Operationalizing Evaluation in Design Science Research: A Structured Framework for Artifact Assessment in C2-System Development

Jan Lundberg[1,2,*]

[1]Swedish Defence University, Drottning Kristinas väg 37, 114 28 Stockholm
[2]Stockholm University, Borgarfjordsgatan 12, 164 55 Kista, Sweden

Abstract

Emerging technologies such as artificial intelligence (AI) and autonomous systems challenge the design of military command- and control systems (C2-systems). While architectural frameworks like NAF and DoDAF provide structural consistency, they offer limited methodological guidance for system development in evolving socio-technical environments. To address this challenge, this PhD project aims to develop a new framework for C2-system design. Its development is based on a concept- and goal model and it consists of Method Chunks (MCs) and a method navigation guidance. The developed MCs have reached the level of maturity at which they need to be evaluated. Hence, this study proposes a structured evaluation process that targets not only the correctness and alignment of these artifacts, but also their practical utility in C2-system design. The proposed evaluation design is operationalized through a case study simulating the integration of drone-based Intelligence, Surveillance, and Reconnaissance (ISR) capabilities into a legacy C2-system. This setup enables a holistic analysis of the MCs and navigation guidance in a realistic, simulated C2 context. The study contributes to the operationalization of evaluation in DSR and offers methodological insights for improving military C2-system design.

Keywords

Evaluation, Design Science Research (DSR), Nato Architecture Framework (NAF)

1. Introduction

Emerging technologies such as AI, autonomous systems, and data-driven decision support are transforming the design of military command-and control systems (C2-systems) [1], [2], [3]. However, while architectural frameworks like the NATO Architecture Framework (NAF) and the Department of Defense Architecture Framework (DoDAF) provide structural consistency, they offer limited methodological support for adapting to system design in complex, evolving design environments. In particular, they lack practical method guidance and tools for addressing socio-technical challenges and technology integration [4].

This study builds on prior research that developed an initial concepts and goal-oriented framework to support C2-system development [5], [6] . This includes MCs [7], [8] and a MAP-based [9], [10] model for guidance of the method process (navigation guidance). The MCs and navigation guidance together aim to help designers select and apply design strategies grounded in several factors such as stakeholder goals, project goals and requirements, as well as emerging technologies. These artifacts are the product of a Design Science Research [11] (DSR) project that engages enterprise modeling [12] with decomposition of high-level goals to requirements, the construct of MCs following Situational Method Engineering [8], [13] (SME) and validation through stakeholder engagement [14].

Despite the focus of design and evaluation cycles s DSR, scholars have emphasized that evaluation practices often lack coherence and the need for evaluation strategies that are context-driven and applicable to iterative, real-world scenarios [15], [16], [17], [18], [19].

PoEM2025: Companion Proceedings of the 18th IFIP Working Conference on the Practice of Enterprise Modeling: PoEM Forum, Doctoral Consortium, Business Case and Tool Forum, Workshops, December 3-5, 2025, Geneva, Switzerland
*Corresponding author.
jan.lundberg@dsv.su.se (J. Lundberg)
0009-0008-0485-6005 (J. Lundberg)

The framework under development has now reached the level of maturity that can be regarded as the first iteration. Its parts have been validated with stakeholders. The next step in this development cycle is evaluation. Hence, this study addresses that gap by operationalizing evaluation as a structured, multi-dimensional process and uses case study-based evaluation design, in which stakeholders apply the MCs and the navigation guidance to a simulated capability development scenario. The focus lies on assessing not just the correctness and relevance of the artifact, but also its usability and utility in supporting real-world design decisions. The goal of this paper is to present the evaluation design.

The remainder of this article is organized as follows; Section 2 provides the research foundation and introducing the challenge of artifacts evaluation. Section 3 presents the proposed evaluation approach. Section 4 outlines the expected contributions, and Section 5 concludes the study and outlines directions for future work.

2. Current Research and Challenge

2.1. Current Research

Existing architectural frameworks that specifically support military capability development, such as Nato Architectural Framework (NAF), [20] and Department of Defence Architectural Framework (DoDAF) [21], provide structural consistency in the design process but they lack guidance on how to adapt to emerging technology and evolving socio-technical and tactical realities. This PhD research project addresses these shortcomings by proposing a conceptual framework grounded in Design Science Research (DSR) [22].

The research project has progressed through seven articles related to the DSR phases outlined in [23]. Figure 1 shows an overview of the research process and related publications. In the first article [24], C2-system development is framed as a System-of-Systems with Socio-Technical challenges, highlighting the need for integration of new technology into legacy systems. The second article [25], clarifies that military capability is a socio-technical phenomenon, underpinning the need for reasoning beyond technical systems. The third article [26] presented an interview-based study identifying stakeholder needs and challenges and outlines high-level goals that serve as inputs for later modelling activities. Building on this, the fourth article [27] introduces a goal-oriented modeling approach using for enterprise modeling method (4EM) [12], combining a concepts model and a stakeholder-driven goal model. The fifth article [28] deepened this reasoning by a deeper goal decomposition and refined conceptual structures. This lays the foundation for linking high-level design goals to corresponding MCs, enabling traceability and support of design decisions. The sixth article [6] employed focus group-based validation to assess coherence and practical relevance, further refining the models. Finally, the seventh article (not yet published) [5] proposes a design of reusable MCs following Situational method engineering (SME) [29], [30], [13] principles, as well as a MAP-based navigation guidance, further developing the framework.

The proposed envisioned artifact consists of (i) a set of Method Chunks (MCs), (ii) a MAP-based Navigation Guidance, and (iii) an underlying goal- and concept model, validated in [6]. MCs are constructed using SME and capture reusable, domain-specific steps to support C2-system design. The Navigation Guidance functions at a higher abstraction level, providing process support by assists designers select and sequence MCs based on specific design intentions. The framework builds on 4EM by reusing its modeling structures while specializing and tailors them for military capability-development in a C2-system context.

This study contributes to the overarching PhD project by proposing a structured evaluation process, tailored to assess usability, fit, and effectiveness of MCs in real-world or simulated C2-system design settings. While this study focuses on designing the evaluation of MCs and the navigation guidance, it potentially contributes more broadly to ongoing discussions on evaluation practices in DSR.

Several researchers have highlighted that evaluation in DSR occasionally lacks methodological coherence, particularly when applied to complex or methodological artifacts [15], [16]. To address this, recent work emphasizes the need for evaluation strategies that; i) are responsive to artifact type and context [15], [31], ii) consider not only utility but also practical integration [16], and iii) can be adapted

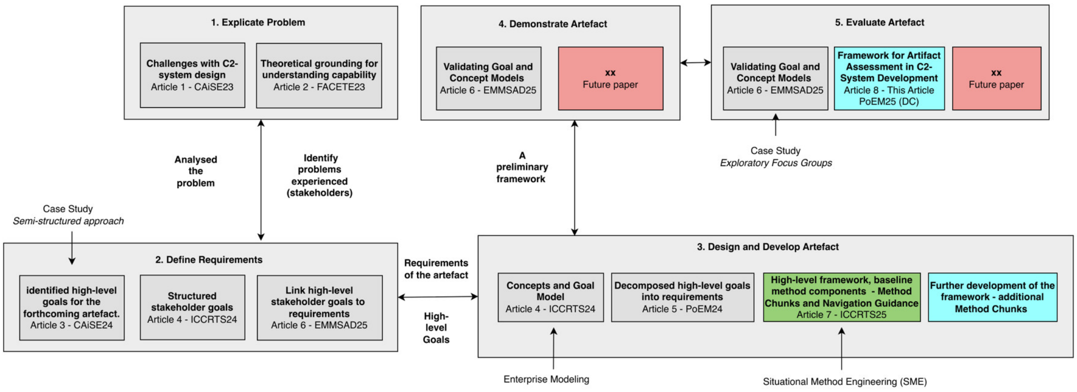

Figure 1: Overview of the research process and related publications

and applied iteratively real-world settings [18], [17]. These insights can be considered requirements for evaluation design, and they also strengthen the relevance of this study by operationalize such principles.

At this stage, the research generated several insights that inform the design of the evaluation and the continued development of the framework. Feedback from stakeholders confirmed that the goal and concept models are relevant, comprehensible, and useful for structuring discussions about capability development [6]. This feedback also indicated that the decomposition from high-level goals to sub-goals and related requirements provides a meaningful foundation for linking goals to method components.

The initial work on Method Chunks (MCs) shows that the structure of intentions, strategies, inputs, and outputs can be expressed coherently and aligned with underlying goal model. Although the full set of MCs is still under development, the early versions outlined in [5], suggest that the approach is suitable for supporting design activities, and that the overall method framework becomes increasingly traceable as the remaining MCs are constructed.

Taken together, these preliminary results indicate that the framework is approaching a level of maturity where a structured evaluation, focusing on practical utility, will be feasible once the remaining MCs have been fully constructed.

2.2. Design Science Research (DSR)

Design Science Research (DSR) is a research framework that focuses on the construction and improvement of artifacts intended to solve real-world problems. Central in DSR is its iterative approach, where the research process cycles through the phases 1) explicate problem, 2), define requirements, 3) design and develop artifact, 4) demonstrate artifact, and 5) evaluate artifact.

Evaluation should not only be understood as a quality control mechanism but also as a source of insight for further artifact development, and the cyclical process enables continuous refinements of the artifact. Without rigorous evaluation, DSR risks producing artifacts that may be relevant in design but misaligned within a real-world application context.

In this study, the artifact under evaluation consists of a set of MCs and the navigation guidance designed to support architectural reasoning in military C2-system development. The DSR framework guides not only the construction of these MCs but also the planned demonstration and evaluation approach.

2.3. The Challenge

This study highlights the challenge of assessing design frameworks, and specifically MCs, not only for modeling-method soundness (verification), and alignment with stakeholder understanding (validation), but also for the framework practical utility in a real-world military context (evaluation). While evaluation in DSR is often framed as a distinct phase [11], [23], the authors in [32] highlights the value of distinguishing between verification, validation, and evaluation (VVE), as each address different yet interdependent aspects of artifact quality.

Verification addresses whether the artefact has been built according to its specification and if it is consistent with existing theories and design principles. This includes structural coherence and cohesion to design rules. Validation assesses whether the artifact leads to outputs that are relevant and meaningful in real-world contexts, i.e., if the output fits the stakeholders' needs and requirements. Evaluation deals with practical utility and explore whether the artifact is effective and worthwhile to use in practice. This includes artifact efficiency and potential added value relative to existing artifacts.

This study adopts the VVE framework as a structured lens for organizing the assessment of MCs and the navigation guidance. MCs are constructed to support reasoning, decision-making, and architectural alignment in C2-system development. However, their effectiveness cannot be assessed easily. Their practical value depends on how they are experienced by stakeholders, if they are understandable, actionable, and seamlessly integrated into existing frameworks.

The challenge, therefore, lies in designing a VVE-informed assessment process that addresses all three quality dimensions. MCs must be verified for internal consistency, validated against stakeholder expectations, and evaluated for their capacity to support an efficient design process. This is particularly important in C2-system development, where supporting frameworks and design tools must operate under time pressure, across socio-technical boundaries, and in environments characterized by rapidly evolving technologies.

The envisioned framework aims to (i) strengthen goal-driven architectural reasoning, (ii) ensure traceability from high-level goals to MC selection and design outputs, (iii) support flexible, intention-driven method enactment, and (iv), because of i-iii, provide structured guidance for integrating emerging technologies into military C2-systems.

3. Proposed Approach

To operationalize the evaluation of the MCs, this study proposes a case study research strategy, anchored in VVE. The focus is not only on the MCs, but also on the navigation guidance, i.e., the previous designed MAP-based model that supports selection of MCs in C2-system design scenarios. Since the navigation guidance guides designers in selecting and sequencing MCs based on design intentions, it becomes important to evaluate both its correctness and utility. In this study, both MCs and the navigation guidance are treated as evaluation subjects across the three VVE dimensions; Verification (V1), Validation, and Evaluation.

The evaluation consists of two phases. First, an assessment targeting the navigation guidance. This phase will engage domain experts to assess whether the navigation guidance structure is logically complete and accurately reflects the intended design pathways, i.e., that each chunk is appropriately linked to each other, and that designer intuitively can traverse the model without confusion. This feedback allows refinement of the Navigation Guidance before its use in the more demanding case study setting. See the Navigation Guidance in Figure 1.

The planned evaluation will be conducted as a case study. While case studies preferably take place in naturalistic settings, conducting one in the context of C2-system development presents notable challenges. The integration of real-world sensors, data flows, decision-making tools, and effectors, spanning all five military domains , is highly complex and not feasible within this research setup. Although other fields also must integrate sensors, data flows, and decision-support tools, military C2-system development includes distinct challenges such as multi-domain coordination (land, air, maritime, cyber, space), adversarial and contested environments (jamming, deception, cyber-attacks), information-classification boundaries, and integration of lethal/non-lethal effectors. These specific challenges necessitate specific method support.

To address this, the scenario used within the case study will be simulated rather than drawn from a real-world setting, designed to realistically reflect relevant design challenges. The evaluation uses a simulated integration of drone-based ISR into a legacy C2 system. While simplified, it reproduces

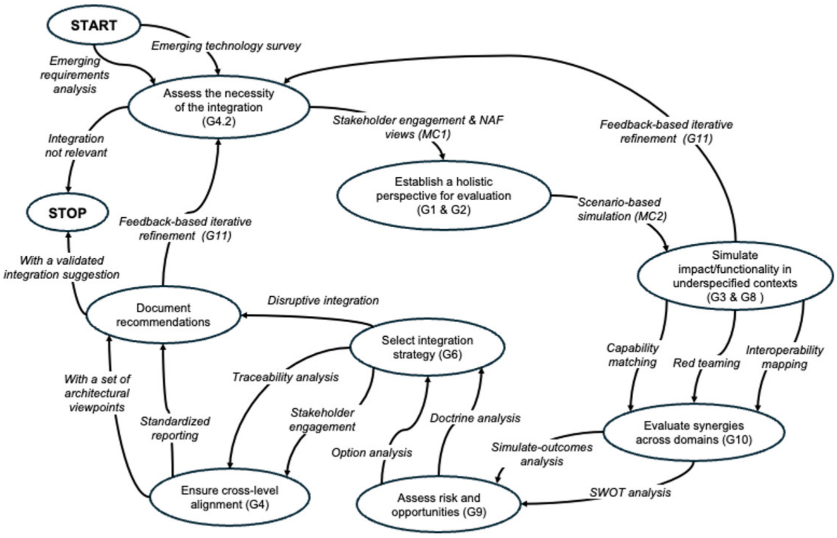

Figure 2: Navigation Guidance Map, reported in [5]

essential characteristics, such as cross-domain information flows, contested-environment, time-sensitive decision cycles and coordination between humans and autonomy systems. This ensures that the scenario remains realistic for capability-development activities while still feasible for controlled evaluation of the MCs and Navigation Guidance. The scenario will centre on relevant capability development challenges, such as mission planning and execution, and integration of new and emerging technologies, allowing MCs and the navigation guidance to be evaluated under authentic conditions. Due to the complexity of actual C2-system environments, the evaluation will focus on a specific process or parts of the system, rather than simulating full system behaviour.

In the scenario, the plan is to use the Navigation Guidance to identify and apply relevant corresponding MCs. Participants will be asked to perform activities such as analysing and refining the method chunks. (e.g., see MC1 in Figure 3).

Throughout the session, empirical data are collected mainly through post-task interviews, and feedback surveys. These are analysed to assess the three dimensions; 1) Verification, Are the MCs and the Navigation Guidance accurate constructed?, 2) Validation, Do the MCs and the Navigation Guidance align with stakeholder needs and design logic in the given scenario?, and 3) Evaluation, Are the MCs and the Navigation Guidance useful, usable, and effective in supporting designers reasoning and structuring work?

Note that VVE as such, consequently, is holistic, assessing not just whether the artifacts work, but whether they are worth using in a real design setting. The table 1 below summarizes how each VVE dimension applies to both the MCs and the Navigation Guidance.

Findings from these VVE activity are used to refine not only MCs and the Navigation Guidance but, if possible, to loop back and refine elements of the goal model and concept model. For example, if a high-level goal, expressed as an intention, in the Navigation Guidance is misunderstood or misapplied, it indicates a need to adjust phrasing or content. In this way, the VVE process does not end, rather it drives an iterative loop of improvement of the framework.

4. Contribution Vision and Next Steps

On a practical level, this study presents an evaluation setup based on a relevant capability development scenario, namely, the integration of an ISR capability into a C2-system. The scenario provides a test bed to explore if MCs and the Navigation Guidance help designers reason through problems and make informed design decisions.

The findings aim not only to refine the envisioned framework but also to contribute insights into

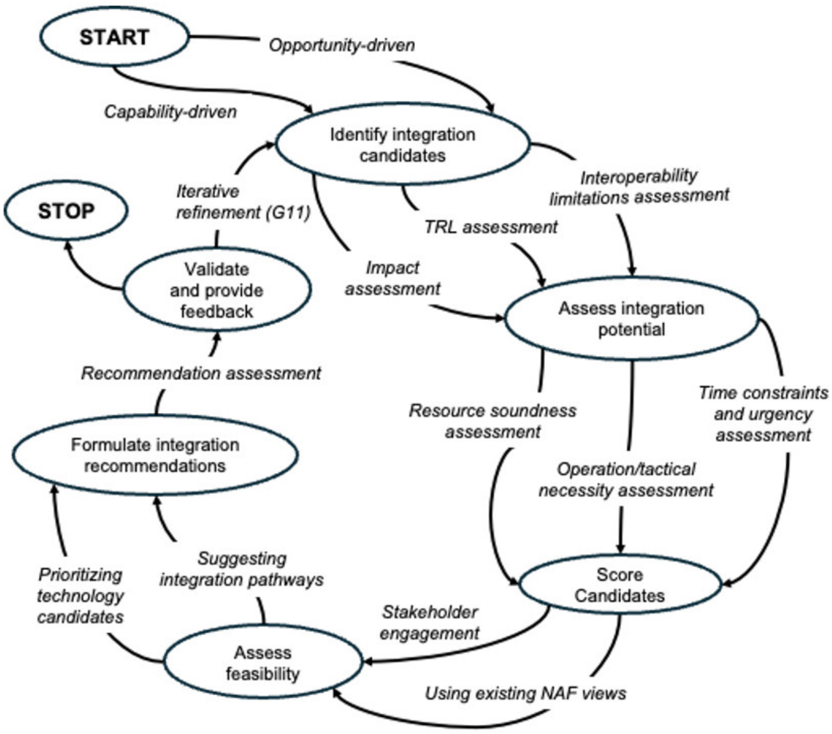

Figure 3: Method Chunk 1 - Integration Assessment, reported in [5]

Table 1
VVE Focus Areas

	Method Chunks – Key Questions	Navigation Guidance – Key Questions
Verification	- Do the method chunks follows the SME principles? - Are all elements (intentions/strategies) complete and traceable?	- Does the Navigation Guidance MAP cover all goals and related method chunks? - Are navigation transitions logically correct and free of gaps?
Validation	- Are the method chunks aligned with stakeholder needs and design challenges? - Do method chunks' outputs address real-world design intentions?	- Are the goal definitions intuitive and semantically aligned with operational logic? - Do users interpret navigation paths as intended?
Evaluation	- Are the method chunks understandable, usable, and are the out-comes relevant? - Do they improve clarity and decision-making in the specific case context?	- Do the Navigation Guidance MAP help users navigate, i.e., find appropriate chunks? - Does the Navigation Guidance MAP support flexible navigation and goal reassessment?

how structured method guidance can be embedded in military C2-system design practices, particularly in contexts aligned with capability development frameworks such as NAF.

This study proposes a structured approach to operationalizing evaluation within DSR, with particular focus on assessing MCs and their supporting Navigation Guidance. The proposed VVE strategy targets the quality of the MCs, aiming to evaluate their usefulness and usability in military C2-system development.

To bridge the gap between theory and practice, the evaluation extends the VVE perspective to include whether the artifact (i.e., the MCs and Navigation Guidance) is worth using in realistic design scenarios. The planned case study, centered on the integration of a drone-based ISR capability into an existing

C2 system, offers a relevant setting for this. It enables the use of both MCs and Navigation Guidance under simulated operational conditions, with active stakeholder engagement in the design process. The evaluation aims to determine whether the artifacts help designers reason effectively and support their design decisions.

Future work will focus on executing the planned case study and refining the framework. This includes revisiting earlier design artifacts, such as the Goal- and Concept Models, the MCs, and the Navigation Guidance.

5. Acknowledgments

This PhD thesis project is a collaboration between Swedish Defence University and Stockholm University. Supervisors of the project are Janis Stirna (Stockholm University) and Kent Andersson (Swedish Defence University).

6. Declaration on Generative AI

During the preparation of this work, the author used Avidnote (https://avidnote.com/) to grammar and spelling check. After using Avidnote, the author reviewed and edited the content as needed and take full responsibility for the publication's content.

7. References

1. J. Weingarten, "Developing future capabilities: Robotics and autonomous systems," 2023. Accessed: Jul. 22, 2024. [Online]. Available: https://www.nato-pa.int/document/2023-robotics-and-autonomous-systems-report-weingarten-034-stctts
2. D. S. Lange, P. Verbancsics, R. S. Gutzwiller, J. Reeder, and C. Sarles, "Command and Control of Teams of Autonomous Systems" Berlin, Heidelberg: Springer, 2012, pp. 81–93. doi: 10.1007/978-3-642-34059-8_4
3. P. Scharre, Army of None: Autonomous Weapons and the Future of War. W. W. Norton & Company, 2018
4. S. Kotusev, "A Comparison of the Top Four Enterprise Architecture Frameworks," 2021, Accessed: Jun. 27, 2025. [Online]. Available: https://www.bcs.org/articles-opinion-and-research/a-comparison-of-the-top-four-enterprise-architecture-frameworks/
5. J. Lundberg, K. Andersson, J. Ralyté, and J. Stirna, "Supporting Future Military Command- and Control (C2) System Design: Defining Method Components to Enhance the Existing Development Framework," in (To appear in) ICCRTS25 - International Command and Control Research & Technology symposium, Stockholm, 2025.
6. J. Lundberg, K. Andersson, and J. Stirna, "Enhancing C2-Systems: Validation of Goal and Concept Models with Stakeholders," Lecture Notes in Business Information Processing, vol. 558 LNBIP, pp. 353–367, 2025, doi: href 10.1007/978-3-031-95397-2_22.
7. J. Ralyté, P. Backlund, H. Kühn, and M. A. Jeusfeld, "Method chunks for interoperability," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 4215 LNCS, pp. 339–353, 2006, doi: 10.1007/11901181_26.
8. B. Henderson-Sellers, C. Gonzalez-Perez, and J. Ralyté, "Comparison of method chunks and method fragments for situational method engineering," Proceedings of the Australian Software Engineering Conference, ASWEC, pp. 479–488, 2008, doi: 10.1109/ASWEC.2008.4483237.
9. C. Rolland, N. Prakash, and A. Benjamen, "A multi-model view of process modelling," Requir Eng, vol. 4, no. 4, pp. 169–187, 1999, doi: 10.1007/S007660050018.
10. C. Rolland, "Capturing System Intentionality with Maps. Conceptual Modelling in Information Systems Engineering," pp. 141–158, 2007, Accessed: Jun. 24, 2025. [Online]. Available: https://hal.science/hal-00706146v1

11. Hevner, March, Park, and Ram, “Design Science in Information Systems Research,” MIS Quarterly, vol. 28, no. 1, p. 75, 2004, doi: 10.2307/25148625.
12. K. Sandkuhl, J. Stirna, A. Persson, and M. Wißotzki, “Enterprise Modeling Tackling Business Challenges with the 4EM Method,” 2014. doi: 10.1007/978-3-662-43725-4.
13. B. Henderson-Sellers, J. Ralyté, P. J. Ågerfalk, and M. Rossi, “Situational method engineering,” Situational Method Engineering, pp. 1–310, Jan. 2014, doi: 10.1007/978-3-642-41467-1.
14. J. Lundberg, K. Andersson, and J. Stirna, “Enhancing C2-Systems: Validation of Goal and Concept Models with Stakeholders,” in Enterprise, Business-Process and Information Systems Modeling. BPMDS EMMSAD 2025, R. Guizzardi, A. Sturm, L. Pufahl, and H. van der Aa, Eds., Vienna: Springer, Jun. 2025.
15. N. Prat, I. Comyn-Wattiau, and J. Akoka, “Artifact evaluation in information systems design-science research - A holistic view,” PACIS 2014 Proceedings, Jan. 2014, Accessed: Oct. 01, 2025. [Online]. Available: https://aisel.aisnet.org/pacis2014/23
16. J. Müller, S. Wurth, T. Schäffer, and C. Leyh, “Toward a Framework for Determining Methods of Evaluation in Design Science Research,” Annals of Computer Science and Intelligence Systems, no. 2024, pp. 231–236, 2024, doi: 10.15439/2024F7208.
17. S. Mdletshe, O. S. Motshweneng, M. Oliveira, and B. Twala, “Design science research application in medical radiation science education: A case study on the evaluation of a developed artifact,” J Med Imaging Radiat Sci, vol. 54, no. 1, pp. 206–214, Mar. 2023, doi: 10.1016/j.jmir.2022.11.007.
18. J. da A. Moutinho, G. Fernandes, and R. Rabechini, “Evaluation in design science: A framework to support project studies in the context of University Research Centres,” Eval Program Plann, vol. 102, p. 102366, Feb. 2024, doi: 10.1016/J.EVALPROGPLAN.2023.102366.
19. J. Venable, J. Pries-Heje, and R. Baskerville, “A comprehensive framework for evaluation in design science research,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7286 LNCS, pp. 423–438, 2012, doi: 10.1007/978-3-642-29863-9_31.
20. Nato, “Nato Architecture Framework Version 4,” 2020.
21. “DODAF - DOD Architecture Framework Version 2.02.” Accessed: Jun. 13, 2024. [Online]. Available: https://dodcio.defense.gov/library/dod-architecture-framework/
22. A. Hevner and S. Chatterjee, “Design Research in Information Systems,” vol. 22, 2010, doi: 10.1007/978-1-4419-5653-8.
23. P. Johannesson and E. Perjons, “An Introduction to Design Science,” 2014. doi: 10.1007/978-3-319-10632-8.
24. J. Lundberg, “Towards a Conceptual Framework for System of Systems,” in CEUR Workshop Proceedings, 2023. Accessed: May 01, 2024. [Online]. Available: https://ceur-ws.org/Vol-3407/paper3.pdf
25. K. Andersson, J. Lundberg, and J. Stirna, “Emerging technology calls for a systemic view on military capability,” G. Poels, J. Van Riel, and R. Fernandes Calhau, Eds., Vienna: CEUR-WS.org, 2023. Accessed: Apr. 04, 2024. [Online]. Available: https://ceur-ws.org/Vol-3645/facete2.pdf
26. J. Lundberg, J. Stirna, and K. Andersson, “Designing Military Command and Control Systems as System of Systems – An Analysis of Stakeholder Needs and Challenges,” Advanced Information Systems Engineering, pp. 336–351, Jun. 2024, doi: 10.1007/978-3-031-61057-8_20.
27. J. Lundberg, J. Stirna, J. Zdravkovic, and K. Andersson, “Beyond Technology: Goal-Oriented Analysis for Integrating Emerging Technologies into Military Command and Control Systems,” in 29th International Command and Control Research and Technology Symposium, London, Sep. 2024.
28. J. Lundberg, S. Hacks, and K. Andersson, “Refinement of a Conceptual Model of a military C2-system through Low-Level Goal Decomposition,” in Companion Proceedings of the 17th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modeling Forum, 2024. Accessed: Dec. 04, 2024. [Online]. Available: https://ceur-ws.org/Vol-3855/forum9.pdf

29. S. Brinkkemper, “Method engineering: engineering of information systems development methods and tools,” Inf Softw Technol, vol. 38, no. 4, pp. 275–280, Jan. 1996, doi: 10.1016/0950-5849(95)01059-9.
30. J. Ralyté and C. Rolland, “An approach for method reengineering,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2224, pp. 471–484, 2001, doi: 10.1007/3-540-45581-7_35.
31. J. Venable, J. Pries-Heje, and R. Baskerville, “A comprehensive framework for evaluation in design science research,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7286 LNCS, pp. 423–438, 2012, doi: 10.1007/978-3-642-29863-9_31.
32. J. Ralyté, G. Koutsopoulos, and J. Stirna, “Verification, validation, and evaluation of modeling methods: experiences and recommendations,” Software and Systems Modeling, 2025, doi: 10.1007/S10270-025-01304-2.