Association for Information Systems
AIS Electronic Library (AISeL)

Research Papers | ECIS 2019 Proceedings

5-15-2019

BECAUSE YOUR TAXONOMY IS WORTH IT: TOWARDS A FRAMEWORK FOR TAXONOMY EVALUATION

Daniel Szopinski
Paderborn University, daniel.szopinski@wiwi.uni-paderborn.de

Thorsten Schoormann
University of Hildesheim, thorsten.schoormann@uni-hildesheim.de

Dennis Kundisch
Paderborn University, dennis.kundisch@wiwi.uni-paderborn.de

Follow this and additional works at: https://aisel.aisnet.org/ecis2019_rp

Recommended Citation
Szopinski, Daniel; Schoormann, Thorsten; and Kundisch, Dennis, (2019). "BECAUSE YOUR TAXONOMY IS WORTH IT: TOWARDS A FRAMEWORK FOR TAXONOMY EVALUATION". In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019. ISBN 978-1-7336325-0-8 Research Papers.
https://aisel.aisnet.org/ecis2019_rp/104

This material is brought to you by the ECIS 2019 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

BECAUSE YOUR TAXONOMY IS WORTH IT: TOWARDS A FRAMEWORK FOR TAXONOMY EVALUATION

Research paper

Szopinski, Daniel, Paderborn University, Paderborn, Germany,
daniel.szopinski@wiwi.uni-paderborn.de

Schoormann, Thorsten, University of Hildesheim, Hildesheim, Germany,
thorsten.schoormann@uni-hildesheim.de

Kundisch, Dennis, Paderborn University, Paderborn, Germany,
dennis.kundisch@wiwi.uni-paderborn.de

Abstract

Taxonomies constitute one fundamental type of artefact in design science, describing and classifying existing or future objects of a domain. Taxonomies support researchers and practitioners with analysing and understanding a domain, which in turn is a prerequisite for theory building. Despite the increasing interest in taxonomies (and methodological guidance for building them), there is hardly any guidance for researchers on how to rigorously evaluate taxonomies. Based on a literature analysis, this study sheds light on the question of whether, when, and how researchers currently evaluate taxonomies. We critically synthesize and comprehensively review 306 articles that are concerned with taxonomies. Surprisingly, we find that taxonomies are rarely evaluated in IS research, nor is there any consistency in terms of methods used for evaluations. We describe the methods used by IS researchers to evaluate taxonomies after taxonomy building has been completed. Being the first to systematically analyse taxonomy evaluation, we propose a preliminary version of a framework for taxonomy evaluation which enables researchers to choose among the wide range of taxonomy evaluation methods available. Our study advances an informed and purposeful evaluation of taxonomies and contributes to bridging the gap between abstract design science evaluation strategies and concrete taxonomy evaluation methods.

Keywords: Taxonomy, Evaluation, Design science research, Research methodology.

1 Introduction

The evaluation of artefacts such as taxonomies is a central and critical part of design science (March and Smith, 1995; Hevner et al., 2004). Beyond that, taxonomies are of interest to researchers from other fields such as biology (e.g., Eldredge and Cracraft, 1980; Sokal and Sneath, 1963) or social sciences (e.g., Bailey, 1994), and not only for research itself but also for general discussion (e.g., to structure a domain of interest) and pedagogical purposes (e.g., Miller and Roth, 1994). Like all artefacts in design science (Gregor and Hevner, 2013), taxonomies require evaluation. Surprisingly, despite the key role played by taxonomies as structure-giving artefacts in the exploitation of new research fields in information systems (IS) (e.g., Kuechler and Vaishnavi, 2008; Galliers and Land, 1990) and the development of novel software artefacts (e.g., Morana et al., 2017), taxonomies are rarely evaluated, nor is there any consistency in terms of methods used for evaluations. The necessity to evaluate taxonomies is not only a requirement of design science research (DSR) in general (e.g., Peffers et al., 2012; Venable et al.,

2016), but also of research on taxonomies. This study seeks to shed light on the question of *whether, when, and how researchers currently evaluate taxonomies* and thereby lays ground for future taxonomy evaluation. Based on a literature analysis we offer suggestions for a framework providing guidance for choosing among the wide range of taxonomy evaluation methods available, and in this way advance an informed and purposeful evaluation of taxonomies.

Taxonomies are artefacts that describe and classify existing or future objects of a domain and, as such, enable academics as well as practitioners to describe and analyze a domain (Nickerson et al., 2013). By helping explain similarities and differences among objects, taxonomies form an important prerequisite to understanding a domain of interest. By structuring and organizing a body of knowledge, taxonomies allow to study the relationships among concepts "with all the potential advantages that brings for the advancement of the field" (Glass and Vessey, 1995, p. 85) and, therefore, to hypothesize about these relationships. For this reason, taxonomies distillate findings and existing knowledge, in turn facilitating discussions on the (past, present, and future) course of research in a domain of interest.

As a result of our study we found that the method proposed by Nickerson et al. (2013) for building taxonomies is a widely accepted method in IS. Previously, researchers have relied on methods borrowed from other disciplines such as the Baily's (1994) method prevalent in the social sciences. Since Nickerson et al. (2013) proposed their method, IS researchers have been able to rely on a rigorous method for systematically developing taxonomies. Over the past five years, this method has been widely applied in IS research and used for developing taxonomies in domains as diverse as crowdsourcing (Geiger et al., 2012), the internet-of-things (Püschel et al., 2016), digital service design techniques (Liu et al., 2016) as well as carsharing (Remane et al., 2016), telemedicine (Peters et al., 2015), cloud computing (Labes et al., 2013), FinTech start-ups (Gimpel et al., 2017), and crowdfunding (Haas et al., 2014). The taxonomies mentioned also illustrate the diversity of IS research (Rai, 2017). As part of our study, we identify 196 articles that have built a taxonomy, noting that, of these, only 61 articles have actually evaluated their taxonomy after completing it. This suggests that the majority of taxonomy building projects do not evaluate their resulting taxonomies. From our careful reading we assume the following possible reasons: First, the lack of methodological guidance is obviously a barrier for authors towards properly evaluating their taxonomies. Second, taxonomies are often built as a new class of objects emerges to structure the domain of these new objects. Evaluation per se is notoriously difficult, but particularly challenging at the beginning of an emerging research field. Third, both criteria for the evaluation of taxonomies are difficult to measure, and criteria for the objects describing taxonomies often do not (yet) exist. Forth, evaluating a taxonomy's usefulness requires to "observe its use over time" (Nickerson et al. 2013, p. 342) which literally takes time. For example, taxonomies are evaluated in the transition of advancing a manuscript from a conference to a journal version (e.g., Szopinski et al., 2017 and Szopinski et al., 2019). The time period between taxonomy building and taxonomy evaluation here is two years.

Despite the widespread adoption of Nickerson et al. (2013)'s method to *build* taxonomies, there is no such equivalent for *evaluating* taxonomies. Rather, there are numerous and varying attempts that apply already existing methods, used for evaluating artefacts other than taxonomies, and adopt them for the purpose of evaluating taxonomies (e.g., expert interviews). This heterogeneity is not problematic per se, but it makes it unnecessarily difficult for researchers to get an overview across the wide range of taxonomy evaluation methods and identify those that suit a taxonomy's particular context. Hence, a comprehensive overview of how to evaluate such taxonomies is still missing. The demand for evaluating taxonomies is evident in the fact that we identified numerous studies that develop a taxonomy and leave the evaluation for future research. For example, Liu et al. (2017) recently found for the domain of gamification "that the existing taxonomies lack rigor and need further development" (Bui et al., 2015 cited in Liu et al., 2017, p. 1013). Likewise, Premkumar et al. (2005) note that "IS research mostly uses existing taxonomies without empirically validating them" (Choudhury, 1997 cited in Premkumar et al., 2005, p. 259).

Typically, DSR distinguishes ex post and ex ante evaluation (e.g., Venable et al., 2016). Applied to this study's context, this refers to the evaluation of taxonomies during (i.e., ex ante) and after (i.e., ex post) completion of building a taxonomy. For example, verifying certain criteria while building a taxonomy can improve its usefulness (e.g., through the ending conditions that are an integral part of the taxonomy

development method proposed by Nickerson et al. 2013). The focus of this study lies in the evaluation of a taxonomy "after [an artefact] has been acquired, designed, constructed, or implemented" (Klecun and Cornford, 2005), i.e., ex post evaluation after taxonomy building has been completed.

Our study aims to help researchers by providing a comprehensive overview of the different approaches that can be pursued for taxonomy evaluation. Our contention is that the gap between abstract design science evaluation strategies and concrete taxonomy evaluation methods makes it difficult for researchers to evaluate their own taxonomies. This difficulty is evident in the numerous calls for taxonomy evaluation that, surprisingly, do not make use of the valuable reservoir of evaluation strategies that DSR actually holds. Furthermore, Nickerson et al. (2013) emphasize that "[a] taxonomy needs to be evaluated for its usefulness" while acknowledging that "determining sufficient conditions for usefulness is difficult" (Nickerson et al. 2013, p. 346). We aim to respond to the call of researchers who find that "the main development cycle [for taxonomies] does not include the evaluation step, leaving it open for researchers" (Chasin et al., 2018, p. 6). This research gap is, for example, also evident in articles, where researchers explicitly call for the evaluation of taxonomies in subsequent research (e.g., Kazan et al., 2018; Zschech, 2018; Krieger and Drews, 2018).

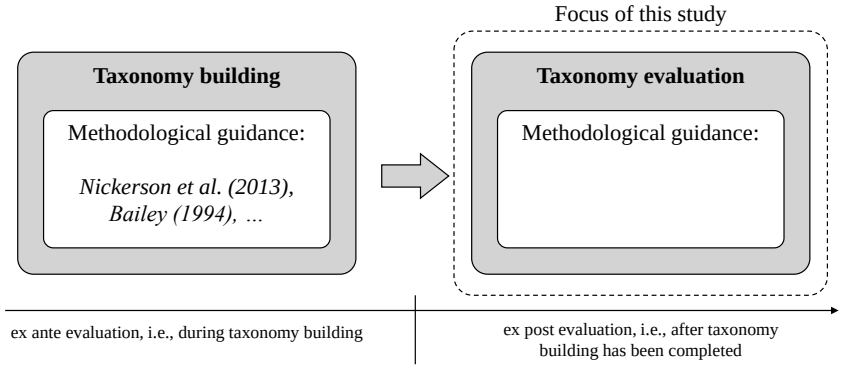

Figure 1. Focus of this study.

To pursue the research question of *whether, when, and how researchers currently evaluate taxonomies*, we synthesize and report on different methods that researchers employ for evaluating taxonomies. Our contribution would prevent researchers from having to revisit each taxonomy building project all over again in order to discuss how to evaluate their taxonomy. This is the issue we are seeking to address' and focus it on the similar issue raised by Peffers et al. (2012, p. 398), i.e. lack of "guidance for how to perform the evaluation, more specifically, what evaluation methods to use". In addition, we provide a basis for future research on the identification of appropriate evaluation approaches for different contexts and types of taxonomies. By providing a systematic overview and suggesting a preliminary version of a framework that captures taxonomy evaluation strategies, we seek to leverage the full potential of taxonomies as a structure-giving artefact for (past, present, and future) IS research.

2 Research background

2.1 Evaluation in design science

Evaluating artefacts such as taxonomies is a crucial activity in any design science project, to ensure the rigorous application of research methods and the robustness of research results obtained in such projects (e.g., Venable et al., 2016; Vaishnavi and Kuechler, 2004; Prat et al., 2015; Sonnenberg and vom Brocke, 2012; Peffers et al., 2012; Cleven et al., 2009; Hevner et al., 2004; March and Smith, 1995). Depending on the nature of the artefact, appropriate evaluation criteria and methods need to be employed (Prat et al., 2015). Evaluation in design science typically seeks to answer questions like, for example: How well does an artefact achieves its expected utility, quality, and efficacy? What makes an artefact work? How does one artefact perform in comparison to others? (Venable et al., 2016).

To answer this kind of question, researchers employ different evaluation strategies. The *Framework for Evaluation in Design Science* (FEDS) captures evaluation strategies and distinguishes them along two dimensions (Venable et al., 2016): First, *why to evaluate* considers the functional purpose of an evaluation which can be either formative (while the artifact is in development, Wiliam and Black, 1996) or summative (at the end of its development, Wiliam and Black, 1996). The second dimension, *how to evaluate*, refers to the paradigm chosen for evaluation, from an artificial evaluation (e.g., to evaluate in the form of laboratory experiments, simulations, or criteria-based analysis) to a naturalistic evaluation (e.g., to evaluate by exploring the performance of an artefact in its real environment). Based on these two dimensions, Venable et al. (2016) describe four strategies ranging from basic to more comprehensive and robust evaluations: *Quick & simple* (for artefacts with low social or technical risks), *human risk & effectiveness* (for artefacts with social or user-oriented risks), *technical risk & efficacy* (for artefacts with technical risks), and *purely technical artefact* (for artefacts without social or human aspects). Applied to this study's object of investigation, all four strategies are potentially suitable for evaluating taxonomies depending on the objects the taxonomies intend to capture: *Quick & simple* (e.g., for a taxonomy on hotel booking apps, Gibbs et al., 2016), *human risk & effectiveness* (e.g., for a taxonomy on virtual reality in healthcare, Almufareh et al., 2018), *technical risk & efficacy* (e.g., for a taxonomy on smart things in the internet-of-things, Püschel et al., 2016), and *purely technical artefacts* (e.g., for a taxonomy on technical characteristics of blockchain systems, Holler et al., 2017).

To operationalize abstract evaluation strategies in design science, the literature already suggests various context-independent (i.e., artefact-independent) methods. Sonnenberg & vom Brocke (2012), for example, provide a framework that assigns evaluation methods (e.g., expert interview or experiment) and evaluation criteria (e.g., consistency or level of detail) to four consecutive evaluation activities. Hevner et al. (2004), as another example, suggest five types of evaluation methods (i.e., observational, analytical, experimental, testing, and descriptive). Peffers et al. (2012), in addition, suggest eight types: Logical argument (face validity), expert evaluation (assessment by one or more experts), technical experiment (performance of an algorithm), subject-based experiment (test involving subjects), action research (use artefact in real world), prototype (implementation of an artefact to demonstrate utility), case study (apply artefact in real world, evaluate effects), and illustrative scenario (apply artefact to synthetic or real-world situation). While most studies provide a list of typically used methods that are conceptually derived, Peffers et al. (2012) derived their evaluation types empirically by reviewing a sample of design science projects. In this study, we follow Peffers et al. (2012), as their contribution provides guidance on evaluation methods with a level of detail that seems to be appropriate to the first investigation of taxonomy evaluation methods (e.g., distinguishing case studies and illustrative scenarios that are often misleadingly used interchangeably in IS research, Prat et al., 2015).

2.2 Taxonomies in information systems

Classifying objects is of great relevance to organizing knowledge (Wand et al., 1995) as well as to understanding and analyzing complex domains (Nickerson et al., 2013). This applies to many areas in IS such as, for example, conceptual modeling, data management and mining, object-oriented systems design, and ontological engineering (Parsons and Wand, 2008). Without classifying objects, researchers would have to perceive each object as unique, and thus, would be overwhelmed by the diversity of what he or she experiences from objects (Lakoff, 1987). To facilitate the classification of objects, schemes can be used that help to, for example, structure concepts and their relationships as well as characterize commonalities of and differences between domains (Bailey, 1994; Wand et al., 1995; Nickerson et al., 2013). Although researchers use different terms for such schemes, two types can be distinguished, namely *typology* and *taxonomy* (Paré et al., 2015; Bailey, 1994). While a typology is usually derived deductively (i.e., conceptual classification), a taxonomy can be derived both deductively and inductively (i.e., empirical classification). Nonetheless, because these terms are often used interchangeably and taxonomies are considered as the prevailing form in IS research (Nickerson et al., 2013; Schmidt-Kraepelin et al., 2018), this study also uses the term *taxonomy* for classifications that are derived both deductively and inductively.

Beyond the descriptive and classifying purpose of taxonomies, they can also serve as *theory for analyzing* which, according to Gregor's classification of theory types, is the most basic form of theory (so-called taxonomic theory, Nickerson et al., 2017), and as such the necessary foundation for more advanced theories (i.e., theories of explanation and prediction) (Gregor, 2006). In this vein, improvements in how to understand a domain through taxonomies can also lead to basic theories that attempt to describe, and allow us to analyze, the objects that play roles in certain phenomena and thereby contribute to theory building (Doty and Glick, 1994). This point is also stressed by Bapna et al. (2004, p. 23), who state that "a robust taxonomy can then be used to perform ex post theory building".

To ensure that taxonomy building is rigorous, Nickerson et al. (2013) propose a method for systematically developing taxonomies. In a nutshell, the method comprises seven steps combining both inductively and deductively developing taxonomies in an iterative manner (see Nickerson et al., 2013 for a detailed description). Recognizing taxonomies as artefacts (i.e., as product of a design science process), taxonomies require being rigorously built and evaluated. As an integral part of the method proposed by Nickerson et al. (2013) there is some guidance on how to evaluate taxonomies while developing a taxonomy (i.e., ex ante evaluation). Therefore, so-called ending conditions determine when taxonomy building is completed. Objective ending conditions define when a taxonomy qualifies as such. Subjective ending conditions ensure that a taxonomy is concise, robust, comprehensive, extendible, and explanatory. However, with regard to the evaluation of a taxonomy after building has been completed (i.e., ex post evaluation), there is hardly any guidance. Nickerson et al. (2013) state that evaluating a taxonomy once building has been completed "may come down to seeing if others use it [...] and speculate on [its] potential use" (Nickerson et al., 2013, p. 346-347).

3 Method

With the goal of providing a comprehensive overview of *how researchers currently evaluate taxonomies*, we collated a literature with 395 articles in three distinct steps (see Table 1): (1) A citation analysis of the most widely used method for taxonomy development in IS, namely the method developed by Nickerson et al. (2013), (2) a keyword search in the *AIS Senior Scholars' Basket of Journals*, and (3) a keyword search in the AIS eLibrary for articles published in the proceedings of the *International* and *European Conference on Information Systems*.

For the citation analysis of the seminal article by Nickerson et al. (2013), we used the citation-indexing service by Google Scholar (step 1). We sought to identify relevant articles of high quality by including in step 1 only articles that cite Nickerson et al. (2013), are written in English, and are published in peer-reviewed journals or conference proceedings. Ending in October 2018, the citation analysis revealed 252 citations. We excluded 15 citations that are not in English and 25 that are published neither in a peer-reviewed journal nor in conference proceedings (e.g., theses, presentations, or working papers). Furthermore, we excluded three duplicates (that Google Scholar indexes multiple times) as well as one article mistakenly indexed as a citation of Nickerson et al. (2013). As a result, we obtained 191 articles that cite Nickerson et al. (2013) and match the aforementioned criteria.

For identifying articles beyond the scope of the method proposed by Nickerson et al. (2013), we followed the recommendations from Webster and Watson (2002) by searching *"taxonomy"* (search string) for *title, abstract, and keywords* (search fields) *up until and including October 2018* (timespan) in the *AIS Senior Scholars' Basket of Journals* (step 2) as well as in the proceedings of the *International* and *European Conference on Information Systems* (step 3). In total, the keyword search in step 2 and 3 reveals 143 articles. We removed 24 duplicates, including all articles which cite Nickerson et al. (2013) and have been already identified in step 1. Furthermore, we excluded five articles that are neither journal nor conference articles. Therefore, step 2 and 3 lead to 114 additional articles that are concerned with taxonomies although not referring to Nickerson et al. (2013). Phase 1 results in 306 identified articles for which we subsequently undertook a full-text review in phase 2 (see Table 1).

Phase 1: Identification of relevant articles					Phase 2: Analysis of relevant articles	
Step	Total number of articles	Number of articles after removing articles that are…			Articles that build a taxonomy	Articles that evaluate a taxonomy
		Non-English				
			Neither journal nor conference articles			
				Duplicates		
(1) Citation analysis of Nickerson et al. (2013)	252	237	199	191	126	45
(2) Keyword search in AIS Senior Scholars' Basket of Journals	70	70	68	63	47	12
(3) Keyword search in conference proceedings (ICIS and ECIS)	73	73	71	52	23	4
Total number of articles	395	380	338	306	196	61

Table 1. Identification of relevant articles.

For preparing the analysis of articles and determining their relevance for answering the underlying research question of this study, the first and second author independently read each article and determined (i.e., coded) whether an article is concerned with the evaluation of a taxonomy in a non-trivial way. Upon our careful reading, we identified 196 articles in which researchers develop a taxonomy and 61 articles where researcher evaluate a taxonomy. We validated the coding of the articles by calculating the interrater reliability. For the coded attributes, the following percentages of agreement were obtained: *taxonomy development* (98 %) (i.e., researchers build a taxonomy), *call for taxonomy evaluation* (98 %) (i.e., researchers demand future taxonomy evaluation), and *taxonomy evaluation* (97 %) (i.e., researchers evaluate a taxonomy). Reassuringly, the interrater agreement between the authors for the selected articles is very high throughout, with values for Cohen's Kappa and Krippendorf's Alpha between 0.8799 and 0.9478 (values of 0.8 signal large interrater agreement, Lombard et al., 2002). In the very few cases of disagreement, the co-authors discussed their opinions to come to a joint verdict about inclusion or exclusion and coding (Paré et al., 2015).

4 Towards a framework for taxonomy evaluation

4.1 Identification of taxonomy evaluation methods

In analysing the collated literature of 395 articles, we identified 196 articles in which researchers develop one or more taxonomies (e.g., Keller and König, 2014 derived two taxonomies in one article). In most of the articles a taxonomy is developed, and in a minority, taxonomies are both built and subsequently evaluated (e.g., Püschel et al., 2016). In some studies, researchers spread reporting about the building (Herterich et al., 2015) and the evaluation (Herterich et al., 2016) across several articles. A majority of 116 articles employ the method proposed by Nickerson et al. (2013). An additional six articles adapt the method by Nickerson et al. (2013), extending the method for validating the taxonomy through a series of case studies to promote the circulation of the taxonomy among relevant stakeholders. Another 81 articles rely on other methods for taxonomy building (e.g., Ackermann et al., 2011 employ iterative grouping and sorting of objects). Out of the 196 articles that are significantly concerned with developing taxonomies, only 61 articles evaluate their taxonomy.

In Table 2 we report on those articles which evaluate at least one taxonomy. In so doing, we distinguish between *taxonomy building* (i.e., how researchers developed a taxonomy), and *taxonomy evaluation* (i.e., how researchers evaluated a taxonomy in that article, relying on the evaluation methods types described by Peffers et al., 2012).

Reference	Tax. building		Taxonomy evaluation										
				Expert Evaluation							Illustrative Scenario		
	Nickerson et al. (2013)	Other methods	Logical argument	Expert inter-view	Focus group	Survey	Delphi echnique	Sorting	Log diary	Case study	Real-world objects	Research about real-world objects	Action research
Addas and Pinsonneault (2015)	-	■	-	■	-	-	-	-	■	-	-	-	-
Agogo et al. (2018)	-	-	-	-	-	-	-	-	-	-	-	■	-
Aksulu and Wade (2010)	-	■	-	-	-	-	-	-	-	-	-	■	-
Almufareh et al. (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Bapna et al. (2004)	-	■	-	-	-	-	-	-	-	-	■	-	-
Bärenfänger et al. (2016)	■	-	-	-	-	-	-	-	-	-	■	-	-
Barn and Barn (2016)	■	-	-	-	-	-	-	-	-	-	■	-	-
Bock and Wiener (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Botha et al. (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Chasin et al. (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Cledou et al. (2018)	-	■	-	-	■	-	-	-	-	-	-	-	-
Conboy (2009)	-	■	-	-	-	-	-	■	-	-	-	-	-
Diniz et al. (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Fellmann et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Fteimi and Lehner (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Gao et al. (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Ge and Gretzel (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Geiger et al. (2012)	■	-	-	-	-	-	-	-	-	-	-	■	-
Gerber et al. (2017)	-	-	-	-	-	-	-	-	-	-	-	■	-
Gibbs et al. (2016)	■	-	-	-	-	-	-	-	-	-	■	-	-
Gimpel et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Gregor (2006)	-	■	-	-	-	-	-	-	-	-	-	■	-
Gutierrez et al. (2008)	-	■	-	-	-	-	-	-	-	-	-	■	-
Hallerstede et al. (2015)	■	-	-	-	-	-	-	-	-	-	■	-	-
Herterich et al. (2016)	■	-	-	-	■	-	-	-	-	-	■	-	■
Herzfeldt et al. (2012)	-	■	-	■	-	-	-	-	-	■	-	-	-
Holler et al. (2017)	■	-	-	-	■	-	-	-	-	■	-	-	-
Hummel et al. (2016)		■	-	-	-	-	-	-	-	-	-	■	-
Jarvinen (2000)	-	■	■	-	-	-	-	-	-	-	-	-	-
Jöhnk et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Keller and König (2014)	-	■	-	■	-	-	-	-	-	-	-	-	-
Khalilijafarabad et al. (2016)	■	-	-	-	-	-	-	-	-	-	-	■	-
King and Sethi (1999)	-	■	-	-	-	■	-	-	-	-	-	-	-
Küpper et al. (2014)	-	■	-	-	-	-	-	■	-	-	-	-	-
Labazova et al. (2019)	■	-	-	-	-	-	-	-	-	-	■	-	-
Larsen (2003)	-	■	-	-	-	-	-	■	-	-	-	■	-
Melas et al. (2017)	-	-	-	-	-	■	-	-	-	-	-	-	-
Nambisan et al. (1999)	-	■	-	-	-	-	■	-	-	-	-	-	-
Nickerson et al. (2009)	-	■	-	-	-	-	-	-	-	-	■	-	-
Oberländer et al. (2018)	■	-	-	-	-	-	-	-	-	-	■	-	-
Prat et al. (2015)	■	-	-	-	-	-	-	-	-	-	-	■	-
Püschel et al. (2016)	■	-	-	-	-	-	-	-	-	-	■	-	-
Raza et al. (2018)	■	-	-	-	-	-	-	-	-	■	-	-	-
Roeder et al. (2018)	-	-	-	-	-	-	-	-	-	-	■	-	-
Schäffer and Stelzer (2017)	■	-	-	■	-	-	-	-	-	-	-	-	-
Schneider et al. (2014)	■	-	-	-	-	-	-	-	-	-	■	-	-
Seyffarth et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Siering et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-

Smith et al. (2014)	■	-	-	-	-	-	-	-	-	-	■	-	-
Snow and Reck (2016)	■	-	-	-	-	■	-	-	-	-	-	-	-
Stoeckli et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Strasser (2017)	■	-	-	-	-	-	-	-	-	-	■	■	-
Strode (2016)	-	■	-	-	-	-	-	-	-	-	■	-	-
Szopinski et al. (2017)	■	-	-	-	-	■	-	-	-	-	-	-	-
Thiebes et al. (2017)	■	-	-	-	-	-	-	-	-	-	■	-	-
Tilly et al. (2017)	■	-	■	-	-	-	-	-	-	-	-	-	-
Tönnissen and Teuteberg (2018)	■	-	-	-	-	■	-	-	-	-	-	-	-
Tsatsou et al. (2010)	-	■	-	-	-	-	-	-	-	-	■	-	-
Werder and Wang (2016)	■	-	-	■	-	-	-	■	-	-	-	-	-
Williams et al. (2008)	-	■	-	-	-	-	-	-	-	-	■	-	-
Zrenner et al. (2017)	■	-	-	-	-	-	-	-	-	■	-	-	-

Table 2. Articles that evaluate a taxonomy (■ addressed; - not addressed).

In the following, we briefly outline each of the identified evaluation methods in respect of the application context of taxonomy evaluation.

4.1.1 Logical argument

Logical argument is an evaluation method that refers to an "argument with face validity" (Peffers et al. 2012, p. 402). In contrast to more formal approaches such as mathematical models, this evaluation method is less formal (Wilde and Hess, 2007). Our sample entails two articles that evaluate a taxonomy by means of logical argument. Both studies map and compare their resulting taxonomies to existing research findings and critically reflect upon them by discussing commonalities and differences. For example, the constituting characteristics of a taxonomy are validated by anchoring them into a well-received framework of the domain the taxonomy intends to capture (Tilly et al., 2017). Another example evaluates comprehensiveness, parsimony, and usefulness of a resulting taxonomy by comparing it with three other already developed related taxonomies (Jarvinen, 2000).

4.1.2 Expert evaluation

Expert interviews entail evaluating an artefact through experts who have proven expertise and experience in a (for the taxonomy relevant) domain. Although time-consuming and resource-intensive, such interviews enable researchers to directly collect data (mostly face-to-face) on a taxonomy (Bhattacherjee, 2012). Our sample entails five articles that evaluate a taxonomy by means of expert interviews. For this, researchers employ unstructured (e.g., Herzfeldt et al., 2012) and semi-structured interviews (e.g., Keller and König, 2014). Furthermore, expert interviews allow to reflect evaluations on taxonomies by academics as well as practitioners (e.g., Schäffer and Stelzer, 2017). Interviewers (i.e., researchers) typically ask questions such as, for example, "Is the taxonomy adequate and complete? Are all relevant objects included in the taxonomy? Would you suggest modifying the taxonomy? Which dimensions and characteristics should be deleted? Which dimensions or characteristics should be added?" (Schäffer and Stelzer, 2017, p. 11). Our sample shows articles where interviews were conducted with between two and 21 experts for 30 to 45 minutes. The criteria that researchers seek to evaluate by expert interviews are comprehensibility, completeness, perceived usefulness of the taxonomy, as well as the level of abstraction of characteristics and dimensions. This may result in changes to the taxonomy with regard to renaming or adding new characteristics and dimensions.

Focus groups are an alternative to expert interviews, in which a group of usually between six to ten interviewees jointly elaborates the evaluation of a taxonomy. By enabling the sharing of views amongst group members, focus groups facilitate collective reflection on the taxonomy and are able to inspire/trigger new thoughts (Tremblay et al., 2010). Our sample includes three articles where a taxonomy is evaluated by means of one or more focus groups (e.g., Holler et al., 2017) with academics as well as practitioners of different seniority (e.g., Cledou et al., 2018; Herterich et al., 2016). The criteria that

researchers seek to evaluate by using focus groups are comprehensiveness, robustness, understandability, extensibility, and wording of the taxonomy's characteristics and dimensions. Furthermore, members of focus groups jointly classify new objects (i.e., those that were not used for building the taxonomy) to validate a taxonomies' completeness (e.g., Cledou et al., 2018; Holler et al., 2017).

Surveys are an easy and cost-effective way (in terms of money, time and effort) of collecting large data sets through standardized questionnaires. Their advantage over expert interviews and focus groups are that surveys allow the respondents to reply at his or her convenience (Bhattacherjee, 2012) whilst involving a large number of respondents (e.g., 1,015 sent questionnaires with 604 responses to test five characteristics and dimensions of a taxonomy in Melas et al., 2017). Our sample entails five articles that evaluate a taxonomy by means of a survey. Surveys on taxonomies typically collect both qualitative (e.g., open and/or closed questions, Tönnissen and Teuteberg, 2018) and quantitative (e.g., Likert-scales, Snow and Reck, 2016) data. In doing so, researchers seek, for example, to find a majority view on the naming of a characteristic or dimension, or on the assignment of a characteristic to a (sub-)dimension (e.g, Tönnissen and Teuteberg, 2018; Snow and Reck, 2016).

The **Delphi technique** is "a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem" and obtain the most reliable consensus among individual group members (Okoli and Pawlowski, 2004, p. 16, van de Ven and Delbecq, 1974). Applied to this study's context, one key and complex issue is the evaluation of a taxonomy. Unlike focus groups, the Delphi technique avoids direct confrontation of group members. To evaluate taxonomies, researchers typically apply the Delphi technique in a multi-stage iterative process. Our sample comprises one article that evaluates a taxonomy by means of a Delphi study (Nambisan et al., 1999). In this study, group members were requested to assign objects to (sub-) dimensions of the taxonomy. After each stage, individual opinions are made available to all group members and each group member is asked to indicate whether he or she concurs with the group opinion, and if not, their reason for differing. This procedure is repeated until a consensus is reached (e.g., Nambisan et al., 1999 needed three rounds to evaluate their taxonomy).

Sorting is a procedure to validate taxonomies with a dedicated focus on the structure of the taxonomy's (sub-) dimensions. Our sample contains four articles that apply this method. Usually several experts receive a taxonomy and are required to assign its characteristics to its (sub-) dimensions. The aim of the sorting procedure is to challenge the (sub-) dimensions derived (whether empirically or conceptually) by the builders of the taxonomy (e.g., Werder and Wang, 2016). Sorting is used both for taxonomies with comparatively few (sub-) dimensions (e.g., four dimensions in Küpper et al., 2014) and for taxonomies with comparatively many (sub-) dimensions (e.g., 70 dimensions in Larsen, 2003). Since different experts usually evaluate the taxonomy during sorting, researchers often also calculate the interrater reliability among the experts. Sorting may be supplemented by Q-sorting that captures complex opinions and attitudes from a subjective perspective and, for that reason, is also applied in IS research (Moore and Benbasat, 1991; Petter et al., 2007). In this study's context, "the taxonomy's usefulness [is determined] in terms of construct validity and reliability" (Oberländer et al., 2018, p. 491). Thus, Q-sorting helps to describe and understand the variety of subjective assignments of characteristics to a taxonomy's (sub-) dimensions (e.g., Oberländer et al., 2018).

Log diaries allow evaluating taxonomies directly in the relevant application environment. This method is particularly suitable for taxonomies whose objects are already used in practice. Our sample counts one article that evaluates a taxonomy by mean of log diaries (Addas and Pinsonneault, 2015). In this study, academics as well as practitioners were asked to look for information technology interruptions (i.e., the object of the taxonomy) during their everyday work. The taxonomy to be evaluated intends to capture the relevance and structure of such interruptions and their consequences upon performance. An online logbook is available to the judges for the evaluation. The online logbook consists of open-ended questions and closed-ended statements supplemented by Likert-type statements. The judges describe each interruption promptly after the occurrence of an interruption on a maximum of one single page. The authors interpreted the responses and evaluated whether the taxonomy is able to capture information technology interruptions (i.e., the object of the taxonomy) as intended.

4.1.3 Action Research

Action research is an evaluation method which assumes that "phenomena are best understood by introducing interventions or 'actions' into those phenomena and observing the effects of the actions" (Bhattacherjee, 2012, p. 40) and in this way helps "to recognize that the artefact emerges from interaction with the organizational context" (Sein et al., 2011, p. 37). Applied to this study's context, action research allows to directly deal with the academics' intervention (e.g., by introducing a taxonomy) and study the effects of this intervention (e.g., the usefulness of a taxonomy in pursuing a certain task) in an organizational real-world situation (Recker, 2013). Our sample comprise one article which evaluates a taxonomy through action research (Herterich et al., 2016). In following guidelines for action design science (Sein et al., 2011), Herterich et al. (2016) evaluate the usefulness and relevance of their taxonomy in terms of strategic digitization and service innovation projects of two case organizations.

4.1.4 Case study

Case studies "involve intensive research on a phenomenon (a case) within its natural setting (one or more case sites) over a period of time" and are the most popular form of qualitative method in IS (Recker, 2013, p. 95; Yin, 2017). Our sample includes four articles which use this evaluation method, primarily to demonstrate a taxonomy's validity, applicability and usefulness (e.g., Conboy, 2009; Holler et al., 2017). For data collection on taxonomy evaluation, case studies often also make use of other methods, such as focus groups with several rounds over several months (e.g., Conboy, 2009), one-time workshops, and semi-structured interviews with potential users of a taxonomy in organizations (e.g., Zrenner et al., 2017; Holler et al., 2017). This allows researcher to evaluate whether the intentions ascribed to the taxonomies are valid in real-world situations, i.e., to facilitate analysis (e.g., Conboy, 2009) and decision-making (e.g., Zrenner et al., 2017).

4.1.5 Illustrative scenario

Illustrative scenarios apply "an artefact to a synthetic or real-world situation [and thereby aim at] illustrating suitability or utility of the artefact" (Peffers et al., 2012, p. 402). Such scenarios differ from case studies by not only applying to real-world situations, but also to artificial situations specifically designed for evaluation purposes. Furthermore, illustrative scenarios do not apply an artefact (i.e., a taxonomy) over a time period. Our sample comprises 43 articles that evaluate a taxonomy by this method, which makes this the prevailing evaluation method for taxonomies. Illustrative scenarios employ two distinct types of objects: Real-world objects and existing research about real-world objects.

Apply a taxonomy to real-world objects. Researchers who evaluate a taxonomy by applying real-world objects to this taxonomy (32 articles of the sample fall into this category), and first identify relevant objects. As diverse as these taxonomies are, so are the objects. Hence, there is a wide range of objects (e.g., mobile applications in Nickerson et al., 2009, sharing platforms in Chasin et al., 2018, blockchain-based systems in Labazova et al., 2019, virtual reality tools in Almufareh et al., 2018, and cloud computing services in Gao et al., 2018). Some researchers evaluate taxonomies with a few selected objects (e.g., two objects in Hallerstede et al., 2015); others with hundreds of objects (e.g., 522 objects in Chasin et al., 2018). After identifying objects, these objects are then classified by using the characteristics of the taxonomy to be evaluated. Both participants who were already involved in the taxonomy building (e.g., Oberländer et al., 2018) and participants who were not involved (e.g., Püschel et al., 2016) classify the previously identified real-world objects. If two or more participants perform the classification, often hit ratios and interrater reliabilities are calculated to measure the agreement among participants (e.g., Oberländer et al., 2018; Püschel et al., 2016). Authors of taxonomy articles (e.g., Gimpel et al., 2017), other researchers (e.g., Oberländer et al., 2018) as well as practitioners (e.g., Herterich et al., 2016) not involved in a taxonomy research project are involved in taxonomy evaluation. Applying a taxonomy to real-world objects allows researcher to evaluate their practical applicability and usefulness for classifying, differentiating, and comparing objects (e.g., Gimpel et al., 2017; Püschel et al., 2016) as

well as to evaluate their robustness, utility, efficacy, stability, and completeness (e.g., Chasin et al., 2018; Stoeckli et al., 2017).

Apply a taxonomy to existing research about real-world objects. Researchers also evaluate a taxonomy by applying existing research about real-world objects (eleven articles of the sample fall into this category). This is useful, as not only the real-world objects itself are helpful to evaluate taxonomies, but also research about these objects (i.e., research articles whose goal is to understand a certain type of objects). Again, we found various types of objects, evaluations with a comparatively small (e.g., three articles in Strasser, 2017) or large number of objects (e.g., 50 articles in Gregor, 2006), as well as classifications performed by several independent researchers (e.g., Gregor, 2006; Larsen, 2003). Typical examples of evaluation criteria in this kind of articles are usefulness, effectiveness, unambiguousness, understandability, and completeness of the taxonomies (e.g., Gerber et al., 2017; Gregor, 2006). Applying a taxonomy to existing research about real-world objects also enables to reflect on the current state of research on a certain type of object (e.g., Khalilijafarabad et al., 2016), to uncover commonalities and differences between studies on this type of object (e.g., Agogo et al., 2018) as well as to identify potential research gaps (e.g., Hummel et al., 2016).

4.2 Formulation of a framework for taxonomy evaluation

Our study reveals a diversity of methods employed by researchers in IS to evaluate taxonomies. The evaluation of taxonomies revolves around three questions: The *who* (i.e., subject), *what* (i.e., object), and *how* (i.e., method) of taxonomy evaluation. These three questions were inductively derived from the taxonomy evaluations reported in our sample of articles. For this, we analyzed the descriptions of the evaluation attempts (i.e., we employed reasoning to involve moving from a set of specific instances of taxonomy evaluations to a more general understanding of taxonomy evaluation, Recker 2013). Therefore, two researchers independently coded the sample of articles in order to label and determine key ideas and key activities for evaluating taxonomies. The results were iteratively consolidated and then clustered. Reassuringly we found that the three questions we propose for synthesizing the findings of this study are also anchored and reflected in the general discourse on evaluation in IS (e.g., the *who* in Gemino and Wand, 2004; the *what* in Sonnenberg and vom Brocke, 2012; and the *how* in Peffers et al., 2012 and Venable et al., 2016). Examples of articles include Herterich et al. (2016) and Larsen (2003) (for a detailed description of the *who*), Püschel et al. (2016) and Gimpel et al. (2017) (for a detailed description of the *what*), and Chasin et al. (2018) and Oberländer et al. (2018) (for a detailed description of the *how*). We organize these questions in a preliminary version of a framework for taxonomy evaluation, which we seek to advance in the future (see Figure 2 with orthogonal taxonomy evaluation characteristics).

Who (i.e., subject): Taxonomy evaluation can be carried out by people who were already involved in the taxonomy building or by people who are newly introduced to a taxonomy research project for the particular purpose of evaluation. The people involved may be academics or practitioners. Having people with a proven experience in the taxonomy's domain or with a taxonomy development method is useful. This allows unveiling inconsistencies and/or errors in the use of a method.

What (i.e., object): Taxonomy evaluation can be carried out on different types of objects. This includes the evaluation of real-world objects (e.g., blockchains) as well as research about such real-world objects (e.g., research about blockchains). A combination of both enables aligning the different speeds at which real-world objects and corresponding research evolve. Furthermore, the same objects can be used for the taxonomy evaluation that were already used for building the taxonomy. Alternatively, the same objects can be used at a different (later) point in time, as can entirely new objects.

How (i.e., method): Taxonomy evaluation can be carried out by means of different methods that we have previously identified and described. Different subjects can apply these evaluation methods for different kinds of objects. The combinations of different subjects, objects, and methods result in different taxonomy evaluation strategies.

WHO? Subject of evaluation	Involvement in taxonomy building	Subject has been involved in taxonomy building			Subject has not been involved in taxonomy building		
	Background	Academics			Practitioners		
	Experience	Domain			Method		
WHAT? Object of evaluation	Type	Real-world			About real-world		
	Involvement in taxonomy building	Object has been used in taxonomy building		Object has been used in taxonomy building and evolved since then		Object has not been used in taxonomy building	
	Coverage	Exhaustive		Selective		Representative	
HOW? Method of evaluation	Approach	Quantitative			Qualitative		
	Method	Logical argument	Expert interview	Focus group	Survey	Delphi technique	Sorting
		Log diary	Case study	Illustrative scenario	Action research	Experiment	…

Figure 2. Preliminary version of a framework for taxonomy evaluation.

In summary, the framework indicates the solution space for potential avenues of taxonomy evaluation methods. In this sense, the framework facilitates the selection of a suitable method to pursue taxonomy evaluation, depending on the type and context of a taxonomy. Furthermore, and beyond surveying the solution space for taxonomy evaluation, we hope that such a framework will lead to more taxonomies being evaluated, and that these are evaluated more comprehensively, rigorously, and/or effectively. Taxonomies as classification schemes are also part of calls for papers (e.g. Steininger et al., 2018) and, therefore, the framework can be used for reviewing taxonomy projects in the future.

4.3 Demonstration of the framework for taxonomy evaluation

To demonstrate the framework's practical applicability and capability of characterizing taxonomy evaluation strategies, we selected two illustrative examples from the literature, namely, Püschel et al. (2016) (see left part of Figure 3) and Larsen (2003) (see right part of Figure 3). First, the taxonomy on *smart things in the internet-of-things* (object) that Püschel et al. (2016) has developed. For evaluating their taxonomy, *two co-authors of that study* (subject involvement, background, or experience) have applied the taxonomy by means of *illustrating* (method) *50 real-life objects* (type) that *have not been used during the taxonomy building* (object involvement). The *selection* of these objects is aimed at covering all of domains of smart things (coverage). Second, the taxonomy on *information systems success* (object) that Larsen (2003) has developed and subsequently evaluated by using *sorting* (method). For this, Larsen (2003) *illustratively applied the taxonomy* (method) on a set of *224 research articles* (i.e., articles about IS success) (type), which already *have been used during building* (object involvement). In doing so, *two other researchers*, such as, for example, a doctoral student and a faculty member (subject involvement, background, experience) have sorted a *subset of the sample* (coverage).

WHO?	Involvement in taxonomy building	Subject has been involved in taxonomy building			Subject has not been involved in taxonomy building		
	Background	Academics			Practitioners		
	Experience	Domain			Method		
WHAT?	Type	Real-world			About real-world		
	Involvement in taxonomy building	Object has been used in taxonomy building		Object has been used in taxonomy building and evolved since then		Object has not been used in taxonomy building	
	Coverage	Exhaustive		Selective		Representative	
HOW?	Approach	Quantitative			Qualitative		
	Method	Logical argument	Expert interview	Focus group	Survey	Delphi technique	Sorting
		Log diary	Case study	Illustrative scenario	Action research	Experiment	…

WHO?	Involvement in taxonomy building	Subject has been involved in taxonomy building			Subject has not been involved in taxonomy building		
	Background	Academics			Practitioners		
	Experience	Domain			Method		
WHAT?	Type	Real-world			About real-world		
	Involvement in taxonomy building	Object has been used in taxonomy building		Object has been used in taxonomy building and evolved since then		Object has not been used in taxonomy building	
	Coverage	Exhaustive		Selective		Representative	
HOW?	Approach	Quantitative			Qualitative		
	Method	Logical argument	Expert interview	Focus group	Survey	Delphi technique	Sorting
		Log diary	Case study	Illustrative scenario	Action research	Experiment	…

Figure 3. Demonstration of the framework by applying Püschel et al. (2016) (left side) and Larsen (2003) (right side).

5 Conclusion

In line with DSR, evaluation is required after construction of an artefact has been completed (i.e., ex post evaluation), which applies to taxonomies similarly as to any other type of artefact. DSR provides an equally comprehensive and valuable body of knowledge on evaluation strategies. At the same time, there is a gap between abstract strategies for evaluating artefacts and the concrete implementation of these strategies for evaluating a certain type of artefacts, namely taxonomies. Unlike for the building of taxonomies, there is hardly any methodological guidance on how to evaluate taxonomies. The need for artefact-specific guidelines for evaluation purposes (here for taxonomies) has already been proven helpful for other types of artefacts (e.g., for software artefact instantiations, Lukyanenko et al., 2015). In critically synthesizing and comprehensively reviewing evaluation endeavors on taxonomies, this study is, to the best of our knowledge, the first to place the evaluation of taxonomies at the center of the investigation and thereby makes the knowledge base on evaluation in DSR accessible to researchers who seek to develop taxonomies. Building on this, we propose a framework whose primary objective is to guide DSR researchers who are in need of evaluating taxonomies. Bridging abstract evaluation strategies in design science with concrete taxonomy evaluation methods, the taxonomy evaluation framework intends to serve as a basis for evaluating taxonomies and provides an initial contribution to help researchers searching for answers to the who, what and how of taxonomy evaluation.

A number of limitations apply to our study. First, the keyword search is limited in scope to journals and conferences. Second, we focus on the ex post evaluation after taxonomy building has been completed. Third, the preliminary version of the framework requires extensive evaluation. Beyond the demonstration of the framework with two illustrative examples, we suggest three different ways that future research could take up to evaluate the framework: (1) By compiling a larger sample of examples to demonstrate the framework's capability of classifying and describing taxonomy evaluation endeavors. For a rigorous compilation we suggest the sample to be made up of three equal parts comprising the most-cited articles (to include highly recognized taxonomies), the most-recent articles (to include late-breaking taxonomies), and randomly selected articles (to include further taxonomies). (2) Following Gregor and Hevner (2013), we suggest that the framework be tested with researchers who were not involved in its development. This could be done either in groups or individually. (3) Finally, we suggest that the authors of the articles evaluating taxonomies could be asked whether the framework covers all relevant characteristics needed to decide on how to evaluate a particular taxonomy.

We believe that the further advancement of the preliminary version of a framework for taxonomy evaluation entails several steps that we would like to undertake in future research. This includes the empirical and conceptual anchoring of the framework (e.g., by developing a taxonomy for taxonomy evaluation), the identification of evaluation criteria for taxonomies, and ultimately the development of prescriptive knowledge that explains which evaluation method is best suited in a given context. Such a context includes, for example, the development of a taxonomy for theory building. Taxonomies hold the potential to serve as theories (Gregor, 2006) or for theory building (Doty and Glick, 1994; Bapna et al., 2004). However, "some taxonomies are taxonomic theories and some are not" (Nickerson et al., 2017, p. 1). Consequently, the question arises as to which criteria can be used to determine whether a taxonomy is a taxonomic theory and what this means for the evaluation of taxonomies. Also, whether these criteria should be included in taxonomy evaluation and if so, which evaluation methods are conducive to determining taxonomic theory. As classifications can inform taxonomies (and vice versa), the principles for defining such classifications, for example from the field of conceptual modeling, would form a valuable starting point (Parsons and Wand, 2013). Furthermore, experience from other fields such as biology and social science are conducive to maturing taxonomy evaluation. We therefore see this study as a prelude to further developments of the framework.

With the preliminary version of the framework, we intend to take a very first step towards developing a framework for taxonomy evaluation and to mature taxonomy evaluation. Thereby we seek to methodologically augment DSR in general, and taxonomy research in particular. The purpose is to not only directly improve taxonomies, but also to indirectly improve research that builds on such taxonomies.

Acknowledgements

This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Center "On-The-Fly Computing" (CRC 901, project number 160364472SFB901) and the European Regional Development Fund (ERDF) within the research project "SmartHybrid – Process Engineering" (ZW 6-85003451).

References

Ackermann, T., A. Miede, P. Buxmann and R. Steinmetz (2011). "Taxonomy of technological IT outsourcing risks. Support for risk identification and quantification." In: *Proceedings of the 19th European Conference on Information Systems (ECIS 2011)*. Helsinki, Finland.

Addas, S. and A. Pinsonneault (2015). "The many faces of information technology interruptions. A taxonomy and preliminary investigation of their performance effects." *Information Systems Journal* 25 (3), 231–273.

Agogo, D., T. J. Hess, D. Te'eni and S. McCoy (2018). "'How does tech make you feel?' A review and examination of negative affective responses to technology use." *European Journal of Information Systems* 27 (5), 570–599.

Aksulu, A. and M. Wade (2010). "A comprehensive review and synthesis of open source research." *Journal of the Association for Information Systems* 11 (11), 576–656.

Almufareh, M., D. Abaoud and M. Moniruzzaman (2018). "Taxonomy development for virtual reality technologies in healthcare sector." In: S. Chatterjee, K. Dutta and R. P. Sundarraj (eds.) *Designing for a digital and globalized World*, pp. 146–156: Boston, MA, USA: Springer.

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques.* Thousand Oaks, CA, USA: Sage.

Bapna, Goes, Gupta and Jin (2004). "User heterogeneity and its impact on electronic auction market design. An empirical exploration." *MIS Quarterly* 28 (1), 21–43.

Bärenfänger, R., E. Drayer, D. Daniluk, B. Otto, E. Vanet, R. Caire, T. S. Abbas and B. Lisanti (2016). "Classifying flexibility types in smart electric distribution grids. A taxonomy." In: *Proceedings of the CIRED Workshop (CIRED 2016)*. Helsinki, Finland.

Barn, R. and B. Barn (2016). "An ontological representation of a taxonomy for cybercrime." In: *Proceedings of the 24th European Conference on Information Systems (ECIS 2016)*. Istanbul, Turkey.

Bhattacherjee, A. (2012). *Social science research. Principles, methods, and practices.* Tampa, Florida, USA, Tampa, Minneapolis, USA: Scholar Commons, University of South Florida; Open Textbook Library.

Bock, M. and M. Wiener (2017). "Towards a taxonomy of digital business models. A conceptual dimensions and empirical illustrations." In: *Proceedings of the 38th International Conference on Information Systems (ICIS 2017)*. Seoul, Korea.

Botha, A., M. Weiss and M. Herselman (2018). "Towards a taxonomy of mHealth." In: *Proceedings of the International Conference on Advances in Big Data, Computing and Data Communication Systems (ICABCD 2018)*. Durban, South Africa.

Bui, A., D. Veit and J. Webster (2015). "Gamification. A novel phenomenon or a new wrapping for existing concepts?" In: *Proceedings of the 36th International Conference on Information Systems (ICIS 2015)*. Fort Worth, Texas, USA.

Chasin, F., M. von Hoffen, M. Cramer and M. Matzner (2018). "Peer-to-peer sharing and collaborative consumption platforms. A taxonomy and a reproducible analysis." *Information Systems and e-Business Management* 16 (2), 293–325.

Choudhury, V. (1997). "Strategic choices in the development of interorganizational information systems." *Information Systems Research* 8 (1), 1–24.

Cledou, G., E. Estevez and L. Soares Barbosa (2018). "A taxonomy for planning and designing smart mobility services." *Government Information Quarterly* 35 (1), 61–76.

Cleven, A., P. Gubler and K. M. Hüner (2009). "Design alternatives for the evaluation of design science research artifacts." In: *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2009)*. Philadelphia, PA, USA.

Conboy, K. (2009). "Agility from First Principles. Reconstructing the concept of agility in information systems development." *Information Systems Research* 20 (3), 329–354.

Diniz, E. H., E. S. Siqueira and E. van Heck (2018). "Taxonomy of digital community currency platforms." *Information Technology for Development* 16 (3), 1–23.

Doty, D. H. and W. H. Glick (1994). "Typologies as a unique form of theory building. Toward improved understanding and modeling." *Academy of Management Review* 19 (2), 230–251.

Eldredge, N. and J. Cracraft (1980). *Phylogenetic patterns and the evolutionary process. Method and theory in comparative biology*. New York, NY, USA: Columbia University Press.

Fellmann, M., S. Robert, S. Büttner, H. Mucha and C. Röcker (2017). "Towards a framework for assistance systems to support work processes in smart factories." In: *Proceedings of the IFIP International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE 2017)*. Reggio di Calabria, Italy.

Fteimi, N. and F. Lehner (2018). "Analysing and classifying knowledge management publications. A proposed classification scheme." *Journal of Knowledge Management* 22 (7), 1527–1554.

Galliers, R. D. and F. F. Land (1990). "Choosing appropriate information systems research approaches. A revised taxonomy." In: *Proceedings of the IFIP TC8 WG8.2*, 317–335.

Gao, F., S. Thiebes and A. Sunyaev (2018). "Rethinking the meaning of cloud computing for health care. A taxonomic perspective and future research directions." *Journal of Medical Internet Research* 20 (7), 1-27.

Ge, J. and U. Gretzel (2018). "A taxonomy of value co-creation on Weibo. A communication perspective." *International Journal of Contemporary Hospitality Management* 30 (4), 2075–2092.

Geiger, D., M. Rosemann, E. Fielt and M. Schader (2012). "Crowdsourcing information systems. Definition, typology, and design." In: *Proceedings of the 33rd International Conference on Information Systems (ICIS 2012)*. Orlando, FL, USA.

Gemino, A. and Y. Wand (2004). "A framework for empirical evaluation of conceptual modeling techniques." *Requirements Engineering* 9 (4), 248–260.

Gerber, A., R. Baskerville and A. van der Merwe (2017). "A taxonomy of classification approaches in IS research." In: *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS 2017)*. Boston, MA, USA.

Gibbs, C., U. Gretzel and J. Saltzman (2016). "An experience-based taxonomy of branded hotel mobile application features." *Information Technology & Tourism* 16 (2), 175–199.

Gimpel, H., D. Rau and M. Röglinger (2017). "Understanding FinTech start-ups. A taxonomy of consumer-oriented service offerings." *Electronic Markets* 22 (4), 1–20.

Glass, R. L. and I. Vessey (1995). "Contemporary application-domain taxonomies." *IEEE Software* 12 (4), 63–76.

Gregor, S. (2006). "The nature of theory in information systems." *MIS Quarterly* 30 (3), 611–642.

Gregor, S. and A. R. Hevner (2013). "Positioning and presenting design science research for maximum impact." *MIS Quarterly* 37 (2), 337–356.

Gutierrez, A., J. Orozco and A. Serrano (2008). "Developing a taxonomy for the understanding of business and it alignment paradigms and tools." In: *Proceedings of the 16th European Conference on Information Systems (ECIS 2008)*. Galway, Ireland.

Haas, P., I. Blohm and J. M. Leimeister (2014). "An empirical taxonomy of crowdfunding intermediaries." In: *Proceedings of the 35th International Conference on Information Systems (ICIS 2014)*. Auckland, New Zealand.

Hallerstede, S., F. H. Aysha Beevi, C. F. Pedersen and S. Wagner (2015). "Data quality oriented taxonomy of ambient assisted living systems." In: *Proceedings of the 1st IET International Conference on Technologies for Active and Assisted Living (TechAAL 2015)*. London, United Kingdom.

Herterich, M. M., T. Buehnen, F. Uebernickel and W. Brenner (2016). "A taxonomy of industrial service systems enabled by digital product innovation." In: *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS 2016)*. Koloa, HI, USA.

Herterich, M. M., M. Holler, F. Uebernickel and W. Brenner (2015). "Understanding the business value. Towards a taxonomy of industrial use scenarios enabled by cyber-physical systems in the equipment manufacturing industry." In: *Proceedings of the International Conference on Information Resources Management (Conf-IRM 2015)*. Ottawa, Canada.

Herzfeldt, A., M. Hausen, R. O. Briggs and H. Krcmar (2012). "Developing a Risk Management Process and Risk Taxonomy for Medium-Sized IT Solution Providers." In: *Proceedings of the 20th European Conference on Information Systems (ECIS 2012)*. Barcelona, Spain.

Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design science in information systems research." *MIS Quarterly* 28 (1), 75–105.

Holler, M., F. Uebernickel and W. Brenner (2017). "Defining archetypes of e-collaboration for product development in the automotive industry." In: *Proceedings of the 25th European Conference on Information Systems (ECIS 2017)*. Guimarães, Portugal.

Hummel, D., S. Schacht and A. Maedche (2016). "Determinants of multi-channel behavior. Exploring avenues for future research in the services industry." In: *Proceedings of the 37th International Conference on Information Systems (ICIS 2016)*. Dublin, Ireland.

Jarvinen, P. H. (2000). "Research questions guiding selection of an appropriate research method." In: *Proceedings of the 8th European Conference on Information Systems (ECIS 2000)*. Vienna, Austria.

Jöhnk, J., M. Röglinger, M. Thimmel and N. Urbach (2017). "How to implement agile IT setups. A taxonomy of design options." In: *Proceedings of the 25th European Conference on Information Systems (ECIS 2017)*. Guimarães, Portugal.

Kazan, E., C.-W. Tan, E. T.K. Lim, C. Sørensen and J. Damsgaard (2018). "Disentangling digital platform competition. The case of UK mobile payment platforms." *Journal of Management Information Systems* 35 (1), 180–219.

Keller, R. and C. König (2014). "A reference model to support risk identification in cloud networks." In: *Proceedings of the 35th International Conference on Information Systems (ICIS 2014)*. Auckland, New Zealand.

Khalilijafarabad, A., M. Helfert and M. Ge (2016). "Developing a data quality research taxonomy. An organizational perspective." In: *Proceedings of the 21st International Conference on Information Quality*. Ciudad Real, Spain.

King, W. R. and V. Sethi (1999). "An empirical assessment of the organization of transnational information systems." *Journal of Management Information Systems* 15 (4), 7–28.

Klecun, E. and T. Cornford (2005). "A critical approach to evaluation." *European Journal of Information Systems* 14 (3), 229–243.

Krieger, F. and P. Drews (2018). "Leveraging big data and analytics for auditing. Towards a taxonomy." In: *Proceedings of the 39th International Conference on Information Systems (ICIS 2018)*. San Francisco, CA, USA.

Kuechler, B. and V. Vaishnavi (2008). "On theory development in design science research. Anatomy of a research project." *European Journal of Information Systems* 17 (5), 489–504.

Küpper, T., R. Jung, T. Lehmkuhl and A. Wieneke (2014). "Features for social CRM technology. An organizational perspective." In: *Proceedings of the 20th Americas Conference on Information Systems (AMCIS 2014)*. Savannah, GA, USA.

Labazova, O., T. Dehling and A. Sunyaev (2019). "From hype to reality. A taxonomy of blockchain applications." In: *Proceedings of the 52th Hawaii International Conference on System Sciences (HICCS 2019)*. Wailea, HI, USA.

Labes, S., K. Erek and R. Zarnekow (2013). "Common patterns of cloud business models." In: *Proceedings of the 19th Americas Conference on Information Systems (AMCIS 2013)*. Chicago, IL, USA.

Lakoff, G. (1987). *Women, fire, and dangerous things. What categories reveal about the mind.* Chicago, IL, USA: University of Chicago.

Larsen, K. R.T. (2003). "A taxonomy of antecedents of information systems success. Variable analysis studies." *Journal of Management Information Systems* 20 (2), 169–246.

Liu, X., K. Werder and A. Mädche (2016). "A taxonomy of digital service design techniques." In: *Proceedings of the 37th International Conference on Information Systems (ICIS 2016)*. Dublin, Ireland.

Liu, D., Santhanam, R., & Webster, J. (2017). "Toward meaningful engagement. A framework for design and research of gamified information systems." *MIS Quarterly* 41 (4), 1011–1034.

Lombard, M., J. Snyder-Duch and C. C. Bracken (2002). "Content analysis in mass communication. Assessment and reporting of intercoder reliability." *Human Communication Research* 28 (4), 587–604.

Lukyanenko, R., J. Evermann and J. Parsons (2015). "Guidelines for establishing instantiation validity in IT artifacts. A survey of IS research." In: *Proceedings of the 10th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2015)*. Dublin, Ireland.

March, S. T. and G. F. Smith (1995). "Design and natural science research on information technology." *Decision Support Systems* 15 (4), 251–266.

Melas, C. D., L. A. Zampetakis, A. Dimopoulou and V. S. Moustakis (2017). "An empirical investigation of technology readiness among medical staff based in greek hospitals." *European Journal of Information Systems* 23 (6), 672–690.

Miller, J. G. and A. V. Roth (1994). "A taxonomy of manufacturing strategies." *Management Science* 40 (3), 285–304.

Moore, G. C. and I. Benbasat (1991). "Development of an Instrument to measure the perceptions of adopting an information technology innovation." *Information Systems Research* 2 (3), 192–222.

Morana, S., S. Schacht, A. Scherp and A. Maedche (2017). "A review of the nature and effects of guidance design features." *Decision Support Systems* 97, 31–42.

Nambisan, S., R. Agarwal and M. Tanniru (1999). "Organizational mechanisms for enhancing user innovation in information technology." *MIS Quarterly* 23 (3), 365.

Nickerson, R., J. Muntermann, U. Varshney and H. Isaac (2009). "Taxonomy development in information systems. Developing a taxonomy of mobile applications." In: *Proceedings of the 17th European Conference on Information Systems (ECIS 2009)*. Verona, Italy.

Nickerson, R. C., U. Varshney and J. Muntermann (2013). "A method for taxonomy development and its application in information systems." *European Journal of Information Systems* 22 (3), 336–359.

Nickerson, R. C., U. Varshney and J. Muntermann (2017). "Of taxonomies and taxonomic theories." In: *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS 2017)*. Boston, MA, USA.

Oberländer, A. M., M. Röglinger, M. Rosemann and A. Kees (2018). "Conceptualizing business-to-thing interactions. A sociomaterial perspective on the internet of things." *European Journal of Information Systems* 27 (4), 486–502.

Okoli, C. and S. D. Pawlowski (2004). "The Delphi method as a research tool. An example, design considerations and applications." *Information & Management* 42 (1), 15–29.

Paré, G., M.-C. Trudel, M. Jaana and S. Kitsiou (2015). "Synthesizing information systems knowledge. A typology of literature reviews." *Information & Management* 52 (2), 183–199.

Parsons, J. and Y. Wand (2008). "Using cognitive principles to guide classification in information systems modeling." *MIS Quarterly* 32 (4), 839–868.

Parsons, J. and Y. Wand (2013). "Extending classification principles from information modeling to other disciplines." *Journal of the Association for Information Systems* 14 (5), 245–273.

Peffers, K., M. Rothenberger, T. Tuunanen and R. Vaezi (2012). "Design science research evaluation." In: *Proceedings of the 7th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2012)*. Las Vegas, NV, USA.

Peters, C., I. Blohm and J. M. Leimeister (2015). "Anatomy of successful business models for complex services. Insights from the telemedicine field." *Journal of Management Information Systems* 32 (3), 75–104.

Petter, Straub and Rai (2007). "Specifying formative constructs in information systems research." *MIS Quarterly* 31 (4), 623.

Prat, N., I. Comyn-Wattiau and J. Akoka (2015). "A taxonomy of evaluation methods for information systems artifacts." *Journal of Management Information Systems* 32 (3), 229–267.

Premkumar, G., K. Ramamurthy and C. S. Saunders (2005). "Information processing view of organizations. An exploratory examination of fit in the context of interorganizational relationships." *Journal of Management Information Systems* 22 (1), 257–294.

Püschel, L., M. Röglinger and H. Schlott (2016). "What's in a smart thing? Development of a multilayer taxonomy." In: *Proceedings of the 37th International Conference on Information Systems (ICIS 2016)*. Dublin, Ireland.

Rai, A. (2017). "Editor's comments. Diversity of design science research." *MIS Quarterly* 41 (1), iii–xviii.

Raza, U., W. Ahmad and A. Khan (2018). "Transformation from manufacturing process taxonomy to repair process taxonomy. A phenetic approach." *Journal of Industrial Engineering International* 14 (2), 415–428.

Recker, J. (2013). *Scientific research in information systems.* Berlin, Heidelberg, Germany: Springer.

Remane, G., A. Hanelt, J. F. Tesch, R. C. Nickerson and L. M. Kolbe (2016). "A taxonomy of carsharing business models." In: *Proceedings of the 37th International Conference on Information Systems (ICIS 2016)*. Dublin, Ireland.

Roeder, J., D. R. Cardona, M. Palmer, O. Werth, J. Muntermann and M. H. Breitner (2018). "Make or Break. Business model determinants of FinTech venture success." In: *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI 2018)*. Lüneburg, Germany.

Schäffer, T. and D. Stelzer (2017). "Towards a taxonomy for coordinating quality of master data in product information sharing." In: *Proceedings of the 22nd MIT International Conference on Information Quality*. Little Rock, AR, USA.

Schmidt-Kraepelin, M., S. Thiebes, M. C. Tran and A. Sunyaev (2018). "What's in the game? Developing a taxonomy of gamification concepts for health apps." In: *Proceedings of the 51st Hawai International Conference on System Sciences (HICSS 2018)*. Waikoloa, HI, USA.

Schneider, S., J. Lansing, F. Gao and A. Sunyaev (2014). "A taxonomic perspective on certification Schemes. Development of a taxonomy for cloud service certification criteria." In: *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS 2014)*. Waikoloa, HI, USA.

Sein, M. K., O. Henfridsson, S. Purao, M. Rossi and R. Lindgren (2011). "Action design research." *MIS Quarterly* 35 (1), 37–56.

Seyffarth, T., S. Kühnel and S. Sackmann (2017). "A taxonomy of compliance processes for business process compliance." In: J. Carmona, G. Engels and A. Kumar (eds.) *Business process management forum*, pp. 71–87, Berlin, Heidelberg, Germany: Springer.

Siering, M., B. Clapham, O. Engel and P. Gomber (2017). "A taxonomy of financial market manipulations. Establishing trust and market integrity in the financialized economy through automated fraud detection." *Journal of Information Technology* 32 (3), 251–269.

Smith, S., D. Winchester and V. Pang (2014). "The construction of identity offences taxonomy. An australian context." In: *Proceedings of the 25th Australasian Conference on Information Systems (ACIS 2014)*. Auckland, New Zealand.

Snow, N. M. and J. L. Reck (2016). "Developing a government reporting taxonomy." *Journal of Information Systems* 30 (2), 49–81.

Sokal, R. R. and P.H.A. Sneath (1963). *Principles of numerical taxonomy.* San Francisco, CA, USA: W.H. Freeman and Company.

Sonnenberg, C. and J. vom Brocke (2012). "Evaluations in the science of the artificial. Reconsidering the build-evaluate pattern in design science research." In: *Proceedings of the 7th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2012)*. Las Vegas, NV, USA.

Steininger, D., T. Hess, Ortiz de Guinea Lopez de Arana, Ana and S. Sarker (2018). *Call for papers. Track on business models and the value of IS in a digital world*. URL: http://ecis2019.eu/programme/research-tracks/business-models-and-the-value-of-is-in-a-digital-world (visited on 11/27/2018).

Stoeckli, E., F. Uebernickel and W. Brenner (2017). "Capturing functional affordances of enterprise social software." In: *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS 2017)*. Boston, MA, USA.

Strasser, A. (2017). "Delphi method variants in information systems research. Taxonomy development and application." *Electronic Journal of Business Research Methods* 15 (2), 120–133.

Strode, D. E. (2016). "A dependency taxonomy for agile software development projects." *Information Systems Frontiers* 18 (1), 23–46.

Szopinski, D., T. Schoormann, T. John, R. Knackstedt and D. Kundisch (2017). "How software can support innovating business models. A taxonomy of functions of business model development tools." In: *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS 2017)*. Boston, MA, USA.

Szopinski, D., T. Schoormann, T. John, R. Knackstedt and D. Kundisch (2019). "Software tools for business model innovation. Current state and future challenges." *Electronic Markets*, forthcoming.

Thiebes, S., G. Kleiber and A. Sunyaev (2017). "Cancer genomics research in the cloud. A taxonomy of genome data sets." In: *Proceedings of the 4th International Workshop on Genome Privacy and Security (GenoPri 2017)*. Orlando, USA.

Tilly, R., O. Posegga, K. Fischbach and D. Schoder (2017). "Towards a conceptualization of data and information quality in social information systems." *Business & Information Systems Engineering* 59 (1), 3–21.

Tönnissen, S. and F. Teuteberg (2018). "Towards a taxonomy for smart contracts." In: *Proceedings of the 26th European Conference on Information Systems (ECIS 2018)*. Portsmouth, United Kingdom.

Tremblay, M. C., A. R. Hevner and D. J. Berndt (2010). "The use of focus groups in design science research." In: A. Hevner and S. Chatterjee (eds.) *Design research in information systems*, pp. 121–143. Boston, MA, USA: Springer.

Tsatsou, P., S. Elaluf-Calderwood and J. Liebenau (2010). "Towards a taxonomy for regulatory issues in a digital business ecosystem in the EU." *Journal of Information Technology* 25 (3), 288–307.

Vaishnavi, V. and B. Kuechler (2004). *Design Research in information systems*.

van de Ven, A. H. and A. L. Delbecq (1974). "The effectiveness of nominal, delphi, and interacting group decision making processes." *Academy of Management Journal* 17 (4), 605–621.

Venable, J., J. Pries-Heje and R. Baskerville (2016). "FEDS. A Framework for evaluation in design science research." *European Journal of Information Systems* 25 (1), 77–89.

Wand, Y., D. E. Monarchi, J. Parsons and C. C. Woo (1995). "Theoretical foundations for conceptual modelling in information systems development." *Decision Support Systems* 15 (4), 285–304.

Webster, J. and R. T. Watson (2002). "Analyzing the past to prepare for the future. Writing a literature review." *MIS Quarterly* 26 (2), xiii–xxiii.

Werder, K. and H.-Y. Wang (2016). "Towards a software product industry classification." In: *Proceedings of the 15th International Conference on Intelligent Software Methodologies, Tools and Techniques (SoMeT 2016)*. Larnaca, Cyprus.

Wilde, T. and T. Hess (2007). "Forschungsmethoden der Wirtschaftsinformatik." *Wirtschaftsinformatik* 49 (4), 280–287.

Wiliam, D. and P. Black (1996). "Meanings and consequences. A basis for distinguishing formative and summative functions of assessment?" *British Educational Research Journal* 22 (5), 537–548.

Williams, K., S. Chatterjee and M. Rossi (2008). "Design of emerging digital services. A taxonomy." *European Journal of Information Systems* 17 (5), 505–517.

Yin, R. K. (2017). *Case study research and applications*. *Design and methods*. Thousand Oaks, CA, USA: Sage.

Zrenner, J., A. P. Hassan, B. Otto and J. C. Marx Gómez (2017). "Data source taxonomy for supply network structure visibility." In: *Proceedings of the Hamburg International Conference of Logistics (HICL 2017)*. Hamburg, Germany.

Zschech, P. (2018). "A taxonomy of recurring data analysis problems in maintenance analytics." In: *Proceedings of the 26th European Conference on Information Systems (ECIS 2018)*. Portsmouth, United Kingdom.