting.

Finally, this paper has several limitations that provide directions for future work. First, since the implementation of the proposed reference architecture during the period of this TAR has just reached the deployment stage, the size of the economic benefits resulting from the implementation is not yet visible. Therefore, the next step would be to measure the actual economic impact following the deployment of the rural smartness platform in West Java, which is currently in progress. For this, a longitudinal study is necessary to understand how the impact of the platform changes overtime depending of speed and scale of the adoption. Second, we did not study the willingness of rural businesses to use the actual rural smartness platform as part of their business activities. This limitation leaves room for future work to explore motivations and factors affecting the use behavior of the platform.

References

1. Zhang, X.Q.: The trends, promises and challenges of urbanisation in the world. Habitat Int. **54**, 241–252 (2016)
2. Naldi, L., Nilsson, P., Westlund, H., Wixe, S.: What is smart rural development? J. Rural. Stud. **40**, 90–101 (2015)
3. Castañeda, A., Doan, D., Newhouse, D., Nguyen, M.C., Uematsu, H., Azevedo, J.P.: A new profile of the global poor. World Dev. **101**, 250–267 (2018)
4. Senyo, P.K., Liu, K., Effah, J.: Digital business ecosystem: literature review and a framework for future research. Int. J. Inf. Manage. **47**, 52–64 (2019)
5. Google: Temasek and Bain: e-Conomy SEA (2020)
6. Google & Temasek/Bain: e-Conomy SEA (2019)
7. Kshetri, N.: Rural e-commerce in developing countries. IT Professional **20**, 91–95 (2018)
8. McKinsey & Company: The Digital Archipelago: How online commerce is driving Indonesia's economic development. McKinsey & Company (2018)

9. Katara, S.K.: Envisioning smart villages through information and communication technologies – a framework for implementation in India. In: Chugunov, A.V., Bolgov, R., Kabanov, Y., Kampis, G., Wimmer, M. (eds.) DTGS 2016. CCIS, vol. 674, pp. 463–468. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49700-6_46
10. Said, M., Supriyono, B., Muluk, M., Haryono, B.: The trade-offs budget of archipelagic local government in Indonesia. Australas. Account., Bus. Financ. J. **14**(1), 58–70 (2020). https://doi.org/10.14453/aabfj.v14i1.6
11. Philip, L., Williams, F.: Remote rural home based businesses and digital inequalities: understanding needs and expectations in a digitally underserved community. J. Rural. Stud. **68**, 306–318 (2019)
12. Mukti, I.Y., Firdausy, D.R., Aldea, A., Prambudia, Y., Sinderen, M.v., Iacob, M.E.: Architecting a service-oriented rural platform : Improving rural economic climate through participation in the digital business ecosystem. In: 2021 IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC), pp. 92–103 (2021)
13. Mukti, I.Y., Firdausy, D.R., Aldea, A., Iacob, M.E.: Architecting rural smartness: a collaborative platform design for rural digital business ecosystem. The Electron. J. Inform. Syst. Developing Countries **89**, e12236 (2022). https://doi.org/10.1002/isd2.12236
14. Wieringa, R.J.: Design Science Methodology for Information Systems and Software Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43839-8
15. The Open Group: SOA Reference Architecture (2011)
16. Li, L., Du, K., Zhang, W., Mao, J.-Y.: Poverty alleviation through government-led e-commerce development in rural China: an activity theory perspective. Inf. Syst. J. **29**, 914–952 (2019)
17. Jung, M.C., Park, S., Lee, J.Y.: Information network villages: a community-focused digital divide reduction policy in rural Korea. J. Telecommun. Dig. Econ. **2**, 21 (2014)
18. Leong, C.M.L., Pan, S.-L., Newell, S., Cui, L.: The emergence of self-organizing e-commerce ecosystems in remote villages of China: a tale of digital empowerment for rural development. MIS Q. **40**, 475–484 (2016)
19. Kwak, J., Zhang, Y., Yu, J.: Legitimacy building and e-commerce platform development in China: the experience of Alibaba. Technol Forecast Soc Change **139**, 115–124 (2019)
20. Porter, M.E.: Competition and antitrust: toward a productivity-based approach to evaluating mergers and joint ventures. The Antitrust Bull. **46**, 919–958 (2001)
21. The Open Group: https://pubs.opengroup.org/architecture/archimate3-doc/
22. West Java Digital Service: Desa Digital: Menuju Desa Jabar Juara dengan Inovasi Digital. West Java Digital Service (2020)
23. Mukti, I.Y., Henseler, J., Aldea, A. et al.: Rural smartness: its determinants and impacts on rural economic welfare. Electron. Markets 32, 1943–1970 (2022). https://doi.org/10.1007/s12525-022-00526-2
24. Topchyan, R., Woehler, C.: Do teacher status, gender, and years of teaching experience impact job satisfaction and work engagement? Educ. Urban Soc. **53**, 119–145 (2021)
25. Finch, H.: Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. Methodology **1**, 27–38 (2005)
26. Richardson, J.T.E.: Eta squared and partial eta squared as measures of effect size in educational research. Educ. Res. Rev. **6**, 135–147 (2011)
27. Tornatzky, L.G., Fleischer, M.: The Processes of Technological Innovation. Lexington books, MA (1990)
28. Trang, S., Zander, S., Kolbe, L.: E-Business adoption at the firm level: Comparing the predictive power of competing IS adoption models. In: Thirty Fifth International Conference on Information Systems (2014)

TEAR 2022

17th Workshop on Trends in Enterprise Architecture Research (TEAR 2022)

Preface

By adopting a holistic, multi-disciplinary stance to an enterprise, the Enterprise Architecture field contributes to an enterprise's need to adapt increasingly fast to changing customer requirements and business goals. In particular, EA covers an enterprise's entire "business-to-IT stack" by explicitly incorporating business-related artifacts in addition to conventional IS/IT artifacts. The Trends in Enterprise Architecture Research (TEAR) workshop series aims to bring together EA researchers from different research communities and provides a forum to present EA research results and discuss future EA research directions.

The 2022 TEAR workshop constituted the 17th iteration of TEAR. It was the first in-person TEAR workshop since the Covid-19 pandemic, as reflected by the titular "connecting people in a connected world." TEAR 2022 was co-located with the 2022 EDOC conference in Bolzano, Italy. We received 12 submissions, of which six papers were selected for presentation and inclusion in the workshop proceedings. This year's TEAR papers cover diverse topics, roughly divided into papers studying EA-related phenomena and design-oriented EA modeling papers.

When it comes to studying EA, the paper by Plessius, Steenbergen, Ravesteijn and Versendaal reports on the results of a survey among enterprise architects and EA stakeholders to reflect on the perceived value of EA. Meanwhile, the paper by Guo and Gao studies EA artifacts and their contribution to business-IT alignment. Finally, the paper by Rouvari and Pekkola discusses how each EA project is unique and that associated issues of contextuality and temporality require a project-specific set of coping mechanisms.

When it comes to modeling EA, the paper by Calhau and Almeida proposes to extend EA modeling with competence-related concepts. To do so, among others, they employ the UFO reference ontology. Alsufyani and Gill propose an ontology to establish a clear relationship between digital EA maturity and performance measures tracking said maturity. Finally, the paper by Grabis presents a workplace topology model of employee interactions to analyze the spread of infectious deceases. The model is developed as a multi-layer graph, whereby a subset of EA concerns constitutes the layers.

We want to thank the TEAR program committee members for their reviewing effort, thus helping us to select the above papers, and thank the authors for their exciting contributions, leading to a lively and constructive discussion at the 2022 TEAR workshop. Moreover, we want to thank both the TEAR Steering Committee and the EDOC 2022 Workshops Chairs for providing us with the opportunity to organize the workshop.

October 2022

Sybren de Kinderen
Dominik Bork

Organization

Workshop Chairs

Sybren de Kinderen	Eindhoven University of Technology, The Netherlands
Dominik Bork	TU Wien, Austria

Program Committee

Agnes Nakakawa	Makerere University Kampala, Uganda
Andreas L. Opdahl	University of Bergen, Norway
Aurona Gerber	CAIR, University of Pretoria, South Africa
Ben Roelens	Open University of the Netherlands & Ghent University, Belgium
Colette Rolland	University Paris 1 Pantheon-Sorbonne, France
Dirk Stelzer	TU Ilmenau, Germany
Dominik Bork	TU Wien, Austria
Elena Kornyshova	Conservatoire National des Arts et Métiers, France
François Coallier	École de Technologie Supérieure, France
Geert Poels	Ghent University, Belgium
Giuseppe Berio	Université de Bretagne Sud and IRISA UMR 6074, France
Jānis Grabis	Riga Technical University, Latvia
João Paulo A. Almeida	Federal University of Espírito Santo, Brazil
Juergen Jung	Frankfurt University of Applied Sciences, Germany
Marc Lankhorst	BiZZdesign, The Netherlands
Maria-Eugenia Iacob	University of Twente, The Netherlands
Markus Buschle	zeb.consulting AB, Sweden
Mathias Ekstedt	KTH Royal Institute of Technology, Sweden
Paul Drews	Leuphana University of Lüneburg, Germany
Peter Loos	IWi at DFKI, Saarland University, Germany
Pierre-Martin Tardif	Université de Sherbrooke, Canada
Pontus Johnson	KTH Royal Institute of Technology, Sweden

Q. Neo Bui	Rochester Institute of Technology, USA
Rainer Schmidt	Munich University of Applied Sciences, Germany
Remco Dijkman	Eindhoven University of Technology, The Netherlands
Rik Eshuis	Eindhoven University of Technology, The Netherlands
Simon Hacks	University of Southern Denmark, Denmark
Stefan Strecker	University of Hagen, Germany
Stephan Aier	University of St. Gallen, Switzerland
Ulrik Franke	RISE, Sweden
Ulrike Steffens	HAW Hamburg, Germany
William Schiano	Bentley University, USA

Steering Committee

Henderik A. Proper	Luxembourg Institute of Science and Technology & University of Luxembourg, Luxembourg
Florian Matthes	Technische Universität München, Germany
James Lapalme	École de Technologie Supérieure, Canada
João Paulo A. Almeida	Federal University of Espírito Santo, Brazil
Marc Lankhorst	BiZZdesign, The Netherlands
Mathias Ekstedt	Royal Institute of Technology, Sweden
Pontus Johnson	Royal Institute of Technology, Sweden
Stephan Aier	University of St. Gallen, Switzerland

Areas Where Enterprise Architecture Contributes to Organizational Goals – A Quantitative Study in The Netherlands

Henk Plessius[1](✉), Marlies van Steenbergen[2], Pascal Ravesteijn[2], and Johan Versendaal[2]

[1] Eduples, Amersfoort, The Netherlands
henk@eduples.nl
[2] University of Applied Sciences Utrecht, Utrecht, The Netherlands
{marlies.vansteenbergen,pascal.ravesteijn, johan.versendaal}@hu.nl

Abstract. Nowadays, many organizations have adopted an agile way of working where agile teams are responsible for the architecture, design and implementation of transformations in business processes. To get some recent empirical data on how this influences the value of EA as perceived in organizations, a survey has been created based on the Enterprise Architecture Value Framework (EAVF), a model to categorize value items. The survey has been distributed among (enterprise) architects and stakeholders of EA. Only small differences were found between the answers of these groups and the overall picture is that the respondents find the contribution of EA (average) important. A more detailed exploration of the outcomes shows that in areas which have a long-standing tradition with EA such as compliance, risk prevention, data management and information systems, the contribution of EA is perceived as (very) important, while in areas such as sustainability, market strategy and technology research the contribution of EA is assessed as less important. The results also suggest that the maturity of the EA processes can be improved.

Keywords: Enterprise architecture · Enterprise architecture value · Enterprise architecture value framework · Enterprise architecture value survey

1 Introduction

Since the proliferation of agile practices in organizations, the discussion about the usefulness and value of Enterprise Architecture (EA), has revived [1, 2]. Some authors claim that EA has to adapt itself to new ways of working [3–5], while others state that members in agile teams should be able to think like an architect [6] or even that EA has outlived its usefulness [7]. The discussion about the value of EA originated around the turn of the century with the emergence of EA as a means to achieve better alignment between the business and the information technology function in an organization [8, 9]. In practice EA has many interpretations that result in major differences between the way

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 149–165, 2023.
https://doi.org/10.1007/978-3-031-26886-1_9

EA is organized and governed in organizations [10, 11], making the value of EA, in the words of Kaisler and Armour [12], an 'elusive question'.

In discussions about the value of EA, it is important to keep in mind that value is not restricted to financial value alone, but has many more dimensions [13]. EA can bring value in areas such as risk reduction, innovation capability, logistics management, compliance and many more. While some of these areas are measurable (given adequate accounting), many are not quantifiable [8]. Moreover, while the visible outcomes of EA are mainly documents, its real value lies in what is done with the artifacts created [14]. Because many different stakeholders are involved in the process leading to implementation, it is difficult to say to what extent success can be explained by EA alone [8]. To mitigate these limitations, we decided to ask (enterprise) architects and stakeholders of architecture how they assess the value of EA. While the results of such an approach are subjective, literature shows us that self-assessments are a reliable instrument [15] and can be quite useful in practice where architects get feedback on their efforts and may use the results to optimize the alignment of their activities to the strategy and goals of their organization. Moreover, as empirical data about the value of EA are scarce [9, 16], the results may provide a detailed insight in the current state of EA, especially in an agile world. Motivated by the need for empirical data, this paper addresses the research question: "*Where can the most important contribution to the value of enterprise architecture be found, according to architects and stakeholders of enterprise architecture*"?

To answer the research question, we created a survey consisting of 62 questions about the perceived value of EA, complemented with 10 questions about the background of the respondents. The questions about the perceived value of EA are based on our previously published Enterprise Architecture Value Framework (EAVF) [17, 18], in which benefits and costs of EA are classified along two axes: organizational goal and architectural activities. We discuss this model in the next section, followed by a short overview of relevant literature. In Sect. 4, the research method including the construction of the questionnaire is explained and Sect. 5 is dedicated to the outcomes of the survey. We end the paper with a discussion of the results.

2 The Enterprise Architecture Value Framework

After a structured literature research, Boucharas et al. [19] conclude that EA value concepts such as goal, benefit and cost are not defined in most research papers and that the way in which EA benefit categories are derived, lacks transparency. This makes it almost impossible to compare studies into the value of EA or to develop a common set of metrics [19–23]. Hence, we started our earlier research into the value of EA [17] with definitions of the basic concepts of EA value, based upon definitions of these concepts in business literature, particularly the definitions as given by Renkema and Berghout [13]. For example, an EA benefit/cost is defined as "*The positive/negative contribution from (one or more) EA activities towards the desired state of affairs for an organization as stated by some goal of that organization*" where an EA activity is defined as: "*The work that a company or organization performs to create a certain output that is carried out by the EA function of the organization*".

From this definition we concluded that EA benefits and EA costs can be classified by organizational goal and EA activity. Peppard and Ward [24] argue that organizational goals and performance measures are often made explicit by means of a balanced scorecard (BSC) analysis [25], so we decided to use the four goal-perspectives of the BSC to classify organizational goals. This decision is supported by the fact that the BSC is widely used in practice [26] and by the research of Boucharas et al. [27] who have assessed several frameworks for classifying organizational goals and found the BSC the most suitable in the context of EA value.

To classify EA activities, we used the three organizational processes to which EA activities according to Ahleman and El Arbi [28] are closely related*:* strategic planning in which the EA is *developed*, the project life cycle in which the EA is *implemented* and operations and monitoring in which EA *exploitation* activities take place. Based on these classifications we have created a two-dimensional framework to classify EA benefits and costs: the Enterprise Architecture Value Framework (EAVF) as depicted in Fig. 1.

Organizational goal / EA Activity	Financial and accountability	Customer and partnerships	Internal processes	Learning and growth
EA Development				
EA Implementation				
EA Exploitation				

Fig. 1. The enterprise architecture value framework [17]

A more substantial discussion on the EAVF and its background can be found in a previous published paper [17]. In that study it is also shown that the EAVF complies with the necessary conditions for a taxonomy as formulated by Nickerson, Varshney and Muntermann [29] and that it can be used as a reference model for other classifications of EA benefits as well. An important aspect of the EAVF is that it is based on the outcomes of EA activities and not on the way these activities are carried out, making the EAVF independent of how the EA function is organized or which methods and tools architects use.

In a follow-up study [18], the four categories of organizational goals were subdivided in 31 goal-subcategories where a contribution of EA may be expected. In Table 1 these goal-subcategories are summarized (in this table ‘Costs’ should be read as: goals concerning costs’, etcetera). Definitions of the various goal-subcategories can be found in [18], including their validation by a panel of 13 (Dutch) EA experts in a Delphi study. While, according to these experts, EA may contribute to all goal-subcategories, it remains unsure whether they are complete.

Table 1. Abbreviations of the goal-subcategories in the EAVF [18]

Financial and accountability	Customer and partnerships	Internal processes	Learning and growth
Costs Revenues Investments Compliance Governance Risk management Societal responsibility	(Customer) experience (Customer) relationships Product position Market strategy Ecosystem	Logistics Procurement Business processes Marketing and sales Service delivery Data management Information mgt Technology (nonIT) General mgt Quality mgt HRM Innovation	Competences Culture Communication and knowl. mgt Alignment Agility Technology research Evaluation and re-use

3 Related Work

As stated in the introduction, there are many interpretations of what EA is (or should be), both in practice as well as in the literature. Interesting overviews can be found in [30, 31]. In this study, building on the definitions given in [31], we view EA as '*a discipline that directs enterprise transformations*', which implicates that we are effectively accepting a very broad range of interpretations of the concept of EA. This is in line with the diversity of EA implementations in practice [11].

When studying the value of EA, we have to take into account both the benefits of EA and the costs of EA [32, 33]. Papers on the benefits of EA are numerous, including a number of meta-studies. However, papers on the costs of EA are almost non-existent. If costs are mentioned, it is in the context of cost reductions by the implementation of EA [34–36], which we consider a benefit of EA.

As a complete overview of all that is written about EA benefits is out of scope for this paper, we will limit ourselves to meta-studies about the topic. One of the first papers giving an overview of existing literature on the subject of EA benefits is by Niemi in 2008 [8]. In an extensive literature study, he identified 27 classes of EA benefits, which are validated by a focus group. Next, Niemi uses the IS classification model of Giaglis, Mylonopoulos and Doukidis [37] to classify these benefits, resulting in 4 classes of EA benefits. Somewhat later, Boucharas et al. [19, 27] conducted a systematic literature review and they identified 100 mutually exclusive benefits which they classified in the strategy map [38] – an extension of the balanced scorecard [25]. In 2011 Tamm et al. [39] counted 213 benefits in a systematic literature review which they classify in 12 different types of EA benefits, but they do not explain how these categories were developed. More recently, in 2017, Yusuf and Kurnia [40] identified 40 different types of EA benefits which they classify in 5 categories, based on the benefit framework for enterprise systems of Shang and Sheddon [41]. In 2019, Niemi and Pekkola [42] discerned 250 EA benefits which they – without further explanation – classify in 40 types. In the same year, Gong and Janssen [9], based on a structured literature research, discern 9 different categories of EA benefits, without explaining where this classification is based upon.

The meta-studies above are grounded in literature research. However, only a small number of authors have used a survey as a means to get empirical data on EA value. Shanks et al. [16] in 2018 found eight publications where a survey was used as the research method. Only the survey conducted by Foorthuis et al. in 2010 [34] and an earlier survey by us conducted in 2014 [43] are aimed at gathering data about the perceived value of EA. The other studies focus primarily on how EA benefits are achieved, rather than on the benefits themselves.

4 Research Design

As shown in the research of Shanks et al. [16], older literature on EA value/EA benefits is mainly conceptual in nature. Empirical studies about how the value of EA is perceived in organizations are still scarce [16] and we did not find any recent empirical studies even though in the last decade a proliferation of agile implementation methods has occurred [1]. In order to get an overall picture of the current perception about the value of EA and at the same time get an impression of the adaptation of EA to agile implementation methods, we decided to use a survey as our research method. Based on the EAVF, we decided to discern three target groups:

- *EA Developers*: architects who create, adapt and maintain (parts of) the enterprise architecture such as enterprise architects, domain architects, business architects and information architects.
- *EA Implementers*: architects and non-architects who are accountable for the implementation of parts of the enterprise architecture, usually in projects. Examples are solution architects, system architects, program- and project managers.
- *EA Users*: non-architects who in their line of work are confronted with the results of enterprise architecture, such as business line managers, staff and project owners.

As there may exist some overlap between the three groups, in the survey we let respondents decide for themselves whether they are developing, implementing or using EA (or none of these) and in this way choose their viewpoint towards EA.

The survey for each of these groups consists of two parts: the first part contains general questions about the background of the respondents (this part is the same for all respondents) while the second part implements the questions about the value of EA. As we wanted the questionnaire to be based on value items as reported in literature, we started the construction of the second part of the survey with an inventory of EA benefits, using studies as mentioned in the previous section. In this way we gathered 112 specific EA benefits, which were categorized in the 31 goal-subcategories as depicted in Table 1. In order to keep the survey comprehensive, in the overloaded subcategories we combined various benefits into one value item. For example, benefits such as ‘reduce costs in general’, ‘reduce specific costs like IS/IT costs’ and ‘reduce administrative costs’, were combined in one item ‘lower operational costs.’ This resulted in 62 unique value items where each item is a statement about the contribution of EA, for example 'lower operational costs c.q. higher revenues', 'the (expected) effects on customer experience and customer satisfaction' or 'the willingness and ability to cooperate in the organization'.

Next, for each item we established the relevance for the three groups of potential respondents. For example, an item about the manageability of projects is relevant for EA implementers, but not for EA developers. The value items and their distribution over the three groups can be found with the outcomes of the questions (see Sect. 5).

The value items can be scored on a 5-point scale ranging from 'not important at all' (score 1) to 'very important' (score 5), supplemented with the option 'don't know' for items where the respondents are not aware of the value delivered by EA to that item. For each of the three groups of respondents, a base question was formulated asking for the perceived value of EA on the value items. In the version for the EA developers this base question is: 'Please state with a score from 1 to 5 the importance of the next items in developing and updating the overall architecture'. In the version for the EA implementers this is replaced by: 'Please state with a score from 1 to 5 the importance of the next items in preparing solution/systems architectures during implementation processes' and in the version for EA users by: 'Please state with a score from 1 to 5 the importance of the contribution of architecture with respect to the next items'. The items were ordered within the four goal-perspectives of the EAVF and to each set of questions an open question was added asking for the completeness of the questions in the eyes of respondents. To prevent a systematic bias from weariness, the four sets of questions were presented to the respondents in random order. After all questions were answered, the questionnaire ended with the calculated average scores on the four goal-perspectives from the EAVF and respondents could comment on this feedback. Finally, the questions were made ready for distribution using the online tool Limesurvey.

The survey was tested by 2 persons and based on their remarks, 'contribution' was used instead of 'value' as to the testers, 'value' was too strongly associated with financial value alone.

After completion, the survey was accessible for a period of two months in the spring of 2021. In this period potential respondents were approached via different channels such as the 'Nederlands Architectuur Forum NAF', a community of practice for architects, the research groups 'Digital Ethics' and 'Process Innovation and Information Systems' of the University of Applied Sciences Utrecht as well as via colleagues of the authors. Furthermore, it was brought to attention in LinkedIn groups on architecture.

The survey was conducted anonymously, but after completion of the survey respondents were given the possibility to receive the analysis and conclusions. To guarantee anonymity, the email address of the respondents was stored separately from the survey data.

5 Results

In the period the survey was accessible, 256 people opened the link to the survey but only 136 of these started with the questionnaire. This resulted in 105 full responses from which 7 indicated that they were not in any way involved with architecture. All questions and their outcomes can be found on http://eduples.nl/index.php/results2021.

The statistical analysis of the data has been done with the statistical package SPSS, version 28.

5.1 Characteristics of the Organizations of the Respondents

Almost all economic sectors were present in these responses, with an emphasis on the governmental sector (Table 2). Compared to other surveys on EA value in the Netherlands [34, 43], we see more respondents from the industrial sector, but less in the financial and insurance sector. As over the past decade the financial and insurance sector in the Netherlands has diminished considerably, we assume the distribution to be representative.

Table 2. Distribution over economic sector

The organization I work for can be classified in the following economic sector	This survey	[34] (2010)	[43] (2014)
No answer	0%	0%	0%
Agriculture, fishing, forestry and mining	0%	1%	2%
Industry (nutrition and manufacturing) and construction	13%	6%	3%
Energy, water and waste production/processing	4%	5%	5%
Education and research	7%	2%	6%
Health and community work	11%	3%	5%
Government (including Defense)	28%	31%	24%
Financial and insurance services	14%	30%	35%
Information, communication, entertainment/recreation	7%	12%	6%
Trade, transport and other services	15%	10%	13%

In line with other research [34, 43], in terms of size larger organizations are in the majority (Table 3), which is to be expected as smaller organizations usually do not employ architects.

Table 3 also shows the distribution of the number of architects over architectural task areas. As we expect the number of architects to grow with the size of the organization, we performed a correlation test on the variables 'organizational size' against 'number of architects'. In correspondence with the ordinal character of the variables we used Spearman's rank correlation and found moderate positive correlations: $\rho_s = 0.435$ for organizational size vs. number of enterprise/domain architects and $\rho_s = 0.524$ for organizational size vs. number of solution/systems architects; both with $p < 0.001$.

Almost half of the organizations (49%) have more than 10 years experience with architecture, but still 7% of the respondents state that the organization they work for has less than 1 year of experience with architecture. Most respondents (60%) have ample (over 6 years) of experience in their current function but we found no significant correlation with the architectural experience of the organization. Finally, in about 10% of the organizations the focus of architecture is on business and information only, while in one-third the focus is on application and infrastructure. In the remaining half of

Table 3. Distribution over organizational size and architectural task area

How many employees are there in the organization you work for?		How many architects does the organization you work for employ?	Enterprise/domain architects	Solution/system architects
Don't know/No answer	0%	Don't know/No answer	4%	7%
Less than 10	2%	0	4%	11%
10 to 100	6%	1	15%	9%
101 to 500	14%	2 to 5	22%	16%
501 to 2000	27%	6 to 10	22%	11%
More than 2000	50%	11 to 20	14%	15%
		More than 20	19%	31%

the organizations, the respondents indicate an equal focus on business/information and application/infrastructure architecture. These results are comparable to those in [43].

Of the 105 respondents, 56 (53%) indicated they are EA developer, 27 (26%) as EA implementer and 15 (14%) as an EA user. The remaining 7 respondents found they have no or insufficient experience with architecture. Their responses will not be used in the next sections as they did not answer the questions about the contribution of EA. The number of respondents is relatively low, especially in the group of EA users, but when taking the three groups together we have in our outcomes – with a confidence level of 95% – a margin of error less than 10% (using the sample size calculator of Surveymonkey on https://www.surveymonkey.com/mp/sample-size-calculator/).

5.2 Perceived Contribution of EA

In the survey, the questions about the contribution of EA are divided over the four goal-perspectives of the BSC. In Table 4 the averaged results in these goal-perspectives, categorized by group of respondents, are given, together with the percentage of the respondents who found the contribution of EA important to very important (score 4 or 5).

As the differences between the values in Table 4 are very small, not much can be concluded from these results – except the fact that in all four goal-perspectives and for all three groups the contribution of EA to the organizations of the respondents is considered between average important and important. Although the averages of the EA developers are marginally higher than those of the other two groups, the differences are very small and due to the relatively small number of respondents, no hard conclusions can be drawn from the outcomes.

More can be learnt by looking at the answers to the individual questions – especially the outstanding ones, the positive (average score ≥ 4, important to very important) as well as the negative results (average score < 3, less than average important). These results are shown in Tables 5 and 6 and they give a good picture of the items where the

Table 4. Averaged results and positive percentages by group and goal-perspective

Category	Number	Finance and accountability	Customer and partnerships	Internal processes	Learning and growth
EA developers	56	3.6/56%	3.4/54%	3.6/59%	3.6/59%
EA implementers	27	3.6/58%	3.4/53%	3.5/56%	3.3/49%
EA users	15	3.4/50%	3.3/53%	3.5/54%	3.4/52%
All	98	3.5/55%	3.4/53%	3.6/57%	3.5/55%

contribution of EA to the goals of the organization is perceived high (Table 5) or low (Table 6).

Table 5. Outstanding positive scores

Items with average score ≥ 4	Dev N = 56	Imp N = 27	Exp N = 15	All N = 98
Finance and accountability				
Compliance with laws, regulations and internal standards	4.4	4.7	(3.9)	4.4
Prevention of risks in business and information processes	(3.9)	4.2	(3.6)	(3.9)
Customer and partnerships				
The exchangeability of data with partners	(3.9)	(3.8)	4.1	(3.9)
Internal processes				
Digitization of business processes	4.1	(3.7)	4.2	4.0
The quality of stored data	4.3	(3.9)	(3.9)	4.1
The interoperability of data between information systems	4.3	4.2	(3.9)	4.2
The quality of information systems and IT infrastructure	(3.7)	4.4	(3.8)	(3.9)
The security of information, systems and infrastructure	4.4	4.6	(3.5)	4.3
‘Outsourcing’ and ‘cloud’	(3.9)	4.1	(3.4)	(3.9)
The involvement of stakeholders	4.0	(3.6)	(3.4)	(3.8)
Learning and growth				
Insight into current and desired situation and the road map	4.2	4.1	(3.7)	4.1

The outstanding positive scores (Table 5) show that the contribution of EA to the goals of the organization are found in particular with value items that are linked to information management and with compliance, risk prevention and providing insight in planned developments; areas that have a long-standing tradition with enterprise architects and can already be found in older meta-studies on EA benefits [8, 27, 39]. On the other hand, the outstanding negative scores (Table 6) where the contribution of EA to organizational goals is perceived low, are concentrated in more recent areas of interest to architecture such as societal responsibility, markets and market strategy, organizational culture, (agile) project management and technology research. However, value items from evaluation and re-use are found here as well, which may indicate that the maturity of the EA processes can be improved.

Table 6. Outstanding negative scores

Items with average score < 3	Dev N = 56	Imp N = 27	Exp N = 15	All N = 98
Finance and accountability				
Sustainability	(3.1)	2.7	(3.0)	(3.0)
Decent working conditions (internally and with partners)	(3.2)	2.7	(3.5)	(3.1)
Customer and partnerships				
The expected effects on markets and market shares	2.5	2.4	2.3	2.4
Alignment with the market strategy of the organization	(3.4)	(3.0)	2.7	(3.2)
Supply chain integration	2.9	(3.5)	(3.6)	(3.2)
Internal processes				
The support of business processes with logistics software	2.8	2.8	(3.4)	2.9
The “time-to-market” of new products and services	(3.1)	2.9	(3.1)	(3.0)
The use of customer journeys in modelling	–	2.8	–	–
Support with ‘agile’ project implementation	(3.3)	(3.3)	2.9	(3.2)
Learning and growth				
The professionalization of project management	–	2.9	(3.4)	–
The culture in the organization	–	–	2.9	–
Research of and gaining experience with new technology	(3.2)	2.8	(3.4)	(3.1)
Experiences with previous results of architecture	2.8	2.9	–	–
Evaluations of project results	–	2.7	2.9	–
The creation of artifacts for reuse	(3.1)	2.8	–	–

In both tables we observe a reasonable agreement between the three groups. To research if a consensus between the three groups is supported and can be found for all value items, we performed a Spearman’s correlation test. We found moderate positive correlations between the three groups: $\rho_S = 0.689$ for EA developers vs. EA implementers, $\rho_S = 0.538$ for EA developers vs. EA users and $\rho_S = 0.487$ for EA implementers vs. EA users – all with $p < 0.001$, so a moderate degree of agreement between the three groups may be assumed.

Another way to look at the scores is to classify the responses in the goal-subcategories of the EAVF (Table 1). To test whether the outcomes are not the result of coincidence, we performed a one-sided binomial test on these goal-subcategories. We divided the responses by goal-subcategory in two sets: the first set being the responses corresponding with a positive perceived contribution of EA (responses 4 and 5) and the second set where no positive contribution was perceived (responses 1, 2 and 3). We then tested the hypothesis: no positive effect of EA is perceived versus the alternative hypothesis: a positive effect of EA is perceived by the respondents.

The hypothesis is accepted when in the first set (responses 4 and 5) the percentage of responses is not significantly more than 40%. The alternative hypothesis is accepted if significantly more than 40% can be found in this set.

Table 7. Distribution of answers over all respondents (legend below table)

Goal subcategories	≤3 #	>3 #	0 #	≤3 %	>3 %	Sig	<3 %	<3 #
Financial and accountability								
Costs and revenues	38	58	2	39.6	60.4	<0.001	24.0	23
Investments	37	55	6	40.2	59.8	<0.001	16.3	15
Compliance	16	80	2	16.7	83.3	<0.001	5.2	5
Governance	34	61	3	35.8	64.2	<0.001	23.2	22
Risk management	35	62	1	36.1	63.9	<0.001	8.2	8
Societal responsiblity	46	42	10	52.3	47.7	0.086	34.1	30
Customer and partnerships								
Customer experience	35	57	6	38.0	62.0	<0.001	18.5	17
Customer relationships	40	53	5	43.0	57.0	<0.001	19.4	18
Product position	68	21	9	76.4	23.6	>0.999	52.8	47
Market strategy	48	46	4	51.1	48.9	0.049	37.2	35
Ecosystem	29	66	3	30.5	69.5	<0.001	25.3	24
Internal processes								
Logisitics	43	52	3	45.3	54.7	0.003	31.6	30
Business processes	18	80	0	18.4	81.6	<0.001	12.2	12
Marketing and sales	63	33	2	65.6	34.4	0.891	34.4	33
Service delivery	39	58	1	40.2	59.8	<0.001	26.8	26
Data management	19	77	2	19.8	80.2	<0.001	13.5	13
Information mgt	15	82	1	15.5	84.5	<0.001	9.3	9
General management	42	54	2	43.8	56.3	<0.001	19.8	19
Quality management	32	64	2	33.3	66.7	<0.001	19.8	19
HRM	43	53	2	44.8	55.2	0.002	11.5	11
Innovation	41	55	2	42.7	57.3	<0.001	21.9	21

(*continued*)

Table 7. (*continued*)

Goal subcategories	≤3 #	>3 #	0 #	≤3 %	>3 %	Sig	<3 %	<3 #
Learning and growth								
Competences	33	63	2	34.4	65.6	<0.001	19.8	19
Culture	34	62	2	35.4	64.6	<0.001	20.8	20
Alignment	18	79	1	18.6	81.4	<0.001	11.3	11
Agility	27	71	0	27.6	72.4	<0.001	18.4	18
Technology research	63	34	1	64.9	35.1	0.864	24.7	24
Communication and KM	45	52	1	46.4	53.6	0.004	17.5	17
Evaluation and reuse	58	38	2	60.4	39.6	0.572	41.7	40

≤3 # – number respondents scoring 1, 2 or 3
>3 # – number respondents scoring 4 or 5
0 # – number respondents scoring 0
≤3% – perc. respondents scoring 1, 2 or 3
>3% – perc. of respondents scoring 4 or 5
Sig – significance
<3% – perc. respondents scoring 1 or 2
<3 # – number respondents scoring 1 or 2

Given the relatively low number of respondents, especially in the group EA users, we tested on the total population as there are moderate positive correlations between the three groups. In Table 7 the results are given. In the tests, responses 0 (unknown/no answer) were excluded.

For most value items, the tested hypothesis can be dismissed with certainty >95% in favor of the alternative hypothesis with the exception of the goal-subcategories: societal responsibility, product position, marketing and sales, technology research and evaluation and reuse. These results match with the outcomes as found with the outstanding negative scores (Table 6) as in these areas the scores given are generally low.

5.3 Perceived Value and the Characteristics of the Respondents

To determine if a relation exists between the outcomes on the questions about the perceived value of EA and the characteristics of the respondents and their organizations (as discussed in Sect. 5.1), we performed chi-square tests. In these tests, we combined again the three groups of respondents and tested against the four goal-perspectives. Given the relatively low number of responses, the number of 0's in the cells of SPSS crosstabs was in all cases above the threshold for a Pearson's chi-square test, so we used the Fisher-Freeman-Halton Exact Test instead. We found only three relations with $p < 0.05$ (Table 8) which could by all means be accidental.

The outcomes on the questions about the perceived value of EA seem independent of the characteristics of the respondents. Whether this also holds true for the three groups separately cannot be established due to the relatively low number of responses.

Table 8. Relations between the respondents' background and goal-perspective

Respondents' background	Goal-perspective	p (2-sided)
Economic sector	Internal processes	0.008
Number of employees	Financial and accountability	<0.001
Number solution architects	Learning and growth	0.018

5.4 Open Questions

After each of the questions in the four goal-perspectives, an open question was added asking if there were any items missing that could be important in determining the value of the contribution of EA. A few suggestions were given: privacy, deprecating old-fashioned technologies, large projects with specific architectures, commitment of stakeholders and development of architecture as a competence throughout the organization. The current survey is based on benefits as found in literature, but these may change and some of these items could be added in a new version of the survey.

After the questions about the value items, feedback was given on the scores averaged by goal-perspective. Almost 20% of the respondents found these scores did not give a valid and reliable view on the contribution of EA. An interesting remark made here is: '*the average is not interesting, the differences are'*. While this may be true within an organization, in a survey like this we see in many items the full range of possible answers (with the exception of items where the average score is quite high or low), so we choose to show these outstanding high and low scores (Tables 5 and 6) instead.

Other examples of remarks made here are: '*I think having an architect is a bit of old school*', '*the summary above reflects how we value and approach architecture, but not necessarily the priorities*' and '*EA in my organization is an ivory tower*, *out of touch with customers and stakeholders, only concerned with their own bureaucracy and artifacts, self-serving'*. Comments like these suggest a lack of communication between the enterprise architects and the rest of the organization resulting in low scores as well.

6 Discussion and Conclusions

In this paper we have presented the outcomes of a survey concerning the contribution of EA to organizations. When relating these outcomes to the results of other surveys, it should be noted that the questions used in [34, 43] and this survey are not the same, so detailed conclusions cannot be drawn. However, Boucharas et al. [27] in 2010 found no benefits in the customer perspective of the balanced scorecard but in our earlier survey in 2014 [43] and in this survey we see an increasing contribution of EA to value items concerning the customer in this goal-perspective. This could be an indication that the focus of EA has shifted from the internal workings of the organization alone to include the organization's environment as well. The same could happen in the future with areas such as 'societal responsibility', '(organizational) culture' and 'technology research'. This would be in line with the trends identified in recent EA publications in [44]. However, the low scores in the goal-subcategory 'evaluation and re-use' cannot be explained by the available data but could point to a low maturity of the EA processes; in the words of Robertson et al. [45]: "*an EA programme exists but it is executed without complete structure and accountability*".

Foorthuis et al. [34] found that EA creators (EA developers and EA implementors) were more positive about the contribution of EA than EA users. In contrast – and in line with our earlier results [43], we found only small differences between these groups, but a clear conclusion in this aspect cannot be drawn as the number of EA users in this survey is small. Also, we did not find convincing relations between the characteristics of the respondents and the outcomes as categorized in the four goal-perspectives of the balanced scorecard (Table 8), which suggests a commonly accepted view on what may be expected of EA, independent of the organizations the respondents work for.

As empirical data on EA value are scarce [16], this research contributes to the scientific community by providing empirical data about the value of EA, as perceived by architects and stakeholders of architecture. Based on these data, insights about the value items that currently score high and those that score low, is gained. Value items where the contribution of EA to organizations scores high (Table 5) may be characterized as belonging to areas that have a long-standing tradition within EA such as compliance, risk prevention, providing insight and information management. It seems that architects put much effort in these areas. The value items where in the eyes of the respondents the contribution of EA is below average such as societal responsibility, markets, culture, project management and technology research cover in majority areas that are more recently recognized as potential areas of interest to EA. These areas may become more important with time as discussed above.

In practice, organizations can use the questionnaires to assess the contribution of EA as perceived by their architects and stakeholders of architecture. Using the questionnaires in this way gives an organization the opportunity to prioritize some items and/or to add extra questions about aspects that are of interest to that organization. In upcoming research, we have elaborated this line of thought by developing an instrument to assess the contribution of EA in organizations. We are testing this instrument in case studies, where more in-depth qualitative research may give insight into the 'why' of the answers.

Overall, the results give a picture of the current state of EA in the Netherlands: in areas that have a long-standing tradition with EA the contribution of EA is perceived as (very) important, while in areas that have more recently come into the focus of EA the contribution of EA is assessed as less important. We also found indications that the maturity of the EA processes can be improved.

This research has its limitations. First of all, as already stated above, the relatively low number of respondents is responsible for a relatively high margin of error and makes it impossible to say anything reliable about the group of EA users. Secondly, the respondents to our survey are self-selected and as such are not necessary a random sample of those working in or with EA. As a consequence, some bias in the answers may be present, moreover so as the questions ask for the *perceived* value of EA. Finally, as the survey involves only respondents from the Netherlands, care must be taken in generalizing the results.

References

1. Canat, M., Català, N.P., Jourkovski, A., Petrov, S., Wellme, M., Lagerström, R.: Enterprise architecture and agile development: Friends or foes? In: 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW), pp. 176–183. IEEE (2018)
2. Hylving, L., Bygstad, B.: Nuanced responses to Enterprise architecture management: loyalty, voice, and exit. J. Manage. Inf. Syst. **36**(1), 14–36 (2019)
3. Duijs, R., Ravesteyn, P., van Steenbergen, M.: Adaptation of enterprise architecture efforts to an agile environment. In: Bled 2018 Proceedings, 16 (2018)
4. Daoudi, W., Doumi, K., Kjiri, L.: Adaptive enterprise architecture: towards a model. In: Proceedings of the 10th International Conference on Information Systems and Technologies, pp. 1–7 (2020)
5. Kotusev, S.: Enterprise architecture: forget systems thinking, improve communication. J. Enterp. Architect. **1**, 12–20 (2020)
6. Horlach, B., Drechsler, A., Schirmer, I., Drews, P.: Everyone's going to be an architect: design principles for architectural thinking in agile organizations. In: HICSS, pp. 1–10 (2020)
7. McLeod, J.: Enterprise Architecture is Dead (2017). https://jonmcleodea.medium.com/enterprise-architecture-is-dead-33dd0e63cbbf. Retrieved 13 Jan 2022
8. Niemi, E.: Enterprise architecture benefits: perceptions from literature and practice. In: Proceedings of the 7th IBIMA Conference Internet and Information Systems in the Digital Age, 14–16 December, 2006, Brescia, Italy, pp. 1–8 (2008)
9. Gong, Y., Janssen, M.: The value of and myths about enterprise architecture. Int. J. Inf. Manage. **46**, 1–9 (2019)
10. Ross, J., Weill, P., Robertson, D.: Enterprise Architecture as Strategy: Creating a Foundation for Business Execution. Harvard Business Press (2006)
11. Ansyori, R., Qodarsih, N., Soewito, B.: A systematic literature review: critical success factors to implement enterprise architecture. Procedia Comput. Sci. **135**, 43–51 (2018)
12. Kaisler, S., Armour, F.: 15 years of enterprise architecting at HICSS: revisiting the critical problems. In: Proceedings of the 50th Hawaii International Conference on System Sciences (2017)
13. Renkema, T., Berghout, E.: Methodologies for information systems investment evaluation at the proposal stage: a comparative review. Inf. Softw. Technol. **39**(1), 1–13 (1997)
14. Kotusev, S.: Enterprise architecture and enterprise architecture artifacts: questioning the old concept in light of new findings. J. Inf. Technol. **34**(2), 102–128 (2019)
15. Ramos-Villagrasa, P.J., Barrada, J.R., Fernández-del-Río, E., Koopmans, L.: Assessing job performance using brief self-report scales: the case of the individual work performance questionnaire. Revista de Psicología del Trabajo y de las Organizaciones **35**(3), 195–205 (2019). https://doi.org/10.5093/jwop2019a21
16. Shanks, G., Gloet, M., Someh, I.A., Frampton, K., Tamm, T.: Achieving benefits with enterprise architecture. J. Strat. Inf. Syst. **27**(2), 139–156 (2018)
17. Plessius, H., van Steenbergen, M., Slot, R., Versendaal, J.: The enterprise architecture value framework. In: Proceedings of the Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK (2018)
18. Plessius H., van Steenbergen, M.: A study into the classification of enterprise architecture benefits. In: MCIS 2019 Proceedings, vol. 33 (2019)
19. Boucharas, V., van Steenbergen, M., Jansen, S., Brinkkemper, S.: The contribution of enterprise architecture to the achievement of organizational goals: a review of the evidence. In: Trends in Enterprise Architecture Research, pp. 1–15. Springer, Berlin-Heidelberg (2010)
20. Schelp, J., Stutz, M.: A balanced scorecard approach to measure the value of enterprise architecture. J. Enterp. Architect. **3**(4), 8–14 (2007)

21. Lange, M., Mendling, J.: An experts' goal-perspective on enterprise architecture goals, framework adoption and benefit assessment. In: Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2011 15th IEEE International, pp. 304–313 (2011)
22. Niemi, E.I., Pekkola, S.: Enterprise architecture benefit realization: review of the models and a case study of a public organization. ACM SIGMIS Database **47**(3), 55–80 (2016)
23. Kurek, E., Johnson, J., Mulder, H.: Measuring the value of enterprise architecture on IT-projects with CHAOS research. J. Systemat. Cybern. Inform. **15**(7), 13–18 (2017)
24. Peppard, J., Ward, J.: The strategic management of information systems: Building a digital strategy, pp. 146–147. John Wiley & Sons (2016)
25. Kaplan, R., Norton, D.: The Balanced Scorecard - Measures that Drive Performance. Harv. Bus. Rev. **1992**, 71–79 (1992)
26. Hasan, R., Chyi, T.: Practical application of balanced scorecard - a literature review. J. Strategy Perform. Manage. **5**(3), 87–103 (2017)
27. Boucharas, V., van Steenbergen, M., Jansen, S., Brinkkemper, S.: The contribution of enterprise architecture to the achievement of organizational goals: establishing the enterprise architecture benefits framework. Technical Report UU-CS-2010-014, Department of Information and Computing Sciences, Utrecht University (2010)
28. Ahlemann, F., El Arbi, F.: An EAM navigator. In: Ahlemann, F., Stettiner, E., Messerschmidt, M., Legner, C. (eds.) Strategic Enterprise Architecture Management: Challenges, Best Practices, and Future Developments, pp. 35–53. Springer, Berlin (2012)
29. Nickerson, R., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. Eur. J. Inf. Syst. **22**, 336–359 (2013)
30. Bean, S.: Re-thinking enterprise architecture using systems and complexity approaches. J. Enterp. Architect. **6**(4), 7–13 (2010)
31. Saint-Louis, P., Morency, M.C., Lapalme, J.: Examination of explicit definitions of enterprise architecture. Int. J. Eng. Bus. Manage. **11**, 184797901986633 (2019). https://doi.org/10.1177/1847979019866337
32. Berghout, E., Nijland, M., Powell, P.: Management of lifecycle costs and benefits: lessons from information systems practice. Comput. Ind. **62**(7), 755–764 (2011)
33. Rodrigues, L., Amaral, L.: Issues in enterprise architecture value. J. Enterp. Architect. **6**(4), 27–32 (2010)
34. Foorthuis, R., van Steenbergen, M., Mushkudiani, N., Bruls, W., Brinkkemper, S., Bos, R.: On course, but not there yet: enterprise architecture conformance and benefits in systems development. In: ICIS 2010 Proceedings, pp. 1–19 (2010)
35. Poort, E.R., van Vliet, H.: Architecting as a risk-and cost management discipline. In: 2011 Ninth Working IEEE/IFIP Conference on Software Architecture, pp. 2–11 (2011)
36. Miguens, J., da Silva, M.M., Guerreiro, S.: A viewpoint for representing costs in enterprise architectures. In: 2018 IEEE 20th Conference on Business Informatics (CBI), vol. 1, pp. 10–19 (2018)
37. Giaglis, G., Mylonopoulos, N., Doukidis, G.: The ISSUE methodology for quantifying benefits from information systems. Logist. Inf. Manag. **12**(1/2), 50–62 (1999)
38. Kaplan, R., Norton, D.: Transforming the balanced scorecard from performance measurement to strategic management: Part I. Account. Horiz. **15**(1), 87–104 (2001)
39. Tamm, T., Seddon, P., Shanks, G., Reynolds, P.: How does enterprise architecture add value to organisations. Commun. Assoc. Inf. Syst. **28**(1), 141–168 (2011)
40. Jusuf, M., Kurnia, S.: Understanding the benefits and success factors of enterprise architecture. In: Proceedings of the 50th Hawaii International Conference on System Sciences, Hawaii, pp. 4887–4896 (2017)
41. Shang, S., Seddon, P.: Assessing and managing the benefits of enterprise systems: the business manager's goal-perspective. Inf. Syst. J. **12**(4), 271–299 (2002)

42. Niemi, E., Pekkola, S.: The benefits of enterprise architecture in organizational transformation. In: Business Information Systems Engineering, pp. 1–13 (2019)
43. Plessius, H., van Steenbergen, M., Slot R.: Perceived benefits from enterprise architecture. In: Eighth Mediterranean Conference on Information Systems, Verona, pp. 1–14 (2014)
44. Gampfer, F., Jürgens, A., Müller, M., Buchkremer, R.: Past, current and future trends in enterprise architecture – a view beyond the horizon. Comput. Ind. **100**, 70–84 (2018)
45. Robertson, E., Peko, G., Sundaram, D.: Enterprise architecture maturity: a crucial link in business and IT alignment. In: PACIS 2018 Proceedings, 308 (2018)

Achieving Alignment by Means of EA Artifacts

Hong Guo[1](✉) and Shang Gao[2]

[1] Anhui University, No. 111 Jiulong Road, Hefei, People's Republic of China
homekuo@gmail.com

[2] School of Business, Örebro University, Örebro, Sweden
shang.gao@oru.se

Abstract. Enterprise Architecture (EA) was reported as powerful but also often failed to be applied. It is necessary to clarify by which means EA brings its value such as alignment. In this research, we explored how core EA deliverables, namely EA Artifacts (EAAs), helped organizations achieve alignment, by observing and analyzing recommended practices of a leading EA tool vendor. The results indicated that, 1) Among 43 EAAs primarily recommended by the vendor, most of them described "alignment" relationships among organizational elements, and in a quite explicit way; 2) Two types of "alignment" relationships were described in such EAAs; 3) EA users recognized such types of alignment relationships achieved by leveraging such EAAs. Application of the findings might help organizations more consciously define alignment goals by leveraging appropriate EAAs sets, and accordingly, further achieve business success with better EA applications.

Keywords: Alignment · Enterprise architecture · EA artifacts · EA tools

1 Introduction

One of the greatest values that Enterprise Architecture (EA) brings to an organization might be alignment [1]. To improve the less reliable and stable successful application of EA [2], it is necessary to understand more about the mechanism by which EA contributes to alignment. Traditionally, as the smallest deliverables of EA, EA Artifacts (EAAs) were less discussed than EA Frameworks (EAFs). However, more recent empirical studies suggested that EAFs tended to guide a holistic approach to architecture rather than comprehensive modeling. Comparatively, some EAAs were commonly used and have brought concrete values to organizations, even if they were only used separately. Therefore, we argue that EAA is the core deliverable for EA help organizations to achieve alignment. We tried to explore the means which helped EAA play this role. In addition, we wanted to investigate whether alignment relationships described by EAAs reflected organizational alignment goals for EA users or not. In this study, we aimed to address the following Research Questions (RQs).

- RQ 1: If and to what extent do EAAs describe "alignment" relationships?
- RQ 2: What types of "alignment" do EAAs describe?
- RQ 3: To what extent do EA users recognize "alignment" described by EAAs?

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 166–179, 2023.
https://doi.org/10.1007/978-3-031-26886-1_10

To answer these RQs, we observed and analyzed content about EAAs on the website of a leading EA tool vendor. The rest of this paper is structured as below. Section 2 introduces some background information about EA, EAA, and alignment. Section 3 describes the research method. In Sect. 4, the research results are presented. Then, the results are discussed in Sect. 5. Lastly, we conclude the paper in Sect. 6.

2 Background

2.1 EA and EAA

EA is generally defined as "the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution [3]". Here, an enterprise is viewed as a "system" [3].

EA is usually delivered as a set of abstract graphics covering the enterprise's high-level content across areas such as strategy, business, information, and technology. We call these abstractions EA artifacts or EA models as they are usually in graphical forms.

2.2 Alignment

Alignment in dictionaries is defined as "the act of aligning or state of being aligned [4]", "arrangement in a straight line or in correct relative positions [5]", and "the proper positioning or state of adjustment of parts (as of a mechanical or electronic device) in relation to each other [4]".

When alignment is used in organizational contexts, it is defined as "the continuous process, involving management and design sub-processes, of consciously and coherently interrelating all components of the business-IT relationship in order to contribute to the organization's performance over time [6]". Typically, organizational alignment is often referred to but not limited to Strategy Alignment [7, 8], Business-IT Alignment [6, 9, 10], and Partner Alignment [1]. Literally, the main difference among them is the **relationships** that are primarily considered about or strive to achieve. For instance, strategy alignment indicates that organizations' resources shall be appropriately arranged to realize strategies. Business-IT alignment cares about whether business and IT components have supported each other. To summarize, organizational alignment is some kind of expectation about relationships among organization components. It is about a status that needs to be achieved by a continuous process that consists of a series of activities.

EA helps organizations achieve alignment primarily by describing/designing important relationships among relevant elements [11]. By doing so, stakeholders are able to perceive better knowledge and take actions to approach expected alignment status. As indicated by Ross [12], who is a leading organizational EA theorist/ expert in MIT, EA is a management approach and means to design interrelationships between people, processes, and technology to get value.

3 Research Method

With the aim to shed some light on ways to better understand how EAAs help organizations achieve alignment, we explored how EAAs described organizations' alignment relationships (RQ1 and RQ2) and to what extent organizations embraced such EAAs (RQ3).

To do so, we collected evidence from publicly available content on official EA tool vendor websites. We primarily used the grey literature review approach [13] to collect the data and content analysis to analyze the data. Grey literature reviews have been acknowledged as a valid alternative to academic literature reviews, as they can give substantial benefits when the state of the practice is concerned [13].

3.1 Data Collection

The evidence mainly came from the website contents of leading EA tool vendors. There are three reasons for us to review such website contents. First, tools are instrumental and very important in EA discipline (Korhonen et al. 2016). Second, tools in general make it easier for users to accept one technology. For EA, user acceptance was perceived as one of the critical challenges. Thus, we assumed that tool support could facilitate EA application. Third, according to our preliminary observation, the content offered in the tool vendor websites was rich and informative. Many white papers, use cases, and feature descriptions were provided on the vendors' websites to provide knowledge to their potential customers and show vendors' expertise.

The vendors were selected according to the vendor list administered in Gartner's [14] annual report named "Gartner Magic Quadrant for Enterprise Architecture Tools" [15], which includes long-established manufacturers as well as insightful new challengers. According to the report, the EA tool market has been expanding rapidly. This was reflected in the growth of the included vendors whose revenues have grown between 15% and 30% in annual revenue since 2020. While in 2020, all included vendors have already demonstrated their performance, market focus, and market momentum. For example, the vendors have $8 million or more per year in EA tool perpetual licensing revenue or equivalent performance. They usually have around an installed base of at least 50 production customers. In addition, they have had 15 new customers in at least three of the four major geographic regions in 12 months [15].

It was not easy to extract evidence from multiple tool vendor websites, as the concepts and structures often differed. In fact, terminology misalignment in scientific papers is also a known issue [16]. Thus, we started by reading through the contents of the websites and gaining an initial understanding of the overall breadth and depth of the information and supporting evidence. Next, we extracted evidence relevant to the RQs. As similar evidence was presented on multiple websites, we chose the most representative formulations (e.g., clear and complete statements).

As a result, we collected data primarily from the website of one of the leading EA tool vendors, namely Capsifi [17]. Capsifi is one of the topping EA tool vendors and has been rapidly growing during the past three years according to the list. In addition, Capsifi presents quite comprehensive information about EAAs and how such EAAs are embraced by their customers. We believe that the way of how this leading vendor leverages EAAs represents the current trend of first-line EA applications. For RQ1 and RQ2, we analyzed description about "Key Features" in "Solutions" as these sections clarified what such EAAs mean and what benefits such EAAs can bring to an organization. For RQ3, we analyzed description about "Capsifi Products" in "Customer stories" as there were some hints about which EAAs were embraced by organizations. To complement the opinion and information declaimed by Capsifi, we also referred to corresponding information from the websites of other vendors such as Avolution [18], Ardoq [19], and LeanIX [20]. For instance, we read materials on such websites to get a better understanding of the meaning and purpose of some specific EAAs.

3.2 Data Analysis

According to the information from the website, there were 43 different "Features"/EAAs which could be used to support "Solutions" including Strategy, Transformation, Governance, Operations, Innovation, Architecture and Experience. **For RQ1,** we analyzed the textual description of the EAAs by focusing on two aspects of information, namely Intensity and Clarity. **Intensity** means to what extent describing "alignment" relationships constitute EAA's value, while **Clarity** indicates if alignment purpose is described in the EAA in an explicit way or not. We evaluated Intensity and Clarity at three levels as shown in Table 1. *Strong* represented that describing alignment relevant relationships was the primary goal of the EAA. *Medium* represented that describing alignment relationships was one part of the primary goal of the EAA. *Weak* represented that alignment might not or only contribute to some trivial part of the primary goal. We also evaluate Clarity with three levels. *Explicit* represented that alignment was clearly described with specific words like "alignment" and "aligned". *Medium* represented that alignment purpose was indicated by some hints. *Implicit* represented that alignment purpose could be deduced in an implicit manner according to the overall description.

Table 1. Criteria for alignment intensity and clarity.

Intensity		Clarity	
Strong (S)	Alignment as the primary goal of the EAA	Explicit (E)	Alignment clearly indicated with specific words
Medium (M)	Alignment as part of the primary goal of the EAA	Medium (M)	Alignment indicated with some hints
Weak (W)	Alignment as some kind of the goal of the EAA or unclear	Implicit(I)	Alignment indicated implicitly

For RQ2, we read the EAA descriptions and classified different types of "alignment" relationships that were described or pursued in different EAAs. We identified two main Categories of alignment namely "Align towards/with others" (*A2w*) and "Align/fit together" (*A2g*). Here, A2w represented that some element(s) should be aligned(adjusted) according to some other element(s). A2g, on the other hand, represented that some elements should be adjusted with each other according to some implicit criteria (e.g., to achieve the best performance together). These two categories also corresponded to the literature meaning of alignment in the dictionary: arrangement in a straight line (A2w) or in correct relative positions (A2g) [5]. We believe that distinguishing the two alignment types could help clarify the purpose of EA modeling. In this way, modelers could focus more on describing or designing specific categories of relationships among elements. The quality of EA models could be improved, and model-based communication and collaboration could also be promoted.

For RQ3, we analyzed 10 customer stories presented on the website. Among them, 15 "winning features" were highlighted. We identified three ways in which EAAs/features contributed to organizational alignment. Firstly, **Local EAAs** represented (usually traditional) artifacts that were primarily used to describe one type or similar types of organizational elements. Secondly, **Linking EAAs** represented artifacts that aligned elements about different aspects of an organization or at different abstraction levels. Thirdly, **EAA Sets** represented sets of artifacts that were used to align elements in a larger scope.

4 Results

In this section, we present relevant results to answer the proposed three RQs.

4.1 "Alignment" Intensity and Clarity

For RQ1, we investigated if and to what extent EAAs described alignment relationships. According to the different levels of Intensity and Clarity defined in Table 1, we mapped 43 EAAs as shown in Fig. 1 and Fig. 2. Thirty out of the total 43 EAAs set alignment as their primary goal. Only 3 EAAs did not set alignment as their direct goal. The alignment could be recognized quite easily. Moreover, 41 EAAs in total described their functions with specific words such as "align", "alignment", or clear hints such as "fit", and "support".

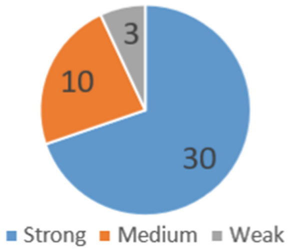


Fig. 1. Alignment intensity of EAAs.

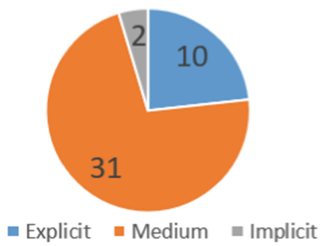


Fig. 2. Alignment clarity of EAAs.

According to our analysis, most EAAs pursued alignment as the main goal, thus were marked as "**Strong**" **Alignment Intensity**. For instance, EAA-002, namely *Strategic roadmaps*, focused on aligning "planned initiatives with desired business outcomes" (SA). Another instance was EAA-011, namely *Products and services*, which focused on aligning technology with "the information, processes, and business capabilities they support" (BITA). We also found that although some EAAs did not pursue alignment as the primary goal (**Weak Alignment Intensity**), they provided views from a specific perspective and therefore supported better collaboration within organizational elements to achieve overall a better performance. For example, EAA-042 named *Value proposition whiteboard* supported to "workshop customer needs & desired benefits" in a collaborative way.

Some EAAs described how organizational elements should be aligned in a quite explicit way (e.g., marked as "Strong" Alignment explicity). They used very typical terminologies such as "align", "enable", "deliver", and "support". Some EAAs were marked as "Medium" as they employed comparatively clear words such as "sequence", "prioritise", "consistent", and "shared". On the other hand, for some other EAAs, we inferred alignment meanings according to the overall description or more subtle hints. For instance, EAA-024, namely *Idea capture*, which aimed to "democratise innovation by capturing ideas from teams" aligns ideas among teams. And EAA-030 which shared "data models with external data sources" indicated that the EAA (i.e., *Schema import/export*) aligned external data sources according to data models.

4.2 Types of "Alignment"

For RQ2, the results showed that there were comparatively fewer EAAs of A2w types compared with those of the other two types. As shown in Fig. 3, 8 EAAs contributed to aligning some organizations' elements according to some other elements and were labelled as **A2w**. Such EAAs were like *Customer segments and personas* (EAA-018) which aligned "products and services" to "customer segments" and *Reference model mapping* (EAA-033) which aligned "pre-packaged reference models" to "industry standards". Moreover, 19 EAAs presented views to align multiple elements by adjusting relationships in between, according to some criteria, and to pursue overall better performance (**A2g**). Two examples are such as *User cases* (EA-001) which aligned business actors and systems according to the sequence in the business transaction execution, and *Organisation structure* (EA-007) which aligned "people and teams" to form specific (e.g., efficient) ways of working.

In addition, we found that there were 16 EAAs which present both "A2w" and "A2g" relationships in one view (as labelled **Mixed**). For instance, *Transformation planning* (EAA-003) aligned (by "budgeting, prioritizing and allocating") resources (A2g) to serve transformation initiatives, and such initiatives were aligned to strategies (A2w). Another instance was *Value streams* (EAA-012). It aligned resources, skills, technology and effort according to the flow of business transactions (A2g) and to deliver business outcomes (A2w).

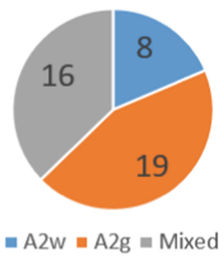


Fig. 3. Alignment types of EAAs.

4.3 Popularity of "Alignment"

In 10 customer stories, 20 features were mentioned 37 times. Some of these features were about EAAs which modeled one single type of organizational elements such as *Business requirements*, *Information*, *Process*, *Form*, *Interaction* and *Service*. They were labelled as **Local EAAs**. Some features were about EAAs which modeled multiple types of elements and link them together. We marked them as **Linking EAAs**. The samples were such as *Capability maps*, *Heatmap*, *Value stream*, and *Roadmaps*. In addition, some features were about using some EAAs at the same time as called as **EAA sets**. One sample was '*System* to *Application* and *Application'* to *'Business capability heatmaps'*. Another sample was '*Capability'* to *'Business operating model'*.

Another interesting finding was that "pain point **impacts**" and "risk points and **dependencies**" have drawn much attention by customers (5 times) although they have not been advocated in general EAA descriptions. In addition, **capability** relevant features have played an important role according to the customers' recognition. They were mentioned 12 times in total, 8 times in Linking EAAs, and 4 times in EAA sets. More information about the features recognized by users is presented in Fig. 4.

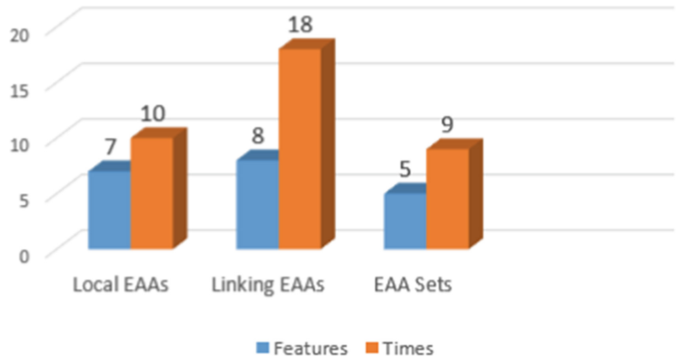


Fig. 4. Users recognized alignment by means of EAAs.

5 Discussion

The results showed that, the primary goal of most EAAs was alignment and was achieved in a fairly explicit manner. This not only showed that existing users might have intentionally used EAAs to achieve alignment, but also demonstrated the potential of EAAs to serve users' needs in a more deliberate way.

Traditional thoughts usually understand alignment, at a very high level, as adjusting organizational elements in one of two ways, either according to certain elements (e.g., SA) or in between (e.g., BITA). However, the research results indicated that alignment in practice might involve quite complex ways of relationship adjustment. Fortunately, EAAs could also do more than expected to help achieve such alignment goals. Some EAAs could be used to present two or even more types of alignments at the same time.

What is more, multiple EAAs have the potential to be used synergistically to serve broader alignment goals.

Last but not least, attitudes of users indicated that alignment brought by EAAs was quite recognized. Alignment was brought by not only local EAAs which traditionally modeled single or a few types of organizational elements, but also emerging EAAs (and even combinations of them) which linked artifacts of multiple types, at different levels, or across domains.

To summarize, this study offered some thoughts which might be different from traditional understanding on the value and usage of EAAs:

- Without having to be part of an EAF defined document, one single EAA could bring independent value.
- Alignment is (if not one of) the most important value that EAA can bring to organizations.
- The description of what to align and how to align in a single EAA has shown considerable diversity. Alignment can be conducted to some elements in connection to some other elements, or be conducted among a set of elements, or be the superposition of the two. This reflected the diversity and complexity of organizations' internal needs for alignment.
- Some traditional EAAs defined by EAFs might bring more immediate and recognized value to users than other traditional ones.
- Emerging EAAs could be used flexibly and within the context of the entire EA without being bound by EAFs. For example, a single EAA, such as a *Roadmap*, might describe organizational elements of multiple aspects and at multiple levels of abstraction. Another example, namely *Capability map*, might describe some abstract elements which do not belong to any level or aspect of organizations proposed by traditional EAFs.
- Artifacts could be used in sets. In this way, alignment scope of EAAs in a set was accumulated/ expanded, and the value of alignment could be further enhanced by analyzing and inferring such as dependencies and influence. Different from large and comprehensive traditional EAFs, such EAA sets could be quite small and flexible. More importantly, they were more purposefully constructed and could be independently measured about the value and cost.

6 Conclusion

This study observed and analyzed EAA relevant content on the website of a leading EA tool vendor and explores by which means EA helps organizations to achieve alignment. The results showed that the majority of 43 commonly used EAAs support organizations to align elements by describing relevant relationships among the elements. There were many types of alignment relationships with varying degrees of complexity. From the perspective of customers, in addition to basic alignment relationships for single and concrete types of elements, alignment relationships for abstract elements such as capabilities, values and other elements, as well as further combination between them, were all pursued. This indicated that EAAs did provide concrete and flexible mechanisms to support organizations to achieve alignment. By clarifying these mechanisms, it was believed that organization might be able to consciously define alignment goals and use appropriate mechanism (traditional/local EAAs, emerging linking EAAs, or EAA sets). They would further achieve business success with better EA applications. The limitation of this study was that the main information came from one website. In the future, we plan to further expand the dataset to include details from other vendors, tool users, and academic papers for better validation and improvement.

Appendix 1. Full List of 43 EAAs and Their Description

	EAA name	EAA description
EAA-01	Use cases	Create structured representations of the sequence of actions and steps between business actors and systems in the execution of common business transactions
EAA-02	Strategic roadmaps	Dynamic roadmap views that provide clear visibility of the **alignment** of planned initiatives with desired business outcomes
EAA-03	Transformation planning	Scenario analysis for budgeting, prioritising and allocating resources for strategically **aligned** transformation initiatives
EAA-04	Capability maps	Develop a consistent, shared view of business capabilities and their inter-dependencies across the operating model
EAA-05	KPI's and metrics	Define, manage, and monitor key measurements and performance indicators for tracking business performance
EAA-06	Heatmaps	Configurable visual cues for tracking diverse, user-defined business measures at a glance, such as performance, maturity, and risk across many of the standard Capsifi visualisations
EAA-07	Organisation structure	A dynamic map for organising business units and territories that **aligns** people and teams into hierarchical, matrix, regional or agile ways of working
EAA-08	Roles & responsibilities	Manage the categorisation and assignment of entitlements and responsibilities to individuals and teams in the roles that they perform in delivering business outcomes

(*continued*)

(*continued*)

	EAA name	EAA description
EAA-09	Skills and competencies	Catalogue the required skills and competencies of individuals and teams, to enable the *business capabilities* that they support
EAA-10	Products and services	An interactive catalogue of the products and services you offer, **aligned** to their features, benefits, and value propositions
EAA-11	Technology portfolio	A consolidated technology landscape of your enterprise systems, their functions, components, performance and **alignment** with the information, processes, and business capabilities they support
EAA-12	Value streams	Visually analyse an inside-out perspective on the stages in the flow of business transactions to **align** resources, skills, technology, and effort in the delivery of business outcomes
EAA-13	Value maps	A structured decomposition of the various dimensions of business value and the levers available to the organization for moving the needle on value delivery
EAA-14	Value trees	Decompose business value dimensions and optimise the allocation of costs and investment opportunities
EAA-15	Business motivation model	A framework for defining and communicating a strategic business plan, based on the BMM template of Desired Results, Courses of Action and Directives
EAA-16	Business requirements	Manage the specification and lifecycle of business needs as they guide the development of enhancements to business capabilities and/or technology systems
EAA-17	Business processes	Manage a catalogue of common business processes defining the sequenced flow logic for tasks and the related **alignment** of roles, information, and systems to support business transactions
EAA-18	Customer segments and personas	Leverage carefully considered customer personas to develop targeted messaging and the personalised delivery of products & services to distinctly defined customer segments

(*continued*)

(*continued*)

	EAA name	EAA description
EAA-19	Customer journey maps	An interactive canvas that tracks the outside-in measurements of the experience of customers on their interaction journeys as they engage with your business across the customer lifecycle
EAA-20	Digital whiteboard	A free-form conceptual modelling canvas for collaborative workshopping of concept models and their properties
EAA-21	Digital wireframes	A drag-and-drop wireframing canvas for the design and configuration of data-driven user interfaces and the flow logic of digital interactions
EAA-22	Intelligent form design	Component-based forms development for assembling form layouts, UI controls and decision logic for web-based forms. Includes a packaged connector for Salesforce Lightning
EAA-23	Interaction flow	Freeform design of the flow of web-based customer interactions including decision tables for the flow-logic of pages and the invocation of components and API's
EAA-24	Idea capture	Democratise innovation by capturing ideas from teams at the coalface of user interactions
EAA-25	Idea board	Prioritise ideas, map design artefacts, define the business case and flow work through to the agile delivery backlog
EAA-26	Conceptual, logical and physical data models	Create structured representations of the sequence of actions and steps between business actors and systems in the execution of common business transactions
EAA-27	Metadata management	Manage the specification and maintenance of data attributes, data types and the ongoing governance and constraints on controlled vocabularies
EAA-28	Dictionaries and glossaries	Create, maintain, reconcile, and publish common global definitions of business terminology across an enterprise
EAA-29	Entity relation diagrams	Visualise the logical schema of the classes and relations in an enterprise data model through the industry-standard template of entity relation diagrams

(*continued*)

(*continued*)

	EAA name	EAA description
EAA-30	Schema import/export	Ingest and publish logical data definitions as XML schemas (XSD) for sharing data models with external data sources
EAA-31	Ontology editing	Support for the design and development of RDF-based ontologies for the semantic expression of enterprise classes and their properties (attributes and relations)
EAA-32	Dataflow diagrams	Visualise the flow of information within an organisation from the role-based inputs and outputs of stakeholders, through process steps, application functions and information repositories
EAA-33	Reference model mapping	Leverage pre-packaged reference models for operating model guidance and alignment with industry standards
EAA-34	API management	Coordinate the usage of a catalogue of enterprise API's, including SWAGGER compliant import/export of interface definitions with mappings to enterprise data model schemas
EAA-35	Capability assessments	Collaboratively assess capability maturity to understand how well the business is performing, identifying pain-points and opportunities for uplifts
EAA-36	Portfolio management	Coordinate and monitor portfolios of change initiatives with strategic alignment of resource allocations, budgets, business cases and the delivery of value-aligned outcomes
EAA-37	Project delivery	Plan, manage and track, waterfall or SAFe-aligned Agile project teams and the flow of work through the stages of a business or technology transformation
EAA-38	Project work schedule	Manage the allocation of resources to optimise the use of skills and effort in the delivery of prioritised work packages
EAA-39	Enablement roadmap	An interactive planning tool that addresses the multi-dimensional challenge of analysing, reconciling, and communicating scope, impacts and timelines across a portfolio of change programs

(*continued*)

(*continued*)

	EAA name	EAA description
EAA-40	Test management	Manage a comprehensive testing plan for technology implementations including the ability to create, assign and monitor the lifecycle and acceptance criteria of defects and enhancement requests
EAA-41	Release planning	Coordinate the scope and planning of release cycles with end-to-end traceability of the functional scope of a release, the requirements addressed and the auto-generation of release notes
EAA-42	Value proposition whiteboard	Collaboratively workshop customer needs & desired benefits as you design product offerings aimed at targeted customer segments
EAA-43	UI component library	Manage and track the design and configuration of UI components and their usage in a library of digital forms

References

1. Niemi, E.: Enterprise architecture benefits: perceptions from literature and practice. Tietotekniikan tutkimusinstituutin julkaisuja **18**, 1236–1615 (2008)
2. Guo, H., Li, J., Gao, S.: Understanding challenges of applying enterprise architecture in public sectors: A technology acceptance perspective. In: 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW). IEEE (2019)
3. ISO/IEC/IEEE: ISO/IEC/IEEE 42020:2019 (2019)
4. Merriam-Webster: Alignment (2021). Available from https://www.merriam-webster.com/dictionary/alignment
5. Oxfordify: Alignment (2021). Available from https://www.oxfordify.com/meaning/alignment
6. Maes, R., et al.: Redefining business-IT alignment through a unified framework. Universiteit Van Amsterdam/Cap Gemini White Paper (2000)
7. Gregor, S., Hart, D., Martin, N.: Enterprise architectures: enablers of business strategy and IS/IT alignment in government. Inf. Technol. People (2007)
8. Henderson, J.C., Venkatraman, H.: Strategic alignment: leveraging information technology for transforming organizations. IBM Syst. J. **38**(2.3), 472–484 (1999)
9. Luftman, J., Brier, T.: Achieving and sustaining business-IT alignment. Calif. Manage. Rev. **42**(1), 109–122 (1999)
10. Luftman, J., Kempaiah, R.: An update on business-IT alignment: "a line" has been drawn. MIS Quart. Exec. **6**(3) (2007)
11. Guo, H., Li, J., Gao, S., Smite, D.: Understanding how enterprise architecture contributes to organizational alignment. In: Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y.K., Pappas, I., Mäntymäki, M. (eds.) I3E 2021. LNCS, vol. 12896, pp. 383–396. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85447-8_33

12. MANAGEMENT, M.S.S.O.: MIT expert recaps 30-plus years of enterprise architecture (2020). Available from https://mitsloan.mit.edu/ideas-made-to-matter/mit-expert-recaps-30-plus-years-enterprise-architecture
13. Garousi, V., Felderer, M., Mäntylä, M.V.: The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. In: Proceedings of the 20th international conference on evaluation and assessment in software engineering (2016)
14. Forbes Media LLC.: Gartner (IT) (2021). Available from https://www.forbes.com/companies/gartner/
15. Gartner: Gartner Magic Quadrant for Enterprise Architecture Tools (2020). Available from https://www.gartner.com/en/documents/3970555/magic-quadrant-for-enterprise-architecture-tools
16. Korhonen, J.J., et al.: Adaptive enterprise architecture for the future: towards a reconceptualization of EA. In: 2016 IEEE 18th Conference on Business Informatics (CBI). IEEE (2016)
17. Capsifi: [cited 2021]. Available from https://www.capsifi.com/
18. Avolution: Avolution (2021). Available from https://www.avolutionsoftware.com/
19. Ardoq AS: Ardoq (2021) [cited 2020]. Available from https://www.ardoq.com/
20. LeanIX: LeanIX (2021). Available from https://www.leanix.net/en/

Check for updates

Contextuality and Temporality of Enterprise Architecture Problems: A Comparative Case Study

Ari Rouvari(✉) and Samuli Pekkola

Tampere University, Kalevantie 4, 33014 Tampere, Finland
{ari.rouvari,samuli.pekkola}@tuni.fi

Abstract. Many enterprise architecture (EA) projects face severe challenges and risks and may even fail. Although the problems are well-known and identified in the literature, the projects keep struggling. It is thus reasonable to ask why they cannot exploit earlier experiences. In this paper, we try to find reasons for these issues by conducting a comparative case study of twelve public sector EA projects in Finland. By using the earlier identified obstacles, risks, and critical success factors as a lens, we show, through participatory observations of the cases, that the problems emerge very differently in different projects, contexts, and the moments of time. This contextuality and temporality necessitates different coping mechanisms as each case is unique and make learning and the application of earlier experiences and practices difficult. We argue that the problems with contextuality and temporality and their inadequate consideration are the main reasons for often failing EA projects.

Keywords: EA work · Problems · Architect · Qualitative case study

1 Introduction

With digital transformation, the organizations strive toward improvement by implementing radical changes with information, computing, communication, and connectivity technologies (Vial 2019). They thus significantly change their current practices and processes, and create new business models, services, and products. Enterprise architecture (EA) has potential to support this transformation (Lamanna and Kurnia 2022; Niemi and Pekkola 2020; Korhonen and Halén 2017). Appropriate EA tools help to reach strategic goals and outcomes of digital transformation.

However, implementing and applying EA into digital transformation is not an easy task. Dang and Pekkola (2017), among others (Banaeianjahromi and Smolander 2019; Seppänen 2014; Kaisler et al. 2005) have identified several problems in using EA in public and private sector organizations. The problems range from communication problems to problems with managers, EA work, EA artifacts, assessment, and EA frameworks (Kaisler and Armour 2017). They hinder EA implementation and use, and may even result a failure. For this reason, obstacles, risks, and critical success factors in EA work

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 180–197, 2023.
https://doi.org/10.1007/978-3-031-26886-1_11

have been studied almost 20 years (Ylimäki 2006; Banaeianjahromi and Smolander 2019; Kaisler and Armour 2017) – but still the problems prevail. Thus, despite the studies on EA problems and their root causes (Dang and Pekkola 2017), it is reasonable to ask why the problems constantly emerge despite the organizations have prepared for them?

We argue the EA problems and failures are highly context specific. Consequently, for example knowing that there typically are some communication challenges is not helpful because the challenges appear differently. These dissimilarities, first, make it difficult to identify the challenges and their root causes early enough, and second, they require different coping mechanisms – depending on the context. This evidently hinders the learning from the past EA projects. In this paper we study this argument through an in-depth comparative case study of twelve public sector EA projects in Finland. We first develop an analysis framework from the literature and then use it as a lens to study the EA projects, their documents and the first author's first-hand observations as their chief enterprise architect or consultant.

The paper is organized as follows. First related research and our framework for analysis are presented. This is followed by research settings and methods. Third, our findings are summarized. The paper ends with discussion and concluding sections.

2 Related Research

Enterprise architecture aims at providing a holistic view of the organization and its processes, data resources, information systems and information technologies, their current and future states and how to reach the goals, and this way improve organizational decision-making (Simon et al. 2013; Kaisler and Armour 2017). Due to its comprehensiveness, complex EA projects are infamous for their failures (Dang and Pekkola 2017). These failures have been studied in various contexts. For example, in the public sector in general (Lemmetti and Pekkola 2012; Dang and Pekkola 2017), in government offices, municipalities and higher education institutions (Seppänen 2014), and in many other settings (e.g. Ahlemann et al. 2012; Löhe and Legner 2013).

The challenges are identified being largely non-technical (Kaisler et al. 2005). Non-technicality is apparent in the public sector, where Banaeianjahromi (2018) categorized 17 obstacles in EA development in Iranian governmental organizations in four groups: management commitment, infrastructure preparation, personnel engagement, and government and politics. Similarly, Kaisler and Armour (2017) identified "classic" critical problems being related to modelling, managing, and maintaining EA, and new ones being related to security and privacy, new technologies and their applications (e.g., cloud computing, opensource software, sensors), big data, and microservices, all concentrating on issues around technologies, not technologies *per se.*

The lack of communication and collaboration seem to be a major problem in many EA projects (Banaeianjahromi and Smolander 2019). The other obstacles include the lack of management support, the lack of knowledge among management, the lack of motivation or knowledge among personnel, resistance to change, EA consultant-related issues and government-related political issues. Niemi (2016) argued that there are potential problems with the stakeholders conflicting views and expectations on EA and EA benefits, while Dang (2018) stated the root causes origin from organizations, people, strategies, and different institutional pressures. To make the situation even more complex, individual obstacles are tightly interwoven (Banaeianjahromi and Smolander 2019), and the root causes are bound to organization conditions and environments.

In addition to obstacles, also critical success factors (CSFs) have been studied. CSF are, by definition, critical for successful outcomes. If they fail, the become obstacles. In the context of EA, determination, destination, and dexterity have been identified being significant for successful EA work. (Seppänen 2014; Ylimäki 2006).

The third point of view for understanding the EA project problems is the risks. Niemi and Ylimäki (2008) studied generic EA risks and factors that may lead to negative outcomes in the EA projects and negative outcomes resulting from these factors. Their extensive list included: disagreement in the EA product and service requirements and their inadequate implementation, their quality and costs, and difficulties in using them, insufficient organizational security, resources or competencies, lack of management support, information quality and access problems, poorly designed EA processes, incompatibility between various work practices, and inadequate planning, control, and management mechanisms.

Obstacles, critical success factors, and risks form a basis for our analysis framework. Table 1 lists 11 obstacles, Table 2 12 critical success factors and Table 3 13 risks, all that may occur and influence the EA projects. With the framework, critical and challenging issues in EA work can be identified.

Table 1. EA obstacles (Banaeianjahromi 2018 unless otherwise stated).

ID	Obstacle	Description of obstacle
O1	Lack of communication and collaboration	Lack of collaboration between other personnel and architects, between members of a team and between organizations
O2	Lack of management support	Lack of management support to prioritize the EA development and to assign enough budget and resources
O3	Lack of knowledge among management	Lack of knowledge about the EA method, the importance or role of EA, using EA, or how to manage and steer EA work
O4	Lack of motivation among personnel	The personnel are not motivated to put efforts and use resources to EA work
O5	Lack of knowledge among personnel,	Lack of knowledge about the EA method, the importance or role of EA, using EA, or the objectives of EA

(*continued*)

Table 1. (*continued*)

ID	Obstacle	Description of obstacle
O6	Resistance to change	Personnel resist the change i.e., they do not adapt to the changes that the EA would introduce due to several reasons: lack of knowledge, lack of trust, or fear of losing jobs
O7	EA consultant-related issues	Internal or external consultants lack skills like EA competence, guiding and mentoring competence, motivating competence, or communication competence
O8	Government-related political issues	Inappropriate definition of operations, lack of long-term goals, and political and governmental changes
O9	Influence of institutional pressure on organizations work life	Institutional pressures make EA to a part of the organization's work life, influencing EA's outcomes. (Brosius et al. 2018)
O10	EA Governance framework	An EA governance framework is a critical tool to ensure that the EA can be used appropriately and in long-term. (Aziz et al. 2005)
O11	Enterprise Architecture Management	Enterprise architecture management concerns the establishment and continuous development of EA. This includes the tasks of planning and controlling business change. (Aier et al. 2011, 645)

Table 2. EA critical success factors (Ylimäki 2006).

ID	CSF	Description of CSF
CSF12	Scope and Purpose	Depth and wide of the EA definition in terms of time and content
CSF13	Communication and Common Language	Understandable architecture definitions, legends for diagrams, Communication channels and timing of communication
CSF14	Business Driven Approach	Business Driven Approach to ensure that EA initiatives are traceable to the business strategy (Schekkerman 2004; Van Eck et al. 2004)

(*continued*)

Table 2. (*continued*)

ID	CSF	Description of CSF
CSF15	Commitment	Commitment (support, leadership, sponsorship, or involvement) of the key stakeholder of project to EA work
CSF16	Development Methodology and Tool	Systematic use of an EA method, frameworks, specifications, and modelling and repository tools
CSF17	EA Models and Artifacts	Usefulness and understandable EA models (e.g., a metamodel) and artifacts (diagrams, tables, matrices, narratives)
CSF18	EA Governance	Well documented EA governance model including roles of stakeholders, period of validity of EA and availability information
CSF19	Project and Program Management	Managing, coordinating, facilitating, and planning the EA projects
CSF20	Assessment and Evaluation	EA evaluation is challenging, because it may take years before the effects and consequences can be measured
CSF21	IT investment and acquisition strategies	Relationship between architectural and investment decisions
CSF22	Skilled team, training, and education	Defined and documented roles of team members, adequate EA competence based on training, education, or work experience
CSF23	Organizational culture	Attitudes towards architecture approach or changes, trusting environment, open communication, and organizational constraints

Table 3. EA risks (Niemi and Ylimäki 2008).

ID	Risk	Description of risk
R24	Lack of use of EA products and services	Lack of use of EA modeling tools, repositories, EA frameworks and EA consultancy
R25	Dissatisfaction of customers	Dissatisfaction with EA services, consultancy, and EA tools

(*continued*)

Table 3. (*continued*)

ID	Risk	Description of risk
R26	Misuse or misinterpretation of EA products	Difficulty in using EA products or services, insufficient competence for using them correctly, inadequate instructions and training
R27	Insufficient realization of EA objectives	Inadequate implementation of EA, inadequate compliance between EA and its implementations, inadequate planning of implementation, inadequate EA guidance to the implementation project
R28	Inadequate EA process performance	Inadequately designed, described, and realized EA working processes
R29	Insufficient predictability of outcomes	Inadequate planning and control mechanisms, insufficient comprehension of objectives, insufficient work practices, insufficient feedback mechanisms
R30	Insufficient documentation	Insufficient EA artefacts (diagrams, matrices, tables, and narratives)
R31	Personnel problems	Lack of skills and competences, instructions and training, lack of motivation and interest, poor conflict management, incompatibility between characteristics of participants and processes
R32	Participant frustration	Inappropriate EA tools and consultancy, insufficient resources, and guidance
R33	Information loss or theft	Insufficient information quality, accessibility, reliability, integrity, presentation, or security, inadequate amount of information
R34	Diminished EA work system performance – environment	Insufficient management support, resources (time, personnel, money), inconsistencies with organizational culture, partners or legislation, incompatibility between environment and the EA work system, between EA and reality, insufficient flexibility, or competence for understanding EA, high level of turmoil and distractions

(*continued*)

Table 3. (*continued*)

ID	Risk	Description of risk
R35	Diminished EA work system performance - infrastructure	Inadequate human support, unclear responsibilities, insufficient competences, separate siloes, inadequate information system or technical infrastructure
R36	Ineffective EA work system performance	Poor alignment between organizational strategy and the EA, unclear or missing "big picture" of EA, inadequate control mechanisms for the strategy changes, inadequate EA strategy, incorrect comprehension of strategy

3 Research Method and Research Settings

Earlier listed obstacles, critical success factors and risks have been identified in certain contexts. Although the researchers have discussed their applicability in broader settings, i.e., they have generalized their findings, the problems have been found to be very context dependent (e.g., Seppänen 2014; Ylimäki 2006). Consequently, to understand why the problems constantly emerge despite they are well-known and documented, we analyzed the problems and their causes in 12 public sector EA projects in Finland. From this perspective, this is a comparative case study of twelve cases (Yin 2009).

The first author has been working several years as chief enterprise architect or consultant in those projects. He thus had a unique chance to gain first-hand data about the problems, and how and why they appeared in the projects. We thus rely on the participatory observations (Clark et al. 2009) where the researcher just observes and reflects different actions and situation. This is contrary to action research (Baskerville 1999) where the researcher actively attempts to improve and change the situation. The use of action research thus puts focus on the changes (improvements) and respective actions. Thus, in this paper, our sample is cross-sectional, which we just observe without trying to change it.

Our research material consists of several published and unpublished architectural definitions and artefacts and different plans. The first author retrospectively analyzed them by using our analysis framework and his own experiences as an actor there. Table 4 lists our EA projects and their different details, including our research material.

The first author started the data analysis by reviewing all the EA projects. He then analyzed and classified the problems by using our analysis framework to list the obstacles, critical success factors, and risks. We illustrate the analysis process in the next chapter through an example of the Case RA06. The analysis results were regularly discussed and reflected by both authors to ensure correct and consistent interpretations.

Table 4. The case EA projects.

ID/period	Architecture descriptions	Domains (Strategy, Business, Data, Applications)	Volume	Type of customer	Number of stakeholders	Size of the EA group
RA01 Q3/2019 ->	Baseline and target	S B D A	65 diagrams 13 tables	Ministry (national)	4	5
RA02 Q1/2021 ->	Baseline	B D	12 diagrams 8 tables	Ministry (national)	6	15
RA03 Q1/2020->	Target	S B D A	27 diagrams 16 tables	Ministry (national)	3	7
RA04 Q3/2018-Q3/2019	Target	S B D A	21 diagrams 12 tables	Ministry (national)	~20	3
RA05 Q4/2020-Q3/2021	Target	S B	18 diagrams 6 tables	Ministry (national)	3	6
RA06 12/2017–11/2019	Target	S B D A	27 diagrams	EU	11	2
EA07 Q2/2021–Q4/2021	Target	S B	19 diagrams 1 table	Ministries (national)	6	3
RA08 Q1/2020–Q2/2021	Target	S B	50 pages 40 diagrams 10 tables	Ministry (national)	15	5
EA09 Q1/2019 ->cont.	Target	S B D A	56 diagrams 14 tables	Agency (national)	4	6
EA10 Q3/2019 ->cont.	Baseline and target	S B D A	41 diagrams 5 tables	Agency (national)	4	5
EA11 Q3/2019–Q4/2021	Baseline	T	14 diagrams 10 tables	Vendor (private)	1	4
EA12 Q2/2019 -> cont.	Target	S B	26 diagrams 6 tables	Department in a ministry)	2	5

4 Observations

We will first present a summary of the cases. Then we will walk through the Case RA06 to show how the problems were spotted and classified, and what effects they had. This is followed by two other cases and their main findings to allow case comparison. Finally, we take a horizonal approach and analyze five problems in all cases to see how they emerged there, and what was the impact of the contexts.

Table 5. The summary of the cases and the problems they faced. (Success -row: E = exceeded expectations, ok = met the expectations, P = met the expectations partially, fail = failed).

	Cases											
Problems	***RA 01***	***RA 02***	***RA 03***	***RA04***	***RA 05***	***RA 06***	***EA 07***	***RA 08***	***EA 09***	***EA 10***	***EA 11***	***EA 12***
Success	ok	ok	ok	E	E	P	E	ok	E	fail	P	P
O1	X		X						(x)	X		X
O2												
O3							X					
O4						X				X		
O5	(x)		(x)		(x)			(x)	(x)	X		(x)
O6						X						
O7	?	?	?	?	?	X	?	?	?	X	?	?
O8						(x)	X		(x)			
O9										(x)		
O10	(x)		(x)		(x)	X	(x)			X		(x)
O11							(x)			X		
CSF12							X			X	(x)	
CSF13						(x)						
CSF14												
CSF15						(x)	?			X		
CSF16												
CSF17												
CSF18	(x)		(x)		(x)	X	(x)			X		(x)
CSF19												
CSF20	X	X	X		(x)		?	(x)	(x)	X	(x)	X
CSF21												
CSF22											X	
CSF23											X	
R24												
R25										X		
R26										X		
R27	(x)	(x)	(x)		(x)	X		X	?			(x)
R28										X		
R29												
R30												
R31										X		
R32												
R33												
R34										X		
R35										X		
R36										X		

Table 5 summarizes the cases and their problems. They faced several problems, and the same problems emerged in many cases. It is thus relevant to ask whether the problems were actually similar. Capital X refers to critical problems, small (x) in brackets to challenging problems, and a question mark to unknown items. Empty cell refers to a good situation. Shaded columns/rows show the cases and problems that are discussed in detail later.

4.1 Case RA06: Reference Architecture for Education Sector

The RA06 was a part of a large, two-year-long EU funded project with eleven partners. One of the deliverables was a reference architecture for education domain. This consisted of strategy, business, data, and application architectures. The project and the reference architecture work were steered by the project plan. When the work was completed by the project team, it was elaborated by numerous national and international actors. The EA project got a lot of positive feedback so many of the architecture's building blocks were used in other architectures. In our analysis, RA06 received nine hits: 6 critical and 3 challenging.

All actors **were not motivated to EA** (O4). Some did not understand its role, and because its notation language looked too technical for them, they were afraid that the EA would hamper communication. This was discovered in the meetings where forthcoming seminars were planned, and when people responsible for communication did not want to show the architecture diagrams to other stakeholders. We classified this obstacle as critical because the EA definition was one of the main project deliverables without which the whole project would have failed. Luckily, the architects gradually gained the other project members' trust, so they began to better understand its importance.

Some project members protested the EA and **resisted to change** their practices (O6). The EA was seen as a new tool for managing the service development. Consequently, they did not actively participate in the EA work. This resulted in the EA project missing some domain competences. Although some of this knowledge was compensated by common understanding (about the education domain), the obstacle was not solved entirely. As the distrust towards the EA remained, we considered this obstacle as critical.

When the project begun, the ministry's support was partly missing **(O8 Government-related political issues)**. This was observed in the steering group meetings where the project was dictated, without further guidance, to a direction where no EA work is needed. The top management support and its absence were thus critical. Later, the support was gained by demonstrating the importance and benefits of EA work. Currently the ministry reuses the architecture artifacts in other projects. This success means we classify this obstacle as "challenging".

EA Governance framework (O10) is critical obstacle because of a missing EA governance model. Without a governance model the architecture definition is nearly useless – it can neither be used nor further developed in a larger scale. We will discuss this more in the Sect. 4.4.

The project used the Archimate modelling language in its work. Some partners considered it difficult to understand. As the architecture definitions should be easily understandable without any interpretation, **Communication and Common Language**

(CSF13) CSF turned out to be very challenging. The situation alleviated when clear and understandable narrative on the EA diagram descriptions was provided.

The project coordinator and main partners were committed to the project**,** but not so much to the EA. The **lack of commitment** (CFC15) emerged similarly to the lack of motivation (O4) and personnel's resistance to change (O6) in the steering group meetings and in the daily work. These three challenges are in fact strongly intertwined, making them hard to distinct. During the project, the partners' commitment increased when they learned and understood the meaning and the importance of EA.

The architecture descriptions have not yet led to organizational or system changes although some mock-ups, proofs of concepts, and pilot solutions have been made. **Insufficient realization of EA objectives** (R27) is thus a critical risk that may have its impacts on the future EA activities. This issue will be discussed more in the Sect. 4.4.

The biggest challenge in the project was the partners' attitude against the EA: some were unfamiliar with the EA method, some were dubious about the utility of EA, and some had unrealistic expectations and narrow understanding of the EA work. These challenges resulted from the EA consultant-related issues (O7), which lead to the project participants' low motivation, commitment, and understanding of the role of the EA.

4.2 EA10: Enterprise Architecture for a Governmental Agency

The ministry funded EA10 project designed a first EA definition for a governmental agency. The EA aimed to cover all business services, providing business, data, and application architecture descriptions. Due to lack of commonly missing strategies in the national agencies, the EA work was based on the national reference architecture of that branch of administration. The EA work was about developing a baseline architecture because the business services were deployed simultaneously when they were modelled. The EA project team of twelve persons constitutes chief architects, customers, and ICT provider representatives. The following problems were evident:

O1 Lack of Communication and Collaboration. The EA team had meetings sporadically. Collaboration and communication were thus very weak.

O7 EA Consultant-Related Issues. The EA consultant failed in motivating the customer to apply the EA. The customer was suspicious not only towards the EA, but it trusted neither the ICT provider nor the architecture consultant. In fact, the customer blamed the ICT provider if everything did not go smoothly.

O11 Enterprise Architecture Management. The EA management was weak due to the customer's inappropriate information on EA as a method and on its objectives.

CSF12 Scope and Purpose. The scope was adequate, but the customer had not yet realized the purpose of EA work. The customer assumed that the EA is only for technology and application design.

R25 Dissatisfaction of Customers. The customer had not realized the importance and meaning of the EA work. They were in a hurry to start to use the EA and felt that it was not promoting the service development quickly enough.

R31 Personnel Problems. The customer lacked motivation, interest, and understanding of the purpose of the EA work. This concretized in the project plan in terms of insufficient human and time resources.

R35 Diminished EA Work System Performance – Infrastructure. The customer was not motivated to participate in the EA work, so they postponed pre-arranged meetings. This decreased the architecture descriptions quality and made the roadmap and schedule development for the EA work difficult.

The problems accumulated when the obstacles and risks materialized. The root cause was the unclear understanding of the importance and role of the EA work. This reduced the customer's motivation and commitment.

4.3 EA07: Reference Architecture for Five Agencies

The EA07 is an ongoing governmental project where a reference architecture for two Finnish collaborative administrative branches is created. The architecture consists of strategy, business, data, and application architectures. Currently five government agencies participate the project, although the number is expected to increase. The project was steered by the project plan and governmental funding.

O3 Lack of Knowledge. Some stakeholders disliked the idea that the EA work changes and manages their domain. They did not understand its role and meaning. This stopped the project. Now, when the understanding has increased, the project is re-started.

O8 Government-Related Political Issues. The EA project was planned to start during the previous government (before 2019). The project was postponed over the parliamentary elections. After the election, there was a significant change in the governmental policies that steered the project. A reason for the delay was the lack of collaboration capability between two ministries. Such collaboration is not supported by the Finnish administrative structures or traditions.

CSF12 Scope and Purpose. At the beginning, the scope and purpose were not explicitly set. This was assumed to take place during the project when the architects and other stakeholders work together and update them. The definitions are expected to be based on various strategies and other documents.

The Finnish governmental structure does not support inter-administrative collaboration and EA work. Legislation is made in and for silos and siloed organizations, and interoperability in business, data and application levels remains incomplete.

4.4 Horizontal Viewpoints on Problems

Next, we will look at the problems and how they emerged in different cases.

Lack of Communication and Collaboration (O1). The lack of communication and collaboration got four critical and one challenging assessments. In the cases RA01 and RA03, initially the EA project teams consisted of only an EA consultant and the ministry representatives. Later both teams were expanded by ICT architects and CIOs. There was no collaboration with education providers or other administrative branches, i.e. domain experts. In the case EA12, the team was lacking collaboration with other departments. The case EA09 did not succeeded in collaborating with the steering ministry, although later some workshops were organized with the architects. The EA10 team had meetings only occasionally.

Poorly understood EA seem to be the main cause for the communication problems. The architecture descriptions are only one of the goals. Equally important is learning by doing the EA since this takes the work into broader use. Wider and deeper collaboration means better learning and better learning means better architecture descriptions and results. There were also practical reasons to not invest in collaboration. The EA work was (and is!) traditionally done in small teams where it is easier and more efficient to work with familiar people, just publish the descriptions, and hope that the other stakeholders adopt and start to use them. The EA teams were not ready to improve their work practices and processes. Collaboration also necessitates coordination and resources. Finally, the EA work was usually perceived as a one-time project, not a continuous process.

EA Consultant-Related Issues (O7). The EA consultant-related issues were critical at least in the cases RA06 and EA10. In all other cases, it was difficult to assess objectively since the first author was the main EA consultant there. However, he has received comments like: "The same enterprise architect is doing all EAs [descriptions]" and "The architecture descriptions look too similar." This means a good architect can create similar descriptions which are then interoperable, comparable, and consistent.

Nevertheless, these comments underline the importance of the consultants. Especially the EA consultants have a big influence on the architecture deliverables. Besides architectural competences, the consultants are expected to have other skills like communicating, collaborating, motivating, or social skills (c.f. Ylinen and Pekkola 2020). Consequently, the EA-consultant related issues emerged because the consultants did not have all necessary skills *in those projects*. For example, they may have missed communication or social skills which may have been adequate elsewhere.

EA Governance Framework (O10) and EA Governance (CSF18). The EA governance and governance framework (O10 & CSF18) was critical in two cases and challenging in five. In the case RA06, the architecture descriptions were planned to be used as input to other EA projects. This was very difficult. In case EA10, the absence of the governance framework almost suspended the EA project and hampered the service development. Although in five cases the EA governance management model was missing, they are projected to develop it later when their EA descriptions are mature enough.

The absence of the EA governance models and frameworks results from their low priority. The customer organizations were keen on utilizing the architectural descriptions without understanding what it requires in a long run. They were not familiar with making governance models or seeing the EA as a holistic tool for managing, steering, and developing operations.

Assessment and Evaluation (CSF20). This CSF got five critical and four challenging assessments. Evaluating architecture descriptions is challenging because it may take years before the effects and consequences emerge and can be measured (Ylimäki 2006). In fact, only two of our projects got good credits in the assessment. The case RA06 was evaluated by an external evaluation agency and the case RA04 in several real-life use cases and seminars and events. In general, the customers seem not to value assessment and evaluation. They are satisfied already when the EA is in use.

Insufficient Realization of EA Objectives (R27). An insufficient realization of the EA objectives refers to the ultimate EA goals: new services, applications, and processes. Architecture descriptions and learning are means to reach these goals. This risk got two critical and five challenging hits. All the projects which got "challenging" assessments are under construction, meaning the risk will most probably not occur. The original objectives of the case RA06 (a reference architecture) was not realized but many of its ideas and architecture descriptions have been reused in other projects. In the case RA08, instructions for implementing the reference architecture were not made. This makes it difficult to realize the objectives.

The EA objectives are usually not realized because of inadequate architecture implementation instructions. The actors seem to believe that architectural descriptions are sufficient for reaching the goals, so no separate EA implementation guidance is needed.

5 Discussions

There seem to be many reasons for successes and failures in the EA work. Three core factors in our public sector EA context are 1) limited collaboration, 2) administrative structure of the Finnish government, and 3) the role of EA consultancy. By managing these root causes, we can eliminate many other problems. However, as we will discuss next, they may not necessarily be the causes in other settings, or can be managed similarly.

Earlier studies have identified that the lack of communication and collaboration is the core obstacle that could explain many other problems (Banaeianjahromi and Smolander 2019). This is not surprising since EA development is mostly about communicating and collaborating with different stakeholders. The problem is also apparent in our EA projects. Eight projects have identified collaboration as essential strategic capability while four ignored it. However, collaboration problems occurred differently and because of dissimilar reasons. The cases RA01 and RA03 had internal problems and misunderstandings, the case EA12 had horizontal communication problems with other departments that were unwilling to collaborate, and the case EA09 vertical problems with its steering actors. The case EA10 had pragmatic problems of setting up meetings. Consequently, communication and collaboration problems emerged very differently, requiring different coping mechanisms.

Organizational structures, legislation, politics, and top-down guidance and governance have been identified as root causes in the public sector (Dang and Pekkola 2017). In our cases, the missing collaboration capability between administrative branches was due to the governmental structure. The Finnish public sector administration follows expertise-based, sectional responsibilities where different ministries and agencies mind their own businesses without interfering the others. Similarly, legislation is made in silos. Interoperability in semantic issues and vocabularies is incomplete. Financial structure does not support interoperability.

Under the circumstances inter-administrative collaboration becomes very difficult so Finnish public sector EA projects are usually not bridging administrative boundaries. For example, the means to manage interoperability challenges between different administrative branches are very limited. Here the case EA07 is a rare positive exception. It is a collaboration project between two ministries, aiming to deliver a new collaboration model to bridge inter-administrative branches. Its utilization remains to be seen as elections and changes in the government or its ministers, or a new government program influence the EA projects and may even terminate them (Banaeianjahromi and Smolander 2019).

The complexity of the EA concept (e.g., Haki and Legner 2021; Lemmetti and Pekkola 2012) often necessitates external interpreters to facilitate the EA work. This emphasizes the role of EA consultants. They can be seen as influential factors, both positively and negatively, behind many problems and how they are coped. An architect plays a major role in the EA development, its maintenance, establishing communication, and providing coordination between the business and the IT teams (Banaeianjahromi and Smolander 2019). Yet this work is difficult since the problems materialize differently in different contexts. This exacerbates the architect's work since his/her earlier experiences, competences, skills, or the social eye, that may have provided successful outcomes in the earlier occasions, may not be appropriate in a new context.

The core factors: collaboration, administrative structure, and consultancy, can thus be considered as root causes in the cases. However, at the same time, they are very contextual. For example, in the case EA10, the problems had accumulated up to the extent that replacing the consultant or investing in collaboration to get a fresh start would not have helped. The window of opportunity (Tyre and Orlikowski 1994) to advance EA in the agency was closed for the time-being. This means the other problems have become more significant root causes. Consequently, the problems are not only contextual, but their importance varies over time. This means that our currently unrealized problems may be realized and hit the fan in different circumstances.

The contextual and temporal characteristic of the problems makes their coping mechanisms challenging. Once a successful strategy and approach may not be successful in other occasion. In this sense, when different internal or external incidents alter the EA project environment, the problems' coping mechanisms need to change respectively (c.f. Smolander et al. 2021). Monitoring and then acting on these incidents can be the responsibility of the EA project, EA architect, or the whole team, depending on the organization and the scope of the project.

Contextuality and temporality also make the learning from previous projects and from the others difficult. Like the problems with transferring the EA practices and tools from one context to another (c.f. Dang and Pekkola 2017), also transferring the experiences and practices how to cope with different challenges is difficult. This means that in every new project and situation, one must learn the context, the stakeholders, the problems, and how they emerge there to be able to handle them appropriately. Learning from previous experiences is difficult.

6 Conclusion and Contributions

In this paper, we have studied EA-related problems and how they appear in EA projects. Although the problems have been studied earlier, their manifestations in individual projects have gained much less interest. Our analysis of twelve EA projects demonstrates that the problems are highly contextual and temporal, their relative importance varying between the projects and over time. This variation puts pressures to EA work which need to be responsive to the changes in the EA project contexts. Once successful preventative actions may be unsuccessful in other times and other contexts. We argue that this contextuality and temporality and their inadequate consideration are the main reasons for often failing EA projects.

This notion of the contextuality and temporality of EA problems is our main contribution. Although it has been implicitly acknowledged earlier, the varying appearances of the communication problems, for instance, demonstrate the need for profound understanding of this variation, and for practical mechanisms to analyze the problems and to acting accordingly. Simple suggestions for coping with certain challenges are too generic claims without significant practical value. For the same reason, we, the authors of this paper, abstain ourselves from making suggestions how to solve various problems.

This notion is useful for researchers as an encouragement for further research on the instantaneous of the problems and their evolution and changes over time, and for practitioners to pay attention to smaller details when trying to survive in the turbulent world of EA work.

The study has some limitations. First, the chosen research method, participative observations, is subjective as the first author was living the daily life of the projects. Although we have minimized the problems of accidental misanalyses by constantly reflecting the findings among the authors, subjectivity is still there. However, as our purpose was not to analyze all problems and their prevalence but to see how they emerged in different situations, subjective bias is minimal. Second, the context, Finnish public sector EA projects, some national, some EU wide, may emphasize some problems. The Finnish administrative structure as a root cause for inter-administrative problems is a good example of such local problems. Again, we were not interested in individual problems but their emergence. In this light, Scandinavian context does not limit the findings.

References

Ahlemann, F., Stettiner, E., Messerschmidt, M., Legner, C. (eds.): Strategic enterprise architecture management: challenges, best practices, and future developments. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24223-6

Aier, S., Gleichauf, B., Winter, R.: Understanding enterprise architecture management design. An empirical analysis. In: Wirtschaftsinformatik Proceedings, p. 50 (2011)

Aziz, S., Obitz, T., Modi, R., Sarkar, S.: Enterprise Architecture: A Governance Framework – Part I: Embedding Architecture into the Organization. Technical Report Infosys. https://cioindex.com/wp-content/uploads/nm/articlefiles/3998-EA-Governance-1.pdf (2005)

Banaeianjahromi, N.: The role of top management commitment in enterprise architecture development in governmental organizations. Complex Syst. Inform. Model. Q. **17**, 95–113 (2018)

Banaeianjahromi, N., Smolander, K.: Lack of communication and collaboration in enterprise architecture development. Inf. Syst. Front. **21**(4), 877–908 (2019)
Baskerville, R.L.: Investigating information systems with action research. Commun. Assoc. Inf. Syst. **2**(1), 19 (1999)
Brosius, M., Aier, S., Haki, K., Winter, R.: Enterprise architecture assimilation: an institutional perspective. In: ICIS 2018 (2018)
Clark, A., Holland, C., Katz, J., Peace, S.: Learning to see: lessons from a participatory observation research project in public spaces. Int. J. Soc. Res. Methodol. **12**(4), 345–360 (2009)
Dang, D.D., Pekkola, S.: Problems of enterprise architecture adoption in the public sector: root causes and some solutions. In: Rusu, L., Viscusi, G. (eds.) Information Technology Governance in Public Organizations. ISIS, vol. 38, pp. 177–198. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58978-7_8
Dang, D.: Enterprise Architecture in the Public Sector. Adoption and Institutionalization. Tampere University of Technology (2018)
Haki, K., Legner, C.: The mechanics of enterprise architecture principles. J. Assoc. Inf. Syst. **22**(5), 1334–1375 (2021)
Kaisler, S.H., Armour, F., Valivullah, M.: Enterprise architecting: critical problems. In: The 38th Hawaii International Conference on System Sciences (2005)
Kaisler, S., Armour, F.: 15 years of enterprise architecting at HICSS: revisiting the critical problems. In: Proceedings of the Hawaii International Conference on System Sciences (2017)
Korhonen, J.J., Halén, M.: Enterprise architecture for digital transformation. In: IEEE 19th Conference on Business Informatics (CBI), pp. 349–358 (2017)
Lamanna, F., Kurnia, S.:Digital Transformation of superannuation and the role of enterprise architecture. In: PACIS 2022 Proceedings, p. 214 (2022)
Lemmetti, J., Pekkola, S.: Understanding enterprise architecture: perceptions by the Finnish public sector. In: Scholl, H.J., Janssen, M., Wimmer, M.A., Moe, C.E., Flak, L Skiftenes (eds.) EGOV 2012. LNCS, vol. 7443, pp. 162–173. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33489-4_14
Löhe, J., Legner, C.: Overcoming implementation challenges in enterprise architecture management: a design theory for architecture-driven IT Management (ADRIMA). IseB **12**(1), 101–137 (2013). https://doi.org/10.1007/s10257-012-0211-y
Niemi, E.: Enterprise Architecture Benefit Realization. Tampere University of Technology (2016)
Niemi, E., Pekkola, S.: The benefits of enterprise architecture in organizational transformation. Bus. Inf. Syst. Eng. **62**(6), 585–597 (2020)
Niemi, E., Ylimäki, T.: Defining enterprise architecture risks in business environment. In: Niemi, E., Ylimäki, T., Hämäläinen, N. (eds.) Evaluation of enterprise and software architectures: critical issues, metrics and practices. University of Jyväskylä, Information Technology Research Institute (2008)
Schekkerman, J.: How to Survive in the Jungle of Enterprise Architecture Frameworks: Creating or Choosing an Enterprise Architecture Framework. Trafford Publishing (2004)
Seppänen, V.: From Problems to Critical Success Factors of Enterprise Architecture Adoption. University of Jyväskylä (2014)
Simon, D., Fischbach, K., Schoder, D.: Enterprise architecture management and its role in corporate strategic management. IseB **12**(1), 5–42 (2013). https://doi.org/10.1007/s10257-013-0213-4
Smolander, K., Rossi, M., Pekkola, S.: Heroes, contracts, cooperation, and processes: changes in collaboration in a large enterprise systems project. Inform. Manage. **58**(2), 103407 (2021)
Tyre, M.J., Orlikowski, W.J.: Windows of opportunity: temporal patterns of technological adaptation in organizations. Organ. Sci. **5**(1), 98–118 (1994)
Van Eck, P., Blanken, H., Wieringa, R.: Project GRAAL: towards operational architecture alignment. Int. J. Coop. Inform. Syst. **13**(3), 235–255 (2004)

Vial, G.: Understanding digital transformation: a review and a research agenda. J. Strat. Inf. Syst. **28**(2), 118–144 (2019)
Yin, R.K.: Case Study Research: Design and Methods, vol. 5. Sage (2009)
Ylimäki, T.: Potential critical success factors for enterprise architecture. J. Enterp. Architect. **2**(4), 2940 (2006)
Ylinen, M., Pekkola, S.: Jack-of-all-trades torn apart: Skills and competences of an enterprise architect. In: Proceedings of the European Conference on Information Systems (ECIS) (2020)

Zooming in on Competences in Ontology-Based Enterprise Architecture Modeling

Rodrigo F. Calhau[1,2(✉)] and João Paulo A. Almeida[1]

[1] Ontology & Conceptual Modeling Research Group (NEMO), Federal University of Espírito Santo (UFES), Vitória, Brazil
jpalmeida@ieee.org

[2] LEDS Research Group, Federal Insitute of Espírito Santo (IFES), Serra, Brazil
calhau@ifes.edu.br

Abstract. Organizations must pay close attention to human resource development in order to be successful. Because of this, competence-based approaches have received increased attention, as the demand for qualified people with the right combination of competences establishes itself as a major factor of organizational performance. This paper examines how competences can be incorporated into Enterprise Architecture modeling: (i) we identify a key set of competence-related concepts such as skills, knowledge, and attitudes, (ii) analyze and relate them using a reference ontology (grounded on the Unified Foundational Ontology), and (iii) propose a representation strategy for modeling competences and their constituent elements leveraging the ArchiMate language, discussing how the proposed models can fit in enterprise competence-based practices.

Keywords: Competences · Ontologies · Competence modeling · Enterprise architecture

1 Introduction

Given the importance of human performance in business management and the transformation of socioeconomic systems in general, it is not surprising that human resource management, education, and training typically receive a considerable interest. The drive to human development has resulted in advancements in fields such as Vocational Education and Training (VET) and Human Resource Management (HRM). One of these advancements has been the gradual change from content-based to *competence-based* methods, which represents a change in Vocational Education and Training from a supply-oriented to a demand-oriented model [17,28].

A focus on competences promotes deeper integration of formal education, vocational training, and professional development, which is aligned with lifelong learning strategies [17]. Further, competence-based methods serve to link an

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 198–213, 2023.
https://doi.org/10.1007/978-3-031-26886-1_12

organization's future requirements to its Human Resources (HR) programs [10]. Personnel selection, development, and performance monitoring, as well as corporate strategy planning, are all examples of competence-based activities in human resource management [27]. By reviewing staff competences, organizations can conduct self-assessment to improve their HR programs, revisiting talent recruitment procedures, performance management systems, training and development tools, employee engagement initiatives, and institutional development plans [10].

The importance of competences to the enterprise has motivated past efforts in which key concepts of Competence Management (CM) were incorporated into Enterprise Architecture modeling [5]. In that work, personal competences were conceived of as "dispositions" of individual business actors that are manifested through their behaviour in organizational contexts. A number of patterns for competence representation in *ArchiMate* were proposed, leveraging on the capability construct. This paper builds up from that baseline and identifies and tackles challenges pursuant to "zooming in" on competences (which were considered as black-boxes in [5]).

The literature on Competence Management reveals that it is indispensable to examine the build up of competences in detail. Over the years, competence has been typically conceived of as the result of the interaction of specific *knowledge* and generic *skills* [24], mediated with *attitudes* [26]. Personal traits, mindset, patterns of thinking, and tacit knowledge are also considered by some authors to be part of competence [8]. While these terms are pervasive in the Competence Management literature, their precise definition has remained elusive. The terms are frequently used interchangeably and are sometimes confused with "competence" itself [26,29].

We argue that conceptual analysis of these notions and their relations is key to their adequate representation in Enterprise Architecture (EA) models. Domain-adequate representations are, in turn, key to support the use of EA models in competence-based practices. We approach the representation of competence elements in this paper by positioning the notions of competence, skill, knowledge, attitude and other personal characteristics through a reference ontology. The reference ontology is then used as a starting point to the representation of competences alongside their constituent elements in *ArchiMate*.

This paper is further structured as follows: Sect. 2 briefly reviews the relevant literature on competences and competence management, stating the key conceptual challenges for "zooming in" on competences, which involves the relations between competences, knowledge, skills, attitudes, and other characteristics. An ontology for these elements is offered in Sect. 3 by specializing the notion of "disposition" in the Unified Foundational Ontology (UFO) [13]. The representation of competence elements in examined in the *ArchiMate* language in Sect. 4. Section 5 discusses related work. Finally, Sect. 6 summarizes our effort and proposes a research agenda, which includes the integration of competence management with other key architectural domains of Enterprise Architecture.

2 Competences

Competence[1] is the general ability to perform well a set of mastery tasks [26]. It is not enough for an individual to have a variety of specific skills for this. Mastery of skills or knowledge does not ensure success in complex and unpredictable environments [29]. In addition to skills, the individual must have a sufficient understanding of the domain in question (knowledge) as well as know how to act appropriately in the context (attitude) [26]. In order to be efficient and effective in such situations, the individual must be able to integrate the most appropriate skills and knowledge for it [29]. As a consequence of this, various authors define competence as a combination of knowledge, skill, and attitude [7,20,24]. Competences, in other words, are highly valued qualifications that are accountable for the effective application of skills and knowledge in specific and complex contexts [29].

2.1 Skills and Competences

In general, *skills*, not unlike competences, allude to the capability to perform actions. The literature provides different definitions for skills emphasizing different aspects of it. For example, Rodriguez et al. [27] defines a skill as the ability of an individual to perform a task (discrete unit of work) well. Esposto [9] defines it as a set of general procedures that underlies the application of knowledge in a domain. Paquette [24] defines skills as processes that act on knowledge in an application domain [24].

There is no agreement on the best criterion for distinguishing competences and skills [29]. One existing distinction is the level of ability awareness. Competences would be more "conscious", while skills would be more "automatic" [29]. However, this distinction is insufficient because conscious actions occur with skills as well [29].

The level of complexity is another criterion that is invoked to differentiate competences and skills. Competences are considered more complex in this case than skills. Indeed, authors argue that skills *structure* competences [24,29]. Competences can be made up of sub-competences, forming an internal hierarchical structure inherent in the individual. In this sense, competence is a complex entity. That is, a competence can be formed by others, which can be formed by others, and so on. As a result, this internal hierarchical structure can be formed by many levels of sub-competences [29]. The *basis* of such an internal hierarchy, however, has not been well understood. Competence decomposition only occurs up to a certain level, where the "basic competences" are. Basic competences are divided into skills after this level. In this regard, it is unclear where basic competences end and skills begin [29]. Even skills can also be divided into different levels, until reaching the "basic skills".

[1] We adopt in this work the term "competence" to refer to an individual's performative ability, and refrain from using the term "competency".

Some Competence Models allow sub-competences or skills that make up a competence to be represented. However, as previously stated, the line between basic competence and skill is not always evident. Due to their similarities, the concepts of competence and skill are frequently misunderstood in definitions and representations. As a result, an important goal of a reference ontology for this domain is to clarify the similarities and differences between the concepts of competence and skill, settling how to position those two notions for a certain context of usage.

2.2 Knowledge

Internal representations of facts, principles, or theories in a specific domain are typically associated with "knowledge" [29]. It is the cognitive outcome of assimilation of concepts, ideas, or figures related to a specific topic [26]. Knowledge is linked to a specific person, the bearer, then it is difficult to transfer and assimilate [4]. Knowledge is assimilated when it becomes a part of the bearer's internal structure. As new information or facts are added, the structure changes [30]. This internal structure is not distinctive to the bearer but is integrated into the internal structure of abilities [30]. Indeed, such internal structures (of knowledge and skill) interact in practical applications and problem-solving [19]. Despite the fact that it changes over time [19], knowledge is a static (passive) entity [29]. It is stored in memory and retrieved using cognitive skills (mental processes) [29].

Many knowledge definitions are similar to skill descriptions as a result of learning. Some authors even consider skills to be a sub-type of knowledge. According [19], skills represent an individual's "practical knowledge" gained through experience. While an individual's interpretations and facts are known as declarative knowledge, the skills (what an individual knows how to do) are known as procedural knowledge [19]. Authors include that skills and knowledge are represented in a similar manner in human mind, via an interconnected internal structure [19].

Understanding an individual's knowledge in the context of CM is important for better understanding their competence. This is particularly useful during the *gap analysis* and *competence assessment* steps. Competence models, which represent a professional's knowledge, can aid in this task. However, confusion between the concepts of knowledge and skill can have an impact on model quality, making it difficult to model knowledge and skill clearly. Despite the similarities described above, skills and knowledge have subtle differences that can interfere with modeling. As a result, a reference ontology for this domain should provide a solid definition of knowledge and clarify the distinction between knowledge and skill.

2.3 Attitudes

In some definitions, attitudes are generally associated with an individual's behavior [20,24]. Others associate them with personality traits or the professional's psychological and emotional nature [26]. Attitude is a tendency to act (or feel) in a given situation [18]. It is based on assumptions, values, and beliefs, so they are non-neutral with respect to actions [18]. In general, definitions of attitude take into account the following characteristics: (i) mental state; (ii) values (beliefs, emotions); and (iii) predisposition to act or behave [1]. That is, it is a concept that is dependent on its context: a situation, an object, or a person. As a consequence, attitude is a disposition toward a specific phenomenon and can be considered a reaction to the context (object, person, or situation) [1]. This type of reaction is bipolar, so it may or may not be beneficial (positive or negative) to the environment in which it is found [1].

Attitudes are regarded as an important aspect of competences, and are included in many competence definitions as one of the key 'KSA' elements (Knowledge, Skills and Attitudes). In contrast to skills and knowledge, attitude is a more general characteristic that is not tied to a specific task or domain [26]. Because they have certain behavioral impact, attitudes are frequently confused with skills, particularly soft skills [20]. Again, as in the case of skills and knowledge, a reference ontology for this domain should clearly position attitudes with respect to the other elements of competence.

2.4 Other Characteristics

Although competence is commonly defined as a set of attitudes, skills, and knowledge, authors consider further types of elements to be components of competences. Personal traits, behavior, mindset, patterns of thinking, and tacit and explicit knowledge are considered by some authors to be part of competence [8]. This is recognized also by Westera [29], for whom competences have additional elements that are not clearly defined. According to Miranda et al. [20], competences are also formed by a set of personal characteristics required to perform tasks in a specific context, leading the authors to consider the KSAO model, a variation of the KSA model that includes "Other Characteristics" as a fourth element to define competence.

According to Westera [29], task analysis is insufficient to establish competences; instead, the individual's characteristics and experience must be considered. Le Deist and Winterton [17] emphasize the importance of focusing on the individual rather than their conduct. The authors explain that, in addition to performance, it is critical to look at traits, motives, attitudes (or values), and knowledge, among other things. Some KSA elements (attitude and knowledge) are considered personal characteristics by the author. Messick [19] extends on this point by stating that the psychological, emotional, social (environmental) situation, and even biomedical information must all be considered.

All of these characteristics, as well as behavior (performance, tasks, and outcomes), are evidence of an individual's competence. Hence, they are critical in the Competence Assessment task, which is one of the most demanding in the CM context, specifically measurement, both quantitative and qualitative, because it entails giving value to something that cannot be fully observed.

3 Ontological Analysis of Competence-Related Elements

We explore the multi-faceted phenomenon of competence by proposing a reference ontology for competence and its constituent elements. The issues discussed in Sect. 2 help us to identify focal points for this effort, and ultimately relate competences, knowledge, skill, attitudes and other human characteristics in a coherent representation.

3.1 Baseline

We build up on the work discussed in [5], which used the Unified Foundational Ontology (UFO) [13] to examine competences from an external perspective, not zooming in on its constituent elements. Competences are considered as "dispositions", which, in a nutshell, are objectified properties inherent in an object (or agent) which may manifest themselves in certain situations through events (or actions). (They are also called "powers" in the philosophical literature [21].)

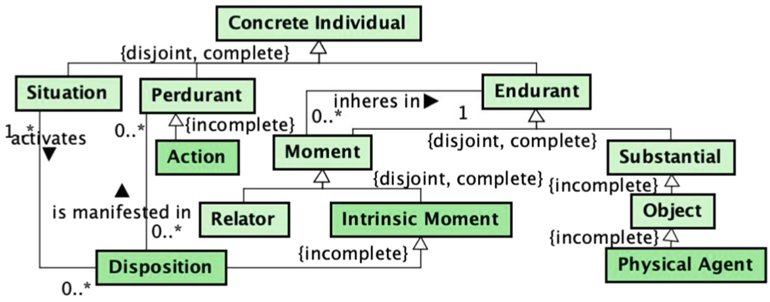


Fig. 1. UFO Fragment (used concepts highlighted)

The domain-independent elements we reuse from UFO are shown with a UML class diagram in Fig. 1. Concrete individuals are partitioned into perdurants (also called *events*), endurants and situations. *Perdurants* are *individuals* who occur in time (i.e. activities, actions, tasks, processes). *Endurants* are individuals that persist in time while retaining their identity (i.e. people, organizations, projects, cars). *Endurants* include *moments* and *substantials*. *Moments* are reified properties that *inhere in* an *endurant* (termed its bearer), on which they are existentially dependent. As endurants, they have a lifecycle of their own, and can be created, destroyed or otherwise change qualitatively in time.

Of special interest to us in this work are those moments called *dispositions*. *Dispositions* are intrinsic moments that can be manifested through the occurrence of *events* (possibly agents' actions, such as Anna's speaking English). In *situations* where *dispositions may manifest*, they are said to be "activated" (e.g., when a magnet is close to some ferrous material, or when Anna is prompted to introduce the topic of a meeting). The literature discusses a number of important features of dispositions; they may fail to manifest when enabled, they may be manifested in tandem with other dispositions in complex events, they may reinforce or cancel each other [21,22]. Reifying (i.e., objectifying) them puts them at the center of our efforts as first-class citizens. As endurants, they can themselves bear moments, and change qualitatively while retaining their identity through time [12].

Figure 1 also shows a few concepts from the UFO-C layer of UFO [14] which are relevant here. *Physical agents* are those objects that, in constrast with non-agentive *objects* bear *intentional moments* such as *beliefs desires* and *intentions* (omitted from the figure). They are capable of *actions*, which are those events that are performed intentionally by *agents*.

3.2 Elements of Competences

In [5], *personal competences* are defined as *dispositions* inhering in a physical agent, with tasks, actions, or behaviors considered as manifestations of those dispositions (*competence manifestations*). The *competence context* that activates the *personal competence* is considered a kind of *situation*. We take this view as a starting point, and extend it to incorporate the internal elements of competence (skills, knowledge, attitudes, etc.), anchoring these elements in the foundational concepts.

Skills and Competences as Human Capabilities. Regarding the skill concept, there are some parallels between it and the definitions of competence [26]. Some authors even argue that such concepts have the same meaning in essence. Competence is conceptually considered a skill sub-type in some cases [29]. Even among those who believe that competence and skill are distinct concepts, there are many similarities between them. In this sense, both are regarded as human abilities that enable satisfactory task performance [26]. Thus, both skill and competence are inherent abilities in a person, the bearer, that enables the performance of specific actions' types. That is, they represent an individual's "know-how". Aside from this fundamental similarity, there are other comparable features in the definitions of these concepts. Both are abilities that can be learned (formally or informally) and developed through practice [19,30]. In this sense, skill and competence can be used to learn new abilities via the transfer mechanism [19,26,30]. In terms of structure, there are also some similarities between skills and competences. Both have a hierarchical structure, according to some authors [26,29]. As a result, they can be aggregated or combined at various levels. Thus, simpler skill/competence forms more complex skill/competence. As a consequence, the complexity of skill and competence can also vary. Another similarity between these concepts is their relationship with the context. Both are associated with a context, environment, area, or domain [26,29]. In this regard, competence and skill can be more generic (domain-independent) or more specific (domain-dependent) [26,29]. Skills and competences frequently rely on favorable conditions to manifest. That is, skills and competences depend on other properties (internal or external) to manifest themselves more effectively. Knowledge, mental states, attitudes, feelings, and so on can all aid in the proper manifestation of a skill or competence, for instance. Finally, in addition to the aforementioned similarities, some authors argue that skills and competences involve similar domains of an individual. According to them, both are related to the bearer's affective, social, physical (or operational), cognitive, and meta-cognitive domains [17,26]. In order to capture the common features of skills and competences, we introduce the notion of Human Capability as shown in Fig. 2. Skills and competences are considered sub-types of the more general notion of Human Capability, which in turn are dispositions inherent in a Person.

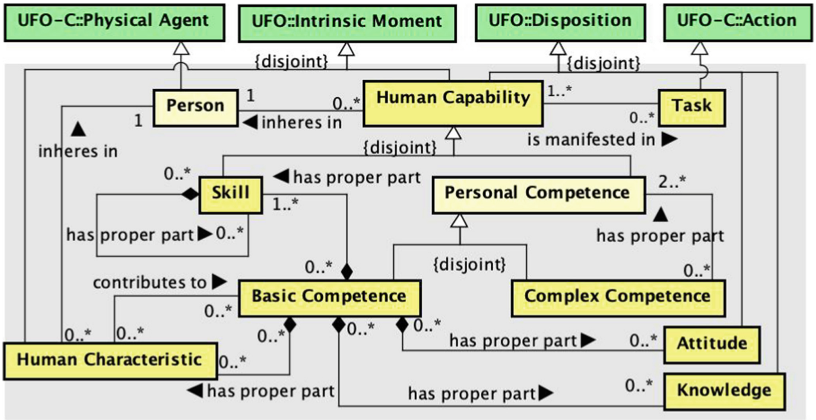


Fig. 2. Competence element's definition (new concepts highlighted)

Human Capability encompasses all human abilities, from those that are innate (inherited) to those that can be learned (formally or not) and is manifested through a task (an action with some goal or a work unit). The task in this work is regarded as the smallest unit of labor. In other words, it is a discrete unit of work that contributes to the production of an output or the achievement of a goal [27].

Competences Versus Skills. In this current work, the main distinction between skills and competences is the structural aspect as revealed by the specialized whole part relations in Fig. 2. Competences are formed by knowledge, skills, attitudes and other human characteristics, whereas skills are formed only by other simpler skills. In this work, a competence is made up of at least one skill that is linked to one or more other competence's elements. Based on [26], another adopted criterion to distinguishes skill from competence is the mode of manifestation. Competence is associated with one or more complex tasks, whereas skill is associated with a simple task (basic unit of work), as in [26]. The whole part relations put forward a hierarchical view of competences. Complex Competences are those composed of other competences; Basic Competences are those at the bottom of the competence decomposition hierarchy, whose elements follow the KSAO model. Skills can also be structured hierarchically for a comprehensive conceptualization.

Knowledge. In this context, knowledge is defined as a justified true belief [16]. Knowledge, while assisting in the realization of skills, differs from skills in that it is a static entity registered in the individual's memory. It is related to the person's knowledge of information, facts, and concepts. It is produced as a result of internal (mental) information processing. Skills, on the other hand, manifest themselves through (cognitive or psycho-motor) tasks and are developed through practical experiences. In this way, knowledge, despite representing external facts or concepts, is existentially dependent on the bearer. Individual knowledge, as a type of belief, can be considered a subjective entity that is difficult to measure or quantify, despite the fact that it may have attributes. Furthermore, knowledge is a mental property that is inherent in the individual that can lead to action. It is manifested alongside other forms of dispositions such as skills to manifest itself in tasks, forming *reciprocal* or *mutual* activation partners [22].

Attitudes. Despite the fact that it is manifested through actions, gestures, postures, and so on, attitude differs from skills in that it is not manifested through tasks. Attitude, on the other hand, can be task-related. For example, a responsible attitude can be present during a developer's completion of the task of fixing a bug in software; an empathetic attitude can be present during the task of negotiating project scope with the client. Attitude in the context of this work is considered a sub-type of *Disposition*, because it is a proclivity to act and behave. Again, like knowledge, it is manifested alongside other forms of dispositions, forming *reciprocal* or *mutual* activation partners [22] with skills and knowledge.

Other Human Characteristics. Human characteristics are particular to an individual and form part of their personality. Some of these characteristics include objective (or measurable) attributes (e.g., sex, age), while others are subjective (non-measurable), such as the individual's motivations, worldviews, values, and beliefs. As previously stated, such human characteristics are regarded in this work as a subtype of *Intrinsic Moment*. A concept of this type includes both objective (measurable) and subjective human characteristics and could be part of a personal competence. As an intrinsic moment, human characteristics can be categorical (e.g. age, gender, etc.) or dispositional (e.g. personality traits). Based on the categorical base of the disposition, the former contributes to competence formation [6]. As illustrated in Fig. 2, the latter are a proper part of the competence. Furthermore, some human characteristics can be included in the competence context, activating the competence manifestation.

4 Well-Founded Competence Representation

Based on the ontological distinctions presented in the previous section, we define an ArchiMate language pattern in this section, with no changes to the Archi-

Mate metamodel. The well-founded representation is proposed to allow modeling of competence and its elements (knowledge, skill, attitude) in the EA context, supplementing the representation proposed in [5]. To summarize, [5] represents: (i) *Person* with a *Business Actor*; (ii) *Personal Competence* with a *Capability* Element related to a *Business Actor*; (iii) *Competence Manifestation* with any ArchiMate *Behavioral Element* related to a *Capability* Element, and; (iv) *Competence Context* with *Plateau* or *Location* Elements related to a *Capability* Element.

Based on this, the elements of competence are represented as follows. *Skill* is represented by the *Capability* Element related to: (i) another *Capability* Element representing a *Personal Competence* with a composition relation, or (ii) a *Business Actor* representing a *Person* (the skill bearer). *Knowledge* is represented by a *Meaning* Element that is linked to a: (i) *Capability* Element that represents a *Personal Competence*, or (ii) *Business Actor* that represents a *Person* (the knowledge bearer). *Attitude* is represented by the *Value* Element related to: (i) *Capability* Element, which represents a *Personal Competence*, or (ii) *Business Actor*, which represents a *Person* (the attitude bearer). Aside from the fact that the ontological model does not establish any relationship between the competence's elements, they could also be linked using the *ArchiMate*'s *Association* relation. This is an appropriate way of representing the relation among competence's element.

Figure 3 depicts a high-level overview of the language pattern. As shown in Fig. 3, John (Person) is a junior developer that works as a front-end developer at a software organization. In this context, he possesses the (complex) competence of full-stack development, which is made up of two (basic) competences: back-end development and front-end development. As shown, Java and SQL coding skills are among John's back-end competences. John's front-end competence includes *HTML*, *CSS*, and *Javascript* (JS) coding skills, besides user interface (UI) prototyping one. Its competence also includes John's responsibility attitude besides knowledge of quality criteria and of UI heuristics.

As the *Plateau* element indicates, the model in Fig. 3 represents John's current situation. Because it represents the current state of individual competences in the organization, this type of model is useful in the Competence Mapping stage of the CM process. On the other hand, it is also possible to represent through this

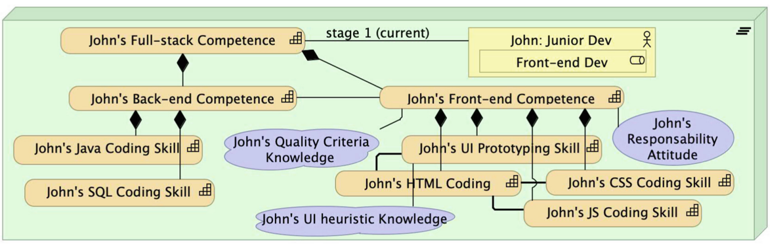


Fig. 3. Skill, knowledge, and attitude representation (current stage)

language pattern the desired competences of the individual for the organization. For this, it necessary specify through a *Plateau* element. This type of diagram is useful for better understanding the desired competences during the Competence Identification step. Figure 4 depicts this situation, where John's desired competences for organization are represented. As shown, it is desirable for the organization that John evolves technically (hard skills) and gains specifically the coding review skill (highlighted one). As a result, John will be able to review the web form code for user story 23 (highlighted one).

In terms of the manifestation of this specific competence, the skills are in charge of the completion of basic tasks (discrete units of works). As shown, *HTML* and UI prototyping skills are responsible for the coding of web form fields coding and web form prototyping manifestation, in context of user story 23 (US23). While these skills are associated with basic task manifestation, John's front-end competence is related to the manifestation of the entire web form development process, which is related to US23.

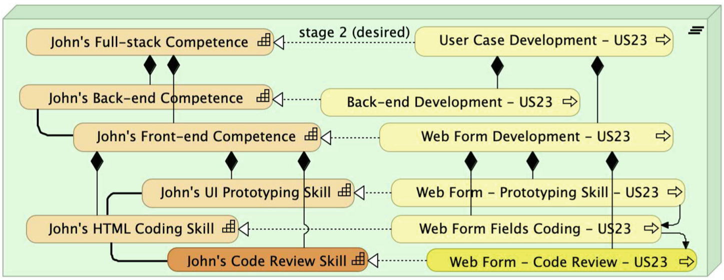


Fig. 4. Skill representation (desired stage)

Following Competence Identification and Mapping, one important step in the CM process is to compare the desired and current organizational states using the Gap Analysis activity. This comparison can concentrate on various aspects, such as the professional's technical or behavioral evolution. Figure 5 depicts a comparison focusing on John's soft skills evolution. In this case, in order for John to advance his front-end competence, he must acquire new soft skills such as communication and problem-solving skills, in addition to the previously mentioned code review skill. In addition to these skills, it is wanted that John will develop a collaborative attitude toward code review (to assist in reviewing college codes). It is also desired that he gain new knowledge about coding best practices and code review techniques.

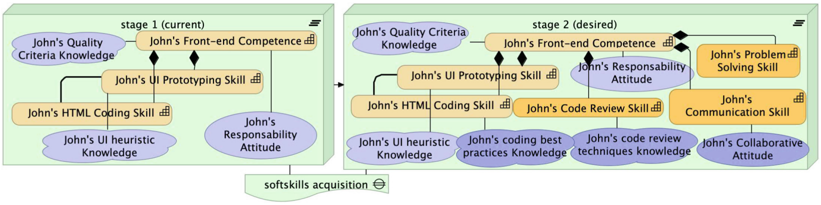


Fig. 5. Competence elements detailing in gap analysis

5 Related Works

Competence models range from simple competence representations to more semantically rich and sophisticated representations [25]. Competence management approaches began to use standardized models, such as XML-based ones, to support specific technological tasks such as data integration and exchange. These models then evolved into more complete conceptual models. Recently, ontology-based models have become more prevalent in CM approaches, incorporating more semantics into competence models [15]. They have been used for a variety of purposes, the majority of which are related to business and education. Some of these works are discussed below.

In the ontology of Zaouga et al. [31], knowledge and skill are considered subtypes of competence rather than elements. The ontology does not cover attitudes. In its place, the authors use the behavior concept with a similar meaning. Paquette's ontology [24] also includes skill and knowledge, but not attitude. In this case, knowledge and skill are components of competence. [24] also connects the concepts of skill and knowledge. Skills are applied to knowledge entities in this case. Skills are classified in the ontology based on taxonomies and complexity levels, and they are also measured using indicators. Miranda et al.'s [20] also incorporate knowledge, skill, and attitude into their model. In its structure, competence consists of these elements. As stated in [20], knowledge and skill are also related concepts. This ontology takes into account not only skill classification but also knowledge and attitude.

In contrast to those works, the proposed ontology treats competence and skill as a compound entity. As a result, they could be represented at various levels of abstraction. They are dispositional concepts, and types of Human Capability, that can manifest themselves through tasks. Knowledge and attitudes share this dispositional nature and can manifest together with tasks through actions, posture, and so on. Aside from the detailed and well-founded representation of competence, the link between competences and Enterprise Architecture is another distinguishing feature of this work.

Some other works such as [3, 23] have also explored foundational ontologies in EA modeling. Both employ UFO to conduct ontological analyses of two concepts closely related to competence: capability and service. [23], for example, views service delivery as the manifestation of competences. [3], on the other hand, conducts an ontological analysis of Capability and is also related to the concept of Competence. [3] briefly discusses the definition of competence based on capability; in the current work, we adopt and expand on that analysis. As discussed here, competences can be placed in the so-called capability bundles [3], connecting individual-level capabilities (competences) with organizational capabilities.

6 Final Remarks and Discussion

The study presented in this paper aimed to improve competence modeling in the context of Enterprise Architecture by using a reference ontology as a semantic foundation. The understanding of skills, knowledge, attitudes, and other characteristics allowed us to zoom in on an individual competence, allowing for a detailed competence representation in the context of Enterprise Modeling.

We investigated the support of Competence Management activities with Enterprise Architecture models with the goal of improving personal competence understanding. From the standpoint of competence detailing and decomposition, the proposed competence representation strategies make it easier to implement Competence Management in EA. The model representation using *ArchiMate*, on the other hand, contributes a set of possibilities to enhancing the Competence Management practice. This distinguishes the current work from other ontology-based competence works in the literature.

As a result, the proposed representation can aid in essential Competence Management (CM) activities such as competence mapping, identification, and gap analysis. In this sense, the proposed representation patterns facilitates CM activities by visualizing modeling competences from various perspectives. It enables the detailing of individual competences in these various representations, assisting with a deeper comprehension of the individual skills, knowledge, and attitudes that comprise these competences. This detailed vision aids in many CM activities such as competence comparison, planning, and assessment, to name a few.

Future research could open up on the concept of competence by investigating how the competences of different individuals can be combined to form organizational and collective capabilities. This study would delve deeper into how organizational capabilities emerge from personal competences. Capabilities are not created by simply combining competences. The combination of high proficiency and competence does not guarantee the formation of a high-performance team. It is a very complex and difficult subject that deserves more investigation. In this regard, we see an opportunity to incorporate *General System Theory* (GST) concepts into the ontological foundation in order to better represent the phenomena of evolution, emergence, and composition in the context of Enterprise Architecture. In this context, we see an opportunity to study theories of dispositions in order to better understand how competences can be related and combined.

We also see the need to develop case studies should be used to validate the proposed competence representation patterns. Although ontological analysis provides the foundation for a well-founded representation (as used here, the foundation incorporates advances in Formal Ontology, Philosophical Logics, Philosophy of Language, Linguistics, and Cognitive Psychology [12]), the pragmatics of a representation in its usage context should be thoroughly assessed. Efforts in this sense have already been made for other UFO-based representation schemes, such as [11,23].

Another area of future research concerns the relationship between competences and other *ArchiMate* perspectives, such as Motivation Elements. In this case, the ontological analysis could include other UFO concepts related to intentions, such as Goal and Proposition, which are related to the organization's strategic goals [2].

Acknowledgment. This study was supported in part by CNPq (313687/2020-0), FAPES (281/2021) and the DSYNE INTPART network (Research Council of Norway project number 309404).

References

1. Altmann, T.K.: Attitude: a concept analysis. In: Nursing forum. vol. 43, pp. 144–150. Wiley Online Library (2008). https://doi.org/10.1111/j.1744-6198.2008.00106.x
2. Azevedo, C.L.B., Almeida, J.P.A., van Sinderen, M., Quartel, D., Guizzardi, G.: An Ontology-Based Semantics for the Motivation Extension to ArchiMate. In: 15th IEEE Int. EDOC Conference. pp. 1–10. IEEE (2011). https://doi.org/10.1109/EDOC.2011.29
3. Azevedo, C.L.B., Iacob, M., Almeida, J.P.A., van Sinderen, M., Pires, L.F., Guizzardi, G.: Modeling resources and capabilities in enterprise architecture: A well-founded ontology-based proposal for ArchiMate. Inf. Syst. **54**, 235–262 (2015). https://doi.org/10.1016/j.is.2015.04.008
4. Brown, J.S.: Learning in the digital age. In: The Internet and the university: Forum. vol. 71, p. 72 (2001)
5. Calhau, R.F., Azevedo, C.L.B., Almeida, J.P.A.: Towards Ontology-based Competence Modeling in Enterprise Architecture. In: 25th IEEE Int. EDOC Conference (EDOC 2021). IEEE (2021). https://doi.org/10.1109/edoc52215.2021.00018
6. Choi, S., Fara, M.: Dispositions. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Spring 2021 edn. (2021)
7. Draganidis, F., Chamopoulou, P., Mentzas, G.: A semantic web architecture for integrating competence management and learning paths. J. Knowl. Manag. **12**(6), 121–136 (2008). https://doi.org/10.1108/13673270810913667
8. Draganidis, F., Mentzas, G.: Competency based management: a review of systems and approaches. Inf. Manag. Comput. Secur. **14**(1), 51–64 (2006). https://doi.org/10.1108/09685220610648373
9. Esposto, A.: Skill: An elusive and ambiguous concept in labour market studies. Australian Bulletin of Labour **34**(1), 100–124 (2008), https://EconPapers.repec.org/RePEc:fli:journl:26132
10. Gangani, N., McLean, G.N., Braden, R.A.: A competency-based human resource development strategy. Perform. Improv. Q. **19**(1), 127–139 (2006)

11. Griffo, C., Almeida, J.P.A., Guizzardi, G.: Conceptual Modeling of Legal Relations. In: 37th Int. Conf. Conceptual Modeling (ER 2018). Lecture Notes in Computer Science, vol. 11157, pp. 169–183. Springer (2018). https://doi.org/10.1007/978-3-030-00847-5_14
12. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.S.: Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. Applied Ontology (Online) **10**, 259–271 (2015). https://doi.org/10.3233/AO-150157
13. Guizzardi, G.: Ontological Foundations for Structural Conceptual Models. No. 15 in Telematica Instituut Fundamental Research Series, Telematica Instituut, The Netherlands (2005)
14. Guizzardi, G., Wagner, G.: Towards ontological foundations for agent modelling concepts using the unified fundational ontology (UFO). In: Agent-Oriented Information Systems II, pp. 110–124. Springer, Berlin Heidelberg (2005). https://doi.org/10.1007/11426714_8
15. HR-XML Consortium: Competencies 1.0 (measurable characteristics) recommendation 2001-oct-16 (2016), http://xml.coverpages.org/HR-XML-Competencies-1_0.pdf
16. Ichikawa, J.J., Steup, M.: The Analysis of Knowledge. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Summer 2018 edn. (2018)
17. Le Deist, F., Winterton, J.: What is competence? Hum. Resour. Dev. Int. **8**(1), 27–46 (2005). https://doi.org/10.1080/1367886042000338227
18. Maze, J.R.: The concept of attitude. Inquiry **16**(1–4), 168–205 (1973). https://doi.org/10.1080/00201747308601684
19. Messick, S.: The psychology of educational measurement. J. Educ. Meas. **21**(3), 215–237 (1984). https://doi.org/10.1111/j.1745-3984.1984.tb01030.x
20. Miranda, S., Orciuoli, F., Loia, V., Sampson, D.: An ontology-based model for competence management. Data & Knowledge Engineering **107**, 51–66 (2017). https://doi.org/10.1016/j.datak.2016.12.001
21. Molnar, G., Bradley, N.: Powers: A study in metaphysics. Clarendon Press (2003)
22. Mumford, S., Anjum, R.: Getting Causes from Powers. Oxford University Press (2011)
23. Nardi, J.C.: A Commitment-based Reference Ontology for Service: Harmonizing Service Perspectives. Ph.D. thesis, Federal Univ. of Espírito Santo (2014)
24. Paquette, G.: An ontology and a software framework for competency modeling and management. Educational Technology and Society **10**(3), 1–21 (2007), http://www.ifets.info/journals/10_3/1.pdf
25. Rezgui, K., Mhiri, H.: Modeling competencies in competency-based learning: Classification and cartography. In: 2018 JCCO Joint Intl Conf ICT in Education and Training, Intl Conf Computing in Arabic, and Intl Conf Geocomputing (JCCO: TICET-ICCA-GECO). IEEE (Nov 2018). https://doi.org/10.1109/icca-ticet.2018.8726208
26. Rodrigues, M., Fernández-Macías, E., Sostero, M.: A unified conceptual framework of tasks, skills and competences. Tech. rep., JRC Working Papers Series on Labour, Education and Technology (2021)
27. Rodriguez, D., Patel, R., Bright, A., Gregory, D., Gowing, M.K.: Developing competency models to promote integrated human resource practices. Hum. Resour. Manage. **41**(3), 309–324 (2002). https://doi.org/10.1002/hrm.10043

28. Sampson, D., Fytros, D.: Competence models in technology-enhanced competence-based learning. In: Handbook on Information Technologies for Education and Training, pp. 155–177. Springer, Berlin Heidelberg (2008). https://doi.org/10.1007/978-3-540-74155-8_9
29. Westera, W.: Competences in education: A confusion of tongues. J. Curric. Stud. **33**(1), 75–88 (2001). https://doi.org/10.1080/00220270120625
30. Wood, R., Power, C.: Aspects of the competence-performance distinction: Educational, psychological and measurement issues. J. Curric. Stud. **19**(5), 409–424 (1987). https://doi.org/10.1080/0022027870190503
31. Zaouga, W., Rabai, L.B.A., Alalyani, W.R.: Towards an ontology based-approach for human resource management. Procedia Computer Science, vol. 151, pp. 417–424. Elsevier (2019). https://doi.org/10.1016/j.procs.2019.04.057

A Knowledge-Graph Based Integrated Digital EA Maturity and Performance Framework

Nujud Alsufyani(✉) and Asif Qumer Gill

University of Technology Sydney, Ultimo, NSW 2007, Australia
nujud.alsufyani@student.uts.edu.au, asif.gill@uts.edu.au

Abstract. Digitalisation is gaining considerable attention from enterprises aiming to improve their maturity and performance using digital enterprise architecture (EA). However, the challenge is how to assess and enhance often disconnected but related digital EA maturity and performance outcome elements. To address this research challenge, this paper proposes an integrated digital enterprise architecture maturity and performance (DEAMP) ontology. This ontology aims to assist organisations in understanding and assessing their digital maturity (DM) level and associated performance outcomes. This is important to understand whether there is a positive change in the performance level (effect) through improving a DM level (cause). A design science research (DSR) method, along with a skeletal enterprise modelling approach, have been used to develop and evaluate the proposed DEAMP ontology. Further, this ontology is represented as a knowledge graph (KG), which can be tailored and used by researchers and practitioners to capture and process DM and performance data for better outcomes as appropriate to their enterprise context and scope.

Keywords: Digital maturity · Digitalisation · Performance · Ontology · Knowledge graph · Enterprise architecture

1 Introduction

Digital technology opportunities and threats led many organisations to go through a digital transformation journey to achieve their profitability, growth or competitiveness goals. Digitalisation or digital transformation is a process of change that involves leveraging emerging digital technology [1] by individuals and enterprises [2], ecosystems, and industries [3]. Digitalisation-related changes may impact or improve performance outcomes [4]. Also, digitalisation maturity levels could be associated with strategic performance gain [5]. However, organisations often encounter challenging situations in understanding and measuring the impact of digitalisation [4] maturity on organisation performance [6].

This paper is a part of ongoing research investigating the integration of digital maturity (DM) and performance outcomes. This ongoing research builds on our earlier systematic reviews, which indicated the need and provided motivation and foundation for integrating often two disjoin but related digital and performance elements [6, 7]. This

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 214–229, 2023.
https://doi.org/10.1007/978-3-031-26886-1_13

earlier work also draws our attention to the need for a theoretical and practical holistic understanding of the integration of DM and performance elements for informed decision-making to uplift digitalisation for performance outcomes. As a result, this research aims to integrate the level of DM and related performance outcomes for well-informed decision-making from the holistic enterprise architecture (EA) perspective. The EA lens provides a layered approach to integrate the DM and performance elements [8]. Also, adopting the EA approach may help understand the maturity of the digital enterprise design components (e.g. people, process, capability), their relationships to each other and performance outcomes [9, 10].

In this paper, we propose an integrated DM and performance ontology, which is represented as a knowledge graph (KG) to link and model the DM levels and performance outcomes. As ontology defines explicit knowledge understanding for better communication and analysis [11], a KG is a knowledge representation of entities, relationships and their instances to capture data and generate insights related to digital maturity and performance [12]. This will help decision-makers who are interested to understand integrated DM and performance outcomes. In summary, KG can be used by the modelling platform developers to support the capturing and processing of the integrated digital EA maturity and performance elements and data. The integration of DM to performance outcomes is important to identify the maturity and related performance gaps, which will then be used to create actions and a roadmap to address those gaps.

The structure of this paper is as follows. Firstly, it discusses the research background and related work. Then, it describes the research method and the development of the digital enterprise architecture maturity and performance (DEAMP) ontology. Finally, before the discussion and conclusion, it discusses the evaluation of the DEAMP ontology's practical relevance and applicability using a KG example.

2 Research Background and Related Work

2.1 Digital Maturity

There are several definitions of DM, and no universally accepted single definition [13]. Here, this paper focuses on the change in maturity levels. The maturity term generally represents an anticipated reality or change to achieve desirable outcomes [14]. Digital maturity, on the other hand, is closely associated with the degree of digital transformation, which could be associated with better organisational performance [13]. DM can mirror the outcomes of digital transformation efforts from technological and managerial aspects [15]. It can be used to evaluate the current and target levels of digitalisation or maturity with a view to navigate the gaps and intended investment decision-making [16]. Achieving a higher level of maturity needs to be aligned with an organisation's strategic goals and key performance indicators (KPIs) [13]. Commonly, maturity models are multilevel frameworks that define different organisational capabilities and development levels [17].

2.2 Organisational Performance

Organisational performance represents the outcomes of effectively achieving the organisational goals and objectives [18]. Those outcomes could cover financial and non-financial outcomes [19]. Hence, there are different methods to model organisational performance outcomes, such as the balanced scorecard (BSC) [20] and strategic measurement analysis and reporting technique (SMART) [21]. On the one hand, the BSC measures customer, internal, innovation/learning and financial performance in a chain of cause-and-effect, leaving the performance measurements to be derived from the organisation's strategy [20]. On the other hand, SMART is an operational performance-oriented framework that consists of four levels. It is a pyramid of objectives and measures that integrate strategy with operational performance-focused measurements. In this paper, along with EA, we used the results and determinants framework [22] as a theoretical lens to study and model performance indicators (PIs) [6]. This framework has two main types of performance elements: results (lagging factors as financial, competitiveness performance measurement dimensions) and determinants (leading factors as resource utilisation, innovation, flexibility, and quality performance measurement dimensions). Each performance dimension has its related types of measures. For instance, financial performance targets profitability, liquidity and market ratios. Thus, the results and determinants framework has been selected because it provides technology-independent six generic dimensions or classes for performance measurement: Competitiveness, Financial, Quality of Service, Flexibility, Resource Utilisation and Innovation. Also, it provides insight into what each performance dimension can measure.

2.3 Enterprise Architecture

EA provides a holistic view of an organisation's architecture design and implementation plan [9, 10] to improve performance outcomes [23–25]. There are several EA frameworks. For example, Zachman's [26] framework provides a generic ontology. In contrast, The Open Group Architecture Framework (TOGAF) [27] provides a generic architecture development method. However, Zachman and TOGAF initiated in a traditional architecture methods and ontologies context. Thus, this study uses the adaptive EA [28] meta-framework, which originated in the digitalisation context and digital ecosystem, and can be used to design situation-specific EA frameworks and capabilities [28]. The adaptive EA has been used because it provides broader coverage and additional layers such as interaction, facility and environment layers when compared to existing frameworks (e.g. TOGAF, Zachman). Also, it consists of explicit performance outcome-driven adaptive EA design layers [29] that can be integrated to attain strategic goals and objectives in the context of digital enterprise [30]. Adaptive EA was used because it provides a systematic layered approach and elements for designing and evolving digitally-enabled enterprises. Also, EA driven approach has been used because it provides a holistic illustration of the organisation's architecture design, underpinning components and their digital transformation roadmap and implementation planning [9, 10].

Adaptive EA [28] offers guidance on six architecture layers: Interaction, Human, Technology, Environment, Facility, and Security layers. Each layer has its underpinning concrete elements as follows: (1) the interaction layer includes the actors and their interactions via different digital touchpoints, channels, and overall journey experience, (2) the human layer covers the business, information, social and professional architecture domains, (3) the technology layer covers infrastructure, application, data and platform architecture domains, (4) the security layer deals with the security concern of every other element or factor across other layers, (5) environmental layer includes PESTEL (Political, Economic, Social, Technological, Environmental and Legal) elements and (6) the facility layer covers heating, ventilation, air conditioning (HVAC), spatial, energy and ancillary elements. Besides these six EA layers, adaptive EA achieves its adaptability through three main activities to identify and analyse changes and then decide the appropriate response to those changes for adaptations across EA layers and elements. Figure 1 illustrates the integration of performance and DM from the Adaptive EA perspective.

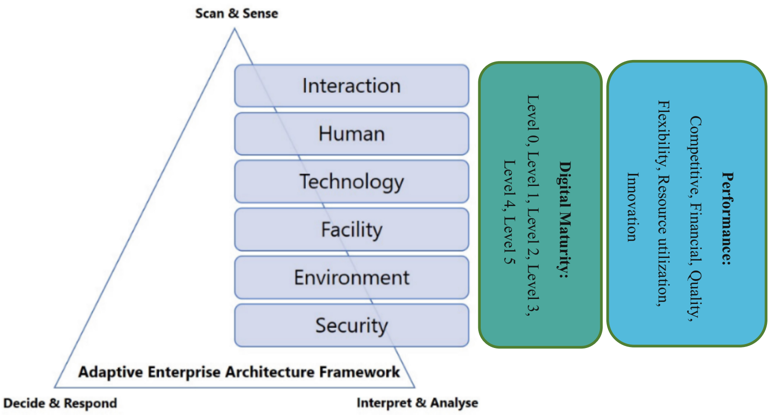


Fig. 1. Adaptive EA layers [8, 28, 31] integrated with six levels of DM [7] and Results and Determinants performance dimensions [22]

2.4 Integrating DM and Performance from EA Perspective

Organisations undergoing digital transformation or digitalisation are required to frame their digitalisation vision to capture the digitalisation need, strategy and future outcomes [32]. Consequently, this may require unpacking the influence of digitalisation on organisational architecture aspects and their performance [33]. The interdependence of digitalisation, strategies and other organisational elements indicates that digitalisation may not clearly specify the complex mechanisms of organisational performance impact [33]. This shows an important knowledge gap or disconnect among digitalisation maturity, organisational performance, strategies and other organisational aspects. It has been reported that assessing maturity and linking it to organisational performance is an important consideration [34]. This is because a higher level of maturity may link to higher performance outcomes [13]. Also, on another note, adopting the EA approach may help understand the organisation's design components, their relationships to each other and performance factors [35]. Thus, we propose to investigate and link maturity level and organisation performance outcomes using the EA perspective. This will help to address a lack of linking and understanding between DM level and organisation performance outcomes.

To unpack and analyse the impact of DM levels on performance outcomes from EA perspective, we used an adaptive EA [28] framework. It offers relevant layers for conceptualising the digital enterprise, which is appropriate to the scope of this study. Also, it originated in the context of digitalisation and the digital ecosystem. Since the adaptive EA does not offer detailed performance outcomes indicators, we adopted the results and determinants framework [22] to cover organisational performance and map it to the adaptive EA layers (Fig. 1). On the other hand, DM levels were adopted from our previous study as it was conducted from EA perspective [7]. Other frameworks or theories might also be used to form similar studies; however, this study is limited to two relevant frameworks (1) adaptive EA and (2) results and determinants. Future studies might use other appropriate frameworks, theories, and perspectives relevant to their study context and scope.

2.5 Ontology and Knowledge Graph

Ontology is "an explicit specification of a conceptualisation" [36]. This conceptualisation can include the definition of a set of concepts, their meanings and relationships [11], as they are the fundamental constructs of conceptual modelling [37]. Generally, ontology provides a consensual understanding of a field for better knowledge communication [38]. Thus, this paper proposes a DEAMP ontology to conceptualise DM and performance elements for explicit knowledge understanding and their integration. It is anticipated that this will lead to effective communication when assessing and improving DM for desired performance outcomes [11]. DEAMP ontology can be represented using several approaches, such as the KG used in the paper, because of its flexible nature in representing real-world entities and their relationships [39]. KG congregates and models real-world knowledge in a network of entities (nodes) and relationships (edges) connecting different entities [12]. As such, ontology displayed in a graph can be considered as a KG if it is populated with instances. Thus, we used the KG because it is useful to represent the connected elements and their instances, such as the integrated performance and maturity elements [39]. To illustrate the ontology, the graph-based approach [40] provides knowledge as labelled nodes representing entities (concepts), and labelled arrows as the relationship between nodes. In summary, using the KG-based approach is deemed appropriate to model the integrated DM and performance ontology elements and their relationships. This also complements the current efforts around KG for conceptual modelling [41, 42].

3 Research Method

This research is conducted using the design science research (DSR) method [43], which is helpful to design, build and evaluate the proposed ontology as an artefact. DSR is used because it enables the development and evaluation of novel artefacts for a purpose [44, 45]. On another note, DSR can provide a systematic foundation to tackle complex organisational design-related problems [46]. To develop the DEAMP ontology, we followed a skeletal enterprise modelling approach [11]. This approach is found to be suitable for integrating two different but relevant concepts of DM and performance. While other approaches such as enterprise ontology can address building EA ontology [11], however, here the focus is on integrating DM and performance elements across EA layers.

The applied DSR includes five steps (Fig. 2). Firstly, we identified the research problem based on the research background and related work. Secondly, we applied creative and analytical thinking to produce an initial tentative design of the DEAMP. Detailed research was conducted to develop the DEAMP design. In the evaluation step, the DEAMP was evaluated. The development and evaluation are iteratively performed to evaluate and evolve the DEAMP. Finally, the outputs and knowledge contributions are identified and reported in the last step to conclude the research.

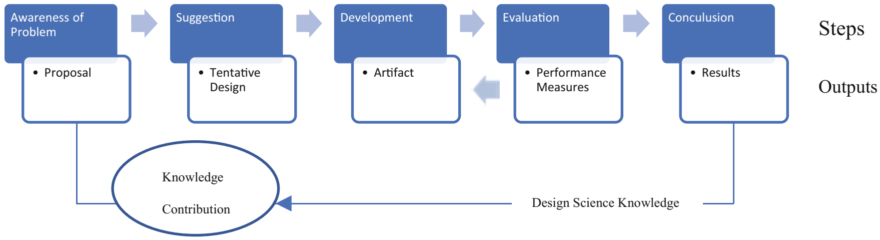


Fig. 2. DSR approach

As indicated earlier, in DSR, this research uses the skeletal enterprise modelling method [11] to construct the DEAMP ontology. It consists of four steps: purpose, scope, build and evaluate/revise [11]. The first and second steps include reasons to build the ontology and set of structured concepts to satisfy identified requirements. The third step requires producing, arranging and structuring concepts' definitions to build the ontology. The fourth step is to evaluate the developed ontology using pre-defined criteria. In summary, the DEAMP ontology is iteratively developed. We organised this ontology development into four iterations. In the first iteration, we reviewed existing studies on DM models and performance outcomes related to digitalisation to identify the key concepts and their relationships. We detailed those concepts in the second and third iterations, and modelled them using the graph modelling approach [40]. Then, we evaluated the ontology using an example scenario (fourth iteration).

4 DEAMP Ontology

This section focuses on the first three iterations that include the purpose of the ontology, key concepts, and the DEAMP graph. In the first iteration, we reviewed the existing literature about DM models and digitalisation performance outcomes separately, as reported in earlier published studies [6, 7]. We synthesised 30 different DM models using a theoretical lens (adaptive EA) [28]. Also, we synthesised digitalisation performance outcomes using two theoretical lenses (adaptive EA [28], results and determinants framework [22]). Subsequently, we integrated the main concepts and relationships extracted from our previous studies [6, 7] to develop the DEAMP. The EA design-driven DEAMP can be used to assess DM and performance outcomes across adaptive EA layers and their underpinning elements (Fig. 3).

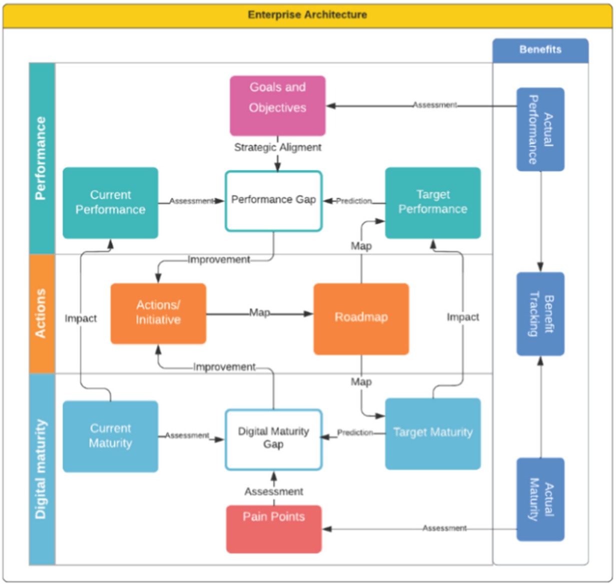


Fig. 3. DEAMP conceptual model

The DEAMP is organised into four areas: performance, digital maturity, actions and benefits. Performance assessment can be done to define the performance gaps in alignment with the organisation's goals and objectives. Similarly, a DM assessment can be conducted to determine the DM gaps based on the underpinning pain points across the EA design. As the current and target DM impact current and target performance, these assessments can then lead to initiate actions to fill the integrated performance and maturity gaps to uplift DM for performance gain. These actions can be included in a roadmap to map to target DM and performance, which is a sequence of actions, timeline, their dependency and priority. Finally, benefits monitor the actual DM and performance outcomes resulting from implementing the actions roadmap (post actions roadmap's implementation). It tracks the resolved pain points and realised goals and objectives according to the desired target maturity and performance levels.

In the second iteration, we adopted a graph-based modelling approach [40] to represent the DEAMP concepts, properties and their relationships. This will serve as an ontology for the EA-driven DM and performance assessment. We used Neo4j graph database to implement the DEAMP graph as a first step to demonstrate the applicability of the proposed work. We used Neo4j because it is a scalable and robust open-source native graph database [47]. Neo4j stores data as nodes and edges that represent entities and relationships. Both nodes and edges can have properties in the form of key and value pairs.

Here, we first focus on the taxonomy aspect of the graph model to highlight the structure of EA-driven DEAMP. The initial graph-based model shows that architecture has an architecture domain and relevant elements. Then, the architecture element is associated with a set of PI (performance indicator) gap, DM gap, objective, action, pain point and benefit. (see Fig. 4).

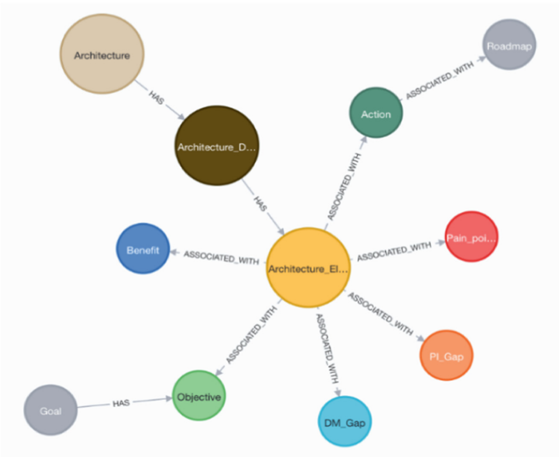


Fig. 4. Initial DEAMP graph-based model

In the third iteration, a full graph was produced (Fig. 5) with concepts and their definitions (Table 1). This captures the ontology aspect of the DEAMP. Additional concepts were added to capture the meanings of the key concepts as classes and attributes. For example, to detect the gap in performance, PI gap uses a set of current and target performance (CPI, TPI), and the same logic goes for DM, current and target (CDM, TDM). The CPI and TPI reflect the impact of CDM and TDM. Also, an action considers PI gap, DM gap, goal and pain point of a specific element and is associated with a roadmap. Moreover, five concepts have a set of dimensions that could be detailed to get additional dimensions. The architecture domain includes business and information as the architecture elements (people, process, capability, information) and roadmap (timeline, dependency and priority). Other domains, such as technology and security, can also be considered. On the other hand, we defined six levels of DM as attributes for the DM level class. For instance, level 0 represents the absence of digitalisation or basic digitisation, whereas advanced level 5 represents innovative, data-driven and adaptable aspects. These levels are further explained in Table 1. Also, PI type includes six different dimensions such as financial, quality and resource utilisations that could be further into more detailed and specific PIs under each PI type. (See Fig. 5, Table 1) Due to the visual constraint of the class labels in Fig. 5, some element labels may not be fully visible. Please see the class label column in Table 1.

Fig. 5. Full DEAMP graph-based model

Table 1. DEAMP concepts and their definitions

Class label	Concept	Definition	References
DM_Gap	DM gap	A level of DM that represents a gap in the DM	[7]
CDM	Current DM	A level of DM that represents the current level of the DM	[7]
TDM	Target DM	A level of DM that represents the target level of the DM	[7]
DM_Level	DM level	Represent the six levels of the DM	[7]
	Level 0	None – Absence of digitalisation or basic digitisation	[7]
	Level 1	Beginner – Digitally aware or ad-hoc digitalisation	[7]

(*continued*)

Table 1. (*continued*)

Class label	Concept	Definition	References
	Level 2	Learner – adopting digital practices	[7]
	Level 3	Intermediate – Consistent, defined, integrated and digitally enabled	[7]
	Level 4	Advanced – Completely developed, predictable and proactive	[7]
	Level 5	Expert – Innovative, data-driven and adaptable	[7]
PI_Gap	Performance gap	A set of PIs that represents a gap in performance outcomes	[6]
CPI	Current PI	A set of PIs that represents current performance outcomes	[6]
TPI	Target PI	A set of PIs that represents target performance outcomes	[6]
PIType	PI Type	Represent the six types of PI	[6]
Financial	Financial	Represent the financial PIs	[6]
Competitiveness	Competitiveness	Represent the competitiveness PIs	[6]
Resource_Utilisation	Resource utilisation	Represent the resource utilisation PIs	[6]
Quality	Quality	Represent the quality PIs	[6]
Innovation	Innovation	Represent the innovation PIs	[6]
Flexibility	Flexibility	Represent the flexibility PIs	[6]
Architecture	Architecture	A business area in the enterprise architecture design	[28]
Architecture_Domain	Architecture domain	A domain of the enterprise architecture design within a business area (business or information of the human layer)	[28]
Architecture_Element	Architecture element	An element of the architecture domain (People, Capability, Process or Information)	[28]
Goal	Goal	A description (statement) of what the organisation wants to achieve	[48]

(*continued*)

Table 1. (*continued*)

Class label	Concept	Definition	References
Objective	Objective	A scalable description of a goal	[48]
Pain_point	Pain point	A problem or solution of an issue that was unfinalised	[28, 49]
Action	Action	An initiative defined by the organisation to improve DM level for desired performance gain	[48]
Roadmap	Roadmap	A sequence of actions based on timeline, dependency and priority	[4, 28]
Timeline	Timeline	A period of time needed to implement an action	[28]
Dependency	Dependency	Represent the dependency between actions	[28]
Priority	Priority	The priority of an action	[28]
Benefit	Benefit	A gain or value realisation via implementing the roadmap	[50]
Actual_M	Actual Maturity	A level of DM that represents the actual DM level after or while implementing the roadmap	[7, 50]
Actual_P	Actual Performance	A set of PIs that represent the actual performance after or while implementing the roadmap	[7, 50]

5 Indicative Evaluation

This study provides an indicative elevation of the DEAMP ontology using an example scenario (fourth iteration). For this example scenario, we created fictitious test data to create the instance of the DEAMP reflecting a real-world marketing domain example as an indicative proof of concept validation.

The scenario is about a fictitious marketing ABC company. This company initiated a digital transformation initiative to improve overall performance by assessing and improving its DM. Their focus is on assessing and improving marketing business domain. Thus, a business architecture layer from the EA is used to understand the business marketing domain capabilities. As such, the marketing business domain can contain several business capabilities, such as advertising, managing customer relationships etc. Here, as an example, we focused on the "manage customer relationships" capability and captured the current and target DM levels, goals, pain points and current and target performance indicators (PIs). Also, the related current, target PIs and gaps are defined in alignment with the company's related goals and objectives. The instance of the example scenario on DEAMP is shown in Table 2 and Fig. 6. It can be observed that manage customer relationships capability (architecture element) indicates a gap (1 level) between the current and target DM state due to the duplicate accounts in the CRM (Customer Relationship Management) system as a pain point. On the other hand, one of the related PI to this capability is income which needs an increase of 15% based on the difference between the current and target income (linked to the company profit/ income objective). Thus, improving the performance management process by implementing a new dynamic online CRM system is the proposed action to uplift the DM to gain an increase in income (performance outcome). This action is anticipated or assumed to take 100 days with a high priority on a digital transformation roadmap. Further, post-action implementation, actual DM levels and performance can be tracked to capture whether the hypothesised (target) performance and DM improvements are reached. This example can be further explored using other related goals, objectives, pain points and PIs. Yet, it has been limited to one related aspect for the purpose of demonstration. Figure 6 shows how the example scenario and data are used to demonstrate the use of the proposed DEAMP ontology as a KG.

Table 2. DEAMP instantiation

Class label	Instances	Instances details
Architecture	Marketing	
Architecture_Domain	Business	
Architecture_Element	Cap01	Capability 01: Manage customer relationships
CDM	L01	Level 01
DM_Gap	1 Level	1 level difference between the current and target DM
TDM	L02	Level 02
CPI	F01	Financial indicator 01: income
PI_Gap	+15%	15% increase to up lift current income to target income
TPI	F01	Financial indicator 01: Income
Pain_Point	PP01	Pain point 01: duplicate accounts in the CRM systems
Goal	G01	Goal 01: improve productivity
Objective	Obj01	Objective 01: increase company profits/income (15%)

(*continued*)

Table 2. (*continued*)

Class label		Instances	Instances details
Action		Ac01	Action 01: implement a new dynamic online CRM system
Roadmap	Timeline	100 days	
	Dependency	NA	
	Priority	High	
Benefit	Actual M	–	–
	Actual P	–	–

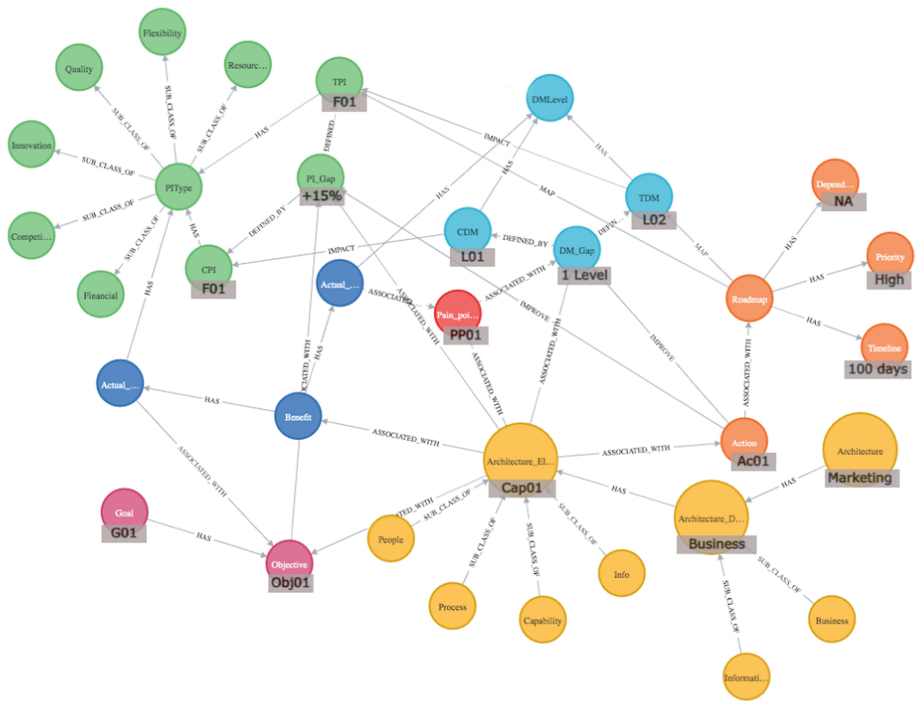


Fig. 6. The instance of DEAMP graph-based model with the example scenario

6 Discussion and Conclusion

Digitalisation is gaining considerable interest from academia and industry. However, there is a lack of understanding the link and impact of digitalisation maturity on organisational performance outcomes. This paper attempts to combine these two important areas into an integrated DEAMP ontology. As a result, this integrated DM and performance indicate several advantages. First, it can provide a systemic approach to link the previous isolated DM and performance concepts via an integrated ontology. Second, once they are integrated, we can then study the impact of a change in maturity on the performance outcomes and vice versa. Third, we can develop a performance outcome-driven approach to set the desired maturity level. Fourth, we can monitor and track the integrated maturity and performance outcomes. Finally, it can provide us with a lens to study the degree of maturity and its relevance to the degree of performance gain.

In conclusion, DEAMP ontology can be used for assessing and navigating DM, as well as linking it to performance results through activities and initiatives from the EA perspective. This intends to improve understanding of DM levels and their impact on performance outcomes when developing digital roadmaps and action plans. The applicability of the proposed DEAMP ontology is demonstrated with the help of an example scenario. This initial evaluation indicates the practical relevance and applicability of the proposed ontology. Future research will conduct additional experiments to evaluate and evolve the proposed DEAMP ontology.

References

1. Morakanyane, R., O'Reilly, P., McAvoy, J.: Determining digital transformation success factors. In: Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 4356–4365 (2020)
2. Gimpel, H., Hosseini, S., Huber, R., Probst, L., Röglinger, M., Faisst, U.: Structuring digital transformation: a framework of action fields and its application at ZEISS. J. Inf. Technol. Theory Appl. JITTA. **19**, 3 (2018)
3. Hinings, B., Gegenhuber, T., Greenwood, R.: Digital innovation and transformation: an institutional perspective. Inf. Organ. **28**, 52–61 (2018). https://doi.org/10.1016/j.infoandorg.2018.02.004
4. Parviainen, P., Tihinen, M., Kääriäinen, J., Teppola, S.: Tackling the digitalization challenge: how to benefit from digitalization in practice. Int. J. Inf. Syst. Proj. Manag. **5**, 63–77 (2017). https://doi.org/10.12821/ijispm050104
5. Westerman, G., Bonnet, D., McAfee, A.: Leading Digital: Turning Technology into Business Transformation. Harvard Business Press (2014)
6. Alsufyani, N., Gill, A.Q.: Digitalisation performance assessment: a systematic review. Technol. Soc. **68**, 101894 (2022). https://doi.org/10.1016/j.techsoc.2022.101894
7. Alsufyani, N., Gill, A.Q.: A review of digital maturity models from adaptive enterprise architecture perspective: digital by design. In: 2021 IEEE 23rd Conference on Business Informatics (CBI), pp. 121–130. IEEE (2021)
8. Gill, A.Q.: Adaptive Enterprise Architecture as Information. World Scientific Publishing Co., Singapore (2022)
9. Lee, M.: Enterprise architecture: beyond business and IT alignment. In: Advances in Intelligent Systems and Computing, pp. 57–66 (2013)
10. Ibrahim Alzoubi, Y.: Distributed agile development communication: an agile architecture driven framework. J. Softw. 681–694 (2015). https://doi.org/10.17706/jsw.10.6.681-694
11. Uschold, M.: Building ontologies: towards a unified methodology. In: 16th Annual Conf. of the British Computer Society Specialist Group on Expert Systems, pp. 16–18 (1996)
12. Hogan, A., et al.: Knowledge graphs. ACM Comput. Surv. **54** (2021). https://doi.org/10.1145/3447772
13. Aslanova, I.V., Kulichkina, A.I.: Digital maturity: definition and model. In: Proceedings of the 2nd International Scientific and Practical Conference "Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth" (MTDE 2020), pp. 443–449. Atlantis Press, Paris, France (2020)

14. Becker, J., Knackstedt, R., Pöppelbuß, J.: Developing maturity models for IT management. Bus. Inf. Syst. Eng. **1**, 213–222 (2009). https://doi.org/10.1007/s12599-009-0044-5
15. Chanias, S., Hess, T.: How digital are we? Maturity models for the assessment of a company's status in the digital transformation (2016)
16. Thordsen, T., Murawski, M., Bick, M.: How to measure digitalization? a critical evaluation of digital maturity models. In: Hattingh, M., Matthee, M., Smuts, H., Pappas, I., Dwivedi, Y.K., Mäntymäki, M. (eds.) I3E 2020. LNCS, vol. 12066, pp. 358–369. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44999-5_30
17. Poeppelbuss, J., Niehaves, B., Simons, A., Becker, J., Pöppelbuß, J.: Maturity models in information systems research: literature search and analysis. Commun. Assoc. Inf. Syst. **29**, 505–532 (2011). https://doi.org/10.17705/1CAIS.02927
18. Cao, M., Zhang, Q.: Supply chain collaboration: impact on collaborative advantage and firm performance. J. Oper. Manag. **29**, 163–180 (2011). https://doi.org/10.1016/j.jom.2010.12.008
19. Wardaya, A., Sasmoko, S. I.G., Bandur, A.: Mediating effects of digital technology on entrepreneurial orientation and firm performance: evidence from small and medium-sized enterprises (SMEs) in Indonesia. Int. J. Eng. Adv. Technol. **8**, 692–696 (2019). https://doi.org/10.35940/ijeat.E1098.0585C19
20. Kaplan, R.S., Norton, D.P.: The balanced scorecard–measures that drive performance. Harv. Bus. Rev. **70**, 71–79 (1992)
21. Cross, K.F., Lynch, R.L.: The "SMART" way to define and sustain success. Natl. Product. Rev. **8**, 23–33 (1988). https://doi.org/10.1002/npr.4040080105
22. Fitzgerald, L, Johnston, R, Brignall, S, Silvestro, R., Voss, C.: Performance measurement in service businesses. Manag. Account. 34–36 (1991)
23. Hinkelmann, K., Gerber, A., Karagiannis, D., Thoenssen, B., Van Der Merwe, A., Woitsch, R.: A new paradigm for the continuous alignment of business and IT: combining enterprise architecture modelling and enterprise ontology. Comput. Ind. **79**, 77–86 (2016). https://doi.org/10.1016/J.COMPIND.2015.07.009
24. Hazen, B.T., Bradley, R.V., Bell, J.E., In, J., Byrd, T.A.: Enterprise architecture: a competence-based approach to achieving agility and firm performance. Int. J. Prod. Econ. **193**, 566–577 (2017). https://doi.org/10.1016/J.IJPE.2017.08.022
25. Alzoubi, Y.I., Gill, A.Q., Moulton, B.: A measurement model to analyze the effect of agile enterprise architecture on geographically distributed agile development. J. Softw. Eng. Res. Dev. **6**(1), 1–24 (2018). https://doi.org/10.1186/s40411-018-0048-2
26. Zachman, J.A.: A framework for information systems architecture. IBM Syst. J. **26**, 276–292 (1987). https://doi.org/10.1147/sj.263.0276
27. Andrew, J.: TOGAF® Version 9.1 – A Pocket Guide (2016)
28. Gill, A.Q.: Adaptive cloud enterprise architecture Intelligent Information Systems. World Scientific Publishing Co., Singapore (2015)
29. Gill, A.Q., Chew, E.: Configuration information system architecture: Insights from applied action design research. Inf. Manag. **56**, 507–525 (2019). https://doi.org/10.1016/J.IM.2018.09.011
30. Gill, A.Q.: Applying agility and living service systems thinking to enterprise architecture. Int. J. Intell. Inf. Technol. **10**, 1–15 (1) (2014). https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/ijiit.2014010101. https://doi.org/10.4018/IJIIT.2014010101
31. Gill, A.Q., Beydoun, G., Niazi, M., Khan, H.U.: Adaptive architecture and principles for securing the IOT systems. In: Advances in Intelligent Systems and Computing, pp. 173–182. Springer, Cham (2020)
32. Westerman, G., Mcafee, A.: A Major Research Initiative at the MIT Sloan School of Management Research Brief the Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry (2012)

33. Park, Y., Saraf, N.: Investigating the complexity of organizational digitization and firm performance: a set-theoretic configurational approach. In: AMCIS 2016 Surfing IT Innov. Wave – 22nd Am. Conf. Inf. Syst., pp. 1–10 (2016)
34. Frederico, G.F., Nailor Pedrini, C., Francisco Frederico, G.: Information technology maturity evaluation in a large brazilian cosmetics industry. Int. J. Bus. Adm. **9**, (2018). https://doi.org/10.5430/ijba.v9n4p15
35. Tamm, T., Seddon, P.B., Shanks, G., Reynolds, P.: How does enterprise architecture add value to organisations? Commun. Assoc. Inf. Syst. **28**, 141–168 (2011). https://doi.org/10.17705/1CAIS.02810
36. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**, 199–220 (1993). https://doi.org/10.1006/knac.1993.1008
37. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.S.: Towards ontological foundations for conceptual modeling: the unified foundational ontology (UFO) story. Appl. Ontol. **10**, 259–271 (2015). https://doi.org/10.3233/AO-150157
38. Guizzardi, G., Halpin, T.: Ontological foundations for conceptual modelling. Appl. Ontol. **3**, 1–12 (2008). https://doi.org/10.3233/AO-2008-0049
39. Bellomarini, L., Fakhoury, D., Gottlob, G., Sallinger, E.: Knowledge graphs and enterprise AI: The promise of an enabling technology. In: Proceedings – International Conference on Data Engineering, pp. 26–37 (2019)
40. Pokorný, J.: Conceptual and Database Modelling of Graph Databases. (2016). https://doi.org/10.1145/2938503.2938547
41. Smajevic, M., Bork, D.: From conceptual models to knowledge graphs: a generic model transformation platform. In: Companion Proceedings – 24th International Conference on Model-Driven Engineering Languages and Systems, MODELS-C 2021, pp. 610–614 (2021)
42. Medvedev, D., Shani, U., Dori, D.: Gaining insights into conceptual models: a graph-theoretic querying approach. Appl. Sci. **11**, 765 (2021). https://doi.org/10.3390/app11020765
43. Vaishnavi, V.K., Kuechler, W.: Design Science Research Methods and Patterns: Innovating Information and Communication Technology. Auerbach Publications, New York (2007)
44. Smuts, H., Winter, R., Gerber, A., van der Merwe, A.: "Designing" design science research – a taxonomy for supporting study design decisions, pp. 483–495 (2022). https://doi.org/10.1007/978-3-031-06516-3_36
45. Gregor, S., Hevner, A.R.: Positioning and presenting design science types of knowledge in design science research. MIS Q. **37**, 337–355 (2013). https://doi.org/10.2753/MIS0742-1222240302
46. Winter, R., Albani, A.: Restructuring the design science research knowledge base: a one-cycle view of design science research and its consequences for understanding organizational design problems. Des. Organ. Syst. An Interdiscip. Discourse. **1**, 63–81 (2013). https://doi.org/10.1007/978-3-642-33371-2_4/FIGURES/10
47. Pokorný, J.: Graph databases: their power and limitations. In: Saeed, K., Homenda, W. (eds.) CISIM 2015. LNCS, vol. 9339, pp. 58–69. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24369-6_5
48. Kaplan, R.S., Norton, D.P., Ansari, S.: The execution premium: linking strategy to operations for competitive advantage. Account. Rev. **85**, 1475–1477 (2010). https://doi.org/10.2308/accr.2010.85.4.1475
49. Yu, Y., Madiraju, S.: Enterprise application transformation strategy and roadmap design: a business value driven and IT supportability based approach. In: Proc. - 2nd Int. Conf. Enterp. Syst. ES 2014, pp. 66–71 (2014). https://doi.org/10.1109/ES.2014.37
50. Serra, C.E.M., Kunc, M.: Benefits realisation management and its influence on project success and on the execution of business strategies. Int. J. Proj. Manag. **33**, 53–66 (2015). https://doi.org/10.1016/J.IJPROMAN.2014.03.011

Workplace Topology Model for Assessment of Static and Dynamic Interactions Among Employees

Jānis Grabis(✉)

Information Technology, Riga Technical University, Zunda Krastmala 10, Riga 1048, Latvia
grabis@rtu.lv

Abstract. Interactions among employees promote spread of infectious diseases at workplaces. Enterprise architecture provides rich information about organizational structure and its relations to other elements in an enterprise. This paper develops a workplace topology model as a sub-set of enterprise architecture to analyze both static and dynamic interactions among the employees and to monitor and to limit spread of infectious diseases. The model is developed as a multi-layer graph combining organizational, facility and sensing layers. The graph analytical methods are elaborated to analyze the interactions. The analysis provides inputs to spread of infectious diseases risk assessment models. The application of the graph analytics methods is demonstrated using an example from an Information Technology consulting company producing both software and hardware products.

Keywords: Organization graph · Graph analytics · Spread of infection

1 Introduction

Propagation of Covid-19 and similar infectious diseases depends on human behavior and working environment [1]. The human behavior concerns the way they interact with each other and the working environment concerns facilities and conditions they interact within. There are many different types of interactions among employees at a workplace making it one of the most frequent places for spread of infectious diseases such as flue or Covid-19 [2, 3]. The organizational structure and processes cause emergence of localized bubbles either promoting or limiting the spread of diseases [4].

The organizational structure typically shows a hierarchy of organizational units and roles and employees working for these organizational units. Modeling of organizational structure is a part of many enterprise architecture management frameworks. The organization structure is relatively static and majority of companies also have dynamic project or product development teams involving interactions across many organizational units. Increasingly, companies are adopting team-oriented approaches. In both cases, interactions among units, teams and employees are naturally represented as a graph [5]. The graph analysis methods provide a powerful tool to analyze the interactions in the organization. There methods include centrality measures, shortest path as well as identification of clusters and major components.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 230–244, 2023.
https://doi.org/10.1007/978-3-031-26886-1_14

While the organization structure shows interactions among employees, it lacks a physical dimension. The physical dimension, i.e., specific work place, is important to understand spread of infectious diseases [6]. The aforementioned features can be captured by adding another layer to the organization's graph. The combined layers are represented as a multi-graph [7]. Multi-layered graphs are well suited to show and to analyzed interactions among various aspects of the modeling problem. The organization and physical layers jointly show employee and their location and interactions among them. In order to limit and prevent, the spread of infectious diseases ambient conditions also should be observed and suitable actions should be carried out. The ambient conditions are determined by sensing and various types of actuators are used to perform the actions. Sensing and enactment capabilities can be represented as another layer in the multi-graph.

Although graphs and physical layout of various are frequently used to analyze spread of infectious diseases up to now there is no comprehensive method to jointly represent various aspects of the interactions and to analyze these interactions to configure scheduling and access systems. It is proposed to use a topological model to represent the organizational, work place layout and sensing concerns. The work environment topology model is a multi-graph jointly representing organizational structure, physical dimension of the work place and sensing capabilities. It is applied to analyze Covid-19 related risks.

The objective of this paper is to formulate the work environment topology model and to define means for analyzing the model using the graph theory and network analysis methods. The topology model can be perceived as a representation of enterprise architecture, which provides rich means for representing organizational structures [8] and recently has been extended to represent aspects specific to sensing enterprises [9]. Both structural and missionary organizational units can be represented [10]. Graph-based analytical methods are shown as a promising technique to gain quantitative insights on the basis of enterprise architecture models [11]. Graph analysis techniques also have been applied to analyze spread of infections in workplaces [12]. The contribution of this paper is elaboration of graph-analytical methods to analyze static and dynamic interactions among employees in the organizational context with an intent to assess the risk of spreading infectious diseases such as Covid-19. Among other factors this risk is affected by intensity of interactions within and outside of organizational bubbles as well as duration of these interaction [13].

The rest of the paper is structured as follows. The background information is discussed in Sect. 2. The topology model is formulated in Sect. 3. Analysis of the topology model is demonstrated in Sect. 5. Section 6 concludes.

2 Background

Enterprise architecture management is an approach for analyzing complex organizations [14]. It is typically represented as a layered structure including business, logical and physical layers. The workplace topology model is a representation of enterprise architecture focusing on interactions among employees and sensing capabilities to analyze these interactions. The three relevant layers are:

1. Organization layers – business layer;

2. Facility layers – physical infrastructure;
3. Sensing layer – digital infrastructure.

2.1 Organizational Layer

ArchiMate states that the business layer elements are used to model the operational organization of an enterprise in a technology-independent manner.[1] It includes active structure elements such as business role, business actor and business collaboration to represent the organization as well as behavior and passive elements. The topology model focuses on active structure elements although the behavior elements also could play a major role in dealing with the spread of infectious diseases.

The elements represented in the organizational layer are:

- Person – a human working at or visiting the organization. They are employees of the company as well as customers, partners and other persons visiting organization on-site. The visitors are also usually registered in the access management system.
- Organization unit – a persistent logical grouping of employees with common goals and activities.
- Team – a temporary organizational unit involving the persons dedicated to a common course and having frequent interactions though not necessarily having specific physical location.

2.2 Facility Layer

The facility layer representing some of the physical aspects of the organization. It assumes that a work place (e.g. office or manufacturing site) is divided into zones. The zones are identified in an access management system and employees have permission to access specific zones.

The elements represent in the facility layer are:

- Organization unit – the same as defined in the organizational layer.
- Zone – a physically bounded and identifiable location in organization's premises.
- Sub-zone (work station) – a constituent part of the zone.

A sub-zone can have sub-zones on its own. Zones and sub-zones form a hierarchy, where an upper level element encompasses lower level elements.

2.3 Sensing Layer

The sensing layer represents digital infrastructure used at the organization to monitor and to limit Covid-19 risks. It focuses on sensors and actuators and their association with specific zones.

Two types of elements are represented in the sensing layer:

[1] https://pubs.opengroup.org/architecture/archimate3-doc/chap08.html.

1. Sensor – a physical or logical device measuring conditions at a workplace. That includes IoT sensors measuring ambient conditions as well sensors measuring properties of wastewater and web sensors retrieving contextual data from internal and external information systems;
2. Actuator – a physical or logical active element, which allows to change conditions at a workplace. These include actuators interacting with building management system to change ventilation parameters or access management system to restrict access to specific zones.

The nodes of the sensing layer correspond to devices in the ArchiMate's technology layer, where additional information about the devices, interfaces and communication networks is provided.

If a sensor or an actuator is attached to a zone having multiple sub-zones, then it is serving all sub-zones, e.g., there is typically one waste water analyzer per object. The wastewater analyzer a specific time of sensor considered in this project as a non-intrusive early warning method [15].

2.4 Integration

The topology model is intended to provide input data to assess the risk of spreading infectious diseases in a workplace. Data provided by various sensors, organizational information systems and wastewater analysis are combined to estimate the risk level, which in turn triggers a corrective action performed by actuators (Fig. 1). The topology model extracts and structures data from the organizational information systems. Both master and transactional data are used. The Time management system manages data about the organizational units and the zones, the Project management system manages information about dynamic project teams and the Enterprise calendar contains scheduled events and their participants. This way the model characterizes interactions among the employees. The intensity and type of interactions is combined with IoT sensor measurements and wastewater analysis results to evaluate the risk. While the topology model based analysis of interactions focuses on planned interactions, the IoT sensor data can capture informal interactions, e.g., anonymized video stream data can be used to apprise interactions taking place at a resting place.

The risk is either predicted or its current value is evaluated. There are several risk evaluation models. For example, Buonanno et al. [13] evaluate the risk of infection depending on duration and type of interactions among people as well as characteristics of the meeting room. The proposed topology model would provide the necessary input data to the risk estimation model. These input data include type of interactions among people and duration of these interactions. Outputs of the risk model are used to trigger risk reduction activities by involving appropriate actuators. For example, ventilation can be adjusted with regards to expected number of the employees in a room and current air quality measurement or events can be rescheduled if the wastewater analysis warning has been triggered.

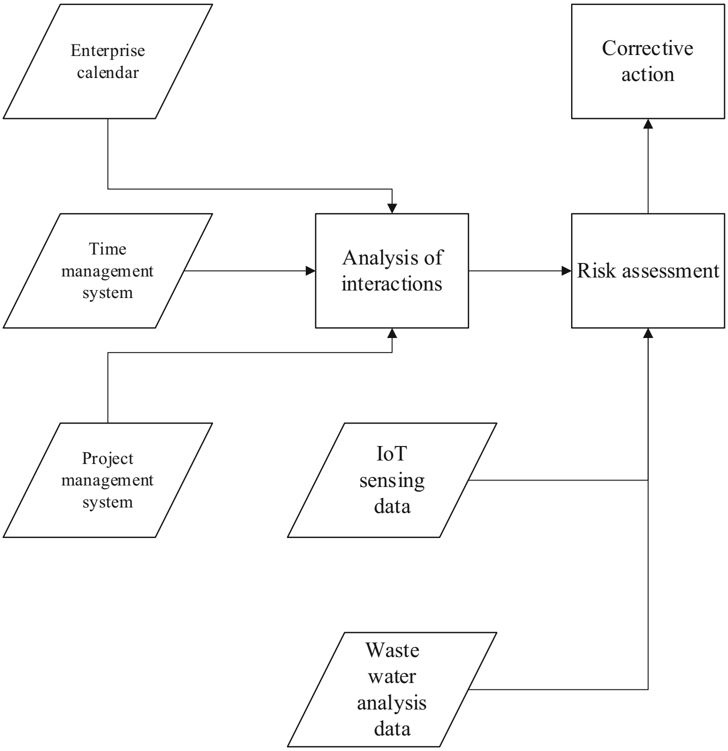


Fig. 1. Integrated risk-based decision-making.

3 Model Formulation

The graph theory is chosen as a method to represent the workplace topology. The graph-based representation can be transformed in a visual representation of the target enterprise architecture framework. The work environment topology model is a multi-layer attribute graph:

$$TM = (\mathcal{Y}, \mathcal{N}, \mathcal{E}, \beta, \lambda), \tag{1}$$

where $\mathcal{Y}$ is a set of layers, $\mathcal{N}$ is a set of nodes, $E : \mathcal{N} \times \mathcal{N}$ is a set of edges connecting the nodes and $\beta : \mathcal{N} \rightarrow \mathcal{Y}$ is a function mapping the nodes to layers. Individual nodes and layers are denoted by n and y, respectively. An edge e_{ij} connects nodes n_i and n_j. Thus, the set of nodes is defined as $N = \{n_i | i \in \mathbb{Z}^*\}$. The nodes have a specific type, and this type is represented by a set of node labels $\mathcal{L}$ is defined as:

$$\mathcal{L} = \{\text{Person, Org. Unit, Team, Zone, Sub} - \text{zone, Sensor, Actuator}\}. \tag{2}$$

There is a function $\lambda(n_i)$ that maps the nodes to their labels. The nodes have properties represented as key-value pairs and a set of the property keys is denoted as $\mathcal{K}$. Every node has the name property as well as other properties as necessary. A set of the property values is denoted as $\mathcal{V}$. For example, sensors have an attribute Type assuming a value "Air quality sensor":

$$\{k_{i1} = \text{type} : v_{i1} = \text{Air quality}\},\ \lambda(n_i) = \text{Sensor}. \tag{3}$$

Sensor attributes are defined according to the Fiware[2] data model and other attributes can be introduced as necessary.

The topology model captures both static and dynamic interactions among employees. In order to represent short term interactions, the Person node has the Calendar attribute. It is a complex attribute including elements such as:

1. Event – name of the calendar event;
2. Event date – date of the calendar event;
3. Event start and end time – specific time slot of the working day the event takes place;
4. Duration – duration of the calendar event;
5. Zone – location of the event;
6. List of participants – persons required or invited to the event;
7. Type – regular meeting, ad hoc meeting etc.

The Calendar attribute is populated by integrating data from Time Management System, Project Management System and Enterprise Calendar. The calendar merges regular working schedule, regular meetings and ad-hoc meetings.

The function β mapping the nodes to layers states that (see [16] on representation of multi-layered graphs):

- Org. unit., Team and Person nodes are present in the Organizational layer;
- Person, Zones and Sub-zones nodes are present in to the Facility layer;
- Zone, Sensor and Actuator nodes are present in to the Sensing layer.

Note that the Person nodes are present in the Organizational layer and the Facility layer, thus linking the two layers and a connection between matching nodes is implied (i.e., if the same node is present in two layers a connection between the nodes can be established). Similarly, the Zone nodes are present to the Facility layer and the Sensing layer and connect the two layers.

The edges connect only specific types of nodes:

- Person to Org. unit – defines an organization unit a person works for;
- Person to Team – defines a project team a person is assigned to;
- Person to Zone – identifies a zone a person has a work place. It can be slowly changing as indicated in the Time Management System;
- Person to Sub-zone – similarly to the previous identifies a sub-zone a person is assigned to;
- Sub-zone to zone – represents the composition of zones;

[2] https://github.com/smart-data-models/dataModel.Device/blob/1c41e5128a54d7b67bf51f3a7daba3c72f497aa0/Device/doc/spec.md.

- Sensor to Zone – shows sensors available for monitoring environmental conditions in a zone;
- Actuator to Zone – shows actuators available in a zone to change the environmental conditions.

The model formulation defines the layers, types of the nodes and permissible edges. A topology model is instantiated for a particular organization, and it shows specific units, employees and other elements for this organization. Typically, most of the information is extracted from various information systems. However, one cannot expect that all information would be readily available without some data transformation and integration effort.

A sample topology model is shown in Fig. 2. All persons are assigned to and organizational unit. Some of the persons are also working for temporary teams. Team membership crosses boundaries of the organizational units. A person works at a specific zone deemed as his or her default work place. However, persons can mover dynamically from one zone to another. That is captured in the calendar or zone access data.

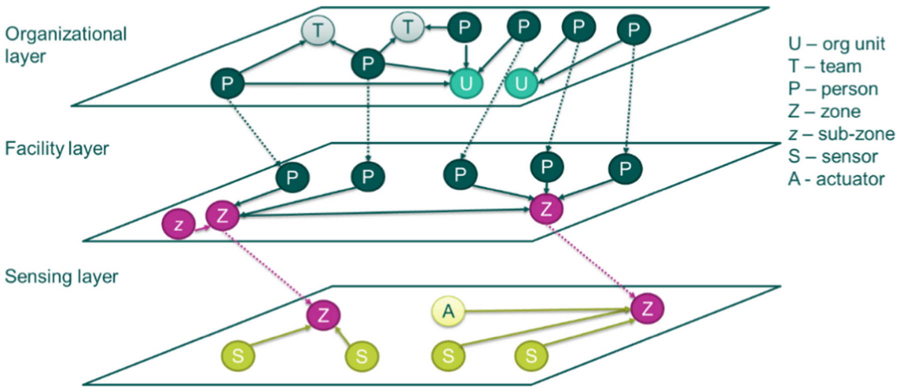


Fig. 2. A visualization of the topology model.

4 Model Analysis

The instantiated (i.e., organization specific) topology model is used to understand interactions in the organization and to evaluate sensing needs. The topology model is used to:

1. Perform static analysis – analyzes interactions using a snapshot of the topology model without using the calendar information;
2. Perform dynamic analysis – analyzes interactions using multiple versions of the topology model or the calendar information;
3. Configure applications used to manage work environment – the access rights to zones, work scheduling and sensing requirements can be derived from the topology model to configure the time management and access management system.

The following sub-sections discuss the static and dynamic analysis while the configuration is beyond scope of this paper.

4.1 Static Analysis

The static analysis is performed using the permanent (more precisely slowly changing data) from the topology model. Several types of static analysis have been identified:

1. The variety of employees in the zone – how many different units relative to the number of persons are present in a single zone;
2. The centrality of person in the zone – how many persons are related to a person;
3. The sensor data availability – how many sensors are available in a zone?
4. The overall graph density in each layer – shows connectivity of nodes.

The variety of employees in the zone (*VE*) that is defined as a ratio between the number of units and the number of persons associated with the zone. Given a node n_i of type Zone, a set of connected nodes of type Person is:

$$\mathcal{N}_i^* = \left\{ n_j | \lambda(n_j) = \text{Person}, \exists e_{ij} \in \mathcal{E} \right\} \tag{4}$$

and a set of organization units these persons belong to is:

$$\mathcal{M}_i^* = \left\{ n_k | \lambda(n_k) = \text{Organization unit}, n_k \in \mathcal{N}_i^* \right\}. \tag{5}$$

VE is calculated as:

$$VE = \frac{|\mathcal{M}_i^*|}{|\mathcal{N}_i^*|}. \tag{6}$$

The centrality of person (CP) indicates connectiveness of the person in the organization and it is calculated:

$$CP = |\mathcal{P}_i^*|, \tag{7}$$

where $\mathcal{P}_i^* = \left\{ n_j | \lambda(n_j) = \text{Person}, \exists e_{kj} \in \mathcal{E} \wedge n_k \in \mathcal{N}_i^* \right\}$ is a set of persons connected to the ith person via its associated unit and teams and $\mathcal{N}_i^* = \left\{ n_j | \lambda(n_j) \in \{\text{Unit, Team}\}, \exists e_{ij} \in \mathcal{E} \right\}$ is a set of units and teams the ith person is associated with.

The sensor data availability is simply calculated as a degree of the zone node calculated withing sensing layer. It can be expressed in both absolute and relative terms. The measure indicated that some of the zones might be lacking sufficient sensing capabilities.

The overall graph density is calculated the ratio between the edges present in one layer of the graph and the maximum number of possible edges in the layer. In the organizational layer, high density suggests a large number of possible interactions. In the sensing layer, high density suggests that the organization is well equipped with sensors and actuators.

4.2 Dynamic Analysis

The dynamic analysis is performed using the frequently changing data from a version of the topology model with a specific timestamp. Several types of dynamic analysis have been identified:

1. Intensity of interactions with other units and teams – characterizes how many events involving persons from other units and teams a person has;
2. Expected duration of interactions – characterizes duration of interactions with persons from other zones;
3. Expected variety – the number of different units or teams a person interacts with.

The intensity of interactions *DP* is calculated for a person and characterizes the number of contacts and the time this person spends with employees working at other organizational units or teams. It is calculated using the following algorithm:

1. Select a node n_j of type person
2. Set $DP_i := 0$
3. For each Event j in the person calendar

    3.1 Set $DP_i := DP_i + kd_j$, where d_j is duration of the jth event and k is a number of participants from a different organization unit or team;

4. Set $DP_i := \frac{DP_i}{D}$, where D is duration of all events for the person (note that events also include the regular scheduled work hour).

DP can exceed one implying that a person is having interactions with multiple persons from different organization units at a single event.

The expected duration of interactions is calculated in a similar manner to *DP* without adding the number of persons multiplier and selecting the persons from other zones in Step 3.1. This measure is often used to calculate the risk of infection. The measure takes into account interactions with persons from other zones assuming than precautionary measures are implemented within the zone.

The expect variety is calculated by counting how many different units or teams are represented in the person's calendar as participants of various events.

5 Application

The topology model is developed for a small and medium size Information Technology and Communications company. The company develops and deploys integrated hardware and software solutions. It has about 35 employees organized in nine units including administration, sales and marketing, software development, electronics and networking services. Five distinctive zones are identified and access is controlled using an access management system. The electronics and technical services in particular cannot work remotely. The company is functionally-oriented and prefers on-site work for all employees. Project teams are established for specific missions like development of new products or onboarding of new customers.

The topology model represents the actual static structure of the case company. However, the team composition and the calendar data are generated for experimental purposes on the basis of interviews with company representatives concerning their work organization. Experimental studies are conducted to compare intensity of employee interactions in the case of functional organization and project-oriented organization. The project-oriented setting is considered to explore possibility for the case company to switch to more project-oriented way of working.

The topology model is shown in Fig. 3 and Fig. 4. It represents a snapshot at a specific time moment. The person to unit and person to zone assignments as well as assignments in the sensing layer are relatively stable. The person to team assignments are changing frequently. The topology model shows that employees belong to specific organization and some of them also belong to project teams. There are five zones in company's office: Office, Electronics, Programming, Common area and Conference room. These zones are treated as sub-zones to the overall office building zone (z0), which has an access gate actuator controlling access to the whole building. It is envisioned that a water monitoring sensor also will be added to z0. The fifth zone has no persons assigned as a permanent work place. It is booked for specific events and dynamic analysis is mainly relevant for this zone. All zones are equipped with air quality sensors, while other types of sensing devices and actuators are sparingly available. This is one of the observations made during the static analysis that sensing capabilities are not sufficient for comprehensive monitoring and enactment of response actions.

The measures used in the static analysis can be calculated. Table 1 reports the calculated values of the variety of employees in a zone. The higher value indicates that employees from many zones are represented in the zone. The largest variety is in z1, where employees from administration, marketing, project management and accounting are located. Larger variety could potentially indicate a greater risk of spreading infections outside an organizational bubble.

The experimental studies are conducted to show an example of dynamic analysis using simulated data. The following assumptions are made in the experimental studies:

1. Participants of events mainly come from the one organizational unit in the case of the functional organization;
2. Participants of events mainly come from one team in the case of the project-oriented organization;
3. The events are classified as unit meeting and team meetings. The unit meetings are more frequent than the project meetings in the functional organization while the team meetings are more frequent in the project-oriented organization;
4. Participants from other units or teams are also invited to events. The share of external participants is larger for the unit meetings than team meetings (because external information requirements are higher).

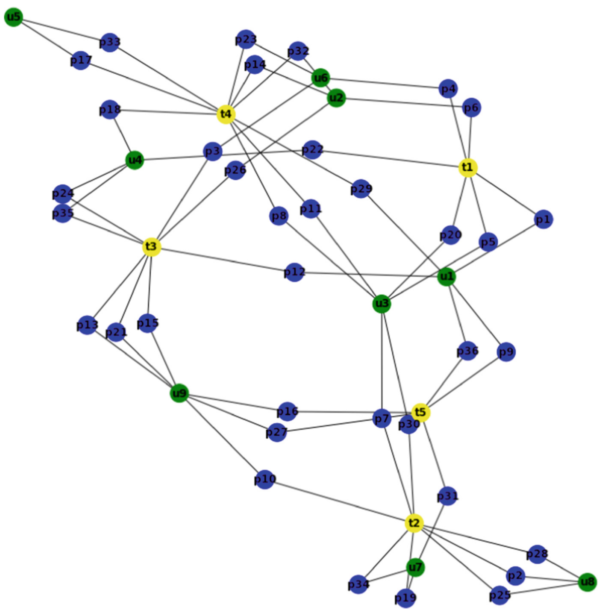


Fig. 3. The organizational layer of topology model for the case company.

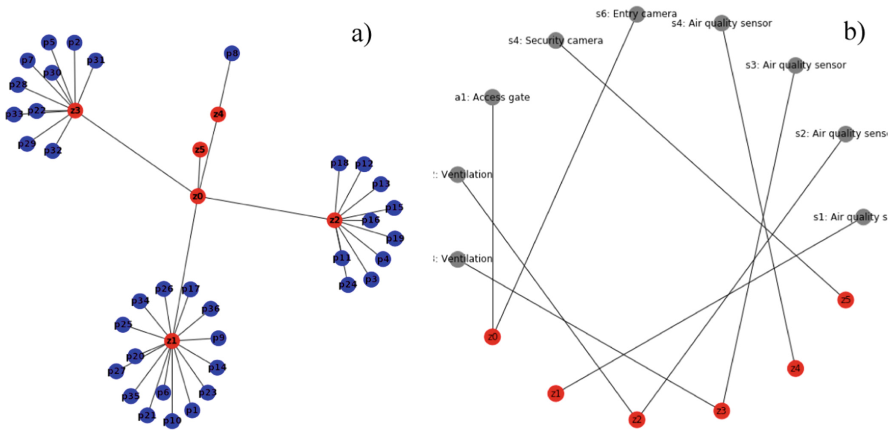


Fig. 4. The topology model of the case company: a) facility layer; and b) sensing layer.

The Intensity of interactions (*DP*) is evaluated in the experiments. The measure is evaluated for the functional organization and the envisioned project-oriented organization, and the share of external participants (4th assumption) is also varied. Each experimental treatment is evaluated for 500 randomly generated person to team assignments and calendar events.

Table 1. Variety of employees in a zone (*VE*) for selected zones

Zone ID	z1	z2	z3
Zone Name	Office	Electronics	Programming
VE	0.27	0.25	0.2

In the first experiment, the share of external participants (*FE*) in each meeting is 50% of all participants. The experimental results (Fig. 5 a)) shows that the intensity of interactions is higher for the functional organization. That is slightly contra-intuitive because

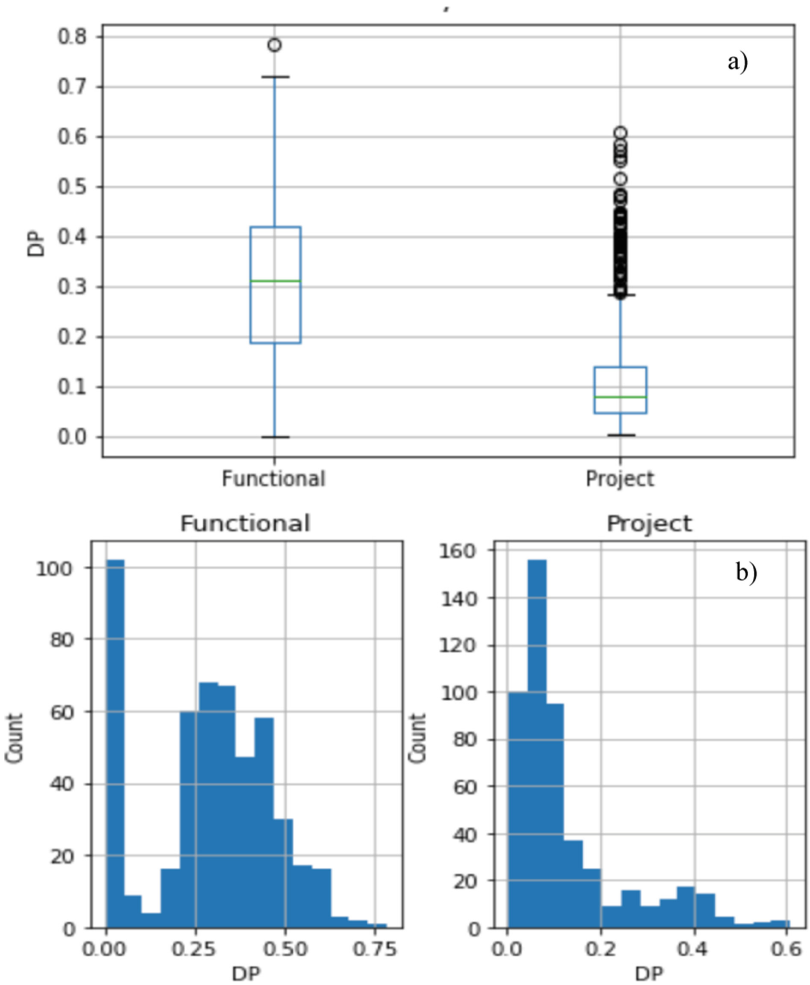


Fig. 5. The intensity of interactions in functional and project-oriented organizations: a) box-plot chart of *DP* and b) distribution of *DP*, $FE = 0.5$.

Table 2. The average value of *DE* for all persons depending on organization type and *FE*

Organization type	$FE = 0.2$	$FE = 0.5$
Functional	0.067	0.288
Project	0.072	0.126
Wilcoxon test p-value	0.03	0.00

less interactions are expected in the functional organization. However, the need for many external advisers is the cause of the high level of interactions. That is also illustrated in Fig. 5 b) showing that *DP* has a bi-modal distribution. The first mode represents meetings with a few external participants while the second mode points towards meetings with many external participants. The project-oriented organization has many outlying cases with a significantly larger number of interactions than usual. With regards to spread of infectious diseases, this observation suggests that the risk prediction is more challenging for the project-oriented organization even though the overall level of interactions might be lower.

To assess the impact of *FE* on the results, the experiments are repeated with $FE = 0.2$ (Table 2). The intensity of interactions was larger for the functional organization if $FE = 0.5$ and the Wilcoxon test confirms that *FE* values differ substantially depending on the type of organization for any level of statistical significance. However, the intensity of interactions is lower for the functional organization if $FE = 0.2$ (at the significance level of 3%). The shape of distributions is similar for both types of the organization (Fig. 6). The results imply that involving fewer external participants significantly reduces the

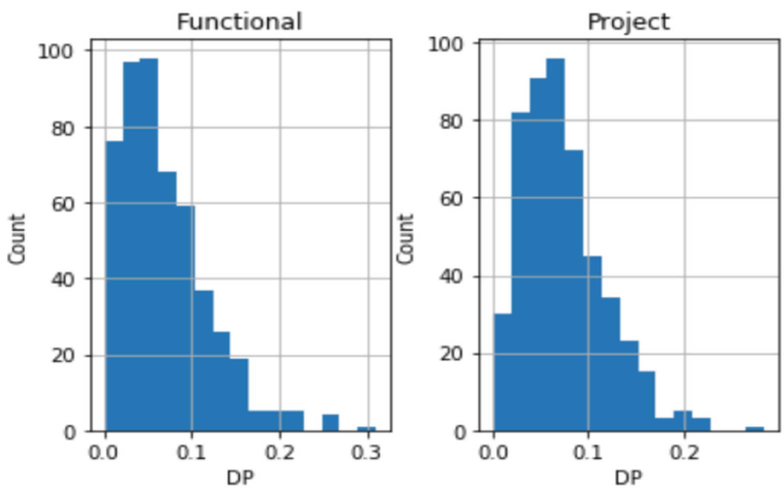


Fig. 6. The intensity of interactions in functional and project-oriented organizations characterized by distribution of *DP*, $FE = 0.2$.

intensity of interactions and this reduction is more profound for the functional organization. In general, involving fewer external participants makes it easier to predict the intensity of interactions.

6 Conclusion

The paper has elaborated a workplace topology model, which provides inputs to risk assessment of spread of infectious diseases. Using the graph theory as a representation method allows to calculate quantitative measures to analyze and to compare characteristics of the organization. Some of the measures serve as inputs to infection risk assessment models. Several static and dynamics measures are defined in the paper while others can be added as necessary. The topology model contains only a sub-set of elements describing the overall enterprise architecture. These sub-elements are selected to allow for quantitative analysis of interactions among employees and infection sensing capabilities.

This research is conducted as a part of applied research project jointly with the aforementioned IT consulting company. The aim of the project is to enhance the company's access management system with features reducing spread of infections such as Covid-19 at workplaces with restricted opportunities for remote work. The further research will focus on identification of the most suitable topology analysis measures from the practical perspective and to gather data for further analysis of dynamical aspects of the topology model.

The topology model itself is dynamic (e.g., person to zone assignments are variable). Therefore, versions of the model and its evolution are tracked. The topology model will be implemented as a part of the access management system, which will also track the model evolution and correlate the topology measures with enterprise performance measures.

Acknowledgements. Project "Platform for the Covid-19 safe work environment" (ID. 1.1.1.1/21/A/011) is founded by European Regional Development Fund specific objective 1.1.1 «Improve research and innovation capacity and the ability of Latvian research institutions to attract external funding, by investing in human capital and infrastructure». The project is co-financed by REACT-EU funding for mitigating the consequences of the pandemic crisis.

References

1. Molinaro, M., Romano, P., Sperone, G.: The organizational side of a disruption mitigation process: exploring a case study during the COVID-19 pandemic, Operations Management Research, in press (2022)
2. World Health Organization Homepage: https://www.who.int/publications/i/item/WHO-2019-nCoV-workplace-actions-policy-brief-2021-1. Last accessed 07 March 2022
3. Koh, D.: Migrant workers and COVID-19. Occupational and Environmental Medicine **9**, 634–636 (2020)
4. Shaw, J., et al.: Working in a bubble: How can businesses reopen while limiting the risk of COVID-19 outbreaks? CMAJ **192**(44), E1362–E1366 (2020)
5. Song, M., van der Aalst, W.M.P.: Towards comprehensive support for organizational mining. Decision Support Systems **46**(1), 300–317 (2008)

6. Arnicans, G., Niedrite, L., Solodovnikova, D., Virbulis, J., Zemnickis, J.: Towards a System to Monitor the Virus's Aerosol-Type Spreading. In: Byrski, A., Czachórski, T., Gelenbe, E., Grochla, K., Murayama, Y. (eds.) Computer Science Protecting Human Society Against Epidemics. ANTICOVID 2021. IFIP Advances in Information and Communication Technology, vol. 616. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86582-5_9
7. Chakravarthy, S., Santra, A., Komar, K.S.: Why multilayer networks instead of simple graphs? modeling effectiveness and analysis flexibility and efficiency!. In: Madria, S., Fournier-Viger, P., Chaudhary, S., Reddy, P. (eds.) Big Data Analytics. BDA 2019. Lecture Notes in Computer Science, vol. 11932. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37188-3_14
8. de Kinderen, S., Kaczmarek-Heß, M.: On model-based analysis of organizational structures: an assessment of current modeling approaches and application of multi-level modeling in support of design and analysis of organizational structures. Softw Syst Model **19**, 313–343 (2020). https://doi.org/10.1007/s10270-019-00767-4
9. Zimmermann, A., Schmidt, R., Sandkuhl, K., Wißotzki, M., Jugel, D., Möhring, M.: Digital enterprise architecture - transformation for the internet of things. 2015 IEEE 19th International Enterprise Distributed Object Computing Workshop, pp. 130–138 (2015). https://doi.org/10.1109/EDOCW.2015.16
10. Pereira, D.C., Almeida, J.P.A.: Representing organizational structures in an enterprise architecture language. In: Proceedings of the 6th Workshop on Formal Ontologies meet Industry, CEUR Workshop Proceedings, vol. 1333 (2014)
11. Smajevic, M., Bork, D.: Towards graph-based analysis of enterprise architecture models. In: Ghose, A. et al. (eds.) Conceptual Modeling. ER 2021. Lecture Notes in Computer Science, vol. 13011. Springer, Cham (2021)
12. Mauras, S., et al.: Mitigating COVID-19 outbreaks in workplaces and schools by hybrid telecommuting, PLoS Computational Biology **17** (2021)
13. Buonanno, G., Stabile, L., Morawska, L.: Estimation of airborne viral emission: quanta emission rate of SARS-CoV-2 for infection risk assessment. Environment international **141** (2020)
14. Lankhorst, M.: Enterprise Architecture at Work: Modelling, Communication and Analysis. Springer Berlin (2013)
15. Zhu, Y., et al.: Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. Science of the Total Environment **767** (2021)
16. Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. Journal of Complex Networks **2**(3), 203–271 (2014)

EDOC Forum

EDOC Forum

Organization

Forum Chairs

João Paulo A. Almeida	Federal University of Espírito Santo, Brazil
Dimka Karastoyanova	University of Groningen, The Netherlands

Program Committee

Aditya Ghose	University of Wollongong, Australia
Andrea Marrella	Università di Roma, Italy
Alan Wee-Chung Liew	Griffith University, Australia
Alexander Knapp	Universität Augsburg, Germany
Alfred Zimmermann	Reutlingen University, Germany
Amin Beheshti	Macquarie University, Australia
Andreas L. Opdahl	University of Bergen, Norway
Andrew Berry	ResMed Inc., Australia
Aniruddha Gokhale	Vanderbilt University, USA
Antonio Vallecillo	Universidad de Málaga, Spain
Artem Polyvyanyy	The University of Melbourne, Australia
Asif Qumer Gill	University of Technology, Australia
Axel Korthaus	Swinburne University of Technology, Australia
Barbara Weber	University of St. Gallen, Switzerland
Ben Roelens	Open Univ. of The Netherlands, The Netherlands & Ghent University, Belgium
Benjamin Yen	The University of Hong Kong, China
Carlos L.B. Azevedo	Federal Institute of Espírito Santo, Brazil
Chiara Di Francescomarino	Fondazione Bruno Kessler-IRST, Italy
Christian Zirpins	Karlsruhe University of Applied Sciences, Germany
Claudenir M. Fonseca	University of Twente, The Netherlands
Claudio Di Ciccio	Sapienza University of Rome, Italy
Colin Atkinson	University of Mannheim, Germany
Cristine Griffo	Free University of Bozen-Bolzano, Italy
Dimka Karastoyanova	University of Groningen, The Netherlands
Dominik Bork	TU Wien, Austria
Fatih Turkmen	University of Groningen, The Netherlands

Fethi Rabhi	The University of New South Wales, Australia
Flavia Santoro	Rio de Janeiro State University, Brazil
Florian Matthes	Technical University of Munich, Germany
Frank Leymann	University of Stuttgart, Germany
Frederik Gailly	Ghent University, Belgium
Georg Grossmann	University of South Australia, Australia
Georg Weichhart	PROFACTOR GmbH, Austria
Giuseppe Di Lucca	University of Sannio (RCOST), Italy
Guido Governatori	CSIRO, Australia
Hans Weigand	Tilburg University, The Netherlands
Henderik A. Proper	Luxembourg Institute of Science and Technology & University of Luxembourg, Luxembourg
Hiroshi Miyazaki	Keio University, Japan
Irina Rychkova	Centre de Recherches en Informatique & Université Paris 1 Panthon-Sorbonne, France
Jaap Gordijn	Vrije Universiteit Amsterdam, The Netherlands
Jan Øyvind Aagedal	Equatex, Norway
João Moreira	University of Twente, The Netherlands
João Paulo A. Almeida	Federal University of Espírito Santo, Brazil
John Mylopoulos	University of Ottawa, Canada
José Raúl Romero	University of Cordoba, Spain
Julio C. Nardi	Federal Institute of Espírito Santo, Brazil
Julius Köpke	Alpen-Adria-Universität Klagenfurt Institute for Informatics Systems, Austria
Lam Son Lê	HCMC Tech, Vietnam
Ljiljana Brankovic	University of Newcastle, Australia
Luis Ferreira Pires	University of Twente, The Netherlands
Luise Pufahl	TU Berlin, Germany
Madhushi Bandara	University of Technology Sydney, Australia
Manfred Reichert	University of Ulm, Germany
Marco Montali	Free University of Bozen-Bolzano, Italy
Maria Teresa Gómez López	University of Seville, Spain
Maria-Eugenia Iacob	University of Twente, The Netherlands
Marten van Sinderen	University of Twente, The Netherlands
Mathias Weske	HPI & University of Potsdam, Germany
Mattia Fumagalli	Free University of Bozen-Bolzano, Italy
Michael Schrefl	University of Linz, Austria

Michael Rosemann	Queensland University of Technology, Australia
Nicolas Herbaut	Université Paris 1 Panthéon-Sorbonne, France
Oscar Pastor	Universitat Politècnica de València, Spain
Paolo Ceravolo	Università degli Studi di Milano, Italy
Peter Bernus	Griffith University, Australia
Peter F. Linington	University of Kent, UK
Pierluigi Plebani	Politecnico di Milano, Italy
Pontus Johnson	KTH Royal Institute of Technology, Sweden
Rainer Schmidt	Munich University of Applied Sciences, Germany
Rajeev Raje	IUPUI, USA
Remco Dijkman	Eindhoven University of Technology, The Netherlands
Renata Guizzardi	University of Twente, The Netherlands
Ronny Seiger	University of St. Gallen, Switzerland
Ruth Breu	Research Group Quality Engineering, Italy
Rüdiger Pryss	University of Würzburg, Germany
Sagar Sunkle	Tata Consultancy Services, India
Saïd Assar	Institut Mines-Télécom Business School, France
Schahram Dustdar	Vienna University of Technology, Austria
Selmin Nurcan	Université Paris 1 Panthéon-Sorbonne, France
Sharmistha Dey	Griffith University, Australia
Simon Hacks	University of Southern Denmark, Denmark
Stefan Tai	TU Berlin, Germany
Stefanie Rinderle-Ma	Technical University of Munich, Germany
Sylvain Hallé	Université du Québec à Chicoutimi, Canada
Tiago Prince Sales	University of Twente, The Netherlands
Ulrich Frank	University of Duisburg – Essen, Germany
Ulrik Franke	RISE, Sweden
Uwe Zdun	University of Vienna, Austria
Vinay Kulkarni	Tata Consultancy Services Research, India
Wolfgang Maass	Saarland University, Germany
Yigal Hoffner	Shenkar College of Engineering and Design, Israel
Zoran Milosevic	Deontik & Best Practice Software, Australia

Additional Reviewers

Barat, Souvik
Berry, Andrew
Bühler, Fabian
Cremerius, Jonas
de Alencar Silva, Patrício
Dexe, Jacob
Gamage, Dimuthu
Guy, Ed
Gökstorp, Simon
Kaczmarek-Heß, Monika
Katsikeas, Sotirios
Khorshidi, Samira
Kuhn, Peter
Lichtenstein, Tom
Murturi, Ilir
Muñoz, Paula
Peregrina Pérez, José Antonio
Ristov, Sasko
Roychoudhury, Suman
Yilmaz, Fatih
Yussupov, Vladimir

Optimized Throttling for OAuth-Based Authorization Servers

Peter Schuller[1(✉)], Julia Siedl[1,2], Nicolas Getto[1,2], Sebastian Thomas Schork[1], and Christian Zirpins[2]

[1] CAS Software AG, CAS-Weg 1-5, 76131 Karlsruhe, Germany
{peter.schuller,julia.siedl,nicolas.getto,sebastian.schork}@cas.de
[2] Karlsruhe University of Applied Sciences, Moltkestr. 30, 76133 Karlsruhe, Germany
{julia.siedl,nicolas.getto,christian.zirpins}@h-ka.de
https://cas-future-labs.de/, https://h-ka.de/iaf/dss

Abstract. Responsiveness is a key requirement for web-based enterprise software systems. To this end, throttling (or rate limiting) is often applied to block illegitimate traffic from outside-facing components and protect their computational resources. OAuth-based authorization servers are among the most popular components facing the web. Alas, the OAuth protocol introduces severe challenges to throttling. The OAuth protocol flow introduces indirections to client requests that make it hard to determine their source to apply rate limits. Moreover, fixed limits perform poorly in cases varying between high and low request loads. In this paper we propose solutions for both issues and provide an efficient solution for throttling in the context of OAuth. This includes integrated methods for a) cooperative throttling of authorization and resource servers as well as b) dynamic rate limiting as part of the throttling algorithm. We evaluate our approach based on a real-world use-case of enterprise CRM.

Keywords: Authorization server protection · OAuth protocol · Cooperative throttling · Dynamic rate limiting · CRM

1 Introduction

Modern web applications, especially those offering services to users over the internet, are facing significant variation regarding system usage [23]. In line with the distributed architecture of such systems, the aspects of authentication and authorization are usually provided as services of dedicated *auth servers* that build on standards like OAuth [16]. By design, such services are directly exposed to vast numbers of requests not only from users and clients, but also from the resource servers they are providing authentication and authorization for. These requests are leading to potential performance challenges on the auth server especially in multi-tenant environments [13]. Under peak load, they may cause significant performance drops or even service outages.

To protect auth servers from such issues, measures to manage system load have to be applied on different levels. Modern cloud-based systems offer standard

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 251–267, 2023.
https://doi.org/10.1007/978-3-031-26886-1_15

mechanisms to scale their capacity as one option to do so [4]. Whilst being useful for increases in legitimate traffic, this approach offers an additional surface to attackers. It prevents performance losses and outages but in turn can create direct financial impact. Thus additional measures to block malicious traffic has to be employed. A technique to do so is *throttling* or *rate limiting*. It is based on per-user limits and blocks all traffic from a user exceeding the defined limit.

The most common approach is generic throttling on the network layer. Whilst being widely available [1–3], this approach has the drawback of operating on limited information, e.g., with respect to source and target IP addresses. Such IP addresses are only partially useful to identify users. Techniques like NAT, VPN or proxies obfuscate an unknown number of users behind a single IP address. Thus, it is unclear, whether a single source IP address identifies a specific user or a group, let alone the group size.

For this reason, it is desirable to enrich the context of throttling mechanisms with additional information from higher system layers. Yet, such information should be still generic enough to make resulting throttling mechanisms widely applicable to different systems and domains, so using application-specific features (like end user profiles) is not a good option. A promising approach is to utilize specific information in the context of auth-servers, as these are closely related to the clients and users of the system but still offer a standardized generic view that is commonly available for many web applications.

Consequently, we tackle the question how throttling can be optimized by means and in the context of OAuth-based systems and break it down as follows:

1. How can meaningful user entities be identified for OAuth-based systems?
2. What are efficient and effective throttling limits for such users?

To answer those research questions we use a case study and evaluate our prototypes on it. As a running example, we use a cloud-based multi-tenant Customer Relationship Management (CRM) system, more specific SmartWe [17]. It provides business users the ability to manage customer-related data such as appointments, e-mails, tasks or documents. Technically, SmartWe is a web application building on OAuth and consequently features the *user* as resource owner, a *client*, a *resource server* and an *auth-server* (see Fig. 1).

Users interact with a web-based client app. As this client is public, the *Authorization Code Grant Type* is used [5, p. 7], meaning that the client redirects the users' browser to the auth-server. After the user logs in, the client retrieves an initial token that it can exchange with the auth-server into a final *access token*. Different grant types exist that have different ways of obtaining such a token, but thereafter, all grant types follow the same interaction pattern.

A client can send a request for a resource (e.g., a data object like an appointment) to the resource server. To signal the users' approval for that access, each request includes the access token retrieved previously. Before returning the resource, the resource server sends this token to the auth-server for validation.

As the example shows, distributed systems may generate internal traffic while processing external requests. In the case of OAuth, resource servers query the

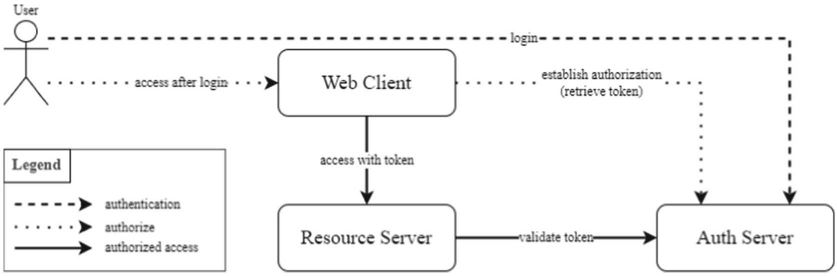


Fig. 1. SmartWe OAuth-based authentication flow

auth-server to validate tokens. Originating from an internal server (and his IP address), those requests cannot be linked to a user on the network layer. It requires additional information from the application layer to resolve this level of indirection and apply proper throttling. Such information can be found in the request itself on the application layer. Throttling on the application layer includes more information on the client but is specific to the used protocol and cannot be implemented generically. Adaptations to the concrete application and use-case are required to make use of it. As we will show, more generic information on the level of OAuth can be used instead.

Beyond matching requests to users, another challenge for throttling is to adequately quantify their limits. Existing throttling solutions mostly require upfront configuration and in order to be effective for resource protection, limits need to be set to rather low rates. This might impose unwanted limitations to legitimate users, especially in situations where the systems overall load is low and resources are available. However, if limits are set rather high and all legitimate traffic is allowed, this might not be strict enough to protect the systems resources in situations of high load. A promising approach to resolve this problem is to dynamically adapt request rate limits to changing levels of system load. To this end, we show how optimized request rate limits for OAuth-based systems can be dynamically computed from the auth-server load.

Concerning the protection of OAuth-based auth-servers, some research already exist and even standards have been published [22]. However, most existing studies focus on protocol security [15, Chapter 9], [11,19] or given implementations [18,21]. To the best of our knowledge, protocol-specific throttling as a protection against overload has not been published yet. As regards dynamic throttling limits, there are also existing approaches. Most of them operate on network layer – either directly [12,20] or embedded in an SDN [7]. Other approaches have implications on the overall system architecture [24] or focus only on outgoing traffic [14]. To the best of our knowledge, an approach to dynamically determine limits for incoming traffic on the application layer is still absent.

Summarizing the above, our approach provides three main contributions:

1. We outline and discuss several ways to identify single user sessions in an OAuth based system. By doing so, we are able to apply throttling also to endpoints that are subject to indirections via resource servers.
2. We propose a mechanism to set dynamic rate limits based on servers' varying load. It maps the system load measured as average response time to an allowed usage-frame, to obtain a load specific throttling limit.
3. We show how to integrate both solutions in a holistic architecture. That way, we obtain balanced throttling that reliably identifies users and applies appropriate limits for every load situation.

The rest of the paper is structured as follows. We describe details of OAuth-based servers that make it hard to control their request traffic in Sect. 2. In Sect. 3, we discuss related work on protecting authorization servers and dynamic request throttling. Our cooperative approach for dynamic throttling of OAuth-based resource access is described in Sect. 4 followed by its evaluation for the case of a real-world CRM system in Sect. 5. Section 6 concludes our research and outlines future work.

2 Foundations of OAuth-Based Authorization Servers

In this paper, we focus on auth-servers that implement the protocol standards OAuth 2.0 (OAuth) [5] and OpenID Connect (OIDC) [10].

The OAuth 2 protocol [5] is about authorization and access rights delegation. It does not require any service (representing an OAuth client as defined in [5, Sect. 2]) to store end user credentials for the purpose of accessing a certain resource. Instead, *tokens* are requested from a system service and issued to grant access to a resource server. OAuth distinguishes between short-lived *access tokens* used as credentials for the actual API consumption and long running *refresh tokens* used to renew access tokens. Access tokens may have a defined lifetime, are limited to a specific access right (scope) and may be revoked at any time. Access tokens are either self-contained or denote an identifier of variable length and consist of alphanumeric digits and some special characters [6].

In contrast to OAuth, OIDC [10] returns an additional *ID token* in form of a JSON Web Token (JWT) that contains actual user information. Thereby OIDC builds upon and extends OAuth with respect to authentication.

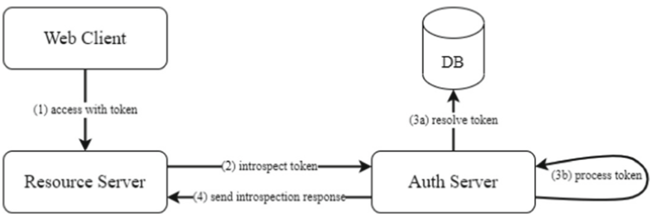


Fig. 2. OAuth token introspection

For validating a token [8], three types of actors are involved: an *authorization server*, a *client* and a *resource server* (see Fig. 2). A client requests a resource from the resource server while presenting an access token (1). The resource server makes a request to the authorization server to validate the access token (2). The authorization server checks if the access token is valid (3) and communicates this back to the resource server (4). Depending on the result, the resource server may give the client access to the resource. The validation of the given access token at the authorization server involves database access and additional processing (3a, 3b) making the token introspection an expensive operation.

The communication is based on HTTP relying on the underlying network and thus sensible to latency and network failures. The system may be able to partially cope with a temporary, short unavailability of the auth-server. For a short period existing sessions may stay active because tokens are cached by the resource server but new login requests cannot be handled. A persisting unavailability on the other hand represents an imminent risk for the overall system availability.

RFC 6749 [5] on the OAuth 2 standard states for the access token requests that *"the authorization server MUST protect the endpoint against brute force attacks (e.g., using rate-limitation or generating alerts)"*. RFC 7662 [8], specifying the token introspection, discusses attack vectors that arise when it is *"left unprotected and un-throttled"*. Thus, the standards are aware of the importance of protecting the resource server and see throttling as an approach to do so.

Especially the token introspection endpoint has further attack vectors if throttling is applied in a generic manner. This mainly arises from the indirection that related requests have to take. From a logical perspective, the request to this endpoint is a direct result of a request that a client sends to a resource server.

3 Related Work

In the following, we summarize existing work that is related to our research problem and solution approach. In particular, this paper deals with *a) protecting authorization servers* and *b) dynamic request throttling*.

Concerning the protection of authorization servers, literature mostly focuses on the security of the protocol itself. The corresponding standard on the "OAuth 2.0 Threat Model and Security Considerations" [22] explicitly considers the communication between authorization server and resource server out of scope. All described attacks are focused on retrieving information or gaining access, but not on causing service outages. Similar arguments can be found in further literature on this topic such as [15, Chapter 9], [11,19]. [11, p. 185] even discusses security considerations specifically for the token introspection endpoint, highlighting its importance still missing any attacks focusing on its availability. Other research in this field adds the aspect of flaws that are introduced by insecure implementations of the OAuth protocol [18,21].

As regards dynamic request throttling, authors of [24] present an architecture for internal service design, in which services are split up into multiple event-driven stages which are all connected by using explicit request queues. This is a problem for already existing services or programmers who don't necessarily want to base their services on the *event-driven architecture* (SEDA). We propose an additional component instead, which produces a dynamic throttling limit.

The importance of such a mechanism can also be seen in [14]. The paper presents an implementation of distributed rate limiters, which enforce a global rate limit across traffic to multiple sites to enable coordinated policing. This can be used to limit global consumption of a priced third-party service to set strict cost limits. While this paper deals with outgoing traffic, it does show the inherent need of such a mechanism in cloud-based services. We use the basis of throttling, but we concern ourselves with incoming traffic.

Another perspective on rate limiting is shown in [20]. The paper introduces a scalable rate limiting method by using a NIC. The host CPU classifies data packets and queues them in a per-class queue, which specifies rate limits for its class. The presented NIC, called SENIC, handles metadata and scheduling. This solution requires additional hardware and its scalability comes from static limits. We propose a software solution instead, that doesn't need specialised hardware.

Authors of [12] propose dynamic rate limiting as a defense mechanism against flooding based (D)DoS attacks. Every edge router in a network that drops a lot of packets gets punished with a strong static limit, while routers that drop few packets may use a weak static limit. This solution works on the network level, while we propose a solution on the application level.

Another dynamic rate limiting mechanism against (D)DoS can be found in [7]. This paper proposes a solution based on a sliding–window algorithm, which works on a weighted network abstraction. The weighting is based on the queue capacity of the controller and the number of rules in the switch. This solution is not using a pure mathematical solution, but is using a deep-learning algorithm for it's dynamics. It focuses security problems that are specific for SDNs. We propose a more general solution instead.

4 A Cooperative Approach for Dynamic Throttling of OAuth-Based Resource Access

In the following, we present our approach for optimized throttling of OAuth-based auth-servers. First, we focus on the problem of applying throttling to the introspection endpoint and define related objectives. As a solution, we discuss considerations for user or client identification and describe protocol adaptations. Second, we describe how dynamic limits can be implemented. We propose a metric for estimating the server load and show the related calculation of limits. Finally, we present a common architecture to combine both parts.

4.1 Protecting the Introspection Endpoint

The first goal is to protect the auth-server from overload via the token introspection endpoint (see Fig. 2). This endpoint is particularly important, since all other servers of the system depend on it for authorizing their clients. Furthermore, this endpoint leads to indirections unlike other auth-server endpoints.

General Considerations. One solution would be to indirectly protect the auth-server by applying throttling at the resource-server only. This has the advantage of evading indirections. However, this approach is not flexible enough as it leaves the auth-server completely dependent on the protection mechanisms of the resource-servers. As a result, this would limit the deployment options of the auth-server. It could only be deployed with trusted resource-servers like those belonging to the same organisation. An additional observation is that a resource-server should be independently protected from the auth-server, because it is publicly available for clients. The approach presented in this paper can be confidently applied to resource-servers outside the control of the organisation if they implement the presented protocol extension.

Another naive approach would be for the auth-server to apply throttling to resource-servers. If the granularity is a whole resource-server this could be fatal. Imagine that a client makes an excessive amount of requests to the resource-server that subsequently reaches its limit for validating access tokens. Then the resource-server would be left in a non-operative state for all clients. Therefore an auth-server has to apply throttling to requests at a finer granularity.

User Identification. A partial conclusion is that there need to be protection mechanisms in both auth-server and resource-server. When designing the approach, we observed that resource-servers can minimize the overhead for classification by letting the auth-server do parts of the data analysis. By cooperating, both servers have sufficient information to classify requests, while keeping the exchanged information at a minimum.

When referring to access tokens, they can be invalid. In the context of this paper, "invalid" only refers to correctly formed access tokens. They may be invalid because they have expired, are guessed randomly by a malicious actor or are damaged due to client implementation errors. Access tokens that have the wrong length or contain illegal characters are not examined here, because they can be filtered out easily by an API-gateway.

When the resource-server receives a request, it can categorize it by access token, client ID and the source IP address + port of the client. To know if the access token is valid, the resource-server has to query its token cache or the auth-server. The resource-server has to query the auth-server for each unknown access token, as the number of possibly valid access token is virtually unlimited.

On the side of the auth-server, this request needs to be categorized in order to apply throttling mechanisms before it is possibly processed. To be efficient, this categorization needs to be faster than the processing itself and thus must be light on resource usage. Therefore, user information associated with an access token should not be used for throttling, as this would allocate the same resources it is meant to protect. Consequently, the categorization can only be done via the access token. Since an access token represents the access rights for a client, this categorization identifies a combination of resource owner and client instance. However, if invalid access tokens are used to classify requests for throttling, the auth-server has to store those potentially unlimited numbers of invalid access tokens. For the auth-server, those invalid access token do not entail any meaning other than they were received by a resource-server. The resource-server however, can map these invalid access token to an IP address.

Approach to Protect the Introspection Endpoint. We use access tokens for classification and enhance them with the client IP address in case of invalid access tokens. Depending on how access tokens and IP addresses are combined, different scenarios are possible at the client. Table 1 classifies these scenarios and depicts their semantics. In particular, there are client errors and malicious attacks. Some combinations are marked "unusual", as for this kind of error to occur, an implementation has to have errors at numerous places.

Table 1. Classification of resource server requests

		Same access token		Different access token	
		Token valid	Token invalid	Token valid	Token invalid
Same IP address	Normal rate	**Single client**	Error/attack	**Many clients at same access point**	Error/attack
	High rate	Error/attack	DOS/unusual error	**Web application**	DOS/ unusual error
Different IP addresses	Normal rate per IP	**Single client switches access point**	Error/attack		
	High rate per IP	Attack/unusual error	DDOS/unusual error		

The idea is that an auth-server counts requests for access tokens and the percentage of invalid requests from IP addresses of clients. If it detects unwanted behaviour, it generates specific errors on these requests. The resource-server has to remember these errors and block those requests instead of forwarding them to the auth-server. Thus, the auth-server needs to manage counters for access tokens and IP addresses. The resource-server needs to manage a collection of currently blocked access token and currently blocked IP addresses of clients. The modified process looks the following (modifications marked as *italics*):

1. A resource-server receives a client-request for a resource with an access token.
2. *If the access token or IP address is marked as blocked, the resource-server may return an error.*
3. Otherwise the resource-server makes a request to the auth-server to validate this access token *while also passing the client IP address with it.*
4. *The auth-server returns a specific error if the limit is reached for the access token or IP address.*
5. The auth-server validates the access token *while registering the use of access tokens and whether an invalid access token originated from that IP address.* Then it returns the result.

6. *If the resource-server received an error it marks the IP address or access token as blocked and may return an error to the client.* Otherwise it answers the client request normally.

The existing protocol of the introspection endpoint changes in two ways. First, the resource-server has to relay the client IP address to the auth server. Second, the resource-server has to handle the two new types of errors the auth-server produces. Those are throttling errors and errors for an IP address, where invalid access tokens originated.

In the auth-server, overhead results from classifying the requests. There are two parameters to configure the classification. First, a limit for the number of requests per access token within a time window needs to be defined. Second, there needs to be a maximum percentage of allowed invalid requests from an IP address. This percentage should also include a threshold for the number of requests from that IP address. Otherwise a new IP address would be blocked immediately if the first request contains an invalid access token. The configuration of these parameters determines the effectiveness of the whole approach. Section 4.2 discusses, how an adaptive approach can be useful here.

At the level of the resource-server, some additional computational resources are needed as well. The collections of blocked access token and IP addresses need to be managed. This data is less complex than the data managed by the auth-server and changes less frequently.

Advantages and Limitations. The advantage of relaying the client IP address to the auth server is, that the case of invalid access token can get handled without needing to store those tokens. The auth-server should not store a collection of invalid tokens indiscriminately, because this list may grow indefinitely. Thereby, the auth-server can protect itself from a broken client that sends invalid token.

If throttling was only implemented in the resource-server, a client could reset its limit by refreshing its access token. The resource-server does not have a mechanism to monitor relations between access token. The presented approach has the advantage that the auth-server can track an access token even if it is refreshed. Since the auth-server also handles token refreshments it can update the corresponding throttling counter when this endpoint is called.

As regards limitations, the presented approach does not define strategies to handle DoS, let alone DDoS attacks. First of all, it presents a simple mechanism to identify its indirect clients at a sensible granularity. To be more explicit, the granularity can be described as the client session. Still, the contribution is that the presented mechanism may be integrated in frameworks that handle attacks, so they become more effective for OAuth scenarios.

Also, the solution could be further improved for the case of invalid access tokens from the same client IP address. If multiple clients are using the same IP and one tries using an invalid access token, all clients might be blocked. This edge-case can get mitigated by dynamically adapting the percentage of allowed invalid requests from the same IP address based on server load. Partly, this can also be mitigated by using a less strict limit if the client is a web application. For accuracy, further research is needed on how to temporarily store invalid access tokens in those edge-cases while limiting the maximum number of stored tokens.

4.2 Dynamic Rate Limiting

The goal of a dynamic throttling limit is to combine the positive effects of a weak limit, such as an increased user experience, with the positive effects of a strong limit, such as the protection of server resources. A dynamic limit, at any given point in time, is therefore derived from the server load at that time. To that end, the average response time is used as an indicator to see, how big the current server load is. From there, the limit, that is required to either bring the server load down or to keep it steady, is calculated.

Server Load. The average response time is calculated on the last responses in a given time frame. While a shorter time frame, such as a few seconds, can depict sudden traffic spikes in the resulting value, a longer time frame, such as a minute, can represent the overall server load situation better. If sudden traffic spikes can be expected, a middle ground value between these two should be considered.

To calculate the server load, an upper and lower limit for the average response time has to be defined beforehand. If the server has an average response time close to or above the upper limit, it is presumed to be busy at that given point in time. Otherwise, if the average response time is under the lower limit, the server is presumed to be underutilized. This relation between a) the average response time of the server and b) it's presumed server load is used to calculate a limit. This limit can be used by a throttling mechanism.

These average-response-time-limits must be defined upfront. In our implementation, the lower limit was set to one second, since that is roughly the point, when a user will start noticing a delay [9]. As a maximum, five seconds was chosen. Since a user will lose concentration after a delay of ten seconds [9], some buffer is present to ensure that the server won't reach such a high delay.

Throttling Limit. Since the goal of a dynamic limit is to counteract a rising server load, the limit has to get stricter accordingly. Nonetheless, the throttling limit can't shrink without any limitations. If it would get too low, a normal user could not use the service anymore. Therefore a minimum and a maximum limit have to be defined based on the expected server behaviour in rising traffic.

Expected is a stable load curve, which can be seen as the bottom line, up until a specific point, at which the curve will grow exponentially. At this point, the throttling limit needs to get stricter. The curve above the load curve shows the defined maximum average response time. The throttling limit needs to be at it's strictest, when the load curve cuts the maximum average response time.

This behaviour is represented in the first two rows of Eq. 1. Here, $f(x)$ is defined as the calculated limit where x is the server load. It will later be used by the throttling mechanism. Furthermore, $Min\_Avg\_Resp\_Time$ is defined as the value, where the load curve starts to grow. $Max\_Avg\_Resp\_Time$ is defined as the value, where the graphs intercept.

$$f(x) = \begin{cases} Max\_Limit & x \leq Min\_Avg\_Resp\_Time \\ Min\_Limit & x \geq Max\_Avg\_Resp\_Time \\ t(x) & Min\_Avg\_Resp\_Time < x < Max\_Avg\_Resp\_Time \end{cases} \tag{1}$$

The third row in Eq. 1 is a function that calculates the limit for a load value in between these set average response time limits. To this end, a linear function $t(x)$ with negative slope is used (see Fig. 3).

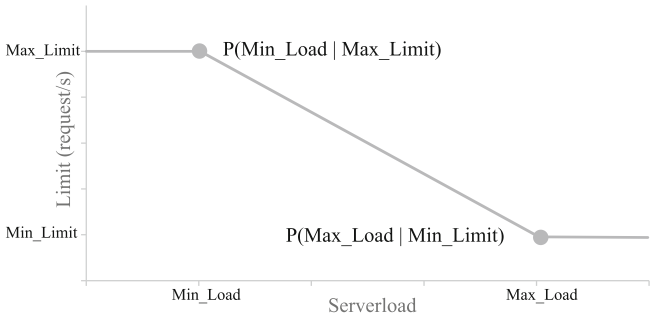


Fig. 3. Graphical representation of throttling function $f(x)$

Since a limit has to be defined for every request, this function has to be called for every request and before the throttling. For clients, this process does obfuscate the throttling limit and create additional complexity for the developer. however, this obfuscation can also be a benefit in attacking scenarios as well, since a potential attacker can't just stay under the throttling limit.

4.3 Common Architecture

Summarising the above, we presented concepts for solving two problems that arise for throttling of OAuth-based systems. First we use tokens in combination with IP addresses to identify users. Once we properly identify the users, the next step is to find good throttling limits. We presented an approach to dynamically link this limit to the current server load.

In the context of the overall system architecture, both partial solutions are implemented directly in the auth-server. This shared context enables us to integrate the parts. When a request arrives two major steps have to be taken:

1. The user for whom this request is to be accounted and the usage is to be calculated, needs to be identified. This is usually done based on IP address. In our approach, this is done as described in Sect. 4.1.

2. Based on this usage and a limit, the decision whether to process the request or not needs to be taken. Usually this limit is statically configured. In Sect. 4.2 we describe our approach to dynamically calculate this limit based on the current system load.

By separating both parts in self-contained components, we enable a modular usage of the overall approach. Not only can the parts be reused, they can also be recombined with enhanced variants for each sub-task.

In a broader systems view, only the first part of our solution has an effect on other components. It requires all resource-servers that query the user introspection endpoint to include the IP address, from which the request originated.

5 Evaluation

In Sect. 4, we outlined a solution for throttling in OAuth-based systems. As our solution addresses two separate issues, our evaluation is split in two parts as well. By using one common example, we underpin the possibility to integrate both solutions, providing a comprehensible and complete approach to the overall research question. For both parts, we show that they fulfill their respective requirements whilst being efficient and scalable.

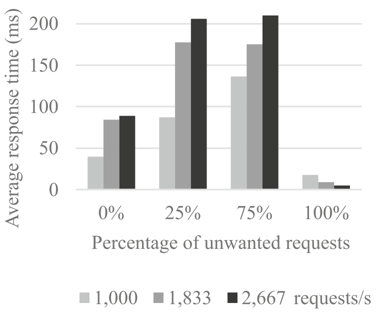
Subsequently, the version of the auth-server without modification will be called *original server* and the server that implements the part of the proposed solution to be evaluated is called *modified server*. Concerning the experimental set-up, we have used *JMeter* to model user interactions with the auth-server. Both the server and *JMeter* were run locally on the same machine to prevent fluctuation in the response time caused by the network connection. Since the auth-server is designed scalable, the actual number of requests is not important since it only reflects the capabilities of the local machine. The most meaningful observations concern comparisons of the original and modified servers.

5.1 Testing the Protection of the Introspection Endpoint

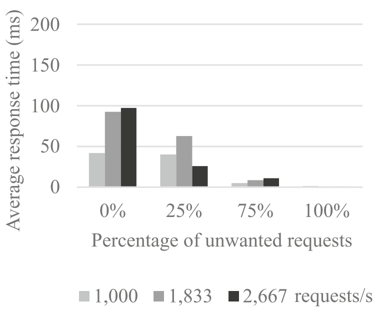
The following test is a comparison of the response times of the auth-server under load with and without throttling. The goal was to validate that the proposed solution for the introspection endpoint produces an acceptable overhead in the auth-server. Along the way, the findings also confirm the precondition that throttling can reduce server load. There were no concrete attack scenarios included in the test, only scenarios where the volume of request exceeds the defined limits. The reason is, that it does not serve the goal of measuring the overhead and throttling is imprecise against attacks by design.

The idea of the testing approach is to approximate the server load by measuring the response time of the auth-server. The testing framework simulates the behaviour of the resource-server. It sends requests for different clients to the introspection endpoint. As an indicator for the event that the auth-server reached its processing limit, it is observed when the response time stops to increase.

Figures 4a and 4b show the resulting response time for sending requests to the introspection endpoint. The different shades represent the requests per second that should be produced by *JMeter*. The percentage of unwanted requests are those requests that produce a throttling error in the modified version of the server. For example, for 25% of unwanted requests, when sending 1,000 request per second, 250 of those requests produced a throttling error in the modified server version. The original server responded to all requests. Wanted requests were modeled in a way that 100 different access tokens where combined randomly for each request with 101 different IP addresses. The unwanted requests were modeled by a combination of one access token and one IP address.



(a) Response times of original server



(b) Response times of modified server

Fig. 4. Response times for different percentages of unwanted requests and different numbers of requests per second to the introspection endpoint

Both diagrams show that the response time increases as more requests are sent. Due to a peculiarity of the server implementation, the response time for the original auth server increases when the percentage of unwanted requests is increased. This does not influence the test results. For 100% of unwanted requests the response time is low for the original server because then the introspection for only one access token is requested and it gets cached.

When comparing the response time of the original server and the modified server for 0% unwanted requests, the overhead of the proposed solution becomes visible. For the same number of requests, the original server has a response time that is about 5% lower. This means that the solution produces measurable linear overhead. The results from the modified server also show that the response time is dramatically decreased when a percentage of the requests get throttled. When all requests get throttled the server has a response time of under 10 ms meaning the classification is efficient.

5.2 Testing the Dynamic Calculation of Rate Limits

A user interacts with the auth server according the OAuth 2 *Authentication Code Grant Flow* and every test models 100 users.

Overhead. To show the overhead of the dynamic component, *JMeter*-Tests were carried out against the original auth-server as well as the modified one with respect to the average response time and the achieved throughput. After every test, the target throughput got increased to represent increasing traffic rates. The results for the average response time are shown in Fig. 5a.

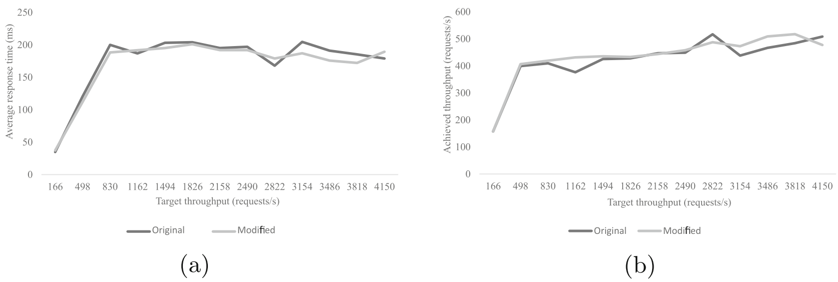
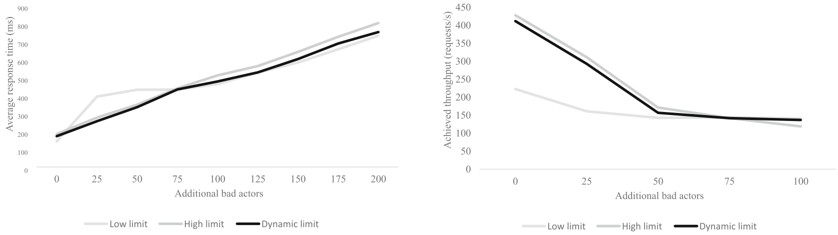


Fig. 5. (a) Comparison: average response time (b) Comparison: achieved throughput

It can be observed that the difference in terms of the average response time is minimal. At a target throughput of 830 request per second, the server is starting to get to it's limit and the difference starts to get lost in noise. The results for the achieved throughput are shown in Fig. 5b. Again, no significant difference can be observed. The two servers are almost identical up until a target throughput of 830, where the limit is reached once again. Random noise can be observed again, but there isn't any noteworthy difference in the reached throughput.

Concluding the results, the overhead of the component doesn't impact the server in terms of a significant slowdown. This applies wit respect to the average response time as well as the number of requests the server can handle.

Throughput. The component was tested against basic throttling algorithms with a high as well as a low static limit. Instead of a rising target throughput, additional bad actors got added. These do not follow the OAuth 2 standard, but instead repeatedly request a resource from the resource server with a forged token to increased server load. The test results with respect to the average response time is show in Fig. 6a.



(a) Average response time

(b) Throughput

Fig. 6. Comparison of static vs. dynamic limits

It is expected, that a low limit grows slower than a high limit, which can be observed from 75 bad actors and above. A spike on the server with the low limit from zero to 75 bad actors can be observed, which happens due to the high average response times for the first few requests. These are the requests for the login page, which take longer to render than a request for a single resource. It also can be observed, that the server with a dynamic limit follows the static limit, which results in the lower average response time. The results regarding the achieved throughput are shown in Fig. 6b.

A higher limit is expected to result in a higher achieved throughput. With a rising server load, both high and low limits result in a lower achieved throughput, but the higher limit decreases stronger. Therefore these two curves cut at 75 additional bad actors. It can be observed that up until this point, the dynamic component results in a similar throughput than the high limit. After this point the results will be similar to the results of the low limit.

With respect to the average response time and the achieved throughput, the dynamic component results in a good middle ground between a high and a low limit to a certain extent. With a rising server load, the component was observed to result in similar values than the two static limits.

6 Conclusion

In this paper we showed how throttling by means and in the context of OAuth-based systems can be optimized.

To overcome the issue of indirection on the auth-servers' token introspection endpoint, we presented a cooperative throttling approach. It combines the clients' source IP address that is forwarded by the resource server with the provided access token to accurately identify user sessions. We showed that with growing amounts of unwanted traffic, response times are dramatically improved.

Regarding the applied throttling limits, we contribute a dynamic throttling solution. It links the throttling limit to the server load, which is measured in terms of the average response time. We showed that our approach is able to combine the advantages of high limits in low load scenarios with the advantages of low limits in high load scenarios.

Putting both parts together, we discussed how the two partial solutions can be integrated to achieve an overall optimized throttling mechanism for OAuth-based systems. Whilst the integrated architecture has been designed, yet no integrated implementation exists. Thus as future work, an implementation and evaluation of the overall approach is to be done.

References

1. Apache mod_evasive module. https://github.com/jzdziarski/mod_evasive. Accessed 15 June 2022
2. limitipconn2 readme. http://dominia.org/djao/limitipconn2-README. Accessed 15 June 2022
3. Rate limiting with nginx and nginx plus. https://www.nginx.com/blog/rate-limiting-nginx/. Accessed 15 June 2022
4. Abbott, M.L., Fisher, M.T.: The Art of Scalability: Scalable Web Architecture, Processes, and Organizations for the Modern Enterprise. Addison-Wesley (2015)
5. Hardt, D. (ed.): RFC6749 - The OAuth 2.0 Authorization Framework (2012)
6. Hardt, D., Jones, M.: RFC6750 - The OAuth 2.0 Authorization Framework: Bearer Token Usage (2012)
7. El Kamel, A., Eltaief, H., Youssef, H.: On-the-fly (D)DoS attack mitigation in SDN using deep neural network-based rate limiting. Comput. Commun. **182**, 153–169 (2022)
8. Richer, J. (ed.): RFC7662 - OAuth 2.0 Token Introspection (2015)
9. Miller, R.B.: Response time in man-computer conversational transactions. In: Unknown (ed.) Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, Part I on - AFIPS 1968 (Fall, Part I), p. 267. ACM Press, New York (1968)
10. Sakimura, N., Bradley, J., et al.: OpenID connect core 1.0 (2014)
11. Parecki, A.: OAuth 2. 0 simplified (2017). Lulu.com
12. Patil, R.Y., Ragha, L.: A dynamic rate limiting mechanism for flooding based distributed denial of service attack. In: Fourth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom2012), pp. 135–138. Institution of Engineering and Technology (2012)
13. Maenhaut, P.-J., Moens, H., Decat, M., et al.: Characterizing the performance of tenant data management in multi-tenant cloud authorization systems. In: 2014 IEEE Network Operations and Management Symposium (NOMS). IEEE (2014)
14. Raghavan, B., Vishwanath, K., et al.: Cloud control with distributed rate limiting. ACM SIGCOMM Comput. Commun. Rev. **37**(4), 337–348 (2007)
15. Richer, J., Sanso, A.: OAuth 2 in Action. Simon and Schuster (2017)
16. Roberto, S.D.O., da Silva, R.C., Santos, M.S., Albuquerque, D.W., Almeida, H.O., Santos, D.F.: An extensible and secure architecture based on microservices. In: 2022 IEEE International Conference on Consumer Electronics (ICCE). IEEE (2022)
17. SE, S.W.: SmartWe World SE | Unabhängige CRM Cloud-Plattform (2022). https://www.smartwe.de/. Accessed 15 June 2022
18. Shernan, E., Carter, H., Tian, D., Traynor, P., Butler, K.: More guidelines than rules: CSRF vulnerabilities from noncompliant OAuth 2.0 implementations. In: Almgren, M., Gulisano, V., Maggi, F. (eds.) DIMVA 2015. LNCS, vol. 9148, pp. 239–260. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20550-2_13
19. Siriwardena, P.: Advanced API Security: OAuth 2.0 and Beyond. Apress (2019)

20. Radhakrishnan, S., Geng, Y., Jeyakumar, V., et al.: SENIC: scalable NIC for end-host rate limiting. In: 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2014). USENIX Association, Seattle (2014)
21. Sun, S.T., Beznosov, K.: The devil is in the (implementation) details: an empirical analysis of OAuth SSO systems. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security, pp. 378–390 (2012)
22. Lodderstedt, T., et al.: RFC6819 - OAuth 2.0 threat model and security considerations (2013)
23. Urdaneta, G., Pierre, G., Van Steen, M.: Wikipedia workload analysis for decentralized hosting. Comput. Netw. **53**(11), 1830–1845 (2009)
24. Welsh, M., Culler, D.: Adaptive overload control for busy internet servers. In: Proceedings of the 4th Conference on USENIX Symposium on Internet Technologies and Systems, USITS 2003, vol. 4, p. 4. USENIX Association, USA (2003)

A Concept and a Multitenant Web Application for Interactive Software Architecture Analysis

Stefan Gudenkauf(✉), Uwe Bachmann, and Niklas Hartmann

Jade University of Applied Sciences, 26389 Wilhelmshaven, Germany
{stefan.gudenkauf,uwe.bachmann}@jade-hs.de

Abstract. The Architecture Tradeoff Analysis Method (ATAM) is a well-known method for the early evaluation of software architecture decisions based on the formulation of quality scenarios. Although widely recognized as beneficial, several limitations of ATAM have become apparent over time that limit its applicability. In this paper, we propose the concept of an interactive web-based application that mitigates some of these limitations while opening the possibility for further analysis based on collected ATAM projects, e.g. based on artificial intelligence. Thereby, we address the following limitations of ATAM: a general need for tool support, the need for an overview of ATAM results, the need for a consistent documentation of an ATAM process, the need for interactive and collaborative process execution among stakeholders, and the benefits of supporting multitenancy for architecture analysis. This applies in particular to systems with microservice architecture in order to document a large number of individual components and to make the documentation accessible to everyone involved in the system. The proposed concept, *Interactive Software Architecture Analysis* (ISAA), consists of a requirements analysis, an analysis model, a concept for interactive visualization, and a set of use cases. We demonstrate the feasibility of our approach by a software prototype in the form of custom modules for the Drupal Content Management System.

Keywords: Software architecture · Architecture evaluation · Tools to support architecture

1 Introduction

Software is playing an increasingly important role in society, with software systems becoming more and more complex [22]. Software architectures help to manage the complexity of software over its lifetime, and the consideration of quality attributes often serves as a basis for the analysis of one or more software architectures [23]. The Architecture Tradeoff Analysis Method (ATAM) is a widely known method for the early evaluation of software architectures regarding concrete quality scenarios [1, 2], and possibly the most prominent representative of *scenario-based* evaluation methods [1, 2, 22, 24]. In our opinion, the following aspects contribute to the popularity of ATAM. (1) ATAM's Quality Attribute Utility Tree supports to visually think about how to clarify quality requirements. Similar to mind maps, Utility Trees are simple organizational

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 268–283, 2023.
https://doi.org/10.1007/978-3-031-26886-1_16

structures that can be clearly arranged, represented, and compared to each other, which fosters accessibility and structure recognition. (2) ATAM analyzes architectural decisions in terms of risks to software quality instead of focusing non-risks. (3) ATAM has a strong emphasis on stakeholder involvement [4].

In this paper, we address the problem of adequate tool support for the analysis of software architectures based on ATAM. To do so, we present a survey of related work on ATAM and tool support, analyze top-level requirements and propose a concept for *interactive software architecture analysis* (ISSA). We discuss ISAA in terms of an analysis model, a concept for interactive visualization, and a set of use cases that tool implementers can consider as functional requirements. While the focus of the paper lies on this concept, we also demonstrate the feasibility of our approach by software prototyping. We also compare our prototype to other ATAM tool support approaches.

In Sect. 2, we present an overview of the ATAM, followed by a goal statement and methodology in Sect. 3, and an overview of related work in Sect. 4. Section 5 then looks at the limitations of the ATAM and distills top-level requirements. To meet these, Sect. 6 presents a concept dubbed *Interactive Software Architecture Analysis* (ISAA). In Sect. 7, we demonstrate the feasibility of our approach by developing a software prototype based on the Drupal Content Management System (CMS). Finally, we draw conclusions and discuss future work in Sect. 8.

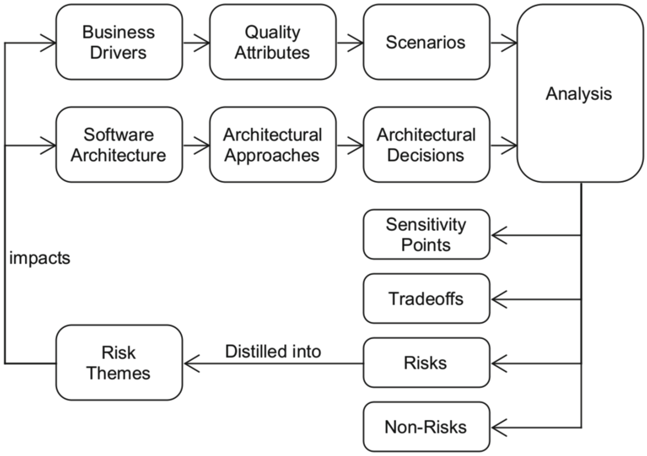


Fig. 1. Overview of the elements of the ATAM process.

2 The Architecture Tradeoff Analysis Method

The Architecture Tradeoff Analysis Method (ATAM) is a method for the early evaluation of software architectures developed at the Software Engineering Institute (SEI) of Carnegie Mellon University [1, 2]. The method relies on the formulation of quality scenarios, which stakeholders shall classify according to their domain importance and technical difficulty. This is supplemented by the presentation of the software architecture, the architectural approaches and decisions made on it. Finally, the stakeholders shall classify the effects of the architecture decisions according to a fixed scheme and form so-called risk themes, which are said to affect the software architecture as well as the original business drivers that led to the development of the software in the first place. Figure 1 shows an overview of the elements of the ATAM process and Table 1 subsumes the individual process steps of ATAM, which are executed in two phases: In the first phase, steps 1–6 are performed in a small meeting, then after a break, steps 1–9 are performed in a larger setting in phase 2.

The application of ATAM promises several benefits for software development [9]. The method focuses on identifying risks early in the software development life cycle. Quality Attributes are made explicit, and architecture documentation is promoted. This fosters the communication among stakeholders, and forms a documented basis for stakeholders to make well-founded architectural decisions. The main benefit therefore is improved software architectures.

Table 1. Steps of the ATAM process according to [1].

Step	Activity	Stakeholders
1	Present the ATAM	Evaluation team, customer representatives, architecture team
2	Present business drivers	See above
3	Present architecture	See above
4	Identify architectural approaches	See above
5	Generate Quality Attribute Utility Tree	See above
6	Analyze architectural approaches	See above
7	Brainstorm and prioritize scenarios	All stakeholders
8	Analyze architectural approaches	Evaluation team, customer representatives, architecture team
9	Present results	All stakeholders

3 Mission Statement and Methodology

We define our goal as follows using the Goal Question Metric (GQM) approach [11, 12]: Provide (issue) tool support (topic) for interactive and collaborative architecture evaluation (process) from a software architecture stakeholder's point of view (viewpoint). Based on this goal, we investigate three questions: (1) What is the current situation on the ATAM process? (2) What is the current situation on tooling (3) Can we provide and maintain *proper* tooling regarding limited resources? We address the first two questions through a literature review and the third question by software prototyping. Additionally, literature research, concept and software prototyping can be regarded as the *plan* and *do* phases of a PDCA cycle [25]. The documentation of our findings in this paper and the gathering of feedback then represents the *check* phase. Finally, the revision of the concept and prototype represents the *act* phase before the next iteration of the PDCA cycle may follow.

4 Related Work

Over time, researchers published various articles on ATAM. We consider the following categories: (1) articles on the application of ATAM, (2) articles on improving or adapting the ATAM process, and (3) articles on ATAM tool support. In the following, we limit ourselves to selected work from the last two categories.

Zalewski argues that ATAM is difficult to apply to the early evaluation of large-scale software architectures [10]. This is because top-level architectural decisions, so called *system organization patterns*, are difficult to relate to the final characteristics of a software system, and because the evaluation of stakeholder expectations would require a sufficiently detailed design. As a framework for evaluating system organization patterns, Zalewski proposes the use of the Goal Question Metric (GQM) approach [11, 12] instead of ATAM. We find the notion of system organization patterns and their aspects as potentially helpful for the documentation of architectural decisions of large-scale software systems within an ATAM process. We also find the goal definition of the GQM as particularly helpful for the definition of the overall business goal related to the business drivers in the ATAM process.

Lionberger and Zhang present ATAM Assistant [5], an application that supports the ATAM workflow focusing on data collection and organization at each phase of the ATAM process. Data collection is realized by windows form based interaction (dialogs and a multiple document interface). Result artifacts, such as the quality attribute utility tree, are presented in tree view structures. At the end of the process, the user can generate a report in HTML format. The author denote the goal of the tool to reduce software architecture evaluation to one person.

In [3], Gabel notes that at the time of writing, there is a lack of tool support for documenting the results of an ATAM process, which to our knowledge is still the case today. He notes that literature often shows the application of ATAM only with spreadsheet-based reports, and that the mapping of risks, tradeoffs, and sensitivity points to the appropriate architectural decisions is often not directly apparent to the stakeholders. As a solution, Gabel proposes the application of software cartography [19], and implemented a single user desktop application to document and report the application of ATAM as a proof of feasibility.

Maheshwari and Teoh analyze the social and environmental aspects of the ATAM process [4]. To address these aspects, they envision a collaborative web-based tool named ACE to allow stakeholders to participate in ATAM processes without having to be physically collocated. Stakeholders communicate synchronously in ACE.

The authors Ågren et al. investigate how architecture evaluation can provide useful feedback during the development of continuously evolving systems and show the importance of dedicated feedback loops during the development of such systems [13].

5 Requirements Analysis

As part of our research for related work, we found that tool support for performing ATAM processes is still scarce. The first requirement is therefore a basic tool support to simplify the execution of the ATAM process for stakeholders (req. 1). This tool support should uniformly document the execution of ATAM processes, so that stakeholders can compare and computers can process individual ATAM processes (req. 2). This opens up the possibility of analyzing architecture variants, and machine learning from process executions that have already taken place, for example, to make suggestions for architectural decisions based on existing business drivers and quality scenarios. Further requirements lie in the documentation of ATAM results. We agree with Gabel [3] that the relationships between risks, tradeoffs and sensitivity points should be presented to the stakeholders in a way that makes them as obvious as possible (req. 3). In this way, stakeholders do not have to resolve these references between tables within reports "in their minds" (see Fig. 2), but can draw direct conclusions from the relationships and can focus on forming risk themes.

Architectural decisions	Risk	Sensitivity	Tradeoff
Backup CPU(s)	R8	S2	
No backup Data Channel	R9	S3	T3
Watchdog		S4	
Heartbeat		S5	
Failover routing		S6	

Fig. 2. Screenshot from Kazman et al. [1] showing a part of an architectural approach documentation. Note that the reader must mentally resolve the references of architectural decisions to risks, sensitivity points and tradeoffs.

However, in contrast to [3], we consider an interactive visualization [6–8] of the results of an ATAM process to be more useful than generating static reports (req. 4). The reason is that a continuous interactive visualization of ATAM results in different iteration stages enables direct feedback loops (cf. [13]) and agile procedures. Additionally, as in [4], stakeholders should be able to participate in ATAM processes without having to be physically collocated, requiring remote work and possibly collaborative editing (req. 5). Finally, we consider multi-tenancy to be an important property of an ATAM tool (req. 6). The following considerations are paramount for this: (1) Cost savings: We only have to provide customers ATAM services using a single tool instance. (2) Leveraging data: Data from all ATAM processes is available in a uniform digital format and will be available for data mining and machine learning in the future. Table 2 subsumes the identified top-level requirements.

6 Concept

As stated in the previous section, scenario-based architecture analysis with ATAM faces several limitations. To address these, we propose the concept of *interactive software architecture analysis* (ISAA) in terms of an ISAA metamodel, a concept for interactive visualization, and an overview of the main use cases for tool support.

Table 2. Top-level requirements for interactive software architecture analysis.

No	Requirements
1	Tool support
2	Consistent and digital process documentation
3	Overview on analysis results
4	Interactive visualization
5	Remote work support
6	Support for multitenancy

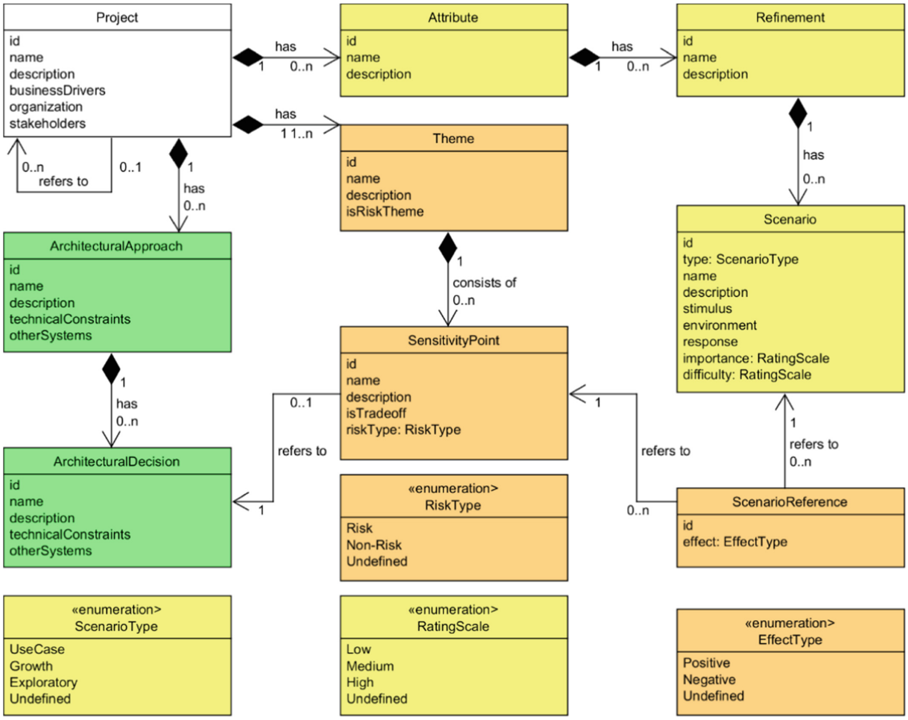


Fig. 3. The ISAA metamodel. Entities that relate to the Quality Attribute Utility Tree are colored yellow, entities that relate to software architecture are colored green, and elements that relate to sensitivity points, treadeoffs, non-risks, and risks are colored beige.

6.1 A Model for Interactive Software Architecture Analysis

The ISAA model consists of the following parts: (1) elements that represent the Quality Attribute Utility Tree, (2) elements that represent architectural approaches and architectural decisions, and (3) elements that represent sensitivity points, tradeoffs, non-risks and risks, and their classification into so-called risk themes. Figure 3 shows the ISAA model in form of an analysis class diagram using the Unified Modeling Language. Since the individual entities refer directly to elements of the ATAM, we refrain from a detailed discussion of all entities. However, we would like to present the following rationale for the model:

1. We consider the three parts of the ISSA model as tree structures, so that later serialization in data exchange formats should be simple. UML composition associations and one-to-many multiplicities between model entities form the respective tree structures.
2. The tree structure of sensitivity points, tradeoffs, risks and non-risks requires that they are initially classified in an anonymous risk theme. During the ATAM process, the stakeholders then assign them to dedicated risk themes.
3. We interpret the explanations of sensitivity points, tradeoffs, risks and non-risks given in [1] and [14] as follows:

    a. A sensitivity point represents the most basic relationship between an architectural decision and a quality scenario.
    b. We automatically denote a sensitivity point as a tradeoff if the architecture decision influences several quality scenarios. This is represented by the Boolean flag `isTradeoff`.
    c. Stakeholders can explicitly classify a sensitivity point or tradeoff as a risk or non-risk. This is represented by the enumeration `RiskType`.

4. Since the impact of sensitivity points, tradeoffs, risks and non-risks on quality scenarios can potentially be regarded as positive or negative, we added an entity `ScenarioReference` to represent impacts using an enumeration `EffectType`.
5. A Quality Attribute Untility Tree is represented by instances of `Attribute`, `Refinement` and `Scenario` for a given `Project` instance. For attributes and refinements, we recommend using well-known characteristics such as these provided by the product quality model of the ISO/IEC 25010:2011(E), see [15].
6. As instances of the entity `ArchitecturalApproach` may represent the *system organization patterns* of large-scale software systems as discussed in [10], we consider the topics proposed in [10] when describing such systems: Subsystem decomposition, geographical and/or organizational allocation of subsystems, data input organization, data storage distribution, data processing organization, distributed data storage management, transaction management, communication framework.
7. Business drivers that lead to the development of a software to be evaluated by ATAM should be documented in the `businessDrivers` attribute of instances of `Project`. The `description` attribute may additionally be used to document business goals formulated according to the GQM approach [11].

8. The entity `Project` represents the root node of an evaluation project. To support the analysis of the architectures of large software systems with many components, a project can refer to other projects as *related*. For example, the evaluation of a single microservices component of a microservices application [20, 21] may refer to the evaluation of the overall macro architecture of the same application.
9. We considered the individual types of quality scenarios as an enumeration `ScenarioType` used by `Scenario`.
10. Each enumeration allows `Undefined` as a value for properties not (yet) defined.

6.2 Interactive Software Architecture Analysis Visualization

"Dynamic, interactive visualizations can empower people to explore the data for themselves.", Murray aptly writes in [6]. Such visualization for ATAM aims that stakeholder can track and examine the results of an ATAM process as it is carried out. In general, cognitive load should be kept as small as possible. To do so, a suitable tool support must expose the references drawn between process elements and make them explorable to the viewer. Our visualization concept consists of several views, which we discuss in terms of how they support interactivity and which benefits we see in them.

Quality Attribute Utility Tree View. This view shows an overview of the Quality Attribute Utility Tree of the ATAM, including quality attributes, refinements, scenarios, and how they are connected. The view shows a typical tree structure that was already used in [1] to statically illustrate a Quality Attribute Utility Tree. However, we consider the project itself as the root of the tree. Quality attributes, refinements and scenarios then form the rest of the nodes. Figure 4 shows an example.

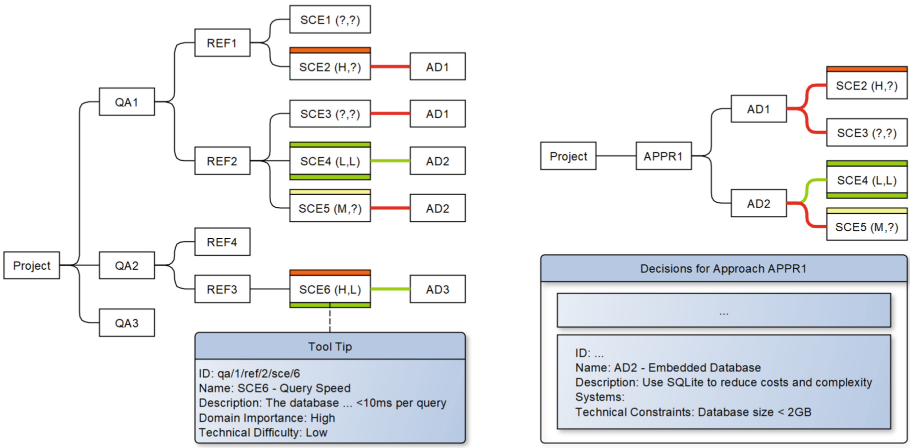


Fig. 4. Examples of a Quality Attribute Utility Tree View (left) and an Architectural Decisions Tree View (right).

Once architectural decisions on sensitivity points, trade-offs, risks, and non-risks have been identified during the process, these architectural decisions should also be explicitly displayed in the Quality Attribute Utility Tree View, associated with the scenarios they affect.

For the sake of clarity, we suggest that the nodes should only display the name of the respective ATAM element. However, we use the following color scheme to encode further information: Scenario nodes can have colored borders above and below their respective node element's name. These indicate the *domain importance* (above) and *technical difficulty* (below) of each scenario once the stakeholders have rated it. A green border represents *low*, yellow represents *medium*, and red/orange represents *high*. In addition, associations between architectural decisions and the scenarios they affect are colored according to the type of impact, if specified. A red color indicates a negative impact and a green color indicates a positive one.

Mouse hover actions should be used to present tooltips with tabular data tailored to the respective node element. More information about each node is made available by clicking on the respective node. This approach adheres to the "Visual Information-Seeking Mantra" *overview first, zoom and filter, then details-on-demand* [7]. It provides a broad overview of how scenarios relate to quality attributes first, which the viewer can supplement with more in-depth information as required.

Architectural Decisions Tree View. The Architectural Decisions Tree View is similar to the Quality Attribute Utility Tree View. Architectural approaches and architectural decisions are presented. If stakeholders already have associated these with scenarios via sensitivity points, tradeoffs, risks and non-risks, the scenarios are also displayed in this view. As in the Quality Attribute Utility Tree View, associations between architectural decisions and the scenarios they affect are colored according to the type of impact, if specified. An example is shown in Fig. 4.

Sensitivity Impact Network View. This view focuses on Sensitivity Points. Instead of a tree structure, we consider the view to be a force-directed graph that visually emphasizes interdependent elements (Fig. 5). The central element of the view is the project as root. The Quality Attribute Utility Tree is placed around the project, with the position of each element being calculated by a force-based algorithm. Architectural decisions that stakeholders have not yet associated with scenarios float as free nodes.

In the Sensitivity Impact Network View, stakeholders can interactively specify the relationships of sensitivity points to architectural decisions and scenarios (e.g. by context menu actions or through drag-and-drop mouse gestures). Stakeholders can also assess the respective impacts as positive or negative. The connections evaluated in this way are colored green or red accordingly. Each new connection then updates the graph's representation using the force-based algorithm. Stakeholders can also identify sensitivity points as risks or non-risks. Risks are then colored red and non-risks green.

Whether a sensitivity point is a tradeoff is automatically determined based on the number of connections to scenarios and the effect types of the connections. With more than one connection, a sensitivity point is considered a tradeoff when different effect types come together (at least one positive and one negative). The sensitivity point can then be marked with a small *T* symbol. Scenarios are highlighted as in the other views.

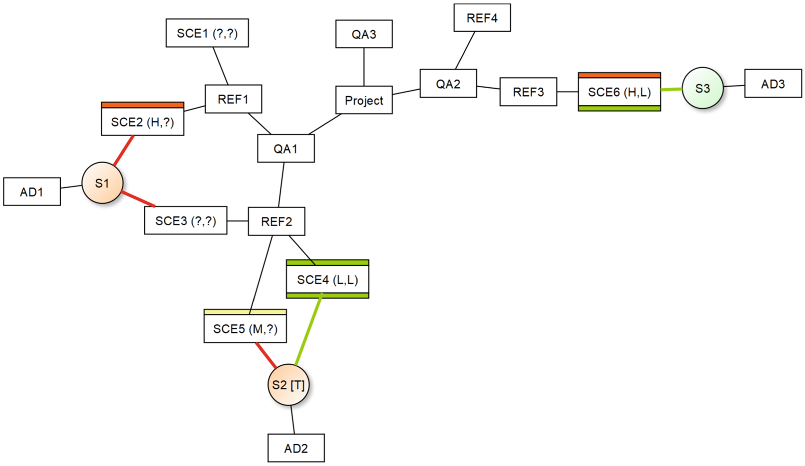


Fig. 5. Example of a Sensitivity Impact Network View from the ISAA prototype.

6.3 Use Cases Model for Tool Support

We consider the following top-level use cases for *interactive software architecture analysis*, arranged by topic (Fig. 6). Implementers can consider these as functional requirements for ISSA tool support.

User Management. Users must be able to register themselves and to authorize themselves once registered. We currently consider the following users: *Guest* users can register themselves. Once registered, *stakeholders* have view rights to the project assigned to them. *Project owners* have full rights to their projects, and may reference other projects, in which they are involved as stakeholders, as *related* (see Sect. 5.1, item 8). They also can assign registered users as stakeholders to their projects. *Domain super users* have view rights to all projects, and *admins* have full rights to all projects and can assign roles to users.

Project Management. Users must be able to manage their projects, which includes creating a new project, opening and updating an existing project, and deleting a project.

Requirements Elicitation. Within a project, users must be able to construct a Quality Attribute Utility Tree including attributes, refinements and scenarios. They also must be able to evaluate a given scenario according to its domain importance and its technical difficulty.

Architecture Analysis. Users must be able to define architectural approaches and decisions. In addition, they must be able to define and classify sensitivity points, as well as to define risk themes.

Result Presentation. The system must provide visual representations of ATAM results according to the interactive visualization concept described in the previous subsection. In addition, while not in the focus of the concept, report generation should also be possible (at least by using a web browser's print function).

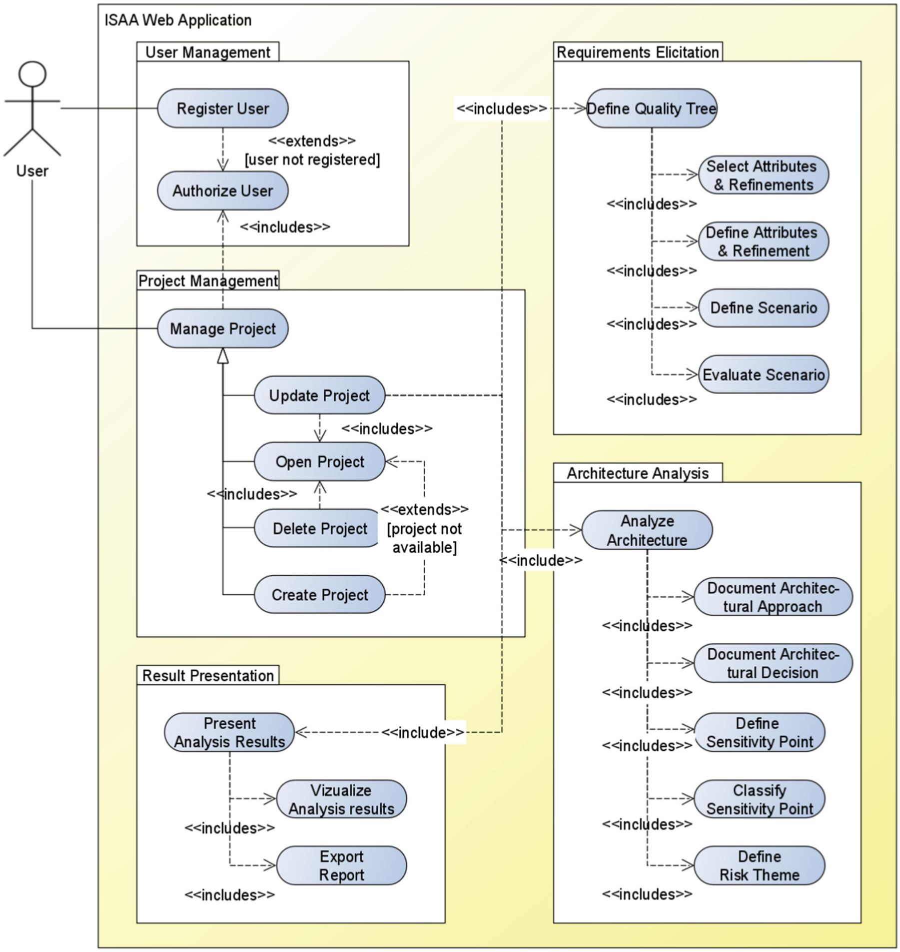


Fig. 6. Use cases of the application.

7 Implementation

As a proof of feasibility, we implemented a software prototype for *interactive software architecture analysis* (ISSA) realized as web application based on the widely-used Linux, Aache, MySQL, and PHP (LAMP) application stack. In addition, we chose the Drupal content management system [16] as a basis for development for the following reasons [17]: Drupal is built upon the PHP web development language. It can use several database systems for persistent data storage, such as MySQL. Drupal also has a large developer community. Finally, as a distributed application framework, it is regarded highly scalable, provides several basic needs of distributed business applications such as user and role management, and is expandable with additional functionalities via a well-established module concept. Therefore, we realized the ISAA web application as a set of custom Drupal modules that handle ATAM process data input and result visualization according to the concept discussed in the previous section. The module architecture of the ISAA web application is presented in Fig. 7. The focal point of the modules is the notion of an evaluation project that the underlying database system persists and that is serialized as a JSON file. The custom ISAA Drupal modules consists of the following components:

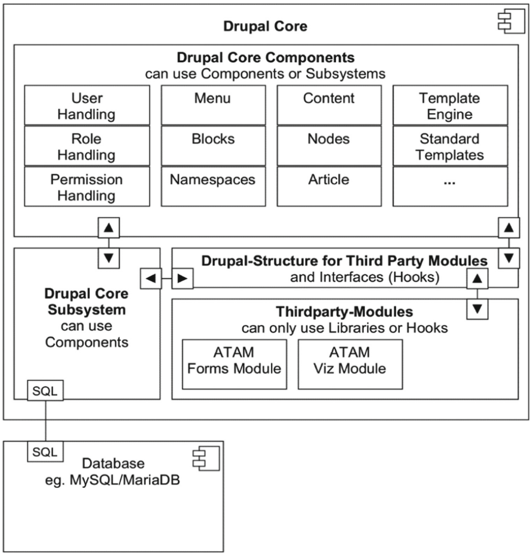


Fig. 7. Module structure of the ISAA web application.

ATAM Forms Module. A Drupal module for data collection and management within an evaluation project via web forms. Table 3 shows the module's structure and Table 4 presents the modules web API regarding the quality attribute utility tree. As of this writing, the module supports data collection related to the overall project and the quality attribute utility tree.

Table 3. Structure of the ISAA prototype's ATAM Forms Module.

Module Structure	Description
atam	Main folder; contains database setup files, libraries, permissions etc.
atam/src	Custom source code
atam/src/Controller	Controllers for the module
atam/templates	Custom templates used by Drupal for page generation
atam/src/Form	Custom forms conforming to the Drupal Form API

(*continued*)

Table 3. (*continued*)

Module Structure	Description
atam/css	Custom layout stylesheets
atam/img	Image files
atam/js	JavaScript files

ATAM Viz Module. A Drupal module for the interactive visualization of evaluation projects. Its module structure is similar to the ATAM Forms Module. A `LocalStorageController` loads data from the database into a viewer's web browser and vice versa for fast manipulation and persistent storage across sessions. Further controller classes handle data transformation and presentation logic. The actual visualization and generation of UI elements is realized at the presentation level and almost exclusively with JavaScript. The visualization thus reacts interactively to the user and can continuously update itself.

Table 4. URL patterns for data collection in the ISAA prototype regarding the Utility Tree.

URL pattern	Meaning
/project	Empty project form
/projects	List of projects
/project/123	Project with identifier 123
/project/123/json	JSON representation of project 123
/project/123/qa	Empty quality attribute form
/project/123/qas	List of quality attributes
/project/123/qa/456	Quality attribute with identifier 456
/project/123/qa/456/ref	Empty refinement form
/project/123/qa/456/refs	List of refinements
/project/123/qa/456/ref/789	Refinement with identifier 789
/project/123/qa/456/ref/789/sce	Empty scenario form
/project/123/qa/456/ref/789/sces	List of scenarios
/project/123/qa/456/ref/789/sce/012	Scenario with identifier 012

8 Conclusion and Future Work

We dedicated this paper to the problem of adequate tool support for the analysis of software architectures based on ATAM. To do so, we researched related work and analyzed top-level requirements. Based on this we proposed a concept for *interactive software* *architecture analysis* (ISSA) that we presented in terms of an analysis model, a concept for interactive visualization, and a set of use cases that tool implementers can consider as functional requirements. While the focus of the paper lies on this concept, we also demonstrate the feasibility of our approach using a software prototype that we realized in form of portable modules for the Drupal content management system. In addition, we also describe the architecture of the prototype. While researching related work, we discovered several software tools to support ATAM. Table 5 compares these to our ISAA software prototypes based on the top-level requirements we have identified. Since the goals of the individual tools are different, we also state them explicitly.

Table 5. Comparison of ATAM software tools.

No	Requirement	ATAM Tool [3]	ACE [4]	ATAM Assistant [5]	Our Prototype
1	Tool support	Yes, but terms of use unclear	Availability unclear	Availability unclear	Yes, but very early prototype
2	Data collection	Yes, form-based	Synchronous comm., utility tree java applet	Yes, form-based, stored using XML	Yes, form-based
3	Result overview	Yes, software cartography, report generation	Not mentioned in list of features	Tree list views, report generation	Yes, by interactive visualizations
4	Interactive visualization	No, static maps	Not mentioned in list of features	No	Early prototype
5	Remote work support	No, single user desktop application	Yes, web application	No, single user desktop application	Yes, web application with different user roles
6	Multitenancy	No, single user desktop application	Not discussed, but also not excluded	No, single user desktop application	Yes
	Goal	*"This software system serves to […] show that there is sensible software support for the ATAM process and that a clear visual representation of the results is possible."* (translated from German)	*"[Provide] a common environment where stakeholders and software evaluators alike can take part in ATAM evaluation without having to be physically collocated."*	*"Support the ATAM workflow and manage the various inputs and outputs of this process." "The goal of this software tool is reduce [software architecture evaluations] to one person."*	Demonstrate the feasibility of our approach, which is interactive software architecture analysis (ISSA) based on the ATAM

Regarding future work, we plan to extend and consolidate the ISAA concept and our web application prototype, and to evaluate it according to Moody's Method Evaluation Model [18]. Students of the Software Engineering course can most likely carry out an initial evaluation. As soon as tooling is sound, we would like to use it in research projects with our industrial partners. We also consider analyzing architectural variants for ISAA tool support using ATAM, which we then like to compare to the architecture of our current ISAA prototype. As soon as we collected a sufficient number of ATAM processes, we also like to investigate how data mining and machine learning can utilize process data to improve future ATAM processes, for example by recommendations on similar business drivers, similar scenarios, and similar architectural decisions. Finally, a closer look at the evaluation of the architectures of large software systems with many components, e.g. microservices architectures, can reveal additional challenges to ATAM and ISAA. How, for example, can ATAM variants cope with the *complexity explosion syndrome* [10] observed by Zalewski? A continued use of ATAM-defined artifacts provided by ISAA throughout the entire software development lifecycle might provide part of an answer, as well as extending the method by the notion of evaluation project groups, hierarchies, or dependency references between evaluation projects (as discussed in Sect. 6.1).

References

1. Kazman, R., Klein, M., Clements, P.: ATAM: Method for Architecture Evaluation. Fort Belvoir, VA (2000). https://resources.sei.cmu.edu/asset_files/TechnicalReport/2000_005_001_13706.pdf. Accessed 2 April 2022
2. Kazman, R., Klein, M., Barbacci, M., Longstaff, T., Lipson, H., Carriere, J.L.: The architecture tradeoff analysis method. In: Proceedings. Fourth IEEE International Conference on Engineering of Complex Computer Systems (Cat. No.98EX193). Fourth IEEE International Conference on Engineering of Complex Computer Systems. ICECCS '98, Monterey, CA, USA. 10–14 Aug. 1998. IEEE Comput. Soc, pp. 68–78 (1998). https://doi.org/10.1109/ICECCS.1998.706657
3. Gabel, A.: Softwareunterstützung für die Evaluation von Softwarearchitekturen mit der Architecture Tradeoff Analysis Method. Carl von Ossietzky Universität Oldenburg, Oldenburg, Thesis (2013)
4. Maheshwari, P., Teoh, A.: Supporting ATAM with a collaborative web-based software architecture evaluation tool. Sci. Comput. Program. **57**(1), 109–128 (2005). https://doi.org/10.1016/j.scico.2004.10.008
5. Lionberger, B., Zhang, C.: ATAM assistant: a semi-automated tool for the architecture tradeoff analysis method. In: Proceedings of the 11th IASTED International Conference on Software Engineering and Applications, November 19–21, pp. 330–335. Cambridge, Massachusetts, USA (2007)
6. Murray, S.: Interactive data visualization for the web: An introduction to designing with D3. O'Reilly, Beijing, Köln (2013)
7. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations: August 14–16, 1996, Blue Mountain Lake, New York. IEEE Computer Soc. Press, LosAlamitos, Calif. (1996)
8. Yau, N.: Visualize this: The FlowingData guide to design, visualization, and statistics. Wiley, Indianapolis, Ind. (2011)
9. Software Engineering Institute: Architecture Tradeoff Analysis Method Collection (2018). https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=513908. Accessed 24 May 2022
10. Zalewski, A.: Beyond ATAM: Architecture Analysis in the Development of Large Scale Software Systems. In: Software Architecture, ed. Flavio Oquendo. Springer Berlin Heidelberg, Berlin, Heidelberg pp. 92–105 (2007)

11. Basili, V., Caldiera, G., Rombach, H.D.: Goal Question Metric Approach. In: Encyclopedia of Software Engineering, pp. 528–532. John Wiley & Sons, Inc. (1994)
12. van Solingen, R., Basili, V., Caldiera, G., Dieter Rombach, H.: Goal Question Metric (GQM) Approach. In: John, J.M. (ed.) Encyclopedia of software engineering. Wiley, New York (2002)
13. Ågren, S.M., et al.: Architecture evaluation in continuous development. J. Syst. Softw. **184**, 111111 (2022). https://doi.org/10.1016/j.jss.2021.111111
14. Masak, D.: Der Architekturreview. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
15. ISO/IEC: ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models (2011). Accessed 22 May 2022
16. Drupal.org: Drupal - Open Source CMS (2018). https://www.drupal.org/. Accessed 12 June 2022
17. Sinha, A.U., Merrill, E., Armstrong, S.A., Clark, T.W., Das, S.: eXframe: reusable framework for storage, analysis and visualization of genomics experiments. BMC Bioinformatics **12**, 452 (2011). https://doi.org/10.1186/1471-2105-12-452
18. Daniel, L.M.: The method evaluation model: a theoretical model for validating information systems design methods. In: ECIS 2003 Proceedings (2013)
19. Wittenburg, A.: Software cartography: models and methods for the systematical visualization of application landscapes. Doctoral thesis. Technical University of Munich, Munich (2007). https://d-nb.info/988065851/34. Accessed 14 June 2022
20. Lewis, J., Fowler, M.: Microservices: a definition of this new architectural term (2014). https://martinfowler.com/articles/microservices.html. Accessed 14 June 2022
21. Wolff, E.: Microservices: Grundlagen flexibler Softwarearchitekturen, 2nd edn. Heidelberg: dpunkt.verlag (2018)
22. Mårtensson, F.: Software architecture quality evaluation: Approaches in an industrial context. Blekinge Institute of Technology, Karlskrona (2006)
23. Bengtsson, P.: Architecture-Level Modifiability Analysis. Blekinge Institute of Technology, Ronneby (2002)
24. Kazman, R., Bass, L., Abowd, G., Webb, M.: SAAM: A Method for Analyzing the Properties of Software Architectures. Fort Belvoir, VA (2007)
25. Arredondo-Soto, K.C., Blanco-Fernandez, J., Miranda-Ackerman, M.A., Solis-Quinteros, M.M., Realyvasquez-Vargas, A., Garcia-Alcaraz, J.L.: A Plan-Do-Check-Act Based Process Improvement Intervention for Quality Improvement. IEEE Access **9**, 132779–132790 (2021). https://doi.org/10.1109/ACCESS.2021.3112948

An Ontology for Software Patterns: Application to Blockchain-Based Software Development

Nicolas Six(✉), Camilo Correa-Restrepo, Nicolas Herbaut, and Camille Salinesi

Centre de Recherche en Informatique (CRI), Université Paris 1 Panthéon-Sorbonne, Paris, France
{nicolas.six,camilo.correa-restrepo,nicolas.herbaut, camille.salinesi}@univ-paris1.fr

Abstract. Ensuring the quality of software design is usually a difficult task. In the blockchain field, the design of an application is particularly important as flaws can lead to critical vulnerabilities and cost overheads. To assist practitioners in this task, software patterns can be used (solutions to repeatable problems in a given context). Some blockchain patterns exist but they are scattered, and described in many different notations and templates. As a result, practitioners can be lost in the selection of adequate blockchain-based patterns. This paper fills the gap by proposing a blockchain-based software pattern ontology. The ontology is composed of two distinct subspaces: first, a set of classes and individuals related to blockchain-based patterns, based on a previous Systematic Literature Review (SLR). It notably reuses a taxonomy of blockchain design patterns, that helps to classify these patterns in relevant categories. Along that, another subspace has been created to further organize the knowledge related to software patterns and allow inferences. A tool is proposed along with the ontology to assist practitioners in finding blockchain-based patterns that fit their needs. An evaluation is performed to assess the usability and the relevancy of the ontology.

Keywords: Blockchain · Software design · Knowledge engineering

1 Introduction

A blockchain is a data structure where each block is linked to the previous one with a cryptographic hash. The addition of new blocks is ruled by a network of peers that each holds a copy of the blockchain (so-called blockchain network). Each block stores a list of transactions, that represent interactions between a user and the blockchain. Following the release of Ethereum in 2015, some blockchain solutions started to support Turing-complete smart contracts, enabling a new range of applications. A smart contract is a computer program that executes predefined actions when certain conditions within the system are met [1]. Hence, users can interact with smart contracts using transactions.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 284–299, 2023.
https://doi.org/10.1007/978-3-031-26886-1_17

The inner workings of blockchain differ from traditional database technologies, granting blockchain many unique qualities. First, blockchain is decentralized: there is no single central actor in charge of the network. The level of decentralization varies depending on the blockchain type [23]. Then, blockchain guarantees the immutability of data[1], as it is impossible to alter blocks after their addition. Finally, blockchain is transparent: any user that has access to a node can see past transactions and stored data. Transparency is not absolute: for instance, some blockchain technologies support private smart contracts, where data and execution are restricted to a set of authorized parties.

Blockchain directly addresses use cases that are difficult to solve with mainstream IS (Information System) technologies, such as ERPs. For instance, supply-chain applications benefit from blockchain immutability for goods traceability [2]. In this use case, trust is improved between supply-chain participants, as malicious actions to alter traceability data will be seen by everybody on the network. Blockchain smart contracts might also enable trustable service automation, such as the ones proposed by DeFi (Decentralized Finance) platforms [13].

Because of the specific properties of blockchain, a growing number of software architects undertake designing blockchain-based applications (BBA). A BBA can either be an application using blockchain as an improved platform to serve existing use cases (e.g., supply-chain traceability), or an application made feasible only by using blockchain technologies (e.g., on-chain decentralized cryptocurrency exchanges). However, they found the creation of BBAs to be a tedious task in both cases. Indeed, BBAs can *suffer* from their own qualities in certain contexts. For instance, transparency and immutability can be a burden when dealing with personal data, that might be subject to data protection regulations such as the General Data Protection Regulation (GDPR). Also, blockchain lacks certain native capabilities due to the way it functions, such as the ability to query external data or store large amounts of data.

Practitioners often employ well-known reusable recipes in the software design phase to improve software quality. Such recipes, known as patterns, are a solution to commonly occurring problems in given contexts. For instance, one of the best-known blockchain patterns is the *Oracle* [22]: as a smart contract cannot query data from outside the blockchain, using this pattern consists of sending fresh data to a smart contract when needed. Up to now, only a few patterns have already been proposed, grouped in collections (e.g., [22]) or standalone (e.g., [18]). Moreover, these patterns are scattered across the literature, making the task of identifying adequate patterns for a specific design difficult. Where a previous Systematic Literature Review (SLR) [17] tackled these issues by gathering these patterns in a collection, more work is needed to structure and formalize this knowledge, as well as facilitate its reuse.

In this paper, the following question is investigated: *How to formalize existing knowledge on blockchain-based software patterns to facilitate its reuse by practitioners?* To address this research question, this paper proposes a blockchain-based software pattern ontology, spanning two interrelated subspaces: (1) blockchain software patterns descriptions and (2) pattern provenance. While

[1] This assumption is only valid if the network is not compromised.

the former focuses on the classification of blockchain patterns established in the aforementioned SLR, the latter gathered knowledge on the original material proposing the pattern and its relationship with the state of the art, hence distinguishing patterns found in the literature from the concept itself. Using an ontology facilitates the inference between linked patterns, using the knowledge found in studies. To explore the ontology, we created an open-source tool[2] that provides, from a series of questions, personalized pattern recommendations. Our ontology validation methodology is twofold. First, we assess that the ontology fulfills specified requirements. Then, we evaluate our ontology-based recommendations over a selection of papers from the literature. The rest of the paper is organized as the following: Sect. 2 introduces existing blockchain-based pattern collections and the blockchain-based software pattern ontology, then Sect. 3 presents the research method employed to build the ontology as well as its requirements. Section 4 describes the resulting ontology, and Sect. 5 presents the validation phase to assess the relevance of the ontology. Finally, threats of validity are discussed in Sect. 6, and Sect. 7 concludes with future works.

2 Related Works

The literature shows that the idea of using ontologies to describe software patterns has already been explored. In [10], Kampffmeyer et al. propose an ontology derived from GoF (Gang-of-Four) design patterns [5]. Each pattern is linked to a set of design problems it solves, along with a tool to help practitioners select patterns without having to write semantic queries. However, their ontology does not bring out any dependency link between the patterns themselves. Our contribution reuses the concept of problem ontology and extends it, as shown in Sect. 4. Another ontology for software patterns is proposed in [6], that encompasses not only design patterns but also architectural patterns and idioms. A pattern is described using different attributes (such as *Problem*, *Context*, *Solution*, etc.), and can be linked to other patterns through a pattern system and relations (e.g., require, use). A similar metamodel for software patterns is proposed in [9]. Some differences can be mentioned, such as the possibility to specify that two patterns conflict with each other, or the *seeAlso* relationship to indicate other patterns related to a specific pattern. In addition, [11] proposes a design pattern repository taking the form of an ontology. The contribution enlights tedious knowledge management and sharing with traditional pattern collections and argues for a structured ontology format. The proposed ontology groups patterns into pattern containers, where one pattern can belong to many containers. Patterns can also be linked to a set of questions and answers, elicited from expert knowledge, through an *answer relevance* attribute. It indicates how relevant a pattern is in addressing a specific question. This contribution follows a similar path to that taken by our ontology by structuring a set of blockchain-based patterns.

Some ontologies have been proposed for modeling the blockchain domain and its components, such as that proposed by De Kruijff and Weigand [4], that

[2] https://github.com/harmonica-project/blockchain-patterns-ontology.

of Ugarte-Rojas and Chullo-Llave [8], and that of Glaser [7]. Another work by Seebacher and Maleshkova [14] focuses on modeling the characteristics of blockchains within corporate networks and their use. However, there is still a gap in the usage of ontologies to store blockchain-based patterns.

Finally, another decision model for the blockchain-based pattern can be mentioned [21]. In this model, a BPNM (Business Process Model and Notation) guides the process of selecting design patterns during the design phase of the software. Each activity in this process is further described using a BPMN-like notation, where the advantages and drawbacks of using a pattern are highlighted on each arrow leading to a pattern. This approach allows fine-grained guidance in the selection of blockchain patterns, but requires more work upfront to include patterns in the model and makes difficult the addition of patterns from outside the scope defined by the main BPMN.

3 Methodological Approach

The proper design of an ontology relies on the usage of a reliable and proven method. To build the ontology, the NeOn method [19] has been chosen due to its inherent flexibility and focus on the reuse of both ontological and non-ontological resources. However, these resources must be handled differently, as they can take multiple forms (e.g. whitepaper, publication, etc.). More information on handling non-ontological resources is given in Subsect. 3.2. NeOn does not force rigid guidelines upon the ontology designer: a set of scenarios is given and the designer is free to select and adapt any scenario that suits their needs.

In this study, we base our approach on two of the scenarios envisaged within NeOn. The first scenario mainly concerns ontology construction from the ground up, to produce a new, standalone, ontology. The principal motivation for this choice is the absence of literature on existing ontology covering blockchain-based patterns. There is also the inability of existing software pattern ontologies to adequately capture the results of the SLR upon which we base our pattern proposal ontology, hence our need to produce a standalone ontology to cover our particular domain of interest. The second addresses the specific aspects of reusing non-ontological resources in the construction of ontologies. This is key since our ontology will be primarily based on the reuse of previous results. The NeOn methodology proposes a set of closely related life cycle models linked to the different scenarios it incorporates. In our case, given our need to reuse non-ontological resources, the six-phase waterfall life cycle has been chosen.

3.1 Ontology Requirements Specification

One important step in the construction of a sound ontology is the specification of requirements through an Ontology Requirement Specification Document (ORSD) [19] that serves as an agreement on which requirements the ontology should cover, its scope, implementation language, intended uses and end users. The ORSD facilitates the reuse of existing knowledge-aware resources in the

Table 1. Ontology competency questions.

CQ1	What are the classes of patterns in the blockchain area and how can they be differentiated and characterized?
CQ2	What are the different propositions of patterns in the academic literature and how can their acceptance by others be quantified?
CQ3	How can we differentiate the concept of pattern from their possible descriptions in different sources?
CQ4	What are the types of relations that can connect two patterns together?
CQ5	What are the different problems bound to the design and implementation of blockchain applications?

creation of new ontologies [19]. Competency questions (CQs) are a way to introduce the functional requirements (FRs) of an ontology; their coverage, ideally in a generalizable manner, allows one to consider the ontology functionally complete. The CQs are not formulated as FRs, but rather as questions that can be translated to requirements afterward (e.g. for CQ4, the ontology shall allow the user to retrieve the possible relations between two patterns). For the sake of brevity, only the CQs are detailed in this paper, listed in Table 1. Nevertheless, the full ORSD is available on GitHub (aforementioned in Sect. 1).

These competency questions define the two main purposes of the ontology. The first purpose is the definition of a sound structure to store software patterns, especially patterns proposed in the academic literature (CQ3). As these patterns might not have been applied enough in real use cases, one objective is to quantify the acceptance by others (CQ2) of a proposed pattern in a study. One possible solution to this problem is the usage of paper citations, described in Sect. 4.1. The relations between these patterns are also an important topic, as patterns are often used together to address larger-scale problems (CQ4). The second purpose is the storage of blockchain-based software patterns, taking their specificities into account. It notably includes their classification into comprehensive categories (CQ1) to guide the reader in the space of blockchain-based software patterns, as well as the problems they address (CQ5). The process outlined in [19] was followed to validate our requirements specification, within the larger framework of the NeOn methodology. Since the ontology was to be built with extensibility in mind, should new requirements arise, the queries that correspond to the competency questions to act as a test suite that ensures the ontology remains conformant as it evolves.

3.2 Reuse of Non-ontological Resources

As the purpose of constructing the blockchain-based software pattern ontology is to formalize the knowledge of a previous SLR, the ontology incorporates knowledge from two different non-ontological resources that can both be found on GitHub[3]. The first is a collection of 160 patterns that were identified during

[3] https://github.com/harmonica-project/blockchain-patterns-collection.

the SLR within 20 different papers; out of which 114 unique patterns have been derived. Each of the collected patterns is described by a set of attributes, e.g., a *Name*, a *Context and Problem*, and a *Solution*. The citations count for each paper that proposes one or more patterns have also been collected. Thus, the reliability of a pattern can be assessed more easily: a pattern proposed in a paper cited multiple times can be considered, to some extent, to be more trustable than a pattern from a non-cited study. More rationale on the usage of these citations to assess pattern reliability is given in Subsect. 4.2. The domain, programming language, implementation examples, and blockchain technology associated with the pattern are also collected if available. Indeed, some patterns may be proposed by paper for a specific programming language (e.g. Solidity), or in the context of a specific domain (e.g., decentralized identity). Also, different types of relations between patterns were identified: *Created from*, *Variant of*, *Requires*, *Benefits from*, and *Related to*. As the application of a specific pattern might require considering other patterns, its relations to others must be made explicit. Further details about these relations are given in Subsect. 4.1. Patterns are classified into one of three categories depending on their general purpose: *Architectural patterns* that regroup patterns impacting the general structure of the application (elements, connections); *Design patterns* that are a way to organize modules, classes, or components to solve a problem; and *Idioms*, solutions to a programming language-related problems. The second non-ontological resource is used to extend this classification: design patterns are classified in subcategories derived from a taxonomy, that emerges from the categorization of the results in the SLR, and is comprised of 4 main categories and 14 subcategories. More details are given about the taxonomy in Sect. 4.

4 Results

4.1 Blockchain-Based Software Pattern Ontology

The primary outcome is the creation of the blockchain-based software pattern ontology, depicted in the conceptual model[4] Fig. 1. This ontology has been written in Turtle, using the OWL language profile. As the figure shows, the driving idea of the ontology is: (a) the explicit distinctions between patterns, variants, and pattern proposals, (b) proposals and their relations with others, and (c) design problems outside the scope of patterns. Using ontologies as a support to store gathered knowledge about patterns also enables the usage of inference engines to find new relations between entities. Also, using an ontology also allows connecting it with other ontologies. Although it is beyond the scope of this study, it is possible to connect patterns with blockchain technologies expressed in other blockchain ontologies.

The central element of this model is the *Proposal* class, that is a pattern introduced within a particular academic paper. In the current form of the pattern proposal ontology, all sources of patterns are academic papers, thus this class is

[4] For the sake of clarity, some subclasses of the ontology are hidden.

not implemented yet. Nonetheless, a class named *Source* will merge other types of sources such as technical documentation or code repositories in the future. This improves the extensibility of the model, as patterns might be found in many different sources. Each paper is linked to a *Identifier*, which can take the form of a DOI (Digital Object Identifier) or an ArXiv ID. For each paper present in the ontology, its citations have been included as identifiers individuals and linked to the paper using a *references* object relation. This system allows inferring the citations of a specific paper, then making enhanced pattern recommendations by computing a "citation score" for each pattern proposal. More rationale on pattern recommendation is given in Subsect. 4.2.

In this model, a proposal is linked to a variant. A variant inherits from a specific `Pattern` and represents one of its possible forms. Indeed, variants are used to express the variability of a pattern: two variants of a pattern might be close enough to address the same problem and solution, but may vary in some aspects (e.g., implementation). As an example, the *Oracle* pattern proposed by Xu et al. [22] is an individual of *Proposal* as it is proposed in the Xu et al. [22] paper, and attached to the *Oracle* variant, an individual of the *Variant* and *Oracle* class. The distinction between the proposals and the resulting pattern is important as in some cases multiple papers proposed the same pattern using different words, templates, and for different domains or blockchains. As such, a *Proposal* inherits from a specific blockchain, domain, or language[5].

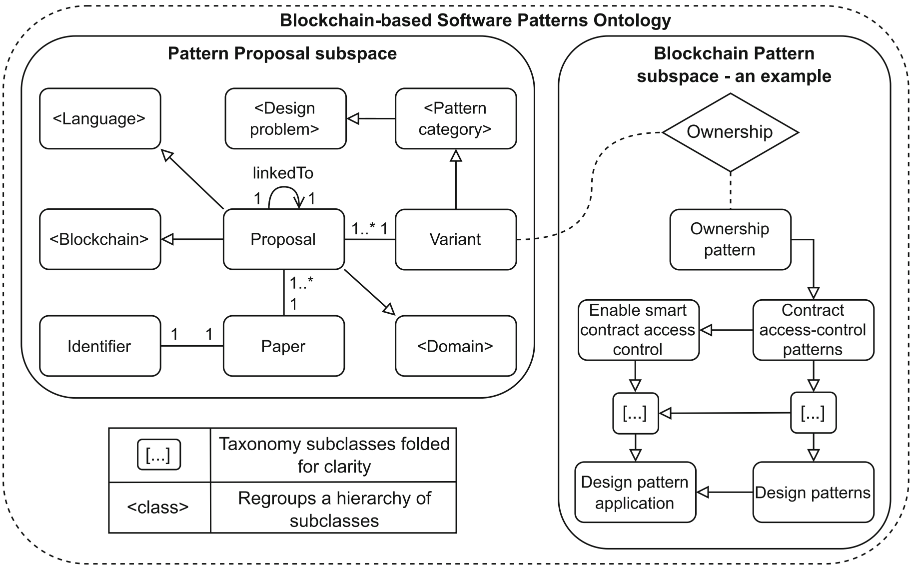


Fig. 1. Blockchain-based software pattern ontology with an exemplified section.

[5] For the sake of clarity, subclasses of blockchain system, domains, and language are hidden (e.g., resp. Ethereum, IoT, or Solidity) in the provided conceptual model.

In this conceptual model, a *Proposal* is described by a *Context and Problem*, that gives a rationale for the purpose of the pattern and addressed problems, and a *Solution* field to introduces the elements composing the pattern solution. This structure for pattern description is derived from the two main pattern formats (GoF and Alexandrian pattern formats [20]), usually employed to express software patterns. Because of the lack of standardization across the literature on the description of patterns, only the context, problem, and solution have been kept to describe a pattern in this ontology. Yet, as the ontology references each proposal paper, the reader can refer to the paper of a specific proposal to learn more about it. *Proposals* can also be linked together, using 5 different relation types that were identified from the SLR:

- *Created from* - when a pattern directly takes its sources in another.
- *Variant of* - when a pattern is a variant of another.
- *Requires* - when a pattern requires another to be applied.
- *Benefits from* - when a pattern might benefit from another when applied.
- *Related to* - to identify weak relations between patterns (e.g., "see also").

By using inference, it is possible to translate these relations from proposals to variants, creating new knowledge about possible relations between patterns. SWRL rules have been written for the inference engine to generate such relations. As an example, the following rule translate a *benefitsFrom* object relation from two proposals to their corresponding variant (Eq. 1).

$$\begin{aligned} \forall\, (p_1, p_2) \in P \; and \; (v_1, v_2) \in V, \\ p_1 \; benefitsFrom \; p_2 \;\cdot\; p_1 \; hasVariant \; v_1 \;\cdot\; p_2 \; hasVariant \; v_2 \\ \Longrightarrow \; v_1 \; benefitsFrom \; v_2 \end{aligned} \tag{1}$$

The subclasses of the *Pattern* class emanate from the reused taxonomy for blockchain-based patterns, built in its related SLR. For instance, the *Oracle* variant from [22] is linked to the *Oracle* pattern class, that inherits from the *Data exchange pattern*, then *On-chain pattern*, *Design pattern*, and finally *Pattern*. Although this hierarchy exists in the blockchain-based software pattern ontology, it is not shown in Fig. 1 for clarity. To further refine this part of the ontology, each *Pattern* addresses a specific *Design problem*. By extension, each subclass of *Pattern* addresses a design *Design problem* subclass. Each problem has been assigned an associated literal question, notably used for recommendations. These questions have been designed along the construction of the design problem taxonomy to give a literal sentence of the problem. The question is presented as an affirmation (here, a user story sentence), that can be answered by yes or no. For instance, the question associated with the *Smart contract usage* design problem, solved by the *Smart contract patterns* is "I want to execute part of my application on-chain". Such an affirmation can be thus presented as a question to the user to guide pattern recommendations.

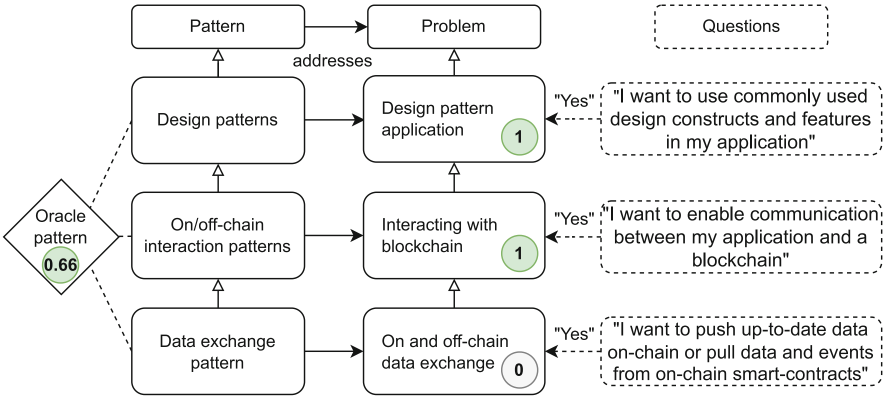


Fig. 2. Pattern scoring based on patterns/problem categories.

4.2 Ontology Querying Tool

In parallel with the ontology, a tool was designed to leverage it without having to use ontology-specific tools (e.g. Protégé). Using this tool eases the access to ontology content for non-experts, but the ontology can also be queried when served using a SPARQL server such as Apache Jena Fuseki[6]. This tool has two main features: the explorer and the recommender. The explorer allows one to dive into the knowledge of the ontology through the presentation of all available patterns in a grid. This section's purpose is to link the solution domain (the list of patterns) to the problem domain (user requirements). Indeed, any user reading available pattern descriptions might find some that suit their goals. The application shows each pattern's name but also the number of linked proposals and variants. By clicking on the pattern card, the user can consult the context, problem, and solution for each pattern variant and proposal. She also has access to a list of linked patterns following the same notation as defined in the pattern proposal ontology. The tool allows filtering patterns out, using the proposal's respective domains, blockchains, and languages. For instance, a user can select Ethereum as the desired blockchain and filter out every non-corresponding pattern. The second part of the tool is the recommender feature. Contrary to the explorer, any user can leverage the recommender to navigate from the problem domain (a set of questions asked by the user) to the solution domain (a set of patterns matching given answers). To personalize pattern recommendations, the user answers a set of questions linked to design problems, as presented in Subsect. 4.1. An illustrative scheme of this process is shown in Fig. 2.

Questions are organized in a tree structure, traversed by a conditional depth-first algorithm. The questionnaire starts with a high-level question (e.g., "I want to use design patterns in my application") Depending on the user's answers, each node of the tree is assigned a score: 1 for "Yes", -1 for "No" and 0 for "I don't

[6] https://jena.apache.org/documentation/fuseki2/.

know", children of nodes with a negative score being skipped. The tool generates the recommendation once the questionnaire is filled up. Patterns constitute the leaf node of the tree. To compute the score S_q for each pattern, the tool sums the score of every parent node and then normalizes the score using the length of the branch, accounting for branch length differences.

Next, we use three different algorithms to compute pattern rankings based on the scores S_q: `NoCitationsAndQS`, `WeightedCitationAndQS` and `UnWeightedCitationAndQS`. `NoCitationsAndQS` simply orders the patterns based on their score, while the other two also take into account an inferred number of citations. For each pattern, this number of citations is computed by summing the number of citations of all papers that proposes the pattern. As an example, if a pattern is proposed by two papers that respectively have 50 and 150 citations, the pattern is given a number of citations of 200. In `UnWeightedCitationAndQS`, we offset the rank, by multiplying S_q by the ratio of the number of citations of a pattern and the number of citations of the most-cited pattern. For `WeightedCitationAndQS`, instead of using the number of citations directly, we use its logarithm. The rationale for this is the extreme ranking skewness in favor of highly cited patterns in `UnWeightedCitationAndQS`. Note that both `UnWeightedCitationAndQS` and `WeightedCitationAndQS` might discriminate newly proposed patterns that do not have many citations yet.

The user can select one of the three algorithms. When `NoCitationsAndQS` outputs a ranking only impacted by the answers, `WeightedCitationAndQS` and `UnWeightedCitationAndQS` also take into account citations, which serves as an indicator of the pattern adoption in the literature. Also, `UnWeightedCitationAndQS` tends to recommend highly cited patterns, considered more reliable, whereas `NoCitationsAndQS` provides a ranking closer to the questionnaire's answers, to the risk of having newly proposed patterns that lack prior reuse. `WeightedCitationAndQS` is compromising between the two others, as it reduces the impact of citations without discarding them. In conclusion, the decision of using one algorithm instead of another is up to the user, depending on its goals: either using recognized patterns or newly proposed patterns that don't have this recognition yet. Nonetheless, the ranking differences between all of those algorithms are evaluated in Sect. 5.

5 Evaluation

To evaluate whether the ontology addresses the initial requirements and if the implemented tool is capable of leveraging the ontology, we conducted a threefold validation. Our first method of validation draws from both the ORSD mentioned above and the more general ontology evaluation methods outlined in the work of Raad and Cruz [12]. In particular, we follow their task-based approach, linking the evaluation of the ontology and the tool itself. We demonstrate the ability of the ontology to cover its requirements by, on the one hand, using SPARQL queries in isolation to answer the CQs, and, on the other, by showing its ability to be used as the central knowledge representation mechanism of our tool, through the validation methods to be covered below. We briefly touch on some of the main evaluation criteria mentioned by Raad and Cruz (cf. [12] for the definition of each concept): **accuracy, completeness, clarity, and conciseness**, though difficult to demonstrate in an absolute sense, are nevertheless covered by the fact that the ontology has been constructed on the basis of an extensive SLR of the field, where care has been taken to isolate only the most relevant aspects; **adaptability** is a consequence of the use of the NeOn methodology and the use of SHACL shapes for automated verification of the ontology and the inferences made thereof, rendering the addition of new patterns to the ontology straightforward; **computational efficiency** is ensured by the compactness of the ontology and the avoidance of recomputing rule-based inferences for every query through pre-compilation of the inferred ontology triples; and, finally, **consistency** is ensured through the use of the aforementioned SHACL shapes for every main class in the knowledge base. As such, SHACL shapes ensure that the current knowledge in the ontology complies with conceptualized rules and that the addition of new knowledge and subsequent OWL reasoning also does.

For the second dimension of our validation scheme, we demonstrate the relevancy of the ontology by addressing the following hypotheses:

- H_1: A practitioner can leverage the tool to navigate from the solution space (blockchain-based patterns) to the problem space (requirements).
- H_2: A practitioner can leverage the tool to navigate from the problem space (requirements) to the solution space (relevant blockchain-based patterns).

Each of these hypotheses will be treated using a specific protocol, both described in the following subsection. For H_1, a survey has been conducted with experts to assess the capability of using the explorer to understand the pattern proposals, and by extension to assess the relevancy of knowledge within our ontology. For H_2, a protocol has been designed to evaluate the recommendations produced by the recommender. Indeed, if the recommendation system is able to suggest adequate patterns, it illustrates the capability of using the ontology to find adequate patterns for specific requirements.

5.1 Protocol

H_1 - To answer this hypothesis, we surveyed a panel of 7 experts from different backgrounds and positions as shown in Table 2.

Table 2. Panel Description (§PhD student, *Lead Tech, † Software Engineer, ‡ Blockchain Engineer).

ID	E1	E2	E3	E4	E5	E6	E7
Role	*	§	†	‡	§	†	§
Blockchain experience (y)	4	4	4	4	2	1	2
Software design experience (y)	5+	1	5	5	2	2	5+

A custom case study has been designed on a blockchain use case. This case study was short enough to ensure participants had the time to assimilate it within the allocated 30" survey timeframe. Organizers proposed 5 patterns $P_{H_1}^{j}$ $(0 < j < 5)$ for each expert n, and the objective was to assess if the expert was able to find and understand the patterns well enough to decide if they were applicable to the case study. This applicability j was rated by each participant n from 0 (non applicable) to 4 (must-have) $R_n(P_{H_1}^{j})$. Then, the survey organizers performed the same exercise. As they worked on the pattern knowledge base and the ontology, they know in detail the patterns presented in the tool and their related papers $\tilde{R}(P_{H_1}^{j})$. Participants' answers were compared to the organizers' own responses and a normalized score for each participant was calculated $S_n^{H_1}$, the average absolute difference between his score and the organizer's score.

$\boldsymbol{H_2}$ - In the second validation step, we aim at evaluating the performance of the various recommender engines, especially those including citation metrics in the ontology. The initial idea was to evaluate the precision and recall of the recommender engine using a set of papers that present blockchain applications. For each paper, the goal was to identify the most suitable patterns $\hat{I}_p$ for the introduced application, then execute the recommendation engine based on the application requirements to retrieve a set of recommended patterns I_p. Using these sets, it is possible to compute the precision and the recall of the recommender system as we respectively assess the proportion of most suitable patterns found in the recommended patterns $\hat{I}_p \in I_p$ over the total amount of most suitable patterns $\hat{I}_p$ and the total number of recommended patterns I_p. However, this approach is very limited for recommender systems: as the set of recommended patterns is very large, the precision will be artificially high. Nevertheless, another approach exists to evaluate the precision and the recall with regards to the ranking, that is precision/recall at cutoff k [3]. Instead of calculating the precision and the recall for all of the recommended patterns, these metrics are computed several times for the k^{th} first patterns, k varying between 1 and the total number of recommended patterns. As a result, it is possible to draw a curve that shows the varying precision and recall depending on k.

To compute these values, the following protocol was undertaken:

- Select a paper p from the literature (n = 13), that propose a blockchain-based application, extract the requirements R_p and identify possible patterns $\hat{I}_p$, which represents the golden standard of patterns for p.
- Answer recommender's questions using only the requirements R_p, and retrieve a set of I_p recommended patterns, and their position/score $S_{p,i}, i \in I_p$.
- Compute the precision and the recall at cutoff k, resp. $\frac{\hat{I}_p \in I_p}{k}$ and $\frac{\hat{I}_p \in I_p}{\hat{I}_p}$.

5.2 Results

H_1 - Fig. 3 shows the descriptive statistics for the score for each panel participant. The mean values for all the questions range from 2.75 to 3.75 with an

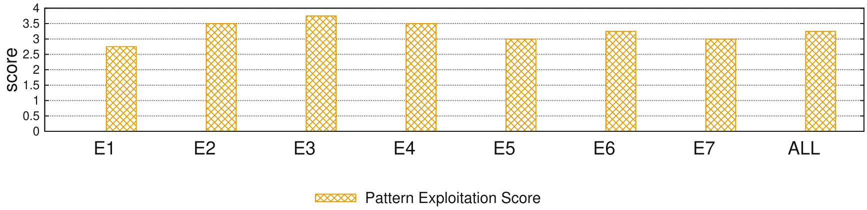


Fig. 3. Panel usecase score $S_n^{H_i}$

average of 3.25/4, which indicates that the participants have successfully navigated the solution space and provided adequate options on the relevance of the proposed pattern. It must also be noted that the most junior profiles having a score of 3, have used the tool effectively despite their lack of proficiency in blockchain application design. The expert panel results show positive mean scores for all metrics, our hypothesis can be considered valid w.r.t. our protocol, despite having room for improvements, essentially in its perceived added value. The small sample size, should also prompt further large-scale surveys, including a pre-flight questionnaire to better quantify prior blockchain background for the respondents, and question its impact on the tool's usability.

H_2 - Fig. 4 and 5 respectively show the precision and the recall at cutoff k for the three recommender systems considered. To interpret the results, it is required to select an adequate k w.r.t. the usage of the recommendation system. As the web platform displays the first 18 patterns on the first page when executing the recommendation system, a value of $k = 18$ has been chosen. Nonetheless, the selection of a suitable k is a difficult issue, discussed in Sect. 6. Regarding the precision, the three algorithms are producing similar results except for a cutoff of $k < 20$, where the inclusion of citations increases the precision. For a cutoff of $k = 18$, the precision is 0.2, meaning that on average 20% of the first 18 recommended are relevant for the considered paper. Regarding the recall, the curves are similar for the three different algorithms, with a small advantage to the *NoCitationsAndQS* algorithm. For a cutoff of $k = 18$, on average 57% of the identified relevant patterns in the papers $\hat{I}_p$ are recommended. This number goes up to 80% for a cutoff of $k = 40$. By extension, it indicates that the majority of the most suitable patterns are ranked at the top by the recommender system.

6 Threats to Validity

Internal Threat to Validity: some aspects of the method used to validate $H2$ can be mentioned. Indeed, the selection of adequate patterns for a given paper has been carried out by two researchers in 13 papers. Although these researchers are experts in blockchain technologies and decentralized applications, it still leaves some space for subjectivity. Several measures have been taken to limit the impact on the results: restricting the selection to the most important patterns,

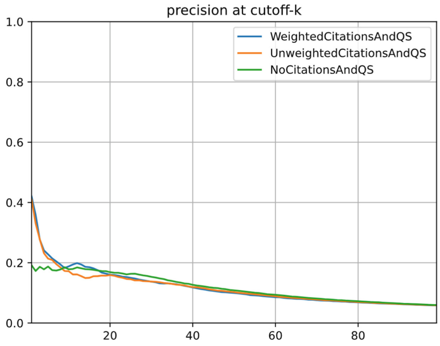


Fig. 4. Average precision at cutoff-k.

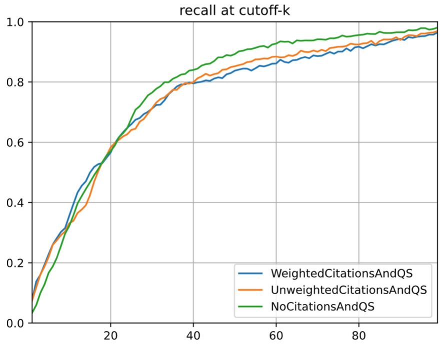


Fig. 5. Average recall at cutoff-k.

and comparing the results between the researchers to evaluate possible discrepancies. The method used to build the ontology can also be a threat to validity. Ontologies, by their very nature, have a degree of subjectivity; nevertheless, by using a structured and proven methodology (NeOn), and building upon a SLR, this risk is mitigated, in conjunction with the formal specification of the ORSD and with the evaluation methodology outlined above. Finally, the selection and retrieval of pattern proposals from the literature, that constitutes the core of the ontology, is also subject to be a threat to validity. However, it is mitigated by the strict method followed to perform the literature review (SLR). More details about possible threats and mitigations are given in another study [17].

External Threat to Validity: the main threat is the generalizability of the ontology. Even if the main purpose of the ontology was its reusability in a tool, careful attention has been made to maximize the ontology reusability. Part of the ontology is inspired by the Design Pattern Intent ontology [10], to bind design patterns (by extension, software patterns) with blockchain design problems. Patterns are also expressed using a shortened pattern format, similar to the GoF pattern format or the Alexandrian form [20]. Future works will refine those patterns to fully comply with one of those two formats. Finally, the ontology has been designed with extensibility in mind. For example, the blockchain class can easily be a connection point between this ontology and other blockchain-related ontologies, such as [4], a blockchain domain ontology.

Conclusion Threats to Validity: we can mention the difficulty to conclude on the recommendation engine relevancy w.r.t. adequate cutoffs k, as its selection mainly depends on the user behavior. Indeed, some users might only read the first 5 patterns, whereas others might fetch all of the recommendations. In our web platform, the first page of the recommender results displays 18 patterns; thus this number might be a good candidate for k. Nonetheless, this number might change depending on the usage of the recommendation engine in the future.

7 Conclusion and Future Work

This paper proposes a blockchain-based software pattern ontology to store, classify, and reason about blockchain-based patterns. The ontology has been built over previous results obtained by performing a systematic literature review (SLR) of the state-of-the-art of blockchain-based patterns. It is composed of proposals that are patterns formalized in the context of an academic paper. 160 proposals have been stored in the ontology, resulting in 114 different software patterns identified. Also, those patterns have been classified using a taxonomy reused from the SLR mentioned above. To best make use of both the categorization and the ontology, a tool has been built. Using it, practitioners can explore the ontology and its collection of patterns, but also use a recommender to get adequate patterns fulfilling their needs. This tool is also meant to be extendable following ontology evolution and support future works. The ontology can also be leveraged as standalone, using SPARQL queries.

This paper paves the way for future works in assisting practitioners in the design of a blockchain application. The different artifacts will be integrated into the Harmonica project[7], a semi-automated framework for the design and implementation of blockchain applications [15]. The integration will notably be done between the tool presented in this paper and BLADE (BLockchain Automated Decision Engine), a decision-making tool for the selection of a blockchain technology [16]. The combination will allow users to select a blockchain, then adequate patterns that are applicable to the chosen technology. Some extensions of this work could also be envisioned in the software pattern domain. Although the pattern proposal ontology is introduced within the scope of blockchain patterns, it could be generalized to all software patterns, such as Internet-of-Things (IoT) or microservices. Finally, existing software patterns in the ontology might be extended to include a formal description using existing pattern formats.

References

1. Belotti, M., Božić, N., Pujolle, G., Secci, S.: A vademecum on blockchain technologies: when, which, and how. IEEE Commun. Surv. Tutor. **21**(4), 3796–3838 (2019)
2. Casado-Vara, R., Prieto, J., De la Prieta, F., Corchado, J.M.: How blockchain improves the supply chain: case study alimentary supply chain. Procedia Comput. Sci. **134**, 393–398 (2018)
3. Croft, W.B., Metzler, D., Strohman, T.: Search Engines: Information Retrieval in Practice, vol. 520. Addison-Wesley, Reading (2010)
4. de Kruijff, J., Weigand, H.: Understanding the blockchain using enterprise ontology. In: Dubois, E., Pohl, K. (eds.) CAiSE 2017. LNCS, vol. 10253, pp. 29–43. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59536-8_3
5. Gamma, E., Helm, R., Johnson, R., Vlissides, J., Patterns, D.: Elements of Reusable Object-Oriented Software, vol. 99. Addison-Wesley, Reading (1995)
6. Girardi, R., Lindoso, A.N.: An ontology-based knowledge base for the representation and reuse of software patterns. ACM SIGSOFT Softw. Eng. Notes **31**(1), 1–6 (2006)

[7] https://github.com/harmonica-project.

7. Glaser, F.: Pervasive decentralisation of digital infrastructures: a framework for blockchain enabled system and use case analysis (2017)
8. Hector, U.R., Boris, C.L.: Blondie: blockchain ontology with dynamic extensibility. arXiv preprint arXiv:2008.09518 (2020)
9. Henninger, S., Ashokkumar, P.: An ontology-based metamodel for software patterns. CSE Technical reports, p. 55 (2006)
10. Kampffmeyer, H., Zschaler, S.: Finding the pattern you need: the design pattern intent ontology. In: Engels, G., Opdyke, B., Schmidt, D.C., Weil, F. (eds.) MODELS 2007. LNCS, vol. 4735, pp. 211–225. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75209-7_15
11. Pavlic, L., Hericko, M., Podgorelec, V.: Improving design pattern adoption with ontology-based design pattern repository. In: ITI 2008–30th International Conference on Information Technology Interfaces, pp. 649–654. IEEE (2008)
12. Raad, J., Cruz, C.: A survey on ontology evaluation methods. In: Proceedings of the International Conference on Knowledge Engineering and Ontology Development, part of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (2015)
13. Schär, F.: Decentralized finance: On blockchain-and smart contract-based financial markets. FRB of St. Louis Review (2021)
14. Seebacher, S., Maleshkova, M.: A model-driven approach for the description of blockchain business networks. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018)
15. Six, N.: Decision process for blockchain architectures based on requirements. CAiSE (Doctoral Consortium), pp. 53–61 (2021)
16. Six, N., Herbaut, N., Salinesi, C.: Blade: Un outil d'aide à la décision automatique pour guider le choix de technologie blockchain. Revue ouverte d'ingénierie des systèmes d'information **2**(1) (2021)
17. Six, N., Herbaut, N., Salinesi, C.: Blockchain software patterns for the design of decentralized applications: a systematic literature review. In: Blockchain: Research and Applications, p. 100061 (2022)
18. Six, N., Ribalta, C.N., Herbaut, N., Salinesi, C.: A blockchain-based pattern for confidential and pseudo-anonymous contract enforcement. In: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 1965–1970. IEEE (2020)
19. Suárez-Figueroa, M.C., Gómez-Pérez, A., Fernández-López, M.: The NeOn methodology for ontology engineering. In: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. (eds.) Ontology Engineering in a Networked World, pp. 9–34. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24794-1_2
20. Tešanovic, A.: What is a pattern. Dr. ing. course DT8100 (prev. 78901/45942/DIF8901) Object-oriented Systems (2005)
21. Xu, X., Bandara, H.D., Lu, Q., Weber, I., Bass, L., Zhu, L.: A decision model for choosing patterns in blockchain-based applications. In: 2021 IEEE 18th International Conference on Software Architecture (ICSA), pp. 47–57. IEEE (2021)
22. Xu, X., Pautasso, C., Zhu, L., Lu, Q., Weber, I.: A pattern collection for blockchain-based applications. In: Proceedings of the 23rd European Conference on Pattern Languages of Programs, pp. 1–20 (2018)
23. Xu, X., et al.: A taxonomy of blockchain-based systems for architecture design. In: 2017 IEEE International Conference on Software Architecture (ICSA), pp. 243–252. IEEE (2017)

Check for updates

Learning-Aided Adaptation - A Case Study from Wellness Ecosystem

Suman Roychoudhury(✉), Mayur Selukar, Deepali Kholkar, Suraj, Namrata Choudhary, Vinay Kulkarni, and Sreedhar Reddy

Tata Consultancy Services Research, Pune, India
{suman.roychoudhury,mayur.selukar1,deepali.kholkar,suraj.39, namrata.choudhary,vinay.vkulkarni,sreedhar.reddy}@tcs.com

Abstract. Modern businesses are being subjected to an unprecedented variety of change drivers that cannot be predicted such as new regulations, emerging business models, and changing needs of stakeholders. This creates new demands on enterprises to meet stated goals in a dynamic and uncertain environment that translate to demands on the enterprise's software systems. In our earlier work, we proposed an architecture for dynamic adaptation under uncertainty. In this paper, we provide validation of our adaptation architecture through a case study of a wellness ecosystem comprising multiple stakeholders with evolving goals and behavior that need to be continuously learnt and adapted. We present experimentation results from our case study that serve as an initial step towards validating our approach.

Keywords: Software adaptation · Digital twin · Reinforcement learning

1 Introduction

Enterprises today operate in a dynamic, uncertain environment where they are continually subjected to unpredictable changes to business, consumer needs, laws and regulations, and even technology. Enterprises need to be able to adapt to these changes and continue to meet their goals despite this dynamism and uncertainty. In current practice analysis of each change in the business environment and making decisions on the enterprise's response to change is done by human experts. On the other hand, modern businesses are driven by software systems that are now ubiquitous and serve as critical growth driver for enterprises. As businesses negotiate the uncertainty in their environment, systems must assist enterprises (i.e., human experts) in adapting to change and meeting their goals. The adaptation process needs to be dynamic and continual, i.e. systems must be capable of detecting, analyzing, and responding to changes in the environment at run-time.

To achieve dynamic adaptation, enterprise systems need to continually acquire knowledge about their environment, evaluate goals, suggest appropriate

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 300–315, 2023.
https://doi.org/10.1007/978-3-031-26886-1_18

courses of action towards meeting the goals, and learn from previous experience. However enterprise systems often have only partial visibility into their environment, also, available information is incomplete and uncertain. For example all usage scenarios of the system cannot be known beforehand. Behaviours and preferences of individual system users can span across a wide spectrum, which are unknown to the system when it begins operation. Moreover behaviour patterns and needs of system actors evolve with time, with life changes and unpredictable events.

In order to understand the adaptation triggers for modern business systems, enterprises need to be viewed in the context of the larger business ecosystem of which they are a part. Business ecosystems comprise multiple interacting stakeholders - business enterprises, suppliers, service, sales, and other partner organizations, consumers, regulators, and also products and processes. Participants in a business ecosystem come together to service the same set of consumers [15]. Working collaboratively enables them to deliver greater value to the customer than they could have individually while increasing their own market access and brand value. Collaboration between stakeholders needs to be managed by a central integrator [16] that orchestrates interactions in the ecosystem. Interactions among stakeholders leads to increased dynamism in the environment of individual systems necessitating their adaptation.

We look at the adaptation problem for businesses in the context of an ecosystem where the central integrator onboards consumers as well as service providers, acquiring the right information and orchestrating services from both so that goals of each stakeholder are met despite dynamism in the environment. The integrator's knowledge at the ecosystem level helps in making orchestration decisions and recommending courses of action to individual stakeholders.

In our earlier work [7] we proposed a learning-aided adaptive architecture for enterprise systems. In this paper, we provide validation of the adaptation learning component of the architecture through a real-life case study from the wellness domain. The key contribution of the paper is demonstrating the utility of adaptation learning to aid stakeholder decision making in the context of business ecosystems. Our approach uses digital-twin based simulation techniques to train a reinforcement learning (RL) agent that has no prior knowledge about stakeholder behaviour. The RL agent acts as the primary learning component and via interactions with various system stakeholders, learns their behavior over time and adapts the system accordingly.

The paper is organised as follows - in Sect. 2, we recap our learning-aided adaptive architecture. In Sect. 3, we detail our wellness ecosystem use case and provide experimental results to validate adaptation learning in enterprise systems. Section 4 summarises our work and outlines future directions.

2 Learning-Aided Adaptive Architecture

In this section, we review our learning-aided adaptive architecture [7] that has four key components as shown in Fig. 1.

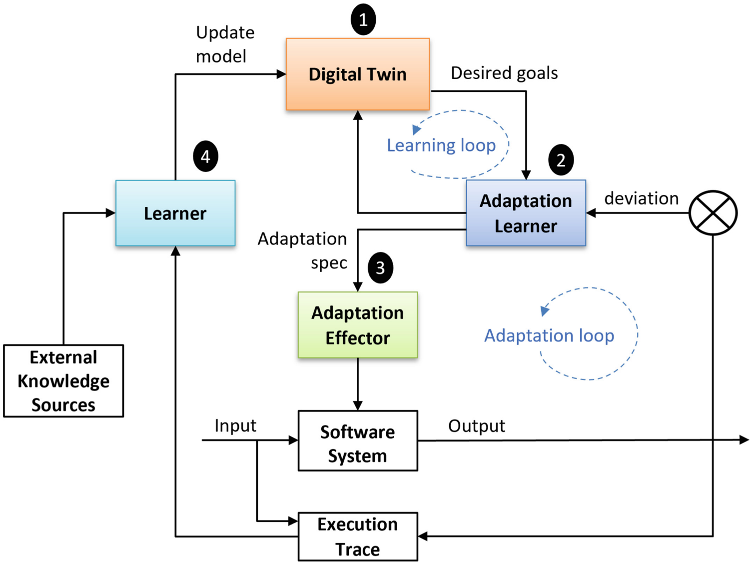


Fig. 1. Learning-aided adaptive architecture [7]

1. A digital twin based decision-making system to model the enterprise system, its environment and goals
2. An in-built reinforcement learning-based adaptation learner system to control or adapt the enterprise model (i.e. the digital twin)
3. A sensor-actuator based adaptation effectation architecture to keep the enterprise model and system in sync
4. A continuous learning module to observe system interaction traces and update the digital twin based on new or updated knowledge

In this section, we review the first two components of the architecture, namely the enterprise model of the system (i.e its digital twin) and the adaptation learning component that are key to realizing the adaptive architecture.

2.1 Digital Twin - System, Environment and Goal Model

At the heart of our adaptation architecture is a System, Environment and Goal (SEG) model manifested as a Digital Twin that is a hi-fidelity machine-processable representation of the software that supports what-if scenario playing [8]. The digital twin not only represents our current understanding about the problem space, business domain and solution but is also continuously updated via behavioral learning and refinement throughout the software life cycle, as well as learning from external knowledge sources.

The SEG metamodel is depicted in Fig. 2. The system interacts with the environment via actions and has specific goals that are predicates over its value space. Actions update the value space and can be composed of other primitive

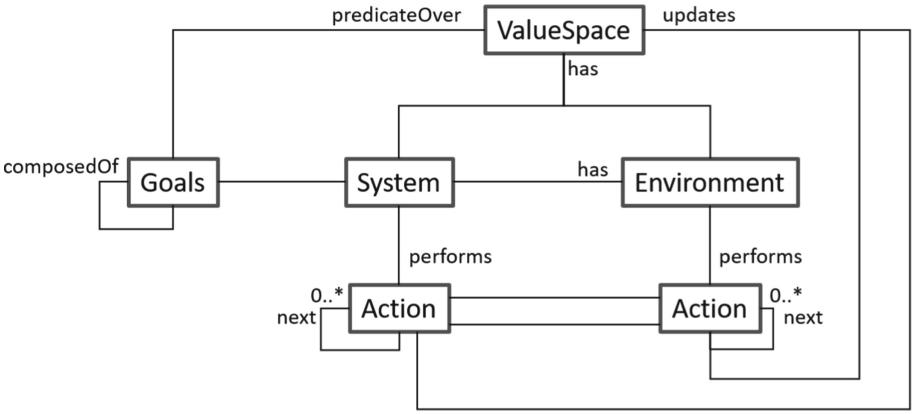


Fig. 2. Digital Twin - System, Environment and Goal Metamodel

actions. An adaptation in the system can be triggered by either a change of actions in the environment or change in goals of the system.

2.2 Adaptation Learner

The Adaptation learner component in Fig. 1 is responsible for

– (i) checking if the system is meeting stated goals when there are changes in the environment,
– (ii) identifying a strategy comprising a sequence of interventions to be introduced so as to meet stated goals, and
– (iii) getting assurance of efficacy of the strategy.

We propose a simulation-based (digital twin) reinforcement learning approach to achieve the three objectives, where the Digital Twin serves as an "experience generator" to train the reinforcement learning (RL) agent. The stated goals help define the reward function of RL agent and the candidate set of primitive actions constitute the action space. The RL agent comes up with a "good enough" policy i.e., a sequence of interventions to be introduced in order to achieve the stated goals in the light of partial knowledge, uncertainty and dynamic changes in the environment of business systems. The next section introduces the case study example from the wellness domain that is used to validate our adaptation learning approach.

3 Case Study - A Wellness Ecosystem

We illustrate adaptation learning using an example of a wellness ecosystem. Wellness needs of individual consumers encompass fitness, healthcare, nutrition, leisure, as well as health insurance. The wellness ecosystem comprises service

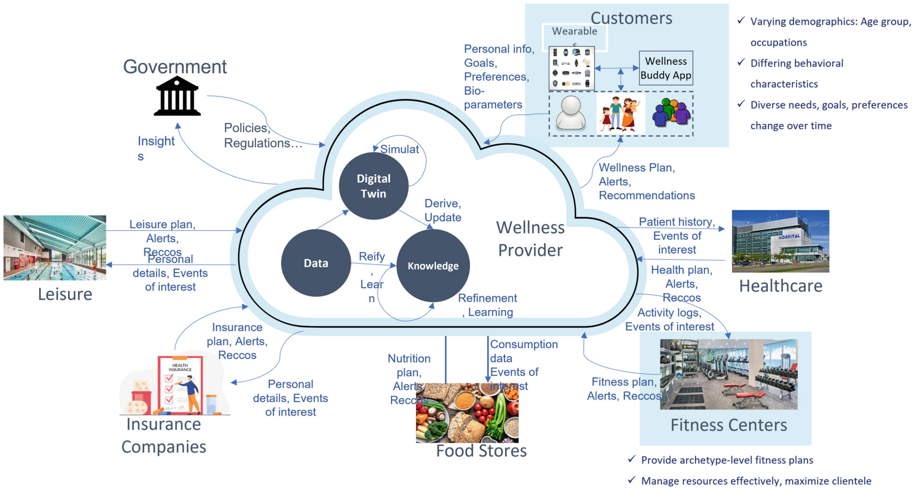


Fig. 3. Wellness ecosystem

providers for each of these aspects with the common objective of fulfilling wellness goals of a consumer population of individuals as depicted in Fig. 3. A central integrator in the form of a wellness provider facilitates and manages services offered by various service providers and acts as a centralized hub so that the consumer no longer needs to separately manage each of his wellness needs. The wellness provider creates a digital twin modeling individual service provider services and customer goals. Incomplete, uncertain, and partially visible information about the environment such as customer preferences and behaviour needs to be learnt in order to be able to meet customer goals. The goal of the wellness provider is to primarily satisfy needs of consumers while meeting goals of each of the stakeholders in the ecosystem. Although the wellness ecosystem of Fig. 3 consists of various stakeholders like “food stores”, “insurance companies”, “fitness centers” etc, for our initial experimentation, we only consider the fitness component of the wellness ecosystem, as highlighted in Fig. 3. This is in tune with current trends with health management systems and smart health devices like mHealth apps that are becoming increasing popular with fitness aspirants [10,11]. For our fitness case study, the stakeholders of interest are individual customers (i.e., fitness aspirants) and fitness centers or gyms that provide fitness facilities to the customers.

Once fitness aspirants register with a fitness center, the individuals are classified into archetypes based on age, gender and occupation. Fitness aspirants within the same archetype are initially given the same fitness plan. In this case study, we consider two such archetypes namely young professionals (Arch1) and senior management (Arch2). The Wellness provider acts as a facilitator between the fitness aspirant and fitness centers and aims to personalize the generic archetype-level fitness plan for every individual by monitoring their behavior in the fitness center.

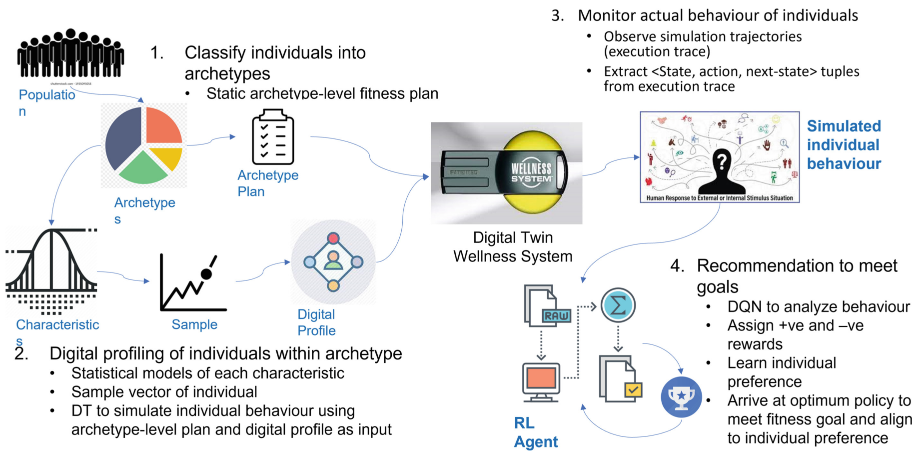


Fig. 4. Experimental setup

For our experimental setup, in the absence of real individual data, we simulate the behavior of individuals in a fitness center by constructing a digital twin of the gym environment as depicted in Fig. 4. To replicate the behavioral characteristics and preferences of individuals in a fitness center we construct a 'digital profile' of fitness aspirants. We assume the 'fitness goal' of each fitness aspirant is dependent on certain variables (e.g., their age, weight, occupation, health condition, food habit, activity level) where each variable is governed by some probability distribution over the entire population. In addition, some variables have causal relationships with other variables (e.g., age, work schedule may determine activity level), hence we model the digital profile of an individual as a probabilistic model as shown in Fig. 5, where each child node conditionally depends on parent nodes. Variables 'goal', 'workout type', and 'workout preferences' are deterministic in nature while all other variables are either discrete or continuous random variables.

Sample digital profiles (only limited variables shown) of two individuals are illustrated in Fig. 6. To generate the given sample, we first sample 'age' from an exponential distribution and then categorize them as young (if age less than 30), mid-level (if age between 30 and 45), and senior (if age greater than 45). Similarly, discrete variable 'occupation' is sampled from a multinomial distribution with values ranging from 'professional', 'self-employed' or 'government'. Since both the above variables (i.e., 'age' and 'occupation') can influence the 'activity level'

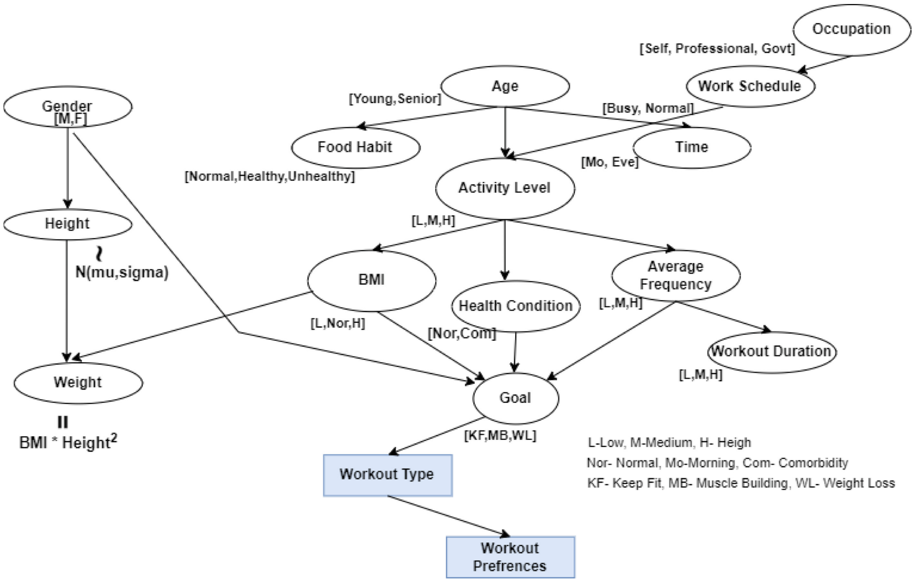


Fig. 5. Digital profile

of an individual, hence 'activity level' is sampled from a conditional multinomial distribution with condition on parent variables 'age' and 'occupation'.

Name	Age	Height (mt)	Weight	Activity Level	Occupation	Frequency per week	Health Condition	Fitness Goal	Workout Plan
Ind_1	29	1.65	63	High	Young Professional	5	Normal	Keep Fit	Default_Plan1
Ind_2	54	1.7	75	Low	Senior Management	3	Comorbid	Keep Fit	Default_Plan2

Fig. 6. Statistical sampling of an individual from digital profile

In this manner, once an individual is sampled from a given population, every individual belonging to an archetype is assigned a generic fitness plan. The default fitness plan for senior management varies from that of young professional due to the difference in exercise intensity and calorie burn requirements. The digital twin encodes variations in fitness behaviour of an individual based upon characteristics recorded in the profile.

Actor-based simulation is used to capture the ecosystem dynamics and uses the digital profile to simulate an individuals behavior in a fitness center. The environment consists of three actors: Individual, Wellness Provider, and Fitness Provider as shown in Fig. 7. Communication between actors is done by sending and receiving messages. Once a message is received, appropriate action are performed at the receivers end. Individual behavior is generated using the digital profile and suggestions are generated by the RL agent. Figure 7 shows the model of the full system where the Environment made up by our Digital Twin (DT) generates state S(t) and passes it to the reinforcement learning (RL) agent, the RL agent then takes an appropriate action A(t) and returns it to the DT. The DT consumes this action and generates the next state S(t+1) and a reward R(t). Over time the agent learns to choose the action that maximizes reward.

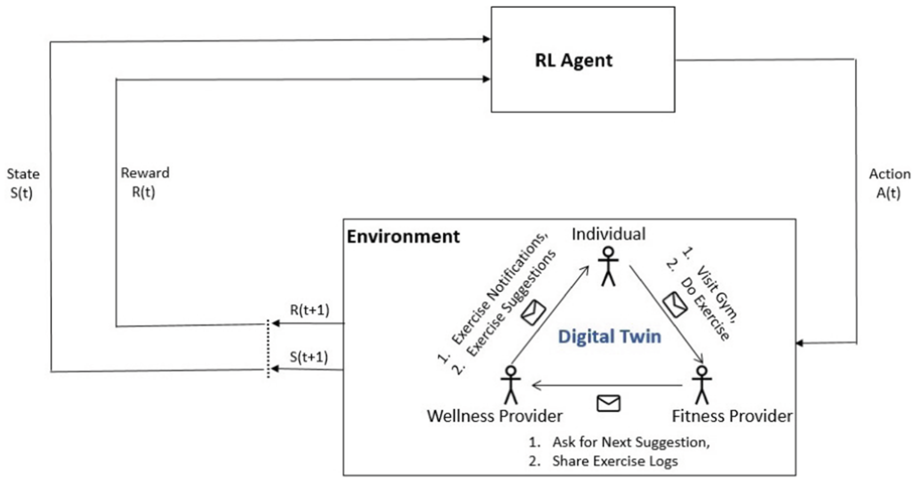


Fig. 7. RL Agent and Digital Twin Interaction

An RL Agent learns about individual preferences by monitoring the actual execution trace of an individual in the simulated gym environment against a given workout. In this way the generic plan is continuously adapted to resemble not only individual preferences but also ensuring fitness goal (in terms of calories burnt) of an individual is met.

We summarize the execution steps of our learning-aided adaptation approach as follows:

1. An individual (fitness aspirant) registers with the wellness provider.
2. The individual is allocated a generic fitness plan by the wellness platform based on their archetype (i.e, young professional or senior management)
3. The digital twin of the fitness center simulates the individual performing exercises prescribed in the plan in accordance with the characteristics of their digital profile, providing feedback in terms of:
   - a. The distance between their exercise preferences and the allocated plan.
   - b. The distance between the level of fitness (calorific burn) and the desired fitness achieved by the plan.
4. Based on the feedback, the RL agent learns the behavior of the individual and then nudges the individual to follow the plan balancing their satisfaction level and also achieving the desired fitness goal (i.e., calorie burn)

5. Eventually the 'best' fitness plan for every individual in a given archetype is adapted from the generic fitness plan and personalized accordingly.

In the next subsection, we provide further technical details of the reinforcement learning solution.

3.1 RL Problem Conceptualization

The goal of the RL agent is to learn preferences of an individual and tailor its recommendations so that they not only help achieve the target calories burnt but are also preferred by the individual. In order to do this the system starts by recommending a generic fitness plan to a new client and observes their behavior. In this case we use a Digital Twin of the Wellness ecosystem as described in the Fig. 7 to generate and record this behavior [8]. By observing the individuals behavior, we are able to encapsulate their preferences, which the RL agent can then use to adapt and create a personalized fitness plan. The generic plan also lends us the basic structure in which the recommendations are to be made. For example, if the generic plan consists of a 3:1:1 split between *cardio*, *upper body* and *lower body* workouts every week, then the consequent recommendation will also follow the same split. For a new individual there is no prior information available, therefore, in order to gather relevant information every individual goes through an exploratory phase at the beginning. During this exploratory phase a generic plan is recommended, this plan equally recommends each possible exercise. Based on the individual's behavior over this period a state is developed which in turn is used to recommend personalized plans. Note that this recommendation problem can be viewed as an Markov-Decision Process (MDP) [18] when

1. The RL agent is called for a recommendation after every exercise.
2. The state encapsulates the preferences of the individual through the activity trace and includes all the exercises performed in the current episode.

This formulation essentially encodes all the information of the individual in the state and thus follows the MDP dynamics.

3.2 Solution Methodology

We now proceed to define the reinforcement learning solution for the problem described. To deal with the diversity amongst the individuals, we map the individual to an appropriate archetype. We train a separate RL agent for each archetype as the goal changes amongst archetypes and even for same goal since the recommendation structure changes based on archetype, the *reward ranges* tend to vary making use of the same agent infeasible. Our approach breaks the larger problem into sub-problems, where the same algorithm can now be applied to each person in a diverse population based on their individual goals and attributes. This is a realistic situation in the wellness domain as people come in with varying expectations and timelines. This parallelization also keeps the agent simpler and speeds up the learning process.

States and Actions. Based on the archetype we get a certain set of exercises that could be performed to achieve the goal. These form the action space for the archetype and based on the action space we get the corresponding state space.

Let's say a person can perform 3 exercises, A_1, A_2, A_3 respectively. Now our state must have enough information for the agent to estimate an individual preference for A_i. To capture this, we create a queue D_i of finite size (here 3) for A_i and initialize it with A_i. After this, whenever A_i is recommended the corresponding action performed by the individual is appended to D_i, and a concatenation of these queues forms the first part of the state. The second part is made of individual specific constants like body mass index (BMI) along with the exercises performed in the current episode. Therefore, for the example of 3 exercises the state size is $3 * 3 + 3 + 1 = 13$ units We work with a much richer action space of 21 exercises so our state space is $21 * 3 + 21 + 1 = 85$ units.

Rewards. The rewards are designed per archetype and in most cases a goal can be expressed using the metabolic equivalent (MET) values of the exercises. For example, for young professionals whose goal is to keep fit, the objective is to maximize calories burnt while sticking to a 3:1:1 split for a given week. To allow the agent to model individual preferences, we associate a penalty when the suggestion is not performed, this allows the agent to achieve the goal of recommending the exercise that the person likes without being biased towards recommending exercises that would burn maximum calories. However, this may lead to a situation where one exercise will be recommended always, so we add a diminishing factor to the preferences so that the reward for recommending the same exercise again is lower, this is in line with how the preferences of individuals are modeled. The reward function is modeled as below-

$$R = \alpha Penalty + \bar{\alpha}\beta^n Preference(e) * Met(e)$$

where $\alpha = 0$ if the recommendation was performed; β = preference diminishing factor; n = number of times the exercise is performed during current day; e = exercise recommended and performed.

Constraints and Action Mask. The situation described so far ignores the various constraints that the wellness provider faces from its stakeholders. The wellness provider must deal with a range of constraints, from the limited gym equipment that remain constant across archetypes, to the personalized health related constraints that are unique to every individual. To deal with constraints, all action masks are collected from the stakeholders involved and we use a simple bitwise *and* operation to generate the final action mask [5]. The action mask is then used to mask out the invalid action by setting its predicted Q-Value to -∞. This relieves our RL agent from the burden of constraint handling as they are now handled externally. It also keeps the system modular, and any addition or deletion of constraints and stakeholders is streamlined.

Neural Network Architecture and Training. We use Deep Q-Learning (DQN) [13] for the computation of individual recommendations given the current state of the individual. The policy network of Fig. 8 is fully connected with layers of (85, 512, 512, 512, 21) neurons respectively with the first number indicating the input size and the last indicating the number of exercises that can be recommended. All the hidden neurons have *tanh* activation, and the output layer has linear activation.

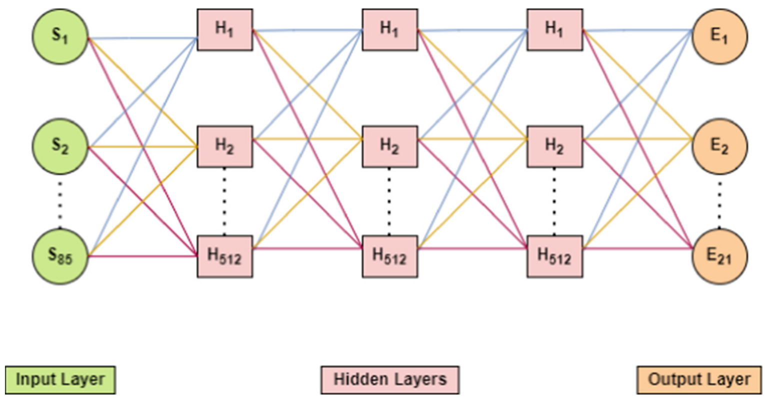


Fig. 8. Neural network architecture

The agent follows a Monte Carlo learning approach [18], which implies that we only have one policy network. The specifics of the learning approach followed are as described in [4]. The policy takes the 85 feature vectors described above and recommends appropriate action. The network is trained using the Pytorch library on Python 3.7 using Adam optimizer with a learning rate of 0.0005 and a batch size of 64, mean squared loss was used with a discount factor of 0.99. Finally, the training calls are only made after the memory buffer contains 1000 training episodes. We use Boltzmann action selection [18] for exploration which implies that our agent is stochastic in nature.

Since the actual problem structure is that of an infinite horizon MDP [18], we terminate the episode at the end of every day. This is done because most of our constraints are either constant, daily or weekly in nature and keeping the length of episode as *daily* makes constraint management simpler. Each training episode begins after an initial exploratory period, which guarantees that the state has enough information regarding the person's preferences. The environment used simulates more than 200 unique individuals to avoid overfitting and to learn the mapping between the state and appropriate actions.

Training Results. We train the agents for 20000 episodes and the episodes start at the beginning of a week. Based on the individual's archetype the recommendations are made following the structure of the generic plan. The episode length is set to a day and Fig. 9 depicts the loss graph for the two archetypes under consideration.

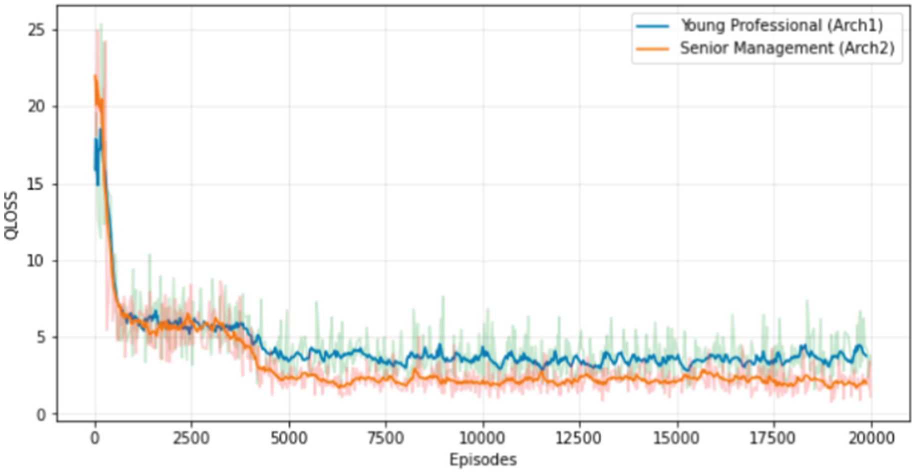


Fig. 9. Loss graph during RL training

3.3 Comparison with Baseline Algorithm

For comparison of our algorithm with a baseline, we propose a greedy upper bound solution as the baseline that can be obtained if we assume we have complete knowledge about individual's preferences. The digital twin provides us with the preferences of an individual at any point of time during the episode. Combining this with the MET values of the exercise, we can run an exhaustive search over an objective function and then take the action that maximizes the return. We define the objective function over reward as follows

$$O = P(e) * \beta^n Preference(e) * Met(e)$$

where P(e) is the probability that exercise e will be performed after it is recommended. This is obtained from the execution trace of the digital twin and β is the diminishing factor defined earlier and preference and MET are the preference of the individual for the exercise and MET value of the exercise.

The objective is nothing but the expected return of recommending exercise *e*. The optimal action thus is the one which maximizes expected return. We present a detailed exploration of the actions taken by both approaches and their effect on the individuals behavior in the subsequent section.

Comparison for a Sample Profile. Figure 10 depicts the recommendations made and their effect on the exercises performed by the same individual. The left column depicts the recommendations made by the greedy Upper bound algorithm and the right column depicts the recommendations from RL over a 120-day period excluding the initial exploratory period of 30 days.

Fig. 10. Results - nudging individual to meet fitness goal by learning their preference (Color figure online)

The blue line depicts the probability of an exercise being performed without any external influence of recommendations and constraints whereas the green line depicts the probability with which the exercise was performed over the 120-day period. The underlying bar graph denotes the number of times the exercise was recommended. Here the blue prior (individuals affinity to an exercise) is not visible to the RL agent and must be inferred from the person's behavior. We also note that the person's choices of exercise to perform are stochastic in nature and the objective function and reward are encoded in such a manner that when an optimal action is recommended and not performed, the same action must be recommended again. Our RL solution is able to map a similar policy where in each case of *Cardio*, *Upper Body* and *Lower Body*, the most prominent recommendation is the same as that of the upper bound. The second and third recommendation do not match as strongly as the difference in upside is minimal. This can be explained as the RL approach is stochastic in nature as compared to the greedy nature of the upper bound solution. In case of *Upper Body* and especially *Lower Body*, the RL recommendations are significantly different than the upper bound. This is because the MET values of most upper body and lower body exercises are approximately equal, which leads to minimal delta between the upsides resulting in different recommendations.

The effect of the recommendations in nudging the user towards their goal is visible in both cases where the probability of the most prominent recommendations is drastically increased. This increase in the probability depicts the nudging behavior of the agent. Since all the probabilities must sum up to 1, the increase is only visible for the most prominent recommendation and in some cases a decrease in the probability of the second and third prominent recommendation is also seen. This is a probable outcome and in line with the expectations. Hence the results presented in this section demonstrate the effect of nudging an individual towards meeting their goal of calorie burn in addition to balancing their preferences for certain exercise routines.

4 Discussion and Related Work

Software adaptation has been an important area of research for some time [2]. Several adaptation architectures have been proposed in the literature [6,9,17] that principally comprise a sensing or monitoring step, followed by analysis and implementation of adaptation. Most approaches use statically built models to guide adaptation, while our approach uses a model built dynamically using deep reinforcement learning.

The adaptation learning scenario presented in this paper is similar to mHealth apps and mobile health management systems [11]. In these systems fitness aspirants log data using a multitude of models from fitness trackers, smart wearable devices to manual entry. These systems then use the data to encourage physical activity and reduce the risk of diseases related to physical inactivity. The intervention mechanism ranges from context aware motivational messages throughout the day [10,12] to recommendation systems, delivering optimal set of activities [3].

In this paper, we have used RL agent as our adaptation-learning mechanism to constantly monitor user preferences and adjust the recommendations based on their dynamic behavior throughout the intervention period. In comparison, previous work in this space [3,10] uses RL to encourage physical activity of an individual, where they use a generic model of an individual instead of learning the model from individuals' dynamic behavior.

The fitness-fatigue model proposed by [1] has been used to predict athletes' performance in a variety of endurance sports. According to the fitness-fatigue model, each practice session has a positive impact on performance by contributing to a "stock of fitness" and a negative impact by contributing to a "stock of fatigue." Fatigue can be measured in terms of calories burned which is proportional to"metabolic equivalents" (MET) of the exercise that we use in our experiment. Therefore, getting maximum fatigue or calorie burn during the exercise can bring maximum fitness for the individual provided individual does not fatigue himself beyond a limit, which we achieve by introducing constraints on calorie burn.

5 Conclusion

In this paper, we demonstrated how adaptation learning using RL can be achieved in a business ecosystem context through the example of a wellness ecosystem comprising multiple stakeholders without prior knowledge of stakeholder behavior. The wellness provider acts as central integrator and learns stakeholder behavior by suitable interactions via a digital twin such that stakeholder goals can be achieved without compromising stakeholder preference. Therefore, balancing stakeholder satisfaction along with meeting stakeholder goal is the key contribution of this paper, thus illustrating learning-aided dynamic adaptation. Use of a digital twin to model and simulate user behaviour helped us to circumvent the problem of unavailability of real users' data that classical learning approaches need for providing recommendations. Although in this paper, we considered goals of stakeholders only from the fitness perspective, in future, we intend to include other wellness perspectives like nutrition, healthcare, recreation, insurance etc. that may require resolution of multiple stakeholder goal conflicts using Pareto optimal techniques [14].

Furthermore, the scope of this paper is limited to validating the Adaptation Learner component of Fig. 1 using a business ecosystem, whereas we would like to extend the validation to other parts of the architecture [7], where adaptation policies learnt on simulated system behaviour are appropriately implemented in underlying software systems and execution traces of real users' behavior are used to update new knowledge resulting in further improvement of the digital twin model.

References

1. Banister, E.W., Calvert, T.W., Savage, M.V., Bach, T.: A systems model of training for athletic performance. Aust. J. Sports Med. **7**(3), 57–61 (1975)
2. Cheng, B.H.C. et al.: Software engineering for self-adaptive systems: a research roadmap. In: Cheng, B.H.C., de Lemos, R., Giese, H., Inverardi, P., Magee, J. (eds.) Software Engineering for Self-Adaptive Systems. LNCS, vol. 5525, pp. 1–26. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02161-9_1
3. Fang, J., Lee, V., Wang, H.: Dynamic physical activity recommendation on personalised mobile health information service: a deep reinforcement learning approach. arXiv preprint arXiv:2204.00961 (2022)
4. Hausknecht, M., Stone, P.: On-policy vs. off-policy updates for deep reinforcement learning. In: Deep Reinforcement Learning: Frontiers and Challenges, IJCAI 2016 Workshop. AAAI Press, New York (2016)
5. Huang, S., Ontañón, S.: A closer look at invalid action masking in policy gradient algorithms. arXiv preprint arXiv:2006.14171 (2020)

6. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. Computer **36**(1), 41–50 (2003). https://doi.org/10.1109/MC.2003.1160055
7. Kholkar, D., Roychoudhury, S., Kulkarni, V., Reddy, S.: Learning to adapt - software engineering for uncertainty. In: 15th Innovations in Software Engineering Conference. ISEC 2022, Association for Computing Machinery, New York (2022). https://doi.org/10.1145/3511430.3511449
8. Kulkarni, V., Barat, S., Clark, T.: Towards adaptive enterprises using digital twins. In: 2019 Winter Simulation Conference (WSC), pp. 60–74 (2019). https://doi.org/10.1109/WSC40007.2019.9004956
9. de Lemos, R., et al.: Software engineering for self-adaptive systems: a second research roadmap. In: de Lemos, R., Giese, H., Müller, H.A., Shaw, M. (eds.) Software Engineering for Self-Adaptive Systems II. LNCS, vol. 7475, pp. 1–32. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-35813-5_1
10. Liao, P., Greenewald, K., Klasnja, P., Murphy, S.: Personalized heartsteps: a reinforcement learning algorithm for optimizing physical activity. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–22 (2020)
11. Liu, Y., Avello, M.: Status of the research in fitness apps: a bibliometric analysis. Telematics Inf. **57**, 101506 (2021)
12. Marcolino, M.S., Oliveira, J.A.Q., D'Agostino, M., Ribeiro, A.L., Alkmim, M.B.M., Novillo-Ortiz, D.: The impact of mhealth interventions: systematic review of systematic reviews. JMIR mHealth and uHealth **6**(1), e8873 (2018)
13. Mnih, V., et al.: Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602 (2013)
14. Moffaert, K.V., Nowé, A.: Multi-objective reinforcement learning using sets of pareto dominating policies. J. Mach. Learn. Res. **15**(107), 3663–3692 (2014). http://jmlr.org/papers/v15/vanmoffaert14a.html
15. Moore, J.: Predators and prey: a new ecology of competition. Harvard Bus. Rev. **71**, 75–86 (1999)
16. Sarafin, G.: What business ecosystem means and why it matters (2021)
17. Shevtsov, S., Berekmeri, M., Weyns, D., Maggio, M.: Control-theoretical software adaptation: a systematic literature review. IEEE Trans. Softw. Eng. **44**(8), 784–810 (2018). https://doi.org/10.1109/TSE.2017.2704579
18. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998). https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf

Demonstrations Track

Demonstrations Track

Organization

Track Chairs

Massimiliano de Leoni	University of Padua, Italy
Ivan Donadello	Free University of Bozen-Bolzano, Italy
Cristine Griffo	Free University of Bozen-Bolzano, Italy

Program Committee

Cristina Cabanillas	University of Seville, Spain
Florian Matthes	Technical University of Munich, Germany
Francesca Zerbato	University of St. Gallen, Switzerland
Francesco Leotta	Sapienza University of Rome, Italy
Jens Gulden	Utrecht University, The Netherlands
Luis Ferreira Pires	University of Twente, The Netherlands
Marco Comuzzi	Ulsan National Institute of Science and Technology, South Korea
Marco Pegoraro	RWTH Aachen University, Germany
Maria d. G. S. Teixeira	Federal University of Espírito Santo, Brazil
Michael Sheng	Macquarie University, Australia
Onur Dogan	Izmir Bakırçay University, Turkey
Riccardo Galanti	IBM & University of Padua, Italy
Ruth Breu	Research Group Quality Engineering, Austria

The Deployment Model Abstraction Framework

Marcel Weller[1(✉)], Uwe Breitenbücher[2], Sandro Speth[1], and Steffen Becker[1]

[1] Institute of Software Engineering, University of Stuttgart, Stuttgart, Germany
{weller,speth,becker}@informatik.uni-stuttgart.de

[2] Institute of Architecture of Application Systems, University of Stuttgart, Stuttgart, Germany
breitenbuecher@informatik.uni-stuttgart.de

Abstract. For the deployment of applications, various deployment technologies, such as Kubernetes and Terraform, are available to automate the deployment of applications. However, to use these technologies, developers must acquire specialized knowledge about these deployment technologies to create, maintain, and understand deployment models, for example, configuration files created with Kubernetes. In this work, we present and demonstrate the Deployment Model Abstraction Framework (DeMAF), a tool that enables transforming *technology-specific* deployment models into *technology-agnostic* deployment models that are modeled based on the Essential Deployment Metamodel (EDMM). The resulting technology-agnostic EDMM deployment models express deployments only by using the general modeling concepts that are supported by the 13 most prominent technologies. Therefore, the target audience for this demonstration includes developers and architects, who will be shown that such transformations can be automated and that the resulting EDMM models can be understood without knowledge of the original deployment technology. We evaluate the general practical feasibility of the approach by a case study that demonstrates a scenario based on the T2-Project and the technologies Terraform, Kubernetes, and Helm.

Keywords: Deployment models · Infrastructure-as-Code · Abstraction · Transformation · Essential Deployment Metamodel

1 Introduction and Motivation

Most deployment technologies enable executing deployments automatically based on *declarative deployment models*, which describe the application's components to be deployed, their configurations, as well as their dependencies [5]. A deployment model often not only describes the deployment of software components but also needs to cover a variety of other aspects, e.g., the provisioning of infrastructure resources such as virtual machines. As a result, many different deployment technologies with special purposes have been developed, which differ significantly from each other regarding supported (i) *deployment modeling languages* and the (ii) *deployment features* they provide [2,5]. Moreover, especially

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 319–325, 2023.
https://doi.org/10.1007/978-3-031-26886-1_19

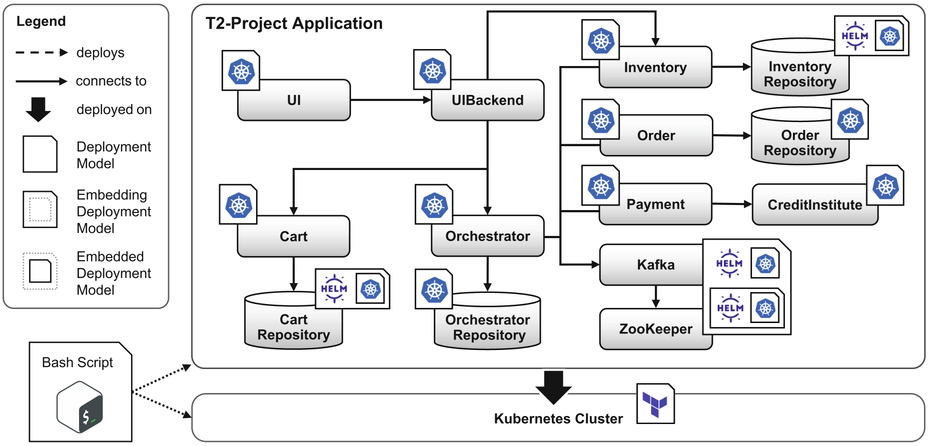


Fig. 1. Overview of the T2-Project and the created deployment model.

for deploying complex applications that consist of various components running in different heterogeneous environments, e.g., a multi-cloud application, often multiple technologies must be combined, which requires deployment models that embed deployment models of other technologies to cover all aspects.

The various available deployment technologies contribute to full automation since almost every composition of components can be deployed by one of these technologies or a combination thereof. For example, the *T2-Project*[1] is a reference architecture for applications following the *microservice architectural style* [3]. It comprises several microservices and databases that together implement an e-commerce webshop for tea as shown in Fig. 1. It contains, e.g., a *UI Service* and several backend services for providing features such as a shopping cart and online payment. The services communicate through asynchronous messaging realized by the *Kafka Service*. Let us assume we want to deploy this composite application on a Kubernetes cluster. We can model the deployment of all business services directly with Kubernetes configuration files. However, for some services, it makes sense to use other deployment technologies that better fit the requirements. For example, for deploying common components such as database systems, the technology *Helm*[2] is better suited since it provides reusable Kubernetes deployment models called *Helm charts*. Thus, such Helm charts can be easily reused, which eliminates the need to define own Kubernetes configuration files. Helm charts always contain an embedded Kubernetes deployment model of the service that they provide. They can also embed other Helm charts, e.g., the Helm chart for the *Kafka Service* embeds another Helm chart for the *ZooKeeper* service. Since Kubernetes itself only enables to describe software deployments based on containerization, we used the Infrastructure-as-Code technology

[1] Implementation of the T2-Project on GitHub: https://github.com/t2-project.
[2] Helm: https://helm.sh/.

Terraform[3] to model the deployment of the Kubernetes cluster itself. Since multiple technologies are involved, we need to provide an overall deployment model that invokes all deployment technologies with the corresponding models, which we implement as a *Bash script*. All mentioned models are available on GitHub[4].

2 Problem Statement

While it makes sense in the scenario described above to select the best fitting deployment technology for each service, it obviously leads to a high number of heterogeneous deployment models and used technologies. Thus, if somebody wants to understand this deployment only based on the provided models, they would have to understand all technical details of all used modeling languages, deployment technologies, and also the overall orchestration. Unfortunately, due to the various available technologies and their technical complexity, the required knowledge for understanding and maintaining such models is immense, especially if they are combined and nested. As a result, an overview of the entire deployment including all services, databases, their configurations, and their dependencies are hard to get from these *technology-specific deployment models*.

3 Deployment Model Abstraction Framework (DeMAF)

In this demonstration, we introduce the *Deployment Model Abstraction Framework (DeMAF)* that addresses the problems described in the previous section by transforming *technology-specific deployment models* into *technology-agnostic deployment models* that are described using the technology-independent Essential Deployment Metamodel (EDMM) [5]. EDMM is the result of a systematic analysis of deployment technologies conducted by Wurster et al. [5] and contains only model entities that can be mapped to 13 of the most prominent declarative deployment technologies such as Terraform, Kubernetes, and AWS CloudFormation. Therefore, EDMM models contain no deployment technology-specific details and only describe the (i) components that get deployed, (ii) their configurations, as well as (iii) their dependencies. Thus, EDMM models provide all information required for understanding the architecture of the deployed system and require no technical expertise about concrete deployment technologies.

An overview of the architecture of DeMAF is shown in Fig. 2. It comprises several components that are independent self-contained microservices. For supporting the transformation of technology-specific deployment models in an extensible way, the DeMAF follows a plugin-based approach: Each plugin is responsible for transforming a deployment model created with a specific deployment technology into entities of EDMM. This way, the framework can be extended to support more technologies, new major versions of already supported technologies, or alternative analysis methods that work better or supplement existing plugins.

3 Terraform: https://www.terraform.io/.

4 Deployment model for the T2-Project on GitHub: https://github.com/Well5a/kube.

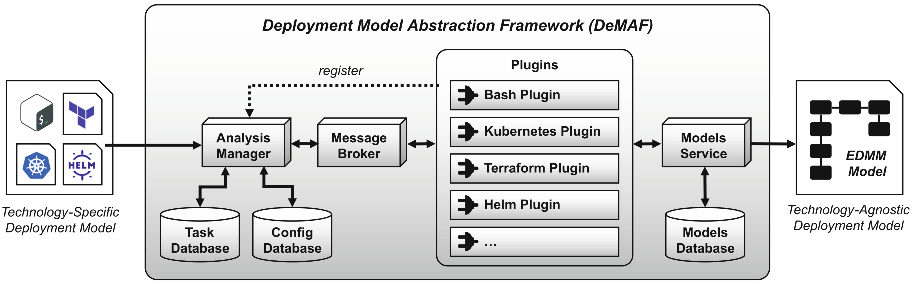


Fig. 2. Concept of the deployment model abstraction framework.

At startup, the plugins register at the *Analysis Manager*, which manages the transformation process, handles user interaction, and persists information about registered plugins in the *Config Database*. We realized the user interface of the Analysis Manager through a command-line interface that provides a command for inputting a deployment model and starting the transformation process. After that, the Analysis Manager determines the deployment technology of the given deployment model, creates a task, and sends it to the appropriate plugin. The tasks are persisted in the *Task Database* to keep track of their status. A RabbitMQ *Message Broker* facilitates asynchronous communication between the Analysis Manager and the plugins. To enable the transformation of technology-specific deployment models that utilize several other deployment technologies, the plugins can detect *embedded deployment models* and the corresponding deployment technologies. An embedded deployment model is a deployment model that is contained in another deployment model, an *embedding deployment model*. The embedded deployment model prescriptively describes the deployment of a specific part of the embedding deployment model. A deployment model can embed any number of other deployment models. The embedded deployment model is possibly created with a different technology than the embedding deployment model and can recursively contain further embedded deployment models. The plugins report detected embedded deployment models to the Analysis Manager. The DeMAF analyzes each embedded deployment model separately by distributing tasks for each embedded deployment model across the plugins.

During the analysis, the plugins transform the found information into an EDMM model, which is persisted in the *Models Database* and managed by the *Models Service*, which provides a common interface for all plugins. The transformation process is finished when the whole given technology-specific deployment model has been analyzed, i.e., including all embedded deployment models. The DeMAF then outputs the created EDMM model as a YAML file[5].

[5] EDMM in YAML Specification: https://github.com/UST-EDMM/spec-yaml.

We implemented a prototype of the DeMAF[6] that currently supports the four deployment technologies Kubernetes, Terraform, Helm, and Bash. We chose these technologies because they cover a variety of use cases and are commonly integrated in deployment models: Terraform focuses on infrastructure components, while Kubernetes is based on containerization and restricted to the deployment of software components. Helm is a package manager for Kubernetes, which provides reusable Kubernetes deployment models. Bash follows an imperative modeling approach. The plugins are implemented in Java and parse the objects of the technology-specific deployment model into an intermediary representation of Java objects that are then transformed into EDMM entities.

4 Evaluation: Case Study

We conducted a case study for validating the DeMAF's ability to transform technology-specific deployment models into EDMM models. We used the deployment model of the T2-Project introduced in Sect. 1 as input, for which we first manually created an EDMM model that contains all information we expect, called the *expected* EDMM model. To ensure that this information is complete, we thoroughly examined the files of the technology-specific deployment model for the T2-Project including all embedded deployment models. We compared this expected EDMM model with the *actual* EDMM model that is automatically generated by the DeMAF and investigated the differences. Thereby, we ignored minor differences in naming, YAML indentation, or the order of elements. The comparison showed that the DeMAF can successfully transform the given technology-specific model into an appropriate EDMM model: The actual EDMM model contains all of the information in comparison to the expected EDMM model. The DeMAF created EDMM components for all services of the T2-Project application shown in Fig. 1. It connected these components with appropriate EDMM relations that accurately describe the dependencies between the services. For the Kubernetes cluster, it created several EDMM components that describe the provided resources and other capabilities such as a container runtime. Additionally, the DeMAF created EDMM relations showing how these components host the T2-Project services. All evaluation results including the expected and actual EDMM model are provided on Zenodo[7]. Moreover, we created a video that is available on YouTube which demonstrates this case study[8].

5 Related Work

Previous work already provided concepts for transforming technology-specific deployment models into technology-agnostic deployment models using the *Topology and Orchestration Specification for Cloud Applications (TOSCA)* [1,4]. Wettinger et al. [4] crawl public code repositories for technology-specific deployment

[6] GitHub organization with the DeMAF prototype: https://github.com/UST-DeMAF.

[7] Zenodo repository with evaluation results: https://doi.org/10.5281/zenodo.6824223.

[8] https://youtu.be/nHl-8zxY-mU.

models and transform them into more generic representations that can be integrated with each other. They crawl technology-specific deployment models created with Chef and Juju and, for each of them, generate TOSCA node types that contain meta information such as the name. However, they do not transform the technology-specific deployment models but rather attach them to the generated TOSCA node types by wrapping them into TOSCA artifacts. Endres et al. [1] follow a similar approach. They crawl public code repositories for technology-specific deployment models created with Chef and derive the structure of the application that is described into so-called *technology-agnostic topology models*, similarly to our work. These topology models are based on the TOSCA standard and are combined with the originating technology-specific deployment models, which are together stored in a repository. This enables the modeling and deployment of applications composed of several technology-specific deployment models without the need to understand the technical details of the used technologies because this information is given by the attached technology-agnostic topology models. However, their tool can currently only transform technology-specific deployment models created with Chef. Moreover, the overall concept does not deal with the transformation of more complex technology-specific deployment models that contain embedded deployment models created with different deployment technologies, which is supported by the DeMAF approach.

6 Conclusion and Future Work

The first prototypical realization of the DeMAF showed promising results on which we can now build on. In future work, we will improve the existing plugins and provide support for further deployment technologies. This may include reworking the transformation logic into a more standardized and verifiable approach, e.g., using the Eclipse Modeling Framework (EMF) to transform the deployment models with transformation rules specified in the MOF QVT (Query/View/Transformation) standard. Additionally, other researchers or developers can use the DeMAF and provide custom plugins. However, developing a system that reliably transforms arbitrary deployment technologies requires substantial effort. Because deployment technologies differ heavily, we would need to implement many plugins which allow the reuse of code to a limited extent. Additionally, deployment technologies evolve by changing existing features or introducing new concepts that make a plugin incompatible. Therefore, it would be beneficial to find more general approaches for analyzing the technology-specific deployment models, like monitoring of network traffic of a deployed application.

References

1. Endres, C., et al.: Anything to topology - a method and system architecture to topologize technology-specific application deployment artifacts. In: Proceedings of the 7th International Conference on Cloud Computing and Services Science (CLOSER 2017), pp. 180–190. SciTePress, April 2017

2. Lu, H., et al.: Pattern-based deployment service for next generation clouds. In: 2013 IEEE Ninth World Congress on Services, pp. 464–471. IEEE (2013). https://doi.org/10.1109/SERVICES.2013.54
3. Speth, S., Stieß, S., Becker, S.: A saga pattern microservice reference architecture for an elastic SLO violation analysis. In: Companions Proceedings of 19^{th} IEEE International Conference on Software Architecture (ICSA-C 2022). IEEE, Mar 2022. https://doi.org/10.1109/ICSA-C54293.2022.00029
4. Wettinger, J., Breitenbücher, U., Kopp, O., Leymann, F.: Streamlining DevOps automation for cloud applications using TOSCA as standardized metamodel. Future Gener. Comput. Syst. **56**, 317–332 (2016). https://doi.org/10.1016/j.future.2015.07.017
5. Wurster, M., et al.: The essential deployment metamodel: a systematic review of deployment automation technologies. SICS Softw. Intensive Cyber-Phys. Syst. **35**, 63–75 (2020). https://doi.org/10.1007/s00450-019-00412-x

Dromi: A Tool for Automatically Reporting the Impacts of Sagas Implemented in Microservice Architectures on the Business Processes

Sandro Speth[1(✉)], Uwe Breitenbücher[2], Sarah Stieß[1], and Steffen Becker[1]

[1] Institute of Software Engineering, University of Stuttgart, Stuttgart, Germany
{speth,stiess,becker}@informatik.uni-stuttgart.de
[2] Institute of Architecture of Application Systems, University of Stuttgart, Stuttgart, Germany
breitenbuecher@informatik.uni-stuttgart.de

Abstract. Distributed transactions that span multiple microservices are more and more realized using the Saga Pattern. However, in case the interaction between microservices fails due to an Service Level Objective (SLO) violation, e.g., insufficient availability, the executed business logic gets significantly impacted when a saga compensates already executed operations. Unfortunately, analyzing such impacts manually and reporting found issues is too slow for modern systems. Therefore, we present *Dromi*, a model-based tool that traces the impacts of SLO violations across a microservice architecture and, if a violation results in compensations caused by sagas, creates an issue report about the violation's location and resulting impacts on the business processes. The target audience of this demonstration includes architects and developers, who will be shown how such impacts are detected automatically by Dromi.

Keywords: Microservices · Impact analysis · SLO · Business processes

1 Introduction

In modern software architectures, often independent microservices work together to implement business processes. However, many processes require distributed business transactions that span multiple services which do not support atomic commit protocols. Therefore, business transactions in microservice architectures are often realized using the *Saga Pattern* [6]. A saga is a sequence of local transactions executed by different microservices. If a local transaction of a microservice fails, the saga executes a series of compensating transactions to undo the changes made by the preceding local transactions that were completed successfully.

However, local transactions executed by microservices can quickly fail due to SLO violations of other services, which promise the quality of a service, such as availability or maximum response time. For example, if a microservice executes a local transaction that depends on the invocation of another service that

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 326–331, 2023.
https://doi.org/10.1007/978-3-031-26886-1_20

is many minutes not available, the transaction fails because the other service violates its SLO guaranteeing a maximum downtime of 10 s. Thus, such SLO violations quickly result in undoing several already completed local transactions of different microservices due to a saga that executes compensating transactions for them. Of course, this heavily impacts the business processes implemented by the microservices since the executed functional business logic gets changed due to the violation of a non-functional SLO. Unfortunately, if many business processes are implemented by a large microservice architecture consisting of many independent services, (i) detecting the impact of SLO violations on the business processes and (ii) reporting analyzed issues is a complex challenge and can hardly be done manually since these tasks are very time consuming and error propagations are hard to detect manually during runtime.

To tackle these issues, in this demonstration, we present *Dromi*, a model-based tool that traces the impacts of SLO violations across a microservice architecture and, if a violation results in compensations caused by sagas, creates an issue report about the violation's location and resulting impacts on the business processes. To enable this, we also introduce the *Dromi Modeling Language (DML)* in this paper, which combines models for microservice architectures, business processes, and sagas. Developers can link the elements of the models with each other to specify their dependencies, which are then used by Dromi to derive impact traces automatically and to generate issues in case an SLO violation has an impact on a business process. Thus, issues are reported automatically without human interaction once the DML model is created. We demonstrate how Dromi and the Dromi Modeling Language can be used in a video[1] and provide all implementations as open source code in GitHub.

2 Motivating Scenario

This section describes the motivating scenario used for explaining Dromi. In the scenario, we consider the T2-Project [9], a microservice architecture for a webshop that implements a business process to order teas. The architecture includes several services to realize an order: (1) the *cart service*, (2) an *orchestrator service*, (3) the *inventory service*, (4) the *order service*, and (5) the *payment service*, which invokes a (6) *credit institution service*. To maintain the consistency of an order, the inventory service, order service, and payment service participate in a saga in which the orchestrator manages to roll back orders in the event of a failure. However, although the saga is implemented, it is typically not directly observable because it is often not modeled in the architecture. In particular, the effects of a failure on the business process itself are not directly visible because the saga logic is often hidden directly in the microservices' code. For example, if the credit institution violates an SLO, e.g., because the service is currently unavailable or the request times out before a response is sent, it cannot be reached by the payment service. Therefore, the payment fails and triggers a saga compensation, which has side effects on the business process. However, these side

[1] https://youtu.be/3E90neB-iUY.

effects are hidden in the code of the individual microservices that the orchestrator controls. Therefore, it is not apparent what impact a non-functional SLO violation has on the functional level of the business process.

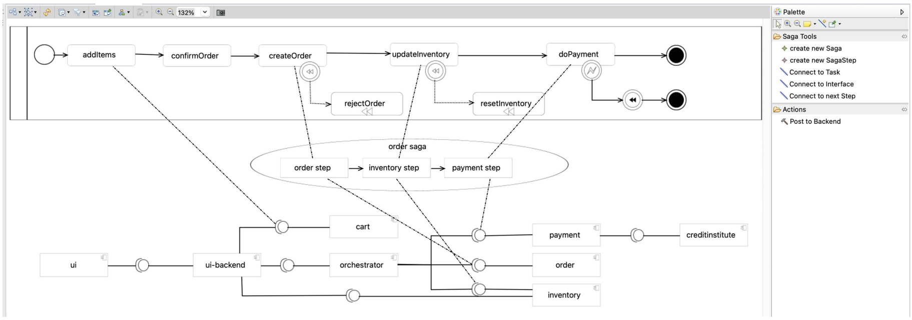


Fig. 1. Dromi Editor showing the motivating scenario modelled in DML.

3 The Dromi Modeling Language

The Dromi Modeling Language (DML) consists of three connected layers: (1) the *Architecture Layer*, (2) the *Business Process Layer*, and (3) the *Saga Layer*. Figure 1 shows a screenshot of our *Dromi Editor*, which is presented in detail in Sect. 4. We explain the layers based on this screenshot to give an overview.

The *Architecture Layer* describes the architecture of the system to be observed. Please note that DML is not bound to microservices but supports any type of component. Therefore, in DML, the Architecture Layer is realized as *UML Component Diagram*, where each component can provide and require multiple interfaces that other components can invoke. The Architecture Layer of the motivating scenario is shown on the bottom of Fig. 1, which describes the microservices of the webshop, their interfaces, and their dependencies.

In the *Business Process Layer*, developers model the business processes implemented by the microservices described in the Architecture Layer. For example, Fig. 1 shows on the top the order process implemented by the webshop microservices of the motivating scenario. Please note that these *business process models* are not executable but just describe the business logic that is implemented by the microservices. Thereby, a single activity contained in the process model can be implemented by one or more microservices. The Business Process Layer is realized using the business process modeling language *BPMN 2.0* since BPMN is widely used in practice. Another reason is that BPMN enables the modeling of *transactional sub-processes* and supports the concept of *compensation*, which is concerned with undoing activities that were already successfully completed because of their effects are no longer desired and need to be reversed. Thus, these BPMN modeling concepts can be used to describe the effects on the business process caused by microservice sagas in the case of compensation.

Business process models are often large. Thus, transactional behavior and compensation activities might be scattered throughout the model, which makes it messy to understand dependencies, e.g., in which situation which activities are compensated. Therefore, DML defines an optional *Saga Layer* located between the two other layers as shown in Fig. 1, which is used for modeling sagas and their steps. Using the Saga Layer, business transactions can be separately modeled as sagas, independently of the Business Process Layer. For example, Fig. 1 shows the *order saga*, which consists of an *order step* that maintains order information, an *inventory step* that removes ordered products from the inventory, and a *payment step* that executes the payment. The steps are executed in the described order.

To enable impact analysis across the layers, developers need to specify the dependencies of the elements in the three layers using *links*, which are shown as dashed lines in Fig. 1. DML allows three ways to link layers and their elements: First, (i) a microservices interface can be directly linked with the BPMN activity realized by this interface. Second, (ii) if the microservice architecture implements one or more saga, links can be directed from a microservice interface to a saga step. Since each microservice interface can be involved in an arbitrary number of sagas, each interface can have multiple links to different saga steps. Finally, (iii) to relate saga steps to BPMN activities, links can be specified between them.

4 System Architecture of Dromi and Demonstration

This section presents Dromi, a tool that automatically detects and reports the impacts of sagas implemented in microservice architectures on the business processes they implement. Thereby, Dromi focuses on the impacts on processes caused by the compensation of sagas that result from SLOs violations of microservices, e.g., high latencies. The source code of Dromi is available on GitHub[2].

As depicted in Fig. 2, Dromi consists of two parts: (i) *Dromi Frontend* and the (ii) *Dromi Backend*. The frontend is a graphical editor for DML models implemented as Eclipse EMF plugin. It enables importing BPMN 2.0 process models created with BPMN tools. Similarly, it enables importing architecture models created using the tool Gropius, which supports a UML Component Diagram-like notation [8]. Finally, the frontend enables drawing sagas in the middle layer and linking elements with each other as described in Sect. 3.

The Dromi Backend consumes DML models. For detecting SLOs violations, the tool *SoLOMON* [7] is integrated. SoLOMON automatically imports the architecture model from Gropius, with which developers map components deployed on Kubernetes to components of the architecture. Then, they model SLOs for the components via SoLOMON's frontend and send them to the SoLOMON backend, where the SLOs are transformed to PromQL queries and monitored through Prometheus. If an alert is triggered, i.e., an SLO is violated, SoLOMON creates an issue for the affected component in Gropius describing the violation. Furthermore, SoLOMON publishes an event that Dromi can subscribe for containing the violated SLO, affected component, time, and the issue ID.

[2] https://github.com/stiesssh/dromi-backend, https://github.com/stiesssh/dromi-models.

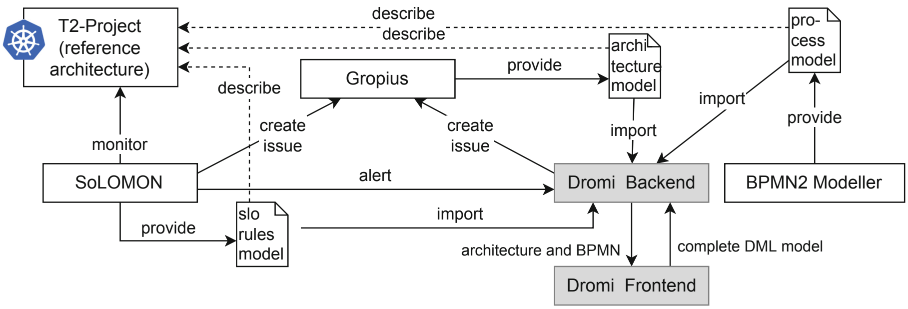


Fig. 2. System architecture of Dromi.

When the Dromi backend receives this event, it executes the *Impact Analysis*, which starts in the Architecture Layer with the microservice that is affected by the SLO violation. In this work, we assume that an SLO violation triggers a saga compensation. If a link exists from the interface of this service to a higher layer, the analysis follows the link until a BPMN task is reached. If no link to a saga or business process exists for the affected microservice, Dromi follows the call chain of the microservice architecture reversely to a service with a link to a higher layer that can be followed. The result of the impact analysis is an *impact trace* that includes the affected microservices, saga steps, and BPMN activities.

After analyzing impacts, the backend automatically executes *Impact Reporting* by creating an issue with the impact trace in Gropius [8], which is a *uniform issue management system* that integrates, e.g., GitHub and Jira. Thereby, it allows reporting issues independently of a microservice's actual issue tracker.

We recorded a video (See footnote 1) to demonstrate the following example with Dromi: Assume the credit institute service in Fig. 1 violates its availability SLO and is unavailable when an order should be checked out. Then the payment service cannot reach it to perform the payment. As a result, the orchestrator service rolls back the saga. Dromi analyses the impact of the SLO violation and reports an issue stating the violated SLO, time, *credit institute service* and *payment service*, the *payment step* of the saga, and the failed business activity, i.e., *doPayment*.

5 Related Work

Hanemann et al. [3] decide on system recovery actions on the cost caused by the SLO violation which caused the failure. They analyze the impact of the violation on the directly dependent services to estimate the cost. Unlike Dromi, they do not analyze the propagation of a violation along the call chain through the architecture. Furthermore, they do not consider impacts on the business process. Mohamed et al. [5] calculate architecture service routes, i.e., sequences of services connected through interfaces, and use them to consider failure propagation. However, they do not consider the process and lack the focus on SLOs, as they consider failures in general. Kleehaus et al. [4] developed the MICROLYZE framework that uses static and dynamic data for dynamic microservice architecture recovery. They also include the business process in their recovery and map tasks to interfaces. For future work, they mention a failure-impact visualization. However, they do not consider patterns. In the domains of requirements engineering and process alignment, authors connect elements from different modeling languages to ensure conformance between process and architecture model [1,2]. However, they use direct connections, which are not sufficient to achieve our objectives, as we must also consider patterns, which none of those works does.

6 Conclusion and Future Work

We showed that Dromi and DML enable the automated impact analysis of a non-functional SLO violation on the functional logic of business processes when transactions span multiple microservices following the Saga pattern. Thus, Dromi is able to trace the violation's impact from the architecture to the business process. In future work, we plan to integrate more microservice patterns in Dromi.

References

1. Aversano, L., Grasso, C., Tortorella, M.: Managing the alignment between business processes and software systems. Inf. Softw. Technol. **72**, 171–188 (2016). https://doi.org/10.1016/j.infsof.2015.12.009
2. Elvesæter, B., Panfilenko, D., Jacobi, S., Hahn, C.: Aligning business and it models in service-oriented architectures using BPMN and SoaML. In: Proceedings of the First International Workshop on Model-Driven Interoperability, MDI 2010, pp. 61–68. Association for Computing Machinery (2010). https://doi.org/10.1145/1866272.1866281
3. Hanemann, A., Schmitz, D., Sailer, M.: A framework for failure impact analysis and recovery with respect to service level agreements. In: 2005 IEEE International Conference on Services Computing (SCC 2005) Vol-1, vol. 2, pp. 49–56 (2005). https://doi.org/10.1109/SCC.2005.10
4. Kleehaus, M., Uludağ, Ö., Schäfer, P., Matthes, F.: MICROLYZE: a framework for recovering the software architecture in microservice-based environments. In: Mendling, J., Mouratidis, H. (eds.) CAiSE 2018. LNBIP, vol. 317, pp. 148–162. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92901-9_14
5. Mohamed, A., Zulkernine, M.: On failure propagation in component-based software systems. In: 2008 The Eighth International Conference on Quality Software. pp. 402–411 (2008). https://doi.org/10.1109/QSIC.2008.46
6. Richardson, C.: Microservices Patterns: With examples in Java. Manning Publications (2018)
7. Speth, S.: Semi-automated cross-component issue management and impact analysis. In: Proceedings of 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 1090–1094. IEEE (2021). https://doi.org/10.1109/ASE51524.2021.9678830
8. Speth, S., Becker, S., Breitenbücher, U.: Cross-component issue metamodel and modelling language. In: Proceedings of the 11th International Conference on Cloud Computing and Services Science (CLOSER 2021). SciTePress (2021). https://doi.org/10.5220/0010497703040311
9. Speth, S., Stieß, S., Becker, S.: A saga pattern microservice reference architecture for an elastic SLO violation analysis. In: Companion Proceedings of 19th IEEE International Conference on Software Architecture (ICSA-C 2022). IEEE (2022). https://doi.org/10.1109/ICSA-C54293.2022.00029

Exploring Enterprise Architecture Knowledge Graphs in Archi: The EAKG Toolkit

Philipp-Lorenz Glaser[1], Syed Juned Ali[1], Emanuel Sallinger[2], and Dominik Bork[1(✉)]

[1] Business Informatics Group, TU Wien, Vienna, Austria
{philipp-lorenz.glaser,syed.juned.ali,dominik.bork}@tuwien.ac.at
[2] Database and Artificial Intelligence Group, TU Wien, Vienna, Austria
emanuel.sallinger@tuwien.ac.at

Abstract. This paper presents the *EAKG Toolkit* that entails a new Knowledge Graph-based representation of enterprise architecture (EA) models and further enables reasoning on EA knowledge. Our developed EAKG Toolkit is unique in the sense that it *i*) transforms ArchiMate models into a KG representation – the Enterprise Architecture Knowledge Graph (EAKG), *ii*) visualizes the EAKG for interactive exploration, and *iii*) extends the EAKG with additional nodes and edges to visually represent detected EA smells.

Keywords: Enterprise architecture · Knowledge graph · Modeling tool · ArchiMate · Archi

1 Introduction

Enterprise architecture models are graphical representations that provide valuable support for, e.g., integrated IT and business decision-making [1], planning future states of the enterprise, and improving the business and IT alignment [2]. To support all these functions, EA models need to be analyzed efficiently. Such EA analysis involves querying models with the aim of evaluating various properties [7]. However, holistic EA models grow in size and complexity, thereby hampering manual human analysis while advanced and automated analysis of EA models is surprisingly underrepresented in research and EA tooling so far [12].

EA modeling tools do not take full advantage of the several structural properties of EA models represented as graphs, such as the differentiation of relations between elements, discovery of paths, clusters, or graph metrics. Current approaches are often tied to a concrete EA approach, offering a limited set of visualization techniques. EA modeling tools offer different features based on the supported EA approach and the analytical capabilities provided and thus, restrict the kind of analysis that they support [6]. A survey from 2016 yielded that *"Modern analysis approaches should combine interactive visualizations with automated analysis techniques"* [5]. The need for proper tool support was pointed out in the past as one EA [10] and business information systems modeling [3]

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 332–338, 2023.
https://doi.org/10.1007/978-3-031-26886-1_21

research gap. Our EAKG Toolkit addresses this gap by utilizing the full potential of the graphical structure of EA models.

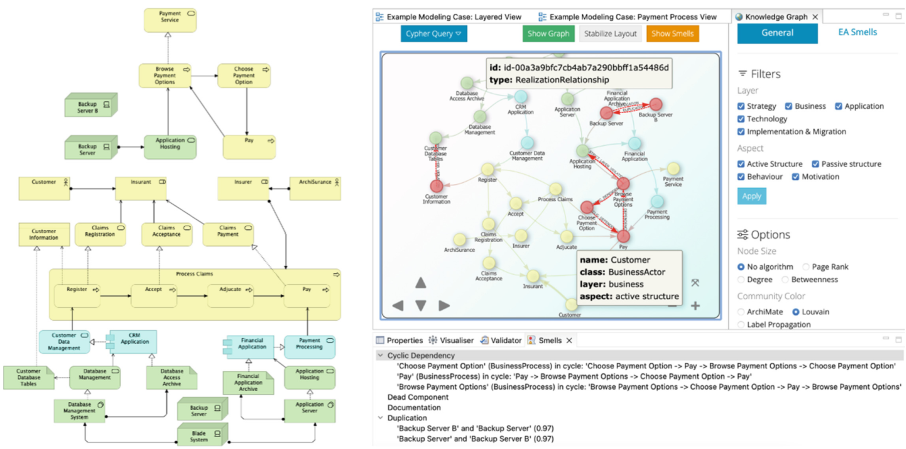
In the context of EA, graph-based formalisms have been applied for representation and reasoning of EA models [9,12] but these works are merely constrained to the explicit knowledge encoded by the EA model (i.e., no further knowledge enrichment) and to basic model analysis (i.e., no KG-based reasoning). In this paper, we present a toolkit for Archi that exploits the benefits of KG-based representation and reasoning in EA, by constructing Enterprise Architecture Knowledge Graphs (EAKGs). The EAKG Toolkit visualizes and analyses the EAKG and supports the EAKG knowledge enrichment. EAKG provides a generic and unified intermediary representation of EAs which makes our approach easily extensible for the integration of other graph-based EA analysis tools.

2 The EAKG Toolkit

The aim of the toolkit is to make KG-based EA analysis available to enterprise architects, i.e., an audience that not necessarily has graph theoretic knowledge. In this section, we first present the features of our toolkit, then we present the internal architecture and implementation details. Eventually, we showcase the usage of our toolkit with a running example (see Fig. 1).

2.1 EAKG Features

The main features of the EAKG toolkit include the visualization of the transformed EAKG and the additional analysis support provided by the graph characteristics and EA smells enrichment of the EAKG. Figure 1b shows the integration of the toolkit within the Archi application, containing both the main *Graph View* (top), and the *Smells Report View* (bottom).



(a) Source EA model (b) Transformed EAKG with toolkit UI

Fig. 1. EAKG toolkit in Archi (Color figure online)

Knowledge Graph Visualization. The main view in Fig. 1b visualizes the EAKG generated from the source EA model shown in Fig. 1a. Nodes denote ArchiMate elements, while edges denote ArchiMate relationships. The transformation maps the properties related to layers (e.g., *Business, Application*) and aspects[1] (e.g., *Active Structure, Passive Structure*) in the original EA model to the properties of the nodes in the resulting EAKG (extended from [11]). The relationship type (e.g., *Realization, Assignment*) is stored in the properties of the relationships in the EAKG. Further properties are exposed by hovering over nodes and edges, as exemplified by the *Customer* element and the relationship between the elements *Database Access Archive* and *CRM Application*.

Graph Characteristics Visualisation. EAKG allows applying graph algorithms (e.g., page rank, degree) on the transformed EAKG to represent centrality and community metrics. The applied algorithms enrich the EAKG with additional properties: graph centrality measures are reflected via the node size, whereas community measures are reflected via node color. EAKG also allows the customization of the EAKG visualization, e.g., by filtering specific ArchiMate layers/aspects and configuring how to represent graph analysis results. Note that in the figure all filters are checked and no graph algorithms are set, thus EAKG visualizes all elements of the source EA model with the same node size and the color according to the ArchiMate layer.

EA Smells Detection. The EAKG Toolkit allows the detection and visualization of EA smells [8,13] in the EAKG. EA smells provide necessary information to the modeler to rectify models designed with bad modeling practices. Our tool visualizes the found smells by means of, e.g., additional relationships and highlighting of affected nodes in red color as shown in Fig. 1b. We contribute here a much richer visualization of EA smells that again uses a Knowledge Graph that allows exploration of the smell in its context. Currently, the EAKG toolkit detects eight different EA smells by running cypher queries on the EAKG. The tool moreover supports the execution of custom cypher queries.

2.2 EAKG Architecture

The EAKG Toolkit (see Fig. 2) is primarily developed with Java and built upon the Eclipse Rich Client Platform (RCP). In the following, we describe the architecture based on our model-based KG creation process [4].

Knowledge Graph Creation. Once the creation process is initiated (from a dedicated action in the menu of Archi), the toolkit creates the *Knowledge Graph Database Manager*, which is responsible for interacting with an *Embedded Neo4j Graph Database*. The manager starts a new database, stored on the local file system, and opens a *Bolt Connector* for remote access (used, e.g., by drivers). After the database is started, the *Knowledge Graph Exporter* uses the current ArchiMate model to iterate over the elements and retrieve its metadata. Each element corresponds to a node and gets stored in the graph database, together

[1] https://pubs.opengroup.org/architecture/archimate3-doc/.

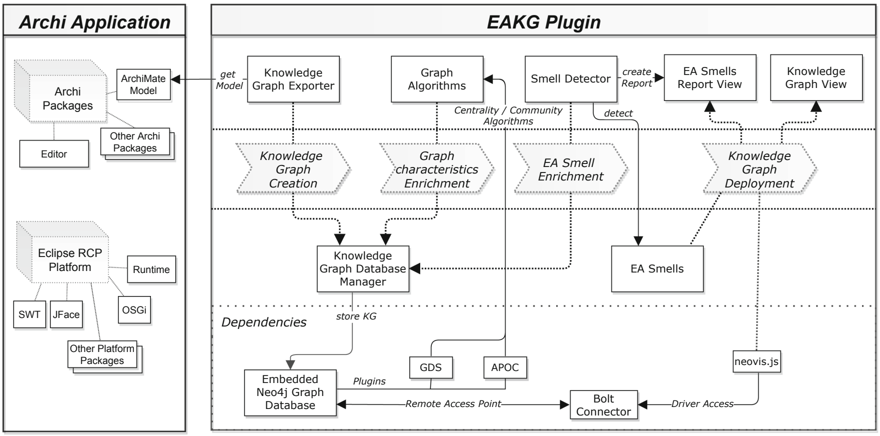


Fig. 2. EAKG toolkit architecture

with initial properties, e.g., layer and aspect. Next, the CSV export provided by Archi is reused to export all the relationships of the model and load the resulting CSV file with a single query into the graph database. The query creates the initial edges between nodes and also stores the relationship type as an edge property.

Graph Characteristics Enrichment. The Graph Database Manager also registers additional procedures, provided by the *Graph Data Science (GDS)* and *Awesome Procedures On Cypher (APOC)* Neo4j plugins, to leverage efficient *Graph Algorithms* in the graph database. The exporter runs the query procedures and sets the corresponding properties in the graph.

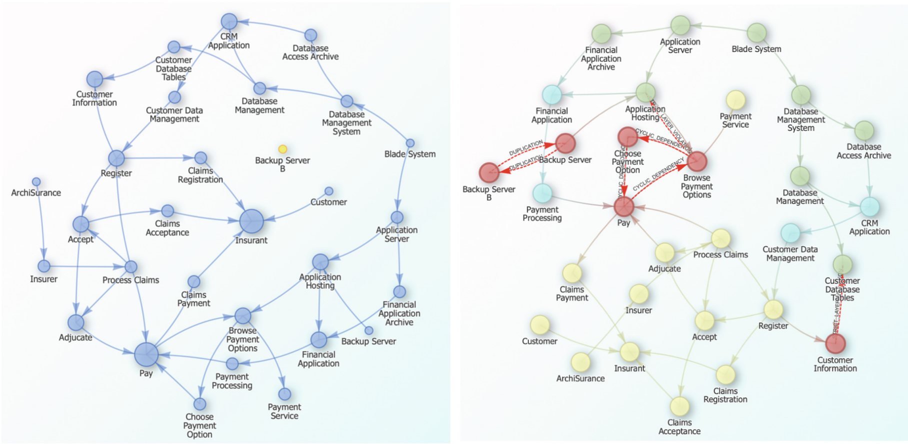
EA Smells Enrichment. The *Smell Detector* runs additional cypher queries to further enrich the EAKG with detected *EA Smells*. The detector also stores additional information such as the affected elements and creates the tree structure for the report.

Knowledge Graph Deployment. The graph database now holds the EAKG enriched both with graph characteristics and EA Smells and can now be fully deployed within Archi. The *EA Smells Report View* incorporates the tree structure created from the smell detector. The *Knowledge Graph View* is part of a web browser component, which simply displays an HTML document with additional CSS, JavaScript, and a neovis.js configuration for the visualization[2].

2.3 EAKG Use

We finally elaborate on the usage of our toolkit in order to demonstrate the aforementioned features. After adding the toolkit to Archi, the Knowledge Graph menu and its items are exposed in the Archi menu bar, where the EAKG creation

[2] neovis.js: https://github.com/neo4j-contrib/neovis.js/.



(a) Graph Analysis View

(b) Smell Detection View

Fig. 3. KG-based EA analysis representations in EAKG (Color figure online)

process can be initiated. Once the database is started and the EAKG creation is finished, the visualization and EA smell report views can be opened (see Fig. 1b). The toolbar at the top allows changing the representation of the graph by running custom cypher queries or visualizing the detected smells.

The right-hand sidebar includes a filter and option menu. Enterprise architects can filter the displayed elements based on specific layers or aspects of ArchiMate. The option menu on the bottom right offers configuration for the *Graph characteristics Knowledge Graph Enrichment*. Figure 3a visualizes the resulting graph after *Node Size* is set to *Degree* and the *Community Color* to *Weakly Connected Components*. Degree denotes the number of connections, and, as can be seen, the size of nodes increases with the amount of incoming and outgoing edges. Similarly, the weakly connected components algorithm detects individual sub-graphs that are rendered in different colors.

The Report view at the bottom lists all detected EA Smells together with the affected elements in the model. In the main view above, the toolbar offers buttons to either show the default graph or to also include EA Smells in the visualization, with affected elements highlighted in red and references to other elements of the smell represented as dashed, red edges. Figure 3b showcases this behavior with nodes and edges that are part of a detected EA Smell highlighted in red with the name of the detected EA Smell as a label. The *EA Smells* tab in the sidebar provides information about each EA Smell, including a visualization, a description, and a solution to fix the smell.

3 Conclusion and Future Work

We presented an Archi-based tool that transforms EA models into Enterprise Architecture Knowledge Graphs (EAKGs) that can be semantically enriched by general graph knowledge and domain-specific enterprise architecture knowledge. Our approach allows full automation for the entire EAKG construction process and provides an efficient and intuitive GUI to explore and analyze the EAKG. The most innovative contribution we make with this tool is that we not only use the KG for EA visualization and analysis but also for representing EA knowledge using, e.g., the added nodes and relationships for EA Smells. Consequently, we propose to not only use KGs for automated analysis of overarching EA models, but also to improve human understandability by appropriate interactive visualizations. The EAKG Toolkit is open source on https://github.com/borkdominik/archi-kganalysis-plugin and a video can be found here: https://youtu.be/a59OawYwiqE.

Acknowledgements. This work has been partially funded through the Erasmus+ KA220-HED project Digital Platform Enterprise (project no.: 2021-1-RO01-KA220-HED-000027576) and the Austrian Research Promotion Agency via the Austrian Competence Center for Digital Production (contract no. 854187).

References

1. Buschle, M., Holm, H., Sommestad, T., Ekstedt, M., Shahzad, K.: A tool for automatic enterprise architecture modeling. In: Nurcan, S. (ed.) CAiSE Forum 2011. LNBIP, vol. 107, pp. 1–15. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29749-6_1
2. Florez, H., Sánchez, M., Villalobos, J.: A catalog of automated analysis methods for enterprise models. Springerplus **5**(1), 1–24 (2016). https://doi.org/10.1186/s40064-016-2032-9
3. Frank, U., Strecker, S., Fettke, P., vom Brocke, J., Becker, J., Sinz, E.J.: The research field "modeling business information systems" - current challenges and elements of a future research agenda. Bus. Inf. Syst. Eng. **6**(1), 39–43 (2014)
4. Glaser, P.L., Ali, S.J., Sallinger, E., Bork, D.: Model-based construction of enterprise architecture knowledge graphs. In: Almeida, J.P.A., Karastoyanova, D., Guizzardi, G., Montali, M., Maggi, F.M., Fonseca, C.M. (eds.) EDOC 2022. LNCS, vol. 13585, pp. 57–73. Springer, Cham (2022)
5. Lantow, B., Jugel, D., Wißotzki, M., Lehmann, B., Zimmermann, O., Sandkuhl, K.: Towards a classification framework for approaches to enterprise architecture analysis. In: Horkoff, J., Jeusfeld, M.A., Persson, A. (eds.) PoEM 2016. LNBIP, vol. 267, pp. 335–343. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48393-1_25
6. Naranjo, D., Sánchez, M., Villalobos, J.: PRIMROSe: a graph-based approach for enterprise architecture analysis. In: Cordeiro, J., Hammoudi, S., Maciaszek, L., Camp, O., Filipe, J. (eds.) ICEIS 2014. LNBIP, vol. 227, pp. 434–452. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22348-3_24
7. Närman, P., Buschle, M., Ekstedt, M.: An enterprise architecture framework for multi-attribute information systems analysis. Softw. Syst. Model. **13**(3), 1085–1116 (2014)
8. Salentin, J., Hacks, S.: Towards a catalog of enterprise architecture smells. In: Gronau, N., Heine, M., Krasnova, H., Poustcchi, K. (eds.) Internationalen Tagung Wirtschaftsinformatik, Community Tracks, pp. 276–290. GITO Verlag (2020)

9. Santana, A., Fischbach, K., de Moura, H.P.: Enterprise architecture analysis and network thinking: a literature review. In: Bui, T.X., Jr., R.H.S. (eds.) 49th Hawaii International Conference on System Sciences, pp. 4566–4575. IEEE (2016)
10. Santana, A., Simon, D., Fischbach, K., de Moura, H.: Combining network measures and expert knowledge to analyze enterprise architecture at the component level. In: 2016 IEEE EDOC Conference, pp. 1–10. IEEE (2016)
11. Smajevic, M., Bork, D.: From conceptual models to knowledge graphs: a generic model transformation platform. In: International Conference on Model Driven Engineering Languages and Systems Companion, pp. 610–614 (2021)
12. Smajevic, M., Bork, D.: Towards graph-based analysis of enterprise architecture models. In: Ghose, A., Horkoff, J., Silva Souza, V.E., Parsons, J., Evermann, J. (eds.) ER 2021. LNCS, vol. 13011, pp. 199–209. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-89022-3_17
13. Smajevic, M., Hacks, S., Bork, D.: Using knowledge graphs to detect enterprise architecture smells. In: Serral, E., Stirna, J., Ralyté, J., Grabis, J. (eds.) PoEM 2021. LNBIP, vol. 432, pp. 48–63. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-91279-6_4

Interactive Design of Time-Aware Business Processes

Keti Lila, Marco Franceschetti, and Julius Köpke(✉)

Department of Informatics-Systems, Alpen-Adria Universität Klagenfurt, Universitaetsstrasse 65-67, 9020 Klagenfurt am Wörthersee, Austria
{keti.lila,marco.franceschetti,julius.koepke}@aau.at
https://www.aau.at/isys/

Abstract. In recent years, the design of time-aware business processes has seen advancements mostly in the theory. Various modeling constructs and techniques to verify the temporal qualities of business processes have been formalized. However, operational support for process designers still lags behind, with only a few tools offering limited time-related functionalities available. Here, we contribute towards closing this gap and propose a modeling tool based on the Camunda modeler. Our tool features an interactive interface that aids the design of time-aware process models with guarantees of temporal correctness in terms of dynamic controllability.

Keywords: Process modeling tools · Time-aware processes · Dynamic controllability

1 Introduction

The design of a business process model requires defining elements from the five process modeling aspects - functional, behavioral, operational, informational, and organizational. However, research around time-aware processes in the last two decades has shown that the temporal aspect ought to be considered as an additional aspect to tackle in the modeling phase [2].

The temporal aspect includes the definition of properties such as durations for tasks, temporal constraints, deadlines, recurring events, etc. as pointed out by extensive prior work (for a comprehensive overview, see the time patterns in [7]). Modeling the temporal aspect, however, makes process design significantly more challenging. This is due to the fact that the temporal aspect may introduce potential conflicts in a process model, which are difficult to recognize and identify for human designers. Examples of such conflicts are temporal constraints between events that cannot be satisfied at the same time or temporal constraints that may be satisfied only for certain task durations that cannot be controlled.

Checking for the absence of conflicting temporal constraints, resp. the possibility of executing a process without violating temporal constraints independent of uncontrollable durations led to the formulation of properties, the most notable

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 339–345, 2023.
https://doi.org/10.1007/978-3-031-26886-1_22

of which are satisfiability and dynamic controllability (DC) [1]. In a nutshell, a process model is satisfiable if it admits at least one trace satisfying all temporal constraints; it is dynamically controllable if, for any observed value for uncontrollable durations and conditions at xor-splits, the executor can dynamically steer the execution satisfying all constraints. Dynamic controllability is considered a highly desirable property since it allows for high flexibility yet with strong guarantees of avoidance of violations.

Typically, DC-checking procedures are applied to fully specified process models at the end of the design phase. However, it would be beneficial for process designers to receive guidance *during* the process design phase for reaching process models that are dynamically controllable, rather than finding out at the end of the design phase whether their processes are dynamically controllable. With such guidance, process designers may enjoy the additional benefit of receiving information about which are the admissible values they may set for the various temporal elements, e.g., which is the maximal allowed bound for a constraint stating the maximum time to elapse between two events that guarantees DC.

To the best of our knowledge, only TimeAwareBPMN-js [8] provides some kind of support during the design phase. However, in TimeAwareBPMN-js it is the responsibility of the designer to decide when to perform a DC-check to receive information regarding the DC property and get informed where conflicts may exist and how they may be fixed. Additionally, TimeAwareBPMN-js does not allow defining temporal parameters, which model external events [3].

Here, we propose TemporalBPMN-ID, an interactive process designer tool, which allows designers to model a business process in a BPMN-like language extended with constructs for the temporal aspect[1]. The tool proactively provides designers information about the DC property of the process being modeled. The tool provides also information about the admissible values for newly added temporal constraints before designers assign them a value. With this proactive approach, designers can focus solely on the modeling of processes, reducing the challenges induced by the temporal aspect. The proposed tool has a microservice architecture, which enables modularization, scalability, and extensibility.

2 Time-Aware Business Processes

We consider business processes to be time-aware when they include elements such as events, activity durations, and temporal constraints between events.

Events may correspond to the start and the end times for control flow elements (e.g., activities, gateways), as well as to time points that do not refer to control flow elements. Events not bound to control flow elements can be encoded through temporal variables, which can be exchanged between local processes in the form of temporal parameters [5]. Our tool supports the definition of events related to control flow elements as well as temporal parameters.

[1] The tool homepage is http://isys.uni-klu.ac.at/pubserv/tools/TemporalBPMN-ID/. It also contains a screen cast demonstrating the tool functionalities.

Traditionally, a distinction is made between contingent and non-contingent durations, i.e., durations that cannot be controlled but only observed (e.g., a bank money transfer, guaranteed to take between 1 and 4 days) versus durations that can be controlled and decided by the process controller (e.g., writing a document) [1]. In our tool, we support the definition of both types of durations.

Temporal constraints restrict the time of occurrence of pairs of events relative to each other. Such a restriction may be either an upper-bound, specifying a maximum allowed time distance, or a lower-bound, specifying a minimum allowed time distance [4]. In our tool, we support the definition of both upper- and lower-bound constraints between events.

Dynamic controllability is among the most relevant properties for time-aware business processes. A time-aware business process is dynamically controllable if, for every possible setting of contingent durations and uncontrollable conditions at xor-splits, the controller can set the values for all subsequent non-contingent events in a way that all temporal constraints can be satisfied. DC is considered the most relaxed notion of temporal correctness that offers guarantees of no temporal constraint violations. Our tool supports the check for dynamic controllability. In order to perform a DC-check of a process model, the tool adopts the established practice of mapping a process model into a temporally-equivalent temporal constraint network such as the Simple Temporal Network with Uncertainty (STNU) or the Conditional STNU (CSTNU) (to represent processes with no conditional executions, resp. processes including conditional executions) and execute a constraint propagation procedure on it, which returns `true` if the network is dynamically controllable, `false` otherwise [1].

3 System Overview

TemporalBPMN-ID is an extension of the open source tool Camunda modeler (https://camunda.com) for modeling BPMN process diagrams. The extension for defining events, temporal parameters, durations, and temporal constraints is realized by annotating the BPMN source. The system has a client-server architecture: the client provides the user interface for defining time-aware process models; the server takes care of executing a DC-check for the model being defined, based on temporal constraint networks. The DC-check is performed whenever a significant change in the process model is detected. Significant changes are changes in the model that determine a change in the temporal aspect, i.e.:

- Adding or modifying a control flow element duration
- Removing a control flow element
- Adding or removing an event
- Adding, removing, or modifying a temporal constraint

The communication between client and server is REST-based. Whenever a significant change in the process model is detected, it is communicated from the client to the server. With the communication of significant changes, the server maintains an up-to-date temporal constraint network that encodes the temporal

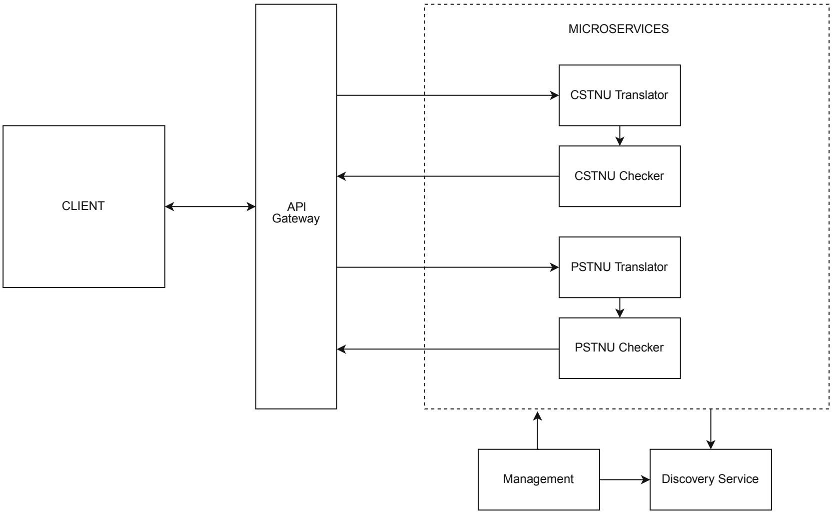


Fig. 1. Architectural diagram of the implemented system.

aspect of the process model by applying the mapping rules for mapping into an STNU with temporal parameters (PSTNU, [3]) or into a CSTNU (e.g., [6]).

For the server, the implementation follows a microservice architecture, which allows for decoupling the different components and features. This allows plugging in different checking algorithms based on different data structures, e.g., PSTNU or CSTNU, which make use of different checking procedures, to be able to formally encode various temporal elements for temporal reasoning.

Figure 1 shows the architectural diagram for the implemented system.

4 Use Case

Here, we give an overview of a use case of modeling the following time-aware business process from the medical domain. The process starts with a blood sampling (taking 2 to 5 min). Then, in parallel, an MRI (Magnetic Resonance Imaging) scan is performed (20 to 45 min) and blood analyses are carried out (30 to 45 min). Finally, a diagnosis report is filled out (3 to 10 min). Usage of the MRI instrumentation must be notified at least 10 min in advance. The whole process, including inter-task delays, must take 90 min at most. All activities are contingent, and gateway executions take no time.

The steps taken by a process designer using our tool are the following:

1. First, the designer adds a pool modeling the clinic.
2. The designer adds the process start and end events to the pool. Each event introduction triggers a communication from the client to the server, requesting to create a new PSTNU node corresponding to the created event.

3. The designer adds various gateways and activities, and specifies the name, duration type, and duration range for the activities.
4. The designer connects the start event to the first activity with a control flow edge. Now that the two elements are connected, the tool detects that the time-aware process is in a new meaningful state and sends a new command to the server. This creates two new PSTNU nodes per activity (one for the start and one for the end of it), a contingent link between each pair of such nodes (representing the duration bounds), and a regular edge connecting the node for the process start event to the node for the first activity start event.
5. The designer adds an upper-bound constraint from the process start event to the process end event with bound 90 to model the requirement of the maximum allowed duration. A command is sent to the server to add a corresponding PSTNU edge between the nodes for the start and end of the process. Afterwards, the server starts the execution of a full constraint propagation procedure, which realizes a DC-check for the process. The procedure returns `true`, signaling that the process is DC, and such a result, along with the derived edges, is sent to the client, which displays *The process is DC*.
6. The designer adds the remaining control flow connectors, each time triggering a communication and a DC-check like above. The process is still DC.
7. The designer adds an upper-bound constraint edge from the process start event to the start of the first activity. Now, due to the received derived edges after the full constraint propagation, the client knows that a derived edge exists between these two events. The corresponding value of 30 is shown automatically to the designer before she starts typing her intended constraint bound. The shown value is the maximum allowed value for such a constraint, i.e., any value larger than 30 violates the DC property of the process. It is now up to the designer to choose whether to keep such a value or to specify a different (smaller) one, in which case a new DC-check would be performed.
8. The designer adds a parameter node P to represent the time of notification of MRI instrumentation use since this is not a process activity. Then, she adds a lower-bound constraint from P to the start event of task M, with a bound of 10. This triggers the addition, to the PSTNU, of a new node and an edge connecting it to the node for the start of M. The server updates the PSTNU and executes a new full constraint propagation to compute potential new implicit constraints, returning `true`.
9. The designer adds an upper-bound constraint edge from process start to P. Again, the constraint is implicitly known from the propagation at the server, and the corresponding bound of 25 is presented to the designer, who accepts the value for the constraint. This value instructs process stakeholders on which is the maximum admissible value w.r.t. the process start for the MRI use notification, given all the other temporal requirements in the model.

Figure 2 shows the view presented to the designer at the end of the above steps.

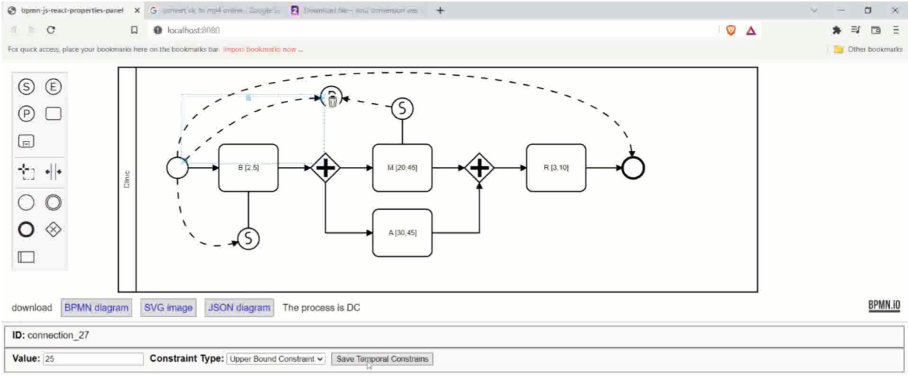


Fig. 2. Screenshot of the tool view after defining the process in the use case.

5 Conclusion

In this paper, we have introduced a novel tool for the interactive design of time-aware business processes. The tool extends the functionalities of the popular tool Camunda modeler with support for modeling the temporal aspect of processes. Additionally, the tool automatically performs an online DC-check on the process model during its design, providing the designer guidance on how to complete the temporal aspect definition. With the proposed tool, we provide process designers with a new means to define time-aware business process models that offer dynamic controllability guarantees. The microservice architecture enables further extensions with new algorithms to be used depending on the meta-model employed for time-aware processes and the temporal qualities to be checked.

References

1. Combi, C., Posenato, R.: Towards temporal controllabilities for workflow schemata. In: 2010 17th International Symposium on Temporal Representation and Reasoning (TIME), pp. 129–136. IEEE (2010)
2. Combi, C., Pozzi, G.: Temporal conceptual modelling of workflows. In: Song, I.-Y., Liddle, S.W., Ling, T.-W., Scheuermann, P. (eds.) ER 2003. LNCS, vol. 2813, pp. 59–76. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-39648-2_8
3. Eder, J., Franceschetti, M., Köpke, J.: Controllability of business processes with temporal variables. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, pp. 40–47. ACM (2019)
4. Eder, J., Panagos, E., Rabinovich, M.: Time constraints in workflow systems. In: Jarke, M., Oberweis, A. (eds.) CAiSE 1999. LNCS, vol. 1626, pp. 286–300. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48738-7_22
5. Franceschetti, M., Eder, J.: Determining temporal agreements in cross-organizational business processes. Inf. Comput. **281**, 104792 (2021)
6. Franceschetti, M., Eder, J.: Computing ranges for temporal parameters of composed web services. In: Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, pp. 537–545 (2019)

7. Lanz, A., Reichert, M., Weber, B.: Process time patterns: a formal foundation. Inf. Syst. **57**, 38–68 (2016)
8. Ocampo-Pineda, M., Posenato, R., Zerbato, F.: TimeAwareBPMN-js: an editor and temporal verification tool for time-aware BPMN processes. SoftwareX **17**, 100939 (2022)

Prosimos: Discovering and Simulating Business Processes with Differentiated Resources

Orlenys López-Pintado, Iryna Halenok, and Marlon Dumas(✉)

University of Tartu, Tartu, Estonia
{orlenyslp,iryna.halenok,marlon.dumas}@ut.ee

Abstract. Prosimos is an open-source tool that discovers business process simulation models from execution data (event logs) and that enables users to perform what-if analysis using the resulting models. Prosimos distinguishes itself from other data-driven business process simulation approaches in the way it models resources. Existing data-driven simulation approaches treat resources as undifferentiated entities, grouped into resource pools, and assume that all resources in a pool have the same performance and availability calendars. In contrast, Prosimos allows resources (within a pool) to have different performance and availability profiles. For example, instead of treating all claims officers in an insurance claims handling process as having the same performance and availability, Prosimos may capture scenarios where senior resources perform some tasks faster than junior ones, or scenarios where some resources work part-time. To this end, Prosimos integrates algorithms for discovering differentiated resource profiles from event logs.

1 Introduction

Business Process (BP) simulation engines allow users to predict how changes in a process can impact its performance [1]. A BP simulation engine takes as input a simulation model, which usually takes the form of a process model (e.g., a diagram in the Business Process Model and Notation – BPMN), enhanced with parameters such as resource availability, activity processing times, inter-arrival rates of new process cases and branching probabilities at decision gateways in the process model. Given such a model, a BP simulation engine produces an event log of a simulated run of the process, as well as aggregate performance metrics, enabling users to compare multiple what-if simulation scenarios.

Traditional simulation approaches treat resources in a process as undifferentiated entities [1,3,5]. These approaches group resources into disjoint resource pools, such that all resources in a pool share the same performance and availability. In these approaches, each activity is assigned to one resource pool. This undifferentiated resource modeling approach implies that the processing time of an activity does not depend on the resource that performs it. Similarly, the undifferentiated treatment of resources implies that all the resources in a pool are

© The Author(s) 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 346–352, 2023.
https://doi.org/10.1007/978-3-031-26886-1_23

available for work during the same periods. In practice, though, each (human) resource exhibits different performance and availability. The assumption that all resources in a pool behave in the same way, introduces approximations that ultimately have an impact on simulation accuracy [1,2,5].

In this demonstration paper, we introduce Prosimos: an open-source simulation tool that implements an approach to BP simulation with differentiated resources. In Prosimos, resources are not grouped into pools but treated as individuals, each with its own resource profile. In particular, the performance of each resource is independent of that of other resources (differentiated performance), and each resource may have its independent availability calendar (differentiated availability). Unlike classic BP simulation models, an activity in a process model may be assigned to multiple resource profiles, and multiple resources can share the same resource profile. The latter also allows us to answer questions like "what if a resource is replaced by another with lower performance?" or "what if a resource changes its availability from full-time to part-time?", not supported by models with undifferentiated resources.

A simulation run in Prosimos produces as output the event log of the simulation, as well as the following performance metrics: mean *waiting time* - the duration from the moment activity is enabled until it is started; mean *processing time* - the duration between the beginning and end of an activity instance; mean *cycle time* - the difference between the end time and start time of a process case; and *resource utilization* - the ratio of the available time of a resource spent executing process activities.

In addition to providing a simulation engine, Prosimos embeds a module to automatically discover simulation models (with differentiated resources) from event logs. This demo paper focuses on the tool architecture and functionality. The description of the underpinning algorithms (including the simulation model discovery algorithms) and other technical details are reported in [6].

2 Prosimos Architecture

Prosimos is a Web-based tool, logically structured into three layers, as shown in Fig. 1. The layer at the bottom, henceforth referred to as Prosimos backend, consists of three groups of components labeled as `SIMOD-DiFF`, `Prosimos Engine` and `Support Modules`. The `Support Modules` consists of a set of components containing supplementary functionalities shared by both the `SIMOD-DiFF` and `Prosimos Engine`. Among those, the `Simulation Model Parser` contains functions for transforming the input files, e.g., BPMN model, XES or CSV event logs, and JSON simulation parameters, into the data structures required by Prosimos. The `BPMN Replayer` implements the token game as specified in the BPMN standard, which serves to execute an event log over the process model, e.g., to estimate branching probabilities and to compute the process state during the simulation. The `Timetables Manager` is a module enclosing the calendar-based operations used for scheduling the inter-arrival time intervals and resource availability. Finally, the `Stochastic Estimator` provides a set of operations to determine and evaluate probability density functions.

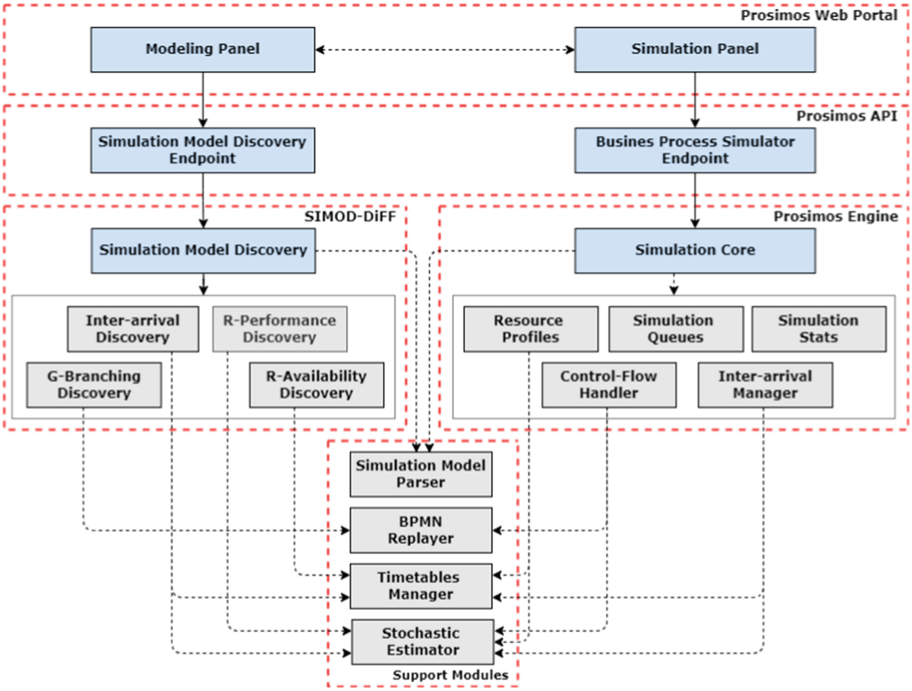


Fig. 1. Prosimos architecture

On the top-left of the back-end, the `SIMOD-DiFF` components discover the simulation parameters given a BPMN model and the corresponding event log written in XES or CSV, respectively. Precisely, the `Inter-arrival Discovery` module estimates the inter-arrival distribution functions adjusted to an arrival calendar, i.e., how often new cases are created and in which time intervals. The `G-Branching Discovery` calculates the branching probabilities once the execution flow arrives at a decision split gateway, i.e., by replaying the input event log over the process model. The `R-Performance Discovery` estimates the probability density functions modeling how long it would take for each resource to execute its allocated tasks. Finally, the `R-Availability Discovery` retrieves the calendars in which each resource is available to perform a task in the process. As a result, the `Simulation Model Discovery` joints and retrieves all the simulation parameters discovered from the event log into a JSON file.

Concluding with the back-end, the `Prosimos Engine` components handle the business process simulation from a given model. The `Simulation Core`, as the name suggests, is the spine of the simulation, engaging the operation of the remaining five modules to produce, from a simulation model received as input, the corresponding simulation log, and performance indicators as output. The `Resource Profiles` handle the differentiated resources, e.g., including resource allocation, performance (according to the `Stochastic Estimator`), and availability (interacting with the `Timetables Manager`). The `Control-Flow`

Table 1. Prosimos REST API

Verb	URI	Description
POST	/api/discover	Discovers the simulation parameters given a BPMN model and an event log
POST	/api/simulate	Performs the simulation from a given simulation model
GET	/api/results	Retrieves the event logs and metrics produced as result the simulation

`Handler` interacts with the `BPMN Replayer` and the `Stochastic Estimator` to compute the state of each simulated process case, i.e., the activities enabled at each moment of the simulation, and to decide which path to follow at each split decision gateway. The `Inter-arrival Manager` uses the inter-arrival distribution functions (`Stochastic Estimator`) and arrival calendar (`Timetables Manager`) to create new process instances to simulate. Prosimos uses a priority multi-queue data structure, in `Simulation Queues`, that handles shared activities and sorts resources according to an allocation input function, i.e., according to resource availability as a default. Finally, the `Simulation Stats` computes the performance indicators of the simulation.

The two layers on top of the architecture, i.e., the Prosimos front-end, consist of the `Prosimos API` and `Prosimos Web Portal`. The Prosimos REST API, in the middle of the architecture, provides three endpoints, grouped into the `Simulation Model Discovery Endpoint` and `Business Process Simulator Endpoint`. Table 1 describes the verbs, URIs, and actions of those endpoints. Although the REST API can span more specific endpoints, e.g., for interacting with the back-end to trigger the `BPMN replayer` or discover calendars for a given resource, for simplicity, we kept the API with the minimum operations required to discover simulation models and to run simulations. On top of the architecture, the `Modeling Panel` provides a web interface for end-users to create, modify or discover (interacting with the Prosimos API) simulation models from event logs. On the right, the `Simulation Panel` allows users to run simulations and retrieve the resulting event logs and performance metrics.

3 Source Code, Evaluation and Demonstration Screencast

End users can discover and simulate business processes with differentiated resources via the software-as-a-service deployment of Prosimos at https://prosimos.cloud.ut.ee. A sample process model and an event log are available at https://shorturl.at/abrsO. The Prosimos Web application can be deployed via a Docker container by using the scripts available at https://github.com/AutomatedProcessImprovement/prosimos-docker. This code repository also includes links to the Prosimos Web Portal and REST API code repositories. The source code of Prosimos back-end can be downloaded from https://github.com/AutomatedProcessImprovement/Prosimos. The code distribution provides the `SIMOD-DiFF`

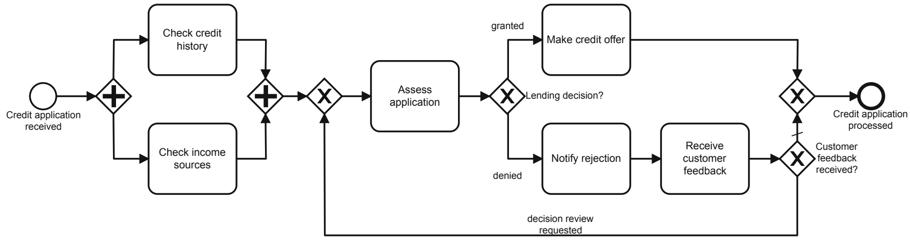


Fig. 2. Simplified credit application process model

components described in the architecture in the folder `bpdfr_discovery`. The `Prosimos engine` and `Support Modules` are in the folder `bpdfr_simulation_engine`.

Prosimos is in its first release and supports the following elements in the standard BPMN 2.0: default start and end events, tasks, inclusive, exclusive, and parallel gateways. Specifically, in the case of the exclusive gateway, Prosimos implements the complete OR join semantic as prescribed by the BPMN 2.0 standard. The selection of those elements in the first release aligns with the output of the existing approaches to discovering BPMN models from event logs. The latter is not a significant limitation as data about other advanced BPMN elements, e.g., throwing/interrupting events, event-based gateway, messages, data objects, and multi-instance attributes, are usually not present in event logs. Thus, they are typically not present in process models discovered automatically from data. Still, we are working to introduce most of those advanced BPMN elements in the next release of the Prosimos simulation engine so that process analysts can simulate processes with a broader spectrum of BPMN. The last updates about the models supported by Prosimos can be accessed from https://shorturl.at/dgnsv.

Prosimos implements the simulation approach with differentiated resources presented and evaluated in the research paper [6]. We empirically assessed the discovered models by simulating them using PROSIMOS and measuring the distance between the simulated logs and the original ones. Specifically, in line with [4], we compare simulated and original event logs by extracting temporal histograms from each event log and computing the Earth Movers' Distance (EMD) between these histograms. The experiments in [6] show that simulation models with differentiated resources produce simulated event logs closer to the original event logs when compared to equivalent models with undifferentiated resources. For further details about the evaluation, we refer the reader to check [6]. Additionally, the folder `testing_scripts` in the Prosimos back-end repository includes all the scripts to run the experiments presented to assess the approach. Finally, some components of Prosimos, like the `BPMN replayer` and the `Timetables Manager`, have been integrated and tested into Simod [3], which is a well-known tool in the BPM community to discover undifferentiated models.

For the demo, we will use the simplified model of a credit application process shown in Fig. 2. Two files must be provided to simulate the process: (i) the process model specified in the BPMN standard and (ii) the simulation parameters written in JSON format. Alternatively, the JSON file with the simulation parameters can be discovered from an event log written in XES or CSV format.

A screencast of Prosimos can be found at https://shorturl.at/FNPS9. Specifically, the demo starts with a case in which the simulation parameters are discovered from an event log in XES format, corresponding to the process in Fig. 2. Prosimos includes web templates to manually create and update the simulation parameters, which are transformed later into the required JSON format. So, once provided the BPMN model and selected the option to specify the simulation parameters, i.e., discovered from a log, loaded from an existing JSON, or created manually from scratch, Prosimos redirects to a view to update the parameters or to start the simulation. We refer the readers to check the Prosimos back-end repository for a more detailed explanation of those simulation parameters, i.e., represented by the tabs: case creation, resource calendars, resources, resource allocation, and branching probabilities.

After running the simulation, Prosimos displays the view simulation results with statistics like the cycle, processing and waiting times, and resource utilization. Business process analysts can use these statistics to detect inefficiencies and decide how to improve their business processes. For example, in the scenario illustrated by the screencast, we can observe the task *Assess Aplication* in Fig. 2 may be a bottleneck due to a very high waiting time. We can also observe (from the discovered simulation parameters) that a single resource executes this problematic task, i.e., the *Credit_Officer_1*, which also works part-time (from 13:00–17:00 on working days). The statistics also show this resource has a very high resource utilization, indicating that the performance of the process may be affected by a resource contention issue. A possible solution would be to add some extra resources or extend the working times of the existing one. Then, after applying the change, the analyst can run a new simulation and quantitatively assess the impact. Additionally, Prosimos produce the simulated event log, which can serve to perform automated analysis and optimization of the process.

The required files to reproduce the demo can be found in the Prosimos docker repository under folder `demo_example` or at https://shorturl.at/abrsO. Each of the above-mentioned code repositories provide instructions to install the required dependencies locally.

Acknowledgment. Work funded by European Research Council (PIX project). Iryna Halenok is also supported by the Estonian Ministry of Foreign Affairs - Development Cooperation and Humanitarian Aid funds.

References

1. Van der Aalst, W.M.P.: Business process simulation survival guide. In: Handbook on Business Process Management 1, 2nd Ed, pp. 337–370 (2015)

2. Afifi, N., Awad, A., Abdelsalam, H.M.: RBPSim: A resource-aware extension of BPSim using workflow resource patterns. In: Proceedings of ZEUS, pp. 32–39 (2018)
3. Camargo, M., Dumas, M., González, O.: Automated discovery of business process simulation models from event logs. Decis. Support Syst. **134**, 113284 (2020)
4. Camargo, M., Dumas, M., Rojas, O.G.: Learning accurate business process simulation models from event logs via automated process discovery and deep learning. In: Franch, X., Poels, G., Gailly, F., Snoeck, M. (eds) Advanced Information Systems Engineering. CAiSE 2022. Lecture Notes in Computer Science, vol. 13295, pp. 55–71. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-07472-1_4
5. Pereira, J.L., Freitas, A.P.: Simulation of BPMN process models: current BPM tools capabilities. In: New Advances in Information Systems and Technologies. AISC, vol. 444, pp. 557–566. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-31232-3_52
6. López-Pintado, O., Dumas, M.: Business process simulation with differentiated resources: Does it make a difference? In: Di Ciccio, C., Dijkman, R., del Rio Ortega, A., Rinderle-Ma, S. (eds) Business Process Management. BPM 2022. Lecture Notes in Computer Science, vol 13420, pp. 361–378. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16103-2_24

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

Doctoral Consortium

Doctoral Consortium

Preface

The Doctoral Consortium (DC) of EDOC is a forum of exchange organized to encourage PhD students to present their early work and exchange with other researchers in their fields. Senior researchers provide feedback and advice on the work and give advice on managing research projects. The DC is also a place to establish a social network with peers in the field of the conference.

For the DC of EDOC 2022, we accepted three submissions for presentation and discussion at the DC session of the conference. Each of the submissions was reviewed by at least three members of the Program Committee and was assigned a senior researcher as tutor to lead the discussion on the research proposal at the DC session. We would like to thank the tutors Lamiae Benhayoun, Dominik Bork, and Paolo Ceravolo for their active participation in the DC session.

In addition to the presentations and discussions, we invited Dr. Claudenir Fonseca to give a brief keynote presentation on the lessons he learned on his PhD journey. It has been an inspiring talk, offering an enjoyable though profound view of the PhD experience, for the PhD students who attended the Doctoral Consortium.

October 2022

Felix Mannhardt
Chiara Di Francescomarino

Organization

Doctoral Consortium Chairs

Felix Mannhardt — Eindhoven University of Technology, The Netherlands
Chiara Di Francescomarino — Fondazione Bruno Kessler, Italy

Advisory Committee

Colin Atkinson — University of Mannheim, Germany
Georg Grossmann — University of South Australia, Australia
Marten van Sinderen — University of Twente, The Netherlands

Program Committee

Andrea Burattin — Technical University of Denmark, Denmark
Andrea Marrella — Sapienza University of Rome, Italy
Claudio Di Ciccio — Sapienza University of Rome, Italy
Dominik Bork — TU Wien, Austria
Fethi Rabhi — The University of New South Wales, Australia
Lamiae Benhayoun — Rabat Business School, Morocco
Luise Pufahl — TU Berlin, Germany
Manuele Kirsch Pinheiro — Paris 1 Panthéon – Sorbonne University, France
Paolo Ceravolo — University of Milan, Italy
Ronny Seiger — University of St. Gallen, Switzerland
Said Assar — Institut Mines-Télécom Business School, France

Data Analytics and Machine Learning for Smart Decision Making in Automotive Sector

Hamid Ahaggach(✉)

LIB Laboratory, University of Burgundy, Dijon, France
Hamid.ahaggach@u-bourgogne.fr

Abstract. The objective of this thesis is to conduct scientific research on the use of data science and artificial intelligence techniques in the practices of automotive dealership companies to assist them in their decision-making processes and to use data-driven methods with modeling approaches for computing these enterprises. By proposing algorithms capable of continuously extracting relevant information from a diverse and multi-structured automotive environment. Due to the large amount of data available within these companies, we will develop algorithms to correctly assess the situation, suggest recommendations for decision-making, develop marketing strategies, and automate manual tasks that cost time, effort, and money.

Keywords: Data science · Artificial Intelligence · Decision-making · Marketing strategies

1 Context and Goals

In our research, we mainly focus on two key aspects of decision support for the automotive sector. First, we consider car commercialization. Here, the aim is to provide an aid to tackle the inventory problems of car dealers. Indeed, car dealerships buy cars from manufacturers and sell them to their customers to make a profit. Dealers cannot just send the unsold cars back to the manufacturer. Keeping unsold cars in the parking lots for a long time is extremely costly and may even threaten the financial prosperity of these companies. Therefore, they will have to find a way to get rid of these cars and get prepared to receive newer models. We propose to use data analytics (DA) and machine learning (ML) to predict the time required to sell car models. Characteristics of cars and sales history are taken into account to make ML models that are capable of making correct predictions in most cases.

The second aspect we consider is damage assessment. Car dealers import vehicles from the manufacturer's lot by ship, truck, or train. On arrival, all vehicles receive a detailed quality control to analyze damage incurred during transport and are stored in the car park. We aim to replace expert inspection, which costs time, effort, and money, with a fast and reliable automatic protocol.

The first work we have conducted in this track is the use of an ontology (OWL) to model the different damage caused to the car, according to the size and type of damage

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 357–363, 2023.
https://doi.org/10.1007/978-3-031-26886-1_24

(Dents, scratches, paint damage, etc.) and the type of vehicle which the damage is to be evaluated. The ontology is built based on the knowledge of insurance experts and a large amount of data in the form of reports that they fill out. This work is the beginning of a project to assess the damage to the car and determine the price required to repair it, based on images captured by high-resolution cameras of vehicles from all angles using image-mining and ML techniques, or based on the available textual data that describes the damage using Natural Language Processing (NLP) techniques and Named-entity recognition (NER).

2 Predicting Car Sale Time

This section presents the problem, related work, suggested solutions, and preliminary results to assist car dealers in their efforts to solve vehicle inventory problems.

2.1 Related Work and Problematic

Due to a lack of available public datasets, there is few research literature on vehicle sales prediction. Most existing works focus on sentiment analysis and the impact of economic factors on vehicle sales, without a comprehensive analysis of vehicle-related attributes. Pai, P. and Liu, C. [1] put forward a model for the prediction of vehicle sales by sentiment analysis of Twitter data and stock market values using least squares support vector regression.

Wijnhoven, F., and Plant, O. [2] test the predictive power of car sales by the ratio of positive to negative tweets, the total number of mentions, Google trends, and the percentage of negative comments. They report that social media sentiments have relatively very weak salience to improve predictions of car sales. Gao, J. et al. [3] proposed a hybrid optimization approach to forecast automobile sales in China by using gross domestic product, consumer price index, highway mileage, and automobile ownership. Wang, F. K., et al. [4] proposed a model based on monthly sales in Taiwan to select the most influential economic indicators such as oil price, current automobile sales, and exchange rate. However, these models only adopt economic indicators to predict nationwide sales. It is not enough to predict the car sales of different brands only based on regional economic indicators.

There are also many works on prediction models based on time-series data in various attributes such as commodity sales forecasting [6] financial market forecasting [7, 12, 13], weather and environmental state prediction [5, 14, 15]. But in our case time series cannot be used as a tool to predict car sales in the coming months, because the sales information, in our dataset, do not follow a precise pattern. For example, if we take a car C which rarely sells, where $\{c_1, c_2 \ldots, c_t\}$ are the monthly sale values, then most of the values are equal to zero. Given the sale values $\{c_s \ldots, c_e\}$ over a period [s, e], where c_s and c_e are respectively the sale value of the start and the end of the period, if we train a network based on sequential models like (*LSTM*, *RNN*...), or based on a statistical analysis model like *ARIMA* to predict $Y_{t+1} = \mathrm{F}\ (c_s \ldots, c_e)$, the sale value for the next period $t + 1$, , then we will get wrong results. Therefore, the research questions that can be asked are: (a) How can we forecast sales using the available data? (b) Which features are important to use in our model? (c) How can we use our model to develop better marketing strategies?

2.2 Methodology

The proposed solution is to use car characteristics to predict selling time. So we have a dataset of pairs D $= \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $X = \{x_1, x_2..., x_n\}$ contains the characteristics of cars such as the color, the price, the power of the engine, also the entry date of the vehicle, because we are using historical data and trying to make predictions to decrease the time to market, so time is a central variable in our model. While $Y = \{y_1, y_2..., y_n\}$ is the time taken to sell the vehicles which can be numerical or categorical. We need to find function $f(x_i) = y_i$ that will be able to predict the time needed to sell a car, this function can be any ML algorithm. The proposed methodology operates in three steps: (i) Data processing: This step is very important and sensitive because it directly affects the results. In our case, we noticed existing of missing data in different attributes, we treat the problem in different ways (Mode, Mean, Regression…) according to the type of attributes and their influence on car sale time via data correlation. In addition, some data elements are anomalous due to recording errors and must be filtered out. We also noticed the presence of data of the same type, but written in different formats data integration step is therefore required to address these kinds of problems. (ii) Dimensionality reduction: We need to select intrinsic features to achieve high prediction accuracy and reduce computation costs. We also need it to visualize data in a reduced space. *Forward selection* gives us a good result compared to other dimension reduction methods. This method keeps only the most important features in a dataset and eliminates the rest; In this case, the features are not transformed. While we used *Principal Component Analysis (PCA)* to visualize the vehicles in 2D space. (iii) Model training: We use "supervised" classification algorithms. We have compared several ML algorithms, but only the below 4 used in this article give a reasonable result. *(1) Support Vector Machines (SVM)* is the best-known form of kernel methods statistical theory of learning. This method searches for the hyperplane that separates samples of each class ensuring that the margin between the closest classes is maximal. *(2) Decision trees (DT)* depend on choosing which attributes to use first to build the tree. *(3) Random forests (RF)* operate by constructing a multitude of decision trees. The output of the RF is the class selected by most trees. *(4) K-Nearest Neighbors (KNN)* stores all available data and classifies new data based on similarity to its neighbors.

2.3 Experimental Results

To test the 4 algorithms and compare their prediction results, we use a large-scale dataset provided by two car dealership companies covering their car sales activities for the period between Oct. 2013 and Nov. 2021. The dataset has 33 attributes and more than 73200 entries. The dataset of the first company contains 40700 among these cars there are 18800 new cars and 21900 used cars, and for the second company, there are in total of 32500 cars, among which there are 18700 new cars and 14000 used cars. We have labeled our data as follows: 0, if selling the vehicle takes less than 3 months. 1, if selling the vehicle takes between 3 and 6 months. 2, if selling the vehicle takes between 6 and

8 months. 3, if selling the vehicle takes between 8 and 12 months. 4, if selling the vehicle takes more than 12 months. Our goal is to predict the time margin that a car will stay in stock before being sold, we will build two models for each company, one model for used cars and the other for new cars. The dataset is randomly split into two parts: the training set (80% of the dataset), is used to train and the test set (20% of the dataset) to evaluate our model using 10-Folds cross-validation. The accuracy and training time are considered as comparison criteria between algorithms on the test set. The accuracy in our case is defined as follows:

Accuracy = The number of well-classified cars / The total number of cars (1).

Table 1. Results of the prediction on the datasets of the two companies.

Company	Vehicle type	Metrics	Models			
			KNN	*SVM*	*DT*	*RF*
1	VN	Accuracy	0.971	0.951	**0.990**	**0.990**
		Training time	**0.096 secs**	8.431 secs	0.996 secs	0.140 secs
	VO	Accuracy	0.849	0.854	0.814	**0.863**
		Training time	0.058 secs	13.52 secs	**0.057 secs**	0.647 secs
2	VN	Accuracy	0.967	0.944	0.987389	**0.994**
		Training time	**0.079 secs**	20.76 secs	0.140 secs	1.145 secs
	VO	Accuracy	0.845	0.862	0.802	**0.870**
		Training time	**0.066 secs**	47.15 secs	0.187 secs	1.484 secs

Table 1 shows that the RF gives much better results in comparison with other models, and this is because RF is composed of several DT that collaborate. In the case of the first company, both the DT and RF give the same accuracy score on a new vehicle because the data in this case are easy to be discriminated by the decision tree. We also note that KNN generally gives good results because it is based on data. SVM takes a lot of time to learn in comparison with other models because generally maximization problems take more time and depend on the performance of the machine used to train the model.

2.4 Discussion and Perspectives

In this paper, we proposed to implement SVM, DT, KNN, and RF to predict the time required for dealers to sell cars. A large-scale car sales dataset provided by two multi-maker dealership companies has been pre-processed to complete missing data and identify the car characteristics that have the greatest impact on car sales. These predictions give companies better ideas about the commercialization of vehicles by providing the characteristics of the car to be purchased and the model will answer the time needed to sell it. This hence helps them to put the right marketing strategy to avoid buying cars that are not easy to sell. In future work, we intend to extend this work by using deep learning techniques because of the large amount of data and the number of characteristics available that could allow the application of such techniques and obtain good results. We also intend to use customer behavior analyses to build a recommendation system based on association rules, to target customers who can buy specific cars based on the profile of former customers.

3 Damage Assessment

Damage assessment in general and damage assessment for cars, in particular, are difficult tasks because there are no specific criteria to assess the damage. We aim to model car damage using an ontology, due to the lack of previous work in the field of cars. Our ontology is based on the insurance experts' knowledge and their description reports. We model the description of all damages through define all concepts, data properties, and object properties and also we take into account the type of car, type and severity of the damage, as well as the part of the car damaged. So that it is evaluated with the same price despite the different experts. This work is the beginning of a project to damage assessment using 9500 images of cars captured by high-resolution cameras through image mining techniques and ML. First, We will use semantic segmentation to identify automotive parts then we will recognize the damaged car parts and model the damage using the proposed ontology and finally estimate the price according to the price database. In the case of using text data of 23000 damage cases. We will use NLP techniques and NER based on rules or based on learning to extract the entities and the relationship between these entities and extract the useful information in the form of ontology to estimate the price.

3.1 Related Work

Works that attempt to assess damage in the automotive field typically do not use ontologies, but rather employ a simple damage assessment (small, medium, major). For this reason, we cannot estimate the exact price necessary to repair the damage. They also do not take into account the location of the damage and the type of vehicle to be assessed, among these works we mention Waqas, U. et al. [8] who have proposed an image-based method of processing automobile insurance. In this regard, they consider the classification of the vehicle damage problem, where the classes include average damage, huge damage, and no damage, based on deep learning techniques, a MobileNet model with transfer learning for classification is proposed. Kyu, P. M., and Woraratpanya, K. [9] implement deep learning algorithms, VGG16 and VGG19, to detect and evaluate vehicle damage based on real-world datasets. Algorithms detect the damaged part of the vehicle and evaluate it by location and then its severity. Initially, CNN models are trained on the ImageNet dataset, followed by fine-tuning, because some of the classes can be very precise to accomplish tasks specified. The learning is then transferred into pre-trained VGG models and some techniques are used to improve the accuracy of the system. However, the severity is only evaluated on three levels (minor, moderate, and severe). Bandi, H. et al. [10] also perform the same work of assessing the extent of car damage using accurate convolutional neural networks, through a high-quality dataset that includes pivotal parameters such as location information and repair costs, but the assessment remains weak because the damage is not accurately described. While Singh, R.

et al. [11] propose a comprehensive system to automate the damage assessment process. This system takes images of the damaged vehicle as input and gives relevant information such as damaged parts and provides an estimate of the extent of damage (no damage, light or severe) for each part. This serves as an indication of the then-estimated repair cost that will be used in deciding the amount of an insurance claim. Popular instance segmentation models such as Mask R-CNN, PANet, and a combination of these two have been used along with transfer learning based on the VGG16 network to perform various tasks of identifying and detecting different classes of fragments and damages.

3.2 Expected Result and Perspectives

This research is a step towards replacing the costly manual damage assessment with the automatic assessment of the damage. The proposed OWL describes accurately the damage on the vehicle that consider the first step to estimate the cost of repair by analyzing the damaged car images and by analyzing the textual reports.

References

1. Pai, P.F., Liu, C.H.: Predicting vehicle sales by sentiment analysis of Twitter data and stock market values. IEEE Access **6**, 57655–57662 (2018)
2. Wijnhoven, F., Plant, O.: Sentiment analysis and Google trends data for predicting car sales (2017)
3. Gao, J., Xie, Y., Gu, F., Xiao, W., Hu, J., Yu, W.: A hybrid optimization approach to forecast automobile sales of China. Adv. Mech. Eng. **9**(8), 1687814017719422 (2017)
4. Wang, F.K., Chang, K.K., Tzeng, C.W.: Using adaptive network-based fuzzy inference system to forecast automobile sales. Expert Syst. Appl. **38**(8), 10587–10593 (2011)
5. Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., Cottrell, G.: A dual-stage attention-based recurrent neural network for time series prediction (2017). arXiv preprint arXiv:1704.02971
6. Zhao, K., Wang, C.: Sales forecast in E-commerce using convolutional neural network (2017). arXiv preprint arXiv:1708.07946
7. Elsworth, S., Güttel, S.: Time series forecasting using LSTM networks: A symbolic approach (2020). arXiv preprint arXiv:2003.05672
8. Waqas, U., Akram, N., Kim, S., Lee, D., Jeon, J.: Vehicle damage classification and fraudulent image detection including moiré effect using deep learning. In: 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2020)
9. Kyu, P.M., Woraratpanya, K.: Car damage detection and classification. In: Proceedings of the 11th international conference on advances in information technology, pp. 1–6 (2020 July)
10. Bandi, H., Joshi, S., Bhagat, S., Deshpande, A.: Assessing car damage with convolutional neural networks. In: 2021 International Conference on Communication information and Computing Technology (ICCICT), pp. 1–5. IEEE (2021 June)
11. Singh, R., Ayyar, M.P., Pavan, T.V.S., Gosain, S., Shah, R.R.: Automating car insurance claims using deep learning techniques. In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), pp. 199–207. IEEE (2019 September)
12. Hsu, M.W., Lessmann, S., Sung, M.C., Ma, T., Johnson, J.E.: Bridging the divide in financial market forecasting: machine learners vs. financial economists. Expert Syst. Appl. **61**, 215–234 (2016)
13. Wang, J., Wang, J., Fang, W., Niu, H.: Financial time series prediction using elman recurrent random neural networks. Computational intelligence and neuroscience (2016)

14. Sapankevych, N.I., Sankar, R.: Time series prediction using support vector machines: a survey. IEEE Comput. Intell. Mag. **4**(2), 24–38 (2009)
15. Koochakpour, K., Tarokh, M.J.: Sales budget forecasting and revision by adaptive network fuzzy base inference system and optimization methods. Journal of Computer & Robotics **9**(1), 25–38 (2016)

Check for updates

To Model or Not to Model? Assessing the Value of Ontology-Driven Conceptual Modeling

Isadora Valle Sousa(✉)

Conceptual and Cognitive Modeling Research Group (CORE),
Free University of Bozen-Bolzano, Bolzano, Italy
ivallesousa@unibz.it

Abstract. Modeling plays an important role in representing and supporting complex human design activities. For example, Ontology-Driven Conceptual Modeling (ODCM) creates concrete artifacts representing conceptualizations of particular domains. However, the development, management, and usage of these artifacts require investments of resources that should be worth it. Often, stakeholders neglected the trade-off analysis of the benefits and investments in ODCM experiences because of the lack of tools to assist them with this task. In this context, the aim of this research is to develop a method to identify when worth investing in ODCM experiences based on an analysis of value. To propose the method, we will develop and correlate knowledge and artifacts regarding the quality and value of the modeling process and product, the return on modeling effort, and domain debt.

Keywords: Return on modeling effort · Domain debt · Value-based analysis · Ontology-Driven Conceptual Modeling

1 Introduction

Modeling plays an important role in representing and supporting complex human design activities [16]. Ontology-Driven Conceptual Modeling (ODCM), for instance, creates concrete artifacts representing conceptualizations of particular domains that support the understanding and communication among stakeholders. More precisely, ODCM involves the use of ontological theories to develop engineering artifacts visioning the improvement of conceptual modeling [2]. For example, the development of new conceptual modeling languages, the improvement of existing languages by adding structuring rules, and the proposition of conceptual modeling patterns and anti-patterns [16]. The use of ODCM can lead to various system engineering benefits such as increased reusability and reliability, sophisticated representation of the domain being modeled, and enhanced domain understanding among its modelers and users [16].

Nevertheless, the development, management, and usage of these artifacts require investments of resources, such as money, time, and workforce. Often,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 364–369, 2023.
https://doi.org/10.1007/978-3-031-26886-1_25

stakeholders neglected the trade-off analysis of the benefits and investments in ODCM experiences because of the difficulties in identifying and quantifying the inputs and outputs involved in it. By modeling experience, we mean the modeling initiatives that involve agents, events, objects, and goals related to the creation, use, and transference of models. Therefore, the analysis of the benefits and investments in ODCM experiences goes beyond the model, it is also associated with the modeling language, the modeling goals, the model designers, and so on. Although some authors had identified the need to reason about this cost-benefit relationship [4,7] - called return on modeling effort (RoME) [11] -, very little work has been conducted to define and explain it.

Due to the involved investments, modeling initiatives should be done to solve specific problems and offering potential returns. Thus, before developing a method to assess ODCM experiences, we first need to reason about the value of these experiences. According to Sales et al. [13], "the value of a thing emerges from how well its affordances match the goals/needs of a given agent in a given context." Therefore, value is a composition of benefits, which emerge from goal satisfaction, and sacrifices, which emerge from goal dissatisfaction [13]. One way to analyze the benefits and sacrifices related to an ODCM experience is through a quality assessment of the models it produces. This can be done via an analysis of the quality dimensions attended (benefits) or not (sacrifices) by the model and its process of development. Despite its importance, the notion of quality in domain modeling is still immature [10], as well as its identification and evaluation.

One example of how quality assessment can be used to analyze the value and the RoME of ODCM experiences is through domain debt (DD). DD is a new notion proposed by Störrle and Ciolkowski [15] in 2019 that means "the misrepresentation of the application domain by an actual system". One reason for this flaw in the system in representing the domain can be the poor quality of the model. Domain debt caused by poor ODCM can require changes in the model that can affect other parts of the system, causing problems that are difficult and costly to solve. The efforts and investments made to solve these problems will directly affect the value and the RoME of the ODCM experience.

In this context, *the core objective of this research is to develop a method to identify when worth investing in ODCM experiences based on an analysis of value*. This new method can enhance enterprises' decision-making processes by offering means to stakeholders to better assess and manage their investments in modeling experiences. To propose the method, we will develop and correlate knowledge and artifacts regarding the value of the modeling process and product, the return on modeling effort, and domain debt.

2 Related Work

Return of Modeling Effort - Guizzardi and Proper identified the need to more explicitly determine the purpose for modeling as well as to reason about RoME [4]. They proposed a taxonomy of modeling-related goals to reason about the purpose for which a model may be created in the context of enterprises. To develop the method to measure RoME, this research will also be based on the methods published in the field of return on investment (ROI) in modeling initiatives. For instance, modeling and simulation [12], data modeling [5], and building information modeling [3].

Quality Evaluation of ODCM - Some frameworks available in the literature address quality in the process and product of modeling. Two fit better with the purpose of this study, the Semiotic Quality (SEQUAL) proposed by Krogstie [8] for the evaluation of the quality of conceptual data models, and the Conceptual Modeling Quality Framework (CMQF) proposed by Nelson et al. [10] in defining the quality attributes of enterprise architecture models. However, both frameworks should be adapted to be used in the quality evaluation of ODCM.

Domain Debt - The term domain debt (DD) was coined in 2019 by Störrle and Ciolkowski [15] to represent technical debts (TD) related to domain-oriented design. In his book [14], Sterling detailed explains software debt, its causes, quality impacts, and management. Kruchten et al. [9] also reasoned about TD and its practices pointing to the need for more tools and methods to identify and manage different types of technical debts. Alves et al. [1] identified thirteen types of technical debt in their ontology of terms on technical debt. However, none of them were related to domain modeling debt.

3 Research Questions

The main question this research aims to answer is *"When is it worth investing in ontology-driven conceptual modeling?"*. Three subjects will be analyzed to answer this question: Quality Evaluation of ODCM, RoME, and DD. The knowledge of each subject will be developed and addressed according to the sub-questions presented below.

- QR1: How do measure the return on modeling effort of ODCM experiences?
- QR2: How to evaluate quality in ontology-driven conceptual models to assess the value of ODCM experiences?
- QR3: How domain debt can be used to identify and quantify the value of ODCM experiences?

4 Research Methodology and Outputs

This research follows a Design Science Methodology [6] since it aims to create novel artifacts in the form of models and methods that will support people in addressing specific problems. The research tasks to be developed will follow the process detailed in Fig. 1. As shown in the figure, TSK 1.1 and TLS 1.2 have already been developed.

Our first step was to conduct a literature review to analyze what is available about RoME and how it can be related to return on investment (TSK 1.1). In

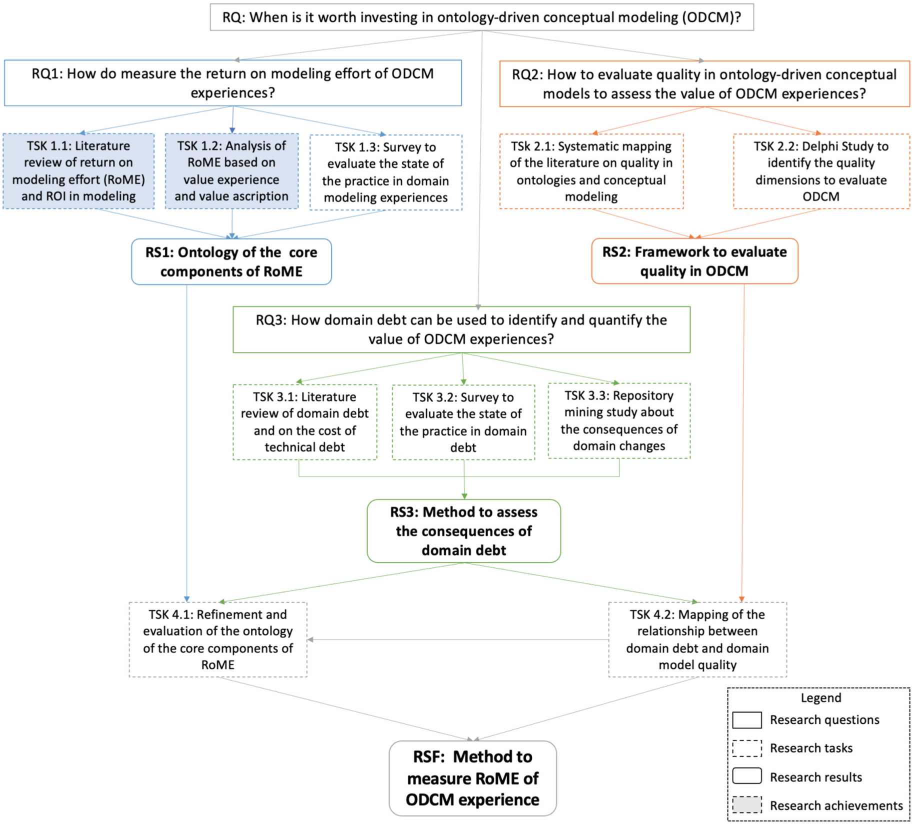


Fig. 1. Methodology detailing.

sequence, we developed an analysis of RoME based on the Common Ontology of Value and Risk (COVER) [13] to understand the value and value ascription of modeling experiences. Both analyses were combined with the taxonomy of modeling-related goals proposed by Guizzardi and Proper [4] to develop a study regarding the value, goals, and affordances that motivate modeling experiences (TSK 1.2). Our next step is to develop and apply an online survey to practitioners of conceptual modeling to analyze the state of the practice in domain modeling experiences (TSK 1.3). The results of the literature review, the value-based analysis, and the survey will generate the inputs to propose an ontology of the core components of RoME (RS1).

To understand how to evaluate quality in ontology-driven conceptual models we will develop a systematic mapping of the literature on quality in ontologies and conceptual modeling (TSK 2.1). In sequence, the plan is to develop a Delphi Study to get consensus among experts about the quality dimensions to evaluate ODCM (TSK 2.2). The quality dimensions to evaluate ODCM defined in the Delphi study will be used to adapt and complete the two existing frameworks: SEQUAL [8] and CMQF [10]. Then, a new framework will be proposed to assess the quality of the product and process of ODCM. The resulting framework (RS2) will be used as an input to study the relationship between domain debt and domain model quality.

We also plan to do a literature review of domain debt and the cost of technical debt (TSK 3.1). Then, we will develop and apply another online survey to people involved in projects that use domain models to analyze the state of the practice of domain debt (TSK 3.2). The last activity is to understand how domain debt affects the RoME of ODCM experiences. Therefore, we will develop an empirical study of repository mining to analyze the consequences of domain changes in real-world projects (TSK 3.3). The goal is to analyze the changes in an ontology-driven conceptual model related to domain debts and their impact on the RoME of the modeling experience. The development of these three tasks will result in a method to assess the consequences of DD (RS3).

Aiming to achieve the main objective of this research project, the final analysis will combine the outputs delivered throughout the project to propose a method for measuring RoME of ODCM experiences (RSF). To do so, first, we will correlate the framework on ODCM quality evaluation and the method to assess the consequences of domain debt to map the relationship between domain model quality and domain debt (TSK 4.1). In sequence, we will combine the knowledge gained about domain model quality and value, domain debt consequences and cost, and RoME's components to refine, evaluate, and proposed a method for measuring the RoME of ODCM experiences (TSK 4.2).

5 Expected Results and Evaluation

RS1: An Ontology of the Core Components of RoME - The ontology will be a representation of the RoME domain encompassing the modeling experience and the modeling value ascription. It will be developed based on theoretical and empirical studies and validated by domain modeling specialists. We also plan to specialize the ontology using real-world examples.

RS2: A Framework to Evaluate Quality in ODCM - The framework will be a tool one can use to evaluate the quality of ODCM according to specific quality dimensions. By correlation, it can also assess the value of ODCM. The quality framework and its dimensions will be evaluated and validated through a Delphi Study with experts in ODCM. After the validation, we will apply it as a tool to measure the value of an ODCM experience in a case study developed later in the research.

RS3: A Method to Assess the Consequence of Domain Debt - This method will help identify the consequences of changes in the domain representation due to problems or mistakes in its modeling. It will consider the artifacts that depend directly and indirectly on the domain entities to estimate the effort needed to repay a debt. The method will be evaluated through a repository mining study in which the changes in the domain representation of a project will be analyzed and categorized. The ones related to problems or mistakes in the

ODCM will be further investigated in a way that their consequences could be identified and, if possible, quantified.

RSF: A Method to Measure RoME of ODCM Experience - It will be the final and validated method to measure RoME of ODCM experiences; more than a formula, a complete value-based analysis method one can apply to identify when it is worth investing in ODCM experiences. This final method will be evaluated and validated through a case study in the same project used in the repository mining study. It will be a complete analysis of the investments, costs, and quality of an ODCM experience.

References

1. Alves, N.S., et al.: Towards an ontology of terms on technical debt. In: 2014 Sixth International Workshop on Managing Technical Debt, pp. 1–7. IEEE (2014)
2. Barcelos, P.P.F., et al.: A fair model catalog for ontology-driven conceptual modeling research. Conceptual Modeling. ER (2022)
3. Giel, B.K., Issa, R.R.: Return on investment analysis of using building information modeling in construction. J. Comput. Civ. Eng. **27**(5), 511–521 (2013)
4. Guizzardi, G., Proper, H.A.: On understanding the value of domain modeling. In: Proceedings of 15th International Workshop on Value Modelling and Business Ontologies (VMBO 2021) (2021)
5. Haughey, T.: The return on investment (roi) of data modeling, pp. 1–18. CA, Erwin, March 2010
6. Johannesson, P., Perjons, E.: An Introduction to Design Science. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10632-8
7. de Kinderen, S., Proper, H.A.: E3rome: a value-based approach for method bundling. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 1469–1471 (2013)
8. Krogstie, J.: Quality of conceptual data models. ICISO, pp. 165–174 (2013)
9. Kruchten, P., Nord, R.L., Ozkaya, I.: Technical debt: from metaphor to theory and practice. IEEE Softw. **29**(6), 18–21 (2012)
10. Nelson, H.J., Poels, G., Genero, M., Piattini, M.: A conceptual modeling quality framework. Softw. Qual. J. **20**(1), 201–228 (2012)
11. Op't Land, M., Proper, E., Waage, M., Cloo, J., Steghuis, C.: Enterprise Architecture: Creating Value by Informed Governance. Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85232-2
12. Oswalt, I., et al.: Calculating return on investment for us department of defense modeling and simulation. Technical report, DEFENSE ACQUISITION UNIV FT BELVOIR VA (2011)
13. Sales, T.P., Baião, F., Guizzardi, G., Almeida, J.P.A., Guarino, N., Mylopoulos, J.: The common ontology of value and risk. In: Trujillo, J.C., et al. (eds.) ER 2018. LNCS, vol. 11157, pp. 121–135. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00847-5_11
14. Sterling, C.: Managing Software Debt: Building for Inevitable Change (Adobe Reader). Addison-Wesley Professional, Boston (2010)
15. Störrle, H., Ciolkowski, M.: Stepping away from the lamppost: domain-level technical debt. In: 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 325–332 (2019)
16. Verdonck, M., et al.: Comparing traditional conceptual modeling with ontology-driven conceptual modeling: an empirical study. Inf. Syst. **81**, 92–103 (2019)

Check for updates

A Proposal for Intent-Based Configuration of ICT Components

Kaoutar Sadouki(✉)

CEDRIC, CNAM, Paris, France
Kaoutar.sadouki@lecnam.net

Abstract. My doctoral research proposal is a cross-field project between Information Systems (IS) architecture and Distributed Systems Computing applied to Industry 4.0 (I4.0) applications. It combines the configuration of Information and Communication Technologies (ICT) (e.g., smart IoT devices) like I4.0 components with a business delivery orientation. The scientific problem of this Ph.D. project is the lack of interconnectivity and alignment of I4.0 ICT to the business strategy and business intents. The notion of intent is very crucial regarding business requirements, as it allows all devices and the equipment under consideration to be connected to business and other needs, the notion has already gained traction in many fields, for instance, Intent-Based Networking (IBN) allows the configuration of the physical and virtual network infrastructure depending on business strategies requirements. To our knowledge, the components of I4.0 applications have not leveraged this important concept yet. Therefore, our objective is to encompass both business strategy and digital technologies deployed to support the I4.0 vision using an intentional perspective, to facilitate the use and configuration of I4.0 components through the automation of administration, and, flexible on-demand reconfiguration.

Keywords: Industry 4.0 · Information and Communication Technology · Intent-Based Approach

1 Introduction

The new industrial revolution, ICT-backed automated and interconnected Industry 4.0, powered by the Industrial Internet of Things (IIoT), Cyber-Physical Systems (CPS), and various sensors, provides interconnected manufacturing systems capable of communicating, analyzing, and driving further intelligent production in the physical world. I4.0 allows new opportunities for business development [1]. However, the successful adoption of these emerging technologies is not obvious, as it requires tools and skills to make the vision of I4.0 a reality.

To meet these challenges, organizations are rethinking their architecture to ensure that the enterprise will achieve the goals and intents defined and mitigate any threats from an adverse environment. Yet, there is a lack of alignment of I4.0 ICTs to business strategies and business intentions. In most cases, intents are neglected, whereas this is

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 370–375, 2023.
https://doi.org/10.1007/978-3-031-26886-1_26

a type of data that can clarify the requirements of business strategies while taking into consideration the internal and external users of I4.0 applications. Therefore, my research goal is to, investigate, and conceptualize intents in the industrial area, to provide an intent-based approach to facilitate the use and configuration of available ICT components. Organizations would be able to adapt and stick, as quickly as possible, to changing business needs and to the requirements of digital transformation.

2 Problem Statement and Related Works

Our research problem is related to a variety of complex technological and organizational challenges. The situation requires an automated configuration as a facilitator for new ICT implementation, as well as the alignment to business intentions. We see that both the business intentions and digital technologies deployed by industries should be encompassed and supported by industrial organizations. We have identified three dimensions to solve this research issue: (i) state of the art of industrial intents and intent-based approaches, (ii) definition of the ICT components within the level of abstraction of our use cases, and (iii) contextual selection and configuration of ICT components using an intentional layer within an I4.0 application. We need to answer the following questions: (a) how are intents captured, translated, and applied regarding an I4.0 infrastructure? (b) Could an intentional layer link both business intents and ICT components of the I4.0 application? (c) How can an intent-based approach solve the ICT configuration issues regarding a given context?

2.1 Intents and Industrial Intention-Based Approaches

Software Networking intents or objectives were introduced first in 2015, in the context of Software Defined Network controllers, then in 2017, Gartner called it the "Next big thing" in networking to help organizations transition to digital [2]. The topic of intent is huge and can be deployed in various areas and many ways. Different dimensions can be considered to categorize or understand intents nature: by the considered role (e.g., Detection, processing, or implementation of intents), by concerns addressed with the defined purposes, by types of origin, intentions can directly be defined by humans or can be automatically generated, etc. Several previous works in different scientific fields have implemented the teleological (intent-based) perspective like Intent-Based Accounting [3], and Web Mining [4].

In the industrial area, intents can have different readings within the ICTs application and depending on the context. For instance, the intent of a business process may be translated as the fulfillment of its described task in the most efficient manner. In [5], intents are defined as business intents "when an intention is declared in a certain context it becomes a stated intent". Intents can be as simple as an atomic intent or as complex as an algorithm of intents that combines a set of sub-intents. In manufacturing applications, within ICT components, an intent is defined as the state that is intended to be achieved In collaborative assembly plans, they have defined two kinds of intents, Designer intent, which means the designer's knowledge about both the individual components as well as the entire assembly process (e.g. geometric information and spatial relationships to generate the assembly plan.) and the operator intents as the human objectives regarding the assembly processes [6].

2.2 ICT Configuration

ICT focuses on two critical aspects: assisting people in converting data into meaningful information and communicating that meaningful information to others. Therefore ICT can be defined as the fusion of methods and processes for the generation, storage, processing, transmission, and perception of information by humans or specialized devices. ICT systems, are based on the traditional components of data center infrastructure components (security, connectivity, power continuity) in addition to an overwhelming amount of internet-delivered content, software, hardware, and support services which include (according to [7]): (a) Internet, a digital communication infrastructure; (b) Data: raw facts and figures, information that is converted to meaningful insights; (c) Hardware components: Physical components that handle the information (creation, transmission, storage, management). (d) Software as a Service (SaaS): software applications over the internet and local client applications that assist with digital design, personal productivity, and process management. (e) Electronic materials facilitate the exchange of digital data. (f) Procedures: like services that support data assets management and customer experience management.

These ICT components are the key technologies underlying the I4.0 technological distributed, highly automated, and highly dynamic enablers, we can mention the most cited ones like IIoT, Big Data, AI, and Robotics, each has interdependencies and characteristics with I4.0.

One of the pressing issues in implementing I4.0 applications is to design uniform interoperability between all of these components. This challenge was globally reported by the World Economic Forum to implement I4.0, that, more than 300 platforms exist for IoT solutions, and four proposals about IoT architectures and reference models, where layers represent the interconnection between applications, services, and physical devices [8]. In the manufacturing industry, interoperability has emerged to become a core concept of I4.0; corresponding to this need, we want to extend the vision of IBN in a larger context. The IBN approach allows a specialized software layer to manage the network, it guarantees that network lifecycle management conforms to the stated objectives, even users could express their network requirements, and the software translates it into network configurations. The IBN systems have operational efficiencies by (a) translating the intents into policies, (b) automated implementation of these policies right after the approval of the network administrator, (c) continuous data analysis, for faults and performance, and (d) Assurance by real-time reporting.

For I4.0 standardization, we refer to the Reference Architectural Model for Industry 4.0 (RAMI 4.0), a service-oriented architecture developed by the German Electrical and Electronic Manufacturers' Association (ZVEI) to support I4.0 initiatives [9], it is used as a reference architecture model based on vertical integration, horizontal integration, and end-to-end engineering. RAMI 4.0 coordinates a three-dimension system, where crucial aspects of I4.0 are associated: (i) Layers, to simplify the IT perspective as a set of smaller manageable parts. (ii) Life Cycle & Value Stream, a layer based on the IEC 62890 standard, which represents the evolution of entities. (iii) Hierarchy Levels, an axis based on the IEC 6224, an international standard for enterprise control system integration. RAMI4.0 added three layers to respond to I4.0, 'Field Device' to introduce intelligence in systems, 'Product' and, 'Connected World' to add collaborative service networks to the factory.

RAMI 4.0 enables common assimilation of I4.0 standards and use cases, however, it does not clarify the details about implementation as needed, and there is no guidance regarding production in the manufacturing process, and communication between devices [10]. In addition, one of the dimensions of this model includes the business layer "intents," still it is reduced to business processes. This lack of a powerful alignment metaphor is a weakness of the RAMI 4.0 proposal and the other existing approaches. Which involves less efficient distribution of resources, less adapted configuration, lower level of business goals achievement, and adaptation to the context.

3 Scientific Method, Expected Results and Plan for the Evaluation

Ongoing Work*:* We are preparing a literature review of industrial intent-based approaches. To our knowledge, a taxonomy of industrial intents still didn't appear in the literature. We use the SMS method to give an overview of a research area by setting categories and counting existing works, then creating classifications and schemes respecting the founded categories [11]. To synthesize the existing literature about intent-based approaches used in industry, we have decomposed our research questions into two main categories. The first category of research questions consists of finding evidences about definitions or existing types and categories of intents in ICTs. The second category involves the configuration and adoption of ICT components, through mining, presenting, transforming, or implementation. The research questions are: (a) how intentions are characterized in the literature of the technological and industrial domain? (b) How intentions and sub-intents are distributed over the building blocks of an industrial architecture? (c) Could a valid taxonomy of intents in the industrial area be defined? (d) What are the strengths and weaknesses of the intent-based approach?

Current Results*:* We have chosen three scientific databases, Dimensions, Web of Science, and Scopus, because they are large multidisciplinary databases covering published material in the sciences, and they provide citation analysis of authors and subject areas. We have used criteria to find relevant papers for our project. Inclusion criteria are described in the following search string: ("industry 4.0" OR "smart industry" OR "manufacturing" OR "production" OR "smart factory" OR "Industrial Internet of Things") AND ("intent-based" OR "intent-oriented" OR "intention-based" OR "intention-oriented") OR ("goal-based" OR "goal-oriented") OR ("objective-based" OR "objective-oriented") AND the published year (between 2010 in 2022) AND the subject categories (Artificial Intelligence and Image Processing, Computer Software, Distributed Computing, Information Systems, Manufacturing Engineering) AND search clause (title and abstract). The applied criteria are given in Fig. 1.

Followed by the exclusion criteria, where an explicit description of approaches based on intents, should be observed in the title or abstract; or at least respond to one of the research questions. We ended with 80 articles divided into the several subdomains of

I4.0. The actual schema of relevant papers is still larger, and the concepts are much wider than expected, (more details could be provided on request).

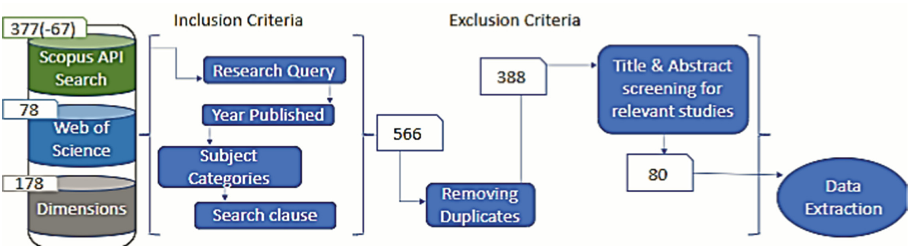


Fig. 1. Inclusion and exclusion criteria applied.

Plan for Future Work: The taxonomy of intent-based approaches from the 1st phase would help us organize the literature review and compare the various industrial intents. During the 2nd phase, we need to work on the characterization and standardization of industrial ICT components, according to the current status, we are going to rely on RAMI4.0; Moving to the analysis of CIM approach and IBN architectures in the 3rd phase, we will adapt each selected intent, in relation with its appropriate context, to the use of the ICT component, in this way we can prepare the configuration of the components of the I4.0 applications through the link with the business strategy. These phases are summed up in Fig. 2.

As a use case for our project, we conducted meetings with the French national electricity board (EDF). The IS structure of EDF is built on the Computer Integrated Manufacturing (CIM) architecture, it is an architecture that enables industrial enterprises to integrate information and business processes. The next step for us is to find common

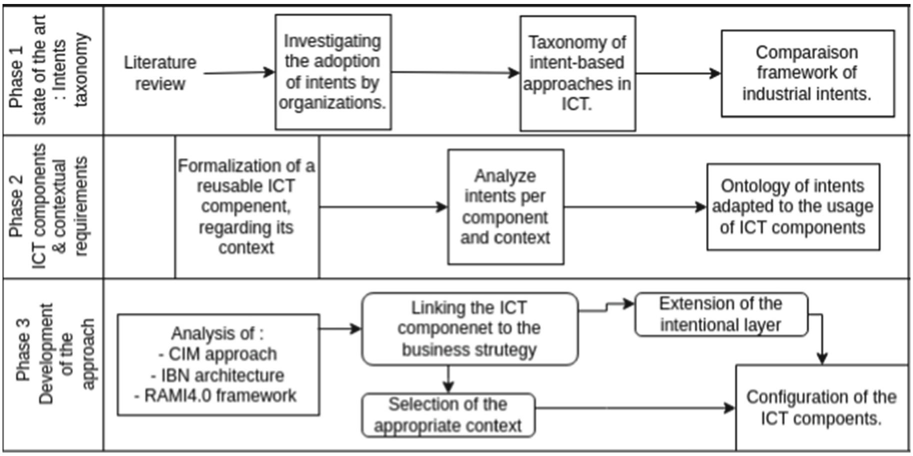


Fig. 2. Research phases.

concepts between CIM and IBN architecture and to study the position of the intentional layer. Then, we propose to apply Situational Method Engineering (SME) inspired methods. SMEs allow building methods adapted to a concrete real use case where reusable method components depend on context factors [12]. Another possible objective for us, obviously an important track, is a direct contribution to the Development Repository for the open Asset Administration Shell (openAAS) project, which works and searches for demonstrable and verifiable reference solutions to strengthen the distribution of I4.0 components (more details could be provided on request).

References

1. Kasakow, G., Aurich Jan, C.: Realising digital connectivity by using interdependencies within a production process. Procedia CIRP **52**, 80–83 (2016). ISSN 2212-8271
2. Han, Y., Li, J., Hoang, D., Yoo, J.-H., Won-Ki Hong, J.: An intent-based network virtualization platform for SDN. In: Proceedings of the 12th International Conference on Network and Service Management (CNSM), pp. 353–358 (2016)
3. Déneckère, R., Kornyshova, E., Hug, C.: A framework for comparative analysis of intention mining approaches. In: Cherfi, S., Perini, A., Nurcan, S. (eds.) Research Challenges in Information Science. RCIS 2021. LNBIP, vol. 415, pp. 20–37. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75018-3_2
4. Zhang, Q., Jiang, X., Sun, J.: Web mining based on VIPS in intention-based information retrieval. In: 2009 International Conference on Natural Language Processing and Knowledge Engineering, pp. 1–5 (2009)
5. Silvander, J., Wnuk, K., Svahnberg, M.: Systematic literature review on intent-driven systems. IET Softw. **14**, 345–357 (2020)
6. Cramer, M., Kellens, K., Demeester, E.: Probabilistic decision model for adaptive task planning in human-robot collaborative assembly based on designer and operator intents. IEEE Robot. Autom. Lett. **6**(4), 7325–7332 (2021)
7. Definition – What does Information and Communications Technology (ICT) mean? Techno-pedia. https://bit.ly/3wyl86N. Accessed 20 Aug 2022
8. Garkushenko, O.N.: Information And Communication Technologies In The Era Of The Emergence Of Smart Industry: Problems Of Definition And Conditions Of Development (2018)
9. Pisching, M.A., Pessoa, M.A., Junqueira, F., dos Santos Filho, D.J., Miyagi, P.E.: An architecture based on RAMI 4.0 to discover equipment to process operations required by products. Comput. Ind. Eng. **125**, 574–591 (2018)
10. R. Langmann, L.F. Rojas-Pena, A PLC as an Industry 4.0 component 2016 13th international conference on remote engineering and virtual instrumentation (REV), IEEE (2016), pp. 10–15, (2016)
11. Adolph, L., Anlahr, T., Bedenbender, H., Bentkus, A., Brumby, L., Diedrich, C., Mehrfeld, J.: German Standardization Roadmap Industry 4.0 (2016)
12. Kornyshova, E., Deneckère, R., Rolland, C.: Method families concept: application to decision-making methods. In: Halpin, T., et al. (eds.) BPMDS/EMMSAD -2011. LNBIP, vol. 81, pp. 413–427. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21759-3_30

Author Index

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2023

T. P. Sales et al. (Eds.): EDOC 2022 Workshops, LNBIP 466, pp. 377–378, 2023.
https://doi.org/10.1007/978-3-031-26886-1