Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 231–236 DOI: 10.15439/2024F7208 ISSN 2300-5963 ACSIS, Vol. 39

Toward a Framework for Determining Methods of Evaluation in Design Science Research

Julia Müller
University of Applied Sciences Heilbronn, Heilbronn, Germany
Email: jmueller8@stud.hs-heilbronn.de
ORCID: 0009-0002-2184-2357

Stefanie Würth
University of Applied Sciences Heilbronn, Heilbronn, Germany
Email: swuerth@stud.hs-heilbronn.de
ORCID: 0009-0009-9469-5213

Thomas Schäffer
University of Applied Sciences Heilbronn, Heilbronn, Germany
Email: thomas.schaeffer@hs-heilbronn.de
ORCID: 0000-0001-8097-286X

Prof. Dr. Christian Leyh
Technical University of Central Hesse (THM) – University of Applied Sciences, THM Business School, Gießen, Germany
Email: christian.leyh@w.thm.de
ORCID: 0000-0003-0535-0336

Abstract—**Evaluation is a key phase of design science research, particularly in design-oriented information systems, one that involves analyzing and solving problems to create artifacts. Because the nature of those artifacts varies based on the problem, they necessitate different methods of evaluation, and selecting an appropriate one first requires identifying appropriate criteria for evaluation. This paper aims to pinpoint those criteria by systematically reviewing the literature, with particular focus on identifying various criteria for evaluation, their frequency, their significance in evaluating artifacts, and their connection to specific methods of evaluation. The findings suggest a framework for choosing the most suitable methods of evaluation based on the defined criteria that can enhance the rigor and relevance of the evaluation phase in design science research.**

Index Terms—**Design Science Research, DSR, Methods of Evaluation, Criteria for Evaluation**

I. INTRODUCTION

IN design-oriented information systems, design science research (DSR) is becoming increasingly important for solving specific problems and developing problem-solving artifacts [1], [2]. DSR is an especially suitable approach to creating artifacts that address a previously unsolved and important business problem and is therefore ideal for developing innovative solutions and prototypes in unexplored areas.

A central phase of DSR is evaluating developed artifacts to prove their usefulness, quality, and effectiveness [3]. Researchers using DSR face the challenge of selecting appropriate methods of evaluation [4]. For example, Ihlström Erikson et al. [5] have emphasized the need for detailed strategies in th e process of evaluation that meet the specific needs of the project at hand. However, despite general agreement on the importance of evaluation in DSR, the literature lacks clear guidance on selecting evaluation-focused strategies and methods [4], [6]. The literature also shows inconsistencies within guidelines for selecting criteria and standards for evaluation [7]. Straßburg et al. [8] have additionally underscored the lack of guidance and the challenges of selecting an appropriate method for evaluating artifacts.

To provide a possible guideline for selecting appropriate methods of evaluation, the relationships between criteria and methods for evaluation have to be determined first. Therefore, this paper examines the following research question:

Which evaluation criteria influence the choice of methods of evaluation in the context of design science research (DSR)?

The chief objective of our research project is to identify the relationships between artifacts, methods of evaluation, and criteria for evaluation based on a systematic literature review. By identifying those relationships, we aim to support decision-making about the selection and application of methods of evaluation in the context of DSR. To that end, criteria have been extracted from the literature, thoroughly analyzed, and summarized. With that summary, the paper provides a framework for selecting appropriate methods of evaluation that can support such decision-making in the selection of evaluation methods based on previously selected criteria in the context of DSR, all to improve the quality and effectiveness of evaluations.

In what follows, the paper first introduces relevant terms to establish a common understanding. Next, it presents the methodological approach used in the systematic literature review, after which it presents the results using the framework. Last, the results are summarized, and implications for future research are derived.

II. THEORETICAL BACKGROUND

This section presents the theoretical foundations and key concepts necessary for understanding our research question and the study guided by it. A firm understanding of those concepts provides a foundation for the analysis and discussion that follow. First, the approach of DSR and its relevance in information systems is explained. Second, the different methods of evaluation in the context of DSR are described in detail, along with their application and significance. Last, the specific criteria for evaluation that are important for evaluating artifacts developed in the context of DSR are discussed. Those principles make it possible to precisely address the research question and systematically justify the selection of appropriate methods of evaluation.

A. Evaluation in Design Science Research

Österle et al. [9] have classified research methods used in the context of information systems into *behavioral* and *design-oriented* approaches. According to both Hevner et al. [1] and Larsen et al. [2], DSR ranks among the design-oriented approaches. In information systems, DSR is understood as a problem-oriented paradigm aimed at creating innovative artifacts. Those artifacts, including models and methods, should enable the effective and efficient analysis, design, implementation, management, and use of information systems [1].

With roots in engineering, DSR typically entails phases such as *problem identification*, *requirements specification*, *design*, *evaluation*, and *communication* [3], [10]. Various approaches to evaluation differ in their applications depending on the type of artifact. Despite such diversity, the overarching goal of evaluation is to assess artifacts for their usefulness and effectiveness [11]. Evaluation is both a specific phase in the design–research cycle [1], [12] and an accompanying activity throughout the process [13]. The targeted assessment of intangible or tangible objects to justify their suitability is central to evaluation in DSR. According to Hevner et al. [1], evaluation is critical to the success of DSR.

B. Methods of Evaluation in DSR

Methods of evaluation are used to structure the evaluation and to transform the developed artifact into a research result [3]. Depending on the type of artifact, some methods of evaluation are more appropriate than others [3]. Different authors have developed different types of methods of evaluation and their application [7]. For example, Hevner et al. [1] divide the methods into five classes: *observational methods*, *analytical methods*, *experimental methods*, *test methods*, and *descriptive methods*. Peffers et al. [3], by contrast, distinguish only between *demonstration* and *evaluation*.

Peffers et al. [3], in discussing the variety of artifacts and methods of evaluation, have identified, among others, *logical reasoning*, *expert evaluations*, *technical experiments*, *subject-based experiments*, *action research*, *prototypes*, *case studies*, and *illustrative scenarios*. Some methods, including *demonstrations*, *literature reviews*, and *expert interviews*, are also commonly used but not explicitly defined in Peffers et al.'s [3], [11], [14] work.

C. Criteria for Evaluating Artifacts

March and Smith [10] have emphasized certain criteria when evaluating artifacts, including *construct*, *model*, *method*, and *instantiation*. *Constructs* are evaluated in terms of completeness, simplicity, elegance, comprehensibility, and ease of use [10], while *models* are evaluated according to their correspondence to real-world phenomena, completeness, level of detail, robustness, and internal consistency [10]. On the whole, operationality, efficiency, generality, and usability are important in evaluating *methods* [10], while *instantiations* are evaluated in terms of efficiency, effectiveness, and impact on the environment and users [10]. Although March and Smith [10] have answered the question of what to evaluate with different criteria, a specific method of evaluation has yet to be assigned to the criteria, and an explicit assignment of criteria to methods of evaluation has not been investigated, either.

III. RESEARCH METHODOLOGY

We conducted a systematic literature review to identify published research addressing the selection of methods of evaluation and criteria in DSR. To ensure that our search for literature was comprehensive and structured, we followed vom Brocke et al.'s [15] approach, which we also supplemented with a forward and backward search as well as a visualization using a concept matrix according to Webster and Watson [16]. Webster and Watson [16] recommend using a concept-centered organizational framework in analyzing literature in order to enable the comparability of different contents within the literature. Vom Brocke et al. [15] have especially emphasized the strict documentation and presentation of the entire search process in scientific papers, for doing so ensures traceability for other researchers and optimal reusability.

Fig. 1 provides an overview of our methodology. The literature search was performed in the databases *AIS Electronic Library*, *IEEE Xplore Digital Library*, *Science Direct*, and *SpringerLink*. We conducted an electronic search of titles, abstracts, and keywords using the search terms [("design science research" AND "DSR" AND "evaluation" AND "criteria")] and [("evaluation method" AND "criteria" AND "DSR")]. The period from 2004 to 2023 was chosen in order to ensure that the results related to the DSR approach, with Hevner et al.'s [1] article, published in 2004, being the oldest included in our review.

The search yielded a total of 903 publications. After we excluded duplicates, the abstracts of 260 publications were reviewed for thematic relevance. After the abstracts, keywords and, if applicable, the full texts were analyzed, 24 publications relevant to our research's focus were identified. In addition to the database search, we performed a forward and backward search following the recommendations of Webster and Watson [16]. That final step raised the final number of publications to 30, all of which were read in detail and subsequently subjected to qualitative content analysis [17].

All publications were analyzed to identify methods of evaluation in relation to the evaluation criteria of the artifacts studied in the context of DSR. Six artifacts, eight methods of evaluation, and ten criteria for evaluation were identified in the 30 publications, all of which have already been analyzed in three works by March and Smith [10], Hevner et al. [1], and Peffers et al. [18]. The findings from the publications allowed extending the characteristics of evaluation, which were listed in a concept matrix next to the respective publications in order to derive correlations. In this paper, they are presented in a framework that serves as a guide for selecting methods of evaluation.

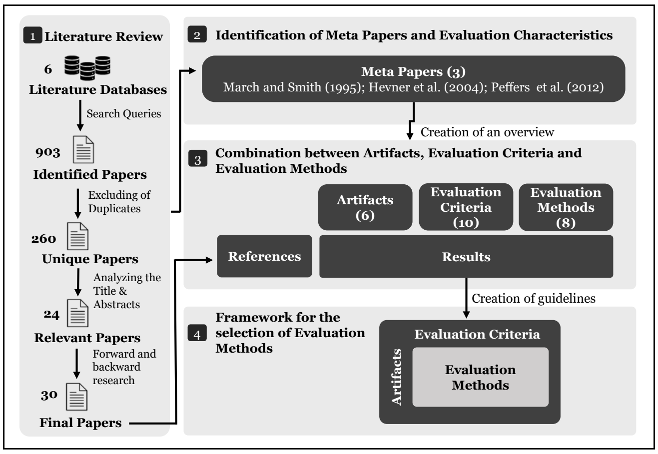

Fig. 1. Overview of the Research Methodology

IV. FINDINGS

In DSR, artifacts play a central role in developing and evaluating innovative solutions in information systems. Those artifacts, ranging from algorithms to frameworks to models, serve different purposes and require specific methods of evaluation in order to assess their effectiveness and applicability. Our literature review revealed six artifacts, ten criteria for evaluation, and eight methods of evaluation. Table I shows the results of the literature review in the form of a shortened concept matrix, as recommended by Webster and Watson [16]. The "SUM" column in Table I indicates the number of publications reviewed that refer to each specific concept.

In this section, the identified artifacts, criteria, and methods of evaluation are presented and explained in a structured manner, using the concept matrix as a visual aid. The developed framework is also presented, which serves as a decision-making aid for selecting methods of evaluation based on the previously selected criteria in the context of DSR.

The full concept matrix is available upon request from the authors.

A. Artifacts, Criteria, Methods of Evaluation, and their Relationships

Evaluation Criteria and their Application to Different Artifacts

Applicability: According to Baskerville et al. [12], *applicability* describes the ability of an artifact to achieve goals beyond its original purpose or to adapt to changing goals. The criterion of applicability should be used for evaluating frameworks and methods, for it allows a variety of methods of evaluation that in turn allow versatile application. The prototype was used most often for that purpose. For other artifacts, the criterion of applicability also allows evaluation with different methods, with the exception of the algorithm artifact.

Appropriateness: According to Akoka et al. [19], *appropriateness* is an criterion of evaluation used to ensure that the developed artifact is appropriate with respect to specific requirements and contexts. According to the definition of *appropriateness*, the methods need to be aligned with the intended outcomes. The criterion of appropriateness provides the broadest range of applications for the framework, with nearly all methods being appropriate in the overall picture of the artifacts, with the exception of the literature review. However, for the artifact of the construct, no method is appropriate for the criterion.

TABLE I.
CONCEPT MATRIX OF THE IDENTIFIED ARTIFACTS, CRITERIA, AND METHODS

		SUM
Artifact	Algorithm	3
	Construct	8
	Framework	9
	Instantiation	3
	Method	7
	Model	8
Criterion	Applicability	8
	Appropriateness	5
	Correctness	6
	Ease of Use	16
	Effectiveness	10
	Feasibility	3
	Functionality	4
	Understandability	4
	Usefulness	17
	Validity	5
Method	Case Study	10
	Demonstration Method	4
	Expert Evaluation	6
	Expert Interview	7
	Literature Review	3
	Prototype	7
	Subject-Based Experiment	3
	Technical Experiment	5
see references [19] – [48]		

Correctness: *Correctness* is defined as the degree of agreement between the results of the artifact and the expected results [9]. The criterion of correctness is covered by most methods of evaluation, with the exception of the expert interview and the literature review. There is no appropriate method of evaluation for the criterion of correctness in an instantiation.

Ease of Use: In the scientific literature, *ease of use* is described as the extent to which the artifact can be used effortlessly [49]. The expert interview is a common method of evaluating the usability of different artifacts. For algorithms, however, the search for suitable methods of evaluation to assess usability remains a challenge.

Effectiveness: In the context of methods of evaluation, *effectiveness* refers to the extent to which an artifact achieves its goal in a narrow sense, without considering situational aspects [50]. Although several methods of evaluation are suitable for most artifacts per the criterion of effectiveness, algorithms present unique challenges that limit the applicability of certain methods, including the case study.

Feasibility: According to Prat et al. [43], *feasibility* is a criterion of evaluation used to ensure that the developed artifact is viable in terms of technical, operational, and economic aspects. Per the definition of *feasibility*, the methods have to correspond with the outcomes of practical implementation and integration. The criterion offers a broad range of applications within the framework, with a variety of methods being suitable for evaluating feasibility, especially case studies, demonstrations, expert evaluations, and expert interviews.

Functionality: The *functionality* of an artifact refers to its ability to provide functions that meet predetermined and expected requirements when the artifact is used under specified conditions [51]. Functionality is often assessed in constructs, frameworks, and models, with expert interviews being a commonly used method. However, finding appropriate methods evaluating functionality in algorithms, instantiations, and methods can be more difficult.

Understandability: *Understandability* refers to the extent to which the artifact can be understood both at the global level and at the detailed level of elements and relationships within the artifact [9]. Understandability is typically evaluated in constructs, methods, and models. By contrast, evaluating understandability in algorithms, frameworks, and instantiations poses significant challenges, for appropriate methods are less likely to be identified.

Usefulness: According to Davis [49], *usefulness* is the extent to which a person believes that using a particular system would improve their work performance and is influenced by perception. Davis [49] has also defined *usefulness* as the extent to which an artifact positively influences an individual's task performance. Usefulness can be assessed for different artifacts, with frameworks and methods being particularly versatile. Case studies are commonly used, particularly for constructs, although identifying alternative methods can improve the assessment process.

Validity: According to Gregor and Hevner [52], *validity* means that the artifact works correctly—that is, achieves its goal in the right way. Validity is not appropriate for the artifacts of algorithms and instantiations because no appropriate methods have been identified. Methods could be analyzed for the remaining artifacts, with prototypes proving to be suitable for all four artifacts. The method and model artifacts, meanwhile, are particularly versatile because they use most of the methods.

Methods of Evaluation for Artifacts in DSR

Case Study: In a case study, the selected artifact is applied to a real situation. For that reason, the effects on the real situation can be evaluated [3].

Demonstration Method: By executing an artifact, the demonstration serves as a preliminary evaluation of the artifact's practical functionality [3].

Expert Evaluation: In an expert evaluation, an artifact is evaluated by one or more experts. Peffers et al. [3] give the example of Delphi studies.

Expert Interview: According to Sonnenberg and vom Brocke [14], expert interviews can be used repeatedly in all iterations of an evaluation. Although the questions within such interviews can be both objective and subjective, the answers are always subjective [11].

Literature Review: A literature review, primarily conducted in the early stages of an evaluation, serves to justify the problem and design objectives [13].

Prototype: A prototype is the implementation of the developed artifact and should serve to demonstrate its suitability and/or usefulness [3].

Subject-Based Experiment: The subject-based experiment method of evaluation is used to test the truth of a claim or hypothesis by using one or more test subjects [3].

Technical Experiment: A technical experiment is used to evaluate the technical performance of an implementation using real data, synthetic data, or no data, and the evaluation does not relate to performance in the real world [3].

B. Framework for Selecting Appropriate Methods of Evaluation

Based on the literature review and the concept matrix, we were able to gain important insights into the selection of methods of evaluation in the context of DSR. Through the structured presentation in the concept matrix, we were also able to determine which methods of evaluation are best suited for evaluating certain artifacts and criteria for evaluation.

To develop those mappings, we analyzed and evaluated the different combinations of artifacts, criteria, and methods of evaluation in an iterative process.

The resulting framework, shown in Table II, provides a systematic basis for selecting appropriate methods of evaluation, represented by numbers 1–8. Special attention was paid to the question of which methods are best suited to evaluate specific criteria. The specific scoring scheme is available upon request from the authors.

Depending on the artifact and the criterion, researchers can choose the appropriate methods to increase the quality and effectiveness of their evaluations. Shaded areas in the table indicate combinations for which no appropriate method of evaluation could be identified.

V. DISCUSSION

The results presented in this paper illustrate the complexity and multifaceted nature of selecting appropriate methods of evaluation in the context of DSR. In many of the reviewed publications, several methods are combined to evaluate an artifact, which contributes to the validation and increased significance of the results. For example, qualitative methods such as expert interviews can be complemented by technical experiments. Such multimethod approaches provide a more comprehensive view of the artifacts and strengthen the validity of the results of evaluation. The developed framework allows identifying versatile criteria for evaluation, including *effectiveness* and *usefulness*, that provide appropriate methods for each artifact.

TABLE II.
FRAMEWORK FOR SELECTING METHODS OF EVALUATION IN DSR

Artifact \ Evaluation Criteria	Applicability	Appropriateness	Correctness	Ease of Use	Effectiveness	Feasibility	Functionality	Understandability	Usefulness	Validity
Algorithm		2, 3	2, 3		1				1, 2, 3	
Construct	2, 3		1, 6, 7, 8	4, 6, 7, 8	3, 4, 6, 7, 8	2	2, 6	6, 7, 8	1	6, 8
Framework	1, 3, 4, 6	1, 4, 6	7	1, 3, 4, 6, 7	1, 4, 6		1, 4		1, 3, 4, 5, 6	1, 3, 6
Instantiation	7, 8	7, 8		4, 8	4, 6, 7	4, 5, 8			4, 5, 8	
Method	1, 2, 6, 7, 8	7, 8	3, 7, 8	1, 2, 3, 4, 6, 8	3, 4, 7, 8	3		1, 3, 4	1, 3, 4, 6	1, 3, 4, 6, 8
Model	4, 5, 6	1, 4	6, 8	1, 4, 8	2, 4, 5, 8		1, 4	1, 3, 4	1, 3, 4	1, 3, 4, 6, 8

Numbers are Methods of Evaluation: 1 - Case Study; 2 - Demontration Method; 3 - Expert Evaluation; 4 - Expert Interview; 5 - Literature Review; 6 - Prototype; 7 - Subject-Based Experiment; 8 - Technical Experiment

The criteria of *applicability*, *appropriateness, correctness*, and *ease of use* proved to be particularly suitable. By contrast, *validity* with two and *functionality* as well as *understandability* with three unassigned artifacts were found to be less versatile.

A major limitation of our work is the assumption that all relevant criteria and methods were explicitly mentioned by the authors. That assumption could lead to bias, for established methods are cited more frequently, whereas newer approaches may be overlooked. The limited number of papers reviewed is another limitation. An extended search including additional databases and conference proceedings could improve the robustness of the results.

This paper does not provide a complete guide to evaluation but serves as a first step in selecting appropriate methods of evaluation based on the identified criteria. A detailed literature review in future studies could uncover more relevant methods and narrow gaps that currently prevent the development of a more comprehensive guide.

Future studies should additionally investigate the practical implications and challenges of the proposed methods through case studies. Doing so would validate the applicability and effectiveness of the methods in real-life scenarios and contribute to improving processes of evaluation.

VI. CONCLUSION

This paper provides important insights for researchers and their future research. The characteristics of evaluation identified in the concept matrix were mentioned several times in the different publications reviewed, which has allowed the creation of an initial guideline for selecting methods of evaluation. That guideline, based on the work of March and Smith [10], is intended to guide the selection of methods of evaluation based on the artifacts and criteria for evaluation.

We have answered our research question by identifying several correlations between methods of evaluation, artifacts, and criteria and by establishing a corresponding framework. To date, no comparable study has considered all three dimensions—the artifact, the evaluation criterion, and the method of evaluation—in DSR. Our paper fills that gap and thereby provides a valuable foundation for future studies.

Overall, our paper demonstrates the need for a systematic, methodical approach to selecting methods of evaluation in DSR. Our paper additionally provides initial support for the selection of appropriate methods of evaluation and provides a basis for further research in the context of DSR. Although the presented framework is an important step, further research is needed. We recommend an extended literature review to identify additional evaluation criteria and their relationship to methods and artifacts. A more comprehensive review would narrow existing gaps in knowledge on the topic and develop a practical basis for a methodology for evaluation.

REFERENCES

[1] A. Hevner, S. March, J. Park, and S. Ram, "Design Science in Information Systems," *MIS Quarterly*, vol. 28, pp. 75-105, 2004.

[2] Larsen, K. R., Lukyanenko, R., Mueller, R. M., Storey, V. C., VanderMeer, D., Parsons, J., and Hovorka, D. S., "Validity in Design Science Research," in *Designing for Digital Transformation. Co-Creating Services with Citizens and Industry. DESRIST 2020*, Lecture Notes in Computer Science (LNCS), vol. 12388, S. Hofmann, O. Müller, and M. Rossi, Eds., Cham: Springer, 2020, pp 272–282.

[3] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *JMIS*, vol. 24, pp. 45-77, 2007.

[4] A. Hevner, N. Prat, I. Comyn-Wattiau, and J. Akoka, "A pragmatic approach for identifying and managing design science research goals and evaluation criteria," in *AIS SIGPRAG Pre-ICIS Workshop 2018*, San Francisco, USA, 2018.

[5] C. Ihlstrom Eriksson, M. Åkesson, and K. Kautz, "Authentic and Concurrent Evaluation – Refining an Evaluation Approach in Design Science Research," in *PACIS 2011 Proc.*, Brisbane, Australia, 2011.

[6] E. Stoeckli, G. Neiditsch, F. Uebernickel, and W. Brenner, "Towards an Understanding of How and Why Design Science Research Scholars Evaluate," in *Proceedings of ACIS 2017*, Hobart, Australia, 2017.

[7] J. Venable, J. Pries-Heje, and Baskerville, R. "FEDS: A Framework for Evaluation in Design Science Research," *Eur J Inf Syst*, vol. 25, pp. 77-89, 2016.

[8] S. Straßburg, S. Kahlert, D. Stöffler, and T. Schäffer, "Identification of Issues in Design Science Research Evaluation – A Literature Review," in *AMCIS 2021 Proceedings*, virtual conference, 2021.

[9] H. Österle et al., "Memorandum on design-oriented information systems research," *Eur J Inf Syst*, vol. 20, pp. 7-10, 2011.

[10] S. March, and G. Smith, "Design and natural science research on information technology," *Dec. Supp. Syst.*, vol. 15, pp. 251-266, 1995.

[11] A. Cleven, P. Gubler, and K. M. Hüner, "Design alternatives for the evaluation of design science research artifacts," in *DESRIST 2009 Proceedings*, Philadelphia, USA, 2009.

[12] R. Baskerville, M. Kaul, and V. Storey, "Genres of Inquiry in Design-Science Research: Justification and Evaluation of Knowledge Production," *MIS Quarterly*, vol. 39, pp. 541-564, 2015.

[13] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, "Action Design Research". *MIS Quarterly*, vol. 35, pp. 37-56, 2011.

[14] C. Sonnenberg, and J. vom Brocke, "Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research," in: *Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012*, Lecture Notes in Computer Science (LNCS), vol. 7286, K. Peffers, M. Rothenberger, and B. Kuechler, Eds., Berlin, Heidelberg: Springer, 2012, pp. 381-397.

[15] J. vom Brocke et al., "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," Comm. of the AIS, vol. 37, pp. 205-224, 2015.

[16] J. Webster, and R. Watson, "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly*, vol. 26, pp. xiii-xxiii, 2002.

[17] P. Mayring, "Qualitative Forschungsdesigns," in *Handbuch Qualitative Forschung in der Psychologie*, G. Mey, and K. Mruck, Eds., Wiesbaden: Springer, 2020, pp. 3-17.

[18] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi, "Design Science Research Evaluation," in *Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012*, Lecture Notes in Computer Science (LNCS), vol. 7286, K. Peffers, M. Rothenberger, and B. Kuechler, Eds., Berlin, Heidelberg: Springer, 2012, pp 398-410.

[19] J. Akoka, I. Comyn-Wattiau, N. Prat, and V. C. Storey, "Knowledge contributions in design science research: Paths of knowledge types," *Dec. Supp. Syst.*, vol. 166, Article 113898, 2023.

[20] M. D. Ahmed, and D. Sundaram, "Design Science Research Methodology: An Artefact-Centric Creation and Evaluation Approach" in *ACIS 2011 Proceedings*, Sydney, Australia, 2011.

[21] J. Barata, P. R. Da Cunha, and A. D. Figueiredo, "Self-reporting Limitations in Information Systems Design Science Research," *Bus Inf Syst Eng*, vol. 65, pp.143-160, 2023.

[22] C. Basile, B. D. Sutter, D. Canavese, L. Regano, and B. Coppens, "Design, implementation, and automation of a risk management approach for man-at-the-end software protection," *Comput. Secur.*, vol. 132, Article 103321, 2023.

[23] M. Bitzer, et al., "Managing the Inevitable – A Maturity Model to Establish Incident Response Management Capabilities," *Comput. Secur.*, vol. 125, Article 103050, 2023.

[24] G. Bou Ghantous, and A.Q. Gill, A. Q., "Evaluating the DevOps Reference Architecture for Multi-Cloud IoT-Applications, "*SN COMPUT. SCI.*, vol. 2, Article 123, 2021.

[25] L. Bunnell, K-M. Osei-Bryson, and V. Y. Yoon, V. Y., "FinPath-light: Framework for an multiagent recommender system designed to increase consumer financial capability," *Dec. Supp. Syst.*, vol. 134, Article 113306, 2020.

[26] B. M. Chaudhry, and J. Smith, "RefineMind: A Mobile App for People with Dementia and Their Caregivers," in: *The Next Wave of Sociotechnical Design. DESRIST 2021*, Lecture Notes in Computer Science (LNCS), vol. 12807, L. Chandra Kruse, S. Seidel, and G. I. Hausvik, Eds., Cham: Springer, 2021, pp 16-21.

[27] Q. Deng, and S. Ji, "A Review of Design Science Research in Information Systems: Concept, Process, Outcome, and Evaluation," *PAJAIS*, vol. 10, 2018.

[28] O F. El-Gayar, and B. D. Fritz, "A web-based multi-perspective decision support system for information security planning," *Dec. Supp. Syst.*, vol. 50, pp. 43-54, 2010.

[29] D. A. Fischer et al., "Towards interactive event log forensics: Detecting and quantifying timestamp imperfections," *Information Systems*, vol. 109, Article 102039, 2022.

[30] J. Forsberg, and T. Frantti, "Technical performance metrics of a security operations center," *Comput. Secur.*, vol. 135, Article 103529, 2023.

[31] T. Matschak, F. Rampold, M. Hellmeier, C. Prinz, and S. Trang, "A Digitization Pipeline for Mixed-Typed Documents Using Machine Learning and Optical Character Recognition," in *The Transdisciplinary Reach of Design Science Research. DESRIST 2022*, Lecture Notes in Computer Science (LNCS), vol. 13229, A. Drechsler, A. Gerber, and A. Hevner, Eds., Cham: Springer, 2022, pp. 195-207.

[32] S. Mdletshe, O. S. Motshweneng, M. Oliveira, and B. Twala, "Design science research application in medical radiation science education," *J. Med. Imaging Radiat. Sci.*, vol. 54, pp. 206-214, 2023.

[33] A. L. Mesquida, and A. Mas, "Implementing information security best practices on software lifecycle processes: The ISO/IEC 15504 Security Extension," *Comput. Secur.*, vol. 48, pp. 19-34, 2015.

[34] T. Mettler, M. Eurich, and R. Winter, "On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework," *CAIS*, vol. 34, pp. 223-240, 2014.

[35] B. Morschheuser, L. Hassan, K. Werder, and J. Hamari, "How to design gamification? A method for engineering gamified software," *Inform. and Soft. Tech.*, vol. 95, pp. 219-237, 2018.

[36] J. A. Moutinho, G. Fernandes, and R. Rabechini, "Evaluation in design science: A framework to support project studies in the context of University Research Centres," *Evaluation and Program Planning*, vol. 102, Article 102366, 2024.

[37] M. Muntean, R-D. Dănăiaţă, and L. Hurbean, "Applying Design Science Research for Developing Business Artifacts," *Proc. Comput. Sci.*, vol. 199, pp. 637-642, 2022.

[38] T. Nagle, C. Doyle, I. M. Alhassan, and D. Sammon, "The Research Method We Need or Deserve? A Literature Review of the Design Science Research Landscape," *CAIS*, vol. 50, pp. 358-395, 2022.

[39] K. Nahar, and A. Q. Gill, "Integrated identity and access management metamodel and pattern system for secure enterprise architecture," *Data & Knowledge Engineering*, vol. 140, Article 102038 2022.

[40] F. K. de Oliveira, M. B. de Oliveira, A. S. Gomes and L. M. Queiros, "RECREIO: Floss as SAAS for sharing of educational resources," in: *Proc. of 12th Iberian Conference on Information Systems and Technologies (CISTI)*, Lisbon, Portugal, 2017.

[41] M. Overeem, M. Mathijssen, and S. Jansen, "API-m-FAMM: A focus area maturity model for API management," *Inform. and Soft. Tech.*, vol. 147, Article 106890, 2022.

[42] R. K. Pallasena, M. Sharma, and V. Krishnaswamy, "A Study of Interaction, Visual Canvas, and Immersion in AR Design: A DSR Approach," AIS Trans. on HCI, vol. 14, pp. 390-425, 2022.

[43] N. Prat, I. Comyn-Wattiau, and J. Akoka, "A Taxonomy of Evaluation Methods for Information Systems Artifacts. *JMIS*, vol. 32, pp.229-267, 2015.

[44] I. G. A. Premananda, A. Tjahyanto, and A. Mukhlason, "Design Science Research Methodology and Its Application to Developing a New Timetabling Algorithm, in: *Proc. of IEEE CyberneticsCom 2022*, Malang, Indonesia, 2022.

[45] J. Pries-Heje, J. Venable, and R. L. Baskerville, "RMF4DSR: A Risk Management Framework for Design Science Research," *Scand. J. of Inform. Syst.*, vol. 26, pp. 57-82, 2014.

[46] S. Pulparambil, Y. Baghdadi, and C. Salinesi, "A methodical framework for service-oriented architecture adoption," *Inform. and Soft. Tech.*, vol. 132, Article 106487, 2021.

[47] A. Shrestha, A. Cater-Steel, M. Toleman, and T. Rout, "The Role of International Standards to Corroborate Artefact Development and Evaluation," in: *Software Process Improvement and Capability Determination. SPICE 2017*, Communications in Computer and Information Science, vol. 770, A. Mas, A. Mesquida, R. O'Connor, T. Rout, and A. Dorling, Eds., Cham: Springer, 2017, pp. 438-451.

[48] P. M. Wiegmann, M. Talmar, and S. B. De Nijs, "Forging a sharper blade: A design science research approach for transition studies," *Environ. Innov. and Soc. Tran.*, vol. 48, Article 100760, 2023.

[49] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, pp. 319-340, 1989.

[50] P. Checkland, and J. Scholes, *Soft systems methodology in action*. Hoboken, New Jersey, USA: John Wiley & Sons.

[51] ISO/IEC/IEEE, *ISO/IEC/IEEE 24765:2017 Systems and software engineering — Vocabulary*, Edition 2, September 2017.

[52] S. Gregor, and A. R. Hevner, "Positioning and presenting design science research for maximum impact," *MIS Quarterly*, vol. 37, pp. 337-355, 2013.