Ewa Ziemba (Ed.)

Information Technology for Management

Ongoing Research and Development

LNBIP 311

15th Conference, AITM 2017
and 12th Conference, ISM 2017, Held as Part of FedCSIS
Prague, Czech Republic, September 3–6, 2017
Extended Selected Papers

Lecture Notes in Business Information Processing 311

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

Ewa Ziemba (Ed.)

Information Technology for Management

Ongoing Research and Development

15th Conference, AITM 2017
and 12th Conference, ISM 2017, Held as Part of FedCSIS
Prague, Czech Republic, September 3–6, 2017
Extended Selected Papers

Springer

Editor
Ewa Ziemba
University of Economics in Katowice
Katowice
Poland

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-319-77720-7 ISBN 978-3-319-77721-4 (eBook)
https://doi.org/10.1007/978-3-319-77721-4

Library of Congress Control Number: 2018934372

Printed on acid-free paper

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

Preface

Two editions of this book appeared in past two years: *Information Technology for Management* in 2016 (LNBIP 243) and *Information Technology for Management: New Ideas or Real Solutions* in 2017 (LNBIP 277).

Given the rapid developments in information technology and its applications for improving management in business and public organizations, there was a clear need for an updated version.

The present book includes extended and revised versions of a set of selected papers submitted to the 12th Conference on Information Systems Management (ISM 2017) and 15th Conference on Advanced Information Technologies for Management (AITM 2017) held in Prague, Czech Republic, during September 3–6, 2017. These conferences were organized as part of the Federated Conference on Computer Science and Information Systems (FedCSIS 2017).

FedCSIS provides a forum for bringing together researchers, practitioners, and academics to present and discuss ideas, challenges, and potential solutions in established or emerging topics related to research and practice in computer science and information systems. Since 2012, the proceedings of FedCSIS are indexed in Thomson Reuters Web of Science, Scopus, IEEE Xplore Digital Library, and other indexing services.

ISM is a forum for computer scientists, IT specialist, and business people to exchange ideas on management of information systems in organizations, and the usage of information systems for enhancing the decision-making process and empowering managers. It concentrates on various issues of planning, organizing, resourcing, coordinating, controlling, and leading the management functions to ensure a smooth operation of information systems in organizations.

AITM is a forum for all in the field of business informatics to present and discuss the current issues of IT in business applications. It is mainly focused on business process management, enterprise information systems, business intelligence methods and tools, decision support systems and data mining, intelligence and mobile IT, cloud computing, SOA, agent-based systems, and business-oriented ontologies.

For ISM 2017 and AITM 2017, we received 48 papers from 18 countries in all continents. From these, after a review process, only nine papers were accepted as full papers and 11 as short papers. In all, 13 papers of the highest quality as ranked by the Program Committee were chosen and the authors were invited to extend their papers and submit them for consideration to the LNBIP publication. Our guiding criterion for including papers in the book was the excellence of publications indicated by the reviewers, the relevance of the subject matter for the economy, and promising results. The selected papers reflect state-of-art research work that is often oriented toward real-world applications and highlight the benefits of information systems and technology for business and public administration, thus forming a bridge between theory and practice.

The papers selected to be included in this book contribute to the understanding of relevant trends of current research on information technology for management in business and public organizations. As with the two previous editions of this book, the first two parts focus on information technology and information systems for knowledge management and business transformation, whereas the third part presents research on implementation and evaluation of information systems.

Finally, the authors and I hope readers will find the content of this book useful and interesting for their own research activities. It is in this spirit and conviction that we offer our monograph, which is the result of the intellectual effort of the authors, for the final judgment of our readers. We are open to discussion on the issues raised in this book, and look forward to the polemical, or even critical, voices as to the content and form.

January 2018 Ewa Ziemba

Organization

AITM 2017

Event Chairs

Frederic Andres	National Institute of Informatics, Tokyo, Japan
Helena Dudycz	Wrocław University of Economics, Poland
Mirosław Dyczkowski	Wrocław University of Economics, Poland
Frantisek Hunka	University of Ostrava, Czech Republic
Jerzy Korczak	Wrocław University of Economics, Poland

Program Committee

Witold Abramowicz	Poznan University of Economics, Poland
Frederik Ahlemann	University of Duisburg-Essen, Germany
Ghislain Atemezing	Mondeca, Paris, France
Kenneth Brown	Communigram SA, France
Agostino Cortesi	Università Ca' Foscari, Venezia, Italy
Beata Czarnacka-Chrobot	Warsaw School of Economics, Poland
Suparna De	University of Surrey, Guildford, UK
Jean-François Dufourd	University of Strasbourg, France
Bogdan Franczyk	University of Leipzig, Germany
Arkadiusz Januszewski	UTP University of Science and Technology in Bydgoszcz, Poland
Rajkumar Kannan	Bishop Heber College (Autonomous), Tiruchirappalli, India
Grzegorz Kersten	Concordia University, Montreal, Canada
Ryszard Kowalczyk	Swinburne University of Technology, Melbourne, Australia
Karol Kozak	TUD, Germany
Christian Leyh	Technische Universität Dresden, Germany
Antoni Ligęza	AGH University of Science and Technology, Poland
André Ludwig	University of Leipzig, Germany
Damien Magoni	University of Bordeaux – LaBRI, France
Krzysztof Michalak	Wroclaw University of Economics, Poland
Mieczyslaw Owoc	Wroclaw University of Economics, Poland
Malgorzata Pankowska	University of Economics in Katowice, Poland
Jose Miguel Pinto dos Santos	AESE Business School Lisbon, Portugal
Artur Rot	Wroclaw University of Economics, Poland
Stanislaw Stanek	General Tadeusz Kosciuszko Military Academy of Land Forces in Wroclaw, Poland

Jerzy Surma	Warsaw School of Economics, Poland and University of Massachusetts Lowell, USA
Stephanie Teufel	University of Fribourg, Switzerland
Edward Tsang	University of Essex, UK
Jarosław Wątróbski	West Pomeranian University of Technology in Szczecin, Poland
Tilo Wendler	Hochschule fur Technik und Wirtschaft Berlin, Germany
Waldemar Wolski	University of Szczecin, Poland
Cecilia Zanni-Merk	INSA de Rouen, France
Ewa Ziemba	University of Economics in Katowice, Poland

ISM 2017

Event Chairs

Bernard Arogyaswami	Le Moyne University, USA
Witold Chmielarz	University of Warsaw, Poland
Dimitris Karagiannis	University of Vienna, Austria
Jerzy Kisielnicki	University of Warsaw, Poland
Ewa Ziemba	University of Economics in Katowice, Poland

Program Committee

Ahmad T. Al-Taani	Yarmouk University, Jordan
Saleh Alghamdi	University of Sussex, UK
Nemeslaki András	National University of Public Service, Budapest, Hungary
Klára Antlová	Technical University of Liberec, Czech Republic
Andrzej Bialas	Institute of Innovative Technologies EMAG, Poland
Zane Bicevska	DIVI Grupa Ltd., Latvia
Boyan Bontchev	Sofia University St. Kliment Ohridski, Bulgaria
Alberto Mora Carreño	Universitat Oberta de Catalunya, Spain
Dimitrios Chatzoudes	Democritus University of Thrace, Greece
Mariusz Chmielewski	Military University of Technology, Poland
Adriana Csikosova	Technical University of Košice, Slovakia
Beata Czarnacka-Chrobot	Warsaw School of Economics, Poland
Robertas Damaševičius	Kaunas University of Technology, Lithuania
Horatiu Dragomirescu	Bucharest University of Economic Studies, Romania
Yanqing Duan	University of Bedfordshire, UK
Helena Dudycz	Wrocław University of Economics, Poland
Ibrahim El Emary	King Abdulaziz University, Saudi Arabia
Susana de Juana Espinosa	University of Alicante, Spain
Francisco Jose Gallego Duran	University of Alicante, Spain
Aleksandra Gaweł	Poznan University of Economics and Business, Poland
Nitza Geri	The Open University of Israel, Israel

Aaron M. Glassman	Embry-Riddle Aeronautical University, USA
Leila Halawi	Embry-Riddle Aeronautical University, USA
Juho Hamari	University of Tampere, Finland
Jarosław Jankowski	West Pomeranian University of Technology in Szczecin, Poland
Dorota Jelonek	Czestochowa University of Technology, Poland
Andrzej Kobyliński	Warsaw School of Economics, Poland
José Paulo Leal	University of Porto, Portugal
Christian Leyh	Technische Universität Dresden, Germany
Krzysztof Michalik	University of Economics in Katowice, Poland
Roisin Mullins	University of Wales Trinity Saint David, UK
Karolina Muszyńska	University of Szczecin, Poland
Daniela Nicklas	University of Bamberg, Germany
Walter Nuninger	Polytech'Lille, Université de Lille, France
Shigeki Ohira	Nagoya University, Japan
Alexander Paar	TWT GmbH Science and Innovation, Germany
Zbigniew Pastuszak	Maria Curie-Sklodowska University, Poland
Elvira Popescu	University of Craiova, Romania
Ricardo Queirós	Politécnico do Porto, Portugal
Jayanthi Ranjan	Institute of Management Technology in Ghaziabad, India
Nina Rizun	Alfred Nobel University, Dnipropetrovs'k, Ukraine
Uldis Rozevskis	University of Latvia, Latvia
Marcin Jan Schroeder	Akita International University, Japan
Martin Sillaots	Tallinn University, Estonia
Alberto Simões	Instituto Politécnico do Cávado e do Ave, Portugal
Andrzej Sobczak	Warsaw School of Economics, Poland
Piotr Sobolewski	Wroclaw University of Science and Technology, Poland
Jaroslav Šušol	Comenius University in Bratislava, Slovakia
Jakub Swacha	University of Szczecin, Poland
Symeon Symeonidis	Democritus University of Thrace, Greece
Marco Temperini	Sapienza University in Rome, Italy
Bob Travica	University of Manitoba, Canada
Manuel Tupia	Pontificia Universidad Catolica del Peru, Peru
Jarosław Wątróbski	West Pomeranian University of Technology in Szczecin, Poland
Janusz Wielki	Opole University of Technology, Poland
Waldemar Wolski	University of Szczecin, Poland
Michal Žemlička	Charles University in Prague, Czech Republic

Acknowledgments

I would like to take this opportunity to express my gratitude to all those who contributed to the ISM 2017 and AITM 2017 research events. First of all, the authors, whose quality work was the essence of the conference, and the members of the Program Committee, who helped us with their expertise and diligence in reviewing the papers. I am deeply grateful to the program chairs of ISM 2017 and AITM 2017, namely, Witold Chmielarz, Helena Dudycz, and Jerzy Korczak, for their extensive organizational involvement in the conferences and the evaluation of papers. I acknowledge the chairs of FedCSIS 2017, i.e., Maria Ganzha, Leszek A. Maciaszek, and Marcin Paprzycki, for putting a lot of effort into organizing the large-scale and excellent research event. I am indebted to the Springer team, led by Ralf Gerstner and Alfred Hofmann, without whom this book would not have been possible.

Contents

Implementation and Evaluation of Information Systems

Information Technology and Systems for Knowledge Management

Conceptualization of a Value Cocreation Language for Knowledge-Intensive Business Services

Christophe Feltus[1(✉)], Lysanne Lessard[2], François Vernadat[3], Daniel Amyot[2], and Erik. H. A. Proper[1]

[1] Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg
{christophe.feltus, erik.proper}@list.lu
[2] University of Ottawa, Ottawa, Canada
lessard@telfer.uottawa.ca, damyot@uottawa.ca
[3] University of Lorraine, Metz, France
vernadat30@gmail.com

Abstract. Knowledge-intensive business services (KIBS) are business-to-business services that are characterized as being knowledge intensive, relying on expert employees, and providing knowledge-based solutions to customers. As a context for service design, KIBS bring unique challenges regarding the need to communicate about value cocreation among companies entering into a service exchange. Unfortunately, until now, there have been limited contributions in the area of modeling languages to support the cocreation of value during business exchanges. In this paper, an abstract language (metamodel) is proposed to support IT designers in understanding value cocreation in the field of IT-related business services. A value creation metamodel is first structured around three dimensions: the nature of the value, the method of value creation, and the business object impacted by the value. Then, value cocreation is modeled as a specialization of the value creation metamodel. This new language is illustrated with a case study related to KIBS in the financial sector.

Keywords: Value cocreation · Service-dominant logic
Knowledge-intensive business services · Metamodel · Language
Service system

1 Introduction

Knowledge-intensive business services (KIBS) are business-to-business services, such as engineering services, management consulting, and information technology (IT) sourcing [1], that are characterized as being knowledge intensive, relying on expert employees, and providing knowledge-based solutions to their clients [2]. These characteristics imply that clients co-define and co-produce desired solutions with KIBS providers [3]. As such, designing KIBS requires paying attention to the value cogenerated amongst KIBS providers, clients, and partners [4]. Value cocreation (VCC) is a concept anchored in marketing theory that explains how value can be co-defined and co-generated during business

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 3–23, 2018.
https://doi.org/10.1007/978-3-319-77721-4_1

exchanges among two or more partners [5, 6]. Many examples of value cocreation in KIBS are gathered in [1]. One of them concerns, for instance, the company PowerDrive, a Swedish manufacturer of hydraulic drive systems that cocreated value with three of its customers based on the collection and analysis of data from an existing remote monitoring system [7]. Another example from [4] concerns KIBS engagements between a team of university professors and their students, and a number of departments within a Canadian municipality, which collaborate on the development of an online event intended to promote city services to a broader range of city residents. In many cases such as PowerDrive and the Canadian municipality, VCC is facilitated thanks to interconnections between the organizations' information systems (IS) and their clients. As a result, understanding how value cocreation happens in this context is essential for the design of information systems that support the development of impactful business-to-business services and information exchanges between the employees who design KIBS, at the provider and client sides. Unfortunately, despite a significant body of empirical research aiming to depict the foundations of VCC (e.g., value in use, value in exchange, etc. [5, 6, 8–12]), and a profusion of languages to express elements of *value creation* (method, natures of the value or type of objects concerned by the value), few contributions have paid attention, until now, to the issue of languages to support the *cocreation of value* during business exchanges, especially in the field of KIBS. This shortcoming makes core concepts of value cocreation difficult to operationalize and, consequently, becomes a risk when KIBS designers must communicate with each other [13] to cocreate value. For example, during the creation of the value proposition, a provider needs to communicate (using a dedicated language) with its customers in order to understand the elements that the latter consider as valuable, and to request access to some parts of the customers' IS architecture.

In that context, this paper proposes a metamodel (language) to support KIBS designers in understanding value cocreation in the field of IT-related business services (e.g., IT outsourcing). The design of this language is illustrated with a case study related to KIBS in the financial sector. The first part of the case study illustrates the *creation of value* in the field of IT outsourcing when a bank outsources the archiving of its customers' data to a datacenter. The second part of the case study illustrates the *cocreation of value* between the bank and the datacenter. Indeed, because both companies have been collaborating for a long time, the datacenter has good knowledge of the bank's information system. For that reason, the bank has decided to outsource the improvement of the privacy of the customers' data to the datacenter. Both have hence started to cooperate to design the privacy improvement service of the customers and therefore the bank has agreed to give information about its information system (architecture, functions, etc.) to the datacenter. In turn, the latter enhances its offer of services and thereby stabilizes its own business. The enhancement is possible as a result of the bank's feedback.

Concretely, in our previous work [14], we have observed through an analysis of the literature from different disciplines and various business sectors, that the three following aspects need to be considered together in order to address the cocreation of value: the nature of value (e.g., privacy and money), the method used to create value (e.g., privacy impact assessment method), and the object concerned by value (e.g., the bank customers' data archived by the datacenter). Based on this previous work, the

paper presents research performed in the context of the design science research approach that aims to produce two artefacts: an improved version of the value creation metamodel proposed in [14] that incorporates stakeholder and resource concepts, and a specialization of the latter to the processes of value cocreation.

The paper is structured as follows: in the next section, the literature related to value cocreation is reviewed and in Sect. 3, the applied research method is presented. Section 4 presents the findings, namely the improved version of the value creation metamodel and its specialization to value cocreation. Section 5 discusses these results and Sect. 6 concludes the paper.

2 Literature Review

The concept of VCC originates from the field of marketing. It aims to define and to explain the mechanism for the co-generation of value during business exchanges amongst two or more companies [5, 6, 8]. Vargo and Lusch [5, 6] formalized it using a framework for defining VCC in the perspective of the service-dominant logic (S-DL). According to the authors, a service is the basis of all exchanges and focuses on the process of value creation rather than on the creation of tangible outputs. As a result, a service system is a network of agents and interactions that integrates resources for VCC [5]. On that basis, Vargo et al. further elaborate on the idea that value is derived and determined in use rather than in exchange, meaning that value is proposed by a service provider and is determined by a service beneficiary. Hence, the firm is in charge of the value-creation process and the customer is invited to join in as a co-creator [5]. For Grönroos [15], this interaction is defined through situations in which the customer and the provider are involved in each other's practices. Consequently, the context (social, physical, temporal, and/or spatial) determines the value-in-use experience of the user in terms of his individual or social environment [16].

Modeling value cocreation in the specific field of the Knowledge-Intensive Business Service has been addressed by Lessard [6] who proposes the value cocreation modeling (VCM) framework to fulfill the requirement emerging from that domain. In parallel, Hastings and Saperstein [10] also define a set of six concepts to design the practice-driven service framework for value creation, namely: customers co-create value with providers, value is created in service systems, modular business architecture, scalable Glo-Mo-So (global, mobile, social) platforms, continuous improvement via learning, and multi-sided metrics. At the analytical level, Storkacka et al. [11] have complementarily proposed to analyze the actors' engagement as a micro-foundation (explanation on a low analytical level) for VCC whereas Frow et al. [12] proposed a framework to assist firms in identifying new opportunities for value cocreation. Therefore, the authors provide a strategically important new approach for managers to identify, organize, and communicate innovative opportunities.

Recently, Chew [17] has argued that, in the digital world, service innovation is focused on customer value creation. Chew proposes an integrated Service Innovation Method (iSIM) that allows analyzing the interrelationships between the design process elements, including the service system. The latter being defined as an IT/operations-led,

cross-disciplinary endeavor. At the information system domains level, Blaschke et al. [18] propose a business-model-based management method encouraging cocreation interactions by reconciling value propositions, customer relationships, and interaction channels. Gordijn et al. [19] explain that business modeling is not about process but about value exchange between different actors. Gordijn et al. propose e^3value to design models that sustain the communication between business and IT groups, particularly in the context of the development of e-business systems. In [20], Weigand extends the e^3value language to consider cocreation. He defines so-called value encounters, which consist in spaces where groups of actors interact to derive value from the groups' resources. In a similar way, Razo-Zapata et al. propose visual constructs to describe the VCC process [21]. These constructs are built on requirements from the service-dominant logic and software engineering communities. They aim to express three cocreation types (co-ordination, co-operation and collaboration) following the three elements of the customer relationship experience: cognition, emotion, and behavior [9]. According to [22], the cocreation may happen through different processes (B2C, B2B, C2B, or C2C) and may refer to different types of value (for the company or the customer).

While existing approaches help to operationalize the concept of VCC, none of them fully considers all the dimensions necessary to cover the VCC domain. This can be seen through recent work presenting the state of the art in the field of VCC. The first one reviews the existing literature through both following perspectives: co-production and value-in-use [23], and the second through two dimensions: theoretical dimension of the cocreation, and collaboration and cocreation between firms and customers [24]. Thus, despite existing contributions, the need to design an effective language to support the management of VCC [19, 20] while considering the nature of the value, the object concerned by this value, and the method used to create the value, has yet to be addressed.

3 Research Methodology

At a methodological level, the research that is undertaken concerns the improvement of value cocreation in the field of knowledge-intensive business services. To achieve this goal, the approach consists in designing a value creation metamodel and in specializing it to express value cocreation. Through this research, we aim to strengthen the organizational capability to improve the design of the information system that sustains this cocreation of value. Hevner et al. [25] explain that the Design Science Research (DSR) paradigm seeks to extend the boundaries of human and organization capability by creating new and innovative artefacts. Practically, provided that we aim to design two new artefacts to support the design of the information system, we acknowledge that this research may plainly be considered in the scope of DSR [26]. Moreover, given that both artefacts are motivated by real problems and rely on the knowledge of the field, we need to involve practitioners all along the artefact building activities. Therefore, we apply the Action Design Research method proposed by Sein et al. [27], whose objective is to strengthen the connections between the practitioners and the researchers

by combining the building, intervention, and evaluation (BIE) activities. Moreover, postulating that the elaboration of the artefacts strongly relates to the IS, we apply an IT-Dominant BIE generic schema (Fig. 1).

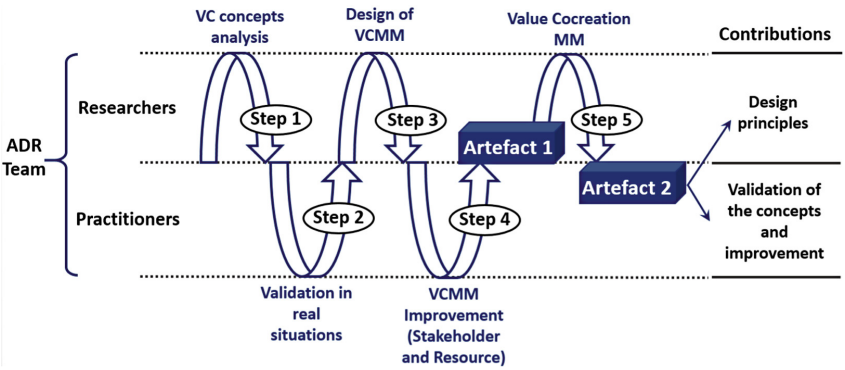

Fig. 1. IT-Dominant BIE generic schema applied to VCC design (adapted from [27])

As advocated by DSR principles [26, 27], the method used to design the value creation model is an iterative approach. Applied to this research, in step 1 (Fig. 1), we (researchers) have analyzed the concepts meaningful to the creation of value from the literature and from different frameworks and we have designed a high-level value creation metamodel, structured along three dimensions (nature of the value, value creation method, and object concerned by the value). In step 2, this value creation metamodel has been tested with regard to real situations with practitioners from different sectors but mostly from healthcare and financial institutions. In step 3, we have formalized the first version of the value creation metamodel (VCMM); the latter was presented in [14].

In this paper, we analyze to what extent this value creation metamodel may be used to model the cocreation of value in KIBS. Therefore, a first statement is that, opposite to value creation, value cocreation implies at least two stakeholders who collaborate. Additionally, the latter may have different roles among which the role to provide resources to support the value cocreation. Consequently, an intermediary step in the approach consists in improving the value creation metamodel proposed in [14] with the concepts of stakeholder and resource. We hence perform a conceptual integration of the value creation metamodel with the model of value presented in [28, 29] (step 4). The resulting integrated metamodel, improved with the stakeholder and resource concepts, consists in artefact 1. Afterwards, in step 5, this artifact 1 is specialized to the processes of value cocreation. The latter is applied to the context of KIBS and is illustrated based on the *Processes of value cocreation* proposed by [6]. The cocreation itself is illustrated with a case of IT outsourcing in the financial sector, namely the outsourcing of *privacy* management. This specialization is artefact 2.

4 Research Findings

In this section, the new version of the value creation metamodel is presented. It corresponds to the version presented in [14] improved based on the value model from [28]. Then, the value creation metamodel is specialized to the processes of value cocreation.

4.1 Value Creation Metamodel

In this section, the metamodel of value creation in the field of IT-related business services is defined according to three dimensions (Fig. 2): the nature of the value, the method of value creation, and the object concerned by the value.

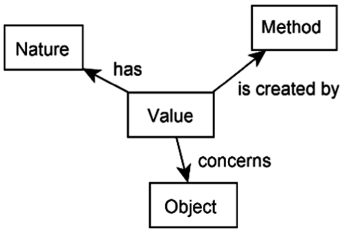

Fig. 2. Three value dimensions

Provided that this research is anchored in DSR, this section presents the last version of the value creation metamodel design iterations. The metamodel is elaborated based on the analysis of value related frameworks [30–48], of scientific literature [3–6, 8–12, 16–23] and on a *performance evaluation methodology for decision support in industrial project* proposed in [28]. The aim of this methodology is to propose a benefit-cost-value-risk based approach to help decision makers in evaluating performance at any stage of an industrial project. The latter allows considering two additional concepts necessary to model the cocreation of value in the field of knowledge-intensive business services: *stakeholder* and *resource*.

In the next sub-sections, each dimension of the value is successively analyzed and modeled, and the integrated value creation metamodel is presented in the last sub-section. Moreover, concepts of the metamodel are illustrated using the first part of the case study related to the outsourcing, by the bank, of the customer's data archiving to a datacenter.

Dimension 1: Nature of the value

To understand and model the nature of the value, first a set of frameworks addressing the different value natures in the field of IT has been reviewed, including security, quality, compliance, privacy, responsibility, and others (Table 1). Based on this review, the most meaningful concepts necessary to express this nature have been extracted. For example, the information systems security risks management (ISSRM [30]) framework, which addresses the IS security (*Nature of the value*), has been analyzed. This framework characterizes security through integrity, confidentiality, availability,

non-repudiation, and accountability (*Value components*), and the latter concerns business assets of the company (*Objects*). Finally, based on a deeper review of the literature, our own definitions of the concepts composing the dimension have been provided in an integrated metamodel of nature of the value (Fig. 3).

Table 1. Nature of the value in the field of IT

Value reference framework	Nature of the value examples		
	Nature of the value	*Component of the nature of the value*	*Concerned object*
ISSRM [30]	IS security	Confidentiality, integrity, availability, non-repudiation, accountability	Business asset
ReMMo [31]	Responsibility	Accountability (e.g., RACI)	Actor
Web quality model [32]	Quality	Functionality, reliability, usability, efficiency, portability, maintainability	Web feature
EA compliance model [33]	Compliance	Correctness, justification, consistency, completeness	Acts of software developers
Privacy metamodel [34, 35]	Privacy	Notice, choice and consent, proximity and locality, anonymity and pseudonymity, security, and access and resource	Sensitive information
VDML [360]	Generic value	Factor of benefit, factor of interest	Business item
HCI [37]	Usability	Learnability, flexibility, robustness	Design rules, design knowledge

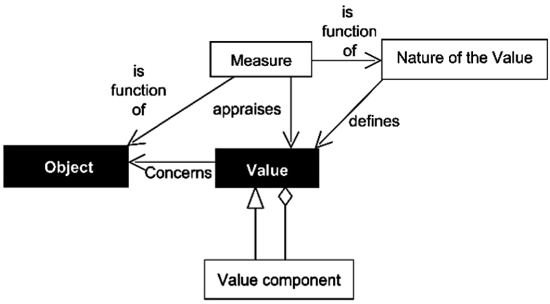

Fig. 3. Nature of the value metamodel

Basically, most reference frameworks [30–37] analyzed focuses on depicting the semantic of value following a given perspective being function of the beneficiary of the value. In practice, due to the quantity of heterogeneous value natures [22], clearly defining the semantic of the latter is laborious. However, we observe that two main perspectives of value nature emerge depending on the context: value at the provider's side vs. value at the customer's side. At the provider's side, the basic rationale for all

companies entering into dyadic exchange relationships is the value capture [49] from a service exchange. This can be in the form of value-in-exchange (e.g., money given by the client), or in the form of value-in-context. In that regard, it is worth noting that considering the provider in the context of the digital society expands this narrow meaning to the consideration of other value elements, such as the information collected on the customers (e.g., analyzing customer data to support the creation of new offerings) which, afterwards, contributes to economic increase [50]. On the customer's side, value generated by a transaction never refers to money but consists in other wealth, which contributes in sustaining and supporting the customer's own business.

According to [28], value is described as the degree of satisfaction of a set of stakeholder expectations or needs, expressed by the appreciation level of a number of performance indicators. Li [29] explains that value can be described by the relative worth, utility, or importance of something. Value increases when the customer's degree of satisfaction increases. The concept of value becomes different depending on the point of view (stakeholder). Accordingly, the *expected value* is the value that the stakeholder would like to get and the *perceived value* is the real value that a stakeholder can finally get. The degree of satisfaction is identified through the comparison of these two elements. According to Zeithaml, value implies some form of *assessment of benefits against sacrifices* [8].

Regarding the case study proposed in the introduction, at the bank's side, the privacy of the customers' data is a legal requirement that has to be fulfilled by each entity processing private information. Having this data privacy generates the benefit of being compliant with regulations, but it is also expensive because the bank needs to deploy an appropriate mechanism to set up this privacy, such as performing privacy impact assessment. At the datacenter's side, offering 24/7 data availability to the bank is a benefit to distinguish the datacenter from its competitors, but this offering is also costly because it requires a very robust infrastructure.

According to this review, the concepts that are relevant to the metamodel for the nature of the value are:

- **Value.** This concept is defined as a degree of worth that concerns something [28, 29] and that improves the well-being of the beneficiary after it is delivered [51].
- **Nature of the value.** Table 1 shows that the nature of the value expresses a domain of interest related to which the value will be delivered (e.g., security of the IS, the cost of a transaction, or the privacy of personal data). As a consequence, the nature of the value **defines** the value to be delivered. In the case of the datacenter that archives the data of the bank customers, the nature of the value generated by the datacenter is the *availability* of the customer's data.
- **Value component.** This concept expresses the different elements that constitute the value, or the pillars that found this nature (e.g., availability, confidentiality, portability, etc.). Hence, the value **aggregates** value components and the latter may also, as a result, themselves be other **types of** value. Regarding the case study, one component of the availability is the *accessibility in real time*.
- **Object.** The object concerned by the value is the element from the information system that has significance and is necessary for a company to achieve its goal, and that is be better off after that value is delivered (e.g., software, process, data).

From a modeling point of view, the value is associated to an object with a relation of type **concerns** or objective to be achieved. In the case study, the object concerned by the value is the *customers' data.*

- **Measure.** The measure corresponds to a property on which calculations can be made for determining the amount of value expected from a value creation method. This measure (e.g., the *% of time data is available*) can result from different factors impacting value. This corroborates the statement made in [28], which argues that the value components are measured by means of estimation methods. Accordingly, there exist an association named **appraises** from the concept of measure to the concept of value, an association named **is function of** between the concept of measure and the type of value, and between the concept of measure and the object concerned by the value. The first expresses that the measure is characterized by the nature of the value and the second that the measure also depends on the object concerned by the value. According to [28], this measure may integrate qualitative and quantitative elementary performance expressions.

Based on the above definitions, the nature of the value is modeled in Fig. 3.

Dimension 2: Method of value creation

A method of value creation corresponds to a set of activities that contribute to the generation of value in the field of IT. Likewise as for the nature of the value, in order to depict the elements relevant for the creation of value, a set of IT-related frameworks about value creation methods have been reviewed (Table 2) and this review has afterwards been completed with elements from the literature. The methods analyzed so far include method by design [35], model driven [28], impact assessment [29], method chunk [40], risk-based [41], and process-based [42] approaches.

Traditionally, value is created through the exchange and use of goods and services [5]. Methods for value creation are the body of techniques and activities that use and generate resources [52]. These correspond, at the corporate level, to a bundle of approaches including the design of strategies, the integration of models, the evaluation of results, etc. (Table 2). By looking more closely at the methods analyzed, it has been observed that each has a dedicated goal, that they are composed of method elements, and that the latter are organized in a sequence of ordinated steps. For instance, by investigating the model-driven approach to interoperability, one can notice that it has for goal to improve interoperability of enterprises' information systems that it is composed of models, and that three steps are required for model-driven interoperability: model design, model integration, and model instantiation. Amongst the other methods reviewed, it is also interesting to highlight that one (method chunk) has for particular objective the creation of methods themselves, using, as chunk of existing methods as method elements, and as method steps the decomposition of existing methods into method chunks and the definition of new method chunks from scratch [40].

As a summary and according to this analysis, the concepts that make the method of value creation are:

- **Method.** The method is a specific **type of** object that defines the means used by the stakeholder to **create** objects and value. According to Table 2, a method is **composed** of a set of activities necessary to achieve a dedicated goal. In the same vein,

Sein et al. [27] explain that the elementary quantitative value expressions (the value components) are aggregated by means of selected aggregation methods and quantitative weights to generate the overall value. The method used to create the availability is *the exploitation of a redundancy system (tools and procedures to guarantee redundancy).*

Table 2. Methods of value creation in the context of IT development

Method reference	Method of value creation examples		
	Method	*Goal*	*Activity*
[35]	By design	Prevent privacy risk from occurring	Project-by-project approach realization
[38]	Model-driven	Improve interoperability of companies' information systems	Models design, model integration, and model instantiation
[39]	Impact assessment	Explore social consequences for social security policies	Scenario design, design of strategies, assessment of impacts, ranking of strategies, mitigation of negative impacts, reporting, stimulation of implementation, auditing and ex-post evaluation
[40]	Method chunk	Method creation	Decomposition of existing methods into method chunks and definition of new method chunks from scratch
[41]	Risk-based	Security strategy development	Analysis of the method elements and identification of the options that exist in investment decisions
[42]	Process-based	Risk management for global supply chain	Step-by-step execution in a function of the dependency amongst them

- **Activity.** The activity is an element of the method that corresponds to a unitary task (e.g., analysis, collect of information, or report). The activities **compose** the method and are organized and coherently articulated with each other (e.g., if-then-else, process elements ordination, etc.). This relation is modeled using an iterative association of a type: activity **follows** activity. The articulation of activities corresponds to the aggregation from [14]. One particular type of activity consists in **generating** resources. For instance: *acquiring a backup tool, maintain the backup tool, etc.*
- **Stakeholder.** A stakeholder is a human, a machine or an organization that is involved in the creation of value at three levels. First, it **performs** the method that generates value (e.g., the risk manager performs a risk analysis); second, it **generates** resources used by the method; and third it **expresses** the value expected after the execution of the method. For example, the *datacenter* is the stakeholder that exploits the redundancy system and the *bank* expresses that it expects availability of the data.

- **Resource.** This element is **a type of** object from the IS that is generated by a stakeholder and that **is used by** an activity composing the value creation method. Resources are typically information and data (e.g., passenger location), but could also consist in computing resources, funding, manpower, etc. For instance, the *backup software* is the resource used by the exploitation of a redundancy system.

Based on the above definitions, the value creation method is modeled in Fig. 4.

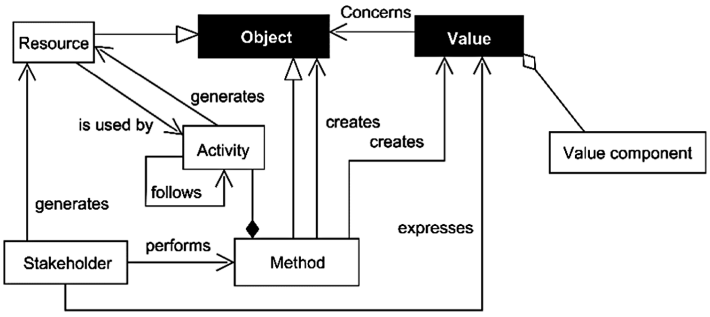

Fig. 4. Value creation method metamodel

Dimension 3: Object concerned by the value

The object concerned by the value corresponds to elements (e.g., information, process, tool, or actor) that exist in a specific environment represented at the information system level by the context. The latter has an influence on the type and the amount of value associated with this object, for instance, a *customer's browsing history* is an object of a data type that has a particular pecuniary value for an airline travel agency that can estimate the value ascribed to a flight ticket for a customer. This value is calculated based on the number of times this flight ticket is viewed on the company's website by the customer. At the opposite, this *customer's browsing history* is not an object of value on a drugstore website with fixed prices. Complementarily, it is also worth noting that this context has no impact on the nature of the value. For example, privacy in healthcare is defined in the same way with the same characteristics as in industry.

To collect and deal with the concepts that are necessary to model the object of value, it has been assumed that each sector of activities, should it be manufacturing, finances, or healthcare, to name a few, is associated with a specific information system. The latter models the objects composing it as well as the relationships between these objects, using a dedicated language.

Sector-specific information systems and enterprise architecture (EA) models and languages are good approaches here because they semantically define generic objects and sometimes concrete languages to express these objects. Numerous frameworks have been designed to model IS and EA of various sectors, e.g., Cimosa [43], ArchiMate® [44], BSE [45], DoDAF [47], and many others (Table 3).

Table 3. Objects of value within an information system

Reference/language	Object concerned		
	Context	*Information system*	*Example of objects*
CIMOSA [43]	Production industry	Industrial information system	Business process, flow, step, function, information, resource and organization aspects, business user, control, capability…
ArchiMate® [44]	Enterprise	Enterprise information system	Service, actor, role, process, function, contract, software, data, capability, role, device, node…
BSE [45]	Enterprise	Business service ecosystem	Service, capability, resource, process, actor…
Demo [46]	Enterprise	Business process, information systems	Models (interaction, business process, action, interstriction, and fact), actor, action…
DoDAF [47]	Military	DoDAF meta-model (DM2)	Guidance, activity, capability, resource, performer, location, information, project materiel, system, service, organization…
ARIS [48]	Enterprise	Business process management	Data, function, organization, material, IT resources, or machine resources…

Regarding the financial case study, the data of the bank's customers is the object concerned by the required privacy (generated by the bank) and concerned by the required availability (generated by the datacenter).

Table 3 provides an overview of some metamodels and languages used to depict the context targeted, the IS under scope, and some examples of objects addressed.

As a summary and according to this analysis, the concepts defining the context and the object concerned by the value are:

- **Information system.** The information system encompasses, and **is composed by,** the objects concerned by the value and the stakeholders that benefit from the value created.
- **Context.** The context represents the surrounding of the IS. It includes (1) the constraints on the system in which the value is created and (2) the definition of the borders of this system (e.g., the sector and the sector purpose of the business entity that is concerned by the IS, the rules and regulations related to the sector or the IS, the institutional arrangements, etc.). Accordingly, the context is associated to the information system with an association named **characterizes**. As stated in [28], the context also allows selecting the *performance components* […] necessary *to define the scope of the performance evaluation problem*. Hence, this selection defines a particular context, or viewpoint, for the evaluation of the value. To model this, the

concept of context is associated to the measure with a relation named **influence**. Regarding the case study in the financial sector, the context is the *financial regulation.*

Based on the above definitions, the object concerned by the value is modeled in Fig. 5.

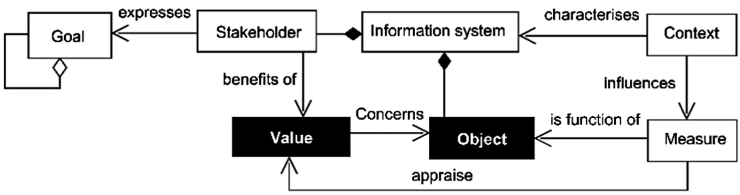

Fig. 5. Object concerned by the value metamodel

Integrated model instantiated to the financial sector case study

In the previous sections, three aspects related to the creation of value have been successively presented and each of these dimensions has been modeled in a dedicated metamodel. All along the description of the concepts, illustrations have been provided regarding a case of IT outsourcing in the financial sector, namely, the archiving of a bank customer's data to a datacenter. Figure 6 presents the integrated value creation metamodel instantiated to the IT outsourcing case.

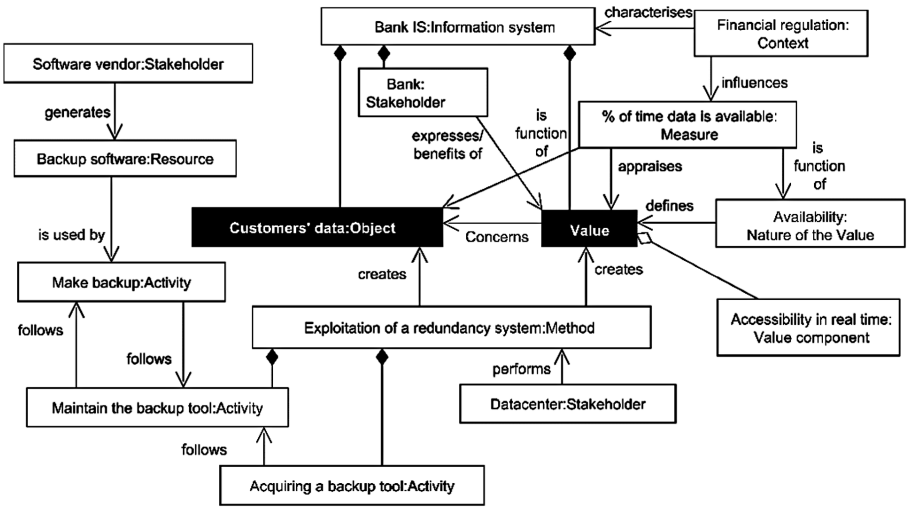

Fig. 6. Value creation metamodel instantiated to IT outsourcing in the financial sector.

4.2 Modeling Value Cocreation as a Specialization of Value Creation

As reviewed in the state of the art, no model for representing the creation of value following the three value dimensions exists yet. This observation is even more relevant for the field of value cocreation, in the context of the service-dominant logic, when two or more stakeholders cogenerate value during business exchanges.

In this respect, as explained in the introduction, the paper proposes an innovative value creation metamodel aiming to support the cocreation of value in the field of knowledge-intensive business services engagements. This section aims to analyze to what extent the value creation metamodel (Figs. 3, 4 and 5) is suitable to model the processes of value cocreation in KIBS proposed in [1] (Fig. 7). To that end, as also explained previously, one specificity of value cocreation is that value is cocreated on the basis of a collaboration between many stakeholders who have different responsibilities during the cocreation, including the generation of the appropriate resources needed for cocreation activities. Consequently, a prerequisite before modeling the value cocreation was to enrich the value creation model with the concepts of the stakeholder and the resources. This improvement was achieved in previous section by integrating the value creation model presented in [14] with the value model proposed in [28, 29].

These processes and generative mechanisms of value cocreation in KIBS engagements are illustrated in Fig. 7. Only the processes dedicated to the alignment within and between actors are considered in the following.

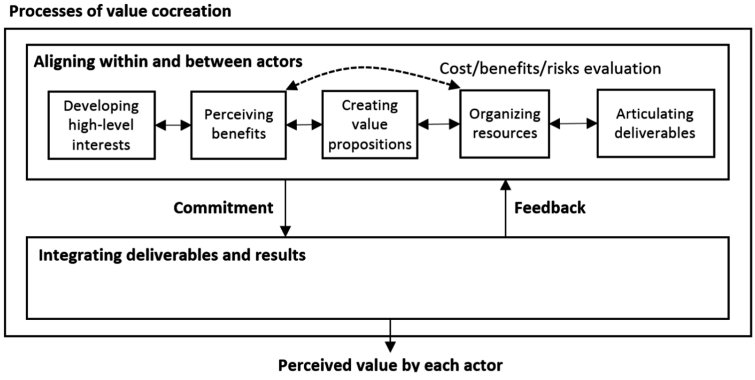

Fig. 7. Processes of value cocreation in KIBS engagements (adapted from [6])

In the following, the value creation metamodel is specialized concept by concept, as illustrated in the object diagram of Fig. 8:

- **Object.** In the cocreation metamodel, like in the creation metamodel, the object concerned by the value is the element from the information system that aims to be better off after that value is proposed and accepted. In the case of the cocreation of privacy between the bank and the datacenter, the object concerned by the value is still the *customer's data* at the bank's side and the *service portfolio* at the datacenter's side.

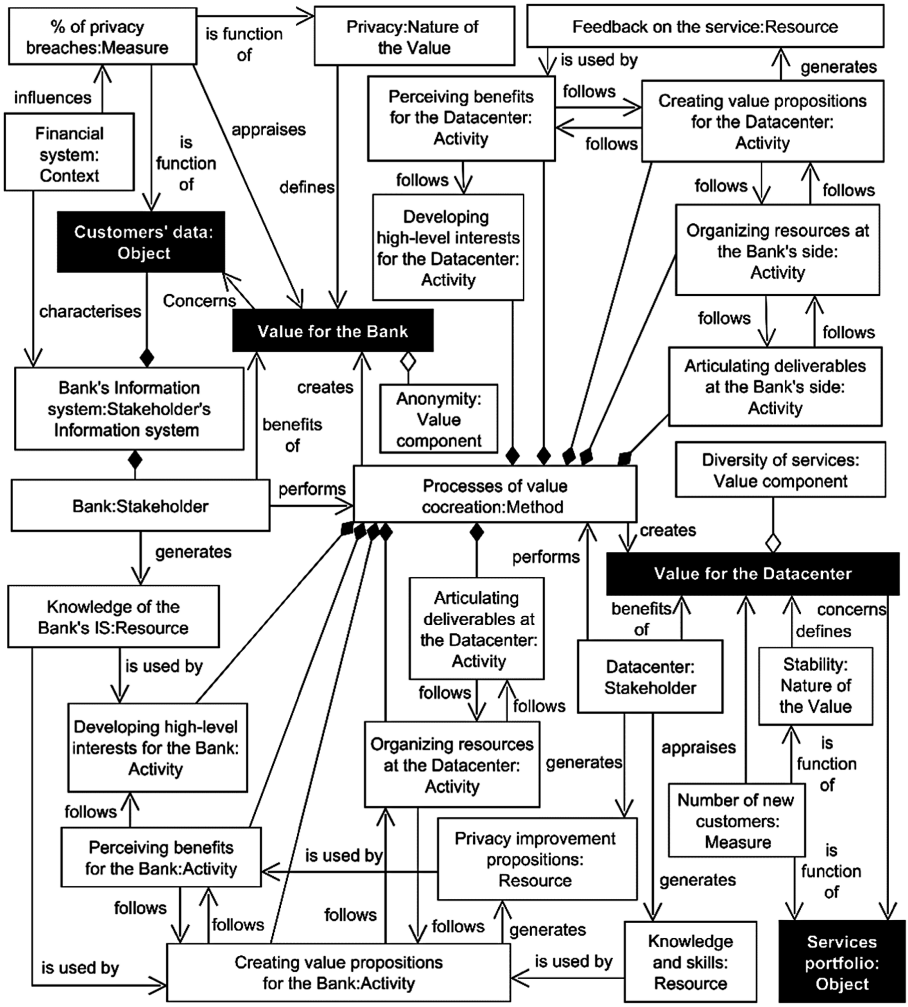

Fig. 8. Specialization of the VCMM to the value cocreation in KIBS engagements

- **Context.** The context of the cocreation is also equivalent to the one from the creation metamodel. In the case study, this context is the *financial regulation.*
- **Nature of the value.** The nature of the value defines the value generated by the creation or the cocreation. In the case study, this nature of the value is *privacy* (for the bank) and *stability* (for the datacenter).
- **Stakeholders.** They are the entities performing the method that cocreates value, who benefit from this value, and who generate the resources used by the method activities. These stakeholders are of three types in the field of KIBS: companies, their customers, and partner organizations. In the case study, the stakeholders are the *bank* and the *datacenter.*

- **Information system.** This concept is not addressed in the processes of value cocreation [6]. However, to keep the specialization of the metamodel coherent, a specialization of the information system is created and named: *Stakeholder's information system*. In the case study, an instance of the information system is the *Bank's information systems.*
- **Value component.** This concept expresses the different elements that constitute the value. At the case study level, an instance of the privacy component is the *anonymity* and an instance of the stability component is the *diversity of services.*
- **Measure.** The measure that appraises the level of privacy is the *% of privacy breaches* and that appraise the level of stability is the *number of new customers.*
- **Method.** The approach followed in [6] to cocreate value is a process-based approach. The first process related to the need for alignment among KIBS actors and the second concerns the integration of the deliverables and results. The case study only focuses on the first part and considers that the integration of the deliverables and results may be achieved similarly. Regarding the case study, the method used to design the *privacy of the bank customers data* could be composed of the same activities as the ones that compose the process of value cocreation in KIBS engagement, namely: *developing high level interest for the bank and the datacenter to cocreated the audit of the leased-line*, *perceiving benefits at each side*, *creating value proposition from the bank and the datacenter business*, etc.
- **Activity.** To be achieved, the method is composed of activities that are articulated with each other. The mechanisms within the processes of value cocreation are considered as a specialization of the concept of activity. Five of them (from the aligning process) are represented in Fig. 8: *developing high-level interests, perceiving benefits, creating value propositions, organizing resources and articulating deliverables*, at the bank's and datacenter's sides.
- **Resource.** According to the definition, a resource is a type of object used by an activity. Many types of resource are needed for the realization of the activities of the value cocreation processes. Example of resources here include: the *privacy improvement propositions* used to perceive benefits, the *knowledge and skills* used to create value propositions, the *knowledge of the bank IS* used to create value propositions, at the bank's side, and *feedback on the services*, at the datacenter's side.

5 Discussion of Findings

The analysis achieved in previous section allows elaborating a new design iteration of the value creation metamodel previously presented in paper [14]. The design of this new iteration is motivated and oriented by the needs to enhance (1) the relations between the value created and the stakeholder that generates and benefits from it, and (2) the relations between the method, the activities composing it, and the resources that are created and used by the latter. Compared to the previous version, this new iteration offers the following advantages:

- It allows expressing the role of each stakeholders involved in the value creation, more especially, it allows expressing who is responsible to perform the method that creates value and who benefits from it. This improvement is mandatory to appraise the value generated and, as a result, to improve the level of that value. Indeed, according to [28], value is appraised in function of the degree of satisfaction of a set of stakeholders' expectations or needs. This set of expectations is introduced in the value creation metamodel by means of the relation: *Stakeholder expresses goal*.
- It allows expressing the resources that are necessary to achieve activities composing the value creation method, but mostly, it allows expressing which resource is generated by which stakeholder during the cocreation activities and which resource is itself generated by the cocreation. For instance, this improved version allows expressing that a stakeholder shares personal information with a service provider in exchange of a service.
- The importance of the context is consolidated by a relation expressing that the context influences the measure of the value created.

The second contribution of the paper lies in specializing the value creation metamodel to the processes of value cocreation in knowledge-intensive business services proposed by [6]. The resulting advantages of that specialization are manifold but mostly, it demonstrates that considering the processes of value cocreation as a type of value creation is justified and, as a result, that cocreation may be handled, at the modeling level, as a specific type of value creation. Acknowledging this, modeling the value cocreation as a type of value creation allows integrating additional elements in the expression of the cocreation, among which:

- The information system. It gathers the elements that are impacted by the (co) creation of value and that are characterized by the context in which this (co)creation happens. These characteristics are, e.g., the IS composition, its structure, the business sector in which it evolves, etc. The cocreation of value impacts the characteristics of the information system. For instance, the cocreation of value generates new collaborations that must be integrated in business processes supported by the information system. These new processes may generate new information that also needs to be managed by the IS (e.g., accessed by the stakeholders, exploited by method, or stored in databases). This impact is not represented in the metamodel.
- The context is a particular type of element that characterizes the information system. Knowing this context is important for the cocreation of value because it may generate constraints to be considered during the design of the cocreation (e.g., regulation, sectorial requirement, institutional arrangement, etc.).
- The stakeholders. They are key players in the cocreation and are the ones that will be better off after value is delivered. First, clearly modeling the stakeholder that performs the cocreation of value is an important management requirement [52] for instance during the assignment of responsibilities [31]. Second, knowing the beneficiary of the value is preponderant to assess the latter, considering that value is evaluated on the basis of the beneficiary's satisfaction [28].
- The resource. Modeling the resources involved in the cocreation is beneficial because this allows expressing the input required for this cocreation. Indeed,

resources often play an important role in cocreation, like when a resource is an information based on which value is created by means of a data mining method. The resource is also an element that may be generated by an activity of the cocreation method and that is worth representing. For example, on Fig. 7, Perceived benefits is an information (resource) generated by the process Perceiving benefits (activity) that is used by the process Valuing during the integration of the deliverables and results.

The metamodel is mainly elaborated based on the review of frameworks from the information system domain [30–48], completed with elements from the scientific literature [3–6, 8–12, 16–23]. This limitation concerning the scope of the domain analyzed is a source of weaknesses for the metamodel, which is currently only valid for use in this area. As a result, further work is necessary to verify the option that has been chosen to ground the metamodel based on the three specific dimensions and to consolidate the latter according to the nature of values, methods, and objects considered in other domains (e.g., healthcare, industry, etc.).

6 Conclusion

The contribution of our research is an enriched version of the value creation metamodel (language) [14] with the concepts of stakeholder and resource [28], and a specialization of that metamodel to the cocreation of knowledge-intensive business service (KIBS). In comparison to the state of the art, despite the impressive amount of literature aiming to explain the concepts and mechanisms of value cocreation, no language has been expressed yet to support the exchange of information related to value cocreation between information system designers. Consequently, this paper contributes in conceptualizing such a language considering three dimensions: the nature of the value, the method of value creation and the object concerned by the value.

The practical implication of our modeling approach is the consideration of four additional elements during the design of value creation and cocreation models: (1) the stakeholder and its role in the processes, (2) the resource and its utility, (3) the information system that is influenced by the cocreation, and (4) the context that dictates the constraints and the institutional arrangement in which the cocreation arises. The impact of the later has been demonstrated in the field of KIBS, but it could also be demonstrated with other types of organizations or businesses, for instance, in finance or information security [53].

Given the limitations of the metamodel (cf. Sect. 5), we intend to further integrate the performance dimensions of the “performance evaluation methodology for decision support” in industrial projects [28] and the process of alignment within and between actors from the value cocreation process described in [6]. We also want to improve the alignment between the concept of measure from the value creation metamodel and (i) the evaluation of the cost/benefits and risks during the alignment within and between actors, and (ii) the outcome and quality metrics of the integration of deliverable and results process from [28]. Finally, the elaboration of the metamodel being performed in the frame of an iterative design approach, further validation of the latter is still

expected, in real settings and in view of concrete business collaborations. This may require the development of a concrete syntax (textual or graphical) for the language corresponding to the metamodel.

References

1. Lessard, L., Okakwu, C.P.: Enablers and mechanisms of value cocreation in knowledge-intensive business service engagements: a research synthesis. In: 2016 49th Hawaii International Conference on System Sciences, HI, USA, pp. 1624–1633. IEEE Computer Society (2016)
2. Muller, E., Doloreux, D.: What we should know about knowledge-intensive business services. Technol. Soc. **31**(1), 64–72 (2009)
3. Bettencourt, L.A., Ostrom, A.L., Brown, S.W., Roundtree, R.I.: Client co-production in knowledge-intensive business services. Calif. Manag. Rev. **44**(4), 100–128 (2002)
4. Vargo, S.L., Lusch, R.F.: Evolving to a new dominant logic for marketing. J. Mark. **68**(1), 1–17 (2004)
5. Vargo, S.L., Lusch, R.F.: Service-dominant logic: continuing the evolution. J. Acad. Mark. Sci. **36**(1), 1–10 (2008)
6. Lessard, L.: Modeling value cocreation processes and outcomes in knowledge-intensive business services engagements. Serv. Sci. **7**(3), 181–195 (2015)
7. Westergren, U.H.: Opening up innovation: the impact of contextual factors on the co-creation of IT-enabled value adding services within the manufacturing industry. Inf. Syst. e-Bus. Manag. **9**(2), 223–245 (2011)
8. Zeithaml, V.A.: Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J. Mark. **52**(3), 2–22 (1988)
9. Payne, A.F., Storbacka, K., Frow, P.: Managing the co-creation of value. J. Acad. Mark. Sci. **36**(1), 83–96 (2008)
10. Hastings, H., Saperstein, J.: A practice-driven service framework for value creation. In: 15th Conference on Business Informatics, pp. 145–152. IEEE (2013)
11. Storbacka, K., Brodie, R.J., Böhmann, T., Maglio, P.P., Nenonen, S.: Actor engagement as a micro foundation for value co-creation. J. Bus. Res. **69**(8), 3008–3017 (2016)
12. Frow, P., Nenonen, S., Payne, A.F., Storbacka, K.: Managing co-creation design: a strategic approach to innovation. Br. J. Manag. **26**(3), 463–483 (2015)
13. Ziemba, E., Kolasa, I.: Risk factors relationships for information systems projects – insight from polish public organizations. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 55–76. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_4
14. Feltus, C., Proper, H.A.E.: Conceptualization of an abstract language to support value co-creation. In: Proceedings of the FedCSIS. ACSIS, vol. 11, pp. 971–980 (2017)
15. Grönroos, C.: Service logic revisited: who creates value? And who co-creates? Eur. Bus. Rev. **20**(4), 298–314 (2008). https://doi.org/10.1108/09555340810886585
16. Chandler, J.D., Vargo, S.L.: Contextualization and value-in-context: how context frames exchange. Mark. Theory **11**(1), 35–49 (2011). https://doi.org/10.1177/1470593110393713
17. Chew, E.K.: iSIM: an integrated design method for commercializing service innovation. Inf. Syst. Front. **18**(3), 457–478 (2016)
18. Blaschke, M., Haki, M.K., Riss, U., Aier, S.: Design principles for business-model-based management methods—a service-dominant logic perspective. In: Maedche, A., vom Brocke, J., Hevner, A. (eds.) DESRIST 2017. LNCS, vol. 10243, pp. 179–198. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59144-5_11

19. Gordijn, J., Akkermans, H., van Vliet, H.: Business modelling is not process modelling. In: Liddle, S.W., Mayr, H.C., Thalheim, B. (eds.) ER 2000. LNCS, vol. 1921, pp. 40–51. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45394-6_5
20. Weigand, H.: Value encounters – modeling and analyzing co-creation of value. In: Godart, C., Gronau, N., Sharma, S., Canals, G. (eds.) I3E 2009. IAICT, vol. 305, pp. 51–64. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04280-5_5
21. Razo-Zapata, I.S., Chew, E.K., Proper, E.: Visual modeling for value (co-)creation. In: 10th International Workshop on Value Modeling and Business Ontologies, Trento, Italy, paper 6 (2016)
22. Alves, H., Fernandes, C., Raposo, M.: Value co-creation: concept and contexts of application and study. J. Bus. Res. **69**(5), 1626–1633 (2016)
23. Ranjan, K.R., Read, S.: Value co-creation: concept and measurement. J. Acad. Mark. Sci. **44** (3), 290–315 (2016)
24. Galvagno, M., Dalli, D.: Theory of value co-creation: a systematic literature review. Manag. Serv. Qual. **24**(6), 643–683 (2014)
25. Hevner, R., March, S.T., Park, J.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
26. 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 (2008)
27. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS Q. **35**(1), 37–56 (2011)
28. Li, F., Etienne, A., Siadat, A., Vernadat, F.: A performance evaluation methodology for decision support in industrial projects. In: Proceedings of the 7th IESM Conference, HTW SAAR, Germany (2017)
29. Li, F.: Performance evaluation and decision support for industrial system management: a benefit-cost-value-risk based methodology. Ph.D. thesis, Arts & Métiers Paritech, France (2017)
30. Matulevicius, R., Mayer, N., Heymans, P.: Alignment of misuse cases with security risk management. In: 3rd International Conference on ARES, pp. 1397–1404. IEEE (2008) https://doi.org/10.1109/ares.2008.88
31. Feltus, C., Petit, M., Dubois, E.: Strengthening employee's responsibility to enhance governance of IT: COBIT RACI chart case study. In: 1st ACM Workshop on Information Security Governance, pp. 23–32. ACM (2009)
32. Calero, C., Ruiz, J., Piattini, M.: Classifying web metrics using the web quality model. Online Inf. Rev. **29**(3), 227–248 (2005)
33. Foorthuis, R.M., Hofman, F., Brinkkemper, S., Bos, R.: Assessing business and IT projects on compliance with enterprise architecture. In: GRCIS 2009. CEUR-WS, vol. 459, paper 6 (2009)
34. Feltus, C., Grandry, E., Kupper, T., Colin, J.N.: Model-driven approach for privacy management in business ecosystem. In: 5th Modelsward 2017, pp. 392–400. SciTePress (2017)
35. Langheinrich, M.: Privacy by design—principles of privacy-aware ubiquitous systems. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 273–291. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45427-6_23
36. OMG: Value Delivery Metamodel, Version 1.0. OMG Document formal/2015-10-05 (2015)
37. Dix, A.: Human-computer interaction: a stable discipline, a nascent science, and the growth of the long tail. Interact. Comput. **22**(1), 13–27 (2001)
38. Bénaben, F., Touzi, J., Rajsiri, V., Truptil, S., Lorré, J.P., Pingaud, H.: Mediation information system design in a collaborative SOA context through a MDD approach. In: MDISIS 2008. CEUR-WS, vol. 340, pp. 89–103 (2008)

39. Becker, H.: Social Impact Assessment: Method and Experience in Europe, North America and the Developing World. UCL Press, London (1997)
40. Ralyté, J.: Towards situational methods for information systems development: engineering reusable method chunks. In: Proceedings of the 13th ISD. Advances in Theory, Practice and Education (2004)
41. Daneva, M.: Applying real options thinking to information security in networked organizations. No. TR-CTI. Centre for Telematics and Information Technology, University of Twente (2006)
42. Manuj, I., Mentzer, J.T.: Global supply chain risk management. J. Bus. Logist. **29**(1), 133–155 (2008)
43. Berio, G., Vernadat, F.: Enterprise modelling with CIMOSA: functional and organizational aspects. Prod. Plann. Control **12**(2), 128–136 (2001)
44. Lankhorst, M.M., Proper, H.A., Jonkers, H.: The architecture of the ArchiMate language. In: Halpin, T., Krogstie, J., Nurcan, S., Proper, E., Schmidt, R., Soffer, P., Ukor, R. (eds.) BPMDS/EMMSAD 2009. LNBIP, vol. 29, pp. 367–380. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01862-6_30
45. Feltus, C., Grandry, E., Fontaine, F.X.: Capability-driven design of business service ecosystem to support risk governance in regulatory ecosystems. CSIMQ **10**, 75–99 (2017)
46. Dietz, J.L.G.: Understanding and modelling business processes with DEMO. In: Akoka, J., Bouzeghoub, M., Comyn-Wattiau, I., Métais, E. (eds.) ER 1999. LNCS, vol. 1728, pp. 188–202. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-47866-3_13
47. U.S. DoD: DoDAF framework, version 2.02 (2010). https://goo.gl/WCA46s
48. Scheer, A.-W., Nüttgens, M.: ARIS architecture and reference models for business process management. In: van der Aalst, W., Desel, J., Oberweis, A. (eds.) Business Process Management. LNCS, vol. 1806, pp. 376–389. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45594-9_24
49. Cox, A.: Business relationship alignment: on the commensurability of value capture and mutuality in buyer and supplier exchange. Supply Chain Manag. **9**(5), 410–420 (2004)
50. Nyman, J.: What is the value of security? Contextualising the negative/positive debate. Rev. Int. Stud. **42**(5), 821–839 (2016). https://doi.org/10.1017/S0260210516000140
51. Korkman, O.: Customer value formation in practice: a practice-theoretical approach. Ph.D. thesis, Swedish School of Economics and Business Administration, Sweden (2006)
52. Ziemba, E., Eisenbardt, M., Mullins, R.: Use of information and communication technologies for knowledge sharing by Polish and UK-based prosumers. In: Ziemba, E. (ed.) AITM/ISM 2016. LNBIP, vol. 277, pp. 49–73. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53076-5_4
53. Feltus, C., Proper, H.A.E.: Towards a security and privacy co-creation method. In: Proceedings of 12th IEEE International Conference for Internet Technology and Secured Transactions (ICITST 2017), Cambridge, United Kingdom (2017)

Semantic Approach to Financial Knowledge Specification - Case of Emergency Policy Workflow

Jerzy Korczak(✉), Helena Dudycz, Bartłomiej Nita, and Piotr Oleksyk

Wrocław University of Economics, Wrocław, Poland
{jerzy.korczak, helena.dudycz, bartlomiej.nita, piotr.oleksyk}@ue.wroc.pl

Abstract. The article presents an approach to integrate a business process knowledge in Decision Support Systems. The main research findings are related to the ontological and procedural issues of knowledge specification. The mathematical rigor used in process descriptions guarantees for precise definition of concepts and relationships in the domain knowledge. It concerns three major aspects of the system design, i.e. formalization of processes predefined in Business Process Modeling Notation, reuse of a domain ontology, and analysis of economic and financial information. Formally specified analytical processes and ontology allowed considerably minimize sources of ambiguity and confusion in the system design and implementation. The described approach is a continuation of the development of the intelligent cockpit for managers (InKoM project), whose main objective was to facilitate financial analysis and evaluation of economic status of the company. The current work and case studies are focused on specification of static (structural) and procedural knowledge of financial analysis in Small and Medium Enterprises. The content of the knowledge covers essential financial concepts and relationships related to the processes of emergency policy. An experiment has been carried out on real financial data extracted from the financial information system.

Keywords: Business Process Modeling · Ontology · Financial analysis
Specification of processes · Emergency policy

1 Introduction

Today Decision Support System (DSS) that should go hand in hand with the knowledge of managers and methods of financial analysis. Surveys have shown that managers of SMEs often do not possess solid background knowledge in financial analysis and IT technology needed to strengthen their competitive position on the market and maintain financial credibility. This problem is often caused by the lack of the knowledge required not only to correctly interpret economic data, but also to properly carry out analytical processes. Moreover, these studies have shown difficulties encountered by managers in skillful use of information systems containing too many functions and tools often exceeding their competencies.

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 24–40, 2018.
https://doi.org/10.1007/978-3-319-77721-4_2

Taking into consideration all managerial requests and complexity of business issues, there is a need for solutions that integrate managerial knowledge in DSSs and support intelligent analysis and decision making [1]. One of the main obstacles for automation of analytical processes within current DSS is the lack of a formal representation of the procedural knowledge within models of business processes. In most of the systems, the operations within the processes are defined diagrammatically, in natural language or pseudo Pascal notation making this representation very informal and ambiguous. Consequently, reasoning tasks and computation are very limited.

The essence of financial analysis is to address the various problems associated with current short-term decision making as well as long-term strategic planning. Both types of decisions made concern the level of debt. Determining the level of debt below which it does not jeopardize business continuity is difficult. Debt level threshold should be set individually for each type of company [2]. Excessive debt often has dire consequences of any company, as it often constitutes a reason for abandoning important business activities as well as does not allow for launching new development activities. Lenders shied away from financing companies that have high level of liabilities in their balance sheet. Lenders' unwillingness to finance debt is often the first signs of looming bankruptcy [3]. Liabilities include necessary sources of business financing, however, they also constitute a serious cause of financial risk that could drive a company to bankruptcy [4]. The most important threat associated with liabilities is the necessity to generate free cash flows; available cash is used to repay accrued liabilities. The lack of ability to timely service current liabilities forces managers to seek other solutions. One of the possibilities is the conversion of liabilities into equity. However, such a transaction could generate a new type of risk: a threat of hostile takeover.

The aim of the paper is to propose an approach that integrates financial knowledge, analytical models, and business reasoning. The analytical models as well as business reasoning rules are known in the literature and can be easily encoded. In the project, it is assumed that financial knowledge is formally defined by domain ontology. The essential part of the work is to propose a solution facilitating automated analysis of information available in financial databases and reports.

The idea of the project is partially inspired by works on modeling business process knowledge, notably by Smith and Proietti [5], De Nicola et al. [6]. In the project, the process of analysis, defined in the Business Process Model and Notation (BPMN) is extended by the domain ontology in OWL and a process knowledge encoded in Business Process Abstract Language (BPAL)[1].

The paper has been structured as follows: the first part briefly introduces the economic and technological background of financial analysis where analytical activities are described in form of a business process diagram in BPMN, enriched by the financial ontology; foundation of automation of analytical processes and the design methodology of process-oriented ontology are discussed in subsequent two sections. The presentation is focused on the formal aspects of procedural knowledge specification. In Sect. 4, the use case is detailed using real life data extracted from a financial

[1] BPAL was developed within EU FP7 *Business Innovation and Virtual Enterprise Environment* (BIVEE): modeling of production processes in manufacturing oriented networked-enterprises.

information system. The analytical activities are specified formally in a language close to BPAL and illustrated by analysis of financial data. The whole analytical process has been broken down into subprocesses, activities, tasks, and completed by the required information resources. Case study specifications focus on one of the key issues of financial analysis related to the processes of emergency policy. Finally, some conclusions are drawn and the future of the project discussed in the last section.

2 Foundations of Automation of Analytical Processes

Managers are obliged to continually evaluate ability of a company to operate in the future, particularly to keep an eye on the company's performance in order to as quickly as possible avoid excessive debt. Negative financial signs should serve as the basis to launch emergency policy. Business practice has developed many ways of reducing liabilities [7]. Among the most common restructuring activities is to strive to eliminate excessive financial liabilities. As a first step, expenses not related to the core business (e.g. purchase of prestigious assets that are related to public relations) should be reduced. In this case, most of the social security benefits for employees are waived. R&D expenses should also be temporarily suspended.

Key Performance Indicators (KPI) analysis is a comprehensive method used by financial analysts to evaluate financial standing of a company and to support managers in their decision making processes. This is due to the fact that set of financial indicators and metrics allow for multifaceted evaluation of the validity and effectiveness of financial activities. In the course of calculating financial ratios, it is possible to compare liabilities with various items included in the balance sheet and in the profit and loss statement as well as cash flow reports. The most commonly used indicator in assessing company equity is *Debt to Equity Ratio* (computed as ratio of *Total Liabilities* to *Equity*). In general, the lower the ratio the better financial position of a company. This ratio is not only used by banks, but by potential investors too. Hence the reason managers want to maintain company equity in excess of financial obligations or to systematically decrease liabilities while simultaneously increasing capital.

The analysis of the *Debt to Equity Ratio* should be concluded by an analysis of the dynamics of liabilities and analysis of changes in the structure of assets. Rise in debt is not desirable; however this may be justified in some cases. For instance, if a company enters into new liabilities due to acquisition of new fixed assets, such assets could generate future profits and financial surplus. Generated profits may be capitalized while surpluses may repay be used to repay liabilities. Consequently, this can significantly improve the *Debt to Equity Ratio*. To assess the likelihood of such a scenario, it is necessary to analyze the changes in asset structure. An increase in liabilities arising from increase in fixed assets and intangible assets is a relatively positive signal. On the other hand, if the main trend is towards an increase in receivables and inventory then this is a sign of financial threat. It usually implies a company has problems with sales and money collection.

The process of financial analysis can be represented as a workflow graph describing the correct sequence of operations, where each operation involves concepts, data items, and the relations between them. To model the analytical operations, Business Process

Modeling and Notation (BPMN) can be applied (http://www.omg.org). Usually, a BPMN model is defined through a Business Process Diagram (BPD), which is a kind of flowchart incorporating constructs to represent the control flow, data flow, the work to be assigned to the participants, and handling of exceptions. In the project, a process of analysis is composed of an interrelated set of subprocesses or activities, where an activity is formed by sub-activities or tasks. A task is an atomic element that cannot be further broken down.

The schema of financial analysis is defined using the BPAL platform. The focus of the study is on the procedural financial knowledge related to one of the key processes: the launch of Emergency Policy. Figure 1 illustrates a main BP Diagram referring to analysis of financial situation of a company. The central panel shows the BP Modeling View, comprising an editor and a set of tools (shown on the right) used to model BP diagrams using the BPMN notation. The left panel of the screenshot lists the resources available in the workspace, including BP models and ontologies. A description of OWL ontology that will be further detailed is found on the right.

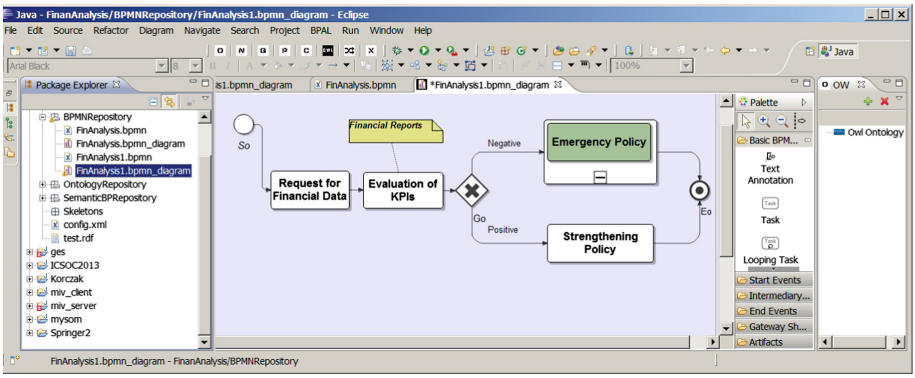

Fig. 1. Diagram of processes of financial data analysis

The process starts with a request for financial data. Detailed information about the sources and market signals will be explained in Sect. 4. Upon receipt of data of the subprocess of KPI's, evaluation starts, among other things, the task of assessing *Debt to Equity Ratio* discussed in the paper. Analysis of the dynamics of liabilities, dynamics of fixed assets under construction are also needed as well as the expenditures on future intangible assets. The branching point, called the gateway G_0, indicates exclusive choice of one of the two processes: Emergency Policy or Strengthening Policy. For instance, if *Debt_to_Equity_Ratio* value exceeds 200%, then this can be treated as the first warning signal to perform the Emergency Policy. The debt analysis is completed by an investigation of the causes of the increase in liabilities.

In general, increase in liabilities may be considered positive if it is intended to purchase new fixed assets or intangible assets. It is also important to check changes in receivables. Lack of increase in fixed assets indicates a significant problem. Usually, increase in liabilities is driven by the need to provide the company with cash. In most

cases, lack of funds is related to lack of receipts. Such a situation requires managers to take corrective actions. Subprocesses of the Emergency Policy will be detailed in Sect. 4.

The workflow model of financial analysis describes the process enactment, however, it does not contain information on the domain knowledge. For this reason, therefore, the financial ontology has been added to the project. The ontology provides semantic annotations of the entities, objects, items and pre- and post-conditions involved in the sub-processes. Ontology Web Language (OWL) was applied in order to define the semantics.

OWL is syntactically layered on RDF. The underlying data model (derived from RDF) is based on statements (or RDF triples) of the form <*subject; property; object*>, which allow for defining a resource (subject) in terms of named relations (properties). Values of named relations (i.e. objects) can be URIrefs of Web resources or literals, i.e. representations of data values.

In literature many research projects show that creating a financial ontology is advantageous in decision making process [1, 8–11]. For our purpose, an ontological framework was designed to represent the area of knowledge of process of emergency policy. The ontology of this study was construed using the approach presented in [12, 13].

Financial ontology has been encoded using the Protégé platform (http://protege.stanford.edu/) in the project. Figure 2 presents a sample visualization of business knowledge focused on the issue of excessive debt. There are two panels on the screenshot. The panel to the left the taxonomic relations, while the one to the right allows for visualization of taxonomic and semantic relations between defined topics (semantic network visualization). There are two types of lines between topics: (1) the solid line represents a relation *subclass-of* and (2) the dashed line represents the experts' defined relationships (for example: *depends on*) on the figure. The key elements of the presented ontology in Fig. 2 are the measures used to assess the appropriate level of financial situation of the company. Visualization in the OntoGraf module in the program Protégé allows for navigation between topics in a highly interactive manner. Interesting nodes can be put in the foreground with zooms, pans, and rotations. In Fig. 2, topics in rectangles which are important in the analysis of threats are highlighted. In this interactive visual process, the manager is able to subsequently concentrate on the interesting elements via filtration and focusing (zooming in). The presented part of the ontology guarantees for correct solution while helping avoid potential risk of misinterpretation. The manager can also add, modify as well as retrieve concepts related to the problem at hand. For example, by expanding *the Debt to Equity Ratio* branch of the semantic network, a manager can see that *the* value of *Debt to Equity Ratio* is determined by *Equity* and *Liabilities* (relation: *depends on*), but *the Debt to Equity Ratio* has influence on *Operational risk* and a three situations: *Hostile* takeover, *Bankruptcy* and *Loss of control*. The manager can see that he should analysis both *Supplementary capital* and *Share capital* (the parts of *Equity*), and *Short term liabilities* as well as *Long term liabilities* (the parts of *Liabilities*).

By means of the ontology, the manager could undertake various actions aimed at reducing operational risk. Failures in the process of analysis may lead to corporate bankruptcy. Reduction of excessive indebtedness can be done either through conversion of liabilities into equity or through issuance of new shares. However, such activities

could also become a source of additional risks associated with a change in ownership structure coming by way of hostile takeover or loss of control over the business.

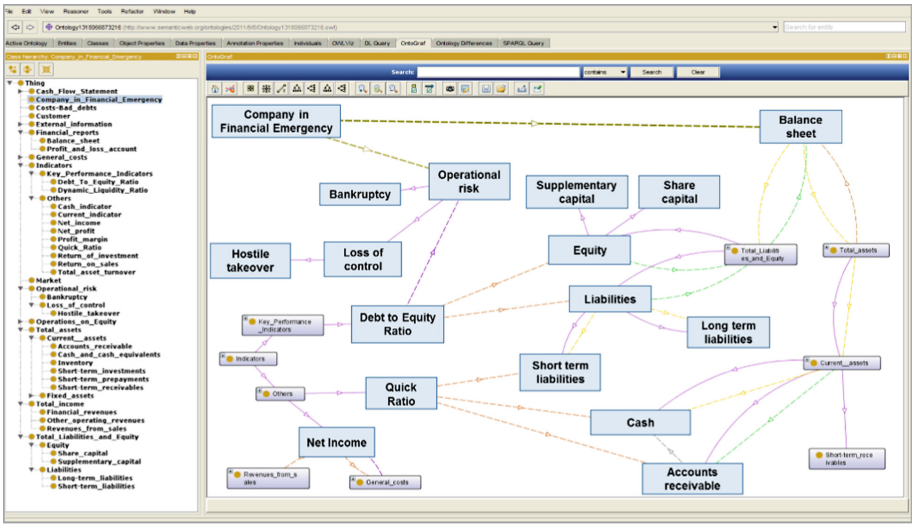

Fig. 2. Part of financial ontology

Business data usually contain a lot of explicit as well as hidden relationships that make their usage difficult. To correctly interpret the values of financial indicators, many measures and ratios that either directly or indirectly influence the final result need to be examined. Explicit visualization not only makes it easy to better understand and interpret indicators, but it also contributes to finding explanations of current values of indicators. Use of ontologies within analytical tools can help in defining business rules in order to get proactive information in the decision-making process.

3 Design of Process Oriented Ontology

Design of process oriented ontology has to provide not only a concise, comprehensive description of business processes, but also to express the semantics of processes in a formal way to be understood both by humans and the computer. In our project, the existing ontology of financial knowledge was enhanced by dynamic and procedural structures of business processes. Note that the diagrammatic representation of BPD is insufficient to be translated into a system that would be able to automatically execute all these analytical processes. It should be also pointed out that many other tasks such as retrieval, verification, or process composition have to be done manually.

There are several languages to describe business processes, for example UML, BPMN, BPEL, PSL, OWL-S, WSML, WISMO, etc. [14–18]. Taking into consideration a rigorous mathematical basis as well as close links with BPMN and modeling facilities, Business Process Abstract Language (BPAL) has been chosen to specify procedural knowledge in processes of data analysis [19, 20].

BPAL provides a number of modeling concepts, symbols, and rules used to define the so-called abstract processes. Syntax and semantics of BPAL constructs can be found in [19, 20]. Looking at specifications of BPAL processes, one may say that there are many similarities to the constructs in BPMN, however BPAL is not just a diagrammatic notation. In general, BPAL Application Ontology is a collection of validated BPAL processes with respect to the BPAL Axioms, where Axioms represent the rules and constraints related to application processes. In our case study, the process of financial analysis has been broken down into subprocesses. A subprocess is a set of logically integrated activities along with concepts, rules, and relations related to them. An activity that represents a unit of work performed within the subprocess can be atomic or compound, where an atomic activity, which is an action that cannot be broken down further, is called a task. Business concepts in an application are defined using unary and relational predicates, called BPAL Atoms.

The process illustrated previously with BPMN can be defined as BPAL abstract specification as follows:

act *(request_for_financial_data),*
act *(analysis_of_KPI),*
act *(evaluation_of_KPI),*
act *(strengthening_policy),*
act *(emergency_policy)*
prec *(request_for_financial_data, analysis_of_KPI),*
prec *(analysis_of_KPI)*
assert*:* ***adec*** *(evaluation_of_KPI),*
prec *(evaluation_of_KPI, strengthening_policy),*
prec *(evaluation_of_KPI, emergency_policy)*
    msg *(reporting_of_KPI_ratios),*
    msg*(policy recommendations)*
part_of *(Balance_Sheet, Assets),*
part_of *(Balance_Sheet,Equity)*
isa *(KPI, Debt_to_Equity_Ratio)*

The key words in the specification have the following meanings [1]: ***act*** represents a business activity, ***prec*** indicates a precedence relation between activities, ***assert*** an assertion, ***and msg*** a message sent and received. ***isa*** is a specialization relation, ***part_of*** – aggregation relation. An exclusive branch ***adec*** leads to the execution of exactly one successor, while an inclusive branch leads to concurrent execution of a non-empty subset of its successors. The set of successors of exclusive or inclusive decision points may depend on conditions that usually take the form of tests on the value of the items that are passed between activities.

The specification describes also physical and informational items produced and consumed by the various activities during execution of a process.

Formally, the created model of business processes has to contain not only a set of ground facts, predicates, but also a set of rules. The rules define among other things: hierarchical relationships within the BPAL predicates, relationships between BPAL elements, properties, and item flow relations. A model of a process should respect

a number of constraints related to representation of activities, events, conditions, and exception handling [20]. Design of analytical processes illustrating these concepts will be detailed in the next section.

4 Use Case

Improvement of the financial situation of a company should constitute the main objective of any manager. In the use case, the analytical process of emergency policy is presented and illustrated by real data extracted from financial information systems. The example is based on general schema of processes related to analysis of financial data (Fig. 1). For the purposes of the study, Emergency Policy process was broken down into seven subprocesses as shown in Fig. 3.

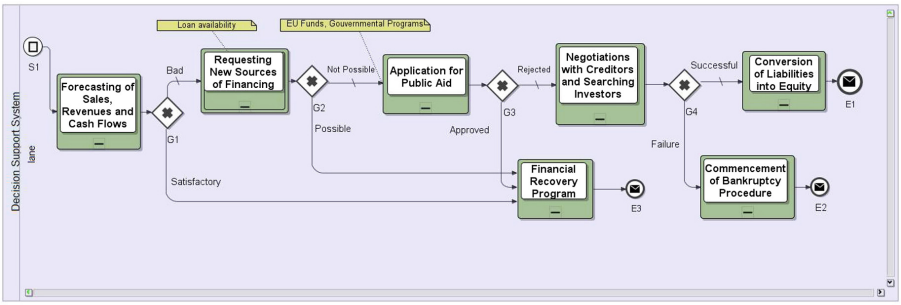

Fig. 3. Diagram of subprocesses of emergency policy

Each of the indicated subprocesses can be described, they comprise input, preconditions, activities, exception handling, postconditions, and output. Due to the limitations of the article length, only selected aspects of process modeling are described.

The first subprocess deals with a request for financial data, in particular the data available in the financial statements. To illustrate the case, the basic information describing the financial situation of the company over five years is shown in Table 1, where the data for 2017 and 2018 represent the predicted values.

The most important components presented in the table point towards the deterioration of financial situation of the company. Moreover, the forecast indicates that there is less probably of any improvement in the future. Among the most important signs of the deteriorating financial situation are:

- losses in all periods (170,000 in 2014, 400,000 in 2015 and 90,000 in 2016),
- gradual reduction of supplementary capital (600,000 in 2014, 430,000 in 2015 and 30,000 in 2016),
- gradual decrease in long-term liabilities pointing to a chance of securing financing for development activities (400,000 in 2014, 360,000 in 2015 and 320,000 in 2016),
- constant growth of short-term liabilities (500,000 in 2014, 900,000 in 2015 and 1,300,000 in 2016),
- gradual decline in free cash (600,000 in 2014, 200,000 in 2015 and 80,000 in 2016).

Table 1. Selected financial information (in k PLN)

Specification	2014	2015	2016	2017	2018
Share capital	600	600	600	600	600
Supplementary capital	600	430	30	0	0
Net income	−170	−400	−90	−260	−370
Long-term debt	400	360	320	280	240
Short-term debt	500	900	1 300	1 380	1 500
Short-term receivables	400	600	800	800	700
Investments and cash	600	200	80	80	20

After obtaining the requested information, the second subprocess focused on evaluation of critical KPIs can begin. The most important measure used in this analysis is *Debt to Equity Ratio* (see Fig. 2). Additionally, liquidity ratios and other measures describing corporate financial stability can be also applied; these include among others:

- the profit dynamics compared with average macroeconomic and market figures,
- the structure and dynamics of fixed assets,
- the dynamics of sales revenue in relation to the increase in receivables as well as the apparent dynamics of losses in relation to bad debts,
- all types of liquidity ratios, e.g. current ratio, quick ratio and cash ratio.[2]

The system provides the data and internal reports which serve as input to the '*Evaluation of KPIs*' process. The data included in the report refers to the most important measures of financial stability. It should be noted that the manager may personally determine the content of the report, or he could also take advantage of the default solution. Assessment of KPIs is a key component of the decision making process, however various commonly used financial analysis may be also be applied.

Subprocess '*Evaluation of KPIs*'[2] can be specified as follows:

input: *Financial Data (share capital, supplementary capital, net income, long-term debt, short-term debt, short-term receivables, and short-term investments including cash)*
preconditions:
debt structure ratio = total liabilities/total assets
debt structure ratio>70%
quick ratio=(accounts receivable + cash) / short term liabilities
quick ratio <1
activities
assessment of debt structure ratio,
assessment of debt to equity ratio,
assessment of liquidity ratios
exception handling: *lack of data values*
postconditions: *negative or positive assessment of financial situation of company*
output: *KPI Evaluation Report*
Recommendation to begin Emergency policy or Strengthening Policy

[2] In the literature there are target ratios discussed. It is generally considered necessary to maintain a ratio of equity to liabilities of 1:1. The quick ratio should be maintained at a minimum level of 120%.

To enrich of a BP with the financial ontology several kind of annotations can be done. Annotations are expressions specifying the preconditions under which processes can be executed. For example, assume that the concept of company is in a financial emergency state is defined in the ontology as *CompanyInEmergency*. So then the activities of *Emergency Policy* are launched when the following expression is true:

(*CompanyInEmergency*
⊓ (∃ *hasDebtToEquityRatio.* integer[> 200])
⊓ (∃ *hasIncreaseInLiabilities*. true)
⊓ ((∃ *hasNoIncreaseInFixedAssets*. true) ⊔ (∃ *hasIntangibleAssets*. true))
⊓ (∃ *hasIncreaseInReceivables*. true)) (*c*)

where *hasDebtToEquityRatio* is an integer-valued data property and *hasNoIncreaseInFixedAssets*, *hasIntangibleAssets*, *hasIncreaseInReceivables* are boolean-valued data properties.

In addition, postconditions such as positive and negative effects of process execution can be defined. A manager can browse hierarchy of ontology concepts, relationships, and annotations.

The forecast could be used to provide warning signals. The values presented in Table 1 (last 2 columns) indicate that the projected *Debt to Equity Ratio* (>200%) significantly exceeds the safety level[3]. Summing up the projected values points to a need to take radical steps aimed at recovering from the poor financial situation of the company.

The most important outcome from this process is the recommendation with respect to the financial standing of the company. In the analyzed company, *Debt to Equity Ratio* of 200% indicates extremely high financial leverage. Thus, the situation of the company is definitely perceived as highly unfavorable, because the company is likely to have serious problems with servicing its financial obligations in the future.

Taking into account the necessity to complete the process, the manager has two options: if *Debt to Equity Ratio* is at an acceptable level, the company should go back to normal activities, however, if the ratio exceeds 200%, then the procedure for reducing financial risk should be initiated. The system based on *Debt to Equity Ratio* analysis generates a signal indicating excessive debt. Such a high level of debt to equity ratio is a clear sign of an extremely unfavorable financial situation of the company.

This suggestion triggers the start of subsequent subprocess called '*Forecasting of sales revenues and free cash flows'*. Sales forecasting as well as the company's ability to generate excess cash are both essential elements of overall company management

[3] According to Table 1, in 2016 *Debt to Equity ratio = [(1300 + 320)/(600 + 30 − 90)] * 100 = 225%.*

A company projected to encounter excessive debt needs to search for corrective solutions. The basis for this subprocess is a report containing financial data and KPIs as described in the previous subprocess.

Subprocess *Forecasting of sales revenues and free cash flows*
***input**: Financial statement data, KPIs values*
***preconditions**:*
debt to equity ratio >200%
***activities**:*
- preparation of the forecast financial statements: the amount of future liabilities, potential minimum amount of free cash flow (linear forecasting model)
- estimation of the debt to equity ratio on the forecast financial statement
- assessment of forecast ratios
***exception handling**: registration of new contracts aimed at increasing sales revenue*
***postconditions**: forecast debt to equity ratio <200%*
***output**: Satisfactory or Bad result of forecasting*
Recommendation to begin Financial recovery program
or Requesting of new sources of financing

Information included in the forecast does not permit for introduction of a financial recovery program and its normal operation, hence it is necessary to look for new solutions created in the subprocess '*Requesting of new sources of financing*'. Acquisition of new sources of external financing depends on prevailing situation in the financial sector. The most common solution is to apply for bank loans. If this is not possible, attempts could be made to issue bonds. However, companies that do not generate profits have significant difficulties with providing bondholders. Loans from credit institutions such as banks could also be considered. However, they are usually very expensive, and require additional security such as a mortgage. This could lead to a buyout of such receivables by competitors with hidden agenda of taking control over the company. This subprocess requires a the manager's decision. Thus it is necessary to obtain information on banking offers, especially on the terms and conditions of bank loans.[4]

[4] Detailed analysis of banking credit conditions leads to search for other solutions. Banks use credit rating as a basis for credit decisions. The results and forecasts presented in Table 1 show that the surveyed company has no ability to repay the loan, and its credit standing is also poor. The rating methodology identifies companies whose liabilities significantly exceed their equity and systematically suffer a loss and thus are "permanently incapable of meeting their liabilities". The root causes of the low credit rating is the data included in Table 1, in particular, growing liabilities and a significant loss. Any bank, following the credit rating methodology, will consider such a result as the basis of issuing a negative credit decision.

Sub-process: *Requesting new sources of financing*
input
Bank and Credit Institution list of offers and financing conditions
Report on corporate credibility
preconditions: unsatisfactory financial forecast
activities:
- analysis of offers and financing conditions
- checking of financial standing of company[4]
- selection of the best offers
- request of manager decision (approve the offer or reject)
exception handling*: obtain of short-term credit lines (3 months)*
postconditions*: acceptance or rejection from bank or credit institution*
output*:Information about the possibility to obtain a credit*
Recommendation to begin Financial recovery program
or suggestion to Apply for public aid

Before issuing an loan, every bank will specify their terms of loan conditions: interest rate, loan period, additional charges, and safety. Analysis of bank offers is weighed against the financial situation of the company. It is necessary to verify the terms of loan in the context of the company's credibility. Securing a loan from a financial institution may entail assignment of certain rights such as mortgages. This is a typical tool of loan security used with respect to unprofitable companies. However this is not rational because it can lead to rapid loss of control over the company. Neither is the possibility of receiving additional funding profitable, as companies in poor financial condition always have their interest rate increased. A willingness of a credit institution to issue a loan could also be a concealed intent of a hostile takeover. If there is more than one suitable offer, a manager has to choose the most favorable loan offer and send the application to the selected bank.

In the case of a negative decision from the bank, the system initiates the next subprocess '*Applying for public aid*', e.g.

- co-financing of new job positions in areas threatened by structural unemployment,
- equipping companies with pro-ecological technologies that eliminate CO2 emissions,
- support for companies employing the disabled,
- subsidized loans used to finance innovative technologies.

As in the previous subprocess, it is necessary to check all available aid funds. Specification of this subprocess is as follows:[5]

[5] Acquisition of aid funds is followed by a complex administrative procedure. Contests are not always announced on a continuous basis and not all entrepreneurs can apply for a particular call. Aid funds very often place additional requirements with regard to the need to transfer ownership rights to the effects of a project's implementation. Moreover, material liability associated with inappropriate use of funds is a problem for managers. Entering the contest may ultimately result in the bankruptcy of the business as a result of the sanctions entered in the contest rules.

Sub-process *Applying for public aid*
input:
EU programs list, governmental funds and programs list, local government funds, and programs list
Application criteria to be met
Financial data report
preconditions: *available financing from banks and other financial institutions*
activities:
- analysis of public aid[5]
- checking of application criteria
- request for manager decision to choose public aid program
- preparing and submitting a project application form
exception handling: *financial aid to adapt the company equipment to meet the needs of people with disabilities.*
postconditions: *acceptance or rejection of submitted application*
output: *Information about Approved or Rejected application*
Recommendation to begin Financial recovery program
or Negotiations with creditors and Searching investors

Financial aid is a solution that only works for a limited number of companies. External institutions offer additional aid funds based on very strict criteria. The manager has to analyze the available opportunities after feeding the system with the required data related to public assistance. If the system generates a support program tailored to the needs of the enterprise, the manager would be asked to prepare and submit applications. If the application is rejected, then it is necessary to execute the next subprocess.

Unfortunately, due to poor financial situation, it may be that a company cannot receive an external financial aid. The basic requirement for applying for public aid is the necessity to present the corporate performance report as well as forecasts for future financial situation (analogous reports as presented in Table 1). Usually evaluation of financial condition in projects co-financed with aid funds is not as restrictive as is the case with banks. However, an enterprise with excessive indebtedness and continuous losses will be eliminated in the first selection phase. The reason is the low score in the section "ability to carry out the project". Poor financial situation, especially if public aid is not available, excludes a company from the possibility of obtaining external financing in form of bank loans. Thus, the last step will be focused on negotiations with creditors with the aim of conversion of debt to equity. Negotiations with creditors and searching investors constitute the last chance for a company to survive. The subprocess can be specified as follows:[6]

[6] Creditors who decide to convert debts into debtor's equity analyze their ability to recover the invested funds. In case of legal insolvency proceedings, they rarely recover more than half of the funds involved. If they decide to convert debt into equity they must be prepared for the possibility of losing the whole. Therefore, the decision to convert is preceded by hard negotiations. A common requirement is to favor shares that will be issued by reducing debt. Existing owners will not always accept these terms, as this can result in loss of control.

Subprocess: *Negotiations with creditors and searching investors*
input:
list of creditors
list of creditors' requirements
list of potential investors
proposal of financial restructuring
preconditions: *not available aid funds*
activities:
- *analysis of creditors' requirements*[6],
- *preparing and submission of restructuring program*
- *negotiations with creditors*
- *preparing the contract on conversion of liabilities into equity*
- *if required searching new investors*

exception handling: *ownership changes of creditors*
postconditions: *list of creditors, list of investors*
output: *Results of negotiations;*
in case of success, recommendation to convert liabilities into equity
in case failure, searching investors
results of searching investors
in case success Financial recovery program
in case of failure, Commencement of bankruptcy procedure

The starting point in any negotiation process is to recognize creditors' requirements. Creditors may expect to acquire a specific number of shares in return for debt reduction. Creditors generally expect the managers to restructure such a company. These two elements serve as the basis for starting the process of conversion of liabilities into equity. An increase in equity could be effected by converting liabilities into shares, i.e. by issuing shares in order to next hand them over to creditors. This method of raising equity is particularly important under difficult financial situation of a company that is unable to meet its obligations or is at risk of bankruptcy [21]. Repayment of liabilities to creditors in the form of shares may contribute to a certain stabilization of the entity. Likewise, any other form of equity increase will, in this case, allow the company to recover, and remain in the market. If the conditions for conversion are accepted by both parties, then parties can then enter into relevant agreement. Negotiations with creditors are the last chance of avoiding bankruptcy.

The creditors' agreement for the conversion of liabilities into equity is the one of possibilities for the company to survive [22]. As shown in the above example, conversion of liabilities into equity should be immediately negotiated even if this is unfavorable for the current owners. Based on the data provided, the most advantageous solution is the conversion of liabilities into equity. In the analyzed case, all creditors are only interested in recovering their debts. In case the creditors are not interested in such arrangements, the second solution is to search for external investors. It is crucial that this decision be followed by an analysis of potential threat of some hostile takeover. This threat does not occur in the analyzed company because its liabilities are spread among various suppliers and banks. They agree to the conversion of debt into equity because it is the only way for them to close these transactions without a loss. If there are no other opportunities to reduce excessive debt, then the company is likely to legally go bankrupt.

The presented use case illustrates an approach to support the processes of Emergency Policy. Integration of external sources of information with the contextual internal data is a way of minimizing uncertainty in the decision process. Identification of analytical activities as well as assigning a minimum set of information is a relatively new approach to analytical workflow automation. The conducted preliminary study may serve as the basis for the use of a process-oriented methodology in financial analysis.

5 Conclusion and Future Works

The aim of the paper to enrich the financial knowledge by presenting formally specified business processes has been achieved. The approach was applied to financial analysis in real-world scenarios. It was made possible to merge the procedural and ontological perspectives as well as to express process-related knowledge by using standard modeling languages such as BPMN, and OWL in the design, and for reasoning and validation using BPAL platform. The mathematical regime used in process descriptions guarantees for precise definition of concepts and relationships in the domain knowledge. Providing formally specified analytical processes and ontology made it possible to considerably minimize sources of ambiguity and confusion.

Currently there are many process modeling notations such as BPMN, EPC, XPDL, and Petri nets. From the functional perspective, they should be interoperable, in order to overcome heterogeneities of different formalisms and map them to one common, machine-interpretable, process ontology. In addition, the solutions should provide interrelations to existing domain knowledge as well as enable query and search facilities.

Giving consideration to a formal view of BPMN, BPAL seemed to us an appropriate choice to automate the use of business process knowledge. The formally written BPAL specifications can be further automatically translated into executable programs BPEL. Once specification of analytical processes and financial ontology is terminated, an analyst using BPAL is in a position to verify the correctness of activities enactment and check whether semantic-constraints are satisfied. If the specification is sound, the analyst can use the reasoning services and retrieve information from the knowledge repository of business processes.

We are convinced that the abstract language of process specification provides a declarative and procedural semantics that can be interpreted, processed, and executed.

Results of experiments carried out are encouraging and reveal practical usability and acceptance by business experts. Processes of emergency policy are frequently carried out intuitively. SME managers do not have the necessary knowledge to integrate financial analysis, market trends, and banking issues. Determining the right moment for a company to implement corrective actions requires complex analysis, and it is not a straightforward procedure. Sequence of actions that should allow a company to return to normal business activity is a multi-faceted undertaking that dependent on individual company. There are many bankruptcy prediction methods based on sophisticated mathematical models presented in the literature. However, they are not very adapted to SME.

An important element is also an identification of the risks that may arise in the course of implementation of corrective actions. Correct integration of selected financial analysis indicators and data from the closest business environment is possible with expert knowledge. Use of appropriate managerial tool is very important for SME. The solution proposed in the article make it possible to pick the appropriate moment for recommending and implementing corrective actions. In addition, it allows managers to choose the best solution based on a company's prevailing situation and external factors. An important element is also an identification of the risks that may arise in the course of implementation of corrective actions.

From the financial perspective, the presented use case leads to the conclusion that the conversion of liabilities into equity should be carried out when there is a dominant owner with high potential for raising capital. The reduction of financial liabilities is highly desirable in order to improve financial standing, however, it is not something that has to be done at all costs.

Further research work will be focused on a comprehensive process-oriented approach to problem solving in enterprises. This would not have been possible without the formally defined knowledge of experienced managers and financial analysts. We are convinced that process oriented approach can contribute to better coordinate of activities spanning contextual and functional boundaries of decision support systems.

Acknowledgement. The authors would like to thank Maurizio Proietti from National Research Council, IASI Rome, for his comments and assistance to run the analytical processes using BPAL platform.

References

1. Korczak, J., Dudycz, H., Nita, B., Oleksyk, P., Kaźmierczak, A.: Attempt to extend knowledge of decision support systems for small and medium-sized enterprises. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, vol. 8, pp. 1263–1271. Annals of Computer Science and Information Systems (2016). https://doi.org/10.15439/2016f181
2. Beaver, W.H., Correia, M.: Financial Statement Analysis and the Prediction of Financial Distress. Now Publishers Inc., Breda (2011)
3. Altman, E.I., Hotchkiss, E.: Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt, 3rd edn. Wiley, Hoboken (2006)
4. Beck, T., Dermirguc-Kunt, A.: Small and medium-size enterprises: access to finance as a growth constraint. J. Bank. Financ. **30**(11), 2931–2943 (2006)
5. Smith, F., Proietti, M.: BPAL: a platform for managing semantically enriched conceptual process models. In: Cunningham, P., Cunningham, M. (eds.) eChallenges e-2014 Conference Proceedings IIMC International Information Management Corporation (2014)
6. De Nicola, A., Lezoche, M., Missikoff, M.: An ontological approach to business process modeling. In: 3th Indian International Conference on Artificial Intelligence, pp. 1794–1813 (2007)
7. O'Brien, T.J., Schmid, K.L., Hilliard, J.: Capital structure swaps and shareholder wealth. Eur. Financ. Manag. **13**(5), 979–997 (2007)

8. Aruldoss, M., Maladhy, D., Prasanna Venkatesan, V.: A framework for business intelligence application using ontological classification. Int. J. Eng. Sci. Technol. **3**(2), 1213–1221 (2011)
9. Cheng, A., Lu, Y.-C., Sheu, C.: An ontology-based business intelligence application in financial knowledge management system. Expert Syst. Appl. **36**(2), 3614–3622 (2009). Part 2
10. Korczak, J., Dudycz, H., Dyczkowski, M.: Design of financial knowledge in dashboard for SME managers. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, vol. 1, pp. 1111–1118. Annals of Computer Science and Information Systems (2013)
11. Neumayr, B., Schrefl, M., Linner, K.: Semantic cockpit: an ontology-driven, interactive business intelligence tool for comparative data analysis. In: De Troyer, O., Bauzer Medeiros, C., Billen, R., Hallot, P., Simitsis, A., Van Mingroot, H. (eds.) ER 2011. LNCS, vol. 6999, pp. 55–64. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24574-9_9
12. Dudycz, H., Korczak, J.: Process of ontology design for business intelligence system. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 17–28. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_2
13. Dudycz, H., Korczak, J.: Conceptual design of financial ontology. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, vol. 5, pp. 1505–1511. Annals of Computer Science and Information Systems (2015). https://doi.org/10.15439/978-83-60810-66-8
14. Abramowicz, W., Filipowska, A., Kaczmarek, M., Kaczmarek, T.: Semantically enhanced business process modeling notation. In: Smolnik, S., Teuteberg, F., Thomasal, O. (eds.) Semantic Technologies for Business and Information Systems Engineering: Concepts and Applications, pp. 259–275 (2012)
15. Martin, D., et al.: OWL-S: semantic markup for web services. W3C Member Submission (2004). http://www.w3.org/Submission/OWL-S/
16. Born, M., Filipowska, A., Kaczmarek, M., Markovic, I., Starzecka, M.: Business functions ontology and its application in semantic business process modeling. In: Proceedings of the ACIS, pp. 136–145 (2008)
17. Fensel, D., Lausen, H., Polleres, A., de Bruijn, J., Stollberg, M., Roman, D., Domingue, J.: Enabling Semantic Web Services: The Web Service Modeling Ontology. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-34520-6
18. Calvanese, D., De Giacomo, G., Lembo, D., Montali, M., Santoso, A.: Ontology-based governance of data-aware processes. In: Krötzsch, M., Straccia, U. (eds.) RR 2012. LNCS, vol. 7497, pp. 25–41. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33203-6_4
19. Smith, F., Proietti, M.: Rule-based behavioral reasoning on semantic business processes. In: Proceedings of the 5th International Conference on Agents and Artificial Intelligence, pp. 130–143. SciTePress (2013)
20. Smith, F., Proietti, M.: Ontology-based representation and reasoning on process models: a logic programming approach (2014). https://arxiv.org/abs/1410.1776
21. Cagetti, M., De Nardi, M.: Entrepreneurship, frictions, and wealth. J. Polit. Econ. **114**, 835–870 (2006)
22. Coates J.C.: Mergers, acquisitions and restructuring: types, regulation, and patterns of practice. In: Oxford Handbook on Corporate Law and Governance. Discussion paper no. 781 (2014, Forthcoming)

Ontological Support for Process-Oriented Competency Management

Ilona Pawełoszek(✉)

Częstochowa University of Technology,
Al. Armii Krajowej 19b, Częstochowa, Poland
ilona.paweloszek@wz.pcz.pl

Abstract. Over the last few decades process-oriented approach has become an essential standard for many companies. This situation requires a new outlook on competency management. The competencies of employees should be considered in the framework of tasks implemented within the business process. According to the author's observations, the greatest challenge is the lack of a consistent view of the overall aspects of human resources in conjunction with business process flow. In particular there is a gap between the process models defined in BPM systems and the competency descriptions defined in HR systems. The aim of this paper is to propose a solution based on a on Software Competence Ontology designed by the author, which can be used to support process-oriented competency management in a software development company. The primary purpose of the solution is to facilitate finding the best contractors to perform the business process tasks. With this purpose in mind a method of comparing candidates' skills and knowledge was developed. The presented approach can be useful in process-based and virtual organizations where the environment is dynamic and frequent changes are needed to preserve the company's competitiveness and agility. An example proposal of practical implementation of ontology-enhanced business process model was presented and illustrated by the process of software configuration management.

Keywords: Competency management · Ontology · Process approach

1 Introduction

With the continuous advancement of technology labor market requirements are also constantly changing. This is particularly noticeable in the case of IT companies, their services need to be more adaptable, agile and flexible than ever before. Therefore the need for new competencies continuously shows up and competencies that are already in existence, change their definitions [1]. This is reflected in a steady growth of interest in competency management systems.

Due to the increased need of agility competency-based management is crucial activity of contemporary business organizations. In this situation information technology support is the core element of the processes such as recruiting the most appropriate candidates, effective planning of employee development programs and project management. In many cases the information about competencies is exchanged between collaborating organizations.

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 41–60, 2018.
https://doi.org/10.1007/978-3-319-77721-4_3

Competency management is not new and has been widely practiced by all sort of companies that develop formal approaches to ensure that they have qualified human resources to meet their business goals. The practice of competency management is based on identifying, defining and measuring differences between individuals in terms of specific work-related constructs, especially the abilities that are critical to successful job performance [2]. Once they are defined, each employee or potential contractor should be described according to the those formal definitions. The definitions can be used in many ways, such as forecasting and scheduling workforce, determining training goals and measuring achievement of these goals.

Traditionally, competencies are position-specific and based on functional role analysis [3]. However popularization of process approach as a standard necessitates changing the ways of defining and identifying requisite competencies, which should be considered in the framework of tasks implemented within the process rather than by position of the employee in organization hierarchy.

The IT tools for supporting process modelling should therefore provide possibility to view the business process from the perspective of competencies required to perform particular tasks. Provision of information describing needed competencies of individuals involved in the process can help not only in workforce planning for the particular process but also in other management tasks such as expert finding, personalization of career paths and staff trainings.

According to Business Process Management (BPM) practices, processes should be modelled and documented in a form of diagrams representing workflow chart. Documenting the business processes aids communication throughout the organization and it is important aspect of many popular quality management approaches such as Six Sigma, Kaizen and LEAN. Moreover business process modelling notation can be extended with some additional user-defined abstract layers describing some important aspects of business operations. Linking business process workflows with semantic descriptions of competencies needed to perform the process would allow for creating new innovative applications supporting decision making and planning in many HR-related areas. A formalized semantic description of a domain is called ontology. Ontologies can be used for creating intelligent applications exploiting automated reasoning engines, which by utilizing interrelations between entities can make intelligent choices in different situations within the domain [4].

According to the author's observations, the greatest challenge is the lack of a consistent view of the overall aspects of human resources in conjunction with business process flow. In particular there is a gap between the process models defined in BPM systems and the competency descriptions defined in HR systems.

The aim of this paper is to show how the ontology describing competencies can be linked with business process models to support process-oriented, dynamic competency management in a company. With this aim in mind a proposal of practical implementation of ontology-enhanced business process model was presented and illustrated by the example of a process form software development domain.

Since the subject of intelligent decision support in the converging areas of human resources and business process management is important to modern organizations there are many research efforts undertaken, which have been described in Sect. 2. The research methodology and foundations for development of the framework for

supporting process-oriented competency management are presented in the Sect. 3. Section 4 presents the Software Competence Ontology and a conceptual framework of a platform for supporting competence management in a software company. The usage scenario of the proposed platform is explained on an example of a process model of software configuration management. The conclusion and future work section outlines research contribution, implications for research and practice as well as directions for discussions and development of the proposed solution.

2 Research Questions and Theoretical Background

The process-based approach is now used routinely by organizations of all sizes, all over the world, in virtually every industry [5]. Any organization that undertakes a process redesign or undergoes an ISO 9000 certification must create some kind of process documentation [6]. However, beyond this formal requirement, business process models can be useful for multiple purposes other than their primary objective, which is workflow representation. Business process models are advantageous for organizing work, understanding problems and business requirements, communicating between team members and preparing documentation.

Business process notations are also flexible and are increasingly often embedded into enterprise systems as tools to represent real flow of process instances. The first such system developed in Poland is Macrologic Merit. The idea of this system is that all processes taking place in the company are presented in a form of diagrams in a real-time. This allows managers to monitor, measure, and modify processes they are running. Process flow and particular operations can be easily redesigned and adapted to needs of the company. System automatically reflects modifications for all employees involved in a process [7]. The idea of extending business process models by competency layer emerged during the author's experience with this innovative solution developed by Macrologic.

Since the aim of this study was to elaborate a conceptual framework, which would allow to connect business processes with information on competency requirements, the study focuses on addressing the following research questions:

Q1: Is there a need to connect competency requirements with business processes models?
Q2: Do some tools or conceptual frameworks already exist that offer functionality of connecting competency descriptions and requirements with business process models?
Q3: Which existing tools can be useful to develop a solution for supporting process-based competency management?

To answer the above questions, the main research methodology used was an exploratory review of achievements so far in related fields to identify the needs and gaps to be addressed, and to collect recommendations for designing. With regard to the first question many researchers and practitioners agree that human resource information systems have to support the process approach. "The company must align its infrastructure, such as information technologies and HR systems, to support the process" [5].

Process-oriented HR-systems incorporate job descriptions based on business process design and incentive systems that emphasize the process' needs. Process-oriented HR systems are identified as important for the maturity of the process orientation in an organization [8]. HR activities either have to generate the job descriptions directly from the processes or take the processes as basis for their task in order to not work on outdated information [9].

Especially in high-tech sector based upon continuously evolving personal skills which require long education and training, competence management is a strategic issue that requires careful planning. The strategic competence management issue must be considered in the frame of a dynamic, process-based view rather than defined in a static way, as structural attributes of actual or potential employees or groups of employees [10, p. 221].

As it can be concluded from literature review, there is a need to connect business process management tools with Human Resource management functions, particularly with competency requirements to support planning and development of human resources and to gain another perspective for business process analysis. An overview commercial offers of BPM software and enterprise systems shows that currently there are no solutions that would connect process models with competency requirements and at the same time allow for extracting competency requirements to integrate them with other systems, such as HR, planning, reporting and decision support.

Analysis of publications in the converged field of competence management and business process management showed that in the dynamic environment of contemporary organizations a Competency-Oriented Business Process Analysis [11] can be the right choice. There are some attempts aiming at enhancing modelling notations by additional information to offer the specific competency-oriented view of the process [12, 13].

The area of enhancing business process models by competency information requires resolving two basic issues:

1. Unfortunately none of the current business process modeling languages or tools support the characterization of the business process in terms of competencies. Therefore the first issue is to find appropriate notation to include competency data in process models, which should be precise, human-readable and allow for automatic processing.
2. The second challenge is to design a formal representation that would be enough expressive to provide the detailed view of the process from the competency perspective.

In the BPM area, the Business Process Modeling Notation (BPMN) is the de-facto standard approved by ISO/OSI [14]. BPMN allows for multi-view and high-level description of business processes. Business Process models are used to communicate a wide variety of information to a wide variety of audiences. However it does not offer standard support for the characterization of the business process in terms of many other specific aspects. These aspects often could be related to the area in which the process is executed, some formal regulations and standards that the process must comply with.

For example Rodríguez et al. [15] propose an extension for including data quality requirements in process models. Bocciarelli and D'Ambrogio describe a BPMN extension for modeling nonfunctional properties of business processes [16]. The extension of

BPMN facilitating security risk management was proposed by Altuhhova et al. [17]. A comprehensive review and classification of BPMN extensions can be found in [18]. As there is a lack of methodological guidelines and standardized approaches for creating extensions of the BPMN, the methods and tools were proposed by Stroppi et al. [19].

The tasks of HR department, such as evaluation and selection of staff or identification of training needs, are complex due to large volumes of scattered information, possibility of misinterpretation and lack of common standards. Therefore the need for ontological approach in competency management is recognized and appreciated by many researchers and practitioners. Ontological frameworks supporting competency management have been proposed by Sucre and Pernalete-Chirinos [20], Kimble et al. [21], Toader [22], Ortega-Gonzalez et al. [23], Janev and Vranes [24].

It should be noted that there are many current activities in the area of modelling and standardization of competency management. The initiatives are oriented on creating and maintaining different competency models, for example:

- IMS-RDCEO - The Reusable Definition of Competency or Educational Objective (RDCEO) specification provides a means to create common understandings of competencies that appear as part of a learning or career plan, as learning pre-requisites, or as learning outcomes. RDCEO provides unique references to descriptions of competencies or objectives for inclusion in other information models [25].
- HR-XML is a library of XML schemas developed by the HR-XML Consortium, Inc. to support a variety of business processes related to human resources management. The competencies schema which is a part of HR-XML allows for capturing of information about evidence used to substantiate a competency together with ratings and weights that can be used to rank, compare, and evaluate the sufficiency or desirability of a competency [26].
- InLOC [27] provides ways of representing intended learning outcomes, including knowledge, skills and competencies, so that the related information may be communicated between and used by ICT tools and services of all kinds, interoperably.
- O*NET [28] – is a database of all occupations in the US economy. It provides taxonomy of competencies and their elements and such as knowledge, skills, abilities and many other. The data was collected from companies operating in United States. The O*NET database can also serve as statistical tool to examine labor market in USA because it contains results of measurement of competency levels. O*NET-SOC can serve as a reference taxonomy of occupations.

The information models specified above can be used to exchange competency definitions between learning systems, HR-systems, repositories of learning content, databases of competency or skills, and other relevant systems. They can provide frameworks and guidelines for developers of detailed domain-oriented solutions.

The focus of this study is on software development domain therefore as a base for creating taxonomy for building Software Competence Ontology the following frameworks have been considered and examined:

- e-CF – European e-Competence Framework [29],
- SFIA – Skills Framework for the Information Age [30],
- ITCM – Information Technology Competency model [31],

- ITC ACM – Information Technology Competency Model Association for Computing Machinery [32],
- UNESCO ICT-CFT - UNESCO ICT Competency Framework for Teachers [33].

The frameworks mentioned above may provide inspiration for creating domain ontologies for specific applications supporting competency management in IT companies. After examining many criteria (such as scope, structure, granularity, orientation, license, compliance with European standards and available documentation) of each of the above frameworks, the European e-Competence Framework was selected. The e-CF encompasses the IT domain sufficiently to be applied in practice, and it seems most suitable for software development domain. The added value of the e-CF is also the fact that it is recommended by European Committee for Standardization. The motivation for creating Software Competence Ontology and its designing process and characteristics are presented in the next section.

3 Research Methodology

3.1 Foundations for Development of Software Competence Ontology

Ontologies are the most expressive form of knowledge representation. They describe a piece of reality as a collection of classes of entities and the relationships among them. Ontologies contain a controlled vocabulary or terminology and a semantic network that encodes the relationships between each term of the vocabulary [34].

As a starting point, a comprehensive literature review was undertaken to explore the approaches to development of ontology-based frameworks. On the basis of several research works [35–38] three main stages of the ontology building process were identified:

1. **Specification** is the process of defining goals, scope of ontology, and resources to be used (existing taxonomies, descriptions of central domain concepts) and use cases.
2. **Conceptualization** can be seen as the process of building controlled vocabulary of concepts, defining hierarchical relations and specification of additional elements such as instances, slots (non-hierarchical relations and properties), facets (restrictions and values).
3. **Formalization** means encoding the elements created while conceptualization in selected knowledge representation language.

Creating a good taxonomy from scratch can be difficult and time-consuming, so it is recommended to find some existing one to use it. In this study the European e-Competence Framework was chosen as a base for construction of Software Competence Ontology which is described in Sect. 4.1. Another good reason for using existing taxonomies is the need for standardization. Standards provide a solid foundation upon which to develop new applications and enhance interoperability of existing ones.

Codification of competencies is the area well-suited to exploit ontologies because competencies can be organized in a form or hierarchy that represents different levels of

abstraction – from very general areas to very detailed pieces of knowledge and skills. Different use cases of Software Competence Ontology are explored and clarified with examples in Sect. 4.

Probably the most widespread use of ontologies is for enhancing searching capabilities of software applications. Ontology of competencies can be important and prevalently used in a wide range of enterprise applications, particularly in the domains such as human resources, manufacturing, project management. By using the logical structure of concepts and their relations it is possible to search for people with similar or complementary competencies regarding the domain and the level of knowledge and skills. There is also a number of other potential applications of ontology of competencies such as analyzing text documents, websites and process models with the aim to identify elements of competency descriptions.

One of the major potential benefits of constructing ontologies is that they can be implemented in reasoning software. The analysis of the graph structure can reveal relations that are not explicitly specified anywhere else. Ontologies can be visualized in many ways. Mapping the ontology to visual form improves understanding of the domain knowledge and helps to reduce many weaknesses of management information systems [39]. Another advantage of the ontological approach is that knowledge is easy to maintain – insert new items, delete inappropriate ones, adding attributes and relationships.

3.2 Extending Business Process Models

In the business process management area, the Business Process Modeling Notation (BPMN) is the de-facto standard approved by ISO/OSI [14], which allows for multi-view and high-level description of business processes. BPMN provides means to describe collaboration, choreography and conversation aspects of business processes. However it does not offer standard support for the characterization of many other important aspects of the business process. These aspects are often related to the area in which the process is executed, some formal regulations and standards that the process must comply with.

BPMN2.0 offers extensibility mechanism for enhancing standard BPMN notation with user-defined attributes and elements. This extensibility feature allows for addition of new types of artifacts. Modeling tools may include features to hide, or show these artifacts. However the operations of adding the artifacts, hiding or showing them do not influence the sequence flow of the BPMN model. This is to ensure that BPMN diagrams always have a consistent structure and behavior [40]. The extension element imports the definition and attributes with their values do the BPMN diagram.

The BPMN2.0 extension element consists essentially of four different classes which are [41, p. 179]:

- ExtensionDefinition - defines additional attributes,
- AttributeDefinition - presents the list of attributes that can be attached to any BPMN element,
- AttributeValue - contains attribute value.

Adding new concepts to the model provides possibility to analyze it from different perspectives. From the point of view of competency management, BPMN models can be enhanced by artefacts representing competencies necessary to run the process. Such an extension would allow for establishing and populating competence requirements across the organization, its business partners and job candidates. BPMN models with references to the descriptions of the required competencies create yet another perspective for analyzing the process performance regarding human factor. Moreover having a unified model for description of competencies allows for addressing them on the stage of process design and further adjusting the process according to the current abilities of human resources.

4 Research Findings and Discussion

4.1 The Concept of Software Competence Ontology

In this study the European e-Competence Framework (e-CF) was used as a reference for formalizing and codifying employees' competencies. The choice was governed by many features of the e-CF, which make it suitable for applying it in software development domain. The e-CF is not based on job profiles but rather task-oriented. The process-based approach is more flexible and suitable for project-oriented companies (which is common in software industry) characterized by dynamic nature of the work environment, where employees are often from different departments and have different job titles. The e-CF provides general and comprehensive specification of e-competencies described in a multidimensional structure which consists of:

- 5 competence areas - derived from the general framework of ICT business process consisting in five phases: (A) plan, (B) build, (C) run, (D) enable, (E) manage,
- 40 competencies,
- 5 generic proficiency levels – each competence has a set of its specific proficiency levels. where 1 denotes the weakest knowledge or skills,
- knowledge and skills examples for each competency – represent the knowledge and skills expected or required. The examples are provided to add value and context and are not intended to be exhaustive [42].

The dimensions of e-CF can be customized into different contexts from ICT business. The classes and relations defined in Software Competence Ontology are presented on Fig. 1. The classes are:

- Person – a person can be an employee or a candidate for job.
- Competence – which is compound of knowledge and skills. A person may have evidenced knowledge and skills, which indicate competency in the given area.
- Knowledge – information acquired by the person.
- Skills – the ability act effectively and to apply knowledge to practice.
- Scale – is a set of reference values describing the level of skill or knowledge.
- Level – describes knowledge or skills presented in the particular scale.

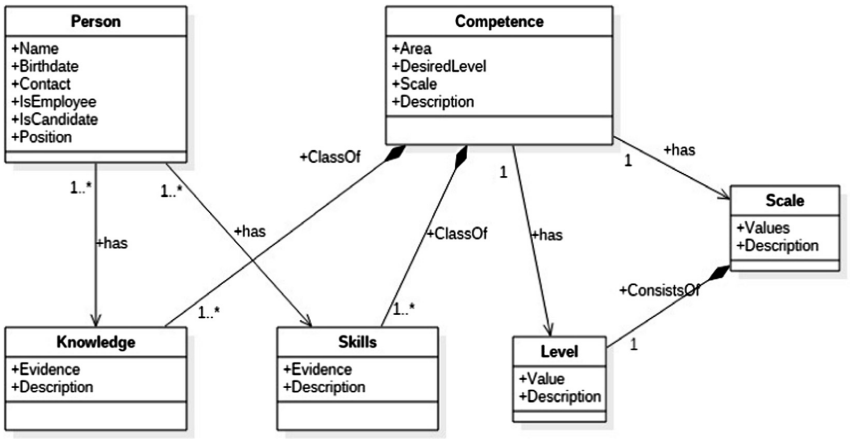

Fig. 1. Classes, attributes and relations of Software Competence Ontology.

Knowledge of the person can be evidenced by certificates, scientific and professional degrees achieved during studies, while skills are evidenced through practice, for example experience on a given position, taking part in projects. In the proposed framework based on Software Competence Ontology we assume that the person cannot have the competency which is not evidenced by knowledge or/and skills. So the competency is derived from evidenced knowledge and skills. Figure 2 presents examples of knowledge and skills for competency "User support".

One drawback of the e-CF model is the lack of level specification for detailed knowledge and skills elements. There is only desired level assigned to competencies. For example, the competence "A.1. IS and Business Strategy Alignment" has desired

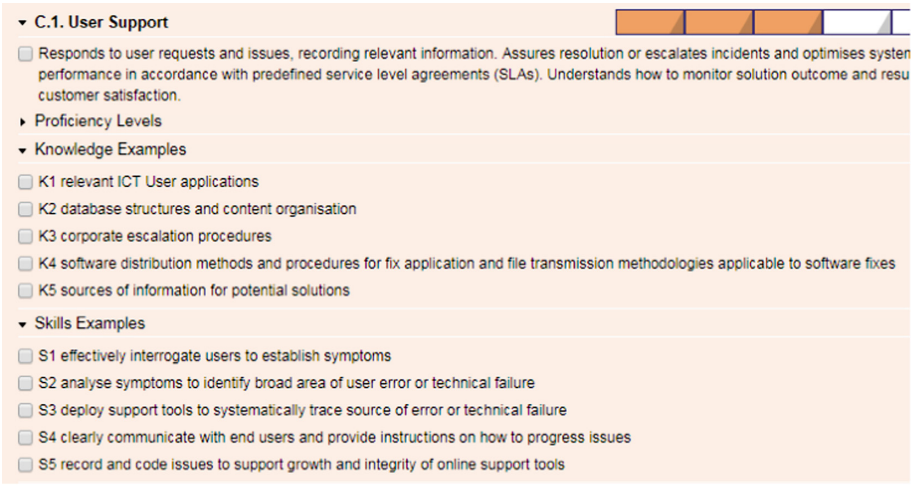

Fig. 2. Knowledge and skills for competency "User support" (source: Screenshot form Web application of e-CF [43]).

proficiency level 4 or 5 (in the 1–5 scale). Therefore to compare the evidenced skills and knowledge of a person with desired level of competency specified in the ontology it is necessary to make assumptions about the desired levels of knowledge and skills elements.

4.2 Evaluation Methods for Decision Support

Software Competence ontology is a body of knowledge describing areas of the domain, knowledge and skills, so it can be helpful in understanding and defining requirements, searching through document bases and making annotations on various documents and process models. However to use the ontology in decision support additional algorithms are needed. Particularly the issue is assessment of one's competency and making comparisons. The framework proposed here allows to measure and compare current competency levels (on the basis of documents, CV, profiles etc.) to make sure that the staff members have the expertise needed to add value to the business process.

The e-CF model contains a number of knowledge and skills examples, and the number is different for each competency. In practice it is often the case that the person does not have all the skills and knowledge defined for the given competency. Therefore there is a need for implementing some evaluation method. The solution to this issue is to calculate the rate of compliance, which represents to what degree the evidenced knowledge and skills of the person are consistent with the definition of competency.

In the case of "User support" competency there is 5 elements of skills and 5 elements of knowledge. However it is also important to take into account the balance between knowledge and skills and their importance. In the prototype solution the weights were used to compute a weighted sum of knowledge and skills and then the calculated sum was interpreted in the scale 1–3 (as there are 3 levels of proficiency defined for this competency). Example calculations are presented in the Table 1.

The weights are assigned to knowledge and skills elements by the manager who should consider the process requirements and goals and ensure that the weights truly reflect the organizational and priorities.

Table 1. Example calculations of the level of proficiency.

Competency elements k- knowledge, s-skills	Weight	Employee 1	Employee 2	Employee 3
k1	1		x	x
k2	1	x	x	x
k3	0.5	x		x
k4	0.5			
k5	0.5	x		
s1	1	x	x	x
s2	1		x	x
s3	1	x		x
s4	0.5		x	x
s5	0.8	x		x
Total	7.8	4.8	4.5	6.8
Level of proficiency	3	2	2	3

Ontologies provide a framework for automatic reasoning, the proposed Software Competence Ontology contains definitions of classes, properties and relations which allow for posing semantic queries, such as:

- find persons who have evidenced level of competencies at least 3 in the area B (Build),
- find a person who can substitute with Manager X in the business process Y.

The first example query can be used to determine which of the competencies evidenced in the employees profiles belong to the area "Build" although the area may not be explicitly specified in their profiles.

In the second case, if the requirements for the business process Y are defined according to the areas and levels specified in the ontology, the aim is to find a person who at least fulfills these requirements (the levels of competencies of the person are equal or close to requirements). Another approach is to find a person who is most similar to Manager X regarding values of her competencies. Similarity can be calculated in many ways e.g. applying selected distance measure and computing distance between levels of competence of each pair of the persons.

The competency values can be represented as a vector, therefore cosine similarity measure can be used. The example calculation is presented in Table 2. The cosine similarity for two vectors A and B is calculated as follows:

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where:

A – the vector representing values of competencies of the Manager X
B – the vector representing values of competencies of other employee

The cosine similarity factor in the above example shows that the most appropriate candidate to substitute Manager X is Employee 3.

Table 2. Finding similarities between competencies of Manager X and other employees.

Competency needed in project Y	C.2. Change support	C.3. Service delivery	D.9. Personnel development	E.8. Information Security Management	D.1. Information security strategy development	Cosine similarity
Manager X	**4**	**3**	**5**	**4**	**3**	–
Employee 1	3	2	5	3	2	0.98631
Employee 2	4	3	2	2	2	0.93017
Employee 3	4	2	5	4	3	0.99369

The ontology of competencies also can be helpful if there is a need to analyze unstructured information such as CVs of candidates or new employees. In such a case semantic similarity measures can be used. The process of measuring semantic similarity is iterative and can be as follows:

1. First, the text of the CV is analyzed to find keywords that are present in the descriptions of ontology classes and properties.
2. The thesaurus is used to find similar words in CV to those that are present in ontology.
3. Each time the keyword is found it is noted as one point for the given area and property of ontology.
4. When no more keywords are found the system displays suggestions of the areas and competencies identified for the given person.
5. The user engagement is needed to evaluate the competencies of the candidate in the scale 1–5 in the each of the areas as the measures may not be specified (or may be specified in different ways) in the text of the CV.

Semantic analysis of the terms used in descriptions of the employees' competencies is based on the use of a lexical database and semantic similarity algorithms. The lexical database WordNet [45] can be used as it is particularly well suited for similarity measures, since it organizes nouns and verbs into hierarchies of is–a relations [46].

The Software Competence Ontology for the prototype of the system was coded using Protégé [47] platform (Fig. 3). The instances of competencies on the base of e-CF were imported from MS Excel file using Cellfie plugin of the Protégé platform.

The ontology classes are displayed in a left-top window, on the left bottom of the figure there are instances of competencies. The "IS and Business Strategy Alignment" competence is selected. On the right top window knowledge and skills elements for the selected competency are visible. Competencies are coded according to the 5 areas specified in the e-CF framework and denoted with letters (A – Plan, B – Build, C – Run, D – Enable, E – Manage) and numbers.

The next possible use for Software Competence Ontology is to support annotation of documents evidencing competency and proficiency levels. The descriptions in employee profile, candidate CV or business process diagram can be compared with detailed knowledge and skills descriptions in the SCO. Then propositions of references to the ontology can be suggested. However it should be noted that practical implementation of such a procedure would require additional dictionaries that contain synonymous words and phrases to those used in ontological descriptions of knowledge and skills.

In the recent years the concept of process approach to management is gaining popularity among companies of all sizes. From the perspective of the process approach the competencies should be seen from the perspective of tasks implemented within the business process. The business processes are visualized in the form of process models. The process models are implemented in many modern enterprise management systems as the tool which depicts the current process instances with connection to generated documents and responsible bodies. Process models are also a formalized way to support analysis and improvement. Therefore it may be said that the models link two different perspectives – business management and the software engineering one.

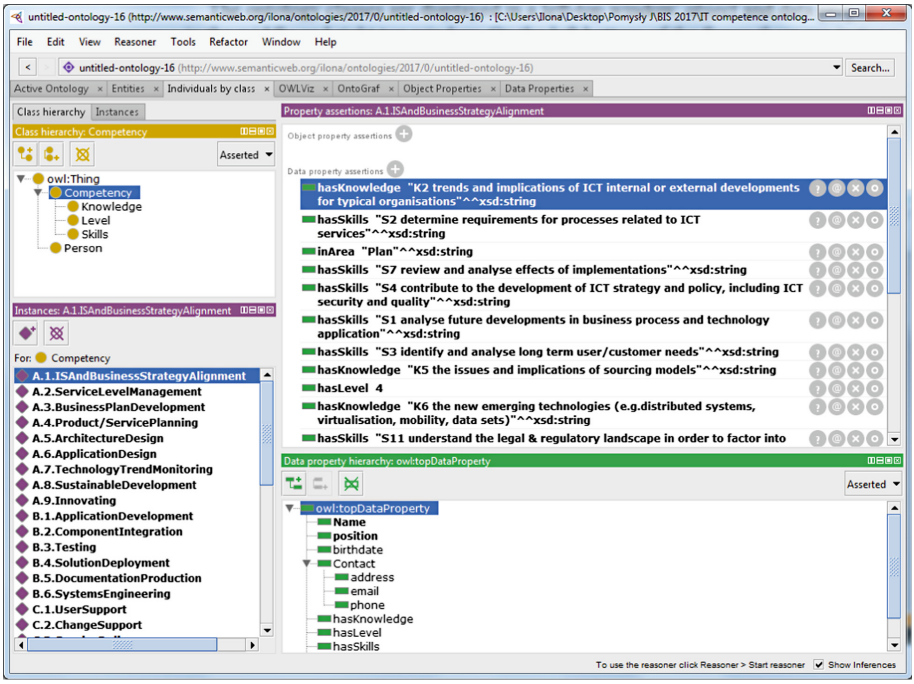

Fig. 3. Software Competence Ontology displayed in Protégé (source: Own elaboration using Protégé [47]).

A business process model is a step-by-step description of what one or more participants should do to accomplish a specific business goal. According to Gartner [48] business process analysis tools are primarily intended for use by business end users looking to document, analyze and streamline complex processes, thereby improving productivity, increasing quality, and becoming more agile and effective. Classical business process analysis is oriented on analyzing and optimizing business processes for better productivity by saving time, costs or creating a more desirable product for customers.

Due to the increased need of agility competency-based management is crucial activity of contemporary business organizations. In this situation information technology support is the core element of recruiting the most appropriate candidates, effective planning of employee development programs and project management. In many cases the information about competencies is exchanged between collaborating organizations.

The IT tools for supporting process modelling should therefore provide possibility to view the business process from the perspective of competencies required to perform particular tasks. Provision of information describing needed competencies of individuals involved in the process can help not only in workforce planning for the particular process but also in other management tasks such as expert finding, personalization of career paths and staff trainings.

4.3 Example Scenario of Software Configuration Management Process

This section describes a proposition of a solution which integrates process models with Software Competence Ontology (SCO). The issue is presented on basis of software configuration management (SCM) process. SCM is one of the processes being integral part of software engineering projects carried out by software companies. A. Leon defines Software Configuration Management as "the art of identifying, organizing, and controlling modifications to the software being built by a programming team [49, p. xxi]. The process of software configuration management consists of identifying and defining the configuration items, controlling the release and change of these items throughout the system lifecycle, recording and reporting the status of configuration items and change requests, and verifying the completeness and correctness of configuration items. It is a knowledge-intensive process that involves cooperation of many participants such as, managers, analysts, developers, testers and end-users.

The goal the SCM process is to successfully deliver a software product to a customer or market in accordance with customer's requirements and software company's business plan [50]. The decisions taken during this process by project managers are usually taken under the pressure of time and require skills from the areas such as: software design, construction, testing, sustainment, quality, security, safety, measurement and human-computer interaction.

The performance of SCM process is essential for software company because it directly impacts the customer satisfaction. Therefore decisions taken during the SCM process should regard both the customer's needs and the business case perspective. Knowledge of related disciplines is very important as well as cognitive skills and behavioral attributes of the team members. The BPMN diagram of software configuration management process enhanced by competence artifacts is illustrated on Fig. 4.

The presented business process model is annotated by the information on competency requirements. The annotations are added as additional artefacts connected to the process tasks with dotted line. To make the diagram more readable added elements contain symbols (for example: A1, B2, C2) which reference to the Software Competence Ontology. The symbols are displayed in a form of hyperlinks so it is possible at any time to look up the detailed descriptions of knowledge and skills needed on each stage of the process.

In the example scenario a project manager wants to find the right people for preparing software and hardware configuration report which is one of the tasks in software configuration management process (Fig. 4).

The manager formulates a query to find persons who have competency denoted in the process model as: B3(S5), which is described in SCO as follows:

- B – area "Build" – consists of competencies needed for building software.
- B3 – Competence no.3: "Constructs and executes systematic test procedures for ICT systems or customer usability requirements to establish compliance with design specifications. Ensures that new or revised components or systems perform to expectation. Ensures meeting of internal, external, national and international standards; including health and safety, usability, performance, reliability or compatibility. Produces documents and reports to evidence certification requirements."
- S5 – Skill no.5 of reporting and documenting tests and results.

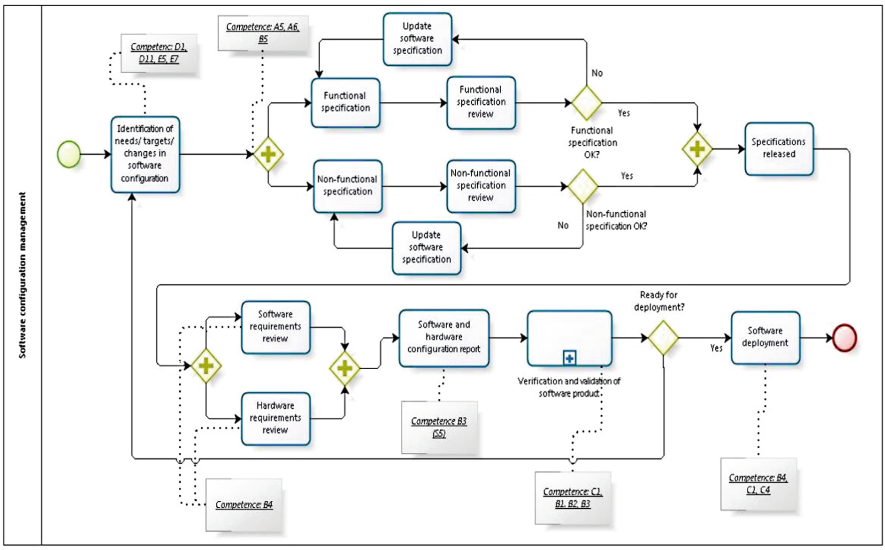

Fig. 4. BPMN diagram of software configuration management process with competency annotations. (source: [44]).

The required level of competence are declared in the ontology using scale 1–5, where 5 means the most advanced knowledge and skills. Meanings of particular levels are also explained in ontology.

The solution of integrating previously presented process with Software Competence Ontology is aimed first of all to facilitate finding the right person to perform a task in the process. Moreover the Software Competence Ontology can act as "common language" to describe the details of employees' profiles, annotating the CVs of job seekers, creating job postings and building a database of existing or potential business partners.

A general scheme of the platform supporting competency management is presented on Fig. 5. The project manager, process owner or HR manager while analyzing business process from the perspective of the competencies can formulate a query and receive the list of potential contractors able to perform particular tasks.

The contractor can be an employee, a job seeker or a business partner who has knowledge and skills fully or partially consistent with the defined requirements. If there is no single person having all the required knowledge and skills for the given task, a team can be formulated consisting of the suggested individuals.

There are many resources of competency data to use. The internal resources contain employees' profiles and CV of job candidates. The external data sources may include information extracted from job hunting websites and databases shared by other organizations. If there are no people with proper competencies among the employees the database of job candidates or external databases exposed by business partners or job hunting portals can be searched through. For the solution to work all the resources should be annotated with ontology concepts and references to knowledge and skills elements specified in the Software Competence Ontology.

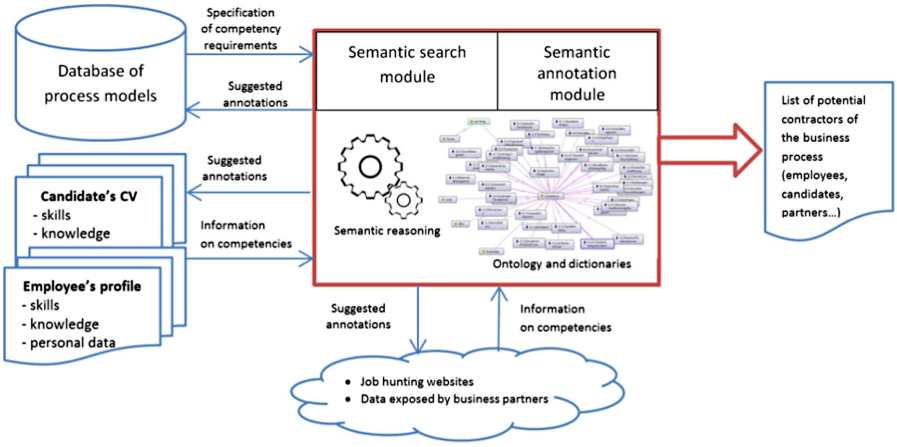

Fig. 5. Modules and resources of competency management framework.

5 Conclusions

5.1 Research Contribution

This work contributes to existing research on competency management and business process management by proposing the new ontology-based solution to integrate competency requirements with business process models. The use of ontology as a conceptual schema for representing competency requirements creates many useful possibilities that have not been proposed in other studies from this field. In particular, the presented framework facilitates reasoning, knowledge sharing and reuse. The possibility to analyze business processes form the competency perspective is useful for planning and selection of staff for particular tasks.

The changes in technology and economic environment create the need for continuous adjustment of business processes and searching or developing new competencies. Human capital is the carrier of the organization's knowledge and skills, therefore developing core competencies of the organization requires developing individual and team competencies. The requirements in this area should follow from the processes and tasks performed by the working group or the whole organization. Therefore the main focus of this study was on individual and team competencies.

5.2 Implications for Research and Practice

The concept of the platform for supporting process-oriented competency management has been proposed and illustrated by the example software configuration management process. IT labor market is, constantly changing due to rapid advancements in technology and innovative products. New competencies emerge and competencies that are already in existence, change their contents.

The dynamic approach to manage competencies on the base of business process flow seems especially significant in the case of knowledge-intensive firms, it also can

be valuable for process-based and virtual organizations where the environment is dynamic and frequent changes are needed to preserve the company's competitiveness and agility, this is often the case of software companies.

The Software Competence Ontology can be used to design and create metadata describing the internal and external resources of competency information. It also can be used to compare candidates and employees against requirements. The methods of calculating the level of proficiency and finding similarities between competency levels of individuals were also proposed and explained by examples.

Competency management, can be seen as one of the most important drivers of business processes performance improvement, therefore there is a growing need for systematic approaches and IT support in this area. The existing competence management systems usually use databases and repositories of knowledge and skills descriptions, which are used in employee profiles to express levels and areas of their expertise. With emergence of Web 2.0 enterprises started to use social networking tools to build and publish employee profiles. The mentioned solutions are very popular however they have two major drawbacks – lack of standardized vocabulary and lack reasoning mechanisms to determine the competencies of individuals' by inferring on different levels of abstraction and detail. These limitations can be addressed by semantic solutions based on ontologies.

5.3 Limitations and Future Works

The Software Competence Ontology was developed in Protégé on the basis of European e-Competence Framework. The proposed ontology will be developed in the future, and extended by new perspectives. As the research shows a very important aspect of ICT projects is competency of communication, in particular cultural and language competencies [51], which were not taken into account in this study at the moment. Communication aspect can become the next element of the SCO in addition to purely technical skills and knowledge.

Increasingly open and collaborative nature of modern organizations involves using external knowledge resources such as job hunters websites or databases of competencies that could be exposed by other business entities. This requires enhanced semantic search algorithms to find similar or the same competencies described with different terms. Therefore the proposed solution can be extended by additional dictionaries and domain ontologies.

References

1. Różewski, P., Małachowski, B., Danczura, P.: Concept of competence management system for Polish National Qualification Framework in the Computer Science area. In: Ganzha, M., Maciaszek. L.A., Paprzycki, M. (eds.) FedCSIS, pp. 759–765 (2013)
2. Vakola, M., Soderquist, K.E., Prastacos, G.P.: Competency management in support of organisational change. Int. J. Manpow. **28**(3/4), 260–275 (2007)
3. Osa, J.O.: Managing the 21st century reference department: competencies. Ref. Libr. **39**, 35–50 (2003)

4. Lundqvist, K.O., Baker, K., Williams, S.: An ontological approach to competency management. In: Proceedings of the iLearn 2007, pp. 1–4 (2007). http://www.eife-l.org/publications/proceedings/ilf07/Contribution110.doc.pdf
5. Hammer, M.: The process audit. Harv. Bus. Rev. **85**(4), 111–123 (2007)
6. Harmon, P.: The state of business process management. A BPTrends Report. https://www.bptrends.com/bpt/wp-content/uploads/2015-BPT-Survey-Report.pdf
7. Wysłocka, E.: Implementation of enterprise processes to IT solution for company. WSEAS Trans. Bus. Econ. **14**, 225–234 (2017)
8. Kohlbacher, M., Gruenwald, S.: Process orientation: conceptualization and measurement. Bus. Proc. Manag. J. **17**(2), 267–283 (2011). https://doi.org/10.1108/14637151111122347
9. Kohlbacher, M.: The effects of process orientation: a literature review. Bus. Process. Manag. J. **16**(1), 135–152 (2010). https://doi.org/10.1108/14637151011017985
10. Lorino, P.: A pragmatic and interpretive approach to competence-based management: the case of a telecommunications company. In: Sanchez, R. (ed.) A Focused Issue on Fundamental Issues in Competence Theory Development. Research in Competence-Based Management, vol. 4, pp. 219–257 (2008)
11. Leyking, K., Angeli, R.: Model-based competency-oriented, business process analysis. Enterp. Model. Inf. Syst. Arch. J. **4**(1), 14–25 (2009)
12. Nolte, A., Sprafke, N., Wiechers, M., Herrmann, T.: Integration von Kompetenz- und Prozessmanagement – Eine Chance für Mitarbeiter und Organisationen. In: GfA (Hrsg.), Arbeit in komplexen Systemen. Digital, vernetzt, human?!, pp. 1–6. GfA Press, Dortmund (2016)
13. Giorgini, F., Zimmermann, V., Faltin, N., Vervenne, L.: Time2Competence: the PROLIX project. In: Nejdl, W., Tochtermann, K. (eds.) EC-TEL 2006. LNCS, vol. 4227, pp. 597–602. Springer, Heidelberg (2006). https://doi.org/10.1007/11876663_59
14. ISO 10303-203:1994: Information technology – object management group business process model and notation (1994)
15. Rodríguez, A., Caro, A., Cappiello, C., Caballero, I.: A BPMN extension for including data quality requirements in business process modeling. In: Mendling, J., Weidlich, M. (eds.) BPMN 2012. LNBIP, vol. 125, pp. 116–125. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33155-8_10
16. Bocciarelli, P., D'Ambrogio, A.: A BPMN extension for modeling non functional properties of business processes. In: Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVS Integrative M & S Symposium, TMS-DEVS 2011, pp. 160–168. Society for Computer Simulation International, San Diego (2011)
17. Altuhhova, O., Matulevicius, R., Ahmed, N.: An extension of business process model and notation for security risk management. Int. J. Inf. Syst. Model. Des. **4**(4), 93–113 (2013)
18. Braun, R., Esswein, W.: Classification of domain-specific BPMN extensions. In: Frank, U., Loucopoulos, P., Pastor, Ó., Petrounias, I. (eds.) PoEM 2014. LNBIP, vol. 197, pp. 42–57. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45501-2_4
19. Stroppi, L.J.R., Chiotti, O., Villarreal, P.D.: Extending BPMN 2.0: method and tool support. In: Dijkman, R., Hofstetter, J., Koehler, J. (eds.) BPMN 2011. LNBIP, vol. 95, pp. 59–73. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25160-3_5
20. Sucre, F.S., Pernalete-Chirinos, D.: Ontological framework for the management of job profiles based on labor competencies. GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología **5**(1), 17–41 (2017)
21. Kimble, C., de Vasconcelos, J.B., Rocha, Á.: Competence management in knowledge intensive organizations using consensual knowledge and ontologies. Inf. Syst. Front. **18**(6), 1119–1130 (2016)

22. Toader, E.A.: Competency-based assessment of IT professionals as a software organization tool. In: Proceedings of the International Scientific Conference eLearning and Software for Education, Bucharest, vol. 3, pp. 13–20 (2015). https://doi.org/10.12753/2066-026x-17-181
23. Ortega-Gonzalez, Y., Carli, G., Grandi, A., Delgado-Fernández M.: The specification of competency questions: an ontological support to match business processes, it and competences of professionals. In: Proceedings of the 5th European Conference on Information Management and Evaluation, pp. 531–539. University of Insubria, Italy (2011)
24. Janev, V., Vranes, S.: Ontology-based competency management: the case study of the Mihajlo Pupin Institute. J. Univers. Comput. Sci. **17**(7), 1089–1108 (2011)
25. IMS Global Learning Consortium: IMS Reusable Definition of Competency or Educational Objective Specification (2017). https://www.imsglobal.org/competencies/
26. HR-XML Consortium: Competencies (Measurable Characteristics) Recommendation (2004). http://www.ec.tuwien.ac.at/~dorn/Courses/KM/Resources/hrxml/HR-XML-2_3/CPO/Competencies.html
27. InLOC: Integrating Learning Outcomes and Competences (2017). http://www.cetis.org.uk/inloc/Home
28. https://www.onetonline.org/
29. European e-Competence Framework: A common European framework for ICT Professionals in all industry sectors (2016). http://www.ecompetences.eu
30. SFIA Version 7 (2017). https://www.sfia-online.org/en
31. ITCM: The Information Technology Competency Model (2018). www.careeronestop.org/competencymodel
32. Hawthorne, E., Campbell, R., Tang, C., Tucker, C., Nichols, J.: ITC ACM. Competency Model of Core Learning Outcomes and Assessment for Associate-Degree Curriculum (2014)
33. UNESCO ICT-CFT: UNESCO ICT Competency Framework for Teachers, version 2 (2011)
34. Masseroli, M., Tagliasacchi, M.: Web resources for gene list analysis in biomedicine. In: Lazakidou, A. (ed.) Web-Based Applications in Health Care and Biomedicine. Annals of Information Systems Series, vol. 7, pp. 117–141. Springer, Heidelberg (2009). https://doi.org/10.1007/978-1-4419-1274-9_8
35. Aussenac-Gilles, N., Biébow, B., Szulman, S.: Corpus analysis for conceptual modeling. In: Proceedings of the EKAW 2000 Workshop on Ontologies and Texts, Juan-les-Pins, French Riviera, vol. 51 (2000)
36. Gomez Pérez, A., Benjamins, V.R.: Overview of knowledge sharing and reuse components: ontologies and problem-solving methods. In: Proceedings of the Workshop on Ontologies and Problem-Solving Methods KRR5 IJCAI 1999, Stockholm, Sweden (1999)
37. Kassel, G.: OntoSpec: une méthode de spécification semi-informelle d'ontologies. In: Proceedings of IC2002, Rouen, France, pp. 75–87 (2002)
38. Uschold, M., Gruninger, M.: Ontologies: principles, methods and applications. J. Knowl. Eng. Rev. **112**, 93–155 (1996). https://doi.org/10.1017/S0269888900007797
39. Dudycz, H., Korczak, J.: Process of ontology design for business intelligence system. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 17–28. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_2
40. White, S., Miers, D.: BPMN Modeling and Reference Guide: Understanding and Using BPMN. Future Strategies Inc., Lighthouse Point (2008)
41. Rekik, M., Boukadi, K., Ben-Abdallah, H.: Specifying business process outsourcing requirements. In: Lorenz, P., Cardoso, J., Maciaszek, Leszek A., van Sinderen, M. (eds.) ICSOFT 2015. CCIS, vol. 586, pp. 175–190. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30142-6_10
42. http://www.cetis.org.uk/inloc/e-CF_categories
43. http://wp1087322.server-he.de/

44. Pawełoszek, I.: Process-oriented approach to competency management using ontologies. In: Ganzha, M., Maciaszek, L., Paprzycki M. (eds.) Proceedings of the 2017 Federated Conference on Computer Science and Information Systems. ACSIS, vol. 11, pp. 1005–1013 (2017). https://doi.org/10.15439/2017f441
45. Princeton University: About WordNet (2010). http://wordnet.princeton.edu
46. Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet: similarity - measuring the relatedness of concepts. In: AAAI 2004, pp. 1024–1025, San Jose, CA (2004)
47. http://protege.stanford.edu
48. http://www.gartner.com/it-glossary/bpa-business-process-analysis-tools/
49. Leon, A.: Software Configuration Management Handbook, 3rd edn. Artech House, Boston/London (2015)
50. Farah, J.: Defining a Software Configuration Management Process to Improve Quality (2012). https://www.cmcrossroads.com/article/next-generation-process-and-quality-0
51. Muszyńska, K.: Patterns of communication management in project teams. In: Ziemba, E. (ed.) AITM/ISM 2016. LNBIP, vol. 277, pp. 202–221. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53076-5_11

Use of a Business Process Oriented Autopoietic Knowledge Management Support System in the Process of Auditing an Organisation's Personal Data Protection

Mariusz Żytniewski(✉)

University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland
mariusz.zytniewski@ue.katowice.pl

Abstract. A key aspect of the operation of organisations is execution of business processes that take place in them. However, completion of specific tasks requires application of specific knowledge defined by the context of its use. While for an expert in a certain field, using specific knowledge is a natural process, for an IT system designer it requires understanding of the purpose of a business process and places where specific knowledge resources are to be used. This paper addresses the issues of integrating business processes and organisational knowledge in the process of building systems designed to support activities connected with the process of an organisation's security audit. In particular, it presents the reasons for integration of BPM and KM, elements of a proposed methodology. The main research findings include: analysis of the theory of integration of BPM and KM, presenting the stages of building a business process oriented autopoietic knowledge management support system and development of a prototype of a system to aid in auditing an organisation's personal data protection.

Keywords: Business process · Autopoiesis · Knowledge management

1 Introduction

One of basic challenges faced by organisations is to create optimal business processes that involve sharing of organisational knowledge [1–4]. In order to facilitate the performance of processes, it is necessary to build IT solutions which, on the one hand, support the performance of a business process, while on the other hand, provide the user with the necessary knowledge that aids them in their activity or even substitute them in decision-making. As indicated by research by Al-Mabrouk [5] and Choy Chong [6], IT technologies are one of key factors of successful use of Knowledge Management (KM) in organisations. Knowledge management in an organisation is a complex process covering various aspects that refer to a whole range of organisational activities.

Akhavan et al. [7] pointed out that the most important key factors of successful implementation of KM are the aspects of knowledge sharing and knowledge storage. Therefore, it is important to search for methods for building IT solutions which will support the process of making organisational knowledge available and storing it. This

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 61–80, 2018.
https://doi.org/10.1007/978-3-319-77721-4_4

results from the required context of the use of information, which in terms of a process that is performed can be viewed as organisational knowledge. While, from the perspective of KM, knowledge is user-oriented, business processes refer to the operation of an organisation as a whole.

When addressing the issue of the integration of Business Process Management (BPM) and KM with reference to IT systems and their impact on participants, it is necessary to indicate three elements of such a system (Fig. 1).

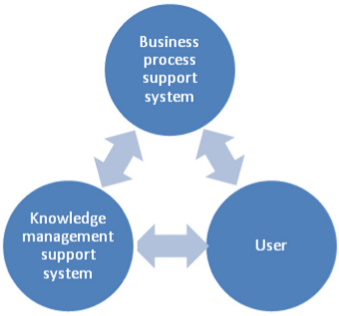

Fig. 1. Elements of a business process oriented knowledge management system

The key element is a user, who, while participating in a business process, should:

- Understand the purpose and course of the business process in which he/she participates,
- know which knowledge is necessary during execution of specific tasks,
- have access to knowledge resources that are necessary for performing a process, and in particular its tasks,
- know effects expected from the execution of a process and measures that can be used to assess it.

As a result, a developed software solution should integrate knowledge as part of the tasks executed in a process and make it available to a user while he/she performs his/her tasks. What's important, it is a continuous process and refers to both tasks and whole business processes. As pointed out by Records [8], from the perspective of the use of KM to support business processes, we have to consider the process flows and steps, we need to fully understand and design the structure and condition of the knowledge that will be required to execute the process.

Research conducted by Metaxios et al. [9] indicated a range of problems connected with the use of IT solutions in knowledge management, one of which was their inclusion in an organisation's processes. In particular, knowledge management mechanisms should, according to respondents, refer to the whole organisation and participants of a process, rather than only a specific division dealing with knowledge management in an organisation.

The author's earlier research [10–13] has shown that the theory of software agent societies and its use in knowledge-based organisations requires a separate view of the

characteristics of multi-agent systems. Especially in the area of the application of methods for knowledge representation in such systems and possibilities of using autopoietic solutions to support the different stages of the life cycle of the process of knowledge management [10, 14]. By applying the theory of autopoiesis, theory of distributed systems and agent-based approach in designing and building business processes supported by knowledge management processes, new functions can be added to a developed system. This aspect will be expanded further in the paper. Power [15] emphasizes that distributed computing systems are characterized by several features. The first of these is the ability to adapt to unforeseen events occurring in a system. In a static system the failure of one of its components may cause the non-functionality of the entire solution. In the case of a distributed system, detection of problems with the functioning of one subsystem may result in its replacement by another. This is due to the assumed separation of various functions of a system into subsystems. A feature of adaptation may also result from the need to adjust the function of a system to changes in its environment. In this case, the adaptation of a system is reflected in the adaptation of this system, as a whole, to the needs of the tasks that are imposed on a given solution. As a result, the functionality of one subsystem may be changed or a new functionality may be added to the system by including a new subsystem.

Research into the methodologies used in designing multi-agent systems [16] indicated a range of features that should be considered in the context of building agent societies designed to be used in knowledge-based organisations. Conducted research [16, 17] found out gaps in the methodologies for supporting design of multi-agent systems considered in the context of agent societies in the area of their application for supporting knowledge-based organisations. The author's current research aims to define the possibilities of using the theory of software agent society in building a business process oriented autopoietic knowledge management support system [18–20]. One of the aspects of creating autopoietic systems to support the integration of business processes and knowledge management is the methodological aspect of the development of such systems [21], which is addressed in the paper.

The aim of the paper is to analyse the issues of modelling the integration of business processes and knowledge management, considered in the context of autopoiesis of IT systems. Section 2 presents reasons for integration of BPM and KM. Section 3 shows an element of a developed methodology for supporting integration processes as part of organisational knowledge business processes. Section 4 presents selected elements of a developed design of software for supporting data security audit, developed based on the methodology presented. The summary will feature the diagnosed advantages and disadvantages of the proposed approach.

2 Problems of Designing the Integration of Business Processes and Knowledge Management

As was pointed out earlier, one of the approaches in the development of IT systems designed to support an organisation's activities is creation of systems aimed at supporting business processes and managing an enterprise's knowledge. Business process support systems are focused on performance of an organisation's ongoing tasks.

Knowledge management systems, in turn, refer to the process of gathering, processing and distributing knowledge in an organisation. Essentially, this knowledge should support performed processes and provide employees with necessary knowledge [22].

Integration of business processes and knowledge management systems as part of a system being developed may refer to the process of the performance of a task undertaken in a business process or support for a decision-maker participating in the process. This paper will address the latter issue. As was pointed out by Bitkowska [23], a knowledge management system can be divided into four sub-systems:

- Databases – which refer to data access and knowledge sharing,
- Organizational language – which allows the terms used in an organisation to be understood,
- Network links – which enable access to information and knowledge within an organisation and beyond, and
- Transfer – which enables transfer of knowledge between individuals.

From the perspective of integration of KM and BPM, these sub-systems have to be subject to contextual integration as part of business processes in which they will be used. For that purpose, it is necessary to develop IT solutions designed to support such processes. In terms of the use of Business Process Management in Knowledge Management, the following postulates of this approach can be formulated [24]:

- business processes, if modelled and captured in business process repository, are a part of codified intellectual capital of the organization,
- knowledge processes in an organisation should be a part of business process repository, and
- business process repository could be used for knowledge creation, sharing and distribution.

The postulates presented here refer to life cycles of knowledge management systems indicating the necessity for connecting them with process-based approach. What's important, the knowledge about a business process should be treated as an organisational knowledge resource. This produces a loopback where the knowledge contained in a knowledge management system supports the execution of a business process, which, improved in this way, affects the organisational knowledge resources. Thus, this process should not only involve the aspect of an organisation management but also processes of designing and building IT systems that will support these processes.

In terms of the support for participants of a business process, it is reasonable to separate the system's elements that are responsible for the implementation of a process, its flows and orchestration from analytical systems that support decision-making processes [25]. This makes it necessary to build integrating solutions which, apart from linking organisational knowledge as part of business processes, will be able to autonomously process and provide decision-makers with knowledge that is necessary for performing tasks they undertake. One of the postulated elements that integrate BPM and KM is the aspect of knowledge codification, which should be ensured by such systems. In this case, problems that can be solved by such systems (against the background of integration of Business Intelligence and KM) include [26]:

- lack of support in defining business rules for getting proactive information and support in consulting in the process of decision making,
- lack of a semantic layer describing relations between different economic topics,
- lack of support in presenting the information of different users (employees) and their individual needs, and
- difficulty in rapidly modifying existing databases and data warehouses in the case of new analytic requirements.

The problem that appears during the design of systems discussed in this paper is specification of business processes and knowledge resources, as well as translating the defined elements of the system into the application of an IT system designed to support a decision-maker's actions. The definition of organisational knowledge resources in the form of knowledge portals forces a decision-maker to search for specific knowledge needed to perform their tasks. On top of that, part of organisational knowledge can be scattered in the organisation, and tasks will take more time to perform. It is thus reasonable to support the process of integration of organisational knowledge as part of the tasks of business processes performed by decision-makers and to facilitate the methods for building IT systems that aid adaptation of knowledge to process participants. In article [21] author presents a range of guidelines on how to define such systems and it's methodology:

- strategic view of the process of developing such a system connected with the necessity of building a technical system and a social one related with the philosophy of an organisation's operation,
- focusing the system on the operation of an organisation, in particular its participants and environment,
- codifying organisational knowledge and inclusion into its resources of the new knowledge created by the system,
- defining the context of using organisational knowledge as part of business processes,
- necessity of codifying meta-knowledge (knowledge on the organisation's knowledge resources) in order to include it in the organisation's business processes,
- necessity of using certain execution standards and mechanisms that enable definition of business processes, storage of their templates as well as launching and evaluation,
- necessity of defining not only information flows but also social and organisational relations within an organisation connected with the transfer of organisational roles to the technical system and business processes,
- separation of both the systems from each other and modelling of their functional relationships as part of the stages of the performance of business processes,
- necessity of defining various kinds of organisational knowledge in terms of business process-oriented knowledge and KM-related knowledge, and
- possibility of relating the architecture of a system being developed to life cycles of BPM and KM in order to better understand the principles of its operation and functional scope.

Notations designed to support business process modelling, such as ARIS [27], BPMN [28] or IDEF0 [29], do not provide ready solutions that specify what knowledge will be provided to a decision-maker during their tasks, its sources or specification. This problem may be solved by using e.g. the KMDL (Knowledge Modelling and Description Language) language [30], but such solution in the case of business analysts requires the use of a new notation when business processes in a company are already documented by means of business process-oriented notations. When a new design notation is used, all the elements of a process have to be mapped to new artefacts. Another problem that appears is specification of knowledge resources. Applied notations usually define their own artefacts describing organisational knowledge without clear indication of how they are codified in the IT system. This results in inconsistency between the definition of knowledge resources in the design and their actual implementation in the IT system. Often, proposed design notations do not address the issues of standards for specification of knowledge resources, defined e.g. by W3C organisation, which indicate what the structure of the ontology describing knowledge resources should look like.

Specification of knowledge resources based on proprietary sets of artefacts does not allow for their direct translation into available standards for knowledge codification. By using standards for describing knowledge resources in the form of ontology description languages OWL, OWL 2, RDF or RDFS, it will be possible to use the developed ontology again in another project and organisation. This is possible thanks to semantic description of the meanings of the terms used in ontologies.

Another aspect is semantic identifiability of the terms used during the design of knowledge resources by a knowledge engineer, which would make it possible to use the ontology for designing purposes (interpretable by the IT system being designed and by process participants) and for the purpose of system implementation. It would allow a once prepared definition of knowledge resources to be an element of design specification, an element of the system being implemented and to be used to integrate an organisation's knowledge resources with other ontologies.

The literature offers [31, 32] a range of studies which define how knowledge is codified based on ontology description languages. The main types of ontology include core, upper-level, domain, task, and application ontologies. The example presented further in the paper represents domain ontology with elements of application ontology. It can be concluded that in terms of integration of business processes as part of knowledge management systems, it is necessary to create solutions that ensure:

- The use of generally accepted standards for describing an organisation's business processes, which will allow already operating organisations with diagnosed business processes to easily integrate the knowledge management system as part of employees' tasks.
- The use of standardised descriptions of an organisation's knowledge resources in the form of ontologies and ontology description languages and possibilities of using already applied standards in the process of defining field ontology.
- Mechanisms that allow for translation of the defined ontologies into a format that can be recognised by IT systems, thus shortening the time it takes to implement the system (the ontology developed at the stage of specifying the system's knowledge

resources will be able to be automatically used during its implementation) and ensuring interoperability of knowledge resources across various projects and organisations.
- Linkage of the process of specifying business processes and organisational knowledge resources with the process of designing not only systems that automate the performance of processes but also systems that support decision-making.
- The use of codified knowledge resources in defining business rules of a business process and rules for the operation of a decision-making support system. Such translation makes it easier to define control mechanisms that control the operation of a system's elements.
- Extension of the architecture of built systems for integration of business processes and knowledge management by autopoietic elements that support decision-makers' actions through processing the codified resources of the system's knowledge.

The first four postulates refer directly to the aspect of integration of BPM and KM. The next two are connected with integration of autopoietic solutions as part of such a solution. The use of the theory of autopoietic systems impacts additional features of the solution being developed. They include: partially open, self-reference, self-control, boundary-generation, self-organisation through self-production. In such systems, business processes are performed dynamically based on system-resident components, and are subject to constant control. By using the theory of autopoiesis in the process of supporting the performance of business processes, system elements are not only subject to self-organisation, but - through the process - performed production processes can be reproduced and then incorporated into the business process being performed. These actions are carried out in a partially open system, where system elements interact with one another, which is equipped with control mechanisms that limit undesired behaviours within the whole system.

The methodology presented further in the paper and the tool developed to support its implementation fulfils the above-mentioned postulates and constitutes a response to the problems pointed out in this chapter.

3 Elements of the Proposed Methodology

The proposed methodology has been developed based on three main stages with a loopback, which involve the specification of business processes of an organisation, its knowledge resources and an autopoietic element that facilitates integration of knowledge resources within a business process [21]. The methodology comprises the following stages: identification and modelling of business processes, identification and modelling of an organisation's knowledge resources and designing and implementation of a process oriented autopoietic knowledge management support system. In accordance with Table 1, the following stages have been identified in the methodology.

The first stage refers to defining the specificity of a business process to be supported. It may turn out that not all stages of a given process can be supported by the system. What's more, this approach helps define information flows between tasks that can be supported by the system and helps specify the recipients of a created software

Table 1. Stages of the developed methodology

Objective	Executed substages	Effects
1. Analysis and development of a business process		
Development of the definition of business processes of an organization	1.1. Specification of organizations involved in the process and the posts performing the tasks 1.2. Determination of relationships inside the organization. At this stage, the relationship is defined within the organizational structure that supports the system. In the case of an organization, it is a structure linking 1.3. Defining the rules of starting and ending the process 1.4. Diagnosing the business process tasks 1.5. Diagnosing the business process events 1.6. Defining the conditions governing decision gates	Making a map of an organisation's business processes constituting the context for further use of a system being developed
2. Identification of organisational knowledge resources		
Making a map of the sources of the organisational knowledge that can be used to build the system	2.1 Identification of codified knowledge sources (e.g. documents, websites, system-generated knowledge developed in earlier iterations) 2.2 Identification of non-codified knowledge sources (e.g. experts) 2.3 Development or update of the ontological model (meta-knowledge) about knowledge resources	Codified meta-knowledge about knowledge sources and their structure
3. Designing and implementation of a process oriented autopoietic knowledge management support system		
Development of the architecture of the autopoietic system (based on the earlier specified demands) that processes organisational knowledge and provides it in the context of selected tasks of a business process. This stage comprises a range of sub-stages connected with the development of a technical system	3.1. Identification of the context of usage 3.2. Analysis of the roles and responsibilities of autopoietic system 3.3 Determining the hierarchical structure of the relationship inside the organization 3.4. Preliminary definition of the architecture of an autopoietic system 3.5. Indication of the impact of control mechanism on the autopoietic system 3.6 Essential definition of the autopoietic element internal architecture 3.7 Designing the interaction autopoietic elements	Development of the architecture of an autopoietic system that is equipped with control mechanisms and uses communication to support the performance of business processes. The effect of its use is development of the elements of a business process oriented autopoietic knowledge management support system

solution. This stage can also be aided by application of Use Case Diagram, which is known from UML notation, but this diagram does not show relations between the tasks undertaken by the users of a process or the direction of information flow. Therefore, it can only be used to help understand the assumed functions of a system. On this basis, the usage context of the system being built is defined. The proposed approach uses the BPMN notation for specification of the process supported by a business process oriented autopoietic knowledge management support system. This allows for the use of this approach to an organisation that already has codified processes without the need to codify them once again.

The definition of a business process alone, though useful in the context of building transaction systems, is not sufficient in the case of supporting KM. The task of a system designer to refer to organisational knowledge resources, indicate them and relate them to the stages of a business process in which they are to be used. As a result, it is necessary to indicate the ontology that defines the scope of terms used by the knowledge management system as well as objects that are defined by means of these terms. In the proposed approach, ontology is defined using the author-developed editor, which uses notation in compliance with OWL 2 specification. The preliminary definition of a business process and knowledge resources can precede the process of designing a business process oriented autopoietic knowledge management support system. The applied methodology refers to creating the context of its use, architecture, impact on the environment, the rules used in the system and its iterations. The reasons for using it in the context of integration of BPM and KM have been outlined in the first part of this paper. The reasons for integration of autopoietic approach have been presented in an earlier paper of the author [20, 21]. The application of elements of the proposed methodology will refer to the process of verification of personal data protection in an organisation.

4 Example of Application

Pursuant to the Act on personal data protection in force on the territory of Poland (Personal Data Protection Act (Journal of Laws of 2016 item 922) and the Resolution of the European Parliament and European Council (UE) 2016/679 of 27 April 2016 on protection of individuals with regard to personal data processing and on free flow of such data, economic entities are obliged to implement a security policy for personal data protection. One of the aspects of this policy is security audits specified in the Act which should be cyclically carried out by an Information Security Administrator. Such audits have to be preceded by establishment of their schedule and approved by a company's Board of Directors.

The first problem with supporting an auditor's actions is connected with the fact that part of the information he/she processes as part of the audit is stored in IT systems of the audited enterprise. The systems where such data is stored do not support business processes connected with personal data protection. The second reason for using the solutions proposed in the paper is the necessity of examining the process of data

processing in the context of physical, technical and organisational security measures. This requires that the person who undertakes such activities not only possesses knowledge about the processes taking place in the organisation, their participants and processed information resources, but also checks whether physical and IT security measures work properly, which often goes beyond the auditor's competencies. Therefore, it is reasonable to apply a system for decision-making support which will facilitate the audit process. Another reason is connected with the auditor's needs regarding knowledge resources. In the case of an audit, knowledge about the organisation often has extend beyond the boundaries of the organisation, because the IT systems used in the organisation can be located at any place (e.g. as a result of using Cloud Computing), and information flows go beyond the organisation. As a result, audit-related activities have to refer to the aspect of IT systems and information flows that are located outside the audited entity. Consequently, we can list a range of premises that indicate the necessity of using the proposed approach to modelling business process oriented autopoietic knowledge management support systems:

- Lack of support of the audit process by IT systems of the audited organisation.
- Necessity of possessing knowledge about organisational, physical and technical aspects of the organisation's operation.
- Necessity of integrating not only IT systems but above all the knowledge about processes in the organisation.
- Necessity of providing the auditor with organisational knowledge about the audit process and knowledge about the organisation itself.
- Necessity of analysing business processes in the organisation and beyond.

Currently available IT systems dedicated to the aspect of personal data protection do not address these issues in a sufficient way and focus on the process of preparing audit documents rather than supporting their preparation. The following sections will present selected aspects of using the proposed methodology in the process of preparing elements of a system designed to support a decision-maker in this process. All the above-mentioned stages are supported by design tools developed by the author.

4.1 Stage 1. Analysis and Development of a Business Process

As was already mentioned, the first stage is indication of the context of the system's operation connected with modelling the structure of the business process that will be supported. For modelling of this stage, the notation BPMN has been used. The initial phase involves establishment of audit schedules which have to be approved by the Board of Directors (Fig. 2).

On this basis, cyclical audits are performed during which technical and physical security measures as well as organisational procedures are checked. This aspect is presented in Fig. 3.

A business process specified in this way indicates only an employee's tasks to be performed and is consistent with BPMN notation. In the next steps, it will be extended by KM elements. This stage has been described in detail in the paper [20].

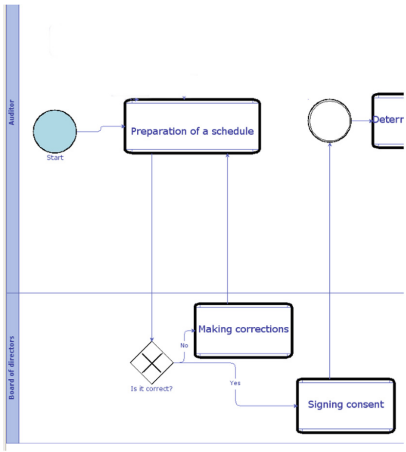

Fig. 2. Initial phase of the process of information security audit

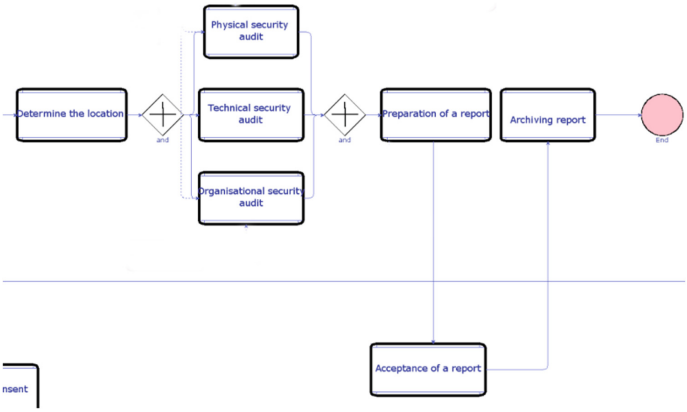

Fig. 3. Phase of an information security audit

4.2 Stage 2. Identification of Organisational Knowledge Resources

Once elements of a business process are defined (in accordance with the developed methodology), it is necessary to define which resources of organisational knowledge can be used during its performance. This process can be supported by preparation of a matrix that defines knowledge resources that will be processed during performance of a given process or specified tasks of the process that will be supported by the system. In the example, a matrix of knowledge resources to be used in the process has been defined. The example matrix has been presented in Table 2.

The knowledge resources diagnosed in this way can be represented in the system as knowledge resources used by the system. For the purpose of specification of knowledge resources, the developed software allows for the development of an ontology

Table 2. Fragment of knowledge resources necessary during an information security audit

Knowledge resource	Source	Type	Description
Leave planning schedule	ERP system	Electronic	Information on the leave planning schedule for audited employees
Building layout	Archive	Electronic/paper	Plan of the rooms in the audited organisation
Site plan	Archive	Electronic/paper	Plan of the audited room
Certificate of the validity of inspections and systems	Organisational unit	Electronic/paper	Documents confirming the validity of the inspection of fire extinguishers, alarm system, fire-extinguishing system, UPC, anti-virus system
Collection of data sets	ABI	Electronic/paper	The enterprise's personal data sets
List of authorisations	ABI	Electronic/paper	Authorisations to process data
…	…	…	…

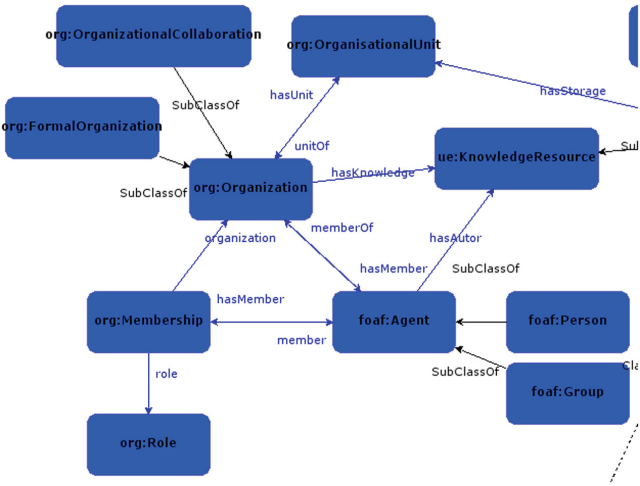

Fig. 4. Elements of the domain ontology of the defined knowledge resources of an organisation based on The Organization Ontology

diagram which uses terms applied in OWL 2 [33]. This enables a knowledge engineer to adapt the specification of organisational knowledge resources to the requirements and apply commonly used ontologies that allow for the description of the area of the problem being modelled. The design presented in the paper uses elements of the specification The Organization Ontology (W3C Recommendation from 2014) [34]. This ontology uses a range of concepts that enable the description of the structure of an organisation, its members and roles. Figure 4 shows elements of this ontology and elements that extent it, as defined by a knowledge engineer. The defined classes and property assertion are consistent with the specification OWL 2.

Knowledge resources defined in this way can be ascribed to a business process and used during specification of an autopoietic system. A knowledge engineer can ascribe certain knowledge resources in the form of class instances to the process defined at stage 1, define the whole ontology or its fragment. Thanks to that, the person performing the process will have access not only to one/several instances of knowledge resources, but the whole database. The aspect of the context of connecting OWL 2 and BPMN notations has been described in detail in the paper [20]. The stages of the verification of the physical resources have been extended by elements of the presented ontology. This makes it possible to indicate which elements of organisational knowledge will be processed as part of a business process. At the same time, elements of knowledge resources, thanks to their semantic codification, can be used in the development of an autopoietic system. The definition covering the semantics of knowledge structures and their implementation can be included into the knowledge database of knowledge management systems and processed by the other IT systems in an organisation.

4.3 Stage 3. Designing and Implementation of a Process Oriented Autopoietic Knowledge Management Support System

In accordance with the proposed methodology (stage 3.1), the process of designing the system begins with definition of a set of tasks performed by the system. To show the elements of the proposed methodology, the following set of tasks, as presented in Fig. 5, has been defined. In accordance with the defined Table 2, the elements of the tasks performed by the autopoietic system refer to selected knowledge resources. The possible defined tasks include:

- Providing knowledge resources concerning the leave planning schedule.
- Providing knowledge resources concerning the building layout.
- Providing knowledge resources concerning the site plan.
- Providing knowledge resources concerning the certificates.
- Providing knowledge concerning data sets.
- Providing knowledge concerning authorisations.

The main tasks, which have been defined in this way, can be presented based on a use case diagram and by means of a diagram of the hierarchy of a system's tasks. Each of the defined tasks is subject to a separate iteration during the development of an autopoietic system and is treated as one case of its usage. Figure 5 shows the results of one iteration of the developed system with defined tasks of an autopoietic system linked with the task of a business process. This diagram is built during the execution of stage 3.2.

The tasks, which are defined in this way, can be combined with roles that are performed by an autopoietic element. These roles may refer to the process of knowledge processing or actions connected with provision of knowledge to the system from external systems. Figure 6 shows an example of defined roles of a system as part of the designed system.

After defining the tasks to be performed by an autopoietic system, the scope of knowledge and roles of the system's elements, it is possible to design a diagram of

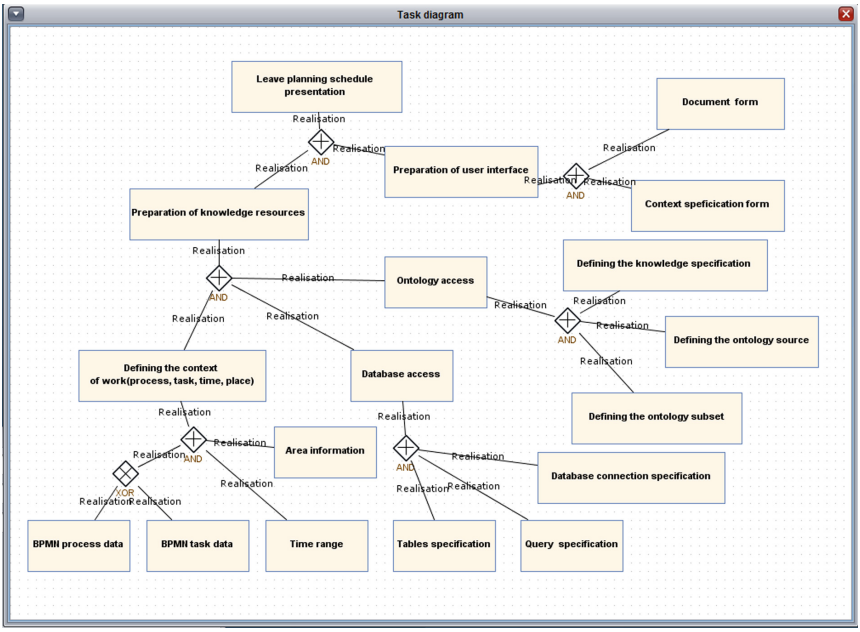

Fig. 5. Diagram of the hierarchy of an autopoietic system's tasks (providing knowledge resources concerning the leave planning schedule)

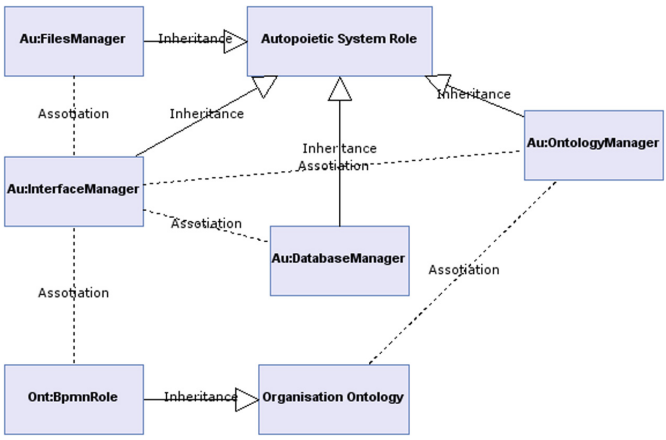

Fig. 6. Diagram of roles of an autopoietic system

programming classes which, thanks to the earlier stages, will be linked to knowledge resources, the system's tasks and roles. On this basis, it is possible to define the code of the autopoietic element. Figure 7 presents a fragment of an autopoietic system class.

As a result, the autopoietic element, which is prepared in this way, is able to provide a decision-maker with specific knowledge when a given task performed by a

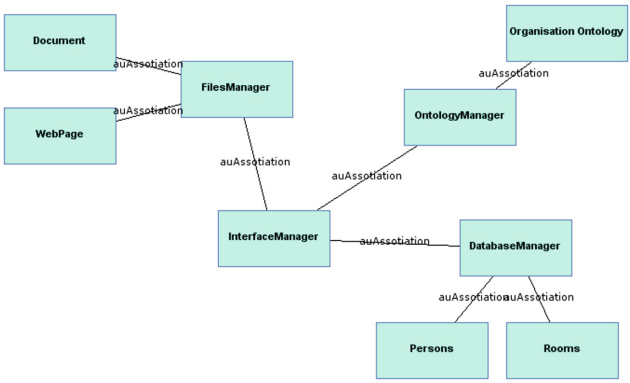

Fig. 7. Diagram of classes of a designed system

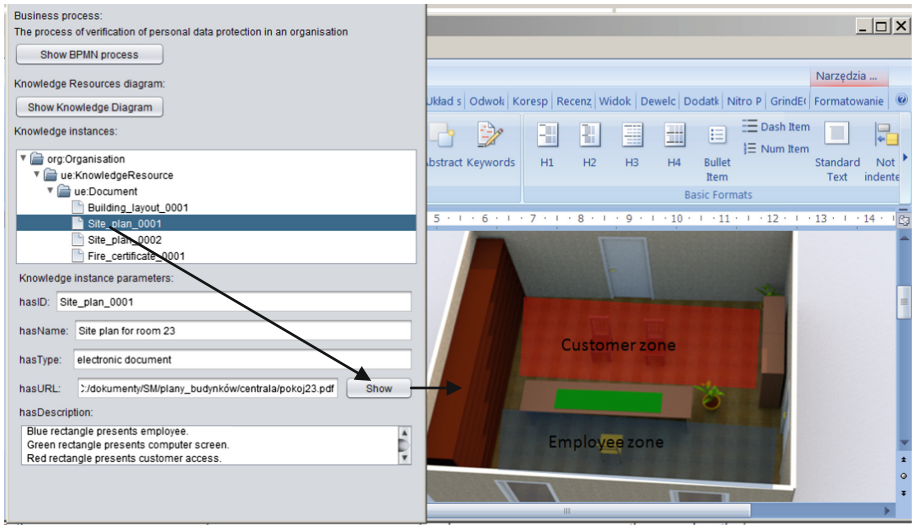

Fig. 8. Example of the operation of the developed system (site plan specification)

business process is triggered. In this example, it supports, through defined knowledge resources, the defined stages of the verification of data processing compliance.

An element of the user interface of the developed system has been presented in Fig. 8.

Figure 8 presents a fragment of a report from an audit process and instances of knowledge resources to which the auditor has access. The design elements presented here fragmentarily address the specification of the system, which can be defined based on the proposed methodology. They do not tackle the aspect of communication between the system's elements or analysis of actions undertaken by autopoietic elements. Examples of the application of these elements of the methodology will be the subject of the author's further research.

5 Conclusion

The issues of integration of business processes and organisational knowledge management require addressing the problems discussed above and have to be considered in the context of participants of this process. The paper showed an example of a methodology for designing and building a business process oriented autopoietic knowledge management support system. The presented methodology for supporting the integration of BPM and KM was extended to include the elements of autopoiesis designed to support a process being executed. This paper included only selected stages of designing a system that integrates BPM and KM. In particular, it presented the process of specifying an organisation's business process, the process of defining organisational knowledge in the context of tasks undertaken in a process, the aspect of extending the specification of a business process to include elements of knowledge resources and elements of an autopoietic system specification.

5.1 Research Contribution

The theories and their practical application as presented in the chapter represent a continuation of the author's research in the area of building business process oriented autopoietic knowledge management support systems. The research presented in this paper as well as the author's earlier papers (indicated in the text of the chapter) addresses the following research topics:

- Analysis of BPM and KM theories in terms of the possibility of integrating both these approaches to support the operation of an organisation.
- Pointing out the main problems (addressed in the literature) of the process of integrating both these approaches with reference to the development of process oriented knowledge management systems.
- Indicating current design methods used in designing process oriented knowledge management systems.
- Analysis of the theory of autopoietic systems and its use for supporting the process of integrating BPM and KM.
- Developing guidelines for building autopoietic systems to support the integration of BPM and KM and the decision maker.
- Formulating the assumptions of the methodology for designing a business process oriented autopoietic knowledge management support system.
- Evaluation of the methodology based on the performance of a developed software solution.

As a result, the research indicates the possibility of extending the theory of integrating BPM and KM (with reference to the development of process oriented knowledge management systems) by new research trends focused on machine processing of an organisation's knowledge and knowledge of its business processes, considered within the theory of autopoietic systems.

5.2 Research Implication

The research presented in the chapter can be useful for those who seek ways to build process oriented knowledge management systems focused on machine processing of an organisation's knowledge. In the presented approach, the process of modelling an organisation's business processes and knowledge resources has been extended to include stages that support the development of an IT system considered in the context of the theory of autopoietic systems. The experience gained during the implementation of the presented project made it possible to point out a range of advantages of using this approach. They include:

- Supporting decision-making processes of decision-makers by providing them with contextual knowledge. This knowledge refers to the business process that is supported and the knowledge resources that are used during the execution of this process.
- Integration of the ontology on organisational knowledge resources, which can be used in the subsequent iterations of the process of building a system. Codified knowledge can be used during the development of new software solutions and can be published on an organisation's corporate portal.
- Possibility of terminological integration of defined knowledge resources within the framework of the terms used in standardised ontologies (the example shows elements of integration of a domain ontology with The Organization Ontology, which is a standard of W3C). Thanks to that, the knowledge codified in this way can be used while building artificial intelligence systems designed to support text analysis.
- The use of elements of BPMN notation and extension of its artefacts by elements used by a knowledge engineer in designing a system. As was pointed out in the paper, one of the problems in designing integration of BPM and KM is the problem related to standards that support their design. In the example discussed in this paper, well-known and applied approaches to designing business processes and organisational knowledge resources were used.
- Indication of methods for integrating autopoietic systems as part of decision-making processes of a decision-maker. The reasons for adopting the autopoietic approach have been discussed in the author's paper [21].
- Iterational character of the approach with respect to certain tasks of a business process. Thanks to that, new functions of a system are added during successive iterations. This contributes to better understanding of the operation of a system and allows it to be tested after new functions have been added.
- Possibility of using this approach to build systems that automate business processes and systems designed to support business decisions. In particular, based on autopoietic systems developed in the context of the theory of software agent societies.

Apart from the advantages connected with the applied designing methodology, the work on the project revealed more advantages related to the support of business processes to be supported by developed software. The execution of the project resulted in the improvement of the understanding of processes that take place in an organisation, the time of their completion and the aspect of sharing knowledge about the process

itself as well as the knowledge necessary for its execution. In particular, the following impact of the implementation carried out on an organisation's operation can be highlighted:

- Definition of a range of business processes related with the process of personal data protection audit, which has contributed to audited persons' better understanding of the principles of the organisation's operation. Thanks to their codification in BPMN, the person subject to an audit process knows its rules and how it is performed. Additionally, business processes defined in this way can become an element of a map of business processes taking place in an organisation and be used by an autopoietic system.
- Development of the ontology of organisational knowledge resources, which allows knowledge resources to be linked with business processes performed in an organisation. Defined knowledge resources can be used in the process of extending the functionality of the developed software solution. Additionally, if further processes taking place in an organisation are diagnosed, they can be re-used and assigned to further tasks of a business process.
- Inclusion of defined knowledge resources into the operation of an organisation's knowledge portal. Thanks to that, knowledge on the performance of the process of verification of personal data protection and other defined business processes can be made available to employees and help them to better understand how the organisation works.

As a result, the proposed solution supports a typical life cycle of the process of knowledge management and applies to knowledge generation, knowledge evaluation, knowledge sharing, knowledge leveraging and knowledge discovery. From the perspective of an auditor, the solution has contributed to acceleration of the process of preparing post audit documentation through its partial automation due to the development of an editor for post audit documents.

5.3 Research Limitation and Future Works

The main limitation of the method presented in the paper is its focus on BPMN and OWL language, which can be used to model an organisation's knowledge resources. As a result, those who model the system need to know both these solutions. Another limitation is that its use is time-consuming due to iterative cycles that are executed to add new features of the system. As a result, development of the BPMN model of a process (stage 1) and knowledge resources used by process participants (stage 2) does not result in completion of these stages, as going to stage 3 may result in the emergence of new knowledge, generated and processed by the autopoietic system. The theory and example of using a solution designed to support an organisation's operations as presented in this chapter do not exhaust the issues addressed in it. The author's future research will focus on defining measures that can be used to evaluate the performance of the system, and on developing a tool to support the processes of simulating the progress of business processes in an organisation using an autopoietic system.

References

1. Jung, J., Choi, I., Song, M.: An integration architecture for knowledge management systems and business process management systems. Comput. Ind. **58**, 21–34 (2007). https://doi.org/10.1016/j.compind.2006.03.001
2. Scholl, W., Konig, C., Meyer, B., Heisig, P.: The future of knowledge management: an international delphi study. J. Knowl. Mang. **8**, 19–35 (2004). https://doi.org/10.1108/13673270410529082
3. Gourova, E., Toteva, K.: Design of knowledge management systems. In: Proceedings of the 8th Nordic Conference on Pattern Languages of Programs (VikingPLoP), pp. 1–15. ACM Press, New York (2014)
4. Scholl, W., Heisig, P.: Delphi study on the future of knowledge management - overview of the results. In: Mertins, K., Heisig, P., Vorbeck, J. (eds.) knowledge management, pp. 179–190. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-24778-4_8
5. Al-Mabrouk, K.: Critical success factors affecting knowledge management adoption: a review of the literature. In: Innovations in Information Technology, pp. 1–6. IEEE (2006)
6. Choy Chong, S.: KM critical success factors: a comparison of perceived importance versus implementation in Malaysian ICT companies. Learn. Organ. **13**, 230–256 (2006). https://doi.org/10.1108/09696470610661108
7. Akhavan, P., Jafari, M., Fathian, M.: Critical success factors of knowledge management systems: a multi-case analysis. Eur. Bus. Rev. J. **18**, 97–113 (2006). https://doi.org/10.1108/09555340610651820
8. Records, L.R.: The fusion of process and knowledge management. BPTrends 1–11 (2005)
9. Metaxiotis, K., Ergazakis, K., Psarras, J.: Exploring the world of knowledge management: agreements and disagreements in the academic/practitioner community. J. Knowl. Manag. **9**, 6–18 (2005). https://doi.org/10.1108/13673270510590182
10. Żytniewski, M.: Modeling of contextual knowledge about the user with the use of the ontology description languages. Studia Ekonomiczne - Uniwersytet Ekonomiczny w Katowicach Informatyka i Ekonometria **216**, 160–169 (2015)
11. Żytniewski, M.: Integration of knowledge management systems and business processes using multi-agent systems. Int. J. Comput. Intell. Stud. **5**, 180–196 (2016). https://doi.org/10.1504/IJCISTUDIES.2016.077131
12. Żytniewski, M., Kowal, R.: Using software agents to enhance the functionality of social knowledge portal. In: Abramowicz, W. (ed.) BIS 2013. LNBIP, vol. 160, pp. 23–34. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41687-3_4
13. Żytniewski, M., Sołtysik, A., Kowal, R.: Creation of software agents' society from the perspective of implementation companies. The advantages of their use, the problems of construction and unique features. Bus. Inform. 162–171 (2013)
14. Żytniewski, M.: Application of the software agents society in the knowledge management system life cycle. Econ. Stud. **153**, 191–201 (2012)
15. Power, J.: Distributed systems and self-organization. In: Proceedings of the 1990 ACM Annual Conference on Cooperation - CSC 1990, pp. 379–384. ACM Press, New York (1990)
16. Żytniewski, M.: Comparison of methodologies for agents' software society modeling processes in support for the needs of a knowledge-based organization. In: Kiełtyka, L., Niedbał, R. (eds.) Wybrane zastosowania technologii informacyjnych zarządzania w organizacjach, pp. 15–26. Publishing House of University of Technology in Częstochowa (2015)

17. Żytniewski, M., Sołtysik, A., Sołtysik-Piorunkiewicz, A., Kopka, B.: Modelling of software agents in knowledge-based organisations. analysis of proposed research tools. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 91–108. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_6
18. Żytniewski, M.: Autopoiesis of knowledge management systems supported by software agent societies. In: Refereed Paper Proceedings - KM Conference 2017 – Novo Mesto, Slovenia, pp. 41–52. A Publication of the International Institute for Applied Knowledge Management (2017)
19. Żytniewski, M.: Gossip and ostracism in modelling automorphosis of multi-agent systems. In: Gołuchowski, J., Pańkowska, M., Linger, H., Barry, C., Lang, M., Schneider, C. (eds.) Complexity in Information Systems Development. LNISO, vol. 22, pp. 135–150. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52593-8_9
20. Żytniewski, M.: Context specification in support of business processes and knowledge management integration. In: Wilimowska, Z., Borzemski, L., Świątek, J. (eds.) ISAT 2017. AISC, vol. 657, pp. 153–163. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67223-6_15
21. Żytniewski, M.: Business process oriented autopoietic knowledge management support system design. In: Paspallis, N., Raspopoulos, M., Barry, C., Lang, M., Linger, H., Schneider, C. (eds.) Information Systems Development: Advances in Methods, Tools and Management (ISD 2017 Proceedings), University of Central Lancashire, Cyprus (2017)
22. Maier, R., Remus, U.: Implementing process-oriented knowledge management strategies. J. Knowl. Manag. **7**, 62–74 (2003). https://doi.org/10.1108/13673270310492958
23. Bitkowska, A.: Knowledge management vs business process management in contemporary enterprises. Ekonomia i Zarzadzanie **8**, 31–37 (2016). https://doi.org/10.1515/emj-2016-0014
24. Bosilj-Vukšić, V.: Business process modelling: a foundation for knowledge management. J. Inf. Organ. Sci. **30**, 185–198 (2006)
25. Slavíček, V.: Enhancing business process management with knowledge. Inf. Manag. 123–134 (2011)
26. Dudycz, H., Korczak, J.: Process of ontology design for business intelligence system. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 17–28. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_2
27. Davis, R.: Introducing ARIS. In: Davis, R. (ed.) Business Process Modelling with ARIS: A Practical Guide, pp. 15–31. Springer, London (2001). https://doi.org/10.1007/978-1-4471-0321-9_2
28. Chinosi, M., Trombetta, A.: BPMN: an introduction to the standard. Comput. Stand. Interfaces **34**, 124–134 (2012). https://doi.org/10.1016/j.csi.2011.06.002
29. Li, Q., Chen, Y.-L.: IDEF0 function modeling. In: Li, Q., Chen, Y.-L. (eds.) Modeling and Analysis of Enterprise and Information Systems, pp. 98–122. Springer, Heidelber (2009). https://doi.org/10.1007/978-3-540-89556-5_5
30. Gronau, N.: Modeling and Analyzing Knowledge Intensive Business Processes with KMDL Comprehensive Insights into Theory and Practice. GITO (2012)
31. Hois, J.: Modular Ontologies for Spatial Information. Logos-Verlag (2013)
32. Roussey, C., Pinet, F., Kang, M.A., Corcho, O.: An introduction to ontologies and ontology engineering. In: Falquet, G., Métral, C., Tweed, C., Teller, J. (eds.) Ontologies in Urban Development Projects, pp. 9–38. Springer, London (2011). https://doi.org/10.1007/978-0-85729-724-2_2
33. Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: the next step for OWL. Web Semant.: Sci. Serv. Agents World Wide Web **6**, 309–322 (2008). https://doi.org/10.1016/j.websem.2008.05.001
34. W3C: The Organization Ontology (2014)

Information Technology and Systems for Business Transformation

Data Mining for Positive Customer Reaction to Advertising in Social Media

Veera Boonjing[1] and Daranee Pimchangthong[2(✉)]

[1] International College, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok 10520, Thailand
veera.bo@kmitl.ac.th
[2] Faculty of Business Administration, Rajamangala University of Technology Thanyaburi, Klong 6, Thanyaburi 12110, Pathum Thani, Thailand
daranee_p@rmutt.ac.th

Abstract. Social media has become a source of large amounts of data that is extremely useful when data are analyzed properly. Data mining is one of the known techniques to analyze data to find hidden information from a large amount of available data without having prior hypotheses. The objectives of this research were to (1) find the most important factors influencing positive reactions from customers after seeing online advertising in social media, (2) find the most important factors influencing purchasing merchandise that are advertised online, (3) identify customer clusters characteristics that have positive reaction after seeing online advertising in social media, and (4) identify customer clusters characteristics that purchase merchandise after seeing online advertising in social media. The sample size of 370 is collected by questionnaires using convenience sampling method. Data mining with cluster analysis is used to analyze data. The findings indicate the characteristics of "product conscious" and "price conscious" clusters for customer's reaction and purchasing after seeing online advertising in social media.

Keywords: Data mining · Social media · Cluster analysis

1 Introduction

With high popularity of social media, several social media websites were developed such as Line, Facebook, Twitter, etc. and the usage rate has increased every year. Social media has become a fast, easy, and less expensive way to reach people in almost every group categorized by age, occupation, education, etc. The usage of social media is widely applicable to various business domains such as marketing, knowledge sharing, customer relationship management (CRM), collaborative activities, organizational communications, education and training, and several other areas [1]. In the highly competitive business world, social media has become a source of large amounts of data that is extremely useful once data are analyzed properly. Social media analytics can be applied to understand the user sentiments about a company or a product [2]. On the other hand, companies can understand the consumer's opinions through social media analytics [3]. Large companies that have adopted advanced data analytics capabilities

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 83–95, 2018.
https://doi.org/10.1007/978-3-319-77721-4_5

outperform their competitors by wide margins [4]. When dealing with data analytics, data mining is one of the known techniques to analyze data to find hidden information from a large amount of available data without having prior hypotheses. Data mining provides a variety of methods such as association, classification, clustering, etc. for analyzing data, but selecting a method to match with the objectives is a challenge.

Marketing campaigns with online advertising are one of the methods that businesses use for increasing purchasing motivation. Finding target customer characteristics and customer reactions to social media advertising helps to reach more customers and is useful information for a marketing campaign. Prior researches [5, 6] attempt to explore customer characteristics and customer buying behaviors using data mining techniques. The retail industry has been using data mining techniques for years to predict what their customers are likely to purchase [7]. Magos and Acatrinei [8] researched on the factors that influence the recipients to open direct emails and make the subsequent actions on purchasing using data mining techniques. However, Cardon and Marshall [9] state that social networking sites are now more commonly used for online communication than email. As social media has started to impact people's lives, a study of its usage would thus be significant [10]. In this research, the postive customer reaction to advertising in social media is examined by using data mining technique for cluster analysis. The purposes of this study are to find the most important factors influencing positive reactions from customers after seeing online advertising in social media, find the most important factors influencing purchasing products that are advertised online, identify customer clusters characteristics that have positive reaction after seeing online advertising in social media, and identify customer clusters characteristics that purchase product after seeing online advertising in social media.

The paper is organized as follows. Section 2 is the literature review on data mining concepts, marketing campaign methods, and social media advertising. Section 3 explains the research methodology using the WEKA data mining software. Section 4 presents the research findings. Section 5 discusses the research findings on the *product conscious* cluster and the *price conscious* cluster. The conclusions are presented in Sect. 6.

2 Literature Review

Data mining is defined as the process to extract knowledge from large quantities of data in order to discover meaningful patterns and rules [11]. Hand et al. [12] defines data mining as the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Han and Kamber [13] defines data mining as a process of knowledge discovery. Leventhal [14] summarizes that data mining contains three key stages: finding patterns, interpreting them in order to check their usefulness, and finally using the patterns to solve business problems. The ultimate goal of data mining is to discover knowledge and it will be useful in several disciplines. In business, data mining is used for strategic benefit such as direct marketing, trend analysis, etc. In direct marketing, data mining is used for targeting people who are most likely to buy certain products and services. For trend analysis, data mining is used to determine trends in the marketplace [15].

Leventhal [14] explains that there are two main types of data mining models as follows:

1. Predictive model: This model is constructed to predict a particular outcome or target variable. Commonly used predictive modelling techniques include multiple regression (for predicting value data), logistic regression (for response prediction) and decision trees (for rule-based value or response models).
2. Descriptive model: This model gives a better understanding of the data, without any single specific target variable. Commonly used descriptive techniques include factor analysis (to extract underlying dimensions from multivariate data), cluster analysis (for grouping a customer database into segments), and association analysis (for discovering relationships between items such as retail products).

Prior data mining researches were performed using a variety of methods concerning customers' behaviors. Zekić-Sušac and Has [5] proposes an integration of two data mining techniques, artificial neural networks and association rules, to discover patterns and customers' profiles in frequent item purchases. Their results demonstrate the efficiently-used data mining techniques for recognizing patterns in shopping behavior and generating new marketing strategies. A study carried out by Noori [6] proposes an integrated data mining and customer behavior scoring model to manage existing mobile banking users. The segmentation model is developed to identify customers groups based on transaction history, recency, frequency, and monetary background. The ability to identify customers by a behavioral scoring facilitates marketing strategy assignment and can support marketing actions to attract more customers, maintain customers, and keep high customers' satisfaction. Pawełoszek [16] explains that cluster analysis is a useful and flexible method to discover structures in a data set, however it does not provide any explanation itself - the semantic interpretation of clustering effects can be performed by combining the data of each cluster with the knowledge of experienced members of the implementation team.

A marketing campaign is a specifically defined series of activities used in marketing a new or changed product or service, or in using new marketing channels and methods. Marketing activities are efforts to increase awareness for a particular product or service. Social media is one of the most popular marketing channels due to the ability to reach large numbers of customers with low cost. Social media advertising helps businesses find new potential customers by using the users' own shared information to identify interest. Rather than reactively targeting users who search for a certain term, social media advertising proactively targets relevant users before they even begin their searches. Research on social media has focused on areas such as brand user segmentation and participation, electronic word-of-mouth communication and online brand communities [17]. Adi et al. [18] proposes that social media browsing intention mediates the relationship between online shopping orientation and electronic word of mouth. When online shoppers browse through social media, this online information search will positively influence consumer attitudes and subsequently promote information sharing with peers while also influencing their purchase intentions.

Prior research that relates to the study topics were reviewed as follows: Magos and Acatrinei [8] researched the factors that influence the recipients to open direct emails and make an action desired by the company and also studies whether and what elements in the

email would influence them to buy the products or services promoted. The results are obtained based on a data mining analysis that includes clustering and classification processes and offers a guide on how organizations should design their email marketing communications in order to have higher response rates. Singh and Peszynski [19] researched on the value of social technologies in organizations based on the 'value focused thinking' approach. The findings highlight innovation of internal processes, creation of organizational identity and new business models, integrated business functions, as well as employee support to be important values of social technology enabled innovation in organizations. Campbell et al. [20] researched on the segmenting of consumer reactions to social network marketing. The purpose of the study is to understand how consumers may be segmented with respect to their reactions to social network marketing. Along with significant covariates such as information search, convenience, entertainment, age, and gender that predict membership, the results identified five segments: Passives (mostly ignore social network marketing and seek entertainment), Talkers (strongly influenced by "word of mouth" and social network marketing), Hesitants (less influenced by social network marketing), Actives (strongly influenced by social network marketing), and Averse (almost entirely ignore social network marketing).

3 Research Methodology

Several aspects related to the factors that influence a respondent to react and purchase a product or service after seeing online advertising in social media are explored from the literature review. Based on the literature review, the questionnaires are developed and distributed to people who used to purchase product or service online through social media such as Facebook, Line, Instagram, etc. The questionnaires are composed of three parts; the first part is about the demographic data of the respondents, the second part is about the respondent's reaction when seeing the marketing campaign in social media, and the third part is open-ended questions about the respondent's opinion. The data has been gathered at the level of samples including 370 respondents aged less than 60 and being either employees, freelancers, entrepreneurs, managers, and/or students. The data has been collected between January and April 2017 through an online survey by using the convenience sampling method.

The research directions include: (O1) determining the most important factors that influence the customers to have a positive reaction after seeing online advertising in social media; (O2) determining the most important factors influencing purchasing products that advertise online in social media; (S1) identifying the customers characteristics that have positive reaction after seeing online advertising in social media; and (S2) identifying the customer clusters characteristics that purchase merchandises after seeing online advertising in social media.

The data mining methodology used was Cross Industry Standard Process for Data Mining (CRISP-DM) and the methodological steps were followed: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [22]. The software used to analyze data is WEKA data mining software [21] and the attribute evaluator is "GainRatioAttributeEval" for selecting attributes and "SimpleKMeans" clustering algorithm for grouping similar customers in groups.

4 Research Findings

The analysis results indicate that most of the respondents are students, females, aged 21–30, single, and educational level of Bachelor's degree, and average income less than 15,000 THB. The highest percentage on the usage of social media is Facebook,

Table 1. Descriptive results from respondent's general information.

Gender	Frequency	Percentage
Males	117	31.6
Females	253	68.4
Total	370	100

Age	Frequency	Percentage	Average income	Frequency	Percentage
< 21 years	69	18.6	<=15,000	243	65.7
21-30 years	244	65.9	>15,000 - 20,000	66	17.8
31-40 years	26	7.0	>20,000 - 30,000	27	7.3
>41 years	31	8.4	>30,000	34	9.2
Total	370	100	Total	370	100

Status	Frequency	Percentage	Education	Frequency	Percentage
Single	317	85.7	<Bachelors'degree	54	14.6
Married	38	10.3	Bachelors' degree	283	76.5
Others	15	4.1	>Bachelors'degree	33	8.9
Total	370	100	Total	370	100

Occupation	Frequency	Percentage	Social media used	Frequency	Percentage
Private	78	21.1	Facebook	269	72.7
Government	41	11.1	Line	50	13.5
Freelance	23	6.2	Instagram	31	8.4
Business owner	13	3.5	Twitter	10	2.7
Student	215	58.1	Others	10	2.7
Total	370	100	Total	370	100

Spent/transaction	Frequency	Percentage	Merchandise	Frequency	Percentage
<= 500 bath	117	31.6	Fashion	200	54.1
>500 – 1,000 bath	138	37.3	Accessories	18	4.9
>1,000 – 2,000 bath	51	13.8	Beauty and health	35	9.5
>2,000 – 3,000 bath	22	5.9	IT and software	54	14.6
>3,000 – 4,000 bath	42	11.4	Airplane ticket	12	3.2
			Entertain	33	8.9
			Others	18	4.8
Total	370	100	Total	370	100

followed by Line and Instagram. The most purchasing is in fashion merchandise, followed by IT and software, and beauty and health merchandise. Spending per transaction is between 500 to 1,000 THB as shown in Table 1.

In order to follow the research directions, the data mining analysis is comprised of an attribute selection and ranking according to a class attribute (A) and a clustering process (B).

To identify the most important factors that influence the customers to have a positive reaction after seeing online advertising in social media (O1), the class attribute Q13 – ReasonForClickingAdvertisementOnSocial Media is selected.

A. Ranking attributes. This step is done in order to identify the relation between the class attribute and the others by using "weka.attributeSelection.CorrelationAttribute Eval", and the results presented in Table 2.

Table 2. Ranked attributes of Q13_ReasonForClickingAdvertisementOnSocialMedia and the others.

0.12727	16 Q15_4	Advertise influence buying: satisfaction with merchandises or services
0.08522	22 Q16_3	Marketing campaign influence buying: special rewards
0.08520	11 Q14_5	Influence of advertisement on social media: not buying
0.08507	4 Q12_4	Interested in seeing advertisement on social media: link for searching more details
0.07962	17 Q15_5	Advertise influence buying: more refer in social media
0.07643	13 Q15_1	Advertise influence buying: interesting price
0.07573	1 Q12_1	Interested in seeing advertisement on social media: merchandise logo
0.07475	18 Q15_6	Advertise influence buying: like marketing campaign
0.07273	9 Q14_3	Influence of advertisement on social media: sent it out for more comments
0.06776	2 Q12_2	Interested in seeing advertisement on social media: text, image, clip
0.06414	24 Q16_5	Marketing campaign influence buying: others
0.05661	21 Q16_2	Marketing campaign influence buying: having complimentary
0.03896	5 Q12_5	Interested in seeing advertisement on social media: package or merchandise
0.03451	20 Q16_1	Marketing campaign influence buying: high percentage of discount
0.03409	7 Q14_1	Influence of advertisement on social media: immediately buy if satisfy
0.03283	10 Q14_4	Influence of advertisement on social media: irritating and don't like advertisement
0.03194	6 Q12_6	Interested in seeing advertisement on social media: others
0.03194	12 Q14_6	Influence of advertisement on social media: others
0.03086	14 Q15_2	Advertise influence buying: popular brand
0.0304	15 Q15_3	Advertise influence buying: exciting new merchandise
0.02497	3 Q12_3	Interested in seeing advertisement on social media: marketing campaign
0.02362	19 Q15_7	Advertise influence buying: others
0.01728	8 Q14_2	Influence of advertisement on social media: saving information for further consideration
0.0063	23 Q16_4	Marketing campaign influence buying: having after sales service

The results present that for the customers, there is a strong correlation between the reason for clicking advertisement on social media and the satisfaction with merchandise (0.12727), special rewards (0.08522), not buying (0.0852), and link for searching on more details (0.08507). Other attributes are ranked below 0.07.

B. Clustering process. The EM (Expectation Maximization) and "SimpleKMeans" clustering algorithms are used for grouping similar customers in groups. The EM algorithm is used to identify the approximated cluster numbers. In this case, the result is two clusters. This value is used as a parameter for the "SimpleKMeans" algorithm. The algorithm results are presented in Table 3.

Table 3. Customers clusters based on positive reaction after seeing online advertising in social media.

No.	Attributes	Cluster 0 (94, 25.4%)	Cluster 1 (276, 74.6%)
1	Q12_1_Interested in seeing advertisement on social media: merchandise logo	0.2128	0.0725
2	Q12_2_Interested in seeing advertisement on social media: text, image, and clip	0.266	**0.6522**
3	Q12_3_Interested in seeing advertisement on social media: marketing campaign	0.2766	0.1739
4	Q12_4_Interested in seeing advertisement on social media: link for searching more details	0.2234	0.0725
5	Q12_5_Interested in seeing advertisement on social media: package or merchandise	0.0213	0.0254
6	Q12_6_Interested in seeing advertisement on social media: others	0	0.0036
7	Q14_1_Influence of advertisement on social media: immediately purchase if satisfy	0.1596	0.1667
8	Q14_2_Influence of advertisement on social media: saving information for further consideration	**0.7447**	**0.6993**
9	Q14_3_Influence of advertisement on social media: sent it out for more comments	0.0426	0.0326
10	Q14_4_Influence of advertisement on social media: irritating and don't like advertisement	0.0532	0.0725
11	Q14_5_Influence of advertisement on social media: not buying	0	0.0254
12	Q14_6_Influence of advertisement on social media: others	0	0.0036
13	Q15_1_Advertise influence buying: interesting price	0.3617	0.3659
14	Q15_2_Advertise influence buying: popular brand	0.1383	0.1268
15	Q15_3_Advertise influence buying: exciting new merchandise	0.0745	0.0507
16	Q15_4_Advertise influence buying: satisfaction with merchandises or services	0.2872	0.3732

(*continued*)

Table 3. (*continued*)

No.	Attributes	Cluster 0 (94, 25.4%)	Cluster 1 (276, 74.6%)
17	Q15_5_Advertise influence buying: more refer in social media	0.1064	0.058
18	Q15_6_Advertise influence buying: like marketing campaign	0.0213	0.0217
19	Q15_7_Advertise influence buying: others	0.0106	0.0036
20	Q16_1_Marketing campaign influence buying: high percentage of discount	**0.6383**	**0.6449**
21	Q16_2_Marketing campaign influence buying: having complimentary	0.1277	0.1522
22	Q16_3_Marketing campaign influence buying: special rewards	0.0319	0.0145
23	Q16_4_Marketing campaign influence buying: having after sales service	0.2021	0.1739
24	Q16_5_Marketing campaign influence buying: others	0	0.0145

Cluster 0. Most customers in this group save information for further consideration (74%) and are interested in high percentage of discount (64%) after seeing advertising in the social media. For this cluster, the advertising should emphasize on merchandise quality as well as price to create more positive reaction.

Cluster 1. Most customers in this group save information for further consideration (70%), are interested in high percentage of discount (64%) after seeing advertising in the social media, and are also interested in seeing text, image, and clip advertising (65%). For this cluster, clip advertising may increase the customer's interest and create more positive reaction.

In order to determine the most important factors influencing purchasing products that advertise online in social media (O2), the class attribute Q1_PurchasingMerchandiseOnline is selected.

A. Ranking attributes. This step is done in order to identify the relation between the class attribute and the others by using "weka.attributeSelection.CorrelationAttributeEval", and the results presented in Table 4.

The results present that for the customers, there is a strong correlation between purchasing merchandise online and saving information for further consideration (0.13879), merchandise logo (0.10179), and immediately purchase if satisfied (0.09747). Other attributes are ranked below 0.09.

B. Clustering process. This process is similar to the one from the previous section. The results are presented in Table 5.

Cluster 0. This group of customers can be called "Product conscious." Most customers in this group save information for further consideration (74%) after seeing advertising on the social media and purchase when satisfied with merchandise or services (55%). Seeing text, image, and clip advertising is also interesting for this group (53%). For this cluster, the advertising should emphasize on quality and brand to justify the price.

Table 4. Rank attributes of Q1_PurchasingMerchandiseOnline and the others.

0.13879	8 Q14_2	Influence of advertisement on social media: saving information for further consideration
0.10179	1 Q12_1	Interested in seeing advertisement on social media: merchandise logo
0.09747	7 Q14_1	Influence of advertisement on social media: immediately purchase if satisfy
0.08843	10 Q14_4	Influence of advertisement on social media: irritating and don't like advertisement
0.08675	3 Q12_3	Interested in seeing advertisement on social media: marketing campaign
0.06656	13 Q15_1	Advertise influence buying: interesting price
0.05314	16 Q15_4	Advertise influence buying: satisfaction with merchandises or services
0.04368	5 Q12_5	Interested in seeing advertisement on social media: package or merchandise
0.04144	2 Q12_2	Interested in seeing advertisement on social media: text, image, clip
0.03877	11 Q14_5	Influence of advertisement on social media: not buying
0.03841	22 Q16_3	Marketing campaign influence buying: special rewards
0.0357	17 Q15_5	Advertise influence buying: more refer in social media
0.03195	4 Q12_4	Interested in seeing advertisement on social media: link for searching more details
0.03025	18 Q15_6	Advertise influence buying: like marketing campaign
0.02892	24 Q16_5	Marketing campaign influence buying: others
0.02329	15 Q15_3	Advertise influence buying: exciting new merchandise
0.02039	19 Q15_7	Advertise influence buying: others
0.01614	20 Q16_1	Marketing campaign influence buying: high percentage of discount
0.01488	14 Q15_2	Advertise influence buying: popular brand
0.0144	12 Q14_6	Influence of advertisement on social media: others
0.0144	6 Q12_6	Interested in seeing advertisement on social media: others
0.00799	23 Q16_4	Marketing campaign influence buying: having after sales service
0.00383	9 Q14_3	Influence of advertisement on social media: sent it out for more comments
0.0038	21Q16_2	Marketing campaign influence buying: having complimentary

Cluster 1. This group can be called "Price conscious." Everyone in this group is concerned about price before purchasing merchandise. Interesting price (100%) and high percentage of discount (71%) will get these customers' attention. Customers in this group also like to save information for further consideration (65%) and are interested to see text, image, and clip advertising (58%). For this cluster, the opportunity to purchase at a low price should be emphasized on the advertisement online.

Table 5. Customers cluster based on purchasing merchandises after seeing online advertising in social media.

No.	Attributes	Cluster 0 (235, 63.5%)	Cluster 1 (135, 36.5%)
1	Q12_1_Interested in seeing advertisement on social media: merchandise logo	0.0894	0.1407
2	Q12_2_Interested in seeing advertisement on social media: text, image, and clip advertising	**0.5362**	**0.5852**
3	Q12_3_Interested in seeing advertisement on social media: marketing campaign	0.2383	0.1333
4	Q12_4_Interested in seeing advertisement on social media: link for searching more details	0.1021	0.1259
5	Q12_5_Interested in seeing advertisement on social media: package or merchandise	0.0298	0.0148
6	Q12_6_Interested in seeing advertisement on social media: others	0.0043	0
7	Q14_1_Influence of advertisement on social media: immediately purchase if satisfy	0.1234	0.237
8	Q14_2_Influence of advertisement on social media: saving information for further consideration	**0.7489**	**0.6444**
9	Q14_3_Influence of advertisement on social media: sent it out for more comments	0.0468	0.0148
10	Q14_4_Influence of advertisement on social media: irritating and don't like advertisement	0.0596	0.0815
11	Q14_5_Influence of advertisement on social media: not buying	0.017	0.0222
12	Q14_6_Influence of advertisement on social media: others	0.0043	0
13	Q15_1_Advertise influence buying: interesting price	0	**1**
14	Q15_2_Advertise influence buying: popular brand	0.2043	0
15	Q15_3_Advertise influence buying: exciting new merchandise	0.0894	0
16	Q15_4_Advertise influence buying: satisfaction with merchandises or services	**0.5532**	0
17	Q15_5_Advertise influence buying: more refer in social media	0.1106	0
18	Q15_6_Advertise influence buying: like marketing campaign	0.034	0
19	Q15_7_Advertise influence buying: others	0.0085	0
20	Q16_1_Marketing campaign influence buying: high percentage of discount	**0.6043**	**0.7111**
21	Q16_2_Marketing campaign influence buying: having complimentary	0.1362	0.163
22	Q16_3_Marketing campaign influence buying: special rewards	0.017	0.0222
23	Q16_4_Marketing campaign influence buying: having after sales service	0.2255	0.1037
24	Q16_5_ Marketing campaign influence buying: others	0.017	0

5 Discussion of Findings

The finding from this research are composed of two major issues. First, factors that influence positive reaction after seeing online advertising in social media and the cluster characteristics of this group. Second, factors that influence purchasing merchandise after seeing online advertising in social media and the cluster characteristics of this group. Based on the findings from this research, the most important factors influencing positive reactions from customers after seeing online advertising in social media are the satisfaction with merchandise and special rewards. For customer clusters characteristics, the findings indicate two clusters with the following cluster characteristics as common characteristics for both clusters: saving information for further consideration, and interested in high percentage of discount. However, the findings also indicate that one cluster has the cluster characteristic for being interested in seeing text, image, and clip advertising. E-commerce business should draw attention to the content that is emphasized in the advertisement online. If the contents are attractive and interesting, then customers may feel encouraged to share it on other social networks. These reposts have more value among the customers since they come from a trusted source [23]. Finally, merchandise quality and price create a reaction that is more positive for both clusters.

The strong correlations between *purchasing merchandise online* and *saving information for further consideration*, *merchandise logo*, and *immediately purchase if satisfied* indicate that most customers want to reconsider before purchasing products unless they are familiar with merchandise logo. However, if they are satisfied with the products, then they will immediately purchase. For customer clusters that purchase merchandise after seeing online advertising in social media, there are two clusters: the *product conscious* cluster and the *price conscious* cluster. For the product conscious cluster, price does not matter as much for shoppers and they decide to purchase based on satisfaction with the merchandise after thoroughly considering. Merchandise quality alone can get the product conscious group to purchase or have positive reaction after seeing online advertising in social media. For the price conscious cluster, they are looking for an interesting price and a high percentage of discount. Offering interesting price will draw high attention from the price conscious cluster. However, both clusters are interested in the advertisement online, and the advertisement that can draw customer's attention should provide interesting text, image, and clip advertisement. The advertisement should encompass a call-to-action so that it would convince the customers to click on the advertisement and continue to purchase the merchandise. The advertisement should also emphasize the high percentage of discount, which can draw attention from both clusters from positive reaction and lead to the purchasing of merchandise after seeing online advertising in social media.

6 Conclusion

Data mining with cluster analysis method is one of the useful tools for analyzing large amounts of data to extract group characteristics. This research is able to identify two important group characteristics that have positive reaction and purchase merchandise

after seeing online advertising in social media. The major findings are the product-conscious and price-conscious clusters, and proposing the interesting ways to draw customer attention based on mined information. The major limitation of the research consists in the low number of respondents (only 370), so this exploratory research should be followed by a conclusive one to verify the conclusions of the present research. In addition, the majority of the respondents were students, aged 21–30, with low income and low spending per transaction, which may have influenced the positive reaction after seeing online advertising in social media. Some concerns about the data mining process in this research are the amount of data and methods, because this research only used cluster analysis.

Future directions of research may include: (1) using additional methods to analyze data such as association, classification, and neural network, which might find more interesting patterns that will be helpful for specific circumstances, (2) adding some other information about respondents like time spent in front of the computer and time spent in using social network, and analyzing the influence of those factors on positive reactions and the purchasing of merchandise.

References

1. Ngai, E.W.T., Moon, K.K., Lam, S.S., Kim, E., Tao, S.: Social media models, technologies, and applications - an academic review and case study. Ind. Manag. Data Syst. **115**(5), 769–802 (2015)
2. Mosley, R.: Social media analytics: data mining applied to insurance Twitter posts. Casualty Actuarial Soc. E-Forum **2**, 1–36 (2012)
3. Yoosin, K., Rahul, D., Jie, Z., Seung Ryul, J.: Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. Online Inf. Rev. **40**(1), 42–61 (2016)
4. Pearson, T., Wegener, R.: Big data: the organizational challenge. http://www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx. Accessed 5 Nov 2016
5. Zekić-Sušac, M., Has, A.: Data mining as support to knowledge management in marketing. Bus. Syst. Res. **6**(2), 18–30 (2015)
6. Noori, B.: An analysis of mobile banking user behavior using customer segmentation. Int. J. Global Bus. **8**(2), 55–64 (2015)
7. Astudillo, C., Bardeen, M., Cerpa, N.: Editorial: data mining in electronic commerce – support confidence. J. Theor. Appl. Electron. Commer. Res. **9**(1), I–VII (2014)
8. Magos, R.I., Acatrinei, C.A.: Designing email-marketing campaigns – a data mining approach based on consumer preferences. Annales Universitatis Apulensis Series Oeconomica. **17**(1), 15–30 (2015)
9. Cardon, P.W., Marshall, B.: The hype and reality of social media use for work collaboration and team communication. Int. J. Bus. Commun. **52**(3), 273–293 (2015)
10. Balan, S., Rege, J.: Mining for socia media: usage patterns of small businesses. Bus. Syst. Res. **8**(1), 43–50 (2017)
11. Berry, M.J.A., Linoff, G.: Data Mining Techniques for Marketing, Sales, and Customer Relationship Management, 3rd edn. Wiley, Hoboken (2011)
12. Hand, D.J., Mannila, H., Smyth, P.: Principles of Data Mining. MIT Press, Cambridge (2001)
13. Han, J., Kamber, M.: Data Mining Concepts and Techniques, 2nd edn. Morgan Kaufmann Publishers, Burlington (2006)

14. Leventhal, B.: An introduction to data mining and other techniques for advanced analytics. J. Direct, Data Digit. Mark. Pract. **12**(2), 137–153 (2010)
15. Negnevitsky, M.: Artificial Intelligence – A Guide to Intelligent Systems, 3rd edn. Pearson Education Limited, London (2005)
16. Pawełoszek, I.: Data mining approach to assessment of the ERP system from the vendor's perspective. In: Ziemba, E. (ed.) Information Technology for Management. Lecture Notes in Business Information Processing, vol. 243, pp. 125–143. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_8
17. Rambe, P., Jafeta, R.: Impact of social media advertising on high energy drink preferences and consumption. J. Appl. Bus. Res. **33**(4), 653–668 (2017)
18. Adi, P., Faizal, W., Rabiatul, W.: The role of social media browsing intention for behavioral outcomes of young consumers. Trziste = Market **29**(1), 39–57 (2017)
19. Singh, M., Peszynski, K.: Organisational value of social technologies – an Australian study. Electron. J. Inf. Syst. Eval. **17**(1), 088–099 (2014). www.ejise.com
20. Campbell, C., Ferraro, C., Sands, S.: Segmenting consumer reactions to social network marketing. Eur. J. Mark. **48**(3/4), 432–452 (2014)
21. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISPDM 1.0 step-by-step data mining guide. Technical report, CRISP-DM (2000)
22. Waikato: Weka 3 – Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/. Accessed 15 May 2017
23. Mimi, U., Poulose, K.: Opinion mining for monitoring social media communications for Brand Promotion. Int. J. Adv. Res. Comput. Sci. **8**(3), 608–611 (2017)

E-government Usage by Polish Citizens

Witold Chmielarz(✉) and Oskar Szumski

Faculty of Management, University of Warsaw, Warsaw, Poland
{witold.chmielarz,oskar.szumski}@uw.edu.pl

Abstract. The aim of this article is to present results of the research in the state of usage of basic e-government manifestations by individual users in Poland in 2017. Article presents the results of research on the popularity, use and influence of e-government functions that support their application in reality. Seventeen core e-government functions, most popular among respondents, were selected based on prior consultation as basis of following analysis. Authors conducted the CAWI analysis to evaluate the distinguished e-government functions on a selected sample of university students. A group of over two hundred and fifty randomly selected people from the university environment was examined. This approach was guided by the structure of the article consisting of the presentation of the research hypothesis, the description of the methodology and the research sample, and the analysis of obtained results and their discussion, together with the conclusions. The major findings from this research are in line with other researches carried over in Poland in the area of e-government. Results of 35% share of using e-government systems are similar to 31–33% in other studies. The research proves that the level of electronic services in e-government area in Poland is considerably lower than in most of European countries. The most popular use of digital services is used for taxation purposes. Some revival though can be noted in the social sphere (culture, public opinion, public health services, etc.), the basic problem in using of e-government services is lack of ability to handle the whole matter from start to finish via the Internet. The results can be used by people involved in the creation and development of e-government systems.

Keywords: E-government · Functions of e-government · Public e-services

1 Introduction

Introduction of a new technology also influence changes in the approach that government and other public agencies adopt to emerging, modern technology to be more efficient, transparent or closer to citizens. The major impact on the current form of communication between Government and Citizens had the development of internet. Building of consciousness of information society among citizens become the starting point to deploy the new idea of communication between government and people living in the country called e-government. Although the idea itself was crafted over 50 years ago when Statistical Offices were using first mainframes to collect statistical data of citizens and economy the name e-government understood in the way is defined currently was given in 90's of previous century when e-government became the new phenomena.

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 96–113, 2018.
https://doi.org/10.1007/978-3-319-77721-4_6

E-governance, meaning 'electronic governance' is using information and communication technologies (ICTs) at various levels of the government and the public sector and beyond, for the purpose of enhancing governance [1]. It was also the starting point of research done in the area. The e-government strategy in Poland was written for the first time in 1999 [2–4]. Studies and empirical activities including implementation projects focused on e-government have been developed from 2000 [5]. Analysing the readiness of e-government in Poland in 2016 Poland was rated as one of 50 (36) most developed countries in the area of e-government in the world with E-Government Development Index (EGDI) = 0.7211 [6, p. 151]. The E-Government Development Index [7] incorporates the access characteristics, such as the infrastructure and educational levels, to reflect how a country is using information technologies to promote access and inclusion of its people. The index is based also on the maturity of offered services and confidence of the users [8, 9].

What is interesting Poland is ranked high in the e-participation index in 2016. According agreed approach "The goal of e-participation initiatives should be to improve the citizen's access to information and public services; and promote participation in public decision-making which impacts the well-being of society, in general, and the individual, in particular" [10].

In 2016 Poland was listed at 14th position [6, p. 56] of all countries and indicated the raise of over 51 position since 2014 but the use of e-government services was rated unsatisfactory and needs some improvement [8, 11].

The goal of this article is to analyze the state of usage of basic e-government services by Polish citizens in 2017 all together with available web portals, that are supposed to be the single point of contact for Citizens. As per research done by other authors success of e-commerce portals and their quality is an important success factor of e-government services [12–14]. Based also on other studies, that show that in practice the interest in e-government is high but on the one hand, its usage is low [9, 15] authors wanted to track the trends among chosen group of students, whether the success factors of e-government have reflection in following study.

Authors performed analysis of popularity, use and influence of e-government functions that support communication between public administration and individuals via digital media. The authors selected, a priori, the range of e-government services that are the most popular for the analyzed population and are consistent with the findings and definitions of the e- public administration used in definition of the European Union recommendations. Authors performed a review of studies executed by other projects within last five years to build up the baseline for analysis and possible comparison of current state of use of public services carried in following research. Based on the defined e-government literature authors defined the research hypothesis that Polish citizens present similar level of use of e-government as status noticed 3–4 years ago. In addition, it was attempted to find out whether the current state of e-government in Poland satisfies people who are willing to use it and to what extent. The working thesis of this research was defined as "Polish citizens don't use provided e-services at the level represented by other EU countries due to inconsistent ability to handle the whole matter from start to finish via the Internet".

Authors conducted the CAWI analysis to evaluate the distinguished e-government functions on a selected sample of university students. The structure of the research consists of following sections:

- Literature review including analysis of chosen publications all together with other studies used as baseline or comparison to the research described in this publication. This section also includes initial research hypothesis that are investigated in the further sections of this research,
- Research methodology section includes description of used within the research methodology. It also includes the description of research focus group that provided feedback to the questionnaire,
- Analysis and Discussion of Results presents detailed findings based on the responses of group of respondents. The section includes analysis of particular answers received, graphical presentation of gathered results and other forms of results presentation that authors find the most beneficial for the readers and the analysis of obtained results and their discussion, together with the conclusions.

Conclusions section presents a synthetic summary of the conclusions from the whole study accompanied with recommendations for further research and possible influence to perception of e-government within Polish citizens, to achieve higher level of confidence and usage of public services provided via electronic means.

2 Literature Review

There is no definitive and complete definition of the e-government in the literature. According to the European Commission, e-government states for the use of digital tools and information systems to deliver better quality public services to citizens and businesses [16]. Polish equivalent – e-public administration - suggests narrowing this concept to public administration [17], while e-government also includes services offered by the budget sector, that go beyond the scope of the commonly understood administration [18]. This is the reason why the Central Statistical Office has defined this as the use of information and communication technology (ICT), organizational change and new competences in public administration to improve public services and planned democratic processes [19, 20]. Public administration is understood not only as an executive apparatus of the state, but also expressed as activities targeted to organize the conditions and principles of social relations (trade, education (e.g. higher education), culture (e.g. public libraries), urban transport, environmental protection, etc.) [21].

In the meantime, according to the Western definition, this is a more electronic system of information and public services. It would therefore be more appropriate to define e-public administration as an opportunity to exploit the complexity of telecommunications tools and techniques to streamline common administrative and civil services. Gathering in one place, in the virtual space, time and a single platform (public administration portal) all matters relating to a specific category of users (citizen, business) facilitates and expands the ability to handle that [22]. However, one more factor remains, that decides - about the possibility of use of technology - organizational and legal determinants.

In the European Union, public services are identified as guidelines for citizens that can be used via the Internet [23]. For individuals those services are: income tax, job search, social security, identity papers, vehicle registration, construction permits, police admissions, access to resources of public library, birth and marriage certificates, college registration, change of place of residence and access to health services.

In Poland above list is slightly modified and according to project "Wrota Polski" it looks as follow: tax registration, job search and help finding a job, getting a social pension for unemployed, handicapped and retired, obtaining a student scholarship, obtaining a personal ID obtaining a driving license, obtaining a passport, registration of a vehicle, obtaining a building permit, reporting to the police such as theft, access to public library catalogs and searching them, filing Civil status office forms and acts and obtaining required copies of those, college registration, change of place of residency and sign up for a doctor's visit [22]. Even though previously listed – limited elements still face problems in proper and timely realization.

Citizens of the European Union more often use the electronic method of contact with governmental agencies than Polish citizens. If we consider digital interaction of citizens with public institutions (excluding e-mail), Poland with the share of 31% of persons of age 25–64 (in 2015) takes 25 position in the ranking. The result is not comparable far away from Denmark (first position – 92%), behind Poland only three countries have lesser level of electronic interaction: Italy, Romania, Bulgaria. e-Public administration services are divided into three groups: downloading forms, uploading of filled applications, and searching for information. In this case Poland has nothing particular to be proud of because also in this area we are away from standard of remaining European countries (2013) we are worse from about 15% of countries in the first two cases (downloading forms, uploading of filled applications) and about 26%, in the area of searching for information [24]. Even worse it looks with regards to listed in the previous paragraph services. While in the top fifteen most developed countries of Europe the percentage of filing tax returns is 32%, in Poland it amounts to only 14.2%; percentage of claims for social benefits in Western Europe is 11% while in Poland 0.7%; use of public libraries in Poland is 3.1% compared to the average of 10% in Western Europe, etc. [25]. The most common reasons for lack of use of e-government interaction is lack of sufficient competence (11%), concern for personal data security (11%) and lack of specific functionality (3%) [23]. The most common reason for not using e-government systems is lack of sufficient competence (11%), concern for personal data security 11% and lack of specific functionality 3%. Identifying the current status of functionality provided via e-government systems may become the basis for the proposed changes in this area.

3 Research Methodology

Statistic analyses provided in the introduction are not optimistic. They indicate rather low interest of potential users of electronic administration systems in the current form. The article did not perform characteristics of those services neither critical analysis of accompanied functionality. As previously stated, additional factors that determine the probability of general success of e-government are additionally organizational and legal

conditions. Those conditions are constantly changing, and through the prism of change it makes it questionable to use at all the e-government systems in its current form and to assess their suitability for the user. A typical example of such activities is for example the rumors around the so-called - electronic signature, or continuous changes in the method that income tax is calculated. The authors thus faced the difficult task of selecting, a priori, the range of e-government services that are the most popular for the analyzed population and from the one hand, are consistent with the findings and definitions of the e-public administration used in definition of the European Union recommendations.

Research method consists of following steps:

1. creation of list with all services that are possible to be provided via e-government systems,
2. conducting the "popularity test" of used public services on a limited group of twenty-five people of the surveyed population,
3. creation of survey questionnaire based on returned from "popularity test" answers, adjusted in its essential part and the language and scope of the question to the respondents' understanding of the basic functions of e-government,
4. using dedicated tools to create an electronic survey and uploading it on servers of the Faculty of Management, University of Warsaw,
5. sending notifications to potential respondents,
6. analysis and discussion of the results,
7. the conclusions of the survey and the consequences therefrom.

CAWI method stands for Computer-Assisted Web Interview that is related to direct interview technique conducted online. Questionnaire is located also on line and respondent is required to self-fulfill provided questions.

Quality type research was executed based on CAWI method on selected sample of university students at the end of February 2017. 254 respondents from academic environment took part in the survey. 197 participants provided full response to the survey, that is 76% of the whole survey population. Despite previous consultation questions were reported to be difficult for respondents. Survey contained following parts:

- introductory questions on the frequency and technology of access to the Internet and frequency of access to e-government,
- the main part consisting of: questions about another e-government service, resulting from the European Commission's assumptions, together with the question of the quality of the service in relation to the same service performed in the traditional way and the results and the degree of satisfaction resulting from it,
- open questions addressing the future of e-government,
- the test sample of the survey questionnaire.

The survey was distributed online via servers of Faculty of Management Of University of Warsaw. Participants were limited only to academic environment and were recruited from students of all types studies (full-time and part-time BA, BSc and MA studies) at Faculty of Management, University of Warsaw, Academy of Finance and Business Vistula.

The survey completed over 250 respondents, who evaluated the whole provided issue. 76% of participants submitted correctly completed full questionnaire. Among the respondents there were 78.17% of women and 21.83% of men, that is specific for economy and management studies.

An average age of the respondent was 21, 39 years (out of range 19–23 years). The age is typical of students of the first years of BA and BSc students, mainly with secondary education (94.42%).

Among respondents 4.06% already finished the Bachelors studies and 0.75% finished Masters level. Approximately 69% of the respondents were students and 30% were working students. Almost 28% of respondents declared the origin of the city with the number of over 500 thousand of inhabitants, 12.18% lived in of the cities of 100–500 thousand inhabitants, over 13% lived in the cities with 50–100 thousand inhabitants and almost 24% of the respondents declared to live in the cities up to 50 thousand inhabitants and 22.34% declared origin of the rural area.

Selection of the sample group was decided after analysis of Batorski research [26], who proved that the highest level of Internet activity is within the age group of 16–24 and 25–34 (almost 70%), Following the http://www.newmarketing.pl/ [27], where 34% of all beneficiaries of all on line services (including mobility) - were coming from the age group of 18–34. Similar outcomes are presented also in research of https://marketingautomagic.pl/ [28]. The accepted assumption about the age of customers is at the same time an advantage of the choice of the research group (the highest percentage of online activities and use of electronic devices), as well as a specific limitation that reduced the potential for generalizations. Undoubtedly, such an approach to the topic also increased the positive results of the analysis.

4 Analysis and Discussion of Results

A significant number of respondents (97.97%) responded to the question about the frequency of use the Internet, that they use it several times a day, and additionally not less than once a day answered 1.51%. Only 0.51% of respondents use it rarely - a few times a month. From the perspective of this research it is important to highlight that opinions of respondents are presented directly as were provided in the questionnaire.

The respondents use the most commonly (44.67%) combinations of two devices of a laptop and a smartphone to access Internet (… I use the Internet most often with a laptop and a mobile device - smartphone, in my opinion those are the two most convenient methods to use the Internet …).

The second rank was assigned to smartphone - 31.98% (… while using the Internet I use the most often smartphone due to the convenience and easiness and quick access to information in any place and time, e.g. in public transport …), on third place is laptop - used four times less 8.63% (… the best device in my opinion is a laptop because mobile version of the site is not always convenient, sometimes the cookie information covers a large part of the page, and not all of items display correctly. Additionally I appreciate the work with keyboard and mouse, the touchpad is just too small to hit the corresponding “key” …). Other devices or their combination most commonly used for Internet access are indicated in 14.72%. The importance of a

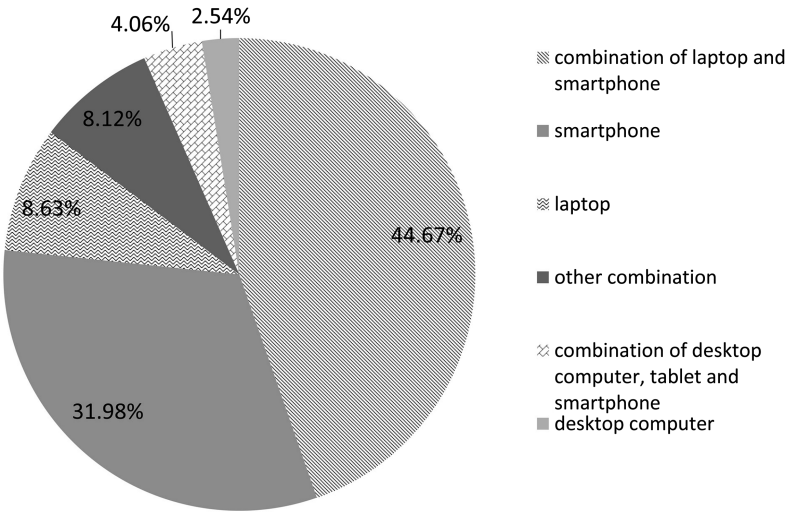

Fig. 1. The most frequent devices used to access internet in examined population

desktop computer has dropped (2.54%), if such device is found it's not the standalone option and it is used in combination with tablet or smartphone (4.06%) (… when I am at home I use a desktop computer, while being away from home I have a smartphone with me …). A detailed illustration of these trends shows Fig. 1. From the presented figures, it can be concluded that this group of respondents who almost constantly access internet and a high percentage of use of mobile devices, is the most promising with regards to use the most common functions attributed to e-government in everyday life.

In the meantime, the answer to the question whether respondents often use functions of the e-government is clearly disappointing. Most of people (39.59%) use such functions very rarely or not at all (18.78%), and additionally 15.23% respondents solve it traditionally, which is a testimony to the fact that almost 75% of respondents does not participate in the benefits of the functionality of this area. What is the reason behind that? Are the respondents not aware of some of the commonly available features offered by the e-government? Are such activities rarely performed within this age group? On the other hand, 14.21% of respondents use e-government systems frequently, at least once a week or several times a month (Fig. 2). Could the education what could be a function of e-government help them to understand what e-government stands for?

The first question concerned the most popular function - tax settlement via Internet. Almost 45% of respondents have confirmed that they have used often (or have used from the beginning of such availability) or several times (… I think this is a great convenience if a personal visit to the US involves problems (e.g., less convenient opening hours …). 8.63% of respondents never used such features (… we never used this type of feature; … for a person who is not capable to fill out the documents on the Internet, it is better to visit the US where the clerk can help. For people familiar with

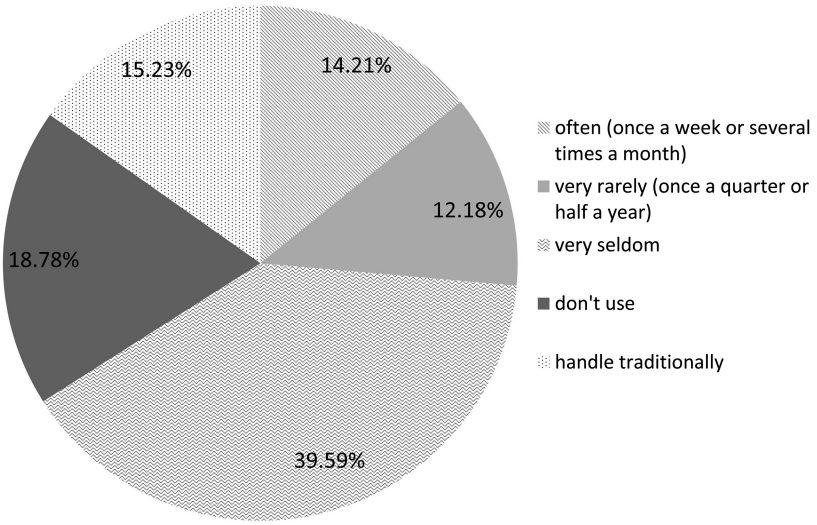

Fig. 2. Frequency of use of functions of e-government

the subject of taxation, the most efficient way is to settle the tax on the Internet …). The traditional way of doing taxes is declared by nearly 30% of respondents. The results are presented in Fig. 3.

Almost 60% of those who declare to use this system, admit that on-line tax settlement is better method than visiting a Tax Office, while the opposite opinion shows 8.12% of respondents. More than a quarter have no opinion on this subject.

Over two thirds of respondents who use tax settlements declared to be fully satisfied or satisfied with the service provided to them. 9.14% of users rated this function as average. Unsatisfied with this type of service was over 22% of those who took advantage of it. There are no opinions yet on how these views have changed after the taxpayers' were given an option - after posting online such will - to settle taxes by the Tax Office (directly by the Tax Office or through, for example selected banks).

At a similar level, was rated the opportunity to use the Internet for job placement. In this case 41% of respondents used this advantage several times, and 16.24% used at least once. Unfortunately, more than half participants did not use it at all, and 8.12% used traditional methods of job-seeking (Fig. 4).

This resulted that only slightly more than 19% of respondents using such form of contact rated online employment searching as better than traditional methods, and almost the same number claimed to be equal. By contrast, over 56% have no opinion on this subject. From the other hand, only 3% believe that online job searching produce worse results than traditional. Overall, 43% believe that the results of using such system are sufficient, almost 12% write about their average satisfaction, and 45% are discouraged by this form of contact.

Generally, similar distribution of responses was obtained in response to the question related to participation in the online social security service for individuals. More than a quarter of respondents declared that they used this form of contact at least once

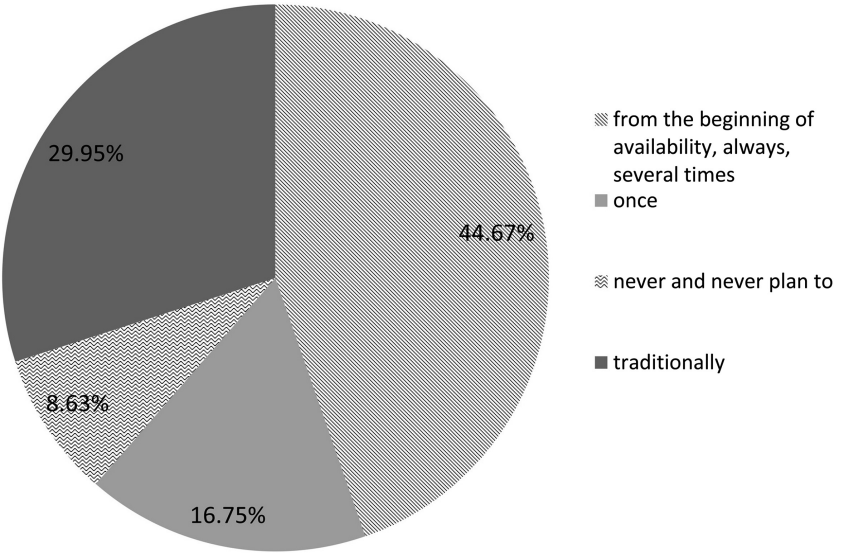

Fig. 3. Frequency of use of on-line tax settlement method

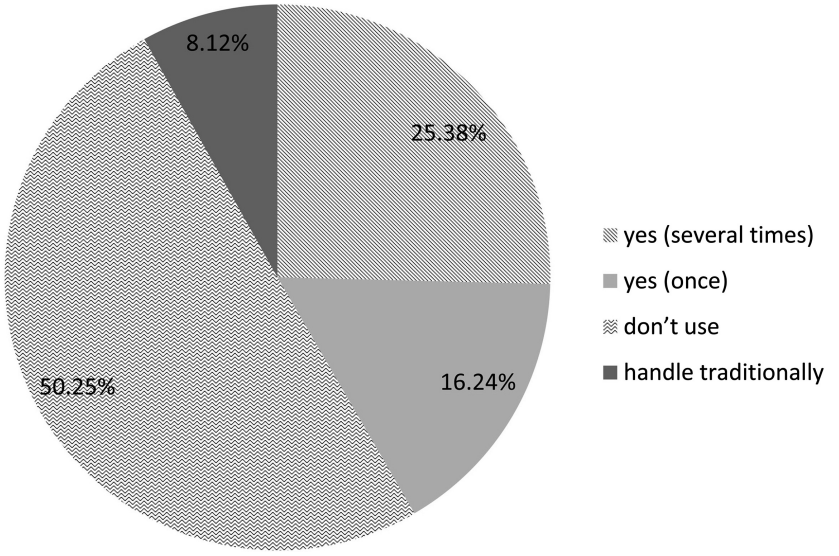

Fig. 4. Frequency of use of online employment searching

(half of this group). But almost the same - 22.34% - claims that handle such cases in a traditional way, and as many as 52.28% do not use at all such services available on the Internet (Fig. 5).

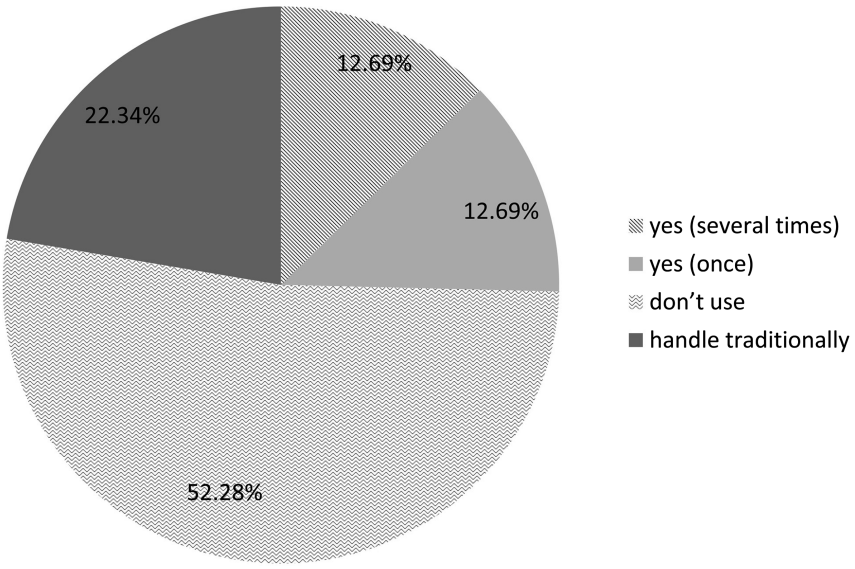

Fig. 5. Frequency of use of social security systems for individuals

People who declared the use of e-government systems that support social security services have claimed that these are forms of communication better or equivalent to traditional methods. On the other hand, more than 69% of respondents did not comment on the subject. For the first group of people, the process of social security services via the Internet was satisfactory or sufficient, 8% were moderately satisfied.

Much better proportions are noticed in the process of using online driving services. More than 58% of respondents used this service at least once, 28% did not use it. Almost 20% (Fig. 6) used traditional methods. More than 56% of respondents consider it to be equal to the traditional, and only 2.5% assess it worse. 38% have no opinion on this subject.

Since March 2015, Personal Identity Card hasn't included information about address of residency. The Ministry of Interior explained this change that the ID was a document confirming identity and Polish citizenship and address data didn't identify the person. Many identified this fact with the abolition of the permanent address registration, that however has not occurred and probably will not occur at all. The full restoration of the full registration obligation is planned, however, with the possibility to register only via the Internet. Respondents were so far not interested in the possibilities offered by the Internet in this regard, additionally were confused by the political shortcomings of successive government administrations on the subject. Perhaps - almost 62% of respondents stated that they do not use this communication tool, and almost 29% of them settle this traditionally. Only over 9% supported their activities with electronic tools. Would the trends change as a result of the change of regulations? For the time being, 83% of respondents have no opinion on this issue, and 73% consider it irrelevant.

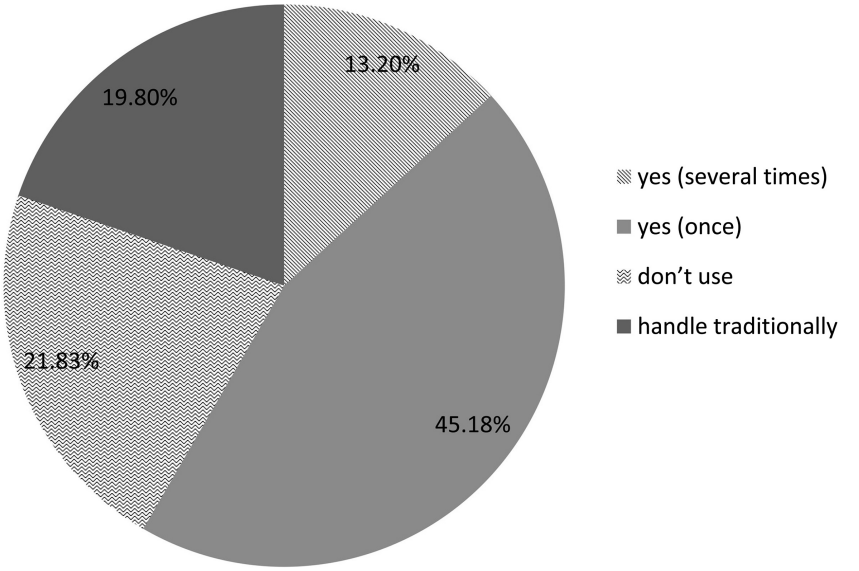

Fig. 6. Frequency of use of on-line service supporting driving licenses application process

Only slightly over 25% of respondents participated in the online process of passport service, more than 38.58% of respondents handle passports traditionally, 36.30% of respondents do not use this online service. Only 14.21% believe that online methods are better than traditional, more than 10% recognize those services as equivalent, but 66.50% do not have any opinion. This is relevant for 27% of respondents who agree that service "suits their needs" and in this case response "not enough" is answered by as many as 62%. Does it mean that e-government services are poorly organized? Or do the respondents expect something different from this service?

Even worse were the results of online personal identification services. At least once, 37.08% of respondents used this possibility out of those two thirds only once. On the other hand, 42.13% of the respondents used traditional methods. Those who used on-line services, claimed that this was a better way than traditional methods (26.90%), but more than half (54%) did not have an opinion on this. Similar percentage of respondents believed that the results of such a service were far away from satisfactory and only 38.07% answered that were sufficient or sufficiently adequate.

Almost 48% of respondents don't use the registration and deregistration of vehicles process, 34% of participants handle it traditionally. Only 18.28% of respondents supported these activities using ICT techniques. Only 18% believe that the results were better or equal to traditional ones. Nearly 75% have no opinion on this subject. Two thirds believe, however, that the results of these activities are almost nonexistent. Also lack of interest in services of participants is caused by the lack of coordination of regulations and documentation.

76% of respondents do not use online services to obtain construction and demolition permits and 20% handle it traditionally. Only 4% participants use the facilities in

this regard. 87% respondents have no opinion about the primacy of web methods, and almost 80% think that the results of the systems used in this field are almost none.

Similar level of results is related to online reporting of negative events to the police. 34% of respondents handle it traditionally, and over 21% of them do not use it at all. No wonder that 84% cannot express their opinion about the need for such a system, and 78% believe that, the impact of ICT on dealing with this type of problem as is currently present is none.

Approximately 19% of respondents used the Internet to obtain the required documents from the Office of Civil Status. Only one quarter deals with this traditionally, and 55% have never had a need to use those services at all. However, less than 10% think that online methods are better than traditional, and 7% recognize the equivalence between traditional and web supported. But still 74% have no opinion on this subject. Two thirds believe that currently the results of use of such systems are almost nonexistent.

The opposite is the situation with the use of on-line health service (medical visits, results preview, etc.). At least 66.7% of respondents used it at least once, and almost all of them considered it at least enough, and only 23.9% handle it traditionally. Therefore, 66% consider such services to be better than traditional, and only 29% of respondents have opposite opinion. Even when treating medical services in the general way as were used in the survey (contact with health services over the Internet), it is not possible to notice that they are at low level. There is no system of unified records of visits to public doctors via the Internet, electronic prescriptions, possibility for the videoconference with the doctor, telemonitoring, telediagnostics, etc. There is no common database not only at national or regional level, but also within the individual private companies providing medical services. The only manifestation of a success is despite such serious and long-term negligence respondents have responded to such a question so relieved.

42.64% of the respondents used the internet voting form via the internet portal (e.g. the civic budget). Almost the same number does not use this form of voting, and 15.7% would prefer to do it traditionally. 33% believe that such methods are better than traditional ones, and only over 50% have no opinion. Almost all internet users rate this form positively, with no opinion of 45%.

The online service of family allowances, nurse care and unemployed job seekers does not apply for more than 86% of respondents. There are almost as many respondents who have no opinion as those who believe that such form of communication is irrelevant.

There is also an advantage of using online public opinion polls and public mailing lists - 53% over the people who don't use such services - 47%. Similarly in the means of the percentage look opinions on the quality of these services and the satisfaction of usage.

On the other hand, on line reading is positioned very well, almost 81% of the respondents experienced this form of communication with culture via the Internet, while 78% consider it to be at least equivalent to traditional methods and 100% are satisfied with it.

Additionally, due to the selected research sample, students were asked about the possibility to apply online for admission to the university, and over 94% of the respondents used this form of communication and considered it better or at least equivalent to the traditional forms as well as they confirmed that participants were satisfied with this form.

5 Conclusion

To sum up the above considerations it should be noted that the results do not differ much from the results described in the survey carried out in Poland and Europe in 2013–2015. It also shows compliance with findings of other research studies in the area of e-government [29, 30] Our results equal 35% share of using e-government systems are similar in comparison to 31–33% in other studies [25, 31], and our 30% of the tax settlement is compared to 31.2% in the European Commission databases [23]. The level of the remaining results is also similar in some cases to the research conducted in Poland in 2013 [32] (Table 1).

The resulting differences may be caused by the methodological discrepancy or characteristics of the selected test group. While for example, in the group of young people it is more common to fill in and send taxation forms via the Internet, education related matters, borrowing publications from the library and job searching issues are at similar levels, comparing to older test group this group often deal with personal and

Table 1. Frequency of use of e-government systems within the test sample in 2017

Function of the e-government	Frequently (always, several times)	Moderate (once, from time to time)	Never (rarely)	Traditionally
Income tax settlement	44.67%	16.75%	8.63%	29.95%
Job offers (employment agency etc.)	25.38%	16.24%	50.25%	8.12%
Social security and insurance (health, social, retirement)	12.69%	12.69%	52.28%	22.34%
Personal cases (change of residency register, national ID cards, passports)	6.09%	18.27%	39.09%	36.55%
Automotive (e.g. vehicle registration and deregistration, driving license)	9.14%	29.19%	34.78%	26.91%
Construction and gardening (obtaining building and demolition permits, cuttings, plantings)	1.52%	2.54%	76.14%	19.80%
Register of events to the police	0.51%	4.06%	61.42%	34.01%
Health care (medical visits etc.)	46.19%	19.80%	10.15%	23.86%
Voting through the internet portal (e.g. citizen's participation budget)	27.92%	14.72%	41.62%	15.74%
Support for family and nursing care allowances and for unemployed	3.05%	5.58%	86.29%	5.08%
Public opinion survey and public mailing list service	36.55%	16.24%	46.70%	0.51%
Book rental from the library	**75.63%**	5.08%	9.14%	10.15%
Science, education (e.g. application for university admission)	**53.81%**	40.61%	1.52%	4.06%

Source: own studies, personal data, automotive and insurance data were averaged to be comparable to other studies.

construction issues once if at all and this is the reason why their perceptions related to the use of the Internet to handle the cases may be limited. The study doesn't include also the development of mobile communication. One of the effects of this development is the increase of the number of connections in the information society, and also e-government as well as better understanding of the concept of easier contact with public administration. The above differences cause that feedback component attached to the survey (opinions and comments of respondents) is even more valuable for the research.

The research proved the working thesis of the research that "Polish citizens don't use provided e-services at the level represented by other EU countries due to inconsistent ability to handle the whole matter from start to finish via the Internet".

In summary, conducted surveys, supplemented by comments from respondents (the respondents' statements, syntax and original spelling were given in the bracket, according to the respondents' feedback), give the following conclusions proving the thesis:

- the level of electronic services in e-government area in Poland is considerably lower than in most of European countries. In the official sphere this situation has not changed since the last three years,
- some revival can be noted in the social sphere (culture, public opinion, public health services, etc.), but the opinion expressed by respondents shows that respondents do not always distinguish public administration services from private services,
- the basic problem in using of e-government services is lack of ability to handle the whole matter from start to finish via the Internet. The small range of available services are often limited to the ability to print out a document (e.g. a request), which is even more disconcerting than attracting potential users,
- another element that deduces the benefits of the e-government is a need to appear personally in the office despite the settlement of some cases over the Internet (*... why would I fill something on the Internet, if still I have to wait in the que to deliver this document, and then again to pick it up ...*),
- there is serious level of a mistrust to the local and central administration as well as perception of a potential lack of internet security (*… citizens are accustomed to traditional methods and are more likely to trust them (they are afraid to loose their personal data) ... ... it is important that the database is reliable and 100% secure, Data security is the most common concern of using Internet services ...*),
- for the respondents the decision to use the e-government function is also balanced with the security of their personal data while handling the matter (*... still most people do not trust computers and prefer to have everything "on a paper ..." ... Polish still have little trust in the online transaction (...) we are aware of the possibility of hacker attacks, we fear that our personal data will be stolen, and somebody lied behind us and take a credit on our behalf ...*),
- there are gaps in information how to handle a given matter or the information is not clear for the user (*... one of the barriers that will be difficult to pass is to convince citizens that electronic services are too complicated. This may be due to illegible websites, many forms to complete or manual that user has to acknowledge before using the web service ..., ... sometimes user is overloaded with mass of information and links on the home page, which can frighten those users who are not familiar with similar websites ...*),

- there are technical problems (... *if the e-government website does not contain errors, it makes it easier rather than difficult to do things ...) ... technical problems downside of the e-government and can discourage people to use them e.g. entering the e-government systems, problems to login to own profile might be also connected to the wonts of traditional approach. Systems must be very easy to use, transparent and stable, and this might lead e-government systems to dominate Poland ...*),
- sometimes there is no response from the officials or lack of answer to the question (... *I did not receive any information about my application within the deadline, and after many phone calls made by me, I received an immediate negative response ...*),
- e-government services are incorrectly designed from the user's point of view and are described in an incomprehensible language (... *but e-government sometimes does not improve it at all, because many functions on websites or in applications are unclear and cause errors when filling in forms - instructions sometimes do not provide clear understanding of a issue, what certainly the person would clarify, ..., unfortunately a large part of those systems still need to be amended to become even more intuitive for ordinary people ...*),
- the low competence of officials informing how to fill out the web forms (... *not the best because of the staff who generally did not know how to handle the application themselves, and the site presented quite low quality ...*),
- the lack of proper marketing and information about possibility to handle public administration issues over the Internet and education about related benefits (... *most people do not even realize that there is such possibility.) They should be better informed about such possibilities and work on the manuals to those sites, so that people are not afraid to use such solutions ..., but citizen must be encouraged and persuaded to use it, because young and well-informed people may not have problems using it, but for the elderly and those who do not trust the Internet and technology ... the problem is the lack of common awareness about e-government ...*),
- in this situation it is not surprising the large share (27–40%) of traditional methods in dealing with official matters (... *there is a "coded" traditional method of dealing with officials ..., ... sometimes it is necessary to contact an official to clarify the rules or procedure to fill the document, this may indicate that the e-government services are not adjusted to customer needs and owned knowledge in this matter ...).*

The above conclusions show a number of potential postulates for e-government, the most important are:

- the ability to provide universal functionality to fill applications and documents without the need for an electronic signature as is in its current form, accompanied with feedback on whether the document has been properly filled out, or reverse information how to correct errors,
- the dissemination of information regarding e-government, as many people simply do not know about the possibility of settling official affairs and education on how to use online services and forms to convince unconverted to use this method to contact with the public offices or institution,

- proper design of e-government services, making them in a user-friendly language, not necessarily professional, maximizing the simplification of forms and minimizing of information taken from the citizen, especially when already present in the databases of the office or institution,
- web portals are very important elements of success of the e-government services, current research indicated some mismatch between what is expected from the web portals seen as single point of contact and current state of offered interactions and content
- and above all, the unification, simplification and integration of numerous legal and organizational regulations and restrictions that hinder contact with public administrations and the use of e-government.
- Test group that participated in the research indicates the latency to the status of use of e-government presented in the global research reports [6]. It might indicate the situation when information about the possibility of use of e-government is not populated accordingly among different users it might indicate the need for more aggressive promotion of digital services within young people and students.

The presented study has preliminary character and was concerned on the distinguished population. Its nature was based on the results obtained from test group and should therefore be extended to include a broader social cross-sectional survey and focus more on the use of e-government by citizens rather than on the current state of its activity and acceptance.

References

1. Holmes, D.: eGov: eBusiness Strategies for Government. Nicholas Brealey, London (2001)
2. eEurope - An Information Society for All, Communication of 8 December 1999 on a Commission Initiative for the Special European Council of Lisbon, 23–24 March 2000. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Al24221
3. COM (2010) 2020 Final, EUROPE 2020 A Strategy for Smart, Sustainable and Inclusive Growth, European Commission. https://www.eea.europa.eu/policy-documents/com-2010-2020-europe-2020
4. MSWiA: Strategia Rozwoju Społeczeństwa Informacyjnego w Polsce do roku 2013, Ministerstwo Spraw Wewnętrznych i Administracji (2008). https://www.mswia.gov.pl/pl/aktualnosci/6278,Strategia-rozwoju-spoleczenstwa-informacyjnego-w-Polsce-do-roku-2013-w-celu-podn.html
5. Ziemba, E., Papaj, T., Descours, D.: Assessing the quality of e-government portals - the polish experience. In: Proceedings of 2014 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1259–1267 (2014). https://doi.org/10.15439/2014f121
6. UN E-Government Survey 2016: E-Government in Support of Sustainable Development. New York (2016). http://workspace.unpan.org/sites/Internet/Documents/UNPAN97453.pdf
7. UN E-Government Knowledge Database: E-Government Development Index (EGDI). https://publicadministration.un.org/egovkb/en-us/About/Overview/-E-Government
8. Ziemba, E., Papaj, T.: A pragmatic approach to the e-government maturity in Poland - implementation and usage of SEKAP. In: Proceedings of European Conference on e-Government, University of Isubria, Como, Italy, pp. 355–363 (2013). http://www.academia.edu/6454355/A_Pragmatic_Approach_to_the_e_Government_Maturity_in_Poland_Implementation_and_Usage_of_SEKAP

9. Ziemba, E., Papaj, T., Żelazny, R.: A model of success factors for e-government adoption - the case of Poland. Issues Inf. Syst. **14**(2), 87–100 (2013)
10. UN E-Government Knowledge Database: E-Participation Index. https://publicadministration.un.org/egovkb/en-us/About/Overview/E-Participation
11. Ziemba, E., Obłąk, I.: The survey of information systems in public administration in Poland. Interdisc. J. Inf. Knowl. Manag. **9**, 31–56 (2014). https://doi.org/10.28945/1939
12. Ziemba, E., Papaj, T., Descours, D.: Factors affecting success of e-government portals: a perspective of software quality model. In: Proceedings of 14th European Conference on e-Government – ECEG 2014, pp. 252–262 (2014). https://doi.org/10.13140/rg.2.1.4787.3125
13. Ziemba, E., Papaj, T.: E-government application at the regional level in Poland - in case of SEKAP. In: Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1047–1054 (2012)
14. Ziemba, E., Papaj, T., Będkowski, J.: Egzemplifikacja e-government w Polsce – analiza porównawcza SEKAP i ePUAP. Roczniki Kolegium Analiz Ekonomicznych **29**, 427–445 (2013)
15. Ebrahim, Z., Irani, Z.: E-government adoption: architecture and barriers. Bus. Process Manag. J. **11**(5), 589–611 (2005). https://doi.org/10.1108/14637150510619902
16. European Commission: e-Government & Digital Public Services. https://ec.europa.eu/digital-single-market/en/public-services-egovernment
17. Grodzka, D.: E-administracja w Polsce. In: Grodzka, D. (ed.) Społeczeństwo informacyjne, Studia BAS, vol. 3, no. 19, pp. 57–81 (2009)
18. Marciniak, M.: Rozwój i wykorzystanie rozwiązań e-government w polskiej administracji publicznej. In: Kinal, J., Rykiel, Z. (eds.) Kultura i administracja w przestrzeni społecznej Internetu. Wydawnictwo Uniwersytetu Rzeszowskiego, Rzeszów (2013)
19. Bogucki, D.: e-government w Unii Europejskiej. Elektroniczna Administracja, vol. 1 (2005)
20. Główny Urząd Statystyczny: Pojęcia stosowane w statystyce publicznej - e-administracja. https://stat.gov.pl/metainformacje/slownik-pojec/pojecia-stosowane-w-statystyce-publicznej/1769,pojecie.html
21. Izdebski, H., Kulesza, M.: Administracja publiczna, zagadnienia ogólne. Liber, Warszawa (2004)
22. Encyklopedia Zarządzania: E-Government. https://mfiles.pl/pl/index.php/E-Government
23. European Commission: eGovernment in Poland, June 2016, Edition 18.1. https://joinup.ec.europa.eu/sites/default/files/ckeditor_files/files/eGovernment_Poland_June_2016_v4_01.pdf
24. Eurostat: Digital economy and society. https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/comprehensive-database
25. Śledziewska, K., Zięba, D.: E-Administracja w Polsce na tle krajów Unii Europejskiej. Digital Economy Lab UW, Warszawa (2016)
26. Batorski, D.: Technologies and media in households and lives of poles. social diagnosis 2015, the objective and subjective quality of life in Poland. Contemp. Econ. **9**(4), 367–389 (2015). https://doi.org/10.5709/ce.1897-9254.192
27. NowyMarketing: Handel mobilny w Polsce rośnie 3 razy szybciej niż e-commerce ogółem (2015). http://nowymarketing.pl/a/5207,handel-mobilny-w-polsce-rosnie-3-razy-szybciej-niz-e-ommerce-ogolem
28. MarketingAutomagic. Polski handel mobilny rozwija się szybciej niż e-commerce (2015). https://marketingautomagic.pl/2015/03/polski-handel-mobilny-rozwija-sie-szybciej-niz-e-commerce
29. Ziemba, E., Papaj, T., Żelazny, R., Jadamus-Hacura, M.: Factors influencing the success of e-government. J. Comput. Inf. Syst. **56**(2), 156–167 (2016). https://doi.org/10.1080/08874417.2016.1117378

30. Ziemba, E.: Factors affecting the adoption and usage of ICTs within Polish households. Interdisc. J. Inf. Knowl. Manag. **11**, 89–113 (2016)
31. Dziennik Internautów: Co trzeci polski internauta korzystał z e-administracji - badania MAiC (2014). http://di.com.pl/co-trzeci-polski-internauta-korzystal-z-e-administracji-badania-maic-49366
32. Ministerstwo Administracji i Cyfryzacji: E-administracja w oczach internautów – 2013 (2015). https://www.gov.pl/documents/31305/0/e-administracja_2013_wersja_zaadaptowana+%281%29.pdf

Analyzing Industry 4.0 Models with Focus on Lean Production Aspects

Christian Leyh[1(✉)], Stefan Martin[1], and Thomas Schäffer[2]

[1] Technische Universität Dresden, Chair of Information Systems, Esp. IS in Manufacturing and Commerce, Helmholtzstr. 10, 01069 Dresden, Germany
Christian.Leyh@tu-dresden.de

[2] Faculty of Business Administration, University of Applied Sciences Heilbronn, Max-Planck-Str. 39, 74081 Heilbronn, Germany
Thomas.Schaeffer@hs-heilbronn.de

Abstract. Nearly all enterprises have to face enormous challenges when dealing with digitalization topics such as Industry 4.0/Industrial Internet. To support companies to handle these challenges and therefore, to be able to "move" in an Industry 4.0 environment several frameworks or reference models already exist. Within this paper we provide results of a detailed analysis of selected Industry 4.0 models. However, we show that not all models are dealing with this topic in a holistic way but rather focusing on specific aspects or requirements of Industry 4.0. Additionally, we focus in our analysis on lean production aspects since the basic principles of lean management/lean production offer since the 1980s appropriate measures to optimize production and therefore, can be/should be addressed by Industry 4.0 models as well. Hence, it became obvious that those principles are not often addressed in Industry 4.0 models. Despite the fact that those aspects are often seen as a basis for Industry 4.0 implementation this is mostly not integrated in the respective models. Therefore, the contribution of our paper consists of the classification of 31 Industry 4.0 models/frameworks as well as the identification of needs for further research to enhance existing Industry 4.0 models to a more holistic approach.

Keywords: Lean management · Lean production · Industry 4.0
Industrial Internet · Digitalization · Digital transformation
Reference models · Maturity models

1 Motivation and Objectives

Since the beginning of the first Industrial Revolution in the middle of the 18th century and the development of steam engines as well as the increased use of hydropower, strong efforts have been made by industrial nations such as Germany or the United States of America to create a basis for their economy's growth [1]. One of the most important foundations was and still is the linking of value chains within a factory and even beyond the companies' borders [2]. To achieve this in an appropriate way still large planning effort is necessary which can be supported by lean management concepts. However, these concepts can only partially deal with the challenges of appropriate

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 114–130, 2018.
https://doi.org/10.1007/978-3-319-77721-4_7

linking value chains. When focusing lean management especially in light of the ongoing digitalization of business the need for new and adequate forms of communication in order to support these value chains or even whole value networks within and throughout all partners in the value networks still emerges up to date [2]. Hence, the implementation of lean management concepts is not primarily based on technical solutions, which are nevertheless elementary, but rather on structures in the form of architectures and standards that simplify and standardize business processes.

However, considering the evolution of technology, digitalization provides manifold opportunities to support or even renew business processes. These advanced technological opportunities, especially the merging of the physical with the digital world, result in new fundamental paradigm shifts affecting all industry sectors. Companies must handle global digital networks, improve automation of individual or even of all business processes, and reengineer existing business models to gain momentum in digital innovation. The prevailing and steady high dynamics in everyday business show that constant changes and adjustments (to which also digitalization belongs) will be no exception, but rather the rule in the future economy [3–6].

To appropriately deal with this, adjusted management and communication concepts have become highly important. In broad parts of society, the Internet of Things (IoT) has already established itself as an interlinked communication network to connect people, "things" and also whole value chains. Examples include package tracking, vital data logging via Smart watch or Smart Home control within the domestic environment. This development is accompanied by increasingly short and individual life cycles of products which consequently lead to new production requirements. Transferring the approaches of IoT to companies resulted in the concept of Industry 4.0 by connecting especially the production itself with the internet; leading to an increasing digitization of products and systems associated with their interconnectedness [7–9].

An analysis using the "Google Trends" tool (Fig. 1) shows that the interest in the field of Industry 4.0 has never been as significant as in the last couple of years. This also indicates the increasing international perception of the term Industry 4.0.

However, especially for those companies willing to use/integrate Industry 4.0 aspects in their production this is not a trivial task. Therefore, to support companies acting in the field of Industry 4.0 different reference models, frameworks or Industry 4.0 architectures have emerged. Using these "tools" should enable companies to structure their business processes appropriate regarding Industry 4.0 requirements.

This is where the present paper comes in. As extended paper of Leyh et al. [10] we set up a study to analyze selected architectural/reference models of Industry 4.0. Moreover, those models should be characterized according to the basic principles of lean management/lean production since these approaches offer since the 1980's appropriate measures to optimize production. Thus, in our opinion, those principles should be addressed by/included in Industry 4.0 models as well. Our study was driven by two research questions:

Q1: What organizational and technical reference models for Industry 4.0 exist?
Q2: What relationship can be established between the encountered models and Lean Production?

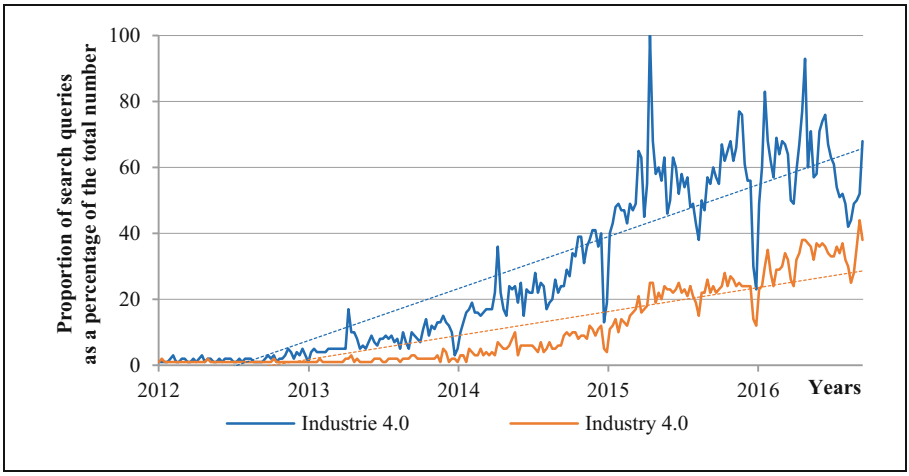

Fig. 1. Search queries for the terms "Industrie 4.0" and "Industry 4.0" on Google since 2012

In order to answer these questions, we set up a study based on a systematic literature analysis. The aim of the literature analysis is to describe, summarize, evaluate, clarify or integrate aspects focusing Industry 4.0 and/or lean production. Selected study results will be presented in this paper. Therefore, the paper is composed of four sections: Following this introduction, the second section provides a conceptual background of the key terms Industry 4.0 and lean production. The third section will be the main part of this paper where our literature review is described in its methodology and selected review results are presented and discussed. The paper concludes with a short summary, as well as an outlook for future research.

2 Conceptual Background

2.1 Industry 4.0

The term Industry 4.0 or the Industrial Internet is characterized as the fourth stage of the industrial revolution (Fig. 2) and consists of an increasing interconnectedness of products and systems. Focusing on the enhancement of the automation, flexibility, and individualization of products, production, and the connected business processes, Industry 4.0 aims at connecting the physical and virtual worlds [11, 12]. From a production perspective, Industry 4.0 is understood as the movement of intelligent workpieces that independently coordinate their paths through the factory. Machines are able to "realize" these tracks and communicate in real time with the corresponding warehouse. If necessary, orders are automatically triggered by means of targeted calculations, errors can be computed at an early stage of production, and the right interpretation enables a machine to change its production order. Information is primarily used to assess and control current processes [13]. Thus, an essential feature of Industry 4.0 can be seen in information aggregation in engineering and operations across different projects, plants, and plant operators [2].

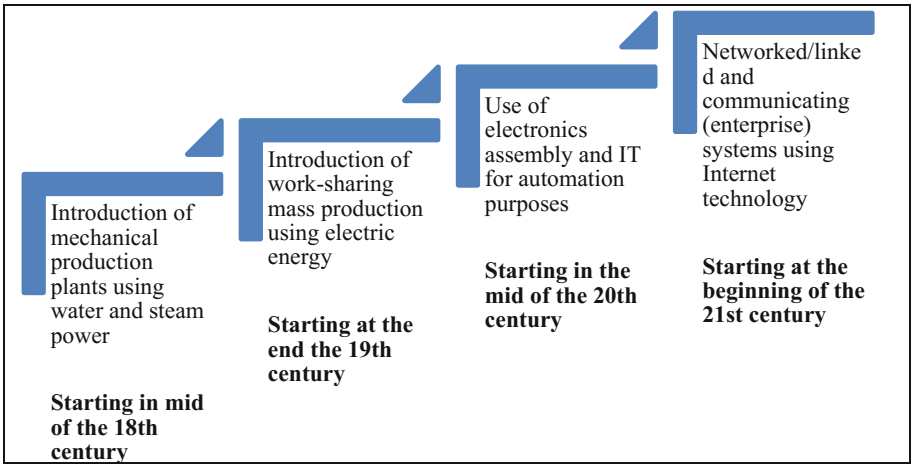

Fig. 2. Historical development of the industrial revolutions [14]

A universal definition for the term Industry 4.0 does not exist. Therefore, we deduced a working definition to serve as the foundation for our research which we also used in other related Industry 4.0 studies of ours (e.g. [3, 8]):

> *Industry 4.0 describes the transition from centralized production towards one that is very flexible and self-controlled. Within this production, the products and all affected systems, as well as all process steps of the engineering, are digitized and interconnected to share and pass information and to distribute this along the vertical and horizontal value chains and beyond in extensive value networks.*

However, the fact that companies have not yet implemented many parts of Industry 4.0 is shown in Table 1.

Table 1. Comparison of today's factory and an Industry 4.0 factory [15]

Characteristics		Today's manufacturing	Industry 4.0 manufacturing
Component (e.g. sensor)	Key attributes	Precision	Independent action based on own predictions
	Key technologies	Smart sensors and fault detection	Degradation monitoring and remaining useful life prediction
Machine (e.g. controller)	Key attributes	Producibility and performance (quality and throughput)	Independent action based on own predictions and comparison with inventory data
	Key technologies	Condition-based monitoring and diagnostics	Operating time recording with predictive health monitoring
Manufacturing System (e.g. MES)	Key attributes	Productivity and overall equipment effectiveness (OEE)	Independent configure, maintain and organize
	Key technologies	Lean operations: work and waste reduction	Low-maintenance, self-adapting production systems

2.2 Lean Production

Lean production/lean management already existed before the concept of Industry 4.0 was introduced/has been established. This form of the production management was first seized by Taiichi Ōno in 1978, who was responsible for the production of the Japanese automotive manufacturer Toyota [16]. After the end of World War II, Toyota noticed that the American car manufacturers were able to produce nine times more in the same time because they manufactured large batch sizes in order to compensate long set-up times. This was not possible for Toyota at this time because their production volumes were too small. Thus, Toyota adjudicated to develop their own philosophy that conforms the high quality of the label "Made in Germany" and competes with the faster and higher productivity levels of the US manufacturers at the same time [17]. After successfully implementing the respective measure to achieve a leaner production, a study including over 90 production centers that was conducted by the Massachusetts Institute of Technology in 1985 showed that Japanese car manufacturers performed better in all performance and quality parameters than the American and European manufacturers [18]. In total, this concept led to a paradigm shift and to the fact that lean production is now defined as a third production system design, since it is neither mass production nor manual work [19].

The basic principle of lean production is based on the avoidance of eight causes of waste. These are summarized by Ōno as *transport*, *storage*, *accessibility of processes*, *unnecessary movement*, *waiting times*, *overproduction*, *tight tolerances*, *defects* and, above all, *unused skills of the employees* [17]. In addition, Oeltjenbruns [20] classifies three central principles of lean production. This classification, shown in Fig. 3, mainly refers to the influence of these basic ideas on the company.

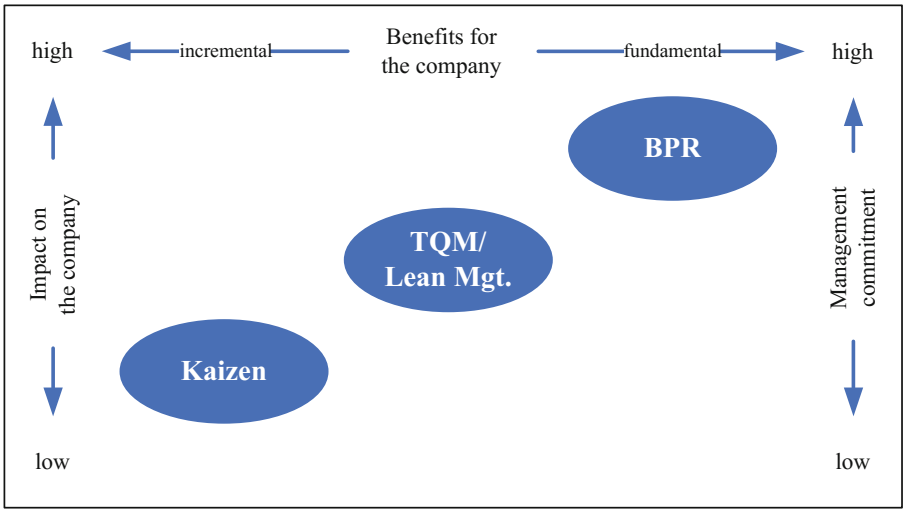

Fig. 3. Mutual classification of Kaizen, TQM/Lean Management and BPR [20]

The term *Kaizen* describes the continuous improvement process, which never ends and the sum of the changes contributes to the long-term success of the company. The *Total Quality Management (TQM)* is also leads to a long-term change in values of the entire workforce, which contributes to process improvement. The approach of *Business Process Reengineering (BPR)* requires the highest management deployment but brings the greatest benefit to the company at the same time. It refers to the radical change and a fundamental reorganization of processes with the aim of achieving greater synergies and the avoidance of waste. For the reorganization, the organizational structure of the company plays a central role. Due to the size of the changes, it can be assumed that this approach is only used for major strategic transformations such as the introduction of completely new product ranges, a fundamental reorientation of the company or even a structural change in the entire production system [20].

2.3 From Lean Production to Lean Automation to Industry 4.0

Kolberg and Zühlke [21] describe Industry 4.0 as a further development of Computer Integrated Manufacturing (CIM) and thus as a network approach, which is complemented by CIM through communication and information technology. This approach is supported by the integration of Cyber Physical Systems (CPS) [21]. These systems are a combination of two essential elements. These elements are the control of processes with the help of integrated software systems and the network of these software systems. However, the network is not limited to a single production line, factory or company, but to the global value chain or whole value networks [22].

With these systems lean automation can be implemented in order to support and expand the approaches and concepts of lean production. Lean automation automates a process with as little waste as possible. The objectives of short lead times with minimal costs and the highest quality remain unchanged. Consequently, it is possible to provide a company-wide representation of the actual situation in real time and to enable simulation-based optimization measures based on decentralized control systems. Each workpiece is thus clearly identifiable and the enterprise systems have information of the customer-specific mass product. Optimization measures and new services can be created from the resulting data collection. Ōno also describes the fact that lean production and automation are not mutually exclusive but the monitoring and use by the employee is elementary and does not work as a replacement [16]. The employee becomes the smart operator of production. The smart operator is, for example, notified by means of e-mail or SMS in the event of a fault reported by sensors. Thereby, the time from the occurrence of the error to the noticing of the error reduces. At the same time, the system makes suggestions for troubleshooting. In addition, modern technology such as Augmented Reality can provide a better representation of the process flows. The employee has access to previously available data such as cycle times in his immediate field of view [21].

However, the variety of possibilities in the linking, networking and interconnectedness requires a systematization of Industry 4.0 aspects, which shows the possibilities for the support and adapting of lean production.

3 Literature Analysis

As shown in Sect. 2 Industry 4.0 and lean management/lean production are both complex concepts which seem to have some connecting points and similarities. To investigate these aspects, we set up a study approach to contrast those two concepts. Since lean production is a mature concept and Industry 4.0 is an emerging topic we set up a systematic literature review to identify current papers dealing with the aspects of Industry 4.0. More specifically, a distinction is made between narrative articles and those that examine statistical and mathematical questions according to [23]. Narrative articles provide definitions for the most important terms and concepts. Mathematical-statistical papers provide comprehensive insights into existing research results and support a more deductive approach [24]. This resulted in a systematic literature review that is explained in the following subsection. After identifying and analyzing the Industry 4.0 articles we compared and contrasted identified Industry 4.0 frameworks and models with important aspects and approaches of lean management/lean production which is discussed later-on.

3.1 Methodology

The systematic literature analysis is intended to answer research question 1 and is based on four steps according to [25] and [26]:

Step 1 – Selection of databases and search terms: To get a broad overview of the topic we selected the databases ScienceDirect (www.sciencedirect.com) as well as Academic Search Complete (www.ebsco.com/products/research-databases/academic-search-complete) and Business Source Complete (www.ebsco.com/products/research-databases/business-source-complete). In addition, we used Google Scholar to identify articles may be not listed in scientific databases. The search fields for the database search were limited to abstract, title and keywords. The search terms themselves resulted from a short preliminary search according [26] and were afterwards discussed with researchers at the respective university institutes. This resulted in the following search string:

TITLE-ABSTR-KEY(“industrie 4.0” OR “industry 4.0” OR “fourth industrial revolution” OR “smart factory” OR “digital factory”) and TITLE-ABSTR-KEY (“framework” OR “scheme” OR “structure” OR “model”)

Step 2 – Implementation of practical screening criteria: With the help of step 2 and step 3, journal papers, conference papers and reports shall be classified. Considering the practical screening criteria, no temporal restriction was applied. In addition, the search was focused on articles in German as well as English, in which a general reference model for Industry 4.0 (or at least to a large extent Industry 4.0 concepts, frameworks or approaches) are presented. Therefore, articles were excluded which deal only indirectly with Industry 4.0 or only with a partial specific aspect of Industry 4.0 such as *Big Data* and do not classify this into a reference scheme. All identified papers were transferred in the literature management program *Zotero* (www.zotero.org). Afterwards, using the tools functionality a duplication check was performed.

Step 3 – Implementation of methodological screening criteria and Step 4 – Synthesis of the results: In these steps, a deeper analysis of the articles, that were not excluded during practical screening was conducted. First, the papers were classified according basic criteria:

1. Manufacturing environment: Does the model/the paper focus the manufacturing industry?
2. Industry 4.0 concept: Does the paper present/discuss/evaluate a reference model which covers the aspects of Industry 4.0 in total? Or are solely partial aspects of Industry 4.0 addressed?
3. Does the model address software and/or hardware aspects of Industry 4.0 requirements?
4. To what extent are lean production principles included and addressed in the reference model?
5. To what extent are business applications or enterprise systems explicitly addressed in the model?
6. Can the paper be classified as narrative article or merely as examining statistical and mathematical aspects (according to [23])?
7. Is an evaluation presented and discussed regarding the suitability and fit of the model in terms of Industry 4.0 requirements?

To assess the papers according these criteria/questions we used Harvey Balls with the differentiation shown in Table 2.

In addition to the seven merely general criteria we also assessed the models regarding four specific implementation requirements of Industry 4.0 postulated by different German national associations (e.g., VDMA – Mechanical Engineering Industry Association; Bitkom – Federal Association for Information Technology, Telecommunications and New Media; ZVEI – German Electrical and Electronic Manufacturers' Association) [9]:

8. The extent of horizontal integration across value networks
9. The extent of vertical integration in the company
10. The extent of product lifecycle management (PLCM) and consistency of engineering
11. The extent of the "human factor" – the employee as a conductor in the value networks

The results of the literature review and the analysis of the papers and models will be presented according to our classification scheme in the next section.

Table 2. Criteria classification

Symbol	Description
○	Criterion is not addressed
◔	Criterion is addressed indirectly
◑	Criterion is mentioned
◕	Criterion is partially addressed
●	Criterion is fully addressed

3.2 Selected Results

The search in the aforementioned databases with the presented search string scored a total of 166 papers. 96 of the articles found are published in conference proceedings which would be excluded by focusing the search solely on high ranked journals. Therefore, this is a first approval of our selected search methodology. In addition, nine out of the 166 were duplicates listed in more than one database. Therefore, those papers were excluded from deeper screening. After the practical screening of the remaining 157 papers, 31 papers could be identified that deal with an Industry 4.0 framework or model according to our criteria. Those 31 articles than were screened in depth to assess the criteria of step 3. Those articles are published not earlier than 2012 which again emphasizes the relevance and topicality of Industry 4.0.

Selected results will be discussed in the following paragraphs. Table 3 gives a short summarization of the results.

The total assessment of the remaining articles as well as the reference list for those papers are provided in the Appendix. As shown in Table 4 in the Appendix, nearly all articles (27 out of the 31) are addressing the manufacturing industry. The remaining articles deal with, for example, the service sector (Tables 4 and 5, No. 18) or the construction sector (Tables 4 and 5, No. 16). During the deeper analysis, it became clear that 15 of the 31 articles presented or discussed an Industry 4.0 approach with a holistic focus (see column 2 in Table 4) whereas 16 papers addressed specific partial aspects of Industry 4.0. In addition, it is striking that a discussion of software architectures is mainly provided in the articles. Hardware issues and aspects are not discussed without software aspects. Furthermore, it becomes apparent that most of the articles address the topic of enterprise architecture at least somehow. Therefore, emphasizing the enterprise and business process structure is important when a company wants to successfully "move" in an Industry 4.0 environment. These considerations of a generally valid architecture are accompanied by a few statistical or mathematical models, as those discussions often constitute a higher degree of detail as in papers with a more narrative focus (according to [23]).

Table 3. Short Categorization of the identified articles

Category of articles	Number of papers
Relevant in the sense of the research questions	31
↳No holistic Industry 4.0 reference model included	16
↳Holistic Industry 4.0 reference model included	15
↳Lean Production principles as main topic addressed	3
↳No Lean Production principles as main topic addressed	11
↳Lean Production principles are addressed in a medium to high extent but not as main topic	1

Considering lean production only three articles (Tables 4 and 5 in Appendix, Nos. 4, 6, and 11) are actively addressing those principles and incorporating them into an Industry 4.0 setting to a full extent (in regard to our criteria). Although lean production is often seen (as discussed in Sect. 2) as one of the foundations for Industry 4.0 most related concepts in the identified 31 articles are either taken on only marginal aspects of lean production or are not at all addressing these principles in combination with Industry 4.0. However, only one article (Tables 4 and 5, No. 11) out of the mentioned three articles fosters a production environment and actually incorporates lean production principles and approaches in an Industry 4.0 reference model. Also those authors present those principles as main point of their model.

In the second part of the in-depth analysis of the papers regarding the Industry 4.0 implementation requirements (criteria 8–11) it is noticeable that considering all articles, that provide a holistic Industry 4.0 model, vertical integration is the main subject in 13 out of those 15 articles. Whereas, the integration and consideration of employees as the main paper topic is addressed the least. Thus, it can be assumed that the automation and therefore often the replacement of human labor are often given more emphasis. However, in the discussion of the concept of Industry 4.0 human replacement is not the main goal. Moreover, the employees should be qualified to work with the Industry 4.0 technologies and instruments and be a "valuable support and cooperation partner" for the smart production floor. The aspects of product life cycle management and especially the consistency of engineering from factory planning to the final stage of the product life cycle are included and discussed regarding at least a certain section-by-section consistency in four articles with a really high emphasis (Tables 4 and 5, Nos. 4, 13, 15 and 18). Especially in the description of the Digital Eco-Factory by Matsuda et al. (Tables 4 and 5, No. 13 and 15) this is especially explained as the main subject in the production sector. However, this small number of hits shows that there is still a need for further discussion and higher emphasis of the specific aspect of Industry 4.0.

4 Discussion and Conclusion

Summing this up, we could identify several models and frameworks addressing the complex field of Industry 4.0 and therefore, provide a first answer to research question Q1. However, not all models are dealing with this topic in a holistic way but rather focusing on specific aspects or requirements of Industry 4.0. Hence, a common goal could be identified throughout all papers. The (explicit or indirect) stated goal is always to reduce the cost per unit produced. It was also crucial for all models and often discussed in the papers that for a further development and appropriate implementation of Industry 4.0, the communication in the three relationships man-man, machine-man and, above all, machine-machine are seen as especially important. From this, the machine-machine communication gains an even larger impact since in Industry 4.0 environments this communication and information sharing are the essential foundation for the autonomous machine decisions. As a first conclusion from these aspects the use of appropriate information and communication technology (ICT) is a crucial factor in Industry 4.0 environments as it is also stated by several authors (e.g., [4, 6, 8, 11]).

Regarding our research question Q2 it became obvious that lean management/lean production principles are not often addressed in Industry 4.0 models. Despite the fact that those aspects are often seen as a basis for Industry 4.0 implementation this is not integrated in the respective models nor is it discussed in connection with these models. Thereby, as shown in Table 4 merely the vertical integration is the main aspect in the identified models mostly also in combination with horizontal integration aspects. This also supports the fact that appropriate ICT is essential for Industry 4.0 environments.

From this, further research needs arise. First of all, considering the ICT it will be a challenge for enterprises which want to "move" in the field Industry 4.0 to identify and implement the appropriate ICT for the own company. Therefore, in addition to the identified models general maturity models are needed (with regard to Industry 4.0 requirements). Several models regarding these issues already exist. Those models deal for example with enterprise system landscapes for Industry 4.0 (e.g., [8]), with organizational aspects (e.g., [13]) or system-specific aspects in detail (e.g., [27]). However, a mapping of these maturity models would be necessary to combine their different points of view. For example, different maturity level assignments and dimensions between these models should be developed to enable companies to fully classify themselves in terms of Industry 4.0 requirements in all levels of their enterprise. With this work, companies would be able to determine their overall maturity in the field of Industry 4.0.

Furthermore, since the aspects of consistency of engineering and the employees themselves as well as the evolution of work (often named as Work 4.0) should be addressed in those models as well since they are essential issues of Industry 4.0. Therefore, also for further research it would be interesting to enhance and further develop existing lean production methodologies such as *Kanban* or *Kaizen* with regard to their Industry 4.0 suitability, since these approaches are already designed in their structure for self-organization and automation (that is a key issue of Industry 4.0). In addition to automation, the human factor must be better integrated into existing models, since employees will remain an essential part of the business processes. Therefore, the cooperation between employees and automated machines should be addressed in more detail in future research.

Closing this, despite the fact that there exist already several frameworks and reference models that are considering Industry 4.0 environments, there are still a couple of issues that can be seen as unsolved or at least not adequately addressed. Therefore, still further research is necessary to combine the existing approaches with additional key aspects of Industry 4.0 that are not addressed in a good extent yet.

Appendix

Additional possibility to download the assessment of all articles and the respective reference list: https://tu-dresden.de/bu/wirtschaft/isih/ressourcen/dateien/isih_team/pdfs_team/Supplementary-Material.pdf.

Table 4. Categorization of the identified articles according to the classification criteria

		General criteria							Industry 4.0 implementation aspects ([9])			
No.	**References**	Manufacturing environment	Holistic Industry 4.0 concept	Software (S) / hardware (H) consideration	Lean production principle	Business application	Mathematical / statistical aspects	Assessment of Industry 4.0 suitability	Horizontal integration across the value network	Vertical integration (e.g., in a factory)	PLCM / consistency of engineering	Employees as a conductor in the value networks
1	Ayadi et al. 2013	●	○	S&H	◔	●	○	○	◕	◕	◕	◑
2	Azevedo et al. 2010	●	●	S	◔	◑	○	◑	◑	●	◕	◑
3	Bagheri et al. 2015	●	●	S	◑	●	◔	◔	◔	●	◕	◔
4	Brettel et al. 2016	●	○	S	●	◔	◔	○	◕	◑	●	○
5	Debevec et al. 2014	●	○	S&H	○	◕	◑	○	◑	◑	○	○
6	Diez et al. 2015	●	○	S	●	◕	●	◕	◑	●	◔	◕
7	Flatscher and Riel 2016	●	○	S	◔	◕	○	○	●	◑	○	●
8	Francalanza et al. 2017	●	○	S	◔	●	◕	◔	◔	●	◕	◕
9	Ivanov et al. 2016	●	○	S	◕	◔	●	◔	◕	●	◔	◑
10	Jufer et al. 2012	●	●	S&H	◑	●	◔	◕	◕	●	◑	◑
11	Kolberg and Zühlke 2015	●	●	S&H	●	◕	◔	◔	◑	●	◔	◕
12	Long et al. 2016	●	●	S	◑	◑	◕	◕	◑	●	○	◑
13	Matsuda and Kimura 2015	●	●	S	○	◕	○	◑	◔	●	●	◑
14	Matsuda et al. 2012	○	○	S	○	●	○	◔	◔	●	○	◑
15	Matsuda et al. 2016	●	●	S	◔	◕	○	◑	◔	●	●	◕

16	Oesterreich and Teuteberg 2016	○	○	-	◑	◑	○	○	◕	◑	◑	◑
17	Qin et al. 2016	●	●	S&H	◔	●	○	◕	◑	●	◔	◑
18	Rennung et al. 2016	○	●	-	○	○	●	●	◑	◑	●	◕
19	Rix et al. 2016	●	○	S	◔	●	○	◑	◔	◕	○	◑
20	Schuh et al. 2015	●	○	S&H	◔	○	◔	◔	○	●	○	●
21	Schuh et al. 2014a	●	○	S	◔	◔	●	○	●	◑	◑	○
22	Schuh et al. 2014b	○	○	-	◔	●	◑	○	◕	◕	◑	●
23	Schuh et al. 2014c	●	●	S&H	◔	◕	◑	●	◔	●	○	◕
24	Schumacher et al. 2016	●	●	S	◑	◑	●	●	◔	●	◑	◑
25	Shafiq et al. 2015	●	○	S	◔	◑	◔	◔	◔	●	◔	◑
26	Sivard et al. 2016	●	●	-	○	◕	◔	◔	◑	●	○	◑
27	Stef et al. 2013	●	○	S	◔	◑	○	○	◕	◕	◕	◑
28	Terkaj and Urgo 2015	●	○	S	○	◕	○	◑	◔	●	○	◔
29	Tolio et al. 2013	●	●	S&H	◔	◑	○	○	○	●	◑	◔
30	Veza et al. 2015	●	●	S	◔	◑	◕	◕	●	◑	◕	◑
31	Wang et al. 2016	●	●	S	◔	◕	◕	◔	◑	●	◔	◕

Table 5. List of identified articles

No.	Reference
1	Ayadi, M., Costa Affonso, R., Cheutet, V., Masmoudi, F., Riviere, A., Haddar, M.: Conceptual Model for Management of Digital Factory Simulation Information. International Journal of Simulation Modelling **12**(2), 107-119 (2013). doi: 10.2507/IJSIMM12(2)4.233
2	Azevedo, A., Francisco, R.P., Bastos, J., Almeida, A.: Virtual Factory Framework: An Innovative Approach to Support the Planning and Optimization of the Next Generation Factories. IFAC Proceedings **43**(17), 320-325 (2010). doi: 10.3182/20100908-3-PT-3007.00069
3	Bagheri, B., Yang, S., Kao, H.-A., Lee, J.: Cyber-physical Systems Architecture for Self-Aware Machines in Industry 4.0. IFAC-PapersOnLine **48**(3), 1622–1627 (2015). doi: 10.1016/j.ifacol.2015.06.318
4	Brettel, M., Klein, M., Friederichsen, N.:The Relevance of Manufacturing Flexibility in the Context of Industrie 4.0. Procedia CIRP **41**, 105–110 (2016). doi: 10.1016/j.procir.2015.12.047
5	Debevec, M., Simic, M., Herakovic, N.: Virtual Factory as an Advanced Approach for Production Process Optimization. International Journal of Simulation Modelling **13**(1), 66-78 (2014). doi: 10.2507/IJSIMM13(1)6.260
6	Diez, J.V., Ordieres-Mere, J., Nuber, G.: The HOSHIN KANRI TREE. Cross-plant Lean Shopfloor Management. Procedia CIRP **32**, 150-155 (2015). doi: 10.1016/j.procir.2015.02.120
7	Flatscher, M., Riel, A.: Stakeholder integration for the successful product–process co-design for next-generation manufacturing technologies. CIRP Annals - Manufacturing Technology **65**(1), 181–184 (2016). doi: 10.1016/j.cirp.2016.04.055
8	Francalanza, E., Borg, J., Constantinescu, C.: A knowledge- based tool for designing cyber physical production systems. Computers in Industry **84**, 39–58 (2017). doi: 10.1016/j.compind.2016.08.001
9	Ivanov, D., Sokolov, B., Ivanova, M.: Schedule coordination in cyber-physical supply networks Industry 4.0. IFAC-PapersOnLine **49**(12), 839-844 (2016). doi: 10.1016/j.ifacol.2016.07.879
10	Jufer, N., Politze, D.P., Bathelt, J., Kunz, A.: Performance Factory -- a new approach of performance assessment for the Factory of the Future. Estonian Journal of Engineering **18**(1), 42–57 (2012). doi: 10.3176/eng.2012.1.04
11	Kolberg, D., Zühlke, D.: Lean Automation enabled by Industry 4.0 Technologies. IFAC-PapersOnLine **48**(3), 1870–1875 (2015)
12	Long, F., Zeiler, P., Bertsche, B.: Modelling the production systems in industry 4.0 and their availability with high-level Petri nets. IFAC-PapersOnLine **49**(12), 145-150 (2016). doi: 10.1016/j.ifacol.2016.07.565
13	Matsuda, M., Kimura, F.: Usage of a digital eco-factory for sustainable manufacturing. CIRP Journal of Manufacturing Science and Technology **9**, 97–106 (2015). doi: 10.1016/j.cirpj.2014.12.003
14	Matsuda, M., Kashiwase, K., Sudo, Y.: Agent Oriented Construction of a Digital Factory for Validation of a Production Scenario. Procedia CIRP **3**, 115-120 (2012). doi: 10.1016/j.procir.2012.07.021
15	Matsuda, M., Sudo, Y., Kimura, F.: A Multi-agent Based Construction of the Digital Eco-factory for a Printed-circuit Assembly. Procedia CIRP **41**, 218-223 (2016). doi: 10.1016/j.procir.2015.12.061
16	Oesterreich, T.D., Teuteberg, F.: Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. Computers in Industry **83**, 121–139 (2016). doi: 10.1016/j.compind.2016.09.006

17	Qin, J., Liu, Y., Grosvenor, R.: A Categorical Framework of Manufacturing for Industry 4.0 and Beyond. Procedia CIRP **52**, 173–178 (2016). doi: 10.1016/j.procir.2016.08.005
18	Rennung, F., Luminosu, C.T., Draghici, A.: Service Provision in the Framework of Industry 4.0. Procedia - Social and Behavioral Sciences **221**, 372-377 (2016). doi: 10.1016/j.sbspro.2016.05.127
19	Rix, M., Kujat, B., Meisen, T., Jeschke, S.: An Agile Information Processing Framework for High Pressure Die Casting Applications in Modern Manufacturing Systems. Procedia CIRP **41**, 1084-1089 (2016). doi: doi.org/10.1016/j.procir.2015.12.134
20	Schuh, G., Gartzen, T., Rodenhauser, T., Marks, A.: Promoting Work-based Learning through INDUSTRY 4.0. Procedia CIRP **32**, 82-87 (2015). doi: doi.org/10.1016/j.procir.2015.02.213
21	Schuh, G., Potente, T., Varandani, R., Schmitz, T.: Global Footprint Design based on genetic algorithms – An "Industry 4.0" perspective. CIRP Annals - Manufacturing Technology **63**(1), 433–436, (2014). doi: 10.1016/j.cirp.2014.03.121
22	Schuh, G., Potente, T., Varandani, R., Witthohn, W.: Consideration of Risk Management in Global Production Footprint Design. Procedia CIRP **17**, 345-350 (2014). doi: doi.org/10.1016/j.procir.2014.01.048
23	Schuh, G., Potente, T., Wesch-Potente, C., Weber, A.R., Prote, J.-P.: Collaboration Mechanisms to Increase Productivity in the Context of Industrie 4.0. Procedia CIRP **19**, pp. 51-56 (2014). doi: 10.1016/j.procir.2014.05.016
24	Schumacher, A., Erol, S., Sihn, W.: A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises. Procedia CIRP **52**, 161-166 (2016). doi: 10.1016/j.procir.2016.07.040
25	Shafiq, S.I., Sanin, C., Szczerbicki, E., Toro, C.: Virtual Engineering Object / Virtual Engineering Process: A specialized form of Cyber Physical System for Industrie 4.0. Procedia Computer Science **60**, 1146-1155 (2015). doi: 10.1016/j.procs.2015.08.166
26	Sivard, G.F., Eriksson, Y., Florin, U., Shariatzadeh, N., Lindberg, L.: Cross-disciplinary Design Based on the Digital Factory as a Boundary Object. Procedia CIRP **50**, 565-570 (2016). doi: 10.1016/j.procir.2016.05.040
27	Stef, I.D., Draghici, G., Draghici, A.: Product Design Process Model in the Digital Factory Context. Procedia Technology **9**, 451-462 (2013). doi: 10.1016/j.protcy.2013.12.050
28	Terkaj, W., Urgo, M.: A Virtual Factory Data Model as a Support Tool for the Simulation of Manufacturing Systems. Procedia CIRP **28**, 137-142 (2015). doi: 10.1016/j.procir.2015.04.023
29	Tolio, T., Sacco, M., Terkaj, W., Urgo, M.: Virtual Factory: An Integrated Framework for Manufacturing Systems Design and Analysis. Procedia CIRP **7**, 25-30 (2013). doi: 10.1016/j.procir.2013.05.005
30	Veza, I., Mladineo, M., Gjeldum, N.: Managing Innovative Production Network of Smart Factories. IFAC-PapersOnLine **48**(3), 555-560 (2015). doi: 10.1016/j.ifacol.2015.06.139
31	Wang, S., Wan, J., Zhang, D., Li, D., Zhang, C.:Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Computer Networks 101, 158–168 (2016). doi: 10.1016/j.comnet.2015.12.017

References

1. Qin, J., Liu, Y., Grosvenor, R.: A categorical framework of manufacturing for Industry and beyond. Procedia CIRP **52**, 173–178 (2016). https://doi.org/10.1016/j.procir.2016.08.005. Proceedings of the 6th International Conference on Changeable, Agile, Reconfigurable and Virtual Production – CARV 2016
2. Bauernhansl, T.: Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Bauernhansl, T., ten Hompel, M., Vogel-Heuser, B. (eds.) Industrie 4.0 in Produktion, Automatisierung und Logistik, pp. 5–35. Springer, Wiesbaden (2014). https://doi.org/10.1007/978-3-658-04682-8_1
3. Bley, K., Leyh, C., Schäffer, T.: Digitization of German enterprises in the production sector – do they know how "digitized" they are? In: Proceedings of the 22nd Americas Conference on Information Systems, AMCIS 2016, San Diego, USA (2016)
4. Mathrani, S., Mathrani, A., Viehland, D.: Using enterprise systems to realize digital business strategies. J. Enterp. Inf. Manag. **26**(4), 363–386 (2013). https://doi.org/10.1108/JEIM-01-2012-0003
5. Pagani, M.: Digital business strategy and value creation: framing the dynamic cycle of control points. MIS Q. **37**(2), 617–632 (2013)
6. Sambamurthy, V., Bharadwaj, A., Grover, V.: Shaping agility through digital options: reconceptualizing the role of information technology in contemporary firms. MIS Q. **27**(2), 237–263 (2003)
7. Bitran, G.R., Gurumurthi, S., Sam, S.L.: The need for third-party coordination in supply chain governance. MIT Sloan Manag. Rev. **48**(3), 30–37 (2007)
8. Leyh, C., Schäffer, T., Bley, K., Forstenhäusler, S.: Assessing the IT and software landscapes of Industry 4.0-Enterprises: the maturity model SIMMI 4.0. In: Ziemba, E. (ed.) AITM/ISM -2016. LNBIP, vol. 277, pp. 103–119. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53076-5_6
9. BITKOM, VDMA, ZVEI: Umsetzungsstrategie Industrie 4.0: Ergebnisbericht der Platform Industrie 4.0. Plattform Industrie 4.0, Berlin (2015)
10. Leyh, C., Martin, S., Schäffer, T.: Industry 4.0 and lean production – a matching relationship? An analysis of selected Industry 4.0 models. In: Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017, Prague, Czech Republic (2017)
11. Kagermann, H., Wahlster, W., Helbig, H.: Umsetzungsempfehlungen für das Zukunftsprojektindustrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0. Frankfurt am Main (2013)
12. Lemke, C., Brenner, W.: Einführung in die Wirtschaftsinformatik: Band 1: Verstehen des digitalen Zeitalters. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-44065-0
13. Kaufmann, T.: Geschäftsmodelle in Industrie 4.0 und dem Internet der Dinge: der Weg vom Anspruch in die Wirklichkeit. Springer, Wiesbaden (2015). https://doi.org/10.1007/978-3-658-10272-2
14. Roth, A.: Industrie 4.0 – Hype oder Revolution? Einführung und Umsetzung von Industrie 4.0, pp. 1–15. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-48505-7_1
15. Lee, J., Bagheri, B., Kao, H.-A.: Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. In: Proceedings of the International Conference on Industrial Informatics (INDIN 2014), Porto Alegre, Brazil, (2014)
16. Ōno, T.: Das Toyota-Produktionssystem [das Standardwerk zur Lean Production], 3rd edn. Campus, Frankfurt am Main (2013). translated from Hof, W

17. Ōno, T.: Toyota production system: beyond large-scale production. Taylor & Francis, Portland (1988). translated from Rosen, C.B
18. Womack, J.P., Jones, D.T., Roos, D.: The machine that changed the world: The story of lean production. Harper Paperbacks, New York (1991)
19. Tommelein, I.D.: Journey towards lean construction: pursuing a paradigm shift. J. Constr. Eng. Manag. **141**(6) (2014). https://doi.org/10.1061/(asce)co.1943-7862.0000926
20. Oeltjenbruns, H.: Organisation der Produktion nach dem Vorbild Toyotas: Analyse, Vorteile und detaillierte Voraussetzungen sowie die Vorgehensweise zur erfolgreichen Einführung am Beispiel eines globalen Automobilkonzerns. Shaker, Aachen, Germany (2000)
21. Kolberg, D., Zühlke, D.: Lean Automation enabled by Industry 4.0 Technologies. IFAC-PapersOnLine (15th IFAC Symposium on Information Control Problems in Manufacturing: INCOM 2015), vol. 48, no. 3, pp. 1870–1875 (2015)
22. Broy, M., Kagerman, H., Achatz, R.: Agenda Cyber Physical Systems: Outlines of a new Research Domain. Acatech, Munich (2010)
23. Wolf, F.M.: Meta-Analysis: Quantitative Methods for Research Synthesis (Quantitative Applications in the Social Sciences). Sage Publications, Newbury Park (1986)
24. Wilde, T., Hess, T.: Forschungsmethoden der Wirtschaftsinformatik: Eine empirische Untersuchung. Wirtschaftsinformatik **49**(4), 280–287 (2007). https://doi.org/10.1007/s11576-007-0064-z
25. Fink, A.: Conducting Research Literature Reviews: from the Internet to Paper, 4th edn. Sage Publications, Los Angeles (2014)
26. Fettke, P.: State of the art of the state of the art — a study of the research method "Review" in the information systems discipline. Wirtschaftsinformatik **48**(4), 257–266 (2006). https://doi.org/10.1007/s11576-006-0057-3
27. BMWi: Erschliessen der Potenziale der Anwendung von "Industrie 4.0" im Mittelstand. Studie im Auftrag des Bundesministerium für Wirtschaft und Energie (BMWi). agiplan GmbH, Mülheim an der Ruhr, Germany (2015)

Automation of Privacy Preserving BPMS in Collaborative Cloud-Based Business Processes

Sergei Makarov[1(✉)], Björn Schwarzbach[1], Michael Glöckner[1], Bogdan Franczyk[1], and André Ludwig[2]

[1] Information Systems Institute, Leipzig University, Leipzig, Germany
{makarov, schwarzbach, gloeckner, franczyk}@wifa.uni-leipzig.de
[2] Kühne Logistics University, Hamburg, Germany
andre.ludwig@the-klu.org

Abstract. Collaboration in business environments is an ongoing trend that is enabled by and based on cloud computing. It supports flexible and ad-hoc reconfiguration and integration of different services, which are provided and used via the internet, and implemented within business processes. This is an important competitive advantage for the participating stakeholders. However, trust, policy compliance, and data privacy are emerging issues that result from the distributed data handling in cloud-based business processes. In an earlier paper, we have discussed the requirements that a Business Process Management System (BPMS) should meet in order to enable privacy preserving business process as a service. This paper presents the necessary steps for implementing such a BPMS for an architecture to enable management of privacy-preserving collaborative business process.

Keywords: Privacy · Business process · BPMS · Cloud · Implementation

1 Introduction

Data privacy compliance, its handling as well as its security are the most concerning aspects for companies that collaborate in business environments, especially in the distributed ones such as cloud environments. One of the core components of such an environment are Business Process Management Systems (BPMS). An architecture for such an environment is proposed in [1, 2].

BPMS evolved from workflow and document management systems. They are able to cover full process management life-cycles in order to provide process design, modeling, execution, monitoring, and optimization [3]. However, focusing on the business process execution remains the most important task, as the other ones can be delegated to and taken over by the other architecture's components. The "divide-and-conquer" principal reflects the idea of service oriented architecture as it presets to decompose the problem into smaller manageable sub-problems. However, its management is one of the most crucial issues [4]. This comprises especially the

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 131–150, 2018.
https://doi.org/10.1007/978-3-319-77721-4_8

reorganization of BPMS' core functions such as the facilitated communication between system elements and parallel providing of appropriate process execution without violation of both privacy and security.

The central role of BPMS according to [5] refers to the integration layer of distributed components. Such a distribution can be found in cloud architectures. The cloud is an appropriate environment for companies with a high necessity for collaboration. The combination of BPMS and cloud solutions can amplify the effectiveness of business collaborations by services-on-demand [6].

The key contribution of the paper is the design and implementation of a fully automated BPMS component in such a privacy-preserving cloud architecture. Furthermore, process design and performance optimization are addressed.

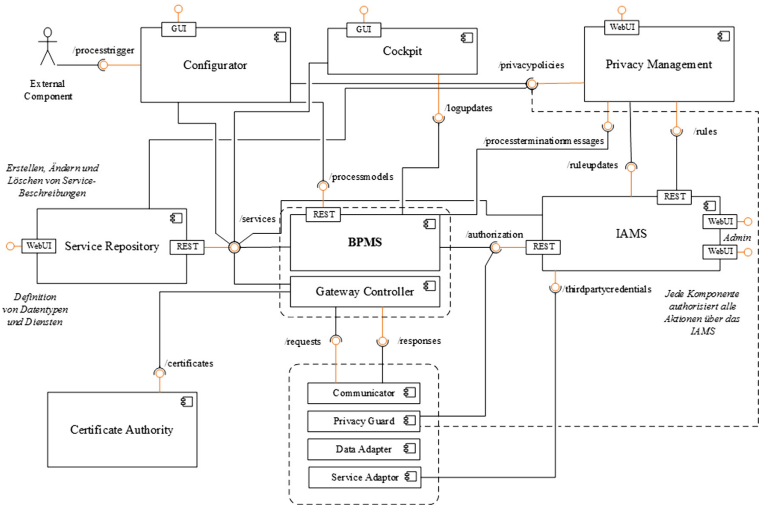

Fig. 1. Architecture for cloud-based collaborative BPaaS (Source: [1])

The first part of BPMS selection and its methodology was presented in [7]. This comprises project documentation analysis and literature research in order to determine requirements regarding software, BPMS, and further project-specific aspects. The result of that paper was the choice of jBPM as the BPMS tool that meets the requirements, and a first prototypical implementation. The architecture is shown in Fig. 1. The following sections present the implementation of automated collaborative processes enabling privacy compliance regarding the requirements defined in [1]. The architecture follows the SOA paradigm with comprehensive architecture and loose coupled components, which in turn can be denoted as web services in order to enable their automation.

The remainder of this paper is structured as follows: After the introduction, the theoretical background is briefly shown. Subsequently, the applied research methodology is outlined. Section 4 presents the implementation of the automated BPMS. Next, the results are discussed. The final section concludes the paper and gives an outlook on future research issues.

2 Theoretical Background

The essential objective of Business Process Management (BPM) is the planning and allocation of effective and efficient process solutions [8]. Identification and deployment of BPM focus on achieving business targets and customer benefits. Reference [3] emphasizes the scope on process identification in terms of business unit processes, content, and direction. In contrary, [8] refers to process allocation as institutional, permanent, goal-oriented planning, implementation, control and improvement during the entire process management lifecycle.

Concerning BPMS being deployed and utilized in Cloud environment, [6, 9] highlight the following advantages: scalability on IaaS layer (e.g. scalability of process instances); steadily available on-demand performance (e.g. CPU or storage capacity, especially in conjunction with data/predictive analytics for each process task); broad range of available SaaS providers for business activity monitoring tools to control activities of process instances; elasticity and availability. In addition, [10] consider BPMS independent from the implemented services. Thus, the changes in the process can be done without affecting target components, e.g. via business rules. Referring to [6], three following categories of systems with cloud-based models can be defined: interconnecting systems with IaaS and PaaS solutions for information exchange through EDI-systems, adaptive systems with SaaS models for monitoring of internal operations, and specialized systems such as CRM on SaaS layer.

The most proper steps for the companies' existing processes on the way to the cloud are outlined by [11] and comprise five following phases: Knowledge Externalization as representation of cloud services in man-computer readable form; BPaaS-Design via modular blocks; BPaaS Allocation as workflow orchestration of services, systems and components or even development of cloud-specific applications regarding user demands (can be referred to as Workflow-as-a-Service); BPaaS Execution as cloud orchestration between cloud services on the top abstract level (knowledge-based execution through the rules based on DMN format); BPaaS Evaluation as collection of log files from BPaaS deployment environment. However, it is not the only problem that should be addressed referencing migration of business processes. Further major concerns, addressed by [12], are data privacy and how private data can be prevented from unauthorized access, conflict of interests as well as physical resource isolation in such distributed environments with its scalability, service abstraction and local transparency. Moreover, data privacy and data integrity cannot be ensured only by securing business process operations but also security of the cloud environment itself should be taken into account. Hence, authentication and preventive measures are a major concern. Two problems are defined by [13–15] regarding data security. First, business processes are designed with architecture-specific restrictions in mind without concern of security requirements. Second, security aspects are integrated into an application in an ad-hoc manner - either during process implementation, administration or via outsourcing. The authors of [15] define the following five perspectives to be considered during process design in order to guarantee process integrity and consistency: Information Perspective for structuring and building relationships between information units; Function Perspective for representation of process activities

and data flow between them; Dynamical Perspective for state representation of both information unit and status transition; Organization Perspective to determine by whom and where the process was executed; Business Process Perspective to represent the whole process as activity, and information data flow.

Regarding ad-hoc problematic, [14] consider following three basic causes: none or minor security aspects' specification during design, faulty infilling of security characteristics by non-security-experts, and impairment of cryptographic methods due to facilitated validation. The last one claims to find a trade-off between proper process model extension by complex security components such as cryptographic protocols and utilization of formal methods in order to provide straightforward implementation on abstract level. This problem can be solved by building a single secure environment in which the process has to be executed.

Unfortunately, common measures are not always sufficient. This problem is addressed in [12] by a four-tier model with security components provided on each level, namely Software, Platform, Infrastructure Security as well as Auditing and Compliance. The latter one has been extended with such components as User and Authorization Management systems as well as Access and SLA Management ones. Furthermore, [16] consider Trusted Third Party (TTP) services within cloud environment as ideal solution to warrant integrity, confidentiality and credibility. TTPs are operative, associated certificate paths which provide acquaintance about Public Key Infrastructure (PKI) and support the following security aspects: strong Authentication, Authorization, as well as data confidentiality. The major benefit of PKI, referring to [16], is coupling of directories such as Lightweight Directory Access Protocol (LDAP). Such directories in conjunction with PKI can be utilized among others for messaging as well as for private key definition. The most common approach in cloud computing is to use PKI together with Single Sign On technique in order to facilitate user experience with multiple applications within a single architecture.

3 Research Methodology

The evaluation of the presented artifacts is based on the Framework for Evaluation in Design Science Research (FEDS), developed by Venable [17]. Artifacts that have been created by a methodology based on Design Science in Information Systems Research [18] can be evaluated by this framework in order to ensure rigor. The framework offers several strategies the could be pursued depending on the characteristics of the designed artifacts. Generally, the FEDS regards evaluation as an ongoing process during design science research in order to improve the artifacts iteratively. Several characteristics influence the evaluation's purpose (why?), point of progress of the design process (when?), strategy (how?) and the artifact itself (what?). The characteristics and resulting strategies are briefly introduced. By outlining the characteristics of the current research, a strategy is chosen and the resulting methodological steps are described.

The framework distinguishes between formative and summative evaluation [19]. The main purpose of formative evaluation is to improve the results of an artifact in the ongoing research process. On the contrary, summative evaluations have the purpose to create a shared meaning of the artifact concerning distinct contexts of application.

The question about the point of progress of design evaluation can be chosen ex-ante or ex-post [19] during the continuous design process. While ex-ante evaluations are more predictive in order to e.g. select a certain technology alternative, ex-post approaches are used to assess developed artifacts in terms of applicability or degree of achievements of objectives. With this, a greater likelihood of ex-post evaluation can be expected for summative evaluations but is not obligatory [17]. Goals of evaluations can be for different purpose: either achievement of environmental utility, or usefulness of solving a specific problem, or comparative advantage over existing solutions, or a complex composite of criteria (e.g. functionality, completeness, consistency), or other impacts (side effects), or reason artifact's functioning.

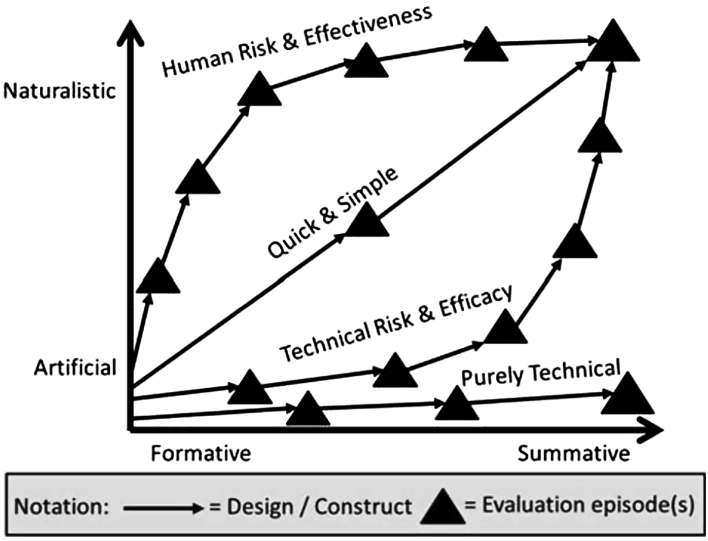

Fig. 2. Framework for Evaluation in Design Science (FEDS) with evaluation strategies (Source: [17])

The framework is displayed in Fig. 2, it comprises two dimensions. On the x-axis, the distinction between already described formative and summative evaluation purpose is located. The y-axis contains a distinction on how to evaluate with either artificial or naturalistic setup. While an artificial setup is used to prove general functionality of a concept, naturalistic evaluations prove an artifacts functionality in real environments, i.e. real people, real systems, and real settings [20]. Different *strategies* can be pursued that are displayed in Fig. 2 as well. Depending on the needs, available resources and circumstances, a strategy is chosen for and possibly changed during evaluation. The fastest strategy with the lowest costs is found in the 'quick\&simple' approach with a very limited number of iterations bears the risk of being not reasonable. A '*Purely Technical*' approach is suitable if naturalistic data and behavior is irrelevant and human users are not focus of the artifact. The other two strategies are used for either facing

'*Human Risk & Effectiveness*' or '*Technical Risk & Efficacy*'. A more detailed description of selecting a suitable strategy depending on specific circumstances can be found in Table 1.

Table 1. Circumstances for selecting a relevant DSR evaluation strategy [17]

DSR evaluation strategy	Circumstances selection criteria
Quick & simple	If small and simple construction of design, with low social and technical risk and uncertainty
Human risk & effectiveness	If the major design risk is social or user oriented *and/or* If it is relatively cheap to evaluate with real users in their real context *and/or* If a critical goal of the evaluation is to rigorously establish that the utility/benet will continue in real situations and over the long run
Technical risk & efficacy	If the major design risk is technically oriented *and/or* If it is prohibitively expensive to evaluate with real users and real systems in the real setting *and/or* If a critical goal of the evaluation is to rigorously establish that the utility/benet is due to the artifact, not something else
Purely technical artefact	If artifact is purely technical (no social aspects) or artifact use will be well in future and not today

The FEDS proposes 4 particular steps during the evaluation process [17]:

1. *Explicate the goals:* 4 goals of the evaluation were distinguished:
   (a) *Rigor* focuses on confirming that the artifact directly produced a certain effect (more likely to be shown with artificial evaluation) or that an instantiation of the artifact works thoroughly in a real situation (more likely to be shown with naturalistic evaluation). A summative evaluation provides the greatest rigor and reliability of the produced knowledge [17].
   (b) *Uncertainty and risk reduction* focuses on reducing either human and social risks or on reducing technical risks, which influences the choice of strategy (see Table 1).
   (c) *Ethics* focuses on reduction of potential risks to animals, people, or the public society. With this, especially potential stakeholders should not be put into risk.
   (d) *Efficiency* focuses on balancing the aforementioned goals in case of resource shortage. Hence, a more formative evaluation is proposed.
2. *Choose a strategy or strategies for the evaluation:* Depending on the aforementioned goals and the described circumstances of Table 1, one or more strategies have to be chosen. This can be done with a 4-step heuristic: (1) evaluate and prioritize design risks (either social/user oriented or technical or both). (2) Estimation of costs for real users, real systems and real settings. If human feedback is

available for a reasonable price, the '*Human Risk & Effectiveness*' strategy is suitable. If the price is too high or serious health concerns exist for users, the '*Technical Risk & Efficacy*' strategy is favorable. (3) If the artifact is purely technical and potential usage lies in remote future, the 'Purely Technical' strategy appears to be suitable or a naturalistic evaluation is just impossible. (4) If the construction that is to be evaluated is of rather small and simple extent, and none of the above-mentioned risks apply, the '*Quick & Simple*' strategy is the best choice.

3. *Determine the properties to evaluate:* the general set of features, goals and requirements of the artifacts that are to be evaluated are chosen. Again, a heuristic with 4 steps is proposed: (1) determine a list of potential evaluands (examples are given in [20–23]), (2) evaluands are to be aligned with the chosen goals, (3) depending on the chosen strategy of step 2, the evaluands should be of rather naturalistic or technical character and (4) determine the final list of evaluands.
4. *Design the individual evaluation episode:* the 3 heuristic sub-steps comprise: (1) derived from the environmental constraints, availability of resources determines their usage. (2) Priority shall be given to essential and more important aspects and resource are to be (re-)allocated. (3) Determination of number and structure of evaluation episodes and the according responsibility.

4 Research Findings

As already mentioned in introduction, we skip both the part of BPMS selection and requirements identification, which have been described in detail in [7]. However, we discuss thoroughly the implementation of those functional requirements, namely Service Selection, Remote Invocation, Process Activity Logging, external Security Provider as well as Cloud Readiness concerning BPMS from the architecture components' perspective. We also focus on additional components we had to embed in order to fulfil the objectives required by project, and which are not a part of the initial project's architecture shown in Fig. 1. The relationships and data exchange format between BPMS and project's components are described in Table 2.

Table 2. Architecture's components in association with BPMS

Component	Relationship/Direction	Format	API
Configurator	Configurator -> BPMS Controller -> BPMS	BPMN/XML	REST
Service-Repository (SR)	BPMS -> SR	JSON	REST
Privacy Management System (PMS)	BPMS -> PMS	XML	REST
Gateway	BPMS <-> Gateway	XML	REST
Cockpit	BPMS -> Log Collector <-Cockpit	JSON	REST
IAMS	BPMS (Security Provider) -> IAMS	XML/XACML	REST

Due to component's loose coupling, the services have a little concern about the process itself, which is entirely encapsulated within BPM engine. As a result, their communication happens either via BPMS' external API such a REST one or via additional middleware, provided complementarily, in order to act between two or more components within the entire system. We refer to such a broker as BPMS-Controller and denote an interface for log data deposition as Log-Collector (both for BPMS and IAMS log data). Additionally, we implement one more external element, namely XML-database (BaseX), to store temporally both BPMN model and I/O mapping data used for process task definition and execution. This concept decouples independent elements of the system, such as Configurator and BPM engine, as well as considerably facilitates process design. The last one implies much clearer process modelling, minimizing of possible errors as well as providing significant speed improvements. All these are possible without the necessity to have some programming background by the process designer.

The tool we have chosen during evaluation phase, which process is thoroughly described in [7], is jBPM BPM system, the open source product of Red Hat, Inc. with comprehensive API possibilities. Thus, its two following interfaces were specified for project objectives' realization: external one to provide the communication with other architecture's components (mainly via REST API) and internal one for implementation of the whole process logic and data processing (mainly via *WorkItem* API). The availability of customizable BPMN activities or tasks, which behavior can be defined by the user self by utilizing corresponding interface, was one of the criteria for choosing BPMS and is referred to as Service Selection, in the sense of reusing process activities. In jBPM such tasks are provided by *WorkItemHandler* (WIH) interface as well as abstract class called *AbstractWorkItemHandler*. The last one implements in turn WIH interface and extends its functionality through *StatefulKnowledgeSession* object being injected during subclass initialization as constructor's parameter. Listing 1 exemplifies the building of custom work item arbitrary defined as *ExecuteGenericTask*.

```
(1)  public class ExecuteGenericTask extends AbstractWorkItemHandler {
(2)    public ExecuteGenericTask(StatefulKnowledgeSession ksession) { super(ksession) }
(3)    public void executeWorktItem(WorkItem wi, WorkItemManager wim) { … }
(4)    public void abortWorkItem(WorkItem wi, WorkItemManager wim) {} }
```

Listing 1. Implementation of WorkItem API

According to API documentation [24], *StatefulKnowledgeSession* object provides the most common way to interact with process engine. The integration of custom-built work item into jBPM KIE Server, a standalone execution server, as jar-file has been thoroughly described in [7] as well.

4.1 Configurator

The role of Configurator is to provide a design of business processes as well as their subsequent deployment to BPMS engine. Referring to [25], the process elements such its META-data description, process itself, etc. should be embedded by the project, which will be compiled as *kjar*-file and deployed to KIE-server, in order to make it executable and provide a possibility to administrate it from KIE-workbench. This

concern can be applied not only to a new process deployment, but also regarding multi-tenancy problem in the cloud, providing service usage for multiple users at a time. For these purposes, the independent data source, persistence and other requirements should be concerned. The idea is to generate new project (i.e. *kjar*-file) for each new user or organization by the first process start and then subsequently use this project as deployment unit to generate new process instances. In both cases it is not sufficient to use only jBPM REST API but also build management tool's command interface should be involved. For our project, Apache Maven was utilized since this building tool is used by jBPM itself. The provided programmatical approach consists of three following steps: temporal storing of either new or modified process in XML-database (since it should be designed from outside of jBPM workbench), project preparation as well as subsequent project's deployment to KIE execution server. Though the first step is optional and can be eliminated in case no any new or modified process model has been provided, the already existing process may be used.

(1) Storing BPMN model in XML-databank

The storage process utilizes BaseX REST API to save the process model as BPMN/XML file into the database. Simultaneously, the name of deployment unit provided by Configurator is to be sent to BPMS-Controller in order to be utilized for subsequent steps. They are used for the preparation of project artifacts such as *pom.xml* with appropriate variables for name definition, triggering the project's undeployment command since it is important to guarantee that the process with the same name has not been already deployed to jBPM execution server, as well as project's deployment itself. Due to asynchronous calls nature, according to [24], - the request may be always accepted by KIE Server, also in the case the undeployment process itself has been failed – it is important to check twice, namely before and after deployment, whether it has been indeed succeeded. The list of all deployed units is possible to obtain over REST with suffix "rest/deployment" added to URL with jBPM container. Listing 2 provides how the project saves in BaseX utilizing its REST API.

```
(1)  public void saveNewBPMNFile(String fileContent) {
(2)    String deploymentUnit = this.groupId + ":" + this.artifactId + ":" + this.version;
(3)    ByteArrayInputStream content = new ByteArrayInputStream(fileContent.getBytes());
(4)    InputStreamEntity is = new InputStreamEntity(content);
(5)    HttpPut createDb = new HttpPut(host + "/basex/rest/" + this.dbName);
(6)    HttpPut put = new HttpPut(URL + this.dbName + "/" + deploymentUnit + ".xml");
(7)    Trigger.send(this);}
```

Listing 2. API for Configurator for storing BPMN model in BaseX-DB

The *send()*-method (line 7, Listing 2) provide concurrently with process model storage operation the values for project name definition to BPMS-Controller (Listing 3). With those variables BPMS-Controller is able to execute project artifacts' preparations in order to trigger process undeployment command as well as provide *kjar* deployment. Both operations are described in detail below.

```
(1) public void send(BaseXmlService basex) {
(2) HttpPost post = new HttpPost(hostname + port + "/project-deployer/" + basex.groupId + "/" +
    basex.artifactId + "/" + basex.version);}
```

Listing 3. Transferring name definitions to BPMS-Controller

(2) Project Preparation

The full name of the deployment unit consists of project and process names as well as version, which in turn, from Maven perspective, denoted as *groupId*, *artifactId* and *version* accordingly. At least one of them should be changed to make it possible to deploy the project to execution server without any name conflict. Regarding the multi-tenancy problem, it makes sense to change only group id according to user or organization name, if the process remains unchanged in the project. The designation of variables to be changed in *pom.xml* file is provided by Listing 4 and should be saved as external file outside the project to be deployed to execution server.

```
(1) <groupId>#{groupId}</groupId>
(2) <artifactId>#{artifactId}</artifactId>
(3) <version>#{version}</version>
```

Listing 4. Denoting variables to be changed in *pom.xml*

Listing 5 provides programmatical approach to substitution of those variables with the proper values. The values of variables to be set in new *pom.xml* file are to be derived either from BPMN model prepared by designer in Configurator or from the filename saved into XML-databank. Otherwise, full name of user started the process can be used. Since we interact with BPMS-Controller for further operations, these variables are to be retained from URL as parameters sent to it from Configurator or Privacy Management System accordingly.

```
(1) Path path = Paths.get(System.getProperty("user.home") +
"/generic-project/pomToGenerate.xml");
(2) String content = new String(Files.readAllBytes(path));
(3) String replaced = content.replace("#{artifactId}", this.artifactId).replace("#{groupId}",
    this.groupId).replace("#{version}", this.version);
(4) Files.write(Paths.get(System.getProperty("user.home") + "/generic-project/project-to-
    deploy/pom.xml"), replaced.getBytes());
```

Listing 5. Exchanging previously denoted variables in *pom.xml* file

The second important step during project preparation is the exchanging of BPMN-file with new or modified process stored in BaseX-database (this step can be eliminated in case the process model remains unchanged). Listing 6 provides code snippet for such a task.

```
(1) DeploymentUnit dUnit = new DeploymentUnit(this.groupId, this.artifactId, this.version);
    //variables coming from configurator
//retrieving process model from xml-db
(2) String content = baseXService.getXMLDocContent(dUnit.toString());
//saving those data as bpmn.file in project to be deployed
(3) Files.write(Paths.get(System.getProperty("user.home") + "/generic-project/project-to-
    deploy/prestige/prestige-generic-standalone/src/main/resources/" + "prestige-generic-
    standalone.bpmn2"), content.getBytes());
```

Listing 6. Exchanging content of BPMN-file with actual project from BaseX-DB

(3) Project's Deployment

Having both project artifacts been prepared and undeployment unit been verified, project deploying can be initialized as the next step (Listing 7). This time both Maven and jBPM APIs are to be implemented.

```
//execute 'mvn clean install deploy' command
(1) Runtime.getRuntime().exec("mvn -f " + System.getProperty("user.home") + "/generic-
    project/project-to-deploy/prestige/prestige-generic-standalone/" + " clean install deploy");
(2) //Runtime.getRuntime().exec("mvn -f " + System.getProperty("user.home") +
    "/Downloads/prestige-bpms-project-deployer/" + " clean install deploy");
//waiting for jar compilation
(3) Thread.sleep(30000);
//execute deployment of the process
(4) HttpPost request = new HttpPost(hostname + port + "/jbpm-console/rest/deployment/" + groupId
    + ":" + artifactId + ":" + version + "/deploy");
```

Listing 7. Project compilation as jar-file and its deployment to KIE server

During project's deployment phase, KIE Server searches, first of all, in local Maven repository installed on the virtual machine it is deployed to. Secondly, it looks in an external repository, namely Maven Central. However, it is also possible to provide an URL to own repository being used and utilized privately by organization for storing their private artifacts. We believe it is a good practice to have an external deposition place for artifacts additionally to application's one since the project migration may be concerned. For such a purpose, Apache Archiva was installed and utilized in the production environment. To grant an access to it, the credentials are to be provided in settings.xml file in the default Maven folder called ".m2" (Listing 8). In order to make it possible to execute Maven "deploy" phase after "install" one (line 2, Listing 7), the URL and ID for external repository, the artifact has to be deployed to, are to be provided in *pom.xml* file (Listing 9).

```
(1) <server>
(2)  <id>external </id>
(3)  <username>admin</username>
(4)  <password>admin</password>
(5) </server>
```

Listing 8. Providing credentials to external repository in .m2/settings.xml file

```
(1) <distributionManagement>
(2)   <repository>
(3)   <id>external </id>
(4)   <url><hostname>:8580/repository/external /</url>
(5) …
(6) </distributionManagement>
```

Listing 9. Providing URL and ID for external repository in *pom.xml* file of project folder

Approach described in this section was implemented for automatic deployment once new or modified process is available in the database. However, we also consider this approach, as it has been already mentioned initially, regarding new user registered on the system, namely during its first process start. This can provide segregation of resources concerning various users in order to grant independent process execution for each of them. This has a particular meaning, due to impossibility to generate a new process instance during work item task execution.

4.2 Cockpit

Cockpit should provide graphical representation of operational process data. Those should be gained from Log-Collector, they were pushed in by each work item and BPMS-Controller during process execution. The second concern regarding this component is to provide a link for process continuation from a particular node it was suspended due to privacy rules violation. The objective of both requirements is to provide logging of each process step during process execution and it has an exceptional meaning for process automatization. The process log data should provide information about who has triggered the process, at what time and what service (activity) is being executed at the time among others. The challenge of such a task is that process engine persistence module commits no any transaction until the process is either finished or stopped (atomicity property). However, there are some elements in BPM notation, namely signal events and human-acting tasks, which provide intermediate state persistence. They are also known as Safe Points in terms of workflow engine. The other possibility associated with modifying isolation layer to the low state is not to be considered due to possible incorrect state output accordingly to [26]. However, after having paused the process, we need to trigger its execution so it can be processed with the next subsequent activity and so on. Mostly, this requires a human interaction in order to send a signal to the process in case of signal events or the execution of a particular task in case of human tasks implementation. Moreover, the problem becomes more severe if the process execution needs to be continued only if some particular requirements have been fulfilled - for instance, if the process was suspended by some service due to a privacy violation. In the last case, the process can either be triggered by the service user after providing additional information or automatically once privacy properties have been changed. The decision regarding process suspension depends on the Boolean variable called “suspended”. It has to be sent both to the BPMS-Controller

along with other relevant process information for logging and to the XOR-Gateway within the process to decide which direction the process should follow after the trigger signal was sent from the BPMS-Controller (Listing 11). A snippet of the process model is provided in Fig. 3 and demonstrates the implementation. Listing 10 provides insights for Boolean programming of edges.

```
(1) return KieFunctions.isTrue(suspended_var)
(2) return KieFunctions.isFalse(suspended_var)
```

Listing 10. Programming of node edges between XOR-Gateway and WorkItems

```
(1) HttpPost post = new HttpPost(URL + "/" + this.deploymentUnit +"/process/instance/" +
    this.processInstanceId + "/signal" + "?signal=" + String.valueOf(this.nodeInstanceId()+1) +
    "&event=" + this.suspended ());
```

Listing 11. Triggering Signal Event via REST API with variable for XOR-Gateway

When the BPMS-Controller gets a request from within a task node, the value of the node instance id and not of the signal event is sent. Thus, as it can be seen in Listing 11 after "signal"-prefix, this value has to be increased by one in order to denote the subsequent process node id.

Since ensued edges can be programmed logically, we specify the process flow as follows: the last executed task should be accomplished repeatedly in case "suspended" set to "true". This can be triggered by the users after changing their privacy rules. This step requires reverification of privacy settings newly provided or modified by user. Otherwise, further process execution continues automatically (in case of negative value of "suspended"). In both cases, the process generates log data due to "suspender" (signal event), which in turn is to be read off by BPMS-Controller from a jBPM database via REST API and to be sent further to Log-Collector.

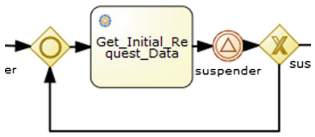

Fig. 3. Process flow accordingly to XOR-Gateway and variable named "suspended"

As already mentioned, this approach of manual trigger operation can be substituted with fully automatic process continuation direct after changing privacy rules, providing that the Privacy Management System takes over the handling of the signal event by itself. However, this was not considered by the project requirements so that the users can decide themselves, when some particular process instance should be continued.

4.3 Log-Collector

Log-Collector provides REST interface for storing both IAMS-data about the user, who started the process, and process META-data into relational database so that cockpit can gain those from it. The data, which collector has to be fed with, are to gain from a process within with the help of the following methods provided by *StatefulKnowledgeSession* object:

- *getProcessIntanceId(workItem)* method, its value is a number of the particular process instance. This value is particular useful for process handling via REST API. It has also been used to denote particular container for deposition of XML-files within BaseX with both process META-data and operational one during process execution.
- *getNodeInstance(workItem).getId()* or *getNodeInstance.*
- *(workItem).getNode().getId()* are both identical methods that return the node instance id as a sequence number of each node provided in the process and is strongly important in order to properly utilize signal events in jBPM. Unfortunately, a bug was detected in version 6.4 that generates new ids automatically after each task repetition (see section "Cockpit") and provides no correct response to them. Since we have utilized this particular version, we were not allowed to use this method but we had to provide task ids manually through I/O-parameters. This is redundant and impairs the idea of process design facilitation. We believe that this bug will be fixed in coming software versions.
- *getNodeInstance(workItem).getNodeName()* method provides information about the node name as a String type, its value has to be used for process monitoring dashboard – in the architecture known as Cockpit – in order to define on which point the error or privacy inconsistency had occurred. Since the node name reflects service description which the task executes, it facilitates further debugging, problem detection or provides user with more correct information about privacy compliance.
- *getNodeInstance(workItem).getNode().getMetaData().get("Lane").toString()* method helps to retrieve the name of the lane. However, since our process consists of a single swimlane, this method can be neglected.
- *KieContext.getVariable(String name)* is the method of *KieContext* interface, its implementation is injected in the KIE server during process execution and can be accessed via *kcontext* from within jBPM script task provided by the process modeler in the design palette. Since the process start has to be triggered from within the Privacy Management System and along with the user email, also other process relevant information have to be transferred, there is no other possibility to eliminate such a script task in order to facilitate the process model. Otherwise, Keycloak API or other libraries for token parsing can be utilized. It's also possible to gain this value from the KIE context, if the process would be started from the KIE workbench, given that the email is set up as a principal attribute of the Keycloak subdomain in the configuration file of the application server (see section "IAMS").

However, there is also another approach in jBPM to read full process log data off via REST API after persistence transaction has been proceeded (either at the end of process or at safe points). In contrast, this method does not allow direct access of the

data during the task execution as *AbstractWorkItemHandler* does, but provides comprehensive information about the whole process instead. Listing 12 demonstrates snippets for log data querying via jBPM REST API from within BPMS-Controller in order to get process, node and variable logs of every particular process.

```
(1) Process process = get(host + "jbpm-console/rest/history/instance/" + id);
(2) List<Process> processes = get(host + "jbpm-console/rest/history/instances");
(3) List<Node> nodes = get(host + "jbpm-console/rest/history/instance/" + processInstanceId +
    "/node");
(4) List<Variable> variables = get(host + "jbpm-console/rest/history/instance/" + processInstanceId +
    "/variable");
```

Listing 12. jBPM REST API for Log Queries

4.4 Service Repository

Process design can be very time-consuming and elaborative for its designer, especially by providing I/O-mapping between work item tasks. In order to facilitate this process, both service object description retrieved from service repository and Java XML-builder library were utilized within each work item. Thus, these require only one variable to be provided manually by each work item during the process design, namely service id. This way, XML-nodes and their values gained from gateway controller may be created automatically during process execution. For traversing through the document or building a new one there, both *w3c.dom* and *javax.xml* libraries were used. As BPMS and Gateway Controller communicate with each other on the basis of REST API and XML-payloads, it makes sense to allocate some sort of temporary storage during process execution, namely XML-database and grant the rights both for writing in and reading from for each work item. With such an approach, the I/O-exchange has been fully automated and needs no any human-involvement. Having implemented BaseX for such a role, a very concise API for managing database via REST is provided. Listing 13 demonstrates the simplicity with which both may be created. As container name, we have utilized a distinguished process instance id, while each XML-file overtakes the name of particular service (service id) provided by repository.

```
(1) HttpPut putDb = new HttpPut(hostname + "/basex/rest/" + this.processInstanceId);
(2) HttpPut putDocument = new HttpPut(hostname + "/basex/rest/" + this.processInstanceId + "/" +
    this.serviceId + ".xml");
(3) putDocument.setEntity(inputStreamEntity);
```

Listing 13. Creating both database and XML-document with BaseX REST API

Furthermore, XQuery can be used over REST against BaseX as a query language. It provides a possibility to search for particular values by providing either XML-node name or its attribute. The syntax for both methods has been provided in Listing 14.

```
(1) HttpGet getByAttribute = new HttpGet(hostname + processInstanceId + "?query=//" + element +
    "[@" + attribute + "=" + value + "]");
(2) HttpGet getByNode = new HttpGet(hostname + processInstanceId + "?query=//" + rootNode + "/"
    + nodeName);
```

Listing 14. Retrieving values from XML-Document via BaseX API

With all methods described in this section, integration of XML-databank as additional component of BPMS-architecture can thoroughly facilitate data exchange between work items and thus provide high automatization level of process execution.

4.5 Gateway

Gateway consists of multiple various modules (Fig. 1) in order to provide data privacy handling during service invocations. This requires a high level of security and can be extremely time-consuming due to the amount of services arising with time in the service repository. Hence, process optimizing is required. Although a cloud architecture can provide a solution to the problem with its performance on-demand, the process itself should implement an asynchronous approach. Since Gateway is implemented with modern architecture and parallel processing in mind, multiple requests can be transferred to it from BPMS concurrently, without idly awaiting each REST response one after another, but rather acquiring them once they are proceeded. Listing 15 shows how this task can be achieved with Java default concurrency library.

```
(1)  List<Callable<Map<String, String>>> callsToGWC = new HashMap<>();
try { PostService postService = new PostService();
// get guid according to your request
(2)   callsToGWC.add( () ->
          postService.post(EncodingService.getBytes(payload), serviced));
} catch (Exception ex) { e.printStackTrace();}
(3)  List<Future<Map<String, String>>> gwcGuids = null
(4)  try {gwcGuids = executorPool.invokeAll(callsToGWC, 10000, TimeUnit.SECONDS); } catch
     (InterruptedException e) { e.printStackTrace();}
(5)   executorPool.shutdown();
```

Listing 15. Utilizing Java Concurrency Library for parallel requests

According to [27], the formula for calculation of threads quantity results from the number of CPUs and I/O intensity coefficient. For-intensive tasks, this value is almost 0, while it approximates to 1 the more I/O intensive task is. Since our process is I/O intensive, we calculate the number of threads as Listing 16 provides.

```
(1)  ExecutorService executorPool = Executors.newFixedThreadPool(
       (int) (coresNumber/(1-0.9) ));
```

Listing 16. Threads' number calculation

The strong performance improvements were detected after parallel processing implementation. However, the process (not the task execution) cannot be continued before all the requests are proceeded by Gateway. That is, further task execution can succeed gradually as responses are coming back but it has to wait up to the last one in order to finish the current one. Since the execution time differs thoroughly depending on the service, the transaction module of application server is to be set up accordingly. In Wildfly 8.x, such a setting can be provided in a standalone-full.xml file (Listing 17), to avoid the interruption of the task execution after five minutes' expiration set up by default (the value is to be provided in seconds).

```
(1)  <subsystem xmlns="urn:jboss:domain:transactions:2.0">
(2)  …
(3)    <coordinator-environment default-timeout="43200"/>
(4)  </subsystem>
```

Listing 17. Changing default timeout of transaction module in standalone-full.xml

4.6 Privacy Management

With the security measures integrated into the cloud environment, the process has been deployed to, the transfer of operational parameters to jBPM from other components, such as Privacy Management System, can be run via URL utilizing jBPM REST API. This has been used to start or stop the process as well as to transfer some user relevant data, such as email, ip address, etc. All transferred variables are to be denoted with "map_"-prefix as provided in line 1, Listing 18 and can be retrieved with the help of KIE context via script task from within a process as it is provided by Listing 20. Immediately after process start, the response provided in payload comes with the relevant information about the started process. The actual process instance id can be retrieved in order to utilize it for further process handling such as process interruption (Listing 19).

```
(1)  HttpPost post = new HttpPost(https://" + this.hostname + ":" + this.port + "/jbpm-
     console/rest/runtime/Prestige:prestige-generic-standalone:1.0/process/prestige-prestige-generic-
     standalone/start?map_ip= + this.ip + "&map_email=" + this.email");
(2)  post.setHeader("Authorization", "Basic " + this.encoded);
(3)  HttpClient client = HttpCli-
     ents.custom().setSSLSocketFactory(SecurityConfig.getSSLConnectionSocketFactory()).build();
```

Listing 18. Utilizing jBPM RESTAPI for remote process start

```
(1)  HttpPost post = new HttpPost(https://" + this.hostname + ":" + this.port + "/jbpm-
     console/rest/runtime/Prestige:prestige-generic-standalone:1.0/process/instance" + this.instance +
     "/abort");
```

Listing 19. Providing URL for process instance termination

```
(1)  kcontext.setVariable("email_var", kcontext.getVariable("email"));
(2)  kcontext.setVariable("ip_var", kcontext.getVariable("ip"));
```

Listing 20. Retrieval of user data from within the process context and its assignment to process variables

4.7 IAMS

It is necessary to provide a consistent and protected environment for the service users during its utilization. However, not only the customers can benefit from such a uniform environment but also platform developers. The security unit of the entire architecture

can facilitate the system tests and provide homogeneous methods. However, the integration of single sign-on processes can be very complex and challenging. Thus, one of the criteria for BPMS selection was the availability of security tools (in [7] known as Security Provider), which can be easily integrated within the platform. Keycloak is such a tool for jBPM, that extends its security functionality. It can be installed on the same application server through the appropriate adapter. Its installation provides extensions in *standalone-full.xml* file by adding an authentication subsystem with a flag set up to "required". Having server been acquainted about Keycloak module, further security settings can be provided. The most significant ones are presented in Listing 21.

```
(1)  <subsystem xmlns="urn:jboss:domain:keycloak:1.1">
(2)    <secure-deployment name="kie-wb-6.4.0-Final.war">
(3)    <real>prestige</real>
(4)    <realm-public-key>…</realm-public-key>
(5)    <auth-server-url>https://...</auth-server-url>
(6)    <ssl-required>external</ssl-required>
(7)    <resource>jbpm</resource>
(8)    <enable -basic-auth>true</enable-basic-auth>
(9)    <credential name="secret">…</credential>
(10)   <principal-attribute>email </principal-attribute>
(11)   </secure-deployment>
(12)  </subsystem>
```

Listing 21. Providing security settings for jBPM platform

The settings in Listing 21 provide flexible options accordingly to the needs of developers. For instance, basic authorization can be activated along with security tokens. It can facilitate, for instance, integration tests. Also, principal attributes can be chosen from multiple options (name, email, etc.). This eases the gain of user identification without needs to retrieve this information from the token programmatically (in case, the process has been started from within a KIE workbench).

4.8 Discussion of Findings

Although BPM systems along with workflow ones have been introduced to provide fully automated process execution, some project requirements may impede this idea through human interaction involvement at those places, it could be eliminated. Furthermore, such complex systems, which provide full control over the process lifecycle, consist of multiple complex components, which in turn are hard to control and most of the time are not bug-free. These point to the necessity of finding a solution, that at least is not less efficient than a default one.

The whole process implementation regarding project requirements is shown in Fig. 4. Since both the process logic and the components' communication are encapsulated by work items, it was possible to provide a clear, not overloaded design. Consequently, process modelling can be done faster while avoiding common process design mistakes. The final solution provided by Fig. 4, optimized in terms of performance and design, had met all project's specific requirements and had successfully performed during evaluation phase.

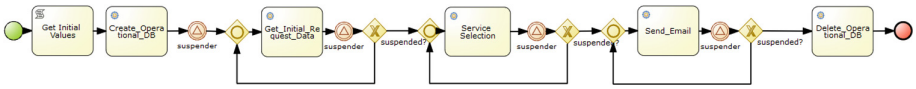

Fig. 4. Final process design and implementation

5 Conclusion

In this paper, the implementation of a fully automated BPMS for privacy-preserving management and execution of collaborative business processes has been discussed. This was the last piece that was needed to implement and evaluate the platform based on the architecture. The implementation and the architecture as a whole has been evaluated in multiple workshops with companies and researchers. The evaluation was positive, all the requirements that have been defined were met and the resulting platform was easy to use and was able to ensure privacy of the business data while executing the business processes.

Hence, the architecture and the fully automated BPMS enable the privacy preservation without the need for human interaction in every situation. Of course, the architecture is not perfect right now. The crucial part for the application of the architecture is the number of services available to the users. The task of implementing service adapters for additional services has to be carried out by the actual platform provider. Since the services adapters are simple, this should not raise any issues.

In the future, we will focus on improving the performance of the platform and optimization of the architecture.

References

1. Schwarzbach, B., Franczyk, B., Petrich, L., Schier, A., Hompel, M.T.: Cloud based privacy preserving collaborative business process management. In: 2016 IEEE International Conference on Computer and Information Technology (CIT), pp. 716–723. IEEE (2016)
2. Schier, A., Petrich, L., ten Hompel, M., Schwarzbach, B., Franczyk, B.: Cloud-Architektur für Privacy-Management in kollaborativen Logistikprozessen. Logist. J. **2192**, 1 (2016)
3. Schmelzer, H.J., Sesselmann, W.: Geschäftsprozessmanagement in der Praxis. Kunden zufriedenstellen, Produktivität steigern, Wert erhöhen: [das Standardwerk]. Hanser, München (2013)
4. Li, Z., Keung, J.: Software cost estimation framework for service-oriented architecture systems using divide-and-conquer approach. In: Fifth IEEE International Symposium on Service Oriented System Engineering (SOSE), Proceedings, 4–5 June 2010, Nanjing, China, pp. 47–54. IEEE, Piscataway (2010)
5. Becker, J., Kugeler, M., Rosemann, M. (eds.): Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. Springer Gabler, Berlin (2012). https://doi.org/10.1007/b138686
6. Hugos, M.H., Hulitzky, D.: Business in the Cloud. What Every Business Needs to Know About Cloud Computing. Wiley, Hoboken (2011)
7. Schwarzbach, B., Glöckner, M., Makarov, S., Franczyk, B., Ludwig, A.: Privacy preserving BPMS for collaborative BPaaS, pp. 925–934. IEEE (2017)

8. Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., Hess, T.: Grundzüge der Wirtschaftsinformatik. Springer, Berlin (2012). https://doi.org/10.1007/978-3-642-30515-3
9. Euting, S., Janiesch, C., Fischer, R., Tai, S., Weber, I.: Scalable business process execution in the cloud. In: 2014 International Conference on Communications and Networking (ComNet), 19–22 March 2014, Hammamet, Tunisia, pp. 175–184. IEEE, Piscataway (2014)
10. Megersa, B.T., Zhu, W.: Cloud-enabled business process management. Int. J. Comput. Theory Eng. **4**, 690 (2012)
11. Woitsch, R., Utz, W.: Business process as a service. Model based business and IT cloud alignment as a cloud offering. In: 2015 International Conference on Enterprise Systems (ES), pp. 121–130. IEEE (2015)
12. Chen, D., Zhao, H.: Data security and privacy protection issues in cloud computing. In: International Conference on Computer Science and Electronics Engineering (ICCSEE), Hangzhou, Zhejiang, China, 23–25 March 2012, pp. 647–651. IEEE, Piscataway (2012)
13. Rodriguez, A., Fernandez-Medina, E., Piattini, M.: A BPMN extension for the modeling of security requirements in business processes. IEICE Trans. Inf. Syst. **E90-D**, 745–752 (2007)
14. Backes, M., Pfitzmann, B., Waidner, M.: Security in business process engineering. In: van der Aalst, W.M.P., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 168–183. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-44895-0_12
15. Herrmann, G., Pernul, G.: Viewing business-process security from different perspectives. Int. J. Electron. Commer. **3**, 89–103 (2015)
16. Zissis, D., Lekkas, D.: Addressing cloud computing security issues. Future Gener. Comput. Syst. **28**, 583–592 (2012)
17. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: a framework for evaluation in design science research. Eur. J. Inf. Syst., 1–13 (2014)
18. Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**, 75–105 (2004)
19. Wiliam, D., Black, P.: Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? Br. Educ. Res. J. **22**, 537–548 (1996)
20. Sun, Y., Kantor, P.B.: Cross-evaluation: a new model for information system evaluation. J. Am. Soc. Inf. Sci. Technol. **57**, 614–628 (2006)
21. Mathiassen, L., Munk-Madsen, A., Nielsen, P.A., Stage, J., Jacksen, M.: Object-Oriented Analysis & Design. Marko, Aalborg (2000)
22. Smithson, S., Hirschheim, R.: Analysing information systems evaluation: another look at an old problem. Eur. J. Inf. Syst. **7**, 158–174 (1998)
23. Stufflebeam, D.L.: The CIPP model for evaluation. In: Kellaghan, T., Stufflebeam, D.L. (eds.) International Handbook of Educational Evaluation, vol. 9, pp. 31–62. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-010-0309-4_4
24. jboss.org. Community Documentation: jBPM Documentation. https://docs.jboss.org/jbpm/release/6.4.0.Final/jbpm-docs/html_single/
25. Fiorini, S.: Mastering jBPM6. Design, Build, and Deploy Business Process-Centric Applications Using the Cutting-Edge jBPM Technology Stack. Packt Publishing, Birmingham (2015)
26. Berenson, H., Bernstein, P., Gray, J., Melton, J., O'Neil, E., O'Neil, P.: A critique of ANSI SQL isolation levels. SIGMOD Rec. **24**, 1–10 (1995)
27. Subramaniam, V., Hogan, B.P. (eds.): Programming Concurrency on the JVM. Mastering Synchronization, STM, and Actors. Pragmatic Bookshelf, Raleigh (2011)

Synthetic Indexes for a Sustainable Information Society: Measuring ICT Adoption and Sustainability in Polish Enterprises

Ewa Ziemba(✉)

Faculty of Finance and Insurance, University of Economics in Katowice,
1 Maja 50, 40-287 Katowice, Poland
ewa.ziemba@ue.katowice.pl

Abstract. This paper is complementary with two published paper [1, 2] and advances information society research by examining and understanding the information and communication technologies (ICT) adoption in enterprises in the context of the sustainable information society (SIS). Its aim is to propose an approach to the measurement of two constructs shaping SIS, i.e. ICT adoption and sustainability in enterprises. ICT adoption is described by four components, i.e. ICT outlay, information culture, ICT management, and ICT quality, whereas sustainability is composed of ecological, economic, socio-cultural, and political sustainability. This study employs a quantitative approach and, additionally, Hellwig's taxonomic measure of development is adopted for multivariate comparative analyses to evaluate ICT adoption and sustainability in enterprises. A value of the Hellwig's synthetic indicator can be in the interval [0, 1] where a higher value of the indicator means that the object is closer to the pattern. The survey questionnaires were used and data collected from 394 enterprises were analyzed. We proposed and calculated five synthetic indexes for measuring ICT adoption, i.e. ICT adoption index and sub-indexes of ICT outlay, information culture, ICT management, and ICT quality as well as five synthetic indexes for evaluating sustainability, i.e. sustainability index and sub-indexes of ecological, economic, socio-culture, and political sustainability. The research revealed that in the largest number of examined enterprises (67%) the levels of ICT adoption and sustainability indexes are at the medium levels, i.e. in the interval (0.39, 0.79]. However, the high levels of ICT adoption and sustainability indexes exceeding value of 0.79 are indicated in, respectively 15.7% and 17.3% of enterprises.

Keywords: ICT adoption · Sustainability · Sustainable development
Information society · Sustainable information society · Enterprises
Poland

1 Introduction

The sustainable information society (SIS) is a new phase of information society development in which information and communication technologies (ICT) are becoming key enablers of sustainability [3–12]. Researchers and various organizations

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 151–169, 2018.
https://doi.org/10.1007/978-3-319-77721-4_9

have explored the areas where the information society, sustainable development, and ICT come together, and identified some correlations between those concepts [13–19]. Overall, the SIS is a multidimensional concept encompassing environmental, economic, cultural, social, and political aspects, all of which could be strongly influenced by adopting ICT by society stakeholders, mainly enterprises, households, and public administration [12].

In general terms, enormous ICT potential for the SIS development can be approached from two angles: ICT as an industry and ICT as a tool [12]. As an industry, ICT have become a major economic driver in the hardware, software, telecommunications, and consulting services sectors. ICT as a tool can be used to transform and improve business, everyday life of people, and public governance.

ICT used as a tool to revolutionize business is examined in this study. Some researchers have identified ICT as one of the most important tools in building sustainable business practices [19] and supporting the success of businesses [20]. It is contended that ICT enable businesses to improve productivity, support innovation, reduce costs, increase the effectiveness of processes services, enhance the efficiency of business decision-making, respond to customers at a faster rate, and acquire new customers [16, 19]. Moreover, ICT adoption in enterprises can yield benefits in environmental preservation by increasing energy efficiency and equipment utilization [6] as well as it can influence social development by making information available to all society stakeholders [9].

All these possibilities make ICT enablers of sustainability in several respects, i.e. environmental protection (ecological sustainability), economic growth (economic sustainability), socio-cultural development (socio-cultural sustainability), and governance (political sustainability) [12].

Following an extensive review of the literature, it can be stated that it did not uncover any deep studies to provide an objective assessment measures of how countries have performed in the field of ICT adoption and sustainability, and by highlighting areas that need further improvement. Moreover, there is a lack of such research on the SIS in less developed European countries, which are called transition economies [21]. The European transition economies are the former Eastern Bloc countries, which, since the early 1990s, have been undergoing transition from the command economy model to the free market model. We can identify the leaders and the followers of the transition process. In the first group there are: Poland, the Czech Republic, Hungary, Slovakia, Slovenia, Lithuania, Latvia, Estonia, Croatia, Romania and Bulgaria. The second group includes Belarus, Russia, Georgia, Moldova, Ukraine, Serbia and Montenegro.

In light of the above limitations, this paper focuses on exploring how the level of SIS can be evaluated. Its aim is to propose an approach to the evaluation of two constructs shaping SIS, i.e. ICT adoption and sustainability in enterprises.

The paper is structured as follows. Section 1 is an introduction to the subject. Section 2 states the theoretical background of ICT adoption, sustainability, and their measurement. Then research questions are posed. Section 3 describes the research methodology. Section 4 presents the research findings on the evaluation of ICT adoption and sustainability in enterprises. Section 5 provides the study's contributions, implications, and limitations as well as considerations for future investigative works.

2 Theoretical Background and Research Questions

2.1 ICT Adoption

ICT are defined as a diverse set of software and hardware, to perform together various functions of information creation, storing, processing, preservation and delivery, in a growing diversity of ways [22]. Based on works about the adoption and implementation of enterprise's information system, the adoption of ICT can be defined as ICT design, implementation, stabilization, and continuous improvement [23]. In this study, ICT adoption is understood as the whole spectrum of activities from the period when enterprises justify the need for adopting ICT until the period when enterprises experience the full potential of ICT and derive benefits from them [1].

Based on a stream of research, Ziemba [1] advanced a model, which categorizes the adoption of ICT into four components: ICT outlay (Out), information culture (Cul), ICT management (Man), and ICT quality (Qua). The component of ICT outlay includes the enterprises' financial capabilities and expenditure on ICT adoption, as well as funding acquired by enterprises from the European funds. The information culture component embraces digital and socio-cultural competences of enterprises' employees and managers, constant improvement of these competences, personal mastery, and incentive systems encouraging employees to adopt ICT. The ICT management component comprises of the alignment between business and ICT, top management support for ICT projects in the entire ICT adoption lifecycle, implementation of law regulations associated with ICT adoption, regulations on ICT and information security and protection. The ICT quality component consists of the quality and security of back- and front-office information systems, quality of hardware, maturity of e-services, and adoption of ERP and BI systems. Table 1 contains the description of each ICT adoption component.

2.2 Sustainability

The definition of sustainable development [24] was, in this paper, taken as a basis for the conceptualization and operationalization of sustainability. According to Schauer [9], sustainable development has four dimensions which are ecological, social, economic and cultural sustainability. In a further study, Ziemba [1] proposed an expanded sustainability which includes four sustainability components, i.e. ecological, economic, socio-cultural, and political. Regarding businesses, the sustainability components are [1]:

- Ecological sustainability (Ecl) is the ability of enterprises to maintain rates of renewable resource harvest, pollution creation, and non-renewable resource depletion by means of conservation and proper use of air, water, and land resources [25, 26];
- Economic sustainability (Eco) of enterprises means that enterprises can gain competitive edge, increase their market share, and boost shareholder value by adopting sustainable practices and models. Among the core drivers of a business case for sustainability are: cost and cost reduction, sales and profit margin, reputation and brand value, innovative capabilities [16, 17, 22];

- Socio-cultural sustainability (Soc) is based on the socio-cultural aspects that need to be sustained, e.g. trust, common meaning, diversity, capacity for learning and capacity for self-organization [27]. It is seen as dependent on social networks, making community contributions, creating a sense of place and offering community stability and security [28, 29]; and
- Political sustainability (Pol) must rest on the basic values of democracy and effective appropriation of all rights. It is related to the engagement of enterprises in creating democratic society [29].

Table 1 presents the description of each sustainability component.

Table 1. Components and primary variables of ICT adoption and sustainability constructs (Source: [1, 2])

Primary variables of the ICT adoption construct				Primary variables of the sustainability construct	
Out1	Financial capabilities	Man16	ICT project team	Ecl1	Sustainability in ICT
Out2	Expenditure on ICT	Man17	Top management support	Ecl2	Sustainability by ICT
Out3	Funding acquired from the European funds	Man18	Management concepts adoption	Eco3	Cost reduction
Cul4	Managers' ICT competences	Man19	Information security regulations	Eco4	Sales growth
Cul5	Employees' ICT competences	Man20	ICT regulations	Eco5	Product development
Cul6	Managers' permanent education	Man21	ICT public project	Eco6	Effective and efficient management
Cul7	Employees' permanent education	Man22	Competitive ICT market	Eco7	Effective and efficient customer service
Cul8	Employees' personal mastery	Qua23	ICT infrastructure quality	Eco8	Effective and efficient work
Cul9	Managers' socio-cultural competences	Qua24	Back-office system quality	Eco9	Acquiring new customers and markets
Cul10	Employees' socio-cultural competences	Qua25	Front-office system quality	Eco10	Increasing customer satisfaction/loyalty
Cul11	Employees' creativity	Qua26	Back-office system security	Soc11	Competence extension
Cul12	Incentive systems	Qua27	Front-office system security	Soc12	Working environment improvement
Man13	Alignment between business strategy and ICT	Qua28	E-service maturity levels	Soc13	Increasing security
Man14	Supporting business models by ICT	Qua29	ERP adoption	Soc14	Reducing social exclusion
Man15	ICT management procedure	Qua30	BI adoption	Pol15	E-democracy
—	—	—	—	Pol16	E-public services

2.3 Measurement of ICT Adoption and Sustainability

Based on SIS theory and the previous research [1, 2] two constructs should be embraced in the measurement of SIS, i.e. ICT adoption and sustainability. In addition, various benchmarks of information society [30, 31] employ two approaches to the quantitative description of such a society. The first one comprises the list of indicators characterizing information society, whereas the other is connected with the so-called synthetic indexes (synthetic indicators) which are based on the chosen set of indicators.

A set of indicators and synthetic indexes for measuring information society can be found in works of many organizations, e.g. Eurostat, ITU (International Telecommunication Union), OECD, World Bank, WEF (World Economic Forum), and IDC (International Data Corporation) [32]. The best known measurement synthetic indexes are the ICT Development Index (IDI) worked out by ITU and Networked Readiness Index (NRI) of the authorship of the WEF. The IDI is a composite index combining eleven indicators on ICT access, use and skills, and it is based on the three-stage model for information society development, i.e. readiness, intensity, and impact [30]. The NRI is composed of four sub-indexes described by fifty-four indicators: an environment sub-index, a readiness sub-index, a usage sub-index, an impact sub-index [33]. Unfortunately, existing indexes of information society do not embrace the sustainable imperative.

Various sets of indicators for measuring sustainability exist already and it seems that constantly new ones are being developed. Van de Kerk and Manuel [34] created a comprehensive Sustainable Society Index embracing 22 indicators grouped into five categories: personal development, clean environment, well-balanced society, sustainable use of resources, and sustainable world. Panda et al. [35] developed a composite Urban Social Sustainability Index for assessing the social sustainable development by Urban India. It includes 27 indicators under seven themes, i.e. demography, education, health, equity, housing, poverty, and safety themes. Orsato et al. [36] examined the usage of Corporate Sustainability Index to assess the sustainability performance of companies listed in stock exchange markets, whereas Laxe et al. [37] formulated and designed Synthetic Index for measuring four dimensions of Port Sustainability, i.e. economic, institutional, environmental, and social.

After analyzing indexes for measuring ICT adoption and sustainability, the overall conclusion is that none of the existing indexes seems to fit the SIS needs completely. In other words, not a one gives a comprehensive and good insight into all relevant issues of SIS in a transparent, simple and easily understandable way, showing at a glance to what extent an information society is sustainable or not and what extant ICT adoption are used for achieving sustainability. So a new index has to be developed, based on a set of indicators in accordance with the nature of SIS, the essence of which is environmental, economic, cultural, social, and political sustainability strongly influenced by adopting ICT by society stakeholders, mainly enterprises, households, and public administration.

2.4 Research Questions

In the previous study Ziemba [1] assessed the quality of the two constructs of SIS, i.e. ICT adoption and sustainability by examining the construct reliability [38], convergent validity [39, 40], and discriminant validity [39, 41]. The following measures were

calculated: the loadings of each item of each component, composite reliability (CR) of all components, average variance extracted (AVE) of all components, Cronabch's Alpha of all components, correlations between the components, the square root of AVE for each component. Overall, the results successfully established the reliability, convergent validity, and discriminant validity of the proposed two SIS constructs and their components [1].

In the other study Ziemba [2] addressed the research questions related to levels of ICT adoption and sustainability as well as ICT adoption impact on sustainability in Polish enterprises. Descriptive statistics were used for assessing the levels of ICT adoption and sustainability, whereas h-Kruskala-Wallisa and Chi-square statistics were used to identify statistically significant differences between the distributions of scores for components of ICT adoption and components of sustainability. In addition the correlation and regression analysis [42, 43] were used to estimate the correlations between a theoretical level of sustainability and the levels of ICT outlay, information culture, ICT management, and ICT quality. Overall, the results revealed that there are significant statistical differences between the highest level of ICT outlay and the lowest ones, namely the levels of ICT quality and ICT management. Moreover, the economic sustainability is at the highest level, whereas the lowest and similar levels are specific to ecological and political sustainability. Finally, it was investigated that the ICT quality, ICT management, and information culture have a significant impact on sustainability in enterprises, whereas the outlay on ICT does not have such an impact.

Nevertheless, it was observed that there is a research gap in the existing body of knowledge related to measuring ICT adoption and sustainability in enterprises using multidimensional statistical methods and on the basis of synthetic indexes of development. In order to bridge the gap this study proposes an approach to the evaluation of ICT adoption and sustainability and focuses on addressing the following questions:

Q1: What synthetic indexes may be used to evaluate ICT adoption in enterprises?
Q2: What is the level of ICT adoption in Polish enterprises?
Q3: What synthetic indexes may be used to evaluate sustainability in enterprises?
Q4: What is the level of sustainability in Polish enterprises?

3 Research Methodology

3.1 Research Instrument

The Likert-type instrument (questionnaire) was developed that consisted of two SIS constructs: ICT adoption and sustainability. The task of respondents was to assess the primary variables describing:

- The four components of the ICT adoption construct, i.e. ICT Outlay (Out), information culture (Cul), ICT management (Man), and ICT quality (Qua) (Table 1). The respondents answered the question: *Using a scale of 1 to 5, state to what extent do you agree that the following situations and phenomena result in the efficient and effective ICT adoption in your enterprise?* The scale's descriptions were: 5 – strongly agree, 4 – rather agree, 3 – neither agree nor disagree, 2 – rather disagree, 1 – strongly disagree; and

- The four components of the sustainability construct, i.e. ecological (Ecl), economic (Eco), socio-cultural (Soc), and political sustainability (Pol) (see Table 1). The respondents answered the question: *Using a scale of 1 to 5, evaluate the following benefits for your enterprise resulting from the efficient and effective ICT adoption?* The scale's descriptions were: 5 – strongly large, 4 – rather large, 3 – neither large nor disagree, 2 – rather small, 1 – strongly small.

3.2 Research Subjects and Procedure

In April 2016, the pilot study was conducted to verify the survey questionnaire. Ten experts participated in the study, i.e. five researchers in business informatics and five managers from five enterprises – leaders in the ICT application. Finishing touches were put into the questionnaire, especially of a formal and technical nature. No substantive amendments were required.

The subjects in the study were enterprises from the Silesian Province in Poland. The choice of this region was driven by the fact of its continuous and creative transformations related to restructuring and reducing the role of heavy industry in the development of research and science, supporting innovation, using *know-how* and transferring new technologies, as well as increasing importance of services. In response to the changing socio-economic and technological environment intensive work on the development of the information society has been undertaken in the region for several years. In the next development strategies of the information society it was and is assumed that the potential of the region, especially in the design, provision and use of advanced information and communication technologies will be increasing [44]. All this means that the results of this research can be reflected in innovative efforts to build a sustainable information society in the region and, at the same time, constitute *a modus operandi* for other regions throughout the country and other countries.

Selecting a sample is a fundamental element of a positivistic study [45]. The stratified sampling and snowball sampling were therefore used to obtain a sample that can be taken to be true for the whole population. The following strata were identified based on the enterprise's size (defined in terms of the number of employees).

The subjects were advised that their participation in completing the survey was voluntary. At the same time, they were assured anonymity and guaranteed that their responses would be kept confidential.

3.3 Data Collection

Having applied the Computer Assisted Web Interview and employed the SurveyMonkey platform, the survey questionnaire was uploaded to the website. The data were collected during a two-month period of intense work, between 12 May 2016 and 12 July 2016. After screening the responses and excluding outliers, there was a final sample of 394 usable, correct, and complete responses. The sample ensured that the error margin for the 97% confidence interval was 5%.

Table 2 provides details about the enterprise's size, type of business activities, and economy sector.

Table 2. Analysis of enterprises profiles (N = 394)

Demographic profile	Number of respondents	Percentage of respondents
Number of employees		
250 and above (large)	78	19.80%
50–249 (medium)	83	21.07%
10–49 (small)	122	30.96%
less than 10	111	28.17%
Economy sector		
I sector – producing raw material and basic foods	27	6.85%
II sector – manufacturing, processing, and construction	83	21.07%
III sector – providing services to the general population and to businesses	238	60.40%
IV sector – including intellectual activities	46	11.68%
Business activities		
ICT (manufacturing, trade, services)	136	34.52%
No ICT	258	65.48%

3.4 Data Analysis

The data were stored in Microsoft Excel format. Using Statistica package and Microsoft Excel, the collected data were analyzed in two stages. The first stage evaluated ICT adoption and sustainability in Polish enterprises, whereas the other stage classified the analyzed enterprises into clusters according to their levels of ICT adoption and sustainability.

In the first stage, the calculations were performed on the basis of Hellwig's taxonomic measure of development also known as Hellwig's synthetic indicator of development [46] which has been adopted for multivariate comparative analyses [47]. This measure is one of classical methods of linear ordering of multivariate objects. It determines the Euclidean distance of each multivariate object from the development pattern (reference object, pattern object) [48, 49].

The procedure of determining Hellwig's synthetic indicator of development is as follows:

1. Selection of diagnostic variables based on substantive and statistical reasons, and gathering of relevant statistics data.
2. Standardization (normalization) of diagnostic variables x_{ij} according to the following formula:

$$z_{ij} = \frac{x_{ij} - \bar{x}_J}{S(x_j)}$$

where:

i – the number of objects (respondents);
j – the number of variables;
$\bar{x}_J$ – the mean value of the variable j;
$S(x_j)$ – the standard deviation of the variable j; and
z_{ij} – the standardized (normalized) value of the variable j for the object i (x_{ij})

3. Determination of the object with the best (highest) values of the diagnostic variables – (the reference object, development pattern) on the basis of the following formula:

$$z_{0j} = \max_i \{z_{ij}\}$$

where:

z_{0j} – the standardized maximum value of the variable z_{ij} when the variable is a stimulant – the so-called pattern (reference)

4. Calculation of the Euclidean distance d_{i0} of each object from the constructed pattern:

$$d_{i0} = \left[\sum_{j=1}^{m} \left(z_{ij} - z_{0j}\right)^2\right]^{\frac{1}{2}}$$

5. Determination of the value of the Hellwig's synthetic indicator of development d_i for the object i according to the following formula:

$$d_i = 1 - \frac{d_{i0}}{d_0}$$

where:

d_{i0} – the Euclidean distance of the object from the reference object;
d_0 – the distance of the object from the pattern determined in accordance with the following formula to normalize d_i in the interval [0, 1]:

$$d_0 = \bar{d}_0 + 3SD(d_0)$$

where:

$SD(d_0)$ – the standard deviation of d_0; and
$\bar{d}_0$ – the average distance between objects and the development determined in accordance with the following formulas:

$$\bar{d}_0 = \frac{1}{n}\sum_{i=1}^{n} d_{i0}$$

$$SD(d_0) = \left[\frac{1}{n}\sum_{i=1}^{n}(d_{i0} - \bar{d}_0)^2\right]^{\frac{1}{2}}$$

The values of the synthetic indicator d_i are in the interval [0, 1] and a higher value of the indicator means that the object is closer to the pattern. The closer the value of a given object to the reference object the higher the level of development is.

Additionally, the analysis of variance (Anova Kruskala-Wallisa) was used to determine if there were statistically significant differences between distributions of d_i for the ICT adoption components and sustainability components in the examined enterprises.

In the other stage, implementation of cluster analysis resulted in grouping of the analyzed subjects – enterprises – into three clusters according to the levels of ICT adoption and sustainability. Two measures: arithmetic mean $\overline{d}$ and standard deviation *SD(d)* of values d_i were used for this purpose:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i; \quad SD(d) = \left[\frac{1}{n}\sum_{i=1}^{n}(d_i - \bar{d})^2\right]^{\frac{1}{2}}$$

The following groups (classes) of enterprises according to their ICT adoption and sustainability development were defined [50]:

- Group 1 with the highest level of synthetic indicator of development (high level of ICT adoption/sustainability) embraces enterprises at a distance from the pattern exceeding $\bar{d} + SD(d)$:

$$d_i > \bar{d} + SD(d)$$

- Group 2 with the medium level of synthetic indicator of development (medium level of ICT adoption/sustainability) embraces enterprises at a distance from the pattern ranking in the interval $(\bar{d} - SD(d), \bar{d} + SD(d)]$:

$$\bar{d} - SD(d) < d_i \leq \bar{d} + SD(d)$$

- Group 3 with the lowest level of synthetic indicator of development (low level of ICT adoption/sustainability) embraces enterprises at a distance from the pattern not exceeding $\bar{d} - SD(d)$:

$$d_i \leq \bar{d} - SD(d)$$

We determined means of d_i as an indicator of development level and after that found a structure of development groups.

4 Research Findings

The statistical analyses enabled the differentiation of enterprises in terms of their levels of ICT adoption and sustainability. The generalized comparison of the levels of ICT adoption and sustainability is followed by a description of the situation of three analyzed development groups in terms of components of ICT adoption and sustainability.

4.1 ICT Adoption Indexes

The following research questions were posed regarding ICT adoption in enterprises:

Q1: What synthetic indexes may be used to evaluate ICT adoption in enterprises?
Q2: What is the level of ICT adoption in Polish enterprises?

The ICT adoption construct includes 30 variables divided into four components, i.e. ICT outlay (Out), information culture (Cul), ICT management (Man), and ICT quality (Qua) (Table 1). Therefore, in order to answer these two questions the following Hellwig's synthetic indicators of ICT adoption were determined and calculated:

- Index of ICT adoption – ICT;
- Sub-index of ICT outlay – ICT(Out);
- Sub-index of information culture – ICT(Cul);
- Sub-index of ICT management – ICT(Man); and
- Sub-index of ICT quality – ICT(Qua).

Table 3 contains the value of such an index and sub-indexes of ICT adoption for Polish examined enterprises.

Table 3. Synthetic indicators of ICT adoption in Polish enterprises determined by Hellwig's method

Index/sub-index	Value of index/sub-index	Standard deviation	Group 1	Group 2	Group 3
ICT	0.59	0.20	15.7%	66.8%	17.5%
ICT(Out)	0.65	0.21	16.0%	69.5%	14.5%
ICT(Cul)	0.60	0.20	17.3%	66.5%	16.2%
ICT(Man)	0.58	0.20	16.0%	68.3%	15.7%
ICT(Qua)	0.61	0.20	16.0%	67.0%	17.0%

The value of ICT index is 0.59 (Table 3) which means that the level of ICT adoption in Polish examined enterprises is distant from the pattern by 0.41. The 2nd group of ICT adoption development – the group with the medium level of ICT adoption – comprises the largest number of examined enterprises (66.8%). The highest and lowest levels of ICT adoption are specific for a smaller and similar number of enterprises, respectively 15.7% and 17.5% of enterprises. In addition, 82.5% of enterprises are characterized by medium and high levels of ICT adoption.

The values of the ICT adoption sub-indexes are in the interval [0.58, 0.65] which means that the levels of ICT outlay, information culture, ICT management, and ICT quality are similar distant from the pattern by, respectively 0.35, 0.40, 0.42, 0.39. In addition, the values of h-Kruskala-Wallisa H(3, N = 1576) = 23.496 and p = 0.000) confirmed significant differences between the distribution of d_i calculated for ICT outlay in enterprises and the distributions of d_i calculated for information culture (p = 0.001) and ICT management (p = 0.000) in such enterprises.

For all ICT components, the 2nd group of ICT adoption development is predominant, while the respective shares of enterprises in groups 1st and 3rd are similar and much smaller than in the 2nd group. The largest number of examined enterprises is in the 1st and 2nd groups combined. In general, the highest percentages of enterprises are in the 1st and 2nd groups of ICT outlay (85.5% of enterprises) and ICT management (84.3% of enterprises). In a slightly smaller percentage of enterprises medium and high levels of information culture and ICT quality are indicated, respectively 83.8% and 83.0% of enterprises. Figure 1 presents the structure of the ICT adoption groups by ICT adoption components.

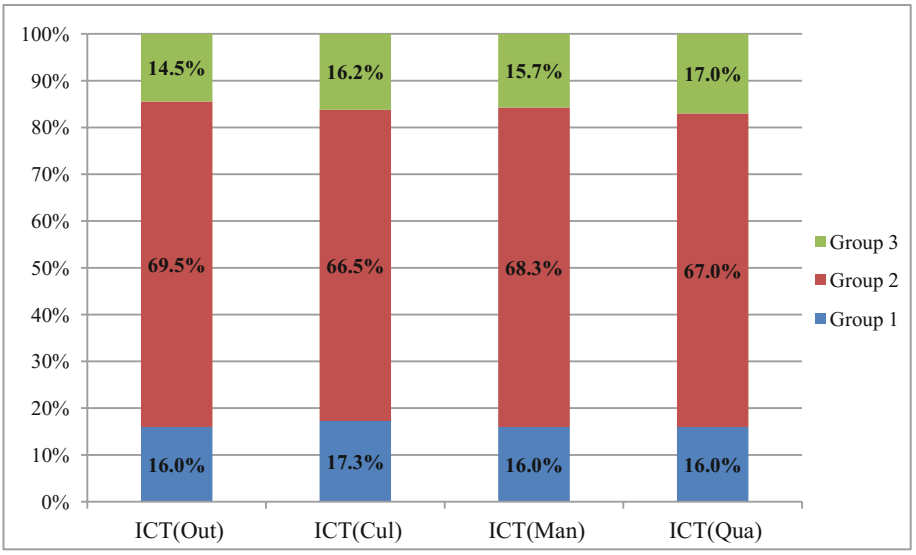

Fig. 1. Percentage of enterprises in the groups of ICT adoption.

4.2 Sustainability Indexes

The following research questions were posed regarding sustainability in enterprises:

Q3: What synthetic indexes may be used to evaluate sustainability in enterprises?
Q4: What is the level of sustainability in Polish enterprises?

The sustainability construct includes 16 variables divided into four components, i.e. ecological (Ecl), economic (Eco), socio-cultural (Soc), and political sustainability

(Pol) (Table 1). Therefore, in order to answer these two questions the following Hellwig's synthetic indicators of sustainability were determined and calculated:

- Index of sustainability – SIS;
- Sub-index of ecological sustainability – SIS(Ecl);
- Sub-index of economic sustainability – SIS(Eco);
- Sub-index of socio-cultural sustainability – SIS(Soc); and
- Sub-index of political sustainability – SIS(Pol).

Table 4 contains the value of such an index and sub-indexes of sustainability for Polish enterprises.

Table 4. Synthetic indicators of sustainability in Polish enterprises determined by Hellwig's method

Index/sub-index	Value of index/sub-index	Standard deviation	Group 1	Group 2	Group 3
SIS	0.59	0.20	17.3%	67.0%	15.7%
SIS(Ecl)	0.65	0.22	9.6%	71.8%	18.5%
SIS(Eco)	0.60	0.20	15.5%	70.6%	14.0%
SIS(Soc)	0.61	0.20	16.5%	66.2%	17.3%
SIS(Pol)	0.65	0.22	8.6%	72.8%	18.5%

The value of SIS index is 0.59 (Table 4) which means that the level of sustainability in Polish examined enterprises is distant from the pattern by 0.41. The 2nd group of sustainability – the group with the medium level of sustainability – comprises the largest number of enterprises (67.0%). The highest and lowest levels of sustainability are specific for a smaller and similar number of enterprises, respectively 17.3% and 15.7% of enterprises. In addition, 84.3% of enterprises are characterized by medium and high levels of sustainability.

The values of the sustainability sub-indexes are in the interval [0.60, 0.65] which means that the levels of ecological, economic, socio-culture, and political sustainabilities are similar distant from the pattern by, respectively 0.35, 0.40, 0.39, 0.35. In addition, the values of h-Kruskala-Wallisa H(3, N = 1576) = 16.005 and p = 0.001) confirmed significant differences between the distribution of d_i calculated for ecological sustainability and the distributions of d_i calculated for economic (p = 0.008) and socio-culture (p = 0.030 sustainability in enterprises as well as between the distributions of d_i calculated for economic and political sustainability (p = 0.033) in enterprises.

For all SIS components, the 2nd group of sustainability is predominant, while the respective shares of enterprises in the 1st and 3rd group are similar and much smaller than in group 2nd. The largest number of examined enterprises is in the 1st and 2nd group combined. In general, the highest percentages of enterprises are in the 1st and 2nd group of economic sustainability (86.0% of enterprises). A slightly smaller percentage of enterprises are characterized by medium and high levels of ecological, socio-culture, and political sustainability, respectively 81.5%, 82.7%, and 81.5% of enterprises. Figure 2 presents the structure of the sustainability groups by sustainability components.

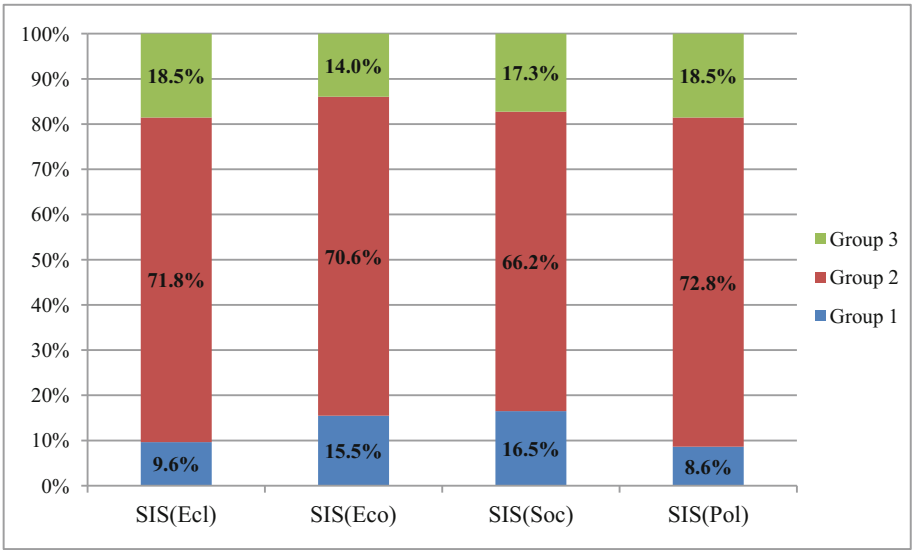

Fig. 2. Percentage of enterprises in the groups of sustainability.

4.3 ICT Adoption and Sustainability Levels

The results of the grouping of enterprises by their levels of ICT adoption and sustainability using Hellwig's method are shown in Fig. 3.

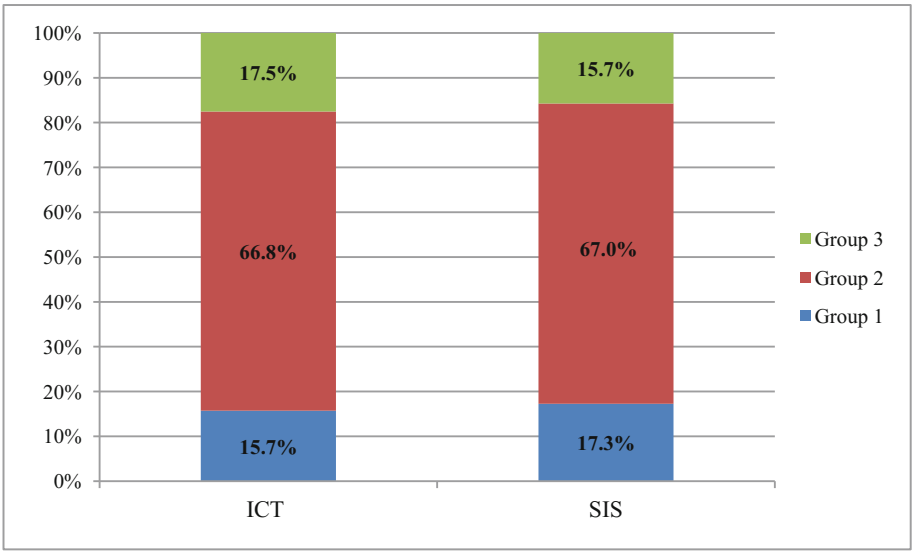

Fig. 3. Percentage of enterprises in the groups of ICT adoption and sustainability.

It has been found that the largest number of examined enterprises (67%) is at the medium levels of ICT adoption and sustainability. These levels are in the interval (0.39, 0.79] according to the indexes calculated based on Hellwig's method. The high levels of ICT adoption and sustainability indexes exceeding value of 0.79 are characteristic for, respectively 15.7% and 17.3% of enterprises. However, 17.5% and 15.7% of enterprises are at the low levels of ICT adoption and sustainability and their ICT adoption indexes are not exceeding the level 0.39 according to the Hellwig's synthetic indicator.

5 Conclusions

5.1 Research Contribution

This work contributes to the existing research on the SIS, especially the evaluation of ICT adoption and sustainability in enterprises by:

- proposing synthetic indexes of ICT adoption in the context of SIS determined by Hellwig's method and indicating the level of ICT adoption in enterprises, also separated into the levels of ICT outlay, information culture, ICT management, and ICT quality by using such indicators; and
- proposing synthetic indexes of sustainability resulting from ICT adoption determined by Hellwig's method and indicating the level of sustainability in enterprises, also separated into the levels of ecological, economic, socio-cultural, and political sustainability by using such indicators.

The new indexes have been developed, based on a set of indicators in accordance with the nature of SIS, the essence of which is environmental, economic, socio-culture, and political sustainability strongly influenced by adopting ICT in enterprises.

With regard to the presented results, it is reasonable to conclude that this study expands the existing research on the SIS provided by Schauer [9], Fuchs [3, 4], Hilty et al. [6, 7], Guillemette and Paré [16, 17], and Curry and Donnellan [14, 15]. Additionally, the findings of this study add new insights into the measurement of ICT adaption [12, 30, 31] and sustainability [34–37].

5.2 Implications for Research and Practice

The research findings can be used by scholars to improve and expand the research on the SIS. Researchers may use the proposed synthetic indicators to do similar analyses with different sample groups in other countries, and many comparisons between different countries can be drawn. Furthermore, the proposed methodology constitutes a very comprehensive basis for identifying the levels of ICT adoption and sustainability, but researchers may develop, verify and improve this methodology and its implementation. For example, components and primary variables of ICT adoption and sustainability constructs may be improved in such a way that measurement of ICT adoption and sustainability allow to gain a more precise view of SIS.

This study offers several implications for enterprises. They may find the results appealing and useful in enhancing the adoption of ICT, experiencing the full potential of ICT and deriving various benefits from ICT adoption like ecological, economic,

socio-cultural, and political. In addition, this study recommends some guidelines for measuring ICT adoption and such benefits resulting from ICT adoption. Furthermore, the findings can help enterprises develop sound ICT adoption plans and receive funding from the European Union that set itself a target of implementing the 2030 Agenda for Sustainable Development [51].

It should be emphasized that this research can be largely useful for the transition economies in Central and Eastern Europe. This is because the countries are similar with regard to analogous geopolitical situation, their joint history, traditions, culture and values, the quality of ICT infrastructure, as well as building democratic state structures and a free-market economy, and participating in the European integration process.

5.3 Limitations and Future Works

As with many other studies, this study has its limitations. First, the ICT adoption and sustainability constructs are new constructs that have yet to be further explored and exposed to repeated empirical validation. Second, the sample included Polish enterprises only, especially from the Silesian Province. The study sample precludes statistical generalization of the results from Silesian enterprises to Polish enterprises. However, early research into the success factors for and the level of adopting ICT in Poland [52] indicated that there is no difference between Silesian enterprises and other Polish enterprises. Therefore, these research findings cannot be limited only to the Silesian enterprises and can be generalized to Polish enterprises. After all, caution should be taken when generalizing the findings to other regions and countries. Finally, the research subjects were limited to enterprises and it is therefore only the viewpoint of enterprises toward ICT adoption for achieving sustainability in the information society. Caution should be taken when generalizing the findings to the SIS.

Additional research must be performed to better understand the SIS, ICT adoption and sustainability construct, and the levels of ICT adoption and sustainability. First, the further validation of the levels of ICT adoption and sustainability should be carried out for a larger sample comprising enterprises from different Polish provinces. Second, research on the measurement of ICT adoption and sustainability in households and government units should be conducted.

Acknowledgement. I wish to express my sincere gratitude to my friend Maria Jadamus-Hacura for helping with statistical analysis.

References

1. Ziemba, E.: The contribution of ICT adoption to the sustainable information society. J. Comput. Inform. Syst. (2017). https://doi.org/10.1080/08874417.2017.1312635
2. Ziemba, E.: The ICT adoption in enterprises in the context of the sustainable information society. In: Ganzha, M., Maciaszek, L., Paprzycki M. (eds.) Proceedings of the 2017 Federated Conference on Computer Science and Information Systems FedCSIS, 3–6 September 2017, pp. 1031–1038. Czech Technical University in Prague, Prague (2017). https://doi.org/10.15439/2017f89

3. Fuchs, Ch.: Sustainable information society as ideology (part I). Inf. Tarsadalom **9**(2), 7–19 (2009)
4. Fuchs, Ch.: Sustainable information society as ideology (part II). Inf. Tarsadalom **9**(3), 27–52 (2009)
5. Fuchs, Ch.: Theoretical foundations of defining the participatory, co-operative, sustainable information society. Inf. Comm. Soc. **13**(1), 23–47 (2010)
6. Hilty, L.M., Aebischer, B.: ICT for sustainability: an emerging research field. Adv. Intell. Syst. Comput. **310**, 1–34 (2015)
7. Hilty, L.M., Hercheui, M.D.: ICT and sustainable development. In: Berleur, J., Hercheui, M.D., Hilty, Lorenz M. (eds.) CIP/HCC - 2010. IAICT, vol. 328, pp. 227–235. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15479-9_22
8. Houghton, J.W.: ICT and the environment in developing countries: a review of opportunities and developments. In: Berleur, J., Hercheui, M.D., Hilty, Lorenz M. (eds.) CIP/HCC - 2010. IAICT, vol. 328, pp. 236–247. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15479-9_23
9. Schauer, T.: The Sustainable Information Society – Vision and Risks. The Club of Rome – European Support Centre, Vienna (2003)
10. Servaes, J., Carpentier, N. (eds.): Towards a Sustainable Information Society. Deconstructing WSIS. Intellect, Portland (2006)
11. Ziemba, E.: The holistic and systems approach to a sustainable information society. J. Comput. Inform. Syst. **54**(1), 106–116 (2013)
12. Ziemba, E. (ed.): Towards a Sustainable Information Society: People, Business and Public Administration Perspectives. Cambridge Scholars Publishing, Newcastle upon Tyne (2016)
13. Avgerou, C.: Discourses on ICT and development. Inf. Technol. Int. Dev. **6**(3), 1–18 (2010)
14. Curry, E., Donnellan, B.: Understanding the maturity of sustainable ICT. In: vom Brocke, J., Seidel, S., Recker, J. (eds.) Green Business Process Management – Towards the Sustainable Enterprise, pp. 203–216. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27488-6_12
15. Donnellan, B., Sheridan, C., Curry, E.: A capability maturity framework for sustainable information and communication technology. IT Prof. **13**(1), 33–40 (2011)
16. Guillemette, M.G., Paré, G.: Toward a new theory of the contribution of the IT function in organizations. MIS Q. **36**(2), 529–551 (2012)
17. Guillemette, M.G., Paré, G.: Transformation of the information technology function in organizations: a case study in the manufacturing sector. Can. J. Adm. Sci. **29**, 177–190 (2012)
18. Seidel, S., Recker, J., vom Brocke, J.: Sensemaking and sustainable practicing: functional affordances of information systems in green transformations. MIS Q. **37**(4), 1275–1299 (2013)
19. Watson, R.T., Boudreau, M.C., Chen, A.J., Huber, M.: Green IS: building sustainable business practices. In: Watson, R.T. (ed.) Information Systems, pp. 247–261. Global Text Project, Athens (2008)
20. Kodakanchi, V., Abuelyaman, E., Kuofie, M.H.S., Qaddour, J.: An economic development model for IT in developing countries. Electron. J. Inf. Syst. Dev. Ctries. **28**(7), 1–9 (2006)
21. Reardon, J., Miller, C., Vida, I., Kim, I.: The effects of ethnocentrism and economic development on the formation of brand and ad attitudes in transitional economies. Eur. J. Mark. **39**(7/8), 737–754 (2005)
22. Nord, J.H., Riggio, M.T., Paliszkiewicz, J.: Social and economic development through information and communications technologies: Italy. J. Comput. Inform. Syst. **57**(3), 278–285 (2017)

23. Ross, J.W., Vitale, M.R.: The ERP revolution: surviving vs. thriving. Inf. Syst. Front. **2**(2), 233–241 (2000)
24. Sachs, J.D.: The Age of Sustainable Development. Columbia University Press, New York (2015)
25. Huang, A.H.: A model for environmentally sustainable information systems development. J. Comput. Inform. Syst. **49**(4), 114–121 (2009)
26. Moldan, B., Janoušková, S., Hák, T.: How to understand and measure environmental sustainability: Indicators and targets. Ecol. Indic. **17**, 4–13 (2012)
27. Missimer, M., Robèrt, K.H., Broman, G.: A strategic approach to social sustainability-part 2: a principle-based definitions. J. Clean. Prod. **149**(1), 42–52 (2017)
28. Hameed, T.: ICT as an enabler of socio-economic development. Information and Communications University, Daejeon (2015). http://www.itu.int/osg/spu/digitalbridges/materials/hameed-paper.pdf. Accessed 12 June 2016
29. Khan, R.: How frugal innovation promotes social sustainability. Sustainability **8**(10), 1034 (2016). https://doi.org/10.3390/su8101034
30. ITU: International Telecommunication Union. Measuring the Information Society Report 2016, International Telecommunication Union, Geneva (2016)
31. Information Economy Report 2017. Digitalization, Trade, and Development. United Nations Conference on Trade and Development UNCTAD (2017)
32. Ziemba, E., Żelazny, R.: Measuring the information society in Poland – dilemmas and a quantified image. In: Proceeding of the 2013 Federated Conference on Computer Science and Information Systems, Krakow, 8–11 September, pp. 1185–1192 (2013)
33. Bilbao-Osorio, B., Dutta, S., Lanvin, B. (eds.): The Global Information Technology Report 2014. Rewards and Risks of Big Data. World Economic Forum, Geneva (2014)
34. Van de Kerk, G., Manuel, A.R.: A comprehensive index for a sustainable society: the SSI—the Sustainable Society Index. Ecol. Econ. **66**, 228–242 (2008). https://doi.org/10.1016/j.ecolecon.2008.01.029
35. Panda, S., Chakraborty, M., Misra, S.K.: Assessment of social sustainable development in urban India by a composite index. Int. J. Sustain. Built Environ. **5**, 435–450 (2016). https://doi.org/10.1016/j.ijsbe.2016.08.001
36. Orsato, R.J., Garcia, A., Mendes-Da-Silva, W., Simonetti, R., Monzoni, M.: Sustainability indexes: why join in? A study of the 'Corporate Sustainability Index (ISE)' in Brazil. J. Clean. Prod. **96**, 161–170 (2015). https://doi.org/10.1016/j.jclepro.2014.10.071
37. Laxe, F.G., Bermúdez, F.M., Palmero, F.M., Novo-Corti, I.: Assessment of port sustainability through synthetic indexes. Application to the Spanish case. Mar. Pollut. Bull. **119**, 220–225 (2017). https://doi.org/10.1016/j.marpolbul.2017.03.064
38. Hinton, P.R., Brownlow, C., McMurvay, I., Cozens, B.: SPSS Explained. Routledge, East Sussex (2004)
39. Gefen, D., Straub, D.: A practical guide to factorial validity using PLS-graph: tutorial and annotated example. Commun. Assoc. Inf. Syst. **16**(5), 91–109 (2005)
40. Hulland, J.: Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strateg. Manag. J. **20**(2), 195–204 (1999)
41. Brown, T.A.: Confirmatory Factor Analysis for Applied Research. Guilford Press, New York City (2006)
42. Miles, J., Shevlin, M.: Applying Regression and Correlation. A Guide for Students and Researchers. Sage Publication, London (2007)
43. Falk, R.F., Miller, N.B.: A Primer for Soft Modeling. The University of Akron Press, Akron (1992)

44. ŚCSI: Strategia rozwoju społeczeństwa informacyjnego województwa śląskiego do roku 2015 [Strategy of information society development in Upper Silesia region]. Śląskie Centrum Społeczeństwa Informacyjnego, Katowice (2009). http://www.e-slask.pl/article/strategia_rozwoju_spoleczenstwa_informacyjnego_wojewodztwa_slaskiego_do_roku_2015. Accessed 12 June 2016
45. Collis, J., Hussey, R.: Business Research. A Practical Guide for Undergraduate and Postgraduate Students. Palgrave Macmillan, New York (2003)
46. Hellwig, Z.: Zastosowanie metody taksono micznej do typologicznego podziału krajów ze względu na poziom rozwoju oraz zasoby i strukturę wykwalifikowanych kadr (The application of the taxonomic method to the typological division of a countries due to their level of development, resources and structure of qualified personnel – in Polish). Przegląd Statystyczny **4**, 307–326 (1968)
47. Hwang, C.-L., Yoon, K.: Multiple Attribute Decision-Making: Methods and Applications. Springer, Heidelberg (1981). https://doi.org/10.1007/978-3-642-48318-9
48. Leń, P., Skrzypczak, I., Oleniacz, G., Mika, M.: The use of statistical methods for the evaluation of land adjustment proposals and elimination of the patchwork pattern of land ownership. In: "Environmental Engineering" 10th International Conference, Vilnius Gediminas Technical University Lithuania, 27–28 April 2017, 1–6 (2017). https://doi.org/10.3846/enviro.2017.214
49. Skordzka, I.: The synthetic measure of the level of smart growth in the European Union countries. Optimum. Studia Ekonomiczne **5**(83), 113–122 (2016). https://doi.org/10.15290/ose.2016.05.83.07
50. Pomianek, I., Chrzanowska, M.: A spatial comparison of semi-urban and rural gminas in Poland in terms of their level of socio-economic development using Hellwig's metod. Socioecon. Ser. Bull. Geogr. **33**, 103–117 (2016)
51. European Commission: The 2030 Agenda for Sustainable Development. https://ec.europa.eu/europeaid/policies/european-development-policy/2030-agenda-sustainable-development_en. Accessed 12 Jan 2017
52. Ziemba, E. (ed.): Czynniki sukcesu i poziom wykorzysatnia technologii informacyjno-komuniakcyjnych w Polsce [Success Factors for and Level of ICT Adoption in Poland]. Warsaw, CeDeWu (2015)

Implementation and Evaluation of Information Systems

Small Brazilian Business and IT Governance: Viability and Case Study

Daniel A. M. Aguillar[1(✉)], Isabel Murakami[1], Pedro Manso Jr.[1], and Plinio Thomaz Aquino Jr.[2]

[1] Instituto de Pesquisas Tecnológicas de São Paulo, Av. Almeida Prado, 532, São Paulo, SP, Brazil
danielaguillar@yahoo.com.br, iaadefm@hotmail.com, pedro.manso@uol.com.br

[2] Centro Universitário FEI, Av. Humberto de Alencar Castelo Branco, 3972, São Bernardo do Campo, SP, Brazil
plinio.aquino@fei.edu.br

Abstract. Small companies have potential to be agile, flexible and informal - this is possible because such companies are usually formed by few members (up to 9 people – statistic average in Brazil). This usually creates more synergy among these professionals because they incline to have more than a single role inside the company. Therefore, it is understandable that those professionals should multi-task, splitting their working hours among different kinds of demands: what may cause difficulties in planning, development, verification, and improvement of internal processes. This article brings a case study where the COBIT 5.0 toolkit (Process Assessment Model) was used to identify internal processes that needed improvement within the studied company. In order to improve the selected processes, ISO/IEC 12207 was tailored concerning company's needs. Additionally, it was applied a continuous improvement cycle – PDCA and it was also proposed the adoption of an agile method – SCRUM, to integrate internal activities and processes.

Keywords: COBIT · Software engineering · SCRUM

1 Introduction

This article presents a viability case study for the application of an Information Technology (IT) corporative governance program in a small software development company. A study [1] shows metrics about development projects – about 56% are concluded within agreed deadlines and about 68% are completed within agreed budget. There are some main factors for the non-completion of projects within deadline and budget, in a scale from 1 to 5 [1]: Poorly defined initial scope (Average: 3.6) and continuous scope changes (Average: 3.5).

In order to meet deadlines and have better budget control, IT should be aligned with business goals. Also, Development and IT areas should be efficient. Solutions must not only meet increasing quality requirements but also be low-cost and easily adaptable [1, 2]. This is a noticeable trend because there is an increasing number of Brazilian

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 173–193, 2018.
https://doi.org/10.1007/978-3-319-77721-4_10

companies pursuing the ISO 9000 certification, at the expense of other software models and standards [3]. Business goals and IT alignment can be planned and controlled with the adoption of a governance model. COBIT (Control Objectives for Information and Related Technology) is a de-facto standard used to provide IT area with a governance model and it helps with understanding and managing IT-related risks [4]. Development and IT efficiency can also be improved with well-defined processes and standardization, using ISO/IEC 12207 [5], the international standard for software lifecycle processes [2].

The process should be easily adaptable and compliant with business' changes. The adoption of agile development is increasing due to the need for fast adaption to changes. The most used agile method nowadays is SCRUM [4, 6, 7]. Considering the Governance context (COBIT), the development model (ISO/IEC 12207) and development method (SCRUM), this article aims to investigate the application viability of COBIT, ISO/IEC 12207 and SCRUM in the context of a small company (up to 9 people, statistic average in Brazil) [8]. COBIT, ISO/IEC 12007 and SCRUM should be tailored considering the limited resources of a small company, where it is common to find human resources assuming multiple roles while working on different projects at the same time. The small IT Brazilian company analyzed in this case study, identified as "studied company" had reported the following main problems: rework due to scope change, consecutive delays due to lack of stakeholders' commitment, stakeholders change and/or multiple stakeholders with inadequate communication. If not well managed, such problems can imply in financial loss, lack of productivity, frequent re-negotiations, and negative impact on other projects and human resources reallocation. Therefore, affecting commercial, development, creation and quality assurance areas – company's key areas. The goal of this study is to assess the application viability of COBIT and consequent verification, analysis and proposition of improvements in a small companies' context, providing a viability analysis of the challenges and benefits observed when applying COBIT, ISO/IEC 12007 and SCRUM in a small company context.

The primary analysis will be done with COBIT toolkit version 5.0 for Process Assessment Model (PAM), maintained by ISACA (Information Systems Audit and Control Association). We will use the term COBIT-PAM in this text to address this framework. Its application helps to verify which are the main processes influenced by the cited problems – knowing such processes will enable the application, within the right adherence level, of standards and/or techniques based in good practices of IT governance and software engineering. By improving the selected processes, some performance enhancement in company's activities is expected, therefore, directly and indirectly affecting IT governance levels in a continuous and positive way, aiming the continuity of all good practices that were proposed.

It is possible to cite as expected contributions of this study internal processes and governance improvement, and the fact that this study is related to a less studied domain: small companies, thus showing the application viability of such practices in this context. The method used in this study is formed by the following activities: (a) Collection of information and problems within the company; (b) Analysis of collected data; (c) Application of COBIT-PAM; (d) Bibliographic survey for possible solutions; (e) Analysis and proposition of a hypothetical solution and (f) Development of paper that documents adopted procedures and conclusions.

2 Literature Review: Governance Models with Agile Models

In order to reach a set of related works, a search combining the keyword COBIT, Small, SME, ISO/IEC 12207, PDCA (Plan, Do, Check and Act), agile and SCRUM was done on Google Scholar, IEEE Xplore, ACM and CAPES, selecting only works published in English or Portuguese since 2012.

Several researchers have investigated the harmonization of governance models with agile development models. The authors of study [9] used concepts of COBIT, PRINCE2, and SCRUM to propose the addition of security tests requirements and penetration tests (pen tests) into the agile development lifecycle. Silva [10] proposed a software acquisition model aligned with COBIT, ITIL and PDCA concepts for continuous improvement, adding concepts such as daily meetings from SCRUM and portfolio management from SAFe. Ozkan et al. [11] observed the migration of a company to agile models from Waterfall, while complying with COBIT – listing challenges and adopted solutions, while losing the big picture in design, a problem solved with the addition of a "zero sprint" step following a formal approval process. Similarly, this work [11] has combined COBIT, PDCA, and SCRUM, however using a different approach, adding COBIT to identify which processes should be improved on a context of small to medium-sized enterprises, using ISO/IEC 12207 for the full software development lifecycle and PDCA for continuous improvement.

In order to combine governance control with the agile model, researchers had worked on mapping COBIT process with SCRUM activities. Gupta [12] mapped COBIT controls with the development processes such as requirement analysis, design, building and testing. Additionally, author 12] also implemented solutions to make agile projects comply with Sarbanes Oxley regulatory requirements. Mahnic and Zabkar [13] validated their proposed model AGIT (AGIle software developmenT), which includes measuring SCRUM based on software development, with information systems auditing criteria, as described in COBIT. Studies [14, 15] indicate a direct connection between team communication and project's outcome: It was observed in a study [14] the impact of enhancing control over project's context and team's communication. This was done using formal control - based on performance evaluation strategy and informal control - based on social and people strategies. In [16], metrics were investigated to find out if they could provide IT management with information regarding the progress of SCRUM-based software development without harming its agility. This work has also mapped COBIT process to both ISO/IEC 12207 processes and the SCRUM model.

3 Research Method: Problem Analysis (Causes × Effects)

It was necessary to collect information and understand the processual problems exposed by this company in order to enable its analysis and propose improvements or solutions. The problem analysis method consisted of:

1. **Brainstorm:** Informal meetings with the board of directors (2 members responsible for the management of Marketing, Sales, Operations and Finances areas) and the technical team (3 members that work assuming multiple roles during software's

lifecycle) in order to detect the main problematic situations affecting the company's management and operations.

2. **Data Collection and Analysis:** All the main topics discussed during the brainstorm were analyzed and the main problematic situations detected were: (A) Rework due to scope change; (B) Consecutive delays due to lack of stakeholder's commitment to the project and its deadlines; (C) Stakeholders change and/or multiple stakeholders with inadequate communication; (D) Money loss due to delays in projects and company's unavailability to work in new projects; (E) Lack of productivity due to extra work caused by unexpected scope changes; (F) Frequent renegotiations due to inadequate process and communication, and scope change; (G) Negative impact on other projects; (H) Human resources reallocation from other projects.
3. **Brainstorm:** Meetings with board of directors to identify possible root causes and effects of each problematic situation detected;
4. **Cause-effect Matrix** (Table 1): Problematic situations were classified into causes or effects of other problems and distributed in a cause-effect matrix. The following matrix shows causes (C) and their effects (E):

Table 1. Cause-effect matrix (Read the matrix as follows: e.g. A-1 is cause/effect of B-2. Grey cells represent the selected main causes.)

Cause/Effect	A-2	B-2	C-2	D-2	E-2	F-2	G-2	H-2	Total Causes
A-1	-	E	E	C	C	C	C	C	**5**
B-1	C	-	E	C	C	C	C	C	**6**
C-1	C	C	-	C	C	C	C	C	**7**
D-1	E	E	E	-	C	E	E	C	2
E-1	E	E	E	E	-	E	E	E	0
F-1	E	E	E	C	C	-	C	C	4
G-1	E	E	E	C	C	E	-	C	3
H-1	E	E	E	E	C	E	E	-	1

After building the cause-effect matrix, an established criterion was applied to allow the identification of the main causes among all problems: it should be causing at least 70% of listed problems. The goal was to define root problems/major causes: A, B, and C were found as such.

The development method being used is based on the waterfall development model. The process is defined as follows: initial project's scope analysis and alignment between its internal manager and stakeholder; characterization of necessary documentation, project development, tests, and delivery - in modular increments.

When a module is delivered, the stakeholder might not be the same who did the initial request, his needs might have changed or it might not be what was expected because of the lack of alignment during the development process (problematic situations "B" and "C"). This will cause the problematic situation "A".

Additionally, whenever alignment is needed, meetings are scheduled without any pre-determined periodicity as there are no formalized alignment milestones.

Considering the need to generate more commitment from stakeholders, the previously proposed adoption of an agile method can be cogitated as an essential part of the solution. This may help to focus on scope's alignment and delivery validations due to frequent alignment during the entire development process.

5. **COBIT 5.0 framework application:** This framework was conceived for technology information management. Its application involves business requirements analysis considering: effectiveness, efficiency, integrity, availability, compliance, and trustability. Its results bring a panoramic view of the general maturity level of the company's IT area and help to understand what needs to be done to reach higher levels.

   It is possible to infer that this framework reduces the complexity of processual analysis to identify possible solution scenarios for company's problems. COBIT-PAM was applied addressing the previously cited problems - the assessment was done for every process described in the framework. It establishes 6 maturity levels, on a 0 to 5 scale. COBIT-PAM was applied following these activities [17]: Identify if each process is achieving its outcomes; Rate each of the outcomes as follows: Not Achieved = 0–15%, Partially Achieved = 15–50%, Largely Achieved = 50–85%, Fully Achieved = 85–100%; To pass a particular level, the process must be rated as Largely or Fully Achieved, to reach the next level, all attributes must be rated as Fully Achieved; Identify the Capability Level Achieved. The problematic processes identified were:

   - There was an initial process of governance being executed to ensure the governance framework setting and maintenance (EDM01 – Rated: Level 1), but it was not being properly managed/maintained - resulting in a lower maturity level.
   - Value optimization governance (EDM02 – Rated: Level 3) was being managed and had a defined process, but needed well-defined quality attributes to measure its effectiveness.
   - Risk optimization governance (EDM03 – Rated: Level 2) was being managed at the start and end of projects but lacked management and measurements during its lifecycle.
   - The budget and costs management process (APO06 – Rated: Level 4) establishes measuring points during the process, that will be analyzed and documented, furnishing information on costs monitoring and control. This is done since the delivery of software's functional and non-functional requirements until development. Software's architecture planning and maintenance phases may be improved with a better analysis of the problem and necessary changes. The validation and quality of tests defined in non-functional requirements must respect planned budget in costs management - always monitoring project's costs.
   - The positive aspect regarding quality management governance (APO11 – Rated: Level 3) lies in the establishment of more rigorous processes during software development (direct or indirectly).
   - A bigger level of quality control should be assured by verifying, validating and making revisions in partnership with stakeholders. This would mitigate uncovered problems in quality metrics that were found during COBIT-PAM's application.

- Risk management governance (APO12 – Rated: Level 2) should be continuous, during the entire project's lifecycle. There is an established process, but it is not managed. The risk management should be done upon the acceptance of services provision, during initial alignments with stakeholders and also during development and maintenance phases, where all problems must be addressed and discussed. Each risk must be analyzed and proper actions should be taken (avoiding, assuming, reducing or transferring risk) and documented.
- Requirement's definition governance (BAI02 – Rated: Level 2) can be improved by using a well-defined acquisition process, suggesting a process to the acquirer (if he does not have it already). Other project's phases will also have impacts, such as provision, development, maintenance, documentation, quality assurance, verification, validation, joint review, and management.
- Upon the establishment of a clearer and more well-defined process, operations management governance (DSS01 – Rated: Level 3) can be more complete, by controlling and monitoring more crucial aspects of projects. To improve these processes - requirements, risks and costs must be traceable and documented. To reach this goal, it was suggested the adoption an ALM (Application Lifecycle Management) tool that automatically registers these steps. A small company's team cannot spend time with bureaucracy that can be automated.

4 Research Findings: Analysis and Solution Hypothesis

Many studies were considered [4, 12–14, 18–20] for the choice and proposition of a viable solution for the studied company. Antón [20] explained how some important principles can help to eliminate flaws during requirement's planning phase: Understanding the problem before defining requirements, Involving stakeholders since the beginning and maintain their involvement during requirement validation phase, and making sure that critical requirements were considered.

Standish Group studies [18, 21, 22] show that 72% of IT projects had failed or were challenged, with a risk of failure. If compared to other years like 1998 and 1996 for example, the study does not show a significant improvement in those scenarios – the reduction rate average is 1–2% every 2 years in failed and challenged projects.

It is believed in another study [19] that failures occur due to a basic principle: the way managers are structuring their projects' plans, with module-based instead of business requirements-based deliveries.

A subsystem-based project is essentially a great increment, first building subsystems and then putting them together. Clients cannot review business functionalities until full integration (when the system is almost fully built). Changes in this phase can generate re-work, costs and delivery delays.

Recent studies [4, 12, 13] have been consecutively showing, throughout the years, improvements in this scenario – 37–39% of all agile projects were considered successful. The same success ratio regarding waterfall projects is around 5–11%. Thus, providing good arguments to propose the adoption of an agile method within the studied company. The main reasons for such success rates were: adoption of agile

methods; modernization (code recovering/conversion and databases), less usage of ERP's and CRM's, and less usage of waterfall development model.

In order to reach a well-based decision about which models, techniques and/or methods could be viable to solve given problems, this analysis considered: (1) Problems faced by the studied company; (2) IT goals impacting processes selected by COBIT-PAM; (3) Principles stated in study [14]; (4) Bigger success rates in agile projects as stated in [4, 12, 13]; (5) Adopted assumptions; (6) Company's policies.

After careful analysis, it was decided by the authors to unite good practices of IT governance with good practices of software engineering, using ISO/IEC 12207 [5]. This was done in order to propose an applicable and viable solution considering money/human resources restrictions. The following conclusions were reached: (I) ISO/IEC 12207 [5]: definition of good adherence level and support of standard's adaptations to attain company's needs; (II) Agile method – SCRUM [2–27]: selection and proposal of key inherent principles; (III) PDCA cycle [28–30]: implementation of continuous improvement cycle to assure continuity of proposed improvements.

Focusing on strategic challenges found after COBIT-PAM application, it was found as a viable solution the definition of an adherence level to the ISO/IEC 12207 standard [5] and the incorporation of processes based on SCRUM and PDCA cycle, as described in a study [27] that shows how to apply ISO/IEC 12207 with SCRUM and agile methods. Such practices may mitigate challenges faced by the studied company, fulfilling its strategic and governance goals by improving its quality, risk, budget and costs, and operations management – Highly IT-related processes, in this case.

Such changes may improve project's indicators, e.g. amount of projects delivered on time and within cost expectations. It is also expected a smaller impact on company's processes from the perspective of reported problems, in such a way that business' goals are affected within a controlled, predictable and/or acceptable limit, or are not affected at all. The next sessions will present the solution proposed by the authors considering a small company context, demonstrating all adjustments made in ISO/IEC 12207 [5] and its relations with software engineering techniques, SCRUM method and impacting IT processes obtained by COBIT-PAM application.

5 Discussion of Findings and Solution Proposition: ISO/IEC 12207, PDCA Cycle and SCRUM

5.1 Software Process Model Preparation

As a technical standard reference, it was adopted the ISO/IEC 12207 [5]. This standard describes a common structure for software lifecycle process. The proposed structure involves processes, activities, and tasks that may be applied simultaneously with the acquisition of software products or development services, addressing phases such as supply, development, operation, and maintenance. This standard provides a general vision of a software's lifecycle. It starts with an idea or need that must be fully or partially addressed by a software and ends with its withdrawal [10].

The ISO/IEC 12207 [5] is highly flexible and may be used along with any lifecycle model (Waterfall, Incremental, Evolutionary, Spiral, etc.); any software engineering method (Object Oriented, Structured Coding, Top-Down Tests, etc.); with any programming language. The standard's processes [5] are defined in three main classes and one extra class: Fundamental, Supporting and Organizational class. The 4th class represents an adaptation of the standard as it may be necessary (called Tailoring class) which was the focus of this study. Every process consists of activities and each activity consists of tasks. A task consumes inputs (data, information, and control) and provides outputs (data, information, and control).

Each process must follow the PDCA cycle: an integrated system of procedures that encourage employees to analyze situations, to stablish plans, to conduct performance auditing and to take appropriate actions, for corrections or progress [27].

In this standard [5], processes, their activities, and their tasks are arranged in a natural sequence, without any mandatory execution sequence. However, it requires certain outputs to be documented, not specifying format, content nor documentation media to be used.

An organization is a group of organized people with a specific goal and its name derives from the process which it is responsible for e.g. it is called acquirer then it is performing an acquisition process. Regardless of the company's size, small or large, depending on its goal, an organization might select a subgroup of processes (with related activities and tasks) to fulfill its purpose [31].

In this research, the following steps were followed for the tailoring process: (a) Lifecycle reference board: premises adopted for the selection of process; (b) Company policies and projects: description of the policies adopted by the studied company; (c) Software lifecycle: selection of the processes and sub-processes; (d) Development model: description of the development model selected; (e) Reference board activities/Software process model: mapping of the adapted ISO/IEC 12207 [5] processes, SCRUM method and the affected COBIT processes.

5.2 Lifecycle Reference Board

Software lifecycle models are used as a high-level standard definition of stages that occur during the development process. The goal is not to provide detailed instructions, but to list main activities and their relations [32]. In order to properly adapt the ISO/IEC 12207 [5], a set of assumptions was adopted:

- Auditing will not be addressed because it regards an external process: other company's responsibility;
- The adoption of a documentation automation tool is needed: to plan releases and enable suggestion and implementation of an agile development method;
- About quality assurance: it was verified, with the board of directors, that there are other professionals working on this topic internally. It is suggested, that another study is done in the future to tackle this topic, to improve and/or define quality attributes necessary for process measurement and control.

5.3 Company Policies and Projects

Main policies adopted by the studied company: (a) Development and implementation of software according to the client's needs; (b) Every project is developed in a centralized way and (c) It is a project-oriented company, that manages its human, computational and other resources in accordance with each project's demand.

Main project's characteristics observed on 50–70% of studied company's projects: (a) Low duration (up to 3 months); (b) The client's profile: diversified with different business drivers; (c) Low reuse of developed components – around 40% rate; (d) Most projects are web projects.

5.4 Software Lifecycle and Development Model

In this section, it will be described which sub-processes from the ISO/IEC 12207 [5] will be used for a better scope management. The other processes were not the focus of this study. Tables 2, 3 and 4 contain which sub-processes were selected from ISO/IEC 12207 [5].

Table 2. Selected fundamental processes

Sub-processes	Responsibility
5.1 Acquisition	Clients
5.2 Supply	Company
5.3 Development	Company
5.5 Maintenance	Company

Table 3. Selected support processes

Sub-processes	Responsibility
6.1 Documentation	Company + clients
6.3 Quality assurance	Company + clients
6.4 Verification	Company + clients
6.5 Validation	Company + clients
6.6 Review	Company + clients

Table 4. Selected organizational processes

Sub-processes	Responsibility
7.1 Management	Company

Examining company's software demands, although they receive numerous types of projects with different complexity levels, the most frequent are small/medium-sized software projects with an intermediate complexity level. Scopes are well defined, but there is a big chance of changes being necessary – mainly asked by stakeholders.

The studied company lacked a formal development process, thus, its need of one led the authors to verify which method/process could be implemented in such small company context - due to its contextual restraints, already addressed in this paper.

With such restrictions, a consensus was reached: these projects will make use of an iterative and incremental development method, so that in each iteration, it will be possible to re-evaluate the software and set up new requirements, increasing client's satisfaction – without negative impacts on the project. Such re-evaluation should be done through sub-processes described in Table 5.

On an iterative-incremental software development model, every functionality delivery happens in a planned way, as a series of partial deliveries - each one with growing completeness - developed within iterations. Every iteration focus on analysis,

Table 5. Evaluations and responsibilities

Sub-processes	Use	Responsibility
Internal process evaluation (tasks from 5.1 to 5.5)	Yes	SCRUM team
Verification (6.4) and validation (6.5)	Yes	Product owner, stakeholder, and SCRUM team
Review (6.6) and auditing (6.7)	Yes–Review No–Auditing	Product owner and Stakeholder
Quality assurance (6.3)	Yes	Product owner and SCRUM team
Improvement (7.3)	Yes	All

planning, implementation and tests of a defined set of requirements. To reach an iterative-incremental (agile) process, it is recommended to follow the Agile Manifesto and its key principles [23].

SCRUM was chosen among other methods because is an easy and low-cost method to implement within a small team context and it also generates fast development. The decision for SCRUM was reinforced by a study [19] that addressed what problems to tackle whenever a company chooses to do modular delivery instead of business goal delivery (as it used to be done by the studied company – most problems were faced by the studied company as well):

- Difficulties in coordinating dependencies between modules – integration will happen at the project's end;
- The client receives only prototypes and not software, due to the risk of not wanting something by the process's end;
- Since the client wants prototypes for demonstration purposes, project might enter a development loop if the project manager always accepts all requirement changes – the system will never be developed;
- The client cannot be sure if business' goals were achieved by the system until it is finished. If the software needs readjustments, it involves rebuilding many modules, bigger costs, and missing deadlines;
- Architecture tests occur only upon integration (project's final phases). Architecture change affects the whole project;
- Complex integration-related problems are found late during the project's development;
- It is hard to plan incremental development when the entire module is seen as one increment only.

Agile Manifesto [23] states "Working product over comprehensive documentation". One of its 12 fundamental principles states "The most efficient and effective method of conveying information to and within a development team is a face-to-face conversation". It is also suggested the adoption of a software to track project lifecycle process by automatically documenting all phases. Some suggested ALM tools were: IBM Rational CLM, HP ALM, MS ALM, Freeware IceScrum, Agilo, and eXPlainPMT.

5.5 Reference Board Activities/Software Process Model

In this section, it is shown the results of ISO/IEC 12207 [5] adaptations and respective mapping, according to applicability, of SCRUM and affected COBIT processes:

(1) **Process Model – Fundamental Process – Acquisition**

- **Activity: 5.1.1 Initiation**
  Model: Description of a concept or a need to acquire, develop or improve software product or service; Define and analyze software requirements; Choose to acquire a new software by means of a contract with the studied company.
- **Activity: 5.1.2 Preparation of Proposal Requirement**
  Model: Acquirer should document acquisition requirements (e.g. proposal request) – its content depends on the acquisition option; Acquisition documentation should include, whenever appropriate: system's requirements; scope declaration of contract-referenced tasks; proposer instructions; software list; terms and conditions; subcontracts' control; technical restrictions (e.g. targeted environment).
- **Activity: 5.1.3 Contract Preparation and Updating**
  Model: Criteria for proposal evaluation and deliberation of supplier's requirements adherence; Prepare and negotiate a contract with the supplier that addresses acquisition requirements, including cost and timetable of software product/service to be delivered; Acquirer must control contract changes through negotiations with the supplier as a part of the changes control mechanism. Contract changes must be investigated regarding their impact on project's plans, costs, benefits, quality, and schedule.
- **Activity: 5.1.4 Supplier Monitoring**
  Model: Acquirer must monitor supplier's activities regarding joint review (item 6.6) and auditing (item 6.7) processes; Acquirer should complement monitoring process with verification (item 6.4) and validation processes (item 6.5), whenever necessary.
- **Activity: 5.1.5 Acceptance and Conclusion**
  Model: Acceptance-based in strategy and defined criteria; Software product or services' review and acceptance test to be handed in – should be accepted by the supplier when criteria are encountered.

Even though the acquisition process is under the client's responsibility, since most are startups, they don't have a well-defined process for acquisition. The previously selected processes are an orientation for the adoption of an acquisition model by the client. The following processes were selected after applying COBIT-PAM, as already cited, since they have direct impact on company's IT goals: **EDM01** – Ensure governance framework and setting, **EDM02** – Ensure value optimization, **EDM03** – Ensure risk optimization, **APO06** – Manage budget and costs, **APO11** – Manage quality, **APO12** – Manage risk, **BAI02** – Define requirements, **DSS01** – Manage operations.

(2) **Process Model – Fundamental Process – Supply**

- **Activity: 5.2.1 Initiation**
  Model: The supplier conducts requirements review, accounting political matters and company's policies. It should consider risks, budget, costs and other criteria that will be adopted to assure product and process quality.
  COBIT: EDM02 EDM03 APO12 APO11 APO06 BAI02
- **Activity: 5.2.2 Response Preparation**
  Model: Supplier should define and prepare a proposal, answering the proposal requirement, including recommendations for technical standard adaptation.
  COBIT: EDM02 EDM03 APO12 APO11 APO06 BAI02
- **Activity: 5.2.3 Contract**
  Model: Supplier must negotiate and hire acquirer to provide IT product/service; Supplier can ask for contract changes as part of change control mechanism, considering budget, costs, and risk.
  COBIT: EDM02 EDM03 APO12 APO11 APO06
- **Activity: 5.2.4 Planning**
  Model: Supplier must: Review acquisition requirements to define a management structure and make sure the project has enough quality to be delivered; Select a lifecycle model for scope, size and project's complexity; Develop and document the project's management plan according to its requirements within the lifecycle model.
  COBIT: EDM02 APO11 BAI02
- **Activity: 5.2.5 Execution and Control**
  Model: Supplier must: Respect development process (item 5.3); Operate and maintain software according to operation (item 5.4) and management (item 5.5) processes, respectively; Monitor and control progress and quality of products/services throughout project's lifecycle, in terms of monitoring technical performance, costs, timetables and project status quo; Identification, register, analysis, and problem/risk resolution.
  COBIT: EDM02 APO11 DSS01 APO06 APO12 BAI02
- **Activity: 5.2.6 Review and Evaluation**
  Model: Supplier must: Coordinate activities of contract review, iterations and communications with acquirer's organization; Conduct or support informal meetings, acceptance meeting/test and joint reviews (according to Item 6.6) with acquirer; Supplier must verify and validate, according to items 6.4 and 6.5, respectively, to check if processes satisfy all requirements; Supply acquirer with evaluation, reviews, tests and problem-solving reports, as agreed in the contract; Provide acquirer with access to supplier's and subcontracted resources, to review software products or services, as specified in the contract and project's plan; Ensure quality according to item 6.3.
  COBIT: EDM02 APO11 BAI02
- **Activity: 5.2.7 Delivery and Conclusion**
  Model: Supplier must: Hand in software product/service as specified in the contract; Provide acquirer with software product/service support, as specified in the contract.
  COBIT: EDM02

(3) **Process Model – Fundamental Process – Development**

- **Activity: 5.3.1 Process Implementation**
  Model: SCRUM and ALM tool selection for requirements management and development, change management, and quality.
  COBIT: APO11 BAI02
- **Activity: 5.3.2 System's Requirements Analysis**
  Model: Defined product backlog in Story-writing workshop, where the product owner, stakeholders and team discuss all requirements, assess the risks and estimate complexity (poker planning). As a result, there will be a generated backlog (analyzed and prioritized) with a delivery plan - sprint planning.
  COBIT: EDM02 BAI02 EDM03 APO12 APO06 APO11
- **Activity: 5.3.3 System's Architecture Project**
  Model: Developed, updated and re-done during iterations; The architect is part of the peripheral project team, that is involved in specific tasks; Non-functional requirements must be considered; Measuring points, control and quality assurance must be planned; The solution must be suitable in terms of cost/budget; Risk must be assessed in architectural decisions (e.g. high availability systems).
  COBIT: BAI02 APO11 APO06 APO12
- **Activity: 5.3.4 Software Requirements Analysis**
  Model: Sprint planning and Sprint backlog
  COBIT: BAI02
- **Activity: 5.3.5 Software Architecture Project**
  Model: Story-writing workshops; Brainstorming and software's design prototyping; Flow simulation design; An initial description is supplied and refactored on next iterations.
  COBIT: EDM02 BAI02 EDM03 APO12 APO06 APO11
- **Activity: 5.3.6 Software Detailed Project**
  Model: SCRUM team principles: Multidisciplinary, self-managed and self-organized; Freedom for developers; Technical problems and risks are discussed and decided in daily meetings.
  COBIT: BAI02 APO12
- **Activity: 5.3.7 Software Coding and Testing**
  Model: Two-people coding (new or complex); Acceptance tests are part of user stories definition, determining if it is complete; Product owner and client are supposed to write acceptance tests; Tests must be part of the process and should be automated whenever possible; Engineering tests are extremely important to assure deliverable's quality; User interface, usability, performance and stress tests; Results must comply with risk policy (accept, mitigate, resolve or ignore).
  COBIT: APO11 BAI02 APO12
- **Activity: 5.3.8 Software's Integration**
  Model: Continuous integration; Integration test is part of acceptance tests.
  COBIT: APO11 BAI02 APO12

- **Activity: 5.3.9 Software Qualification Test**
  Model: Team tests during and after each sprint; End of sprint: potentially deliverable product increment with high quality, tested, complete and ready; Review meeting: team presents to product owner, stakeholder and invited people; Product owner evaluates if the sprint's goal was attained; Corrections/changes generate product backlog's new items;
  COBIT: APO11 BAI02 APO12
- **Activity: 5.3.10 System's Integration**
  Model: Aggregations must be tested when integrated, according to their requirements; Must be part of acceptance tests.
  COBIT: APO11 BAI02 APO12
- **Activity: 5.3.11 System's Qualification Test**
  Model: The delivery of a new sprint characterizes not only software qualification, but also its test qualification according to specified qualification requirements.
  COBIT: APO11 BAI02 APO12
- **Activity: 5.3.12 Software Installation**
  Model: All professionals needed for software installation (DBA, technicians, etc.) are part of the peripheral project team. They are allocated to SCRUM teams for specific tasks.
  COBIT: DSS01
- **Activity: 5.3.13 Software Acceptance Support**
  Model: The client is always available; Support and acceptance are done in review meetings at the end of each sprint.
  COBIT: EDM02 APO01 BAI02

(4) **Process model – Fundamental Process – Maintenance**

- **Activity: 5.5.1 Process Implementation**
  Model: SCRUM for maintenance: Other project's area, apart from development; Has its own backlog and deliveries.
  COBIT: APO11 BAI02
- **Activity: 5.5.2 Problem and Changes Analysis**
  Model: Every maintenance should generate a new item in the product backlog – analyzed during sprint planning; As a result of sprint planning, the project has a sprint backlog with which the maintenance team is committed; Maintainer should analyze the problem report or change request regarding their impacts within the organization, the existing system, and interacting systems, observing the following categories: (a) Kind: e.g. correction, improvement, prevention or adaptation to a new environment; (b) Scope: e.g. modification size, involved cost and change deadlines; (c) Criticality: e.g. design's impact and protection/security.
  COBIT: EDM02 BAI02 EDM03 APO12 APO06 APO11
- **Activity: 5.5.3 Implementation Modification**
  Model: Maintainer should analyze and determine which documentation, software units, and versions need to be modified; Maintainer should use development process to implement modifications; Development process' requirements should be completed.
  COBIT: APO11 BAI02 APO12

- **Activity: 5.5.4 Maintenance Review/Acceptance**
  Model: As in item 5.3.7, acceptance tests must be part of the user story; The client is always available; Support and acceptance must be done in review meetings at the end of each sprint.
  COBIT: APO11 BAI02 APO12 EDM02
- **Activity: 5.5.6 Software Discontinuation**
  Model: Best practice – Discontinuation tasks should be defined in sprint planning.
  COBIT: EDM02 BAI02 DSS01

(5) **Process Model – Fundamental Process – Documentation**

- **Activity: 6.1.1 Process Implementation**
  Model: SCRUM is not against documentation, but preconizes minimal necessary documentation; Automation: using ALM tool; Requirements must be traceable during the entire project's lifecycle.
  COBIT: EDM02 BAI02 EDM03 APO12 APO06 APO11 DSS01
- **Activity: 6.1.2 Project and Development**
  Model: Use of tool's models – may be modified according to team preference (item 6.1.4).
- **Activity: 6.1.3 Production**
  Model: Tool must contain models and pre-defined reports, ready to be used.
- **Activity: 6.1.4 Maintenance**
  Model: Tool allows customization of models and creation of new documents, if necessary.

(6) **Process Model – Fundamental Process – Verification**

- **Activity: 6.4.1 Implementation Process/6.4.2. Verification**
  Model: Verification is integrated to the ALM tool; It is possible to analyze, review and test, blocking or releasing the continuation of an activity; Lifecycle activities and software products that require verification must be determined based on scope, magnitude, complexity, and critical factors' analysis; All problems and non-conformities must be solved; Results of verification activities must be available to acquirer and other organizations that may be involved in the process.
  COBIT: APO11 BAI02 APO12 EDM02 EDM03

(7) **Process Model – Fundamental Process – Validation**

- **Activity: 6.5.1 Process Implementation**
  Model: Use of ALM tool for tests and construction planning, test management functions during entire software lifecycle, integration of requirement management and defect-monitoring tools, test results registry, and project's history – may be used for auditing purposes.
  COBIT: APO11 BAI02 APO12 EDM02 EDM03

- **Activity: 6.5.2 Validation**
  Model: Set up test plan: Define business and test goals; Review and approval processes for test plan and each test case; Manage project's requirements and test cases, establishing interdependencies; Estimate testing effort; Define schedule for each test iteration and follow other important test dates; Make a list of many environments to be tested and generate test configurations; Create test plan snapshot in a moment in time; Define quality goals, admission and output criteria; Create and manage test cases; Check test's progress during execution.
  COBIT: APO11 BAI02 APO12 EDM02 EDM03 APO06

(8) **Process Model – Fundamental Process – Joint Review**

- **Activity: 6.6.1 Process Implementation**
  Model: Sprint review is done at each sprint end with those interested; All required resources to conduct reviews must be agreed between parties; Such resources include: personnel, location, facilities, hardware, software, and tools; Parties should agree with the following items in each review: schedule; software (result of an activity), problems to be reviewed, scope, procedures, and criteria for review's start and end; Detected problems during reviews must be registered and included in problem-solving process, as required; Review results must be documented and distributed. The revising party will present to the revised party the suitability (e.g. approval, disapproval or conditional approval) of review results; Parties must agree with review results and any responsibilities for action items and finishing criteria;
  COBIT: EDM02 EDM03 APO11 BAI02 APO12 APO06
- **Activity: 6.6.2 Project's Management Review**
  Model: Sprints and project follow up are done mainly by sprint burndown graphics and project burndown.
  COBIT: EDM02 BAI02 APO11
- **Activity: 6.6.3 Technical Reviews**
  Model: SCRUM promotes continuous verification. The client is committed to elaborate use cases in story-writing meetings and in sprint deliveries when the team presents the sprint review.
  COBIT: EDM02 BAI02 APO11

(9) **Process Model – Fundamental Process – Quality Assurance**

- **Activity: 6.3.1 Process Implementation**
  Model: Best practice – At each sprint delivery, the product owner checks system's controls, tests its results and checks which are the risks and how to treat them; Quality attributes must be defined, controls implemented and measured regarding given goals; At the end of each sprint, a sprint retrospective should be done: the whole team must evaluate what was good, what could be improved and who's in charge of representing the inspection-adaptation spirit; Using results from other support processes such as verification, validation, joint reviews and problem resolution.
  COBIT: EDM02 BAI02 APO11 EDM03APO12 DSS01

- **Activity: 6.3.2 Product Assurance**
  Model: Product guarantee is documented using the verification tools (item 6.4), validation (item 6.5), joint review (item 6.6) and problem resolution (item 6.8).
  COBIT: EDM02 BAI02 APO11 EDM03 APO12 DSS01
- **Activity: 6.3.3 Process Assurance**
  Model: All processes used in software lifecycle will be documented in the tool and may be verified by the client;
  COBIT: EDM02 BAI02 APO11 EDM03 APO12 DSS01
- **Activity: 6.3.4 Quality Assurance Systems**
  Model: Addresses ISO/IEC 9001:2000 [33] – Better planning and control of work routines, eliminating unnecessary steps; Standardization of tasks and responsibility definition for greater security and work agility; Control system to identify and treat anomalies; Work striving for improvements in quality and clients' satisfaction.
  COBIT: EDM02 BAI02 APO11 EDM03 APO12 DSS01

(10) **Process Model – Fundamental Process – Management**

- **Activity: 7.1.1 Scope Initiation and Definition**
  Model: Product owner is responsible for the project and defines the project's vision, that represents its needs and what should be satisfied at the end of the project; Project's vision is always reviewed during meetings: story writing meeting, sprint planning, sprint review, daily SCRUM, review retrospective; Managing process must be initiated establishing process' requirements; With the established requirements, manager must establish process viability, verifying if required resources (people, material, technology and environment) are available, adequate and appropriate for executing and managing the process – and also if deadlines are attainable; Whenever necessary, with all involved parties approval, process requirements can be modified at this point to attain conclusion criterion.
  COBIT: EDM01 EDM02 EDM03 APO06 APO11 APO12 BAI02 DSS01
- **Activity: 7.1.2 Planning**
  Model: Product backlog planning in story-writing workshops; Elaboration of iterations in sprint planning; ALM tool usage for product backlog. Releases and sprint planning; Project's costs/budget control planning; Project's risk plan.
  COBIT: EDM01 EDM02 EDM03 APO06 APO11 APO12 BAI02 DSS01
- **Activity: 7.1.3 Execution and Control**
  Model: Project and sprint monitoring through project and sprint burndown charts, respectively – both prepared by ALM tool automatically.
  COBIT: APO06 APO11 APO12 BAI02 DSS01
- **Activity: 7.1.4 Review and Evaluation**
  Model: Sprint review at the end of each sprint.
  COBIT: EDM02 EDM03 APO06 APO11 APO12 BAI02 DSS01
- **Activity: 7.1.5 Conclusion**
  Model: Last sprint review – finishing project.
  COBIT: EDM02 EDM03 APO06 APO11 APO12 BAI02 DSS01

6 Conclusion

In this small-company case study, it was applied COBIT-PAM in order to identify which problematic processes needed improvements to reach IT goals.

After improving selected processes, ISO/IEC 12207 was adapted according to the company's needs, and the adoption of PDCA cycle was also suggested for continuous improvement along with SCRUM method implementation to organize company's development processes.

This research's goals can be defined as attained: It was verified as viable the diagnostic of a small company's IT governance using COBIT, processes that impacted strategic goals were identified and solutions were proposed such as the definition of formalized processes, based on ISO/IEC 12207. The PDCA cycle along with SCRUM contributed to the improvement process and it may enable the company to adopt continuous delivery strategies, gaining more competitiveness in the market. This study was limited to processes related to software engineering in ISO/IEC 12207 and IT governance impacting solutions were proposed. These propositions of improvement on processes involved the establishment of a set of controls and processes with positive impacts on the company's management, helping it achieving higher maturity levels.

Concluding, an objective and selective process review is obtained by the application of COBIT-PAM and it is possible to adopt standards and patterns in a tailored level of adherence and complexity within processes, enabling better processes and reduction of efforts/costs – mainly in the studied context.

6.1 Research Contribution

This study's goals (initially defined) are understood as "reached", providing the following contributions:

- It was verified that small-sized companies that want to prepare for growth need to have well-defined processes and can consider the application of COBIT-PAM;
- It was understood as feasible the adoption of standardized processes related to the software engineering lifecycle, regardless of the size of the organization;
- It was shown that using the ISO/IEC 12207 it is possible to standardize software's development processes, regardless of available resources;
- Contributions were identified, within the given context, to the improvement of processes when applying COBIT-PAM, PDCA cycle and SCRUM method together.

6.2 Implications for Research and Practice

The research method used in this study can be followed by any researcher, regardless of the company's size or its actual problems.

In this study, informal meetings were done to retrieve information on the main problems with the board of directors and the technical team. However, researchers may use more formalized methods instead, such as questionnaires with closed questions. Once analysis and comprehension of the main problems are done, researchers may

proceed and fill the COBIT-PAM in order to identify the main process gaps within the studied company.

Following these steps, and then identifying the problematic processes to be improved, researchers may tailor the standard ISO/IEC 12207 and prepare a software development model while practitioners may improve the actual software development model adopted by the company.

6.3 Limitations and Future Work

The application of COBIT-PAM in this company was complex because small companies usually don't have specific departments to manage their internal processes. All roles and attributions, in general, are treated by a reduced amount of people, that accumulate roles, mixing activities in different processes.

It was also noticed that some activities and processes such as the monitoring of costs and risks (during the project's lifecycle) were not being formally executed. Although the company was aware of the consequences and impacts of scope changes in the project, there were no formalized processes to tackle this problem in order to standardize decisions in such scenarios.

During the development of this case study it was verified that COBIT-PAM governance framework, when applied in a small company's context, might be challenging, because it requires some adaptions (given the deepness of its analysis) such as reduction of its high complexity (may be hard to understand and use it), having enough time and human resources on getting it to a reduced scope (for analysis) and bringing together the high volume of information/necessary resources for a precise diagnostic.

It is possible to imply that these might be challenging situations, because in a small company context, there is a reduced amount of professionals to execute certain tasks, and they end up accumulating responsibilities regarding some processes, that may have different maturity levels.

References

1. Guevara, J.K., Hall, L., Stegman, E.: IT Key Metrics Data 2014: Key Applications Multiyear. Gartner, Stamford (2013)
2. Christof, C., Shankar, K.: Industry trends 2017. IEEE Softw **34**(2), 112–116 (2017). https://doi.org/10.1109/MS.2017.558SEBRAE
3. Weber, K., Araújo, E., Rocha, A.R., Oliveira, K., Rouiller, A.C., Wangenhein, C.G., Araújo, R., Salviano, C., Machado, C.F., Scalet, D., Galagarra, O., Amaral, M.P., Yoshida, D.: Melhoria de Processo do Software Brasileiro (MPS.BR): um Programa Mobilizador. CLEI (2015)
4. Vallerão, A.G., Roses, L.K.: Monitoramento e controle de projetos de desenvolvimento de Software com o SCRUM: avaliação da Produção Científica. In: Revista de Gestão e Projetos, vol. 4, no. 2 (2013). https://doi.org/10.5585/gep.v4i2.154
5. ABNT – ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS: NBR ISO/IEC 12207 – Tecnologia de informação - Processos de ciclo de vida de software, p. 35. ABNT, Rio de Janeiro (1998)

6. Ozkan, N.: Risks, challenges and issues in a possible SCRUM and COBIT marriage. In: 2015 Asia-Pacific Software Engineering Conference (APSEC), New Delhi, pp. 111–118 (2015). https://doi.org/10.1109/APSEC.2015.29
7. Bunyakiati, P., Surachaikulwattana, P.: Fit between Agile practices and organizational cultures. In: 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, pp. 1–6 (2016). https://doi.org/10.1109/JCSSE.2016.7748915
8. SEBRAE: Micro e Pequenas Empresas em Número. SEBRAE São Paulo (2017)
9. Tomanek, M., Klima, T.: Penetration testing in agile software development projects. arXiv preprint arXiv:1504.00942 (2015). https://doi.org/10.5121/ijcis.2015.5101
10. Silva, M.S.: GAIA Modelo de maturidade para aquisição de software. Universidade Estadual de Londrina, Paraná, Brazil (2016)
11. Ozkan, N., Tarhan, A., Kucuk, C.: SCRUM at Scale in a COBIT Compliant Environment: The Case of Turkiye Finans IT, Ozkhan, Tarhane Kucuk (2017)
12. Gupta, S.: SOX compliant agile processes. In: AGILE 2008 Conference, Toronto, ON, pp. 140–143 (2008). https://doi.org/10.1109/Agile.2008.48
13. Mahnic, V., Zabkar, N.: Using COBIT indicators for measuring SCRUM-based software development. WSEAS Trans. Comput. **7**(10), 1605–1617 (2008)
14. Persson, J.S., Mathiassen, L., Aaen, I.: Agile distributed software development: enacting control through media and context. Inf. Syst. J. **22**(6), 411–433 (2012). https://doi.org/10.1111/j.1365-2575.2011.00390.x
15. Muszyńska, K.: Patterns of communication management in project teams. In: Ziemba, E. (ed.) AITM/ISM -2016. LNBIP, vol. 277, pp. 202–221. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53076-5_11
16. Concas, G., Marchesi, M., Destefanis, G., Tonelli, R.: An empirical study of software metrics for assessing the phases of an agile project. Int. J. Softw. Eng. Knowl. Eng. **22**(4), 525–548 (2012). https://doi.org/10.1142/S0218194012500131
17. ISACA: COBIT® 5 Supplementary Guide for the COBIT 5 Process Assessment Model (PAM), APMG International (2017)
18. Standish Group International Inc.: Extreme Chaos Report (2001). https://courses.cs.ut.ee/MTAT.03.243/2013_spring/uploads/Main/standish.pdf
19. McGovern, F.: Managing software projects with business-based requirements. IT Prof. **4**(5), 18–23 (2002). https://doi.org/10.1109/MITP.2002.1041174
20. Antón, A.I.: Successful software projects need requirements planning. IEEE Softw. **20**, 44–46 (2003). https://doi.org/10.1109/MS.2003.1196319
21. Standish Group International Inc.: The Chaos Manifesto (2012). https://cs.calvin.edu/courses/cs/262/kvlinden/resources/CHAOSManifesto2012.pdf
22. Hastie, S., Wojewoda, S.: Standish Group 2015 Chaos Report - Q&A with Jennifer Lynch (2015). https://www.infoq.com/articles/standish-chaos-2015
23. Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J., Thomas, D.: Manifesto for agile software development (2001). http://agilemanifesto.org. Accessed 21 Jan 2018
24. SCRUM STUDY: A Guide to the SCRUM Body of Knowledge - SBOK GUIDE. VMEdu, Inc., Phoenix (2013)
25. Larman, C.: Agile & Iterative Development: A Manager's Guide. Addison-Wesley Professional, Boston (2004). ISBN 0-13-111155-8
26. Hendrikse, Z.: Agile Awareness. http://it-essence.xs4all.nl/presentations/moduleAgile/
27. Irrazabal, E., Vásquez, F., Díaz, R., Garzás, J.: Applying ISO/IEC 12207:2008 with SCRUM and Agile Methods, Universidad Rey Juan Carlos, Madrid, España (2011)

28. Quality Assurance Mentor: PDCA Cycle. http://www.quality-assurance-mentor.com/software-quality-assurance.html
29. Reserva em revista: Ciclo PDCA. http://necs.preservaambiental.com/ciclo-pdca-abordagem-de-processo-e-escopo-do-sistema-de-gestao-ambiental/
30. Carvalho, M.M., Paladini, E.P.: Gestão da Qualidade. Editora Campus, Rio de Janeiro (2005)
31. Singh, R.: An Introduction to International Stangaiadard ISO/IEC – Software lifecycle process. Federal Aviation Administration, Washington, DC, EUA (1999)
32. IEEE: SWEBOK: a project of the IEEE Computer Society Professional Practices Committee (2004)
33. ISO – International Standards Organization: ISO/IEC 9001:2000 Quality Management Systems – Requirements. 3rd edn (2000)

Models of Data Quality

Zane Bicevska[1], Janis Bicevskis[2], and Ivo Oditis[1(✉)]

[1] SIA DIVI Grupa, Riga, Latvia
{zane.bicevska, ivo.oditis}@di.lv
[2] University of Latvia, Riga, Latvia
janis.bicevskis@lu.lv

Abstract. The research proposes a new approach to data quality management presenting three groups of DSL (Domain Specific Language). The first language group uses concept of data object in order to describe data to be analysed, the second group describes the requirements on data quality, and the third group describes data quality management process. The proposed approach deals with development of executable quality specifications for each kind of data objects. The specification can be executed step-by-step according to business process descriptions, ensuring the gradual accumulation of data in the database and data quality verification according to the specific use case.

Keywords: Data quality · Domain-specific modeling languages
Executable business processes

1 Introduction

The term "quality" depends highly on the context in which it is applied. The term is commonly used to indicate the superiority of a manufactured good or to attest a high degree of craftsmanship or artistry [1]. In manufacturing industries, quality is viewed as a desirable goal to be achieved through management of the production process.

Data quality is an IT-specific concept, and it can be defined as the degree to which the data fulfills requirements of characteristics [2]. Examples of data quality characteristics are: completeness, validity, accuracy, consistency, availability, and timeliness.

The data quality problem is topical since over 50 years, and many different approaches are discussed in scientific publications. In the major part of sources the central attention is paid to defining of data quality characteristics informally and measuring their values. Mechanisms for specifying of data quality characteristics in formalized languages usually are not considered. The main goal of this research is to provide data quality management mechanisms being able to execute data quality specifications which are defined using formalized domain specific language (DSL).

In order to achieve the goal, two key requirements for specifying the data quality were formulated. Firstly, the ISO 9001:2015 standard considers data quality as a relative concept, largely dependent on specific requirements resulting from the data usage. It means the same data can be of good quality for one usage and completely unusable for another. For instance, to determine a number of students in a high school, only the status of students is significant, no other data such as students' age or gender. The same

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 194–211, 2018.
https://doi.org/10.1007/978-3-319-77721-4_11

data may be checked for its compliance with different quality requirements. It should also be emphasized that many conditions and requirements can not be checked during the data input as they are dependent on values of other data objects that are not entered yet. For instance, at the time of student's enrollment not all information about his/her financial obligations is available and/or entered in the database. This is the reason why high-quality data in practice occurs rarely. Therefore, the following quality requirements are set for a data quality management system:

- Data quality requirements should be defined on several levels: for a separate data object, for a data object in terms of it's attributes, for a data object in the context of it's data base, for a objecy objektam datubāzes kontekstā, in the context of several information systems,
- Data quality requirements should be described by using a graphical DSL which is adaptable to the specific needs of each new information system; the language should be easy-to-use thus ensuring the involvement of business experts instead of IT staff,
- Data quality must be checked in many stages of information processing, each time using the individual description of data quality requirements; data quality requirements should not be consolidated into one comprehensive requirements specification for postponed data quality checking after the whole data set is entered into data base,
- To evaluate the data quality for the specific usage, the requirements for data must be described; the descriptions should be executable, as the stored data will be "scanned" and its' compliance to requirements will be checked.

The solution proposed in the paper meets the previously formulated requirements and has been described in detail in the following sections. The data quality model consists of graphical diagrams, where each diagram describes a certain part of the data quality check. All checks that have to be performed in one business process step are merged into packages, and a set of all packages forms data quality model. The diagrams consist of vertices and arcs. Vertices marked with mnemonic graphic symbols represent elementary actions of data quality management. The vertices are interconnected by arcs that indicate the sequence of actions to be performed. Such diagrams resemble the well-known flowcharts commonly used by both IT and business professionals; they are simple enough to be used not only by IT professionals.

The data quality model can be used on two levels: informal and executable. The informal data quality model contains textual descriptions of the checking activities to be performed, i.e., the diagram symbols contain textual descriptions of the activities. An informal model can be transformed into an executable model, replacing informal texts with program code, SQL sentences, or other executable objects.

The paper deals with the following issues: an overview about the related research (Sect. 2), a description of the proposed solution including an example of the data quality model (Sect. 3), research findings (Sect. 4), and analysis of the proposed solution (Sect. 5). This paper is a continuation of the paper included in the proceedings of the FedCSIS 2017 [3], expanding it by about 40%. The enhancements refer to description of the data quality model (Sect. 3), and analysis of the proposed solution (Sect. 5).

2 Theoretical Background

There are four main research branches present: (1) the total data quality management (TDQM) theory, (2) the data quality defining by using the Object Constraint Language (OCL), (3) the data quality management using SSIS tools, (4) the data quality management using DQS tools. They all are described in this section.

2.1 Total Data Quality Management

The issue of data quality is essential since the very beginning of the IT industry. Numerous studies have led to various definitions of data quality. For instance, data are of good quality if they satisfy the requirements imposed by the intended use [4].

Data are of high quality if they are fit for their intended uses in operations, decision making, and planning. According to Juran and Gryna [5] data is fit for use if they are free of defects (accessible, accurate, timely, complete) and possess desired properties (relevant, comprehensive, proper level of detail, easy to read, easy to interpret) [6].

Data quality can also be characterized by different dimensions. In 1996, Wang and Strong [7] defined 15 data quality dimensions which are confederated in four quality groups: intrinsic, contextual, representational, accessibility.

Redman [6] provides 51 data quality dimensions, arranged in 9 data quality groups. Such in-depth gradation of data quality dimensions may seem an overstatement, especially for practitioners. In 2013 the Data Management Association International UK Working Group possesses only 6 dimensions: Completeness, Uniqueness, Timeliness, Validity, Accuracy, and Consistency.

2.2 Object Constraint Language

The OCL started as a complement of the UML notation with the goal to overcome the limitations of UML in terms of precisely specifying detailed aspects of a system design [8]. Since then, OCL has become a key component of any model-driven engineering (MDE) technique as the default language for expressing all kinds of (meta) model query, manipulation and specification requirements.

Constraints at the model level set conditions that the “data” of the system must satisfy at runtime. Therefore, the implementation of a system must guarantee that all operations that modify the system state will keep the data in a consistent state (a state that evaluates to true all model invariants). Clearly, the best way to achieve this goal is by providing code-generation techniques that take the OCL constraints and produce the appropriate checking code in the target platform where the system is going to be executed.

Typically, OCL expressions are translated into code either as database triggers or as part of the method bodies in the classes corresponding to the constraint context types. Roughly, in the database strategy each invariant is translated as a SQL SELECT expression that returns a value if the data does not satisfy that given constraint. This SELECT expression is called inside the body of a trigger so that if the SELECT returns a non-empty value then the trigger raises an exception. Triggers are fired after

every change on the data to make sure that the system is always in a consistent state. The OCL has many positive qualities:

- OCL is an extension of UML, and it has gained a wide popularity in the computer scientists' community,
- OCL provides a rich range of means of expression, allowing the use of widely used programming constructions.

At the same time the disadvantages of OCL should also be recognized:

- OCL is a declarative language without graphical notation,
- no data read/write operations,
- constraints of OCL are closely related with the data storage in a relational database, no operations for reading and checking of discrete data objects that are not related to database (such operations are necessary for verifying of data entered via screen forms,
- defining and understanding of OCL constraints requires deep knowledge and skills in object-oriented programming; it makes the OCL unsuitable for industry professionals without appropriate IT background.

OCL-based data quality solutions are hard to use practically due to the dynamic data input into database as well as to the complexity of OCL.

2.3 SQL Server Integration Services

As every solution, Microsoft SQL Server Integration Services (SSIS) has various advantages and disadvantages [9]. SSIS offers wide range of features for data migration, and designing of ETL and transformation processes [10]. To cover a broad spectrum of requirements for data migration and ETL processes, SSIS includes both standardized operations for many widely-used database management systems, and add-ons for different import/export formats, and opportunities for developers to use the programming environment Visual Studio.

Furthermore, SSIS is open platform allowing create and use external add-ons. Hence SSIS should be considered as a mature platform that is suitable not only for solving of ETL tasks but also for processing of emails, linear text files, XML files, and other operations. The rich range of included features enables creating of SSIS packages from predefined components or to develop them by programming.

Microsoft has designed this product to provide better approach towards data migration, manipulation and transformation. With the power to define the workflow of process and task, user can easily define how the process should flow and perform some task on different interval. It also provides color codification and real-time monitoring. The main advantages of SSIS are:

- SSIS can handle data from heterogeneous data sources,
- SSIS provides data transformation functionality,
- tightly integrated with Microsoft Visual Studio and Microsoft SQL Server,
- suitable for complex transformations, multi-step operations and structured exception handling.

SSIS has some disadvantages:

- to see package execution report needs Management Studio rather than being published to reporting services,
- SSIS memory usage is high and it conflicts with SQL.

Sarjen [11] assures that usage of SSIS removes need of hardcore programmers as SSIS is apparently easy to understand and manage. In contradiction to [11] the authors of this research believe that the usage of SSIS have some fundamental barriers. The complexity of the approach is high; the usage of the solution for data processing and data quality management require either programmer's level of understanding of process execution, or many years of experience with SSIS.

Although not designed specifically for data quality management, features offered by SSIS provide a number of suitable solutions. Currently there are not known SSIS uses for data quality management which were not related to data migration. However, data quality management elements offered by SSIS are practically usable and should be taken over in further data quality solutions.

2.4 Data Quality Services (DQS)

DQS is one of Microsoft SQL Server components that enables maintaining the quality of data and ensuring suitability of the data for its business usage. DQS is a knowledge-driven solution that provides both computer-assisted and interactive ways to manage the integrity and quality of your data sources.

MS SQL Server 2012 standard edition provided rather limited DQS features. This paper deals with the latest currently available MS SQL Server and DQS version MS SQL Server 2017 [12]. DQS provides the following four functions to improve data quality:

1. DQS Knowledge Bases. To cleanse data, you have to have knowledge about the data. To prepare knowledge for a data quality project, you build and maintain a knowledge base (KB) that DQS can use to identify incorrect or invalid data. DQS enables to use both computer-assisted and interactive processes to create, build, and update knowledge base.
2. Data Matching. DQS data matching process enables to reduce data duplication and improve data accuracy in a data source. Matching analyzes the degree of duplication in all records of a single data source, returning weighted probabilities of a match between each set of records compared. User can then decide which records are matches and take the appropriate action on the source data.
3. Data Cleansing. Data cleansing is the process of analyzing the quality of data in a data source, manually approving/rejecting the suggestions by the system, and thereby making changes to the data. Data cleansing in Data Quality Services (DQS) includes a computer-assisted process that analyzes how data conforms to the knowledge in a knowledge base, and an interactive process that enables the data steward to review and modify computer-assisted process results to ensure that the data cleansing is exactly as they want to be done. Identifies incomplete or incorrect

data in your data source (Excel file or SQL Server database), and then corrects or alerts you about the invalid data and provides two-step process to cleanse the data:

- The computer-assisted process uses the knowledge in a DQS knowledge base to automatically process the data, and suggest replacements/corrections. The data processed by DQS is split into 5 groups: suggested, new, invalid, corrected, correct). The result is displayed to the user for decision making and further processing of the data.
- The interactive process allows the data steward to approve, reject, or modify the changes proposed by the DQS during the computer-assisted cleansing (manually).

4. Data Profiling. Data profiling has two major goals: first, to guide you through data quality processes and support your decisions, and second, to assess the effectiveness of the processes. It provides with automated measurements of data quality. It is dynamic and adjustable.

DQS is as a tool for analysing and improving the quality of data stored in MS SQL Server. It provides users with many necessary features. However, the features available in MS SQL Server 2017 should only be seen as the first step towards creating data quality solutions due to the following DQS shortcomings:

- The list of available data formats for domains is incomplete, since only 17 formats are offered. Moreover, none of the formats support date entries starting with the year, e.g. "yyyy-mm-dd" and "yyyy-dd-mm" entries are not possible.
- Checking of the data stored in Excel tables is only supported by the Microsoft Excel 2003 version, however, most users are using newer versions. As a result, the DQS tool will be hardly used for analysing Excel data.
- In consistency checks the minimum consistency threshold is 80%, which may not always be sufficient for identification and prevention of data quality issues.
- Analysis can be performed only to fields in one table. Simultaneous processing of multiple tables is possible by creating table views containing merged tables. Creation of such table views require proficient knowledge of databases.
- DQS supports analysing of relatively large tables. As supported by experiments the analysis of a database containing a million records requires 40–50 min (CPU, Disk and Memory load up to 100%). Consequently, DQS is not recommended for analysis of large amounts of data.

3 Proposed Data Quality Methodology

3.1 Data Quality System

This section is dedicated to the three main components of data quality system: data object, quality requirements' description, and process of data quality measuring. All three components together constitute the data quality specification, where the description of the data object defines the data to be analysed, the description of the quality requirements defines the conditions to be met to call the data qualitative, and the description of the quality measurement process defines procedure to be followed in order to assess the data quality. All three components will be graphically represented.

The data quality management process (Fig. 1) begins with the retrieval of verifiable data from data sources. This operation can be accomplished using the SQL Server Integration Services (SSIS) tool mentioned in the previous section or its analogue ETL tools. The operation involves data retrieval from many sources with different data formats, supporting data filtering and format transformation. The data retrieval result is recorded into the data object (the exact description of the data object is provided in the next section).

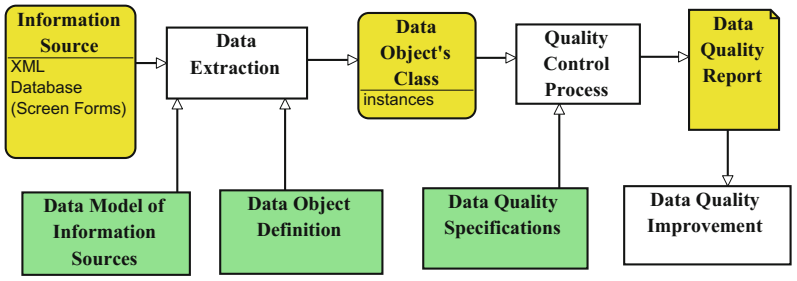

Fig. 1. Architecture of data quality system.

Following the data retrieval, the data object quality specification is prepared. The data object quality specification contains the conditions to be met by correct data. The quality specification conditions are defined using logical expressions and field names of the data object. Thereafter, the quality verification process is carried out including verification whether the retrieved data meet the quality requirements. The results of verification process are reported in order to improve the data quality. Reporting can be performed using the Data Quality Services (DQS) tool described in the previous section. The quality of the data quality architecture is similar to that offered by Becker et al. [13].

3.2 Data Objects

Separate Data Object

Traditionally the notion of a data object is understood as the set of values of the parameters that characterize a real-life object. For instance, the entity Developers has the following attributes (Fig. 3), Dev_ID (the developer to whom the spent time should be referred), Dev_name (developer's name), Dev_surname, Dev_load (the minimum monthly developer's workload).

The quality checking of one of the data object parameters values is reduced to an examination of the properties of the individual values, for instance, whether a text string may serve as a value of the field Dev_name, or value of the field Dev_load vērtība is numerical and potentially credible. Anyway, the checking of parameter values is local and formal process. It does not respect contextual interlinks with other data objects and does not check the compliance of data with the true characteristics of a natural person.

In general, the quality checking of a data object may include not only compliance with the type of particular parameter values but also checking of more complex conditions, for example, checking of digital check code for self-checking correctness.

A specific quality control of a particular data object is a typical component for the input data quality control in every information system. Data is usually entered into an information system by filling in screen form fields, followed by an information quality check and its retention in the database. In cases when the input fields are not filled correctly, the user usually receives an error message and is allowed to adjust the input data. Similarly, an information about a particular data object, such as a person, can be requested from another information system using web services. In this case, the message received contains information about the requested data object, and it is advisable to check the data quality of the information before it's usage.

Data Object's Classes

Information systems deal not only with certain data objects, but also process many data objects in a unified way. In this case, the classes of data objects are used, which represent many objects of the same structure. A data object class has a name, and its elements have the same structure as they all are characterized by the same parameters. Each specific data object may contain parameter values fully or partially.

Data objects are described by the class diagram in Fig. 2. The data objects class consists of many specific data objects – instances. The data object class structure consists of arbitrary number of fields and arbitrary number of other data object classes. Thereby the data objects class forms a tree structure.

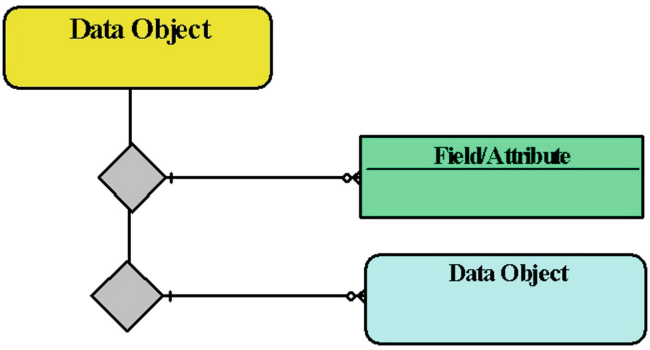

Fig. 2. Meta-model of data object's classes.

A document-oriented database can serve as an example of a data object class. It contains documents of a uniform structure where fields can be filled in partially or completely for each specific data object. If relational data bases are used values of data object's parameters can be arranged in different data base tables. In such a case, actions similar to transformation of the relational database into a document-driven database must be performed in order to retrieve objects of one class.

Data object classes are established in order to enable definition of data quality requirements for data object collections. For example the number of erroneous records in the database containing data about persons may not exceed 1% of the total number of records in the data base. This dimension of quality is measurable by checking all parameter values for all records and dividing the number of erroneous records by the total number of records. Such operation is understandable and easy to implement in document databases (including XML files) when analyzing all the records stored in the database in a row. Much more opaque is performance of this operation in the relational database because the values of the parameters can be "scattered" across many tables and it is necessary to retrieve the information about the values of one specific data object's parameters from many tables. Therefore, the data object and the data object classes have been selected as the data quality evaluation object in this study since many of the popular dimensions of data quality can be clearly defined on this basis and the corresponding operations by which they will be measured can be easily understood.

Example of Data Object's Classes

The proposed solution will be based on a simplified example of an information system. Let us consider a working time tracking (WTT) system having the ER model given in the Fig. 3. There are many active projects in an enterprise (entity Projects); every project has several employees (entity Developers); each employee (developer) may be involved in several projects; the working time spent by an employee (developer) in a specific time frame is aligned to one specific project (entity Work_time).

The entity Projects has the following attributes: Proj_ID (project identifier to specify the project to which the spent working time should be referred), Proj_name (project name), Proj_volume (the estimated work amount of the project in man-hours), Proj_start_date, Proj_end_date, Proj_limits (the maximum allowable work amount of the project in man-hours), Proj_actual (project is active/passive), Proj_leader_ID (project manager).

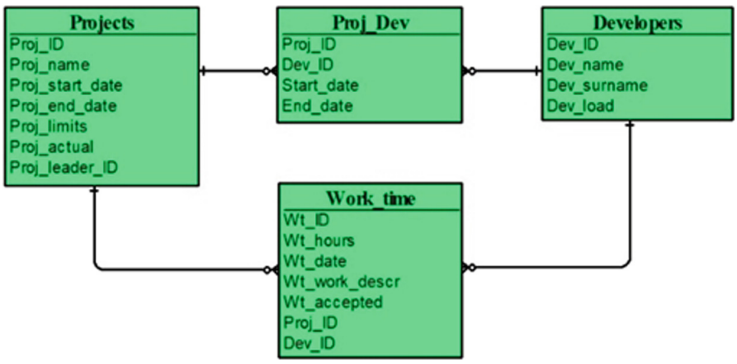

Fig. 3. WTT data model.

The entity Developers has the following attributes: Dev_ID (the developer to whom the spent time should be referred), Dev_name (developer's name), Dev_surname, Dev_load (the minimum monthly developer's workload).

The entity Work_time has the following attributes: Wt_ID (identifier of the spent working time record), Wt_hours (spent working time of the developer), Wt_date (date of the spent working time), Wt_work_descr (description of the performed work), Wt_accept (reported working time is accepted by the project manager, Yes/No), Proj_ID (the project to which the time should be referred), Dev_ID (the developer to whom the time should be referred).

The entity Proj_Dev_time is a junction table for dealing with many-to-many relationships, and it has the following attributes: Proj_ID (the project where the Dev_ID works), Dev_ID (the developer working in the project Proj_ID), Start_date (the date when the developer Dev_ID started to work in the project Proj_ID), End_date (the date by which the Dev_ID will be assigned to the Proj_ID).

3.3 Quality Specification

Data quality requirements for a data object are defined by using logical expressions. The names of data object's attributes/fields serve as operands in the logical expressions; as operations may be used the traditional means of programming languages.

Data Quality Requirements for a Separate Data Object

Let us assume, the developers prepare reports about their working time autonomously and send the reports to data base where all enterprise data from various sources is collected. The procedure receives values of attributes:

$$< \text{Proj\_ID}, \text{Dev\_ID}, \text{Wt\_date}, \text{Wt\_hours}, \text{Wt\_work\_descr} >$$

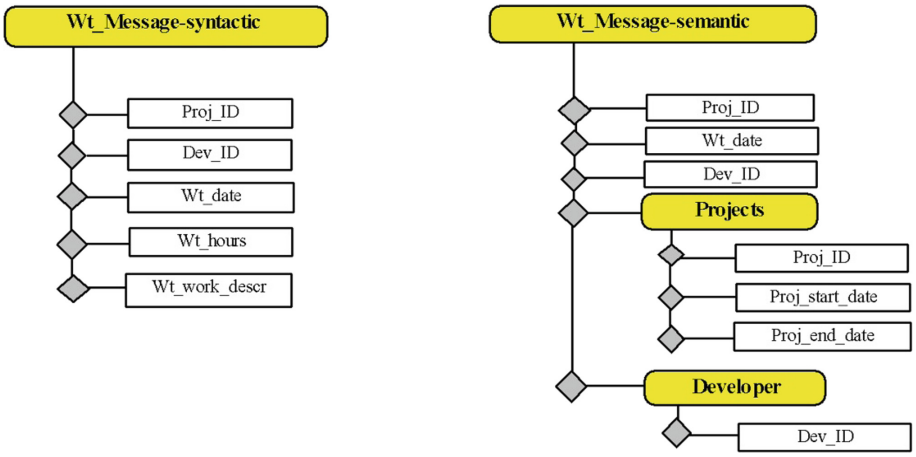

Fig. 4. Data objects Wt_Message-syntactic and Wt_Message-semantic.

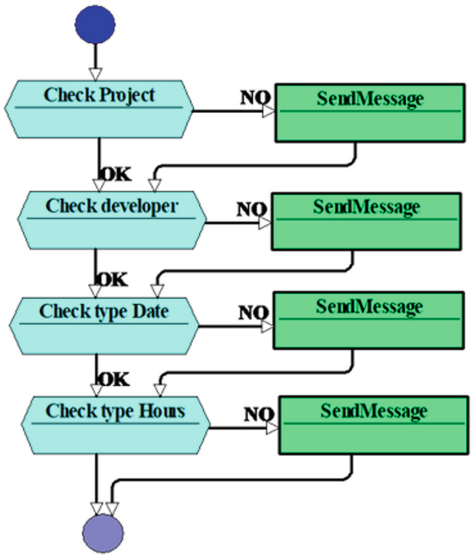

Fig. 5. Example of data object's syntactic control.

First, syntactic control of the data object Wt_Message-syntactic (Fig. 4) will be performed in order to assess the quality of the message to be entered in the database. The quality specification of report shown in the Fig. 5 ensures quality control within one input message: (1) Are all mandatory fields completed (Proj_ID, Dev_ID)? (2) Have input values correct data types (Wt_date, Wt_hours)?

In order to make the quality specification executable, informal texts should be replaced by program routines executing the desired operations.

Contextual Control on Interrelated Data

Contextual control on interrelated data (Fig. 6) ensures quality control using attribute values of mutually interconnected data objects (Fig. 4): (1) Does the message contain object instances with references to other data objects (Project exists, Developer exists)? (2) Are the attribute values of input data in compliance with related data objects?

In order to make the quality specification executable, informal texts should be replaced by SQL statements for data retrieving and control of constraints (Fig. 7). The SQL statements enable defining of data objects using SELECT and defining of data quality requirements using WHERE conditions. Tools like SSIS may be used – these also offer statements for execution of SQL statements and validation of results. However, other notations or languages could also be used instead of SQL statements if they support executable correctness checking for objects of a particular class.

Contextual Control on the Database

Contextual control on the database (Fig. 8) checks the compliance with conditions valid for the whole data base (examples: isn't the maximum of work amount allowed for the project exceeded, do the reports of employee cover the minimal workload of the employee in the time period, etc.).

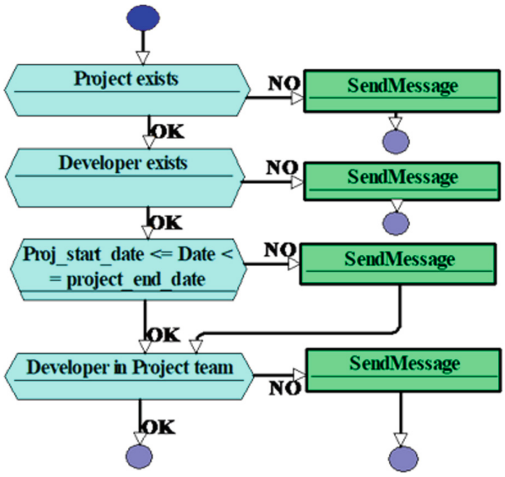

Fig. 6. Example of contextual control on interrelated data

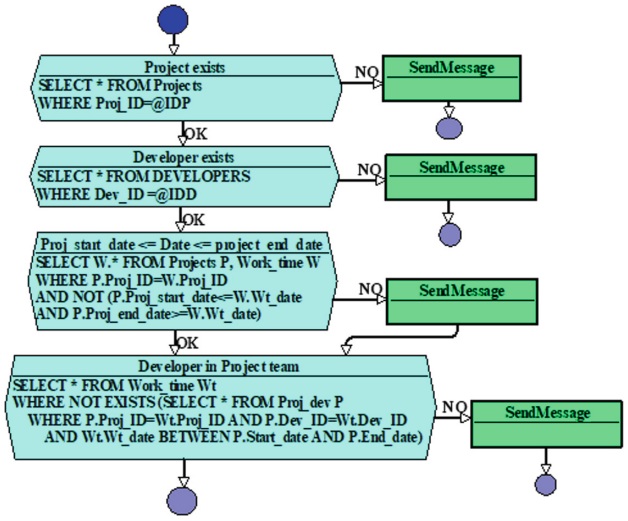

Fig. 7. Example of executable contextual control on interrelated data

In order to make the quality specification executable, informal texts should be replaced by SQL statements for data retrieving and control of constraints (Fig. 9).

The given examples have shown how to define data quality requirements in different levels – starting with a separate message up to quality requirements on the entire data base. Data quality requirements are described by graphic diagrams. The proposed data quality model is a platform independent model (PIM), like UML diagrams serve as a PIM model in the MDA architecture.

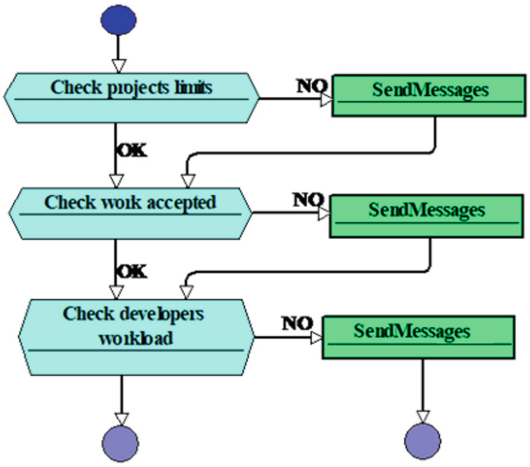

Fig. 8. Example of contextual control on the database

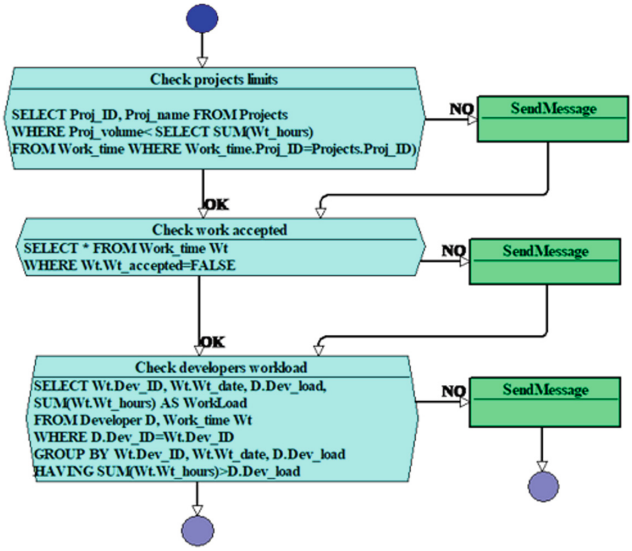

Fig. 9. Example of executable contextual control on the database

Diagrams in Figs. 5, 6 and 8 form an informal data quality specification of working time tracking (WTT) system. It may be useful for industry experts to describe data quality requirements since a more formalized specification of executable controls is not practically applicable without IT skills. Executable data quality specifications can be used in business process steps to check data quality in certain points within processes.

Practical uses of the proposed approach have shown advantages of graphically represented data quality specification as they were more effective in discovering data quality errors caused by information system than the traditionally used informal data quality specifications in the textual form. It reaffirmed advantages of graphic diagrams in comparision to natural language texts in standardized documents.

Additional advantages can be achieved if the data quality specifications are transformed to executable specifications. Although additional programming is needed to ensure the executability, much higher level of data quality can be achieved in an information system as a whole if data quality controls are incorporated in business process steps.

First two of mentioned data controls typically are applied during data input. In case data should be saved anyway it is marked as incorrect. As this one control over database could be rather resource-intensive (time, server memory, processor time, data locking) it can not be executed on every data manipulation. Contextual control on the database usually is executed out of business hours and even not every day. Still the proposed approach is universal, and it is applicable in different cases – during the initial data input in information system, migrating data from one information system to another, performing data transformation to data warehouse.

Contextual Control on Several Databases

The above described approach is also applicable in cases when several information systems in different enterprises are involved. Such case is typical for public institutions with different but interrelated state information systems.

This problem has been addressed in Latvia since 2000. The essential data of public interest are accumulated in different state information systems: Population Register, Business Register, Vehicle register, etc. Each of the registers is managed and maintained by some public body which is responsible for the quality of the accumulated data.

The registers should also mutually exchange data; usually it is organized with the help of web services serving and receiving data – concrete values of data objects' attributes. Each data exchange session may require only few attribute values. When using data quality specifications, it is possible to check and evaluate the quality of received data.

Like the Latvian Integrated State Information System project [14], the described problem is also addressed in Estonia [15], Lithuania [16] etc.

Currently development of various industry-specific state information systems is continuing, and the identified data quality problem persists in each system again and again.

3.4 Quality Verification Process

A specific data object or a class of data objects is used as an input for a quality verification process. In the case of a specific data object, the information is entered into the data object fields and subsequently the conditions describing the data object's quality are verified. In the case of data object classes, instance values of the data object class are selected from the database or from the XML file, then all instances of the

accumulated collection are inspected by verifying quality conditions of each instance. The quality verification process creates a test protocol which identifies data objects nonconforming the quality requirements. SSIS and DQS tools provide tools for data retrieving from information sources and building of test protocols therefore it will not be further discussed here.

A quality check of a particular data object occurs when data on a single physical entity is entered into the information system. The field values are passed on to the quality check and the conditions of the specification are verified. If non-compliance with quality requirements has been detected, the process registers the case to let analyzing the causes of discrepancies and eliminating of them.

The quality analysis process description language may include all verification activities of a particular data object. Additionally, loop constructions may be necessary to process class instances, similarly to processing of collections used in C# (Iterator). This can be required when the main data object class contains subordinated data object classes whose data quality check is included in the verification of the main data object class.

4 Research Findings

Theory of total data quality management (TDQM) describes main principles of data quality and methods of its evaluation. This paper proposes an implementation mechanism of this theoretical methodology using executable data quality models. This was the main goal of the research. Three language families are suitable for creating of executable data quality models:

- Data object defining languages allow defining objects whose quality will be analysed. Data object values may be retrieved from different sources of information making them available to assess the data quality.
- Requirements defining languages allow defining requirements for data object's quality. The data quality requirements should be formulated for several levels of a data object – discrete data object, contextual control on interrelated data, contextual control on the database.
- The quality analysis process description languages define the process of quality evaluation. It can be made as an informal process description, e.g. the data validations could be described in a textual way, or as an automatically executable process where the informal description is substituted by a source code or SQL statements making the informal model executable.

All three language families are consistently designed as graphic languages. According to the MDA ideology [17], the data object and the data quality specification build a platform independent model (PIM) but the data quality evaluation model builds a platform specific model (PSM).

5 Discussion of Findings

5.1 Usage of Data Quality Models

In practice the proposed data quality verification technology is primarily suitable for information systems using relational databases. As the data base structure (ER model) is relatively constant, the data quality model may be used in very effective way. Before the use of the data quality model is executed and all discrepancies with requirements are identified. The user may decide to improve data quality or to use the data as it is. The proposed technology can also be used for the control of input data in very complex screen forms with many interrelated or conditionally linked records. For example, a number of input fields in customs declaration documents may exceed 50 in one screen form. The created data quality model may serve as a precise specification for software development and a testing model for the created software.

A similar situation arises by importing data in data warehouses. Since various mutually unrelated databases may serve as data sources, the compatibility and compliance with requirements of the data sources should be checked before data importing into the data warehouse. The data quality model allows defining compliance requirements precisely and in detail.

The proposed technology is applicable not only for relational databases but also for document-oriented NoSQL databases with XML documents. The main difference: in case of document-oriented database the requirements will be described by XQuery statements (instead of SQL statements), and addressing to the data objects attributes will not be processed via column names of the tables in the ER model.

5.2 Implementation of Data Quality Models

The data quality model, like business processes, is described using some graphical DSL. Since the data quality requirements are different, the used DSLs can also be different. Hence it is advisable to use not only one specific graphical editor supporting one specific DSL but to create your own editor for each used DSL. Currently there are several such platforms in use; one of them, called DIMOD was derived from GrTP [18]. The business process modeling environment DIMOD is intended to:

- Define the DSL using meta-model that is stored into model's repository. DSL parameters may be defined and modified using the separate configuration component Configurator. Once the DSL is defined the corresponding modeling editor is created for the DSL automatically.
- Create and edit data quality diagrams in the DSL. This is usually done by some highly qualified modeling experts in collaboration with domain experts ("clever users").
- Check the data quality models' internal consistency. Both IT and domain experts are involved in it; to publish the created model/diagrams in WEB. It allows a wide range of users to use the data quality model diagrams.

6 Conclusion

The research shows the relative and dynamic nature of data quality. The usage of data implies requirements for data quality; the data are accumulated and verified step-by-step. The research proposes a data quality management solution which meets the requirements. It complements the general guidelines of TDQM with a constructive approach to the data quality management offering user-friendly graphical DSL.

The proposed approach is practically applicable and feasible for different levels – a discrete data object, interrelated data objects, data in a database, data in several databases. Data quality requirements should be specified in an easy-to-understand definition language to ensure that industry experts will be able to formulate data quality requirements with minimal involvement of IT professionals.

The paper is a continuation of research in the area of executable models and DSL [19–23]. The proposed approach and tools for designing of executable data quality specifications in different levels let to design, develop and use the specifications as steps in executable business processes.

The research has managed to separate the data quality PIM model, which consists of the data object and data quality requirements, from the PSM model, which consists of a description of the data quality control process. This is a new and innovative approach in the data quality research. Unfortunately, due to the size limit, the PSP model is only shown on the relational databases. In the future similar results could be achieved looking at data quality problems for document-oriented databases [24] and for data stored in XML and JSON files.

For future research, two directions are of particular interest. The first direction deals with analysing quality models for a variety of domains, for example, in state information systems, medicine, banking, and others; and the second deals with finding optimal implementation of data quality model for different types of data storage, e.g., relationship database, document-oriented database, and other.

Acknowledgments. The research leading to these results has received funding from the research project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/16/A/007 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 1.8 "Data Quality Management by using Executable Business Process Models".

References

1. Veregin, H.: Data quality parameters. In: Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (eds.) New Developments in Geographical Information Systems: Principles, Techniques, Management and Applications, pp. 177–189. Wiley, Hoboken (2005)
2. ISO 9001:2015: Quality management principles. https://www.iso.org/standard/62085.html
3. Bicevska, Z., Bicevskis, J., Oditis, I.: Domain-specific characteristics of data quality. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of the 2017 Federated Conference on Computer Science and Information Systems. ACSIS, vol. 11, pp. 999–1003 (2017). https://doi.org/10.15439/2017f279

4. Olson, J.E.: Data Quality. The Accuracy dimension. Morgan Kaufmann Publishers, Burlington (2003)
5. Juran, J.M., Gryna, F.M.: Juran's Quality Control Handbook, 4th ed (1988)
6. Redman, T.C.: Data Quality. The Field Guide. Digital Press, Boston (2001)
7. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. **12**(4), 5–33 (2016). https://doi.org/10.1080/07421222.1996.11518099
8. OCL 2.0: Object Constraint Language™. Version 2.0. http://www.omg.org/spec/OCL/2.0/
9. SQL Server Integration Services-SSIS. https://www.codeproject.com/Articles/155829/SQL-Server-Integration-Services-SSIS-Part-Basics
10. Features Supported by the Editions of SQL Server 2014. Microsoft Developer Network. http://msdn.microsoft.com
11. What is SSIS? Its advantages and disadvantages. Microsoft Practices. http://www.sarjen.com/ssis-advantages-disadvantages/
12. Microsoft. Data Quality Services Features and Tasks. https://docs.microsoft.com/en-us/sql/data-quality-services/data-quality-services-features-and-tasks
13. Becker, D., Jaster, J., Kuperman, J.: Flexible and generic data quality metadata exchange. In: 12th International Conference on Information Quality, Cambridge, MA (2007)
14. The State Information System Register. http://www.varam.gov.lv/eng/darbibas_veidi/e_gov/?doc=13052
15. Administration system for the state information system RIHA. https://www.ria.ee/en/administration-system-of-the-state-information-system.html
16. State registers and information systems. https://ivpk.lrv.lt/en/activities/state-registers-and-information-systems
17. Model driven architecture. http://www.omg.org/mda/
18. Bārzdiņš, J., Zariņš, A., Čerāns, K., Kalniņš, A., Rencis, E., Lāce, L., Liepiņš, R., Sproģis, A.: GrTP: transformation based graphical tool building platform. In: The 10th International Conference on Model-Driven Engineering Languages and Systems, Models (2007)
19. Bicevskis, J., Bicevska, Z.: Business process models and information system usability. Procedia Comput. Sci. **77**, 72–79 (2015). https://doi.org/10.1016/j.procs.2015.12.361
20. Ceriņa-Bērziņa, J., Bičevskis, J., Karnītis, Ģ.: Information systems development based on visual domain specific language BiLingva. In: Szmuc, T., Szpyrka, M., Zendulka, J. (eds.) CEE-SET 2009. LNCS, vol. 7054, pp. 124–135. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28038-2_10
21. Bicevska, Z., Bicevskis, J., Karnitis, G.: Models of event driven systems. In: Arnicans, G., Arnicane, V., Borzovs, J., Niedrite, L. (eds.) DB&IS 2016. CCIS, vol. 615. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40180-5_6
22. Ziemba, E. (ed.): Information Technology for Management. LNBIP, vol. 243. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8
23. Łobaziewicz, M.: The role of ICT solutions in the intelligent enterprise performance. In: Ziemba, E. (ed.) AITM/ISM 2016. LNBIP, vol. 277, pp. 120–136. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53076-5_7
24. Document Databases—MongoDB. https://www.mongodb.com/

Selecting Project Communication Management Software Using the Weighted Regularized Hasse Method

Karolina Muszyńska(✉) and Jakub Swacha

Faculty of Economics and Management,
University of Szczecin, Szczecin, Poland
{karolina.muszynska, jakub.swacha}@usz.edu.pl

Abstract. Successful realization of projects in dispersed, multinational teams is a difficult task, because of the communication and collaboration issues which may arise due to scarce personal contact, linguistic problems or less influential leadership. In such an environment, it is particularly important to assist the management of project communication processes with appropriate ICT tools. However, due to the huge number of software products aimed at supporting these processes which are available on the market, it is a challenge to choose the best one for a given project and team. In this paper, a four-stage procedure based on the weighted regularized Hasse matrix is proposed as a way of solving the problem of selecting the most suitable project communication management software for a project having specific requirements.

The proposed procedure involves comparing and ranking candidate tools taking into consideration the importance of multiple functional and non-functional features of the project communication management systems with their respective weights based on the results of a questionnaire conducted among members of a dispersed international project team. As illustrated by the provided example, the proposed procedure effectively leads to a complete ranking of candidate tools without the necessity to perform effort-demanding pairwise comparisons of alternatives or complicated calculations required by other methods (such as AHP or outranking methods). For this reason, the presented findings have vast implications for practice, introducing a convenient way of solving the project communication management software selection problem in the case of many candidate solutions and multiple criteria having distinct weights.

Keywords: Software selection · Weighted regularized Hasse method
Project communication management support

1 Introduction

1.1 Motivation

Realization of a project in a dispersed international team is a complicated and demanding task. While, as in other settings, it is vital to follow a project management methodology or a set of best practices to manage its different areas including: scope,

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 212–228, 2018.
https://doi.org/10.1007/978-3-319-77721-4_12

schedule, costs, quality and human resources in any project, yet in dispersed, multinational teams, additional communication and collaboration issues may arise, as a result of linguistic problems, limited trust, scarce direct contact among team members and less influential leadership.

Different research studies show that geographic dispersion may impede effective information sharing, coordination, problem solving, building trust, and constructively resolving conflicts with others on the team [1–4]. Project delivery risks with distributed teams tend to be greater when compared to co-located teams [5]. This is mainly due to the lack or high limitation of face-to-face contact, which hinders interpersonal relations, trust and commitment and causes misunderstandings. Another impeding factor is the fact that project team members often participate simultaneously in many projects what makes work coordination even more complicated.

The physical distance in dispersed multinational project teams augments the importance of the information and communication technologies (ICT) through which most communication and the building of relationships is performed. Thus, the ICT support, along with rules of communication, project management style, goal-setting and managers' competences, becomes one of their critical success conditions [6].

The current trend in project management is to find technology that allows the creation of a professional environment for dispersed teams, similar to the one expected if these teams were co-located [7]. An important role here is played by project management software which can significantly help to better manage projects, reduce costs, shorten schedules and be more responsive to customer demands. This is a broad category of software, covering areas such as scheduling, cost control, budget management, resource allocation, collaboration, communication, quality management and documentation [8].

The project management and communication software may differ in many aspects, e.g. offered features, complexity of handled information, ease of use, or price per user. It is obvious, that in order for the software to help solve the problems of project realization rather than create new problems of its own, it should be well suited to the specific project, taking into consideration its scope, size, management model, workflows and, last but not least, users' expectations. It has been observed that selecting the right tool has a significant effect on the success of the project and effectiveness of teamwork [9].

With tens of available project management suites, and tens of criteria to consider, the choice of the right tool is not trivial. Moreover, such choice should be repeated for every project undertaken, as the specificity of a given project may render unusable a tool used successfully in multiple other projects.

The problem addressed in this paper can therefore be defined as the selection of project management and communication software with the following conditions:

- there are many candidate solutions,
- there are multiple selection criteria having distinct weights,
- the procedure has to be easy-to-use.

By easy-to-use we mean a procedure characterized by simple input data gathering, uncomplicated calculations, and offering highly informative and highly readable form of results. The condition of being easy-to-use follows from the fact that the procedure

will be repeated often, by various people who will most probably have no scientific background, and who will have limited time to learn and apply it.

1.2 Problem Setting

It is a significant challenge to analyze data and make a decision taking into account many different aspects and criteria. The fact that many different indicators must be included simultaneously means that the so-called multi-indicator system or multi-criteria analysis must be used [10].

One way to handle a multi-indicator system is a mathematical mapping of the single indicator values to get a one-dimensional scalar, eventually to be used as the ranking indicator [11]. However such a mapping process, e.g. by using a weighted sum, hides all background information and may also cause unwanted compensation effects [12].

There are several popular methods that could be used to obtain a linear order from a multivariate data matrix, such as AHP, PROMETHEE, ELECTRE family methods and others – their characteristics and application examples are described in the Literature review section.

These methods have been applied to various types of software selection problems, such as selection of multimedia authorizing system [13], ERP system [14–16], SaaS product [17], software supporting design and manufacturing processes [18], simulation software [19] or learning management system [20]. One of these methods (AHP) has also been applied to the project management software selection problem [8, 21, 22].

Careful examination of the solutions proposed in the literature exposes research gap consisting in the lack of an easy-to-use procedure for evaluating and ranking software products, taking into account many candidate solutions and multiple criteria having distinct weights.

All the above-mentioned methods are complicated, and the particular method which was applied to the project management software selection problem in the prior works, AHP, is practically infeasible to solve the stated problem with the defined conditions, because not only the time to make the pairwise comparisons grows with the square of the number of considered solutions (which does not fare well when there are many of them), but it also requires the experts to be able to compare the alternatives (i.e. have experience in using them [17]), whereas it is highly unrealistic to expect the experts to have experience with tens of different project management suites, not to mention to be able to compare their various respective properties among each other using a 1 to 9 integer scale.

1.3 Approach

A less known yet simpler and thus attractive alternative to the methods described in the previous section is the partial order method. It allows not only to rank objects but also to obtain information to what extent a given object is better than another.

The partial order method has been used in many different research studies in environmental sciences, chemical industry, poverty analysis and many others [23]. It has also been successfully applied to software selection problem in the case of digital assets management systems [24].

This approach has, however, a significant weakness. In the case of problems with many criteria, such as the one researched here, often a large number of incomparabilities are observed which leads to a less meaningful representation. Moreover, the original Hasse method considers all criteria as equally relevant in determining the final data structure, and that is not always desired. A comprehensive solution to both these shortcomings has been proposed by Grisoni et al. [25] in the form of the weighted regularized Hasse method. It is this improved method that has been chosen to solve the discussed problem.

The aim of this paper is therefore to define an easy-to-use procedure for evaluating and ranking project management and communication software, taking into account many candidate solutions and multiple criteria having distinct weights, using the weighted regularized Hasse method [25]. As a proof of concept, this procedure is applied to evaluate and rank project management systems with regard to twenty five criteria, with their weights defined by representatives of seventeen partners from eight countries forming the case international project consortium. Due to its territorial dispersion and the multiplicity of communication channels among the project team members, the project represents a valid exemplification of a situation in which a deliberately selected project management and communication system is needed to ensure effective and successful project realization.

The paper is structured as follows. Section 2 reviews literature on alternative decision support methods, analyzing their strengths and weaknesses regarding the software selection problem addressed in this paper. Section 3 provides an insight on the chosen approach, introducing the concepts of partial ordering and Hasse diagrams. It also includes the basic information on the project selected for the exemplification and a description of the applied selection procedure. Section 4 presents the research results, which are discussed in Sect. 5. The final section concludes.

2 Literature Review

Multi-criteria decision analysis (MCDA) methods have been developed to support decision makers in their unique and personal decision process, as they incorporate preference information provided by them. There are four main types of decision that may be identified: the choice problem, the sorting problem, the ranking problem and the description problem [26]. The problem addressed in this paper is mainly a choice problem, although it can also be considered as a ranking problem.

For each type of decision, there are MCDA methods which are acknowledged as most suitable. In case of the choice or the ranking problem the suggested methods include: AHP, ANP, MAUT/UTA, MACBETH, PROMETHEE, ELECTRE I or III and TOPSIS, as well as combinations of these methods [26].

Each of these methods has its limitations, assumptions and characteristics, as well as the input information and the modelling effort required. It is also important to define from the very beginning the type of output needed. And so, for example, the MAUT is recommended if the 'utility function' for each criterion is known. AHP and MACBETH support the approach of pairwise comparisons between criteria and options. The difference is that comparisons are evaluated on a ratio scale for AHP and on an interval

scale for MACBETH. The drawback is that a large quantity of information is needed. Another alternative way is to define key parameters. PROMETHEE only requires indifference and preference thresholds, whilst ELECTRE requires indifference, preference and veto thresholds. If the user has only the ideal and anti-ideal options as input, the TOPSIS method is advisable. If criteria are dependent on each other, ANP can be used.

The modelling effort generally defines the richness of the output. One advantage to defining utility functions is that the options of the decision problem have a global score, and based on it, is possible to compare all options and rank them from the best to the worst, with equal rankings permitted – this is defined as a complete ranking. Outranking methods are based on pairwise comparisons. The preference or outranking degree reflects how much better one option is than another. It is possible for some options to be incomparable. These incomparabilities mean that a complete ranking is not always possible, which is referred to as a partial ranking [26].

In the literature, there are numerous examples of application of the above mentioned methods in a scope of problems similar to the one considered in this paper. Lai et al. used AHP to support selection of a multimedia authorizing system in a group decision environment. Three alternative systems were evaluated [13]. Also Wie et al. used AHP-based approach for selecting a suitable ERP system. In the case described there, as much as 24 attributes were identified, but it was judged too 'impractical' to make comparisons of all of them. So they were divided into 3 smaller groups consisting of a few attributes each, which were evaluated by specialists from a given area. This led to a modified AHP hierarchy including only 9 attributes and 3 alternatives [14]. The aim of yet another study [15] was to propose an application supporting modular capability-based ERP software selection which uses AHP method. The selection procedure involved 2 candidate systems and 8 main criteria. Even with that small number of alternatives and attributes, the calculations needed to reach the final decision were quite extensive. The research presented by Godse and Mulik regarded using AHP method for supporting the selection of a SaaS product [17]. The case involved 16 criteria and 3 alternative products, while the study performed by Ngai, on the use of AHP for selecting the best web site for online advertising [27] included 5 criteria and 5 alternative websites.

The PROMETHEE technique was used by Kazemi and Bardeji in the insurance branches ranking [28] as well as for equipment selection [29], portfolio selection [30], or ERP selection [16].

Even though the problem of selecting the best project management and communication software seems very similar to the ones researched in the mentioned studies, it has some characteristics which makes the use of AHP, PROMETHEE or ELECTRE family methods not quite suitable for dealing with it. First of all, the number of criteria (twenty-five), as well as the considered alternatives (ten) makes using pair-wise comparisons of the criteria (for assigning their relative importance) as well as pair-wise comparisons of the alternatives against all criteria, a very extensive and troublesome task, which causes the process to be inefficient. Besides, outranking methods should be used when a sensitive analysis to resolve disagreements or to support critical decisions about discriminating product features or cornerstone products is needed [31]. This was

not the case in this study, as here the advantages and weaknesses of the compared alternatives were straightforward.

Applicability of the other MCDA methods, like MAUT, ANP, MACBETH or TOPSIS was not considered, because the required input data or the sophisticated mathematic models they use made them irrelevant or too complicated for the issue under investigation.

3 Research Methodology

3.1 Theoretical Background

In partial ordering, to acknowledge object X as better than object Y (written as: $X \geq Y$), there must be at least one indicator value for object X which is higher than the corresponding indicator value for object Y, and no indicator for object X is lower than the corresponding indicator value for object Y. If some indictors for object X are higher and others are lower than the corresponding indicators of object Y, then the objects are recognized as incomparable. A set of comparable objects is called a chain, whereas a set of mutually incomparable objects is called an antichain. If all indicator values for two objects are equal, the objects are considered as equivalent, having the same rank [32].

Partial orders can be visualized with Hasse diagrams, in which comparable objects are connected by a sequence of lines, while incomparable objects are not connected. The levels give approximation to a weak order of the objects from "bad" (bottom) to "good" (top). Before constructing a Hasse diagram, it is essential to make sure that all indicators have a uniform orientation. Partial order method provides a weak order, where tied orders are not excluded. This is obtained by calculating the average order of the single objects, as e.g. described by Bruggemann and Annoni [33].

As mentioned in the Introduction, the original Hasse method is not well suited for problems with many criteria, where a large number of incomparabilities can often be observed, and it does not allow to discriminate the respective criteria in terms of their relevance. This is why the weighted regularized Hasse method by Grisoni et al. [25], free of the shortcomings listed above, has been chosen to solve the project management and communication software selection problem. The details of the performed procedure are provided in Subsect. 3.3; before that, however, the project selected for the study is described.

3.2 The Case Problem

The case problem was to select the best project management and communication software for a consortium carrying out an international cooperation project realized within the Interreg South Baltic Programme 2014–2020 and supported by the European Union from the European Regional Development Fund.

The consortium comprises three scientific partners – University of Szczecin (the Lead Partner), University of Applied Sciences in Stralsund, and Business Academy North in Greifswald; five museum or cultural institution partners – National Marine

Fisheries Research Institute in Gdynia, Lithuanian Sea Museum in Klaipeda, Malmö Museums, NaturBornholm, Experyment Science Centre in Gdynia; and one partner specialized in creative IT-related events – Foundation of Internet Industry Development "Netcamp". Apart from the nine partners, taking part in all project activities, there are also eight associated partners involved only in selected activities. These are: Estonian Maritime Museum from Estonia, Museum Lolland-Falster and Experimentarium from Denmark, Museum of the World Ocean from Russia, IT-Lagune e.V. and Mecklenburg-Vorpommern Tourist Board from Germany, IZITEQ B.V. from the Netherlands and City Culture Institute from Poland. The high number of involved partners is certainly a factor of difficulty in project communication.

The main aim of the project is to develop new IT-enabled tourism products for natural and cultural heritage tourist destinations in the South Baltic Region in a form of multilingual BYOD-guided tours providing an enhanced visitor experience during and after the visit featuring multimedia content and gamification techniques.

The case problem has the following characteristic properties:

- there is no single project management system used a priori by all or even most of the consortium members (the users have thus different experiences and expectations);
- the staff of respective consortium members has very different levels of IT fluency, hence the need for a very easy to use (to avoid putting off the users less fluent in IT), but still highly functional solution (to satisfy the more fluent users and their more sophisticated needs);
- the project is scheduled for three years (so, there is enough time to learn the new software, which makes the prior experience not a critical factor);
- the financial management is done in a separate system prescribed on the European Union programme level (therefore no financial features should be taken into account in the evaluation);
- the intention of the consortium was to find a tool, which would strongly support communication and collaboration activities (hence the focus is on functional aspects relevant to them).

3.3 Selection Procedure

The procedure of selecting the best project management and communication software included four stages. The first one comprised two phases – selection of criteria against which the potential project management systems will be ranked and obtaining weights reflecting the importance of each criterion. The weights were set on the basis of a questionnaire answered by the project partners' representatives. For each criterion, they were asked to assess its importance on a five grade scale. For each value on the scale, a number has been assigned: not important – 0, of little importance – 1, desired – 2, important – 3, absolutely crucial – 4. The weights of the respective criteria were calculated by summing up the numbers obtained from the respondents and then normalizing them to make the sum of weights of all the criteria equal to 1.

In stage two, a set of project management systems to be evaluated was chosen. Because of the huge amount of that type of tools available, a pre-selection phase was needed. The pre-selection was based on the following assumptions: the software is

recognized as popular on the benchmarking lists [34–37] the annual cost of using the tool by 25 users does not exceed the threshold of 600 euro (negotiated with the consortium members), the available disk space (in case of cloud solutions) is not less than 20 GB, and a demo/trial version of the tool is freely available for testing.

In the third stage, the pre-selected project management tools were evaluated with respect to the criteria – features of the system which were defined in stage one. Fourteen of the criteria could be rated using a binary scale: with 1 assigned if a certain criterion was met, and 0 if it was not. Other criteria needed a larger evaluation scale (0 to 2 or even 3), because their scope strongly differed among the tested tools. All functions and features which were evaluated using a non-binary scale are listed in Table 1 (Table 2 presents a full list).

Table 1. Evaluation rules for the non-binary features of the project management systems.

Feature (scale)	Levels (points awarded)
Sharing and co-creating documents (0–3)	No sharing/co-creating (0), place to store and share files (1), place to store and share files with version control (2), sharing files and co-creating documents (3)
Email integration (0–3)	No email integration (0), notifications to external email (1), possibility to send to/receive from external email (2), own mailbox/internal messages (3)
Instant messenger (0–2)	No chat (0), one-on-one chat (1), group chat (2)
Notifications (0–2)	No notifications (0), automatic, but poorly configurable notifications (1), highly configurable automatic notifications (2)
Project schedule (0–2)	No schedule (0), schedule only defined in tasks (no visualization) (1), schedule displayed on a Gantt chart (2)
Managing tasks (0–2)	Flat or two-level task hierarchy (0), at least three level tasks hierarchy (1), multilevel task hierarchy, task dependencies (2)
Dashboard (0–3)	No dashboard (0), dashboard with only recent activities (1), dashboard with tasks, activities, calendar (2), dashboard with graphical visualization of project status (3)
Shared calendar (0–2)	No shared calendar (0), calendar with no integration with external calendars (1), calendar with integration with external calendars and/or meeting planner (2)
Access control (0–2)	No user access management (0), basic user access management (1), advanced user access management (2)
Mobile version (0–2)	No mobile app (0), basic functions mobile app (1), full mobile app (2)
Interoperability (0–2)	No interoperability (0), basic export/import possibilities (1), integration with many different tools (2)

As a result of stage three, the original Hasse matrix and the corresponding diagram (Fig. 1) were obtained.

The goal of the final stage was to determine the complete ranking of the project management and communication systems for the BalticMuseums: Love IT! project team, taking into consideration the weights of the respective criteria. To accomplish that, the approach proposed by Grisoni et al. [25] was followed.

In its first phase, the weighted count matrix t^W was obtained using the following formula [25, Eq. 1]:

$$t_{ij}^W = \sum_k w_k \cdot \delta_{i,j,k} \tag{1}$$

where:

$$\delta_{i,j,k} = \begin{cases} 1 & if\ x_{ik} \triangleright x_{jk} \\ 0.5 & if\ x_{ik} = x_{jk} \\ 0 & if\ x_{ik} \triangleleft x_{jk} \end{cases} \tag{2}$$

and: w_k denotes the weight of criterion k, and x_{ik} the evaluation of object (software system) i with regard to criterion k.

In the second phase, a weighted regularized Hasse matrix H^R was obtained from the weighted count matrix using the following formula [25, Eq. 2]:

$$[H^R(t^*)]_{ij} = \begin{cases} 1 & if\ t_{ij}^W \geq t^* \\ -1 & if\ t_{ij}^W \leq 1 - t^* \\ 0 & otherwise \end{cases} \tag{3}$$

where t^* has been set to the minimum allowed value of 0.55 – i.e. it is enough for object i to be better than object j in 55% of the criteria to set an ordering between the two.

Although Grisoni et al. [25, p. 97] suggested one more phase to obtain a total ordering, in our case it was unnecessary, as the chosen value of t* for the construction of the weighted regularized Hasse matrix was sufficient to eliminate all the incomparability between evaluated tools and to construct the complete ranking. Thus, the complete ranking of the project management and communication systems for the investigated project could be constructed.

4 Research Findings

The first stage of the research procedure resulted in specifying 25 criteria – features of project management tools whose importance was evaluated by the respondents of the questionnaire – the representatives of the project partner organizations (in total, 16 responses were received). Note the simplicity of the data gathering process as compared to, e.g., the AHP method requiring pairwise comparisons [21].

Table 2 presents the criteria together with corresponding normalized weights (the normalization consisted in dividing each weight by the sum of all weights so that the sum of all normalized weights is 1). The values reflect which criteria were indicated as the most important by the majority of the respondents; the five top-ranked were: email integration, sharing and co-creating documents, managing project tasks, project schedule, and dashboard.

Table 2. Normalized weights of the project management software evaluation criteria

Symbol	Criterion (feature of the system)	Normalized weight	Symbol	Criterion (feature of the system)	Normalized weight
C1	Sharing and co-creating documents	0.053493	C14	Risks register	0.034934
C2	Email integration	0.063319	C15	Shared calendar	0.043668
C3	Audio/video conference	0.040393	C16	Poll option	0.028384
C4	Discussion forum	0.032751	C17	Access control	0.036026
C5	Instant messenger	0.029476	C18	Mobile version	0.044760
C6	Notifications	0.046943	C19	Configurability	0.031659
C7	Project schedule	0.049127	C20	Interoperability	0.044760
C8	Managing project tasks	0.052402	C21	Ability to install on an own server	0.025109
C9	Work time register	0.037118	C22	Availability of detailed documentation	0.030568
C10	Wiki pages	0.032751	C23	Availability of tutorials	0.045852
C11	Search engine	0.037118	C24	Helpdesk – technical support	0.039301
C12	Dashboard	0.046943	C25	Ability to withdraw and delete data	0.033843
C13	Issues register	0.039301			

The second stage of the research concentrated on the pre-selection of project management and communication tools for the final evaluation. As a result of the pre-selection process, nine project management systems were chosen. We denote them with symbols T1–T9 as their actual names are irrelevant to the description of the selection procedure, however curious readers may find them in our earlier conference paper [38].

In the third stage, each of the pre-selected systems was evaluated with regard to each of the 25 criteria, what resulted in creation of a source matrix for the Hasse diagram. Note that at this stage the criteria weights were not yet taken into account.

The Hasse diagram, presented in Fig. 1, reveals the dependences among the evaluated project management systems. There are five chains showing the order among some of the systems: T6 $\geq$ T4 $\geq$ T3, T6 $\geq$ T7, T6 $\geq$ T2, T1 $\geq$ T2 and T1 $\geq$ T7. Apart from the listed chains, other systems are incomparable with one another. Due to the existence of incomparable objects, the Hasse diagram does not provide the complete ranking of the evaluated software.

In order to accommodate the criteria weights, in the first phase of the final stage, the weighted count matrix t^W was calculated (Table 3).

Another Hasse diagram (presented in Fig. 2) has been drawn, based on these data. Already at this point, it can be seen which project management tool is better than all others and which is the worst, but there are still a few systems which are incomparable (e.g. T1 and T4).

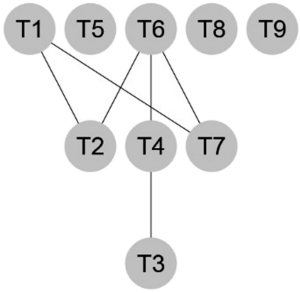

Fig. 1. The original Hasse diagram (Source: own elaboration, obtained using [39]).

Table 3. Weighted count matrix for the selected project management tools

Project management tool	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1	0.500	0.729	0.718	0.574	0.678	0.447	0.778	0.664	0.694
T2	0.271	0.500	0.570	0.362	0.553	0.249	0.585	0.493	0.538
T3	0.282	0.430	0.500	0.310	0.456	0.247	0.510	0.419	0.463
T4	0.426	0.638	0.690	0.500	0.691	0.373	0.723	0.632	0.698
T5	0.322	0.447	0.544	0.309	0.500	0.269	0.554	0.463	0.507
T6	0.553	0.751	0.753	0.627	0.731	0.500	0.800	0.717	0.716
T7	0.222	0.415	0.490	0.277	0.446	0.200	0.500	0.409	0.431
T8	0.336	0.507	0.581	0.368	0.537	0.283	0.591	0.500	0.544
T9	0.306	0.462	0.537	0.302	0.493	0.284	0.569	0.456	0.500

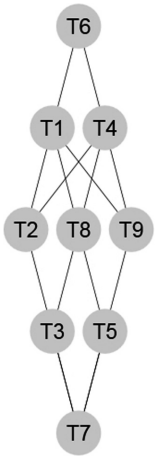

Fig. 2. Hasse diagram resulting from the weighted count matrix t^W (Source: own elaboration, obtained using [39])

Table 4. Weighted regularized Hasse matrix calculated for $t^* = 0.55$

Project management tool	T1	T2	T3	T4	T5	T6	T7	T8	T9
T1	0	1	1	1	1	−1	1	1	1
T2	−1	0	1	−1	1	−1	1	0	0
T3	−1	−1	0	−1	0	−1	0	−1	0
T4	−1	1	1	0	1	−1	1	1	1
T5	−1	−1	0	−1	0	−1	1	0	0
T6	1	1	1	1	1	0	1	1	1
T7	−1	−1	0	−1	−1	−1	0	−1	−1
T8	−1	0	1	−1	0	−1	1	0	0
T9	−1	0	0	−1	0	−1	1	0	0

Fig. 3. Hasse diagram resulting from the weighted regularized Hasse matrix calculated for $t^* = 0.55$ (Source: own elaboration, obtained using [39])

In order to eliminate the remaining incomparabilities, the weighted regularized Hasse matrix $H^R(0.55)$, based on the weighted count matrix, was calculated (Table 4).

The Hasse diagram resulting from this matrix (Fig. 3) reveals the complete ranking of the project communication management systems with respect to the defined criteria and their weights set by the project consortium members.

5 Discussion of Findings

The approach based on the weighted regularized Hasse method applied to the considered case problem of ranking project communication management systems turned out to be very effective and well suited. In several aspects it proved to be simpler than e.g. the AHP method: gathering data regarding the importance of individual criteria used a simple questionnaire with the number of questions equal to the number of criteria – not requiring pairwise comparisons of 25 criteria, neither pairwise comparison of 10 objects against 25 criteria, which would be very arduous and complicated.

The obtained results, thanks to their graphical form of Hasse diagrams, are highly readable and clear. The calculations required to obtain the weighted count matrix as well as the regularized weighted Hasse matrix did not involve any dedicated software (a spreadsheet with a single, few-lines-long automation script was used), and for producing Hasse diagrams based on the calculated matrixes, a freely accessible online tool was applied [39].

6 Conclusion

6.1 Research Contribution

The selection of the most appropriate project management system is one of the most important decisions which influence the realization, communication, collaboration and documentation processes throughout the project. It is, of course, only one of many important factors which determine the success of the project, but using the right IT tools makes all other processes easier to realize.

In this paper, it was shown how the selection of the most appropriate software can be supported using a procedure consisting of four stages: (1) definition of the evaluation criteria and their importance for the project team members, (2) pre-selection of the project management software tools, (3) evaluation of the pre-selected tools against the defined criteria, and (4) establishing the complete ranking of the evaluated tools, using the weighted regularized Hasse matrix.

The proposed procedure has been validated using the case of an international project, realized by a consortium of 17 organizations from 8 countries. The applied procedure led to the final ranking of the project management tools, listing the systems under consideration in the order of preference based on the fulfilment level of the 25 defined evaluation criteria and the criteria weights set by the consortium members.

The proposed approach based on the weighted regularized Hasse matrix was not only effective, but was also found to be much simpler than AHP or outranking methods with regard to input data gathering (stages 1 and 3), required calculations (stage 4) and the way of presenting results (also stage 4 of the procedure).

6.2 Research Implication

The presented findings have vast implications for practice, as the proposed procedure allows to obtain solutions of the software selection problem in the case of many candidate tools and multiple criteria having distinct weights, requiring less effort on data collection and less complicated calculations compared to other methods used for such purpose (referred to in the Literature review section).

Although the procedure has been developed for the project communication management software selection, it can as well be applied to software selection problems in other domains having similar context (many candidate solutions, multiple criteria, criteria having distinct weights), of course with redefined set of criteria and their respective evaluation rules.

6.3 Research Limitation and Future Works

Selection of the project communication management software is certainly an important element of setting up an effective project realization environment. It must be, however, taken into consideration that using even the best software tools for project management and communication is not enough to ensure project success [40]. Equally important is to define how and when the tool should be used and to define appropriate procedures, so that it is used effectively [41, 42].

The presented findings show the feasibility of applying the weighted regularized Hasse method to the problem of project communication management software selection, yet they do not delve into comparison of results of the proposed procedure with others, based on less easy-to-use but more widely applied methods. Such a comparison could only be made after gathering a substantial base of case projects and thus forms an interesting area for future work.

References

1. Hill, N.S., Bartol, K.M., Tesluk, P.E., Langa, G.A.: Organizational context and face-to-face interaction: Influences on the development of trust and collaborative behaviors in computer-mediated groups. Organ. Behav. Hum. Decis. Process. **108**(2), 187–201 (2009). https://doi.org/10.1016/j.obhdp.2008.10.002
2. O'Leary, M.B., Cummings, J.N.: The spatial, temporal, and configurational characteristics of geographic dispersion in work terms. MIS Q. **31**(3), 433–452 (2007)
3. Cramton, C.D., Webber, S.S.: Relationships among geographic dispersion, team processes, and effectiveness in software development work teams. J. Bus. Res. **58**, 758–765 (2005). https://doi.org/10.1016/j.jbusres.2003.10.006

4. Joshi, A., Lazarova, M.B., Liao, H.: Getting everyone on board: the role of inspirational leadership in geographically dispersed teams. Organ. Sci. **20**(1), 240–252 (2009). https://doi.org/10.1287/Orsc.1080.0383
5. Daim, T.U., Ha, A., Reutiman, S., Hughes, B., Pathak, U., Bynum, W., Bhatla, A.: Exploring the communication breakdown in global virtual teams. Int. J. Proj. Manag. **30**(2), 199–212 (2012). https://doi.org/10.1016/j.ijproman.2011.06.004
6. Verburg, R.M., Bosch-Sijtsema, P., Vartiainen, M.: Getting it done: critical success factors for project managers in virtual work settings. Int. J. Proj. Manag. **31**(1), 68–79 (2013). https://doi.org/10.1016/j.ijproman.2012.04.005
7. Oliveira, J., Tereso, A., Machado, R.J.: An application to select collaborative project management software tools. In: Rocha, Á., Correia, A.M., Tan, F.B., Stroetmann, K.A. (eds.) New Perspectives in Information Systems and Technologies, Volume 1. AISC, vol. 275, pp. 467–476. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05951-8_44
8. Kutlu, B., Bozanta, A., Ates, E., Erdogan, S., Gokay, O., Kan, N.: Project management software selection using analytic hierarchy process method. Int. J. Appl. Sci. Technol. **4**(6), 113–119 (2014)
9. Ali, A.S.B., Anbari, F.T.: Project management software acceptance and its impact on project success. In: Wikstrom, K., Artto, K.A. (eds.) Proceedings of the International Research Network on Organizing by Projects VI Project Research Conference, pp. 16–37. Abo Akademi University, Turku (2004)
10. Carlsen, L.: Data analyses by partial order methodology. Chem. Bull. Kazakh Natl. Univ. **2**, 22–34 (2015). https://doi.org/10.15328/cb632
11. Annoni, P., Bruggemann, R., Carlsen, L.: A multidimensional view on poverty in the European Union by partial order theory. J. Appl. Stat. **42**(3), 535–554 (2015). https://doi.org/10.1080/02664763.2014.978269
12. Munda, G.: Social Multi-criteria Evaluation for a Sustainable Economy. Springer-Verlag, Berlin (2008)
13. Lai, V.S., Wong, B.K., Cheung, W.: Group decision making in a multiple criteria environment: a case using the AHP in software selection. Eur. J. Oper. Res. **137**(1), 134–144 (2002). https://doi.org/10.1016/S0377-2217(01)00084-4
14. Wei, C.C., Chien, C.F., Wang, M.J.J.: An AHP-based approach to ERP system selection. Int. J. Prod. Econ. **96**(1), 47–62 (2005). https://doi.org/10.1016/j.ijpe.2004.03.004
15. Karaarslan, N., Gundogar, E.: An application for modular capability-based ERP software selection using AHP method. Int. J. Adv. Manuf. Technol. **42**(9–10), 1025–1033 (2009). https://doi.org/10.1007/s00170-008-1522-5
16. Kilic, H.S., Zaim, S., Delen, D.: Selecting "The Best" ERP system for SMEs using a combination of ANP and PROMETHEE methods. Expert Syst. Appl. **42**(5), 2343–2352 (2015). https://doi.org/10.1016/j.eswa.2014.10.034
17. Godse, M., Mulik, S.: An approach for selecting software-as-a-service (SaaS) product. In: Proceedings of the IEEE International Conference on Cloud Computing, pp. 155–158. IEEE, Bangalore (2009). https://doi.org/10.1109/cloud.2009.74
18. Rao, R.V.: Software selection in manufacturing industries using a fuzzy multiple criteria decision making method. PROMETHEE Intell. Inf. Manag. **1**, 159–165 (2009). https://doi.org/10.4236/iim.2009.13023
19. Azadeh, A., Shirkouhi, S.N., Rezaie, K.: A robust decision-making methodology for evaluation and selection of simulation software package. Int. J. Adv. Manuf. Technol. **47**, 381–393 (2010). https://doi.org/10.1007/s00170-009-2205-6

20. Mastalerz, M.W.: Electre method for choosing an e-learning platform. In: Proceedings of VIth International Conference on Perspective Technologies and Methods in MEMS Design, pp. 168–171. Lviv Politechnic National University, Lviv (2010)
21. Ahmad, N., Laplante, P.A.: Software project management tools: making a practical decision using AHP. In: Proceedings of the 30th Annual IEEE/NASA Software Engineering Workshop, pp. 76–84. IEEE Computer Society, Columbia (2006). https://doi.org/10.1109/sew.2006.30
22. Sutterfield, J.S., Swirsky, S., Ngassam, C.: Project management software selection using analytical hierarchy process. Acad. Inf. Manag. Sci. J. **11**(2), 79–93 (2008)
23. Fattore, M., Bruggemann, R. (eds.): Partial Order Concepts in Applied Sciences. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-45421-4
24. Swacha, J., Komorowski, T., Muszyńska, K., Drążek, Z.: Acquiring digital asset management system for an international project consortium. J. Manag. Finan. **3**(1), 91–102 (2013)
25. Grisoni, F., Consonni, V., Nembri, S., Todeschini, R.: How to weight Hasse matrices and reduce incomparabilities. Chemometr. Intell. Lab. Syst. **147**, 95–104 (2015). https://doi.org/10.1016/j.chemolab.2015.08.006
26. Ishizaka, A., Nemery, P.: Multi-Criteria Decision Analysis: Methods and Software. Wiley, Chichester (2013)
27. Ngai, E.W.T.: Selection of web sites for online advertising using the AHP. Inf. Manag. **40**(4), 233–242 (2003). https://doi.org/10.1016/S0378-7206(02)00004-6
28. Kazemi, M., Bardeji, S.F.: Application of combinational approach of FAHP and PROMETHEE in the insurance branches ranking. Int. J. Procurement Manag. **9**(5), 548–567 (2016). https://doi.org/10.1504/IJPM.2016.078691
29. Yilmaz, B., Dağdeviren, M.: A combined approach for equipment selection: F-PROMETHEE method and zero–one goal programming. Expert Syst. Appl. **38**(9), 11641–11650 (2011). https://doi.org/10.1016/j.eswa.2011.03.043
30. Vetschera, R., Teixeira De Almeida, A.: A PROMETHEE-based approach to portfolio selection problems. Comput. Oper. Res. **39**(5), 1010–1020 (2012). https://doi.org/10.1016/j.cor.2011.06.019
31. Maiden, N.A., Ncube, C.: Acquiring COTS software selection requirements. IEEE Softw. **15**(2), 46–56 (1998). https://doi.org/10.1109/52.663784
32. Bruggemann, R., Patil, G.P.: Ranking and Prioritization for Multi-indicator Systems – Introduction to Partial Order Applications. Springer, New York (2011)
33. Bruggemann, R., Annoni, P.: Average heights in partially ordered sets. MATCH-Commun. Math. Comput. Chem. **71**(1), 117–142 (2014)
34. Capterra. http://www.capterra.com/project-management-software/#infographic. Accessed 14 Nov 2017
35. Creative Bloq. http://www.creativebloq.com/software/best-project-management-71515632. Accessed 14 Nov 2017
36. PCMag. http://www.pcmag.com/article2/0,2817,2380448,00.asp. Accessed 14 Nov 2017
37. Project-Management.com. https://project-management.com. Accessed 14 Nov 2017
38. Muszyńska, K., Swacha, J.: Project management and communication software selection using the weighted regularized Hasse method. In: Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, ACSIS, vol. 11, pp. 919–923 (2017). https://doi.org/10.15439/2017f346
39. pyHasse online tool. http://spyout.pyhasse.org. Accessed 14 Nov 2017

40. Pawełoszek, I.: Data mining approach to assessment of the ERP system from the vendor's perspective. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 125–143. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_8
41. Katainen, T., Nahar, N.: Using methods and IT tools innovatively for the management of international IS development projects. In: Proceedings of Portland International Conference on Management of Engineering & Technology, pp. 1851–1863. IEEE, Istanbul (2008). https://doi.org/10.1109/picmet.2008.4599805
42. Muszyńska, K.: Patterns of communication management in project teams. In: Ziemba, E. (ed.) AITM/ISM -2016. LNBIP, vol. 277, pp. 202–221. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53076-5_11

Hierarchical Representation of Website Evaluation Model Using Survey and Perceptual Based Criteria

Jarosław Wątróbski[1(✉)], Artur Karczmarczyk[2], Jarosław Jankowski[2], Paweł Ziemba[3], and Waldemar Wolski[1]

[1] University of Szczecin, ul. Mickiewicza 64, 71-101 Szczecin, Poland
{jaroslaw.watrobski,wwolski}@wneiz.pl
[2] Department of Computer Science,
West Pomeranian University of Technology in Szczecin,
ul. Żołnierska 49, 71-210 Szczecin, Poland
{akarczmarczyk,jjankowski}@wi.zut.edu.pl
[3] Department of Technology, The Jacob of Paradies University,
ul. Teatralna 25, 66-400 Gorzów Wielkopolski, Poland
pziemba@pwsz.pl

Abstract. The high availability of e-commerce websites which deliver similar services and products, as well as the harsh rivalry between competitors, increased the importance of systematic evaluation of the e-commerce websites' quality, usability and user experience. Multiple methodologies for performing the evaluation are available, however, they are based mainly on survey data. In our previous research, we introduced perceptual measurements from eye tracker (ET) to the set of evaluation criteria. In this paper, we present an approach based on AHP (Analytic Hierarchy Process) to allow a thorough analysis of the complex structure of criteria and its impact on the final evaluation. Additionally, we combine the AHP outputs with the COMET (Characteristic Objects METhod) technique to build a fuzzy rule base that provides a stable model of the entire domain of evaluation criteria. The results of the conducted empirical verification of the proposed approach are presented and discussed. The main research findings show that the rankings obtained with the presented approach are very stable and the probability of a rank reversal phenomenon is low.

Keywords: Website quality evaluation · AHP · COMET
Eye tracking · eQual

1 Introduction

In January 2017, around 50% of the total world population of 7.5 billion people were Internet users, while over 4.9 billion people used mobile devices [1]. Within a period of one year, a growth of 482 million users, i.e. 21%, was observed

E. Ziemba (Ed.): AITM 2017/ISM 2017, LNBIP 311, pp. 229–248, 2018.
https://doi.org/10.1007/978-3-319-77721-4_13

among active social media users. According to January 2016 data [2], 79% of UK population and 71% of USA population at least once within a 30-days period searched online for a product or service to purchase. In 2015, the total value of online sales in Europe was 455 billion euro [3], compared to 131.61 billion Euro in 2013 and 156.28 billion Euro in 2014. It is forecast that by 2020 the online sales in the USA will reach $523 billion [4].

The e-commerce sector is characterized by a very harsh competition. In June 2016, only 650 thousand out of 12 million online stores sold more than $1,000 per year [5]. With such a strong competition, entrepreneurs try to increase their chances by marketing and using analytic tools [6], refactoring the usability of the website [7], providing web content accessibility [8] or building the credibility of the website [9]. The latter is the perception of being trustworthy and believable, and it can be built, among other things, by providing superb user experience and high levels of usability and quality [9]. As a result, it is beneficial for the business owners to periodically evaluate the quality, usability and user experience of their website and compare it with the ones of the competition [10]. There are multiple website and e-commerce evaluation methods, such as eQual [11], SiteQual [12], E-S-QUAL [13], which have been successfully used in the evaluation of e-commerce [14], e-government [15], e-banking [13], information services [16] or university websites [17]. The methods differ in the range of possible applications, assessment scale used, their theoretical basis, verification of solution or the minimum number of evaluators.

Since the website quality evaluation problem encourages a multi-faceted viewpoint, Multi-Criteria Decision Analysis (MCDA) methods can be used to approach it, such as TOPSIS [18], PROMETHEE [19] and AHP in their classic and fuzzy variants [20,21]. Also, a hybrid approach is possible that combines the classic methods with MCDA methods, such as PEQUAL [22].

The aforementioned methods are often based on survey data, which has some disadvantages. First of all, the number of questions needs to be short or the number of alternatives limited, so the survey is not too long for the respondents to complete. Additionally, to make sure the responses are not artificial, the questionnaires need to be directed to an appropriate group of real users. Moreover, the reliability of the collected data needs to be verified, as well as a possible alternatives' ordering or comparison bias needs to be considered [11]. In order to avoid the aforestated disadvantages, research tools based on eye tracking (ET) devices are being developed. Originally, they were used mainly in medicine, however, nowadays we can also find multiple studies on user experience [23], website quality [24] and usability evaluation [25] based on the data collected with ET [26].

This paper is an extended version of the Authors' earlier work [27], where an integrated approach was proposed, in which survey-based and ET-based data was aggregated with the use of PROMETHEE [27] and TOPSIS [28] methods. Nonetheless, our prior research did not address the issue of the proper structuring of the evaluation problem. The TOPSIS method works on a flat structure of criteria, whereas the PROMETHEE method provides the means to aggregate the criteria into groups and clusters only. In order to avoid oversimplification of

the problem, it is important to consider the actual multi-level structure of the criteria. This can be achieved by utilizing the AHP (Analytic Hierarchy Process) method [29]. This method allows to thoroughly study the structure of criteria and their effect on the final evaluation. Unfortunately, the results obtained by this method are vulnerable to a rank reversal phenomenon. For this reason, we propose to combine the AHP method with the COMET method, which provides a stable model for the complete domain of evaluation.

The aforementioned problems, i.e. integration of ET-based and survey-based results in an evaluation approach, considering the complex structure of criteria and taking into account the rank reversal problem constitutes an interesting research gap, which this paper is addressing. The main objective of this paper is to extend the formerly proposed approach for e-commerce websites evaluation [27] by utilizing AHP to allow complex criteria structuring and, furthermore, by integrating COMET method analysis to provide a stable full-domain model. In practical terms, the extended framework is used to evaluate the 10 most popular e-commerce websites.

The paper is split into sections. Section 2 contains the literature review. The methodological framework of the proposed approach is presented in Sect. 3. Section 4 contains empirical study results and their discussion. The conclusions and future directions are outlined in Sect. 5.

2 Literature Review

2.1 Website Quality, Usability and User Experience

Nowadays, when the competitors are just a few clicks away, the task of constant evaluation and improvement of e-commerce websites becomes profoundly important to maintain a profitable online business [9]. Considering the fact that the human's average focus timespan is as short as eight seconds [30,31], it is essential that the e-commerce websites are created in a manner that the users are able to find what they are looking for almost immediately [32]. Therefore, the website needs to be characterized by high levels of quality, usability and user experience.

Quality is related to usability [33], which, in turn, is defined by the ISO 9241-11:1998 standard [34] as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use", and by the ISO 25010:2011 standard [35] as "the ability of software to be in intelligible, easy to learn and use as well as attractive to the user in specified circumstances". The user experience (UX), on the other hand, is a concept related to usability and is defined as "a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service" [36]. Since the system usability is not constant, as it changes depending on the users' preferences [37] or software and hardware evolution [33], it is important to regularly evaluate the website [10]. The methods of usability evaluation can be organized into 5 groups [38]: user testing, inspection methods, inquiry methods, analytical modeling and simulation methods, whereas the methods of website

quality evaluation can be divided into 3 groups: expert evaluation, user traces analysis, interviews and surveys. The latter is a common choice for websites' quality and usability evaluation.

2.2 Websites Evaluation with Surveys and Eye Tracking Devices

Regardless of the type, quantity or structure of the criteria used in an evaluation method, the methods themselves often follow a similar procedure, in which the user's thoughts are initially obtained either locally [39] or on-line [40] by surveys, and, subsequently, the responses are put on a Likert scale. Multiple classic methods have been successfully used to perform such survey-based evaluations, such as the Ahn method [14], SiteQual [41], E-S-QUAL [13], WAES (Website Attribute Evaluation System) [42], WEQ (Website Evaluation Questionnaire) [43], WPSQ (Web Portal Site Quality) [16], WQM (Website Quality Model) [44] or eQual [11]. Based on the number of citations and the wide range of applications, including evaluation of e-commerce, e-government and university websites, the latter is one of the most popular websites' evaluation methods. It is based on 22 criteria grouped into three clusters: Usability (further divided into Usability and Design subcategories), Information Quality and Service Interaction (further divided into Trust and Empathy subcategories).

Recently, an additional item has been added to the group of website usability, quality and user experience evaluation tools: eye tracking (ET) devices. A profound analysis of literature allows to group the applications of the ET devices in websites' evaluation into two groups: studies based on ET results exclusively and approaches combining the ET and survey data. While in the latter the perceptual measurements from ET device are supplemented by usability evaluation surveys, the former provides usability assessment based exclusively on measurable factors, such as AOI [23], TFF, FBT [45], visits and revisits [24] or the time required to complete a given task [46]. The methods from both groups are presented in Table 1 along with references to the research where they were used. A more detailed analysis of the methods can be found in the Authors' prior work [27,28].

Apart from the evaluation methods mentioned above, MCDA-based endeavors at websites' assessment can be found. The MCDA approach is justified since the evaluation of websites requires to consider multiple dimensions and measurements [55]. For example, Chmielarz broadly utilizes the scoring method to assess an extensive variety of business-oriented websites [56–58]. Lee and Kozar applied the AHP method to evaluate e-tourist and e-commerce websites [59]. Sun and Lin used the fuzzy TOPSIS method to evaluate e-commerce websites [20]. Del Vasto-Terrientes et al. used the ELECTRE-III-H method on traveler websites [60]. Furthermore, hybrids of different MCDA techniques can be used [21,61,62]. Apart from constructing a ranking of evaluated websites, the MCDA methods, due to their decision-making process construction [63], allow to perform an analysis of the obtained solution, such as its stability verification [64,65] or the analysis of the preferences of the decision-makers (DM) [66,67].

Table 1. The literature review of websites evaluation with the use of eye tracking (ET), surveys (S) and interviews (I).

#	ET	S	I	Application	Particip.	ET crit.	Ref.
1	Yes			E-commerce websites	21	3	[23]
2	Yes			Online shops, online newspapers, company webpages	40	2	[45]
3	Yes			Social commerce	34	2	[47]
4	Yes			Websites of mobile phone manufactures	17	5	[25]
5	Yes			E-tourism	60	3	[48]
6	Yes			E-commerce	42	3	[49]
7	Yes			Clinical guidelines on the web	14	4	[50]
8	Yes	Yes		E-commerce, B2B	25	7	[51]
9	Yes	Yes		Online banking	10	3	[52]
10	Yes	Yes		E-commerce	38	4	[24]
11	Yes		Yes	Websites of mobile service providers	44	3	[10]
12	Yes	Yes	Yes	E-government websites	9	3	[53]
13	Yes	Yes		Online hotel booking websites	16	3	[54]
14	Yes	Yes		E-commerce	30	2	[46]

In their previous research, the Authors used the PROMETHEE [27] and TOPSIS [28] MCDA methods to propose an approach to websites' evaluation based on a combined set of 22 eQual-based survey criteria and 6 ET-based perceptual criteria. The proposed evaluation model, however, had some limitations regarding the structuring of the evaluation problem. The TOPSIS method ignored the structure of criteria completely and allowed to manipulate with the DM preferences only on the individual criteria level [68]. On the other hand, the PROMETHEE method allowed to aggregate the criteria into groups and clusters and analyze the preference relations between them and the evaluated variants, nonetheless, the obtained model did not allow to address the complexity of the criteria to its full extent [69,70]. Moreover, both utilized methods are prone to a rank reversal [71,72], i.e. a phenomenon in which the relative position of two variants can change if another variant is introduced to the ranking. Therefore, the aforementioned limitations motivate the Authors' contribution to modify the approach presented in [27] to accommodate the need to precisely map the complexities of the criteria structure and to increase the stability of the obtained solution by utilizing rank-reversal free techniques.

3 Research Methodology

3.1 Proposed Framework for Website Evaluation

The proposed framework, visually demonstrated in Fig. 1, derives from the classical survey approach. It utilizes the eQual website evaluation method's set of 22 criteria grouped into three categories (Usability, Information Quality, Service Interaction) and five subcategories (Usability, Site Design, Information Quality, Trust, Empathy) [11]. Moreover, due to the surveys' limitations mentioned above

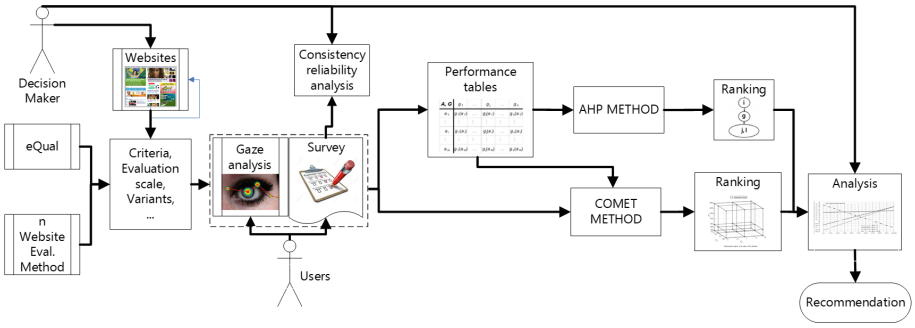

Fig. 1. The methodological framework for websites' evaluation.

and in an endeavor to increase the objectivity of the research by incorporating physically measurable data, the evaluation criteria set was extended with the input values from an eye tracking (ET) device. Such approach extended the retrospective survey research with additional empirical data recorded in real-time. On the other hand, basing the research exclusively on ET data would limit the evaluation range only to a set of selected parts of the system, whereas incorporation of survey data allows a more holistic assessment. Eventually, a set of 28 criteria was created (Fig. 2).

While the 22 survey-based criteria are described in detail in [11], the 6 ET-based criteria require some additional explanation. During the perceptual part of the empirical research, a set of carefully prepared screenshots from the evaluated websites is displayed to the participants. During the preparation phase, the researchers analyze each slide and configure which of its parts are directly related to sales conversions – they mark so-called areas of interest (AOI). During the experiment phase, the participants are given a task to perform on each of the displayed slides, within a specified time limit, thus allowing the ET software to collect the following data:

E1 – viewers
number of people who have visited the configured areas of interest (AOI);
E2 – first view (s)
time elapsed in seconds before the area was noticed for the first time;
E3 – watched time (s)
time spent on a given AOI, expressed in seconds;
E4 – watched time (%)
time spent on a given AOI, expressed in percent;
E5 – revisitors
the number of participants who returned to the AOI;
E6 – revisits
the number of revisits to the AOI.

In case of the Authors' empirical research, a set of 10 carefully selected e-commerce websites was evaluated: Alibaba (A1), Amazon (A2), Apple (A3),

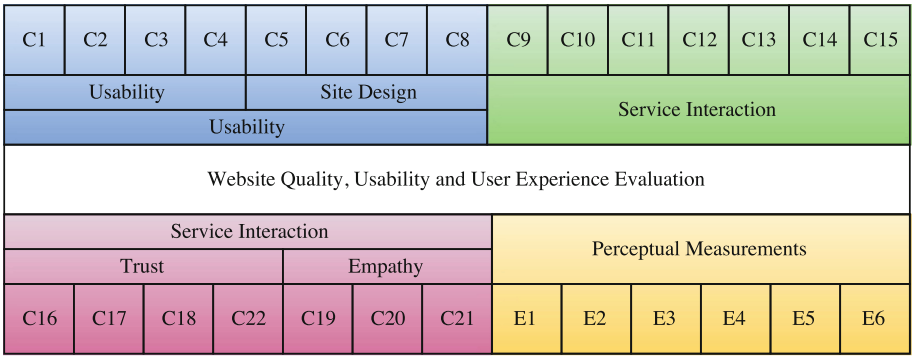

Fig. 2. The combined set of 22 survey-based and 6 ET-based criteria.

BestBuy (A4), eBay (A5), Macy's (A6), Rakuten (A7), Staples (A8), Target (A9) and Walmart (A10). The choice of the websites was based on existing top e-commerce websites' rankings. In the survey part of the research, questionnaires on each evaluated website were collected from 41 users accustomed to online shopping, thus providing 410 surveys with answers expressed on a 7-point Likert scale. Subsequently, in the perceptual part of the research, a group of 20 computer-literate students was presented a set of 3 slides for each evaluated website, with the use of ET device and GazePoint software. The slides contained:

Home page. The front page of each website, containing a product search form and a list of categories;
Product page. A page containing the description, images and price of a single product in offer;
Payment page. The page from the purchase process which provides the possibility to choose the payment method.

Each slide was presented to the participants for a period of 10 s, interleaved with 3-s pauses between slides' transitions. The following AOIs were configured: a piece of electronic (either a smartphone or a watch) on the home pages, the price on the product pages and the PayPal payment method on the payment pages. Eventually, the perceptual data was combined with the aggregated survey responses and constituted an input to the MCDA-based evaluation model.

3.2 AHP Method

In order to allow to study the effect of the criteria structure in its full complexity on the final evaluation result, the AHP (Analytical Hierarchy Process) method by Saaty [29,73,74] was chosen to aggregate the input data. The method is built on three principles [75]: construction of a hierarchy, setting priorities and logical consistency. The decision problem is decomposed and structured into a hierarchy of sub-objectives, attributes, criteria and variants. The structure can

contain many layers. In the next step, the decision maker (DM) compares each two elements at each level of the hierarchy, thus providing their relative priority in the form of the 1–9 scale if one element surpasses the other, or in the form of the $\frac{1}{9}$–1 reciprocal scale otherwise. The values of the scale can be defined as: 1 - equal importance, 3 - moderate importance of one element over another, 5 - strong importance, 7 - very strong importance, 9 - extreme importance. The results of the comparisons are then stored in a pairwise comparison matrix, which allows to obtain the weights of individual elements of the structure. The procedure is repeated on all levels of the hierarchy, eventually providing the final score of each variant. Sometimes, the DM's judgments can be inconsistent, however, the AHP method can accommodate these inconsistencies as a tolerable error in measurement, as long as they do not exceed 10%. Alternatively, the actual ratio of both compared elements' measurement values can be used if available instead of the DM's judgments.

In case of the Authors' empirical study, when comparing elements on the individual criteria level of the hierarchy, a ratio of the two compared elements' values was computed and scaled to the 1–9 scale. In contrast, when comparing the groups within clusters and the clusters with other clusters, initially an equal weight was assigned to all elements. For the perceptual branch of the hierarchy, four separate evaluations were performed. First, only the data from home pages or product pages or payment pages was used. Eventually, in the fourth assessment, the perceptual data from all three evaluated pages was averaged for each website.

The calculation of scores for each of the four scenarios was followed by a sensitivity analysis of the averaged scenario. The sensitivity analysis is the MCDA methods' tool that allows to explore how the possible changes in the DM's preferences would affect the final ranking of variants. Thus, the stability of the obtained rankings was studied.

Unfortunately, AHP and many other MCDA methods such as TOPSIS and PROMETHEE are known to be prone to the rank reversal phenomenon, which means that an introduction of an 11th website to the ranking could change the relative positions of the originally evaluated websites. Therefore the Authors propose to extend the framework by incorporating the COMET method analysis into the research approach.

3.3 COMET Method

The COMET (Characteristic Object Method) method [72,76] is an MCDA method in which the complexity of the algorithm is independent of the number of variants. Similarly to AHP, in COMET the DM performs pairwise comparisons of elements, yet instead of variants, characteristic objects (COs) are compared. The evaluation of the COs allows to form a matrix of expert judgement (MEJ), which, in turn, allows to obtain a fuzzy rule base for determining the preferences. Eventually, Mamdani's fuzzy inference method is used to compute the preference value of each variant and to form the final ranking of variants [77]. The rule base guarantees that the produced results are unequivocal and the bijection makes

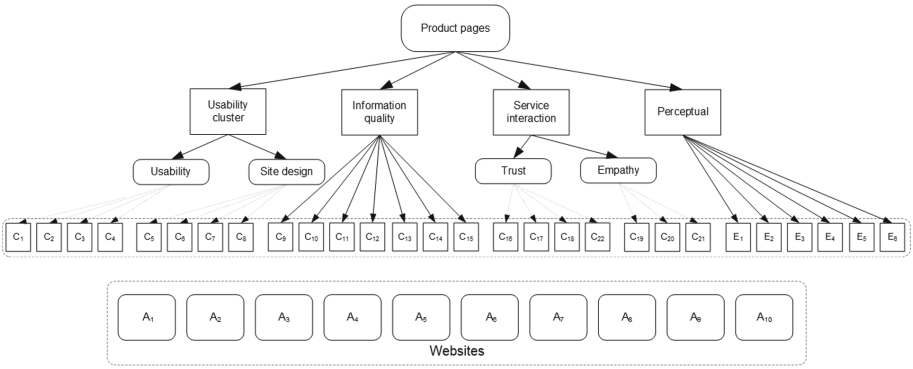

Fig. 3. AHP hierarchy for the e-commerce website evaluation problem.

the obtained ranking completely free of rank reversal. A detailed description of the COMET method can be found in [78,79].

During the empirical research, the output from the clusters' level of the AHP hierarchy was used to produce the input to the COMET method. The minimum, average and maximum values were calculated for each of the four clusters, thus providing three characteristic values for each cluster. The characteristic objects were then generated as a Cartesian product of all the characteristic values. As a result, 81 characteristic objects were produced. In order to create the MEJ matrix, 3240 pairwise comparisons of the characteristic objects needed to be performed. Subsequently, the MEJ matrix was used to produce the fuzzy rule base. Eventually, the variants were evaluated with the obtained COMET model and the generated ranking was compared with the AHP-based rankings.

4 Empirical Study and Results Discussion

In the first step of the empirical research, survey data from 41 users and perceptual data from 20 users evaluating 10 top e-commerce websites was combined and used to build a performance table (see Table 2). The data from the obtained table was then used to build the AHP hierarchy of the evaluation problem. The produced hierarchy is presented in Fig. 3. The analysis of the figure allows to notice the complexity of the structure, which was not fully taken into account during our previous research [27]. The Usability and Site Design groups are aggregated into Usability cluster, and the Trust and Empathy groups are aggregated into Service Interaction cluster, therefore the criteria under these clusters are structured into four levels of hierarchy. On the other hand, in case of the Information Quality and Perceptual clusters, the criteria are structured only into three levels.

Table 2. Performance table for the top 10 e-commerce websites' evaluation.

Cluster	Group	Criterion	Website									
			A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Usability	Usability	C1	4.902	5.610	5.683	5.000	6.024	5.049	4.976	4.927	4.854	5.049
		C2	4.951	5.707	5.415	4.878	5.951	4.976	5.098	4.927	4.756	5.220
		C3	5.000	5.317	5.610	5.000	5.610	4.854	4.805	4.829	4.683	4.829
		C4	4.829	5.390	5.585	4.878	5.634	5.049	4.854	4.659	4.854	5.244
	Site design	C5	4.829	5.024	5.976	4.341	4.683	4.707	4.268	4.512	4.220	4.927
		C6	5.098	5.488	6.024	4.561	5.341	5.049	4.707	4.927	4.707	4.805
		C7	4.829	5.366	5.829	4.537	4.878	4.756	4.439	4.732	4.415	4.805
		C8	4.634	5.146	5.415	4.049	4.512	4.585	4.024	4.220	3.683	4.268
Inf. quality	Inf. quality	C9	5.000	5.537	5.049	5.073	5.634	4.780	4.805	4.780	4.756	4.537
		C10	4.902	5.537	5.902	5.098	5.683	4.902	5.024	4.805	4.902	4.805
		C11	5.585	5.268	5.488	5.122	5.415	5.512	5.488	5.146	5.561	5.317
		C12	4.951	5.463	5.341	5.268	5.537	4.902	4.732	4.854	5.049	4.610
		C13	4.732	5.537	5.561	5.244	5.512	4.878	4.756	4.707	4.902	4.976
		C14	4.854	5.488	5.171	5.098	5.220	4.634	4.659	4.854	5.024	4.488
		C15	4.927	5.390	5.293	4.854	5.488	4.732	4.512	4.829	4.756	4.951
Service inter.	Trust	C16	4.927	5.829	5.927	4.244	5.878	4.512	4.415	4.488	4.195	4.927
		C17	4.732	5.805	6.000	4.537	5.659	4.512	4.293	4.927	4.317	4.951
		C18	4.732	5.610	5.805	4.707	5.561	4.659	4.390	4.780	4.220	4.902
		C22	4.683	5.610	6.171	4.634	5.268	4.756	4.220	4.683	4.220	4.902
	Empathy	C19	3.951	4.927	4.878	3.537	4.049	3.976	3.659	3.756	3.366	3.951
		C20	3.878	4.683	4.293	3.366	3.488	3.439	3.463	3.610	3.146	3.756
		C21	4.780	5.268	5.561	4.829	5.293	4.610	4.268	4.390	4.610	4.732
Perceptual	Home pages	E1	10	15	14	18	16	10	12	16	17	10
		E2	5.210	5.00	4.64	1.85	3.27	4.41	3.35	3.91	3.32	3.65
		E3	0.390	0.53	1.06	0.98	0.60	0.41	0.62	0.65	0.81	0.43
		E4	3.870	5.27	10.57	9.83	6.04	4.06	6.24	6.45	8.11	4.27
		E5	6	13	9	18	13	6	10	11	16	6
		E6	1.800	2.50	5.10	3.40	2.50	3.80	3.10	2.50	4.90	2.80
	Product pages	E1	20	20	20	20	20	20	18	18	19	20
		E2	1.990	2.04	2.73	1.47	1.67	1.39	1.60	1.25	1.94	1.91
		E3	2.030	2.70	3.22	2.83	1.84	2.55	2.31	3.14	1.85	0.80
		E4	20.280	21.70	32.25	28.27	18.41	25.49	23.15	31.45	18.48	7.98
		E5	17	19	15	18	18	19	18	18	16	16
		E6	2.300	5.50	3.50	4.60	5.40	5.10	5.50	5.30	6.50	2.90
	Payment pages	E1	20	20	19	20	18	17	17	20	16	20
		E2	2.350	0.32	1.96	2.93	1.15	2.85	4.54	0.32	3.22	2.88
		E3	0.990	3.67	2.57	2.93	3.86	0.90	2.01	4.97	0.97	1.83
		E4	9.890	36.72	25.69	29.26	38.61	8.96	20.10	49.70	9.70	18.30
		E5	17	20	18	18	18	14	16	20	15	18
		E6	4.20	4.30	6.30	4.30	4.00	6.40	4.50	3.90	4.20	4.80

4.1 Analysis Based on Perceptual Measurements of Individual Pages

After structuring the problem into a four-level hierarchy, the AHP method was used to evaluate the websites using the survey data combined with perceptual measurements of individual pages. All clusters and groups were initially assigned equal weights (their comparison with each other on a single level always returned 1 on Saaty's scale [74], i.e. equal importance). On the other hand, in case of the

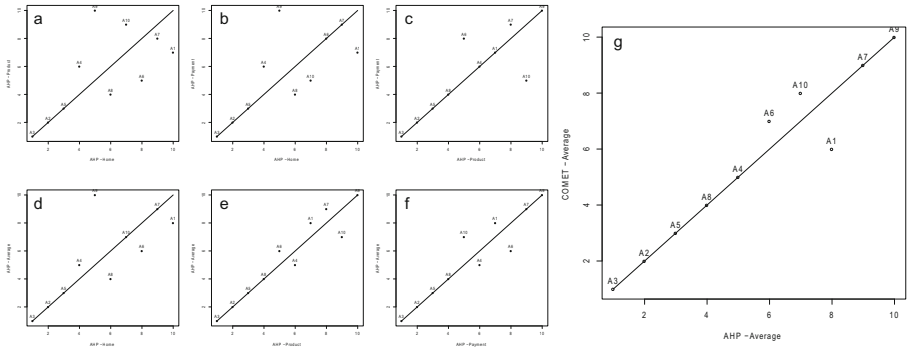

Fig. 4. Comparison of the evaluated websites' ranks in various rankings: AHP for data including ET measurements for home, product and payment pages (a–c); AHP for data including ET measurements for individual pages or averaged measurements (d–f); comparison of rankings based on averaged data with AHP method and AHP + COMET (g).

criteria C1–C22 and E1–E6, the variants were compared pairwise by computing their ratio, which was then scaled to the range 1–9 or $\frac{1}{1}$–$\frac{1}{9}$. The results of the performed evaluations are presented in Table 3a–c. The analysis of the results shows that regardless of the scope of perceptual measurements chosen, the variants A3, A2 and A5 are on the leading positions in the obtained ranking. The ranks of the remaining 7 websites differ depending on the evaluation's perceptual input data. The differences between the ranks that the variants obtained is depicted in Fig. 4a–c.

Figure 4 depicts rank comparison charts. If a variant is plotted on the diagonal line on the chart, there is no difference in rank of that variant in both compared rankings. The variants that are plotted close to the line have a slight change of rank, whereas the variants plotted far from the line suffered a significant change of rank. The analysis of Fig. 4a–c allows to observe that the product and payment page data produced the most similar rankings – only three of the variants, i.e. A6, A9 and A10 changed their positions. For example, if payment page data is used instead of the product page data, the A10 variant advances considerably from rank 9 to rank 5. This means that although the product page of this website is not appealing to the experiment participants, the usability of the payment page places it in the first half of the evaluated websites.

4.2 Analysis Based on Averaged Perceptual Data

In the next step of the research, the perceptual measurements data from home, product and payment pages was averaged and, subsequently, a new evaluation was performed. The obtained ranking is presented in Table 3d and its comparisons to the rankings obtained in Subsect. 4.1 are presented in Fig. 4d–f. Similarly to the rankings based on the individual ET inputs, the variants A3, A2 and A5 were given ranks 1–3 respectively. The analysis of Fig. 4d–f shows that if the rankings based on the home and product pages are compared to the ranking

Table 3. Evaluation results for AHP method (a–d) and COMET method (e) for survey data combined with perceptual data regarding (a) home pages, (b) product pages, (c) payment pages; or averaged perceptual data (d–e).

	a		b		c		d		e	
	AHP - home		AHP - product		AHP - payment		AHP - average		COMET	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank
A1	0.088	10	0.094	7	0.09	7	0.092	6	0.30054524	6
A2	0.113	2	0.117	2	0.118	2	0.118	2	0.86938661	2
A3	0.125	1	0.12	1	0.119	1	0.121	1	0.90972222	1
A4	0.102	4	0.096	6	0.094	6	0.097	5	0.38450486	5
A5	0.111	3	0.11	3	0.112	3	0.112	3	0.7397428	3
A6	0.09	8	0.097	4	0.089	8	0.092	6	0.29244032	7
A7	0.089	9	0.091	8	0.087	9	0.089	9	0.18725644	9
A8	0.093	6	0.097	4	0.106	4	0.099	4	0.38653109	4
A9	0.097	5	0.09	9	0.087	9	0.088	10	0.1712963	10
A10	0.091	7	0.09	9	0.099	5	0.092	6	0.26890561	8

based on the averaged data, half of the websites remain on unchanged positions, whereas in case of the payment page ranking only 4 variants change their positions when the evaluation approach is changed to the averaged one.

4.3 Sensitivity Analysis of the Obtained Solution

One of the major advantages of utilizing MCDA methods to websites' evaluation is the possibility to study how the final evaluation would change if the decision maker (DM) put more importance to some criteria or groups of criteria compared to others. For example, if the DM decided that the criteria from the Perceptual cluster should be moderately more important than the ones from the other criteria (at least 4 on Saaty's scale), the order of the ranking from Subsect. 4.2 would change in a manner that the variant A8 would advance from rank 4 to rank 3, thus outranking the variant A5. Moreover, if the DM gave the perceptual criteria even more importance (7 or more on Saaty's scale), the A8 website would become the leader in the ranking, which means that this variant is the one most supported by the perceptual criteria. On the other hand, if the weight of the Perceptual cluster was to be decreased significantly, the A8 variant would drop to the next to penultimate position with rank 8.

The aforementioned facts can be stated as a result of the performed sensitivity analysis. In each step of the sensitivity analysis, the preferences matrix was modified in such a manner, that the analyzed cluster received consecutively the score of $\frac{1}{9}, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4, 5, 6, 7, 8, 9$, while the results of the comparisons of the other clusters remained unchanged at the value of 1. No change was imposed on the pairwise comparisons of the individual criteria, i.e. the ratio of the compared values was calculated and scaled to the range of $\frac{1}{9} \ldots 1$. Thus, the problem was solved 17 times for each cluster, and the obtained scores and ranks were then plotted. The results of the performed sensitivity analysis are

presented in Fig. 5. The default scenario is marked with the red line. The analysis of Fig. 5e shows that the ranking based on averaged perceptual data is quite stable, and the change of the leading position can be observed only if the weight of the Perceptual or Information Quality cluster is increased to over 7. On the other hand, no change in the weights of the Usability and Service Interaction cluster criteria can change the positions assigned to the three best variants. The analysis of Fig. 5a–d allows to notice that the leading variant A5 is supported by all the survey-based clusters, whereas the Perceptual cluster most strongly supports the variants A8 and A4, but is in conflict with the A1 website, and based on this cluster's weight its evaluation can drop even 6 positions from rank 4 to rank 10.

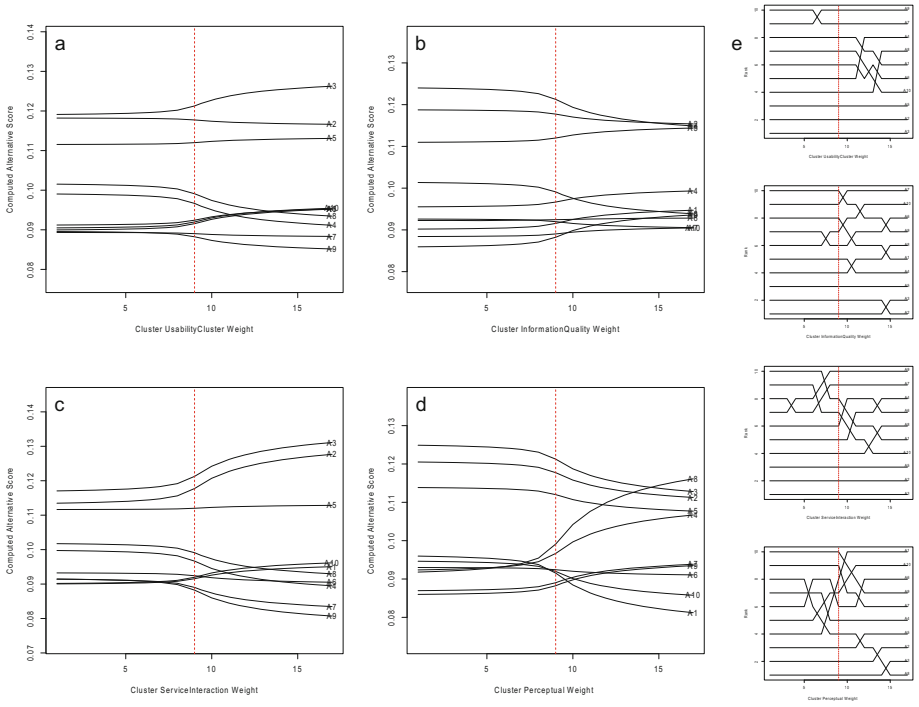

Fig. 5. Sensitivity analysis results for usability (a), Information Quality (b), Service Interaction (c) and Perceptual clusters (d). Score changes (a–d) and rank changes (e).

4.4 Increasing Evaluation Stability with Characteristic Objects

The AHP evaluation was followed by the assessment with the use of the Characteristic Objects METhod (COMET), which, instead of comparing individual variants, compares the characteristic objects (CO), thus allowing the researchers to obtain a stable evaluation model. The COMET evaluation was performed based on the averaged perceptual measurements scenario (Table 3d). Minimum,

Table 4. The output of the AHP method with averaged perceptual measurements.

	Weight	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	MIN	AVG	MAX
	100.00	9.20	11.80	12.10	9.70	11.20	9.20	8.90	9.90	8.80	9.20			
Usability	25.00	2.40	2.90	3.20	2.20	2.80	2.40	2.20	2.30	2.10	2.40	2.10	2.49	3.20
Inf. qual.	25.00	2.40	2.90	2.80	2.50	2.90	2.30	2.30	2.30	2.40	2.20	2.20	2.50	2.90
Serv. inter.	25.00	2.40	3.30	3.40	2.10	2.80	2.20	2.00	2.20	1.90	2.50	1.90	2.48	3.40
Perceptual	25.00	1.90	2.70	2.70	2.80	2.60	2.30	2.40	3.10	2.40	2.10	1.90	2.50	3.10

average and maximum values of the scores obtained by the variants on the cluster level of the AHP hierarchy (see Table 4) were computed and used to create 81 characteristic objects. In total, 3240 pairwise comparisons between the COs were performed to create the MEJ matrix, which allowed to create the fuzzy rule base. A visual representation of the created MEJ matrix is depicted in Fig. 6 and a sample rule is presented in Eq. (1):

$$IF\ C_1 \sim 2.1\ AND\ C_2 \sim 2.2\ AND\ C_3 \sim 3.4\ AND\ C_4 \sim 1.9\ THEN\ P \sim 0.17 \tag{1}$$

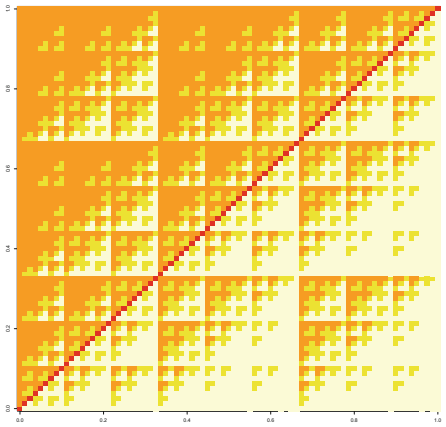

Fig. 6. Graphical representation of the obtained matrix of expert judgment (MEJ).

The fuzzy rule base was eventually used to evaluate the 10 websites. The results of the evaluation are presented in Table 3e. When the obtained ranking is compared with the one based exclusively on the AHP method (Fig. 4g), it can be observed that the top five and the worst two variants do not change their positions. Only slight changes in ranks can be observed for the variants A1, A6 and A10. The Pearson correlation coefficient values between the two rankings are equal to 0.9636 and 0.9962 for the ranks and scores respectively, which shows an almost linear relation between both rankings, thus confirming their similarity. The similarity, in turn, confirms the fact that the obtained ranking is very stable and the probability of a rank reversal phenomenon occurrence for the obtained solution is low.

5 Conclusions

With billion dollars' worth of sales every year, e-commerce is one of the most important sectors of online business. Due to the continuous evolution of software, hardware and preferences of the users, as well as the tough rivalry in the sector, a systematic evaluation of the e-commerce website's quality, usability and user experience is crucial.

The prior approaches to the websites' evaluation were based mainly on survey data. The Authors expanded this approach in [27] to incorporate perceptual data into the evaluation criteria with the use of the PROMETHEE II method. That approach, however, did not fully consider the complexity of the structure of the criteria. Moreover, it was prone to the rank reversal phenomenon. Therefore, the Authors' main contribution in this paper was to extend the formerly proposed approach for e-commerce websites' evaluation to allow extensive criteria structuring and to decrease the probability of rank reversal occurrence in the produced assessments.

An experiment was performed during which 10 most popular e-commerce website were evaluated based on a combined set of survey-based and ET-based criteria. The AHP method was used to perform the websites' evaluations on each level of the criteria structure, while the COMET method was used to avoid the rank reversal of the obtained solution. The performed sensitivity analysis allowed to verify the stability of the produced rankings.

The methodological contribution presented above included the following highlights:

- A unique approach is proposed, which aggregates the data obtained from surveys and perceptual measurements.
- By using the AHP method, a detailed analysis of the structure of criteria and its impact on the final evaluation is enabled.
- By using the COMET method, the stability of the obtained model is increased, thus minimizing the risk of rank reversal.

During the research, possible areas of improvement and future work directions were identified. First of all, the empirical study was based on the perceptual data collected only from three areas of the websites. It would be beneficial to increase the number of areas in the ET experiment, to provide a more comprehensive set of measurements. Moreover, the perceptual data was averaged during the empirical study, whereas in future works the structure of the evaluation criteria could be modified to contain unaggregated measurements from all measured areas, contained in separate groups within the perceptual criteria cluster or, alternatively, forming their own clusters. Last, but not least, the diversity of the perceptual criteria in the model could be increased, for example by including measurements from EEG devices.

References

1. Kemp, S.: Digital in 2017 global overview, January 2017
2. Kemp, S.: Digital in 2016, January 2016
3. Ecommerce News Europe: ecommerce in Europe, September 2016
4. Lindner, M.: Online sales will reach $523 billion by 2020 in the US, January 2016
5. Paul, R.: Just how big is the ecommerce market? you'll never guess!, June 2015
6. Strzelecki, A., Furmankiewicz, M., Ziuziański, P.: The use of management dashboard in monitoring the efficiency of the internet advertising campaigns illustrated on the example of Google analytics. Studia Ekonomiczne **296**, 136–150 (2016)
7. Grigera, J., Garrido, A., Panach, J.I., Distante, D., Rossi, G.: Assessing refactorings for usability in e-commerce applications. Empir. Softw. Eng. **21**(3), 1224–1271 (2016)
8. Sohaib, O., Kang, K.: Assessing web content accessibility of e-commerce websites for people with disabilities (2016)
9. Olsson, M.: Build a Profitable Online Business: The No-Nonsense Guide, 1st edn. Apress, Berkely (2013)
10. Kruger, R., Gelderblom, H., Beukes, W.: The value of comparative usability and UX evaluation for e-commerce organisations. In: CONF-IRM 2016 Proceedings, January 2016
11. Barnes, S.J., Vidgen, R.: The equal approach to the assessment of e-commerce quality: a longitudinal study of Internet Bookstores. In: Web Engineering, p. 161 (2005)
12. Webb, H.W., Webb, L.A.: SiteQual: an integrated measure of web site quality. J. Enterp. Inf. Manag. **17**(6), 430–440 (2004)
13. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: ES-QUAL: a multiple-item scale for assessing electronic service quality. J. Ser. Res. **7**(3), 213–233 (2005)
14. Ahn, T., Ryu, S., Han, I.: The impact of the online and offline features on the user acceptance of internet shopping malls. Electron. Commer. Res. Appl. **3**(4), 405–420 (2005)
15. Barnes, S.J., Vidgen, R.: Measuring web site quality improvements: a case study of the forum on strategic management knowledge exchange. Ind. Manag. Data Syst. **103**(5), 297–309 (2003)
16. Yang, Z., Cai, S., Zhou, Z., Zhou, N.: Development and validation of an instrument to measure user perceived service quality of information presenting web portals. Inf. Manag. **42**(4), 575–589 (2005)
17. Barnes, S., Vidgen, R.: WebQual: an exploration of website quality. In: ECIS 2000 Proceedings, p. 74 (2000)
18. Kabir, G., Hasin, M.: Comparative analysis of topsis and fuzzy topsis for the evaluation of travel website service quality. Int. J. Qual. Res. **6**(3), 169–185 (2012)
19. Brans, J.P., Mareschal, B.: Promethee methods. In: Figueira, J., Greco, S., Ehrogott, M. (eds.) Multiple Criteria Decision Analysis: State of the Art Surveys. ISOR, vol. 78, pp. 163–186. Springer, New York (2005). https://doi.org/10.1007/0-387-23081-5_5
20. Sun, C.C., Lin, G.T.: Using fuzzy TOPSIS method for evaluating the competitive advantages of shopping websites. Expert Syst. Appl. **36**(9), 11764–11771 (2009)
21. Bilsel, R.U., Büyüközkan, G., Ruan, D.: A fuzzy preference-ranking model for a quality evaluation of hospital web sites. Int. J. Intell. Syst. **21**(11), 1181–1197 (2006)

22. Wątróbski, J., Ziemba, P., Jankowski, J., Wolski, W.: PEQUAL-e-commerce websites quality evaluation methodology. In: 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1317–1327. IEEE (2016)
23. Chen, L., Pu, P.: Eye-tracking study of user behavior in recommender interfaces. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 375–380. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13470-8_35
24. Chang Lee, K., Wook Chae, S.: Exploring the effect of the human brand on consumers' decision quality in online shopping: an eye-tracking approach. Online Inf. Rev. **37**(1), 83–100 (2013)
25. Cowen, L., Ball, L.J., Delin, J.: An eye movement analysis of web page usability. In: Faulkner, X., Finlay, J., Détienne, F. (eds.) People and Computers XVI-Memorable Yet Invisible, pp. 317–335. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0105-5_19
26. Bojko, A.: Eye Tracking the User Experience. Rosenfeld Media, Brooklyn (2013)
27. Ziemba, P., Wątróbski, J., Karczmarczyk, A., Jankowski, J., Wolski, W.: Integrated approach to e-commerce websites evaluation with the use of surveys and eye tracking based experiments, pp. 1019–1030, September 2017
28. Wątróbski, J., Jankowski, J., Karczmarczyk, A., Ziemba, P.: Integration of eye-tracking based studies into e-commerce websites evaluation process with eQual and TOPSIS methods. In: Wrycza, S., Maślankowski, J. (eds.) SIGSAND/PLAIS 2017. LNBIP, vol. 300, pp. 56–80. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66996-0_5
29. Saaty, T.L.: What is the analytic hierarchy process? In: Mitra, G., Greenberg, H.J., Lootsma, F.A., Rijkaert, M.J., Zimmermann, H.J. (eds.) Mathematical Models for Decision Support. NATO ASI Series, vol. 48, pp. 109–121. Springer, Heidelberg (1988). https://doi.org/10.1007/978-3-642-83555-1_5
30. McSpadden, K.: You now have a shorter attention span than a goldfish. Time Online Mag. Accessed 7 May 2015 (2016)
31. Weatherhead, R.: Say it quick, say it well-the attention span of a modern internet consumer. The Guardian Online **19** (2012). https://www.theguardian.com/media-network/media-network-blog/2012/mar/19/attention-span-internet-consumer
32. Jankowski, J., Kazienko, P., Wątróbski, J., Lewandowska, A., Ziemba, P., Zioło, M.: Fuzzy multi-objective modeling of effectiveness and user experience in online advertising. Expert Syst. Appl. **65**, 315–331 (2016)
33. Nielsen, J.: Usability Engineering. Elsevier, Amsterdam (1994)
34. International Organization for Standardization: ISO 9241–11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part 11: Guidance on Usability (1998)
35. ISO/IEC 25010: 2011. Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE)-System and software quality models (2011)
36. Hassenzahl, M.: User experience (UX): towards an experiential perspective on product quality. In: Proceedings of the 20th Conference on l'Interaction Homme-Machine, pp. 11–15. ACM (2008)
37. Jankowski, J., Wątróbski, J., Kolomvatsos, K., Kazienko, P.: Fuzzy modeling of user behaviors and virtual goods purchases in social networking platforms. JUCS - J. Univers. Comput. Sci. **22**(3), 416–437 (2016)
38. Fernandez, A., Insfran, E., Abrahão, S.: Usability evaluation methods for the web: a systematic mapping study. Inf. Softw. Technol. **53**(8), 789–817 (2011)

39. Boulding, W., Kalra, A., Staelin, R., Zeithaml, V.A.: A dynamic process model of service quality: from expectations to behavioral intentions. J. Mark. Res. **30**(1), 7 (1993)
40. Barnes, S.J., Vidgen, R.: An evaluation of cyber-bookshops: the webqual method. Int. J. Electron. Commer. **6**(1), 11–30 (2001)
41. Webb, H., Webb, L.: Business to consumer electronic commerce website quality: integrating information and service dimensions. In: AMCIS 2001 Proceedings, p. 111 (2001)
42. La Porte, T.M., Demchak, C.C., Friis, C.: Webbing governance: global trends across national-level public agencies. Commun. ACM **44**(1), 63–67 (2001)
43. Elling, S., Lentz, L., de Jong, M.: Website evaluation questionnaire: development of a research-based tool for evaluating informational websites. In: Wimmer, M.A., Scholl, J., Grönlund, Å. (eds.) EGOV 2007. LNCS, vol. 4656, pp. 293–304. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74444-3_25
44. Ping Zhang, G.M.: User expectations and rankings of quality factors in different web site domains. Int. J. Electron. Commer. **6**(2), 9–33 (2001)
45. Roth, S.P., Tuch, A.N., Mekler, E.D., Bargas-Avila, J.A., Opwis, K.: Location matters, especially for non-salient features-an eye-tracking study on the effects of web object placement on different types of websites. Int. J. Hum.-Comput. Stud. **71**(3), 228–235 (2013)
46. Hu, L., Zhang, W., Xu, Q.: The determinants of online payment method choice: insight from an eye-tracking study. In: WHICEB, p. 80 (2013)
47. Menon, R.V., Sigurdsson, V., Larsen, N.M., Fagerstrøm, A., Foxall, G.R.: Consumer attention to price in social commerce: eye tracking patterns in retail clothing. J. Bus. Res. **69**(11), 5008–5013 (2016)
48. Hernández-Méndez, J., Muñoz-Leiva, F.: What type of online advertising is most effective for etourism 2.0? an eye tracking study based on the characteristics of tourists. Comput. Hum. Behav. **50**, 618–625 (2015)
49. Wang, Q., Yang, S., Liu, M., Cao, Z., Ma, Q.: An eye-tracking study of website complexity from cognitive load perspective. Decis. Support Syst. **62**, 1–10 (2014)
50. Khodambashi, S., Gilstad, H., Nytrø, Ø.: Usability evaluation of clinical guidelines on the web using eye-tracker. Stud. Health Technol. Inform. **228**, 95 (2016)
51. Štrach, P., Slivkin, N.: Adaptation needed: eye-tracking study of cross-cultural differences in perception of B2B websites (2017)
52. Yuan, X., Guo, M., Ren, F., Peng, F.: Usability analysis of online bank login interface based on eye tracking experiment. Sens. Transducers **165**(2), 203 (2014)
53. Albayrak, D., Cagiltay, K.: Analyzing Turkish e-government websites by eye tracking. In: 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), pp. 225–230. IEEE (2013)
54. Pan, B., Zhang, L.: An eyetracking study on online hotel decision making: the effects of images and umber of options (2016)
55. Kim, S., Stoel, L.: Dimensional hierarchy of retail website quality. Inf. Manag. **41**(5), 619–633 (2004)
56. Chmielarz, W., Zborowski, M.: Comparative analysis of electronic banking websites in selected banks in Poland in 2014. In: 2015 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1499–1504. IEEE (2015)
57. Chmielarz, W.: Evaluation of selected mobile applications stores from the user's perspective. Online J. Appl. Knowl. Manag. **3**(1), 21–36 (2015)

58. Chmielarz, W.: Methods of comparative analysis of electronic bankings' websites. case of Poland. In: Chroust, G., Kotsis, G., Risak, V., Rozsenich, N., Zinterhof, P. (eds.) 1-st CEE Symposium on Business Informatics, Osterreichische Computer Gesellschaft, Vienna, pp. 73–84. Citeseer (2009)
59. Zenebe, A., Zhou, L., Norcio, A.F.: User preferences discovery using fuzzy models. Fuzzy Sets Syst. **161**(23), 3044–3063 (2010)
60. Del Vasto-Terrientes, L., Valls, A., Slowinski, R., Zielniewicz, P.: ELECTRE-III-H: an outranking-based decision aiding method for hierarchically structured criteria. Expert Syst. Appl. **42**(11), 4910–4926 (2015)
61. Kaya, T.: Multi-attribute evaluation of website quality in e-business using an integrated fuzzy ahptopsis methodology. Int. J. Comput. Intell. Syst. **3**(3), 301–314 (2010)
62. Huang, J., Jiang, X., Tang, Q.: An e-commerce performance assessment model: its development and an initial test on e-commerce applications in the retail sector of china. Inf. Manag. **46**(2), 100–108 (2009)
63. Guitouni, A., Martel, J.M., Vincke, P., North, P.: A framework to choose a discrete multicriterion aggregation procedure. Defence Research Establishment Valcatier (DREV) (1998)
64. Konys, A., Wątróbski, J., Różewski, P.: Approach to practical ontology design for supporting COTS component selection processes. In: Selamat, A., Nguyen, N.T., Haron, H. (eds.) ACIIDS 2013. LNCS (LNAI), vol. 7803, pp. 245–255. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36543-0_26
65. Chmielarz, W., Zborowski, M.: Aspects of mobility in e-marketing from the perspective of a customer. In: 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1329–1333. IEEE (2016)
66. Chmielarz, W., Zborowski, M.: Comparative analysis of electronic banking websites in Poland in 2014 and 2015. In: Ziemba, E. (ed.) Information Technology for Management. LNBIP, vol. 243, pp. 147–161. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30528-8_9
67. Chmielarz, W., Zborowski, M.: Comparative analysis of e-banking services in Poland in 2016. In: Wrycza, S., Maślankowski, J. (eds.) SIGSAND/PLAIS 2017. LNBIP, vol. 300, pp. 43–55. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66996-0_4
68. Wątróbski, J., Jankowski, J., Ziemba, P.: Multistage performance modelling in digital marketing management. Econ. Sociol. **9**(2), 101 (2016)
69. Wątróbski, J., Ziemba, P., Jankowski, J., Zioło, M.: Green energy for a green city—a multi-perspective model approach. Sustainability **8**(8), 702 (2016)
70. Ziemba, P., Wątróbski, J., Zioło, M., Karczmarczyk, A.: Using the PROSA method in offshore wind farm location problems. Energies **10**(11), 1755 (2017)
71. Verly, C., Smet, Y.D.: Some results about rank reversal instances in the PROMETHEE methods. Int. J. Multicriteria Decis. Mak. **3**(4), 325 (2013)
72. Sałabun, W.: The characteristic objects method: a new distance-based approach to multicriteria decision-making problems: the comet: a new distance-based approach to MCDM problems. J. Multi-Criteria Decis. Anal. **22**(1–2), 37–50 (2015)
73. Saaty, T.L.: How to make a decision: the analytic hierarchy process. Eur. J. Oper. Res. **48**(1), 9–26 (1990)
74. Saaty, T.L.: Decision making with the analytic hierarchy process. Int. J. Ser. Sci. **1**(1), 83–98 (2008)
75. Macharis, C., Springael, J., De Brucker, K., Verbeke, A.: PROMETHEE and AHP: the design of operational synergies in multicriteria analysis.: strengthening PROMETHEE with ideas of AHP. Eur. J. Oper. Res. **153**(2), 307–317 (2004)

76. Sałabun, W., Piegat, A.: Comparative analysis of MCDM methods for the assessment of mortality in patients with acute coronary syndrome. Artif. Intell. Rev. **48**(4), 1–15 (2017)
77. Faizi, S., Sałabun, W., Rashid, T., Wątróbski, J., Zafar, S.: Group decision-making for hesitant fuzzy sets based on characteristic objects method. Symmetry **9**(8), 136 (2017)
78. Bashir, Z., Wątróbski, J., Rashid, T., Sałabun, W., Ali, J.: Intuitionistic-fuzzy goals in zero-sum multi criteria matrix games. Symmetry **9**(8), 158 (2017)
79. Faizi, S., Rashid, T., Sałabun, W., Zafar, S., Wątróbski, J.: Decision making with uncertainty using hesitant fuzzy sets. Int. J. Fuzzy Syst. **3**(16), 1–11 (2017)

Author Index