Connecting the Dots: Towards a Holistic Data Standardization Methodology in Public E-Procurement (HoDS)

Cedric Pauken
University of Koblenz
cpauken@uni-koblenz.de

Andreas Schmitz
University of Koblenz
andreasschmitz@uni-koblenz.de

Renzo Kottmann
Coordination Office for IT-Standards
renzo.kottmann@finanzen.bremen.de

Maria A. Wimmer
University of Koblenz
wimmer@uni-koblenz.de

Abstract

The use of isolated ICT solutions in public e-procurement has led to poor interoperability between organizations, processes, and systems. Consequently, data inconsistencies arise, and data cannot be reused throughout the procedure. While data standardization is recognized as a key enabler to overcome these issues, existing approaches often operate in isolation and thus fail to ensure holistic interoperability. This paper harmonizes four complementary methodological components—the European Interoperability Framework (EIF), the Framework for Interoperable Service Architecture Development (FISAD), Requirement Engineering in complex public sector structures, and Simple Semantic Data Modeling in XML (SeMoX)—into a holistic data standardization methodology (HoDS). Using Design Science Research, HoDS is grounded in rigorous and relevant foundations and is developed iteratively in coordination with European and German standardization bodies. With its modularity and interdependency, HoDS supports agile adaptation to dynamically changing requirements. Applying HoDS to the specification of an e-catalogue standard shows its practical usefulness.

Keywords: Data Standardization, Interoperability, E-Procurement, Methodology, Design Science Research

1. Introduction

Recent OECD statistics reveal that public procurement accounts for roughly 13% of worldwide gross domestic product as well as 29% of government expenditures in 2023 (OECD, 2023). Therefore, major interest is in saving costs and increasing process efficiency (Prier & McCue, 2008). By utilizing electronic means such as standardized data formats and web-based applications, the procurement process can be conducted electronically (Bobowski & Gola, 2018; Tatsis et al., 2006). While current developments such as electronic catalogues have the potential to make public e-procurement more efficient (Mavidis & Folinas, 2022; Olivares et al., 2024), still a variety of challenges must be overcome. Notably, the use of standalone island solutions hampers interoperability between existing processes and technologies (Babica et al., 2019; Ferreira & Amaral, 2016). Furthermore, unstructured process and data transitions cause media disruptions and data inconsistencies throughout the procurement process. Consequently, persistent procurement information must be re-entered and re-verified frequently (Siapera et al., 2024). Recent research has examined the vital role of interoperability (Campmas et al., 2022; Margariti et al., 2022) and standardization (Egorova et al., 2021; Schmitz et al., 2025; Siapera et al., 2024) in overcoming these challenges. However, the specification of useful, cost-effective and interoperable data exchange standards requires holistic considerations on the interoperability of heterogenous architectures and layers. Such standardization efforts must examine legal requirements, organizational processes and systems, semantic accuracy, and technical feasibility (European Commission, 2017; Schmitz et al., 2025). Hence, several guiding frameworks and methodologies have emerged from both research and practice in recent years. Examples include the Framework for Interoperable Service Architecture Development (FISAD) (Schmitz & Wimmer, 2023), the European Interoperability Framework (EIF) (European Commission, 2017), and Simple Semantic Data Modeling in XML (SeMoX) (Kottmann et al., 2024). However, these frameworks and methodologies are designed to function only within their specific contexts. Their integration into a holistic data standardization approach requires further harmonization.

Thus, the objective of this paper is to align existing frameworks and methodologies towards a holistic data standardization methodology in the domain of public e-procurement. Two research questions guide the research: **RQ1:** Which existing frameworks and methodologies are suitable for the specification of interoperable data exchange standards

URI: https://hdl.handle.net/10125/111680
978-0-9981331-9-5

in public e-procurement? **RQ2:** How can existing frameworks and methodologies be harmonized towards a holistic data standardization methodology?

Using Design Science Research (Hevner, 2007; Hevner et al., 2004) to answer these RQs (see Section 2), we provide three key contributions to the field: a), Section 4 provides a consolidated overview of existing frameworks and methodologies for specifying interoperable data exchange standards. b) Section 5 harmonizes this overview in a chronological process that helps researchers and practitioners alike to specify interoperable data exchange standards in public e-procurement. c) We apply the theoretical contributions in the use-case of "XKatalog" in Section 6. This interoperable data exchange standard, developed through the application of HoDS, improves the German procurement process by bridging formerly unstructured process and data transitions along the overall procurement procedure. Finally, we synthesize these key contributions in the context of the research questions and provide indications on limitations and future research in Section 7.

2. Methodology

The harmonization of data standardization in public e-procurement requires sound methodological foundations, rooted in both rigor and relevance. As such, Design Science Research (DSR) is identified as a suitable research paradigm for addressing both research questions (Hevner, 2007; Hevner et al., 2004). The DSR's rigor cycle provides scientific grounding of the harmonization via a systematic literature review, whereas the DSR's relevance cycle ensures identification of the practical needs and the validation of the proposed methodology from standardization experts. The rigor cycle's systematic literature review combines approaches of Webster & Watson (2002) and Carrera-Rivera et al. (2022). It is divided into the phases of planning, conducting, and results. We query SCOPUS, IEEE, and Google Scholar. Search strings and concept-related keywords are derived from the research objective, and a list of preliminary inclusion criteria as well as a final quality assessment criteria checklist are defined to evaluate found literature. In the first cycle of the conducting phase, concept related keywords are used to query the specified sources, such as *"Data Standardization"*. In the second cycle of the conducting phase, a search string to identify current challenges and barriers in public e-procurement that relate to interoperability and standardization is utilized: *(((("electronic procurement" OR "E-Procurement") AND "public") OR ("electronic government procurement" OR "e-GP")) AND ("challenges" OR "barriers" OR "issues") AND ("data" OR "standardization" OR "interoperability"))*

Lastly, based on the identified problem scope of the second cycle, required types of methodological components are queried via another keyword-based search. Found literature of the three cycles is subsequently evaluated based on the list of preliminary inclusion criteria. For an article's preliminary inclusion, its abstract must be concerned with this paper's research objective, it must be a peer-reviewed article, and it must be written in English. Afterwards, the quality of each paper is assessed by analyzing whether research objectives are clearly defined, limitations are discussed, verification of proposed results via use-cases from practice, proof-of-concept or similar means is conducted, and significant contributions in answering this paper's research objective are made.

In the DSR's relevance cycle, input and feedback from experts in data standardization and public e-procurement are gathered in several practice-oriented project workshops following (Ørngreen & Levinsen, 2017). Supplementing the theoretical findings of the rigor cycle with insights into practically relevant problems, needs and solutions aids in grounding the proposed harmonization in the relevant application context of public e-procurement in Germany and the EU. Participants of the workshops include members of the European Committee for Standardization (CEN), the German Coordination Office for IT-Standards (KoSIT), and several federal authorities and private enterprises active in public procurement solutions, especially on catalogue standardization. Finally, the DSR's design cycle considers the results of the rigor and relevance cycles to iteratively harmonize existing frameworks and methodologies towards a holistic data standardization methodology in coordination with European (CEN) and German (KoSIT) standardization bodies. Moreover, the proposed data standardization methodology is applied to the real-world context of specifying a national e-catalogue standard for public e-procurement in Germany. The case of the "XKatalog" verifies the methodology's agility, adaptability and effectiveness in practice.

3. Theoretical background

Public procurement refers to the acquisition of goods, works and services by public sector organizations from other organizations, including private, public and semi-public bodies (Uyarra & Flanagan, 2010). Public e-procurement digitizes this process via the use of electronic means, particularly Web based applications (Tatsis et al., 2006). In the European Union, the public e-procurement process is divided into the pre-award and post-award phases, where the first denotes the part of tendering, where economic operators submit their offers, whereas the second part starts after an economic operator is awarded with a contract by the contracting authority. The pre-award phase consists of e-sourcing, e-notification, and e-tendering, whereas the post-award comprises e-contracting, e-ordering, e-invoicing and e-payment (Bobowski & Gola, 2018; OpenPEPPOL AISBL, 2023; Prałat, 2019).

Interoperability is an important concept within the domain of public e-procurement, as it enables public administrations and economic operators to efficiently share and exchange data with each other (Pardo et al., 2012). In this paper, we rely on the definition of interoperability in the European Interoperability Framework (EIF): "the ability of organizations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organizations, through the business processes they support, by means of the exchange of data between their ICT systems" (European Commission, 2017, p. 5). Interoperability is crucial for the successful digital transformation of public administrations and governments due to its facilitation of seamless services between public and non-public bodies across traditionally often rigid public structures (Margariti et al., 2022; Pardo et al., 2012).

Furthermore, data standardization is a key means for achieving interoperability (DeNardis, 2010; Gal & Rubinfeld, 2019). Standardization requires consensus across all involved parties to regulate aspects of terminology, structure, attributes and use of data in a specific context (Gal & Rubinfeld, 2019). Two important concepts arise within data standardization: semantics and syntax. Data syntax is necessary to standardize the structure of messages (Aasa et al., 1988), whereas data semantics standardize the meaning and the use of their content (Wood, 1985). When both syntax and semantics are standardized, data becomes machine-readable (Janner et al., 2008), forming the basis for automated processing, validation and transformation of data (Bowers & Ludäscher, 2004). However, when data is insufficiently standardized, interoperability in intra-organizational ecosystems becomes severely hampered (Dzierzawa et al., 2025). In the public sector, the degree to which technical standards are open and accessible influences the ability of governments to perform electronic services efficiently and cost-effectively (DeNardis, 2010). Finally, data standardization is a key facilitator for public e-procurement as it assists in overcoming current barriers such as data inconsistencies and unregulated data transitions in the process (Egorova et al., 2021; Schmitz et al., 2025; Siapera et al., 2024).

4. Identification and Analysis of methodological components

In this section, results of the incrementally structured literature review and workshops, in accordance with the rigor and relevance cycles of DSR, serve as the foundation for addressing **RQ1**. We analyze the problem environment of data standardization and map methodological solution components onto identified needs. Evaluations on suitable standardization approaches, interoperability frameworks, and requirement engineering methods are discussed in the following sections, too.

4.1 Data Standardization Approaches

Based on findings in both literature and practice, approaches to data standardization utilize either data modeling, human-readable documents, or a combination of both. On the one hand, the data modelling approach aims to collect semantic data in a structured syntax to enable machine readability (Peckham & Maryanski, 1988). Resulting data models are often based on Unified Modeling Language (UML) (Halpin & Bloesch, 1999) or Extensible Markup Language (XML) (Ma & Yan, 2007), and thus facilitate machine-to-machine communication (Gal & Rubinfeld, 2019) as well as process automation (Jämsä-Jounela, 2007). In German standardization practice of the public sector, the Coordination Office for IT-standards uses XÖV (XML in der öffentlichen Verwaltung), a model-driven approach based on UML, to specify data standards for G2G and G2B. While this enables automatic validation and generation of further technical artifacts, domain experts struggle to read and understand such data models (Bajaj, 2004; Eichelberger & Schmid, 2009; Henning, 2018). Consequently, experts with little technical expertise are unable to comprehensively review the resulting models, which may lead to semantic inconsistencies and errors in the final data standard.

On the other hand, the semantic-driven document approach prioritizes human readability by writing down semantic data in text-based document formats, such as Microsoft Word or PDF (International Organization for Standardization, 1996). In practice, the European Committee for Standardization (CEN) exclusively uses such textual documents to define legislative standards within Technical Committees (TC). In the field of public e-procurement, EN16931 for electronic invoicing has been standardized in a document approach by TC 434, with further European norms under development for pre- and post-award phases by TC 440. These de-jure standards define legislative specifications that European Member States shall adhere to due to the right of primacy of EU-law over national law (Avbelj, 2011). The document format of European Norms contributes to high human-readability and thus ensures that all domain experts can understand and contribute to the norms. However, several other issues arise due to the lack of machine-readability. There are little to no automation possibilities within the standardization process. Hence, data inconsistencies in the semantic data specifications must be validated and maintained manually across the Norm's whole lifespan. In practice, this commonly leads to copy-paste errors and unnecessary labor for both European TC members managing, as well as national standardization officers implementing EN16931. Furthermore, deciding on the right technical syntax for these semantically driven specifications has proven difficult. Currently, the norms semantics must be bound to multiple data syntaxes simultaneously to ensure practical feasibility.

The harmonization of both data model and document-driven standardization efforts has the potential to overcome the barriers inherent in each, thereby facilitating a more holistic approach to data standardization (Schmitz et al., 2025). Consequently, **Simple Semantic Data Modeling in XML (SeMoX)** (Kottmann et al., 2024) was created as an open-source approach to data standardization with the goal of bridging the gap between domain experts and technical implementations. By applying a divide-and-conquer philosophy onto the data standardization process, domain experts can contribute semantic knowledge to the standard without reading structured data models, and technical experts retain full freedom for technical implementation of the semantics. This is possible by the separation of a shareable semantic model in XML from other technical artifacts. At its core, SeMoX' semantic XML model consists of five modular components: "Terms", "Rules" and "Structures" are used to capture semantic knowledge from domain experts, whereas "Semantic Datatypes" and "Syntax-Bindings" are crucial to the standard's technical implementation and aim to bridge the gap between domain knowledge and technical expertise. A generator enables the XML model's conversion into a human-readable PDF specification, which can subsequently be read and reviewed by domain experts. Since the model itself is an XML-instance, the use of XML-Schema helps to automatically identify data inconsistencies in large data structures. Furthermore, the visualization of business documents such as electronic invoices can be rendered into an HTML-based web-editor via the use of further tooling. SeMoX has already proven its viability in practice through its use for implementing the EN16931 on electronic invoicing as well as the "Commission Implementing Regulation (EU) 2022/2303 of 24 November 2022 amending Implementing Regulation (EU) 2019/1780" on electronic public contract notices (eForms) in Germany. Consequently, the incorporation of SeMoX into the data standardization process facilitates the contribution of both domain and technical experts to the standard and overcomes barriers of purely data model or document driven standardization approaches (Kottmann et al., 2024; Schmitz et al., 2025).

4.2 Guiding Interoperability Frameworks

Failing to achieve a sufficient degree of interoperability leads to isolated solutions that are unable to effectively communicate with each other (Babica et al., 2019; Ferreira & Amaral, 2016). The current lack of interoperability in public e-procurement causes media disruptions, data inconsistencies, and consequently, poor reusability of data along the whole procedure (Schmitz et al., 2023; Siapera et al., 2024). While data standardization has been discussed as a potential solution component to address such issues (Egorova et al., 2021; Schmitz et al., 2023; Siapera et al., 2024), standardization requires a high degree of interoperability to achieve meaningful impact in public e-procurement (Paľová & Vejačka, 2022). As a result, guiding interoperability frameworks have emerged from both literature (Schmitz & Wimmer, 2023) and practice (Campmas et al., 2022; European Commission, 2017). Such Frameworks aim to ensure that public sector systems, tools and standards are developed to be interoperable with existing laws, organizations and solutions. Data standards that focus on only a specific layer of interoperability may become blind to the overall environment in which they should operate. Highly technical data standards may lack adoption due to poor semantic and organizational interoperability (Amar et al., 2024). Legal standards may be difficult to bind to existing technical syntaxes and domain accurate semantics (Ali et al., 2024). Semantically driven standards may become organizationally or technically unfeasible in practice (Folmer et al., 2011). Hence, the **European Interoperability Framework (EIF)** was proposed by the European Union to guide public services and standards development to achieve interoperability on all layers of interoperability: legal, organizational, semantic, and technical interoperability, alongside integrated public service governance and interoperability governance. Furthermore, the EIF provides twelve guiding interoperability principles (European Commission, 2017). In addition to the EIF, the European Interoperability Reference Architecture

(EIRA) offers a reference model for solution architectures in public administrations (European Commission, 2024). Since EIF and EIRA aim to facilitate interoperability on a general level, they cannot address all niches of public service development in sufficient detail (Ali et al., 2024). Notably, the use of Enterprise Architecture is crucial for considering the impacts of developed ICT solutions and for achieving architectural interoperability, also in the public sector (Dang & Pekkola, 2017).

Without Enterprise Architecture, data standards that only focus on data interoperability may fail to operate in a productive environment: without baseline and target architecture models on the business layer, data standards cannot be integrated into processes; without considering applications, they cannot be adopted by tools and systems; without evaluating suitable technical components and networks, they cannot be sent and received properly (Schmitz & Wimmer, 2023). For this reason, the **Framework for Interoperable Service Architecture Development (FISAD)** has been proposed in literature (Schmitz & Wimmer, 2023). Rooted in Design Science Research, in the Open Group Architecture Framework (TOGAF) and in the EIF, FISAD provides a domain-specific methodology for interoperable Enterprise Architecture development in the public sector. FISAD has proven to be particularly suitable for data standardization in public e-procurement (Schmitz et al., 2025).

In line with Schmitz and Wimmer (2023), FISAD consists of five main perspectives of architecture development: (1) foundation, (2) business, (3) data, (4) application, and (5) technical. In the (1) foundation perspective, the architecture's problem scope and target vision are established. To do so, scientific literature and relevant practical sources are analyzed alongside legal frameworks and high-level decisions to establish scientific best-practices and legal interoperability. In the (2) business perspective, baseline and target processes as well as their information flows are captured to ensure organizational interoperability. Information flows are subsequently mapped to standardized data specifications in the (3) data perspective, thereby establishing semantic and syntactic interoperability to existing standards. The (4) application perspective aims to identify application components that are suitable for supporting the identified work- and information flows. Together with the (5) technical perspective that defines the target infrastructure, both perspectives enable technical interoperability. The resulting artefacts can be defined as either a solution or a reference architecture (Schmitz & Wimmer, 2023).

4.3 Requirement Engineering

In line with ISO/IEC/IEEE 42010, to ensure that a standard complies with architectural perspectives, requirements should be engineered, and the standard's specification measured against them (International Organization for Standardization, 2022). Requirement engineering requires the incorporation of demands from a vast number of stakeholders, viewpoints, and objectives into the software, system, and standard development process (Pandey et al., 2010). In the context of the public sector, Klier et al. (2017) highlight that ICT requirements evolve dynamically in the requirement engineering process due to their complex and ever-changing sociotechnical environment. In public e-procurement specifically, conflicting demands from legal norms, public and private sector organizations, and legacy and target systems pose a significant challenge to the specification and coordination of requirements (Moe et al., 2017). Consequently, Klier et al. (2017) propose a **holistic view on requirement engineering**. This approach can adapt to evolving requirements by prioritizing iterative and extensive communication between stakeholders and establishing a central requirement steering group. Notably, this steering group is tasked with coordinating the requirements and exercising IT-governance functions that span across different levels of nationality. Hence, resulting requirements exhibit both high quality and high interoperability (Klier et al., 2017). The integration of actors from a European, national, federal and scientific level has been identified as a suitable constellation for coordinating requirements on standardization efforts in public e-procurement in Germany.

In line with Klier et al. (2017), the core requirement engineering process itself consists of six main phases. First, during the requirement elicitation (Pandey et al., 2010), input from the customer/user, the environment, the process and data exchange to be supported, and the technicality of the system should be carefully considered. Afterwards, elicited requirements are further structured, classified and prioritized. A subsequent analysis evaluates whether the consolidated requirements are written in clear language and do not face any obvious legal or technical limitations. Lastly, a highly iterative demonstration, evaluation and re-engineering process loop takes place. Requirements and their prioritization are discussed within the central steering group, and conflicting perspectives on certain requirements are resolved through mediation. The process ends with a list of substantiated requirements that are evaluated to be complete, correct, consistent and feasible in practice (Klier et al., 2017).

5. Towards a Holistic Data Standardization Methodology (HoDS)

The results of **RQ1** are synthesized next to subsequently discuss how existing frameworks and methodologies can be harmonized in a holistic data standardization methodology (HoDS). Data standardization efforts in Public E-Procurement operate in complex, multi-actor environments (Klier et al., 2017; Moe et al., 2017) and depend on the interoperability of several layers (Ali et al., 2024) and architectural perspectives (Schmitz & Wimmer, 2023). Both domain and technical experts must be able to contribute without being forced into unfamiliar tools (Henning, 2018). Creators of data standards should review all layers of interoperability to not become blind to legal regulations or semantic feasibility (Ali et al., 2024; European Commission, 2017). Baseline and target architecture of the standard's environment must be evaluated to understand how integration into both existing and future processes, systems and networks is possible (Schmitz & Wimmer, 2023). To ensure compliance with all considerations, requirements should be formulated, coordinated, and implemented (Klier et al., 2017; Moe et al., 2017). In adhering to these considerations, several methodological components have emerged from literature. SeMoX can be used to bridge the gap between domain and technical experts (Kottmann et al., 2024; Schmitz et al., 2025). The EIF provides guidance in achieving overarching interoperability across legal, organizational, semantic and technical layers (European Commission, 2017). FISAD facilitates the standard's architectural interoperability in accordance with the EIF and along DSR to ensure rigorous, relevant and interoperable development of architectural models (Schmitz & Wimmer, 2023). Requirement engineering enables the documentation, management and implementation of identified requirements (Klier et al., 2017). Yet, when developing data standards, it is imperative to utilize methodological components in a cohesive manner. If not done correctly, data standards will fail to deliver value at scale, and they risk becoming isolated "island solutions" that further contribute to the lack of interoperability in public e-procurement (Babica et al., 2019; Ferreira & Amaral, 2016; Margariti et al., 2022). To overcome these issues and answer **RQ2**, existing methodological components must be aligned and integrated into a coherent and actionable methodology. In line with the DSR's design cycle, the **Holistic Data Standardization Methodology (HoDS)** presented hereafter was incrementally developed based on the rigorous and relevant foundations of **RQ1** and iteratively validated by members of the CEN and KoSIT. An overview of the proposed HoDS methodology is depicted in **Figure 1**.

The **European Interoperability Framework (EIF)** (European Commission, 2017) provides the overarching structure and foundational principles that inform all downstream methodological activities. The framework ensures that standardization initiatives are not limited to purely technical considerations but instead remain aware of the broader sociotechnical context in which they operate. Although not part of the core methodological components, the EIRA (European Commission, 2024) can act as a complementary resource to the EIF by offering standardized interoperability architecture building blocks. While EIF and EIRA offer a strong conceptual foundation, these do not offer sufficient operational guidance for structuring the architecture of standardization endeavors. Therefore, we use the **Framework for Interoperable Service Architecture Development (FISAD)** (Schmitz & Wimmer, 2023) to conduct further architectural evaluations on the standard in a Design Science Research approach to avoid architectural blindness. Since FISAD is rooted within principles of the EIF already, the combined application of both frameworks yields a comprehendsive understanding of the needs, constraints, and systemic interdependencies that should be addressed in a data standard. To ensure that such insights are systematically translated into actionable specifications, a robust **requirement engineering** process as proposed by Klier et al. (2017) is added. To manage the landscape of dynamic policies, heterogeneous stakeholders, and multi-layered standardization approaches in public e-procurement, the proposed methodology adopts an iterative, governance-based requirement engineering approch. A central steering group with an IT-governance function should be established to coordinate the elicitation, negotiation, and consolidation of all requirements across international, national, and federal levels. Through repeated cycles of analysis, validation, and revision, the standard's requirements can be refined until they are complete, consistent, and practically feasible (Klier et al., 2017). This enables the development of the actual data standard using **SeMoX (Simple Semantic Data Modeling in XML)** (Kottmann et al., 2024; Schmitz et al., 2025). SeMoX enables both machine and human-readability of the data standard by writing the core model in XML and using transformation tools to convert markup into PDF specifications, Excel tables and HTML visualizations. Consequently, human-readability of the data standard facilitates its iterative review from non-technical domain experts. Machine-readability results in automatic transformation and validation capabilities.

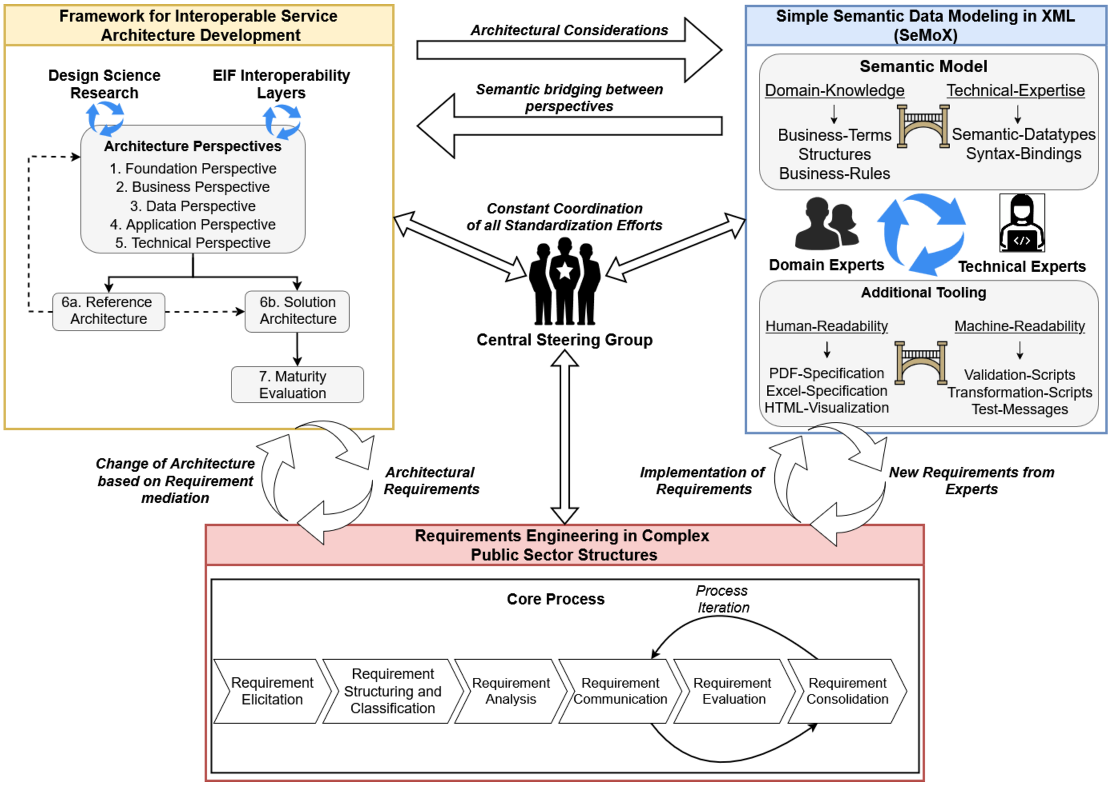

Figure 1. Holistic Data Standardization Methodology (HoDS)

Since the modeling process is directly based on the previously defined requirements, every core component of SeMoX's semantic model (terms, rules, structures, datatypes and syntax bindings) serves a defined purpose in implementing the requirements and aligning the standard with the broader interoperability. The combined application of FISAD and SeMoX enables synergies, as SeMoX can provide a semantic-driven bridging between the different perspectives of FISAD, while FISAD can position SeMoX into a comprehensive architectural consideration. As a result, the final data standard is aligned with the overarching vision of interoperability, can be validated against its consolidated requirements, and reviewed in line with the EIF's interoperability principles. It is only when a standard has been demonstrated to successfully pass all evaluations that it can be regarded as a mature and deployable solution that fosters interoperability in practice.

Given the complex and interdisciplinary nature of data standardization in the public sector, HoDS is designed to be flexible and adaptive, recognizing that change requests and conflicts of interest can occur frequently in the standardization process. The inter-connectivity between its standalone methodological components, achieved through the definition of clear references and dependencies, enables the agile identification of downstream implications whenever upstream changes occur. For example, when new legal requirements arise from the EU, such as the release or revision of an EN, the central steering group can analyze and coordinate necessary changes across the architecture, requirements and SeMoX specification. During SeMoX-based expert workshops, if domain experts identify semantic inconsistencies, new requirements can be formulated, consolidated, and integrated into both the semantic model and enterprise architecture. Consequently, the proposed methodology can quickly and accurately respond to inevitable errors, gaps, or changes during the standardization process without compromising its overall integrity. Through the harmonized, modular, and adaptive approach of HoDS, it becomes possible to not only develop meaningful data standards, but to also maintain and evolve them sustainably over time.

6. Specification of the XKatalog as a use-case from practice

To validate the proposed methodology in practice, the XKatalog has been specified as a national e-catalogue standard for public e-procurement in Germany. In line with legal and scholarly definitions, e-catalogues generally refer to information that is organized electronically in a catalogue-like format for utilization during both pre- and post-award phases of the procurement procedure (European Parliament, 2014; Pekolj et al., 2019; Siapera et al., 2024). Notably, e-catalogue standards are a crucial solution component for regulating the transition between pre- and post-award phases (Siapera et al., 2024). Their implementation has been shown to increase process efficiency and competition while simultaneously reducing administrative costs (Hartati et al., 2020; Olivares et al., 2024; Vaidya & Campbell, 2016; Varney, 2011). Hence, the specification of the XKatalog addresses overarching interoperability issues and contributes towards the harmonization of public e-procurement in Germany. In applying the HoDS methodology to the XKatalog's specification, several implications arise. First, the overarching context of the EIF alongside the architectural considerations of FISAD have been vital in examining the required interoperability of the standard. Legal interoperability was analyzed by consulting European and German law, organizational interoperability was iteratively reviewed with external enterprises and contracting authorities from practice, semantic interoperability was incrementally captured in expert workshops on public e-procurement and e-catalogues, and technical interoperability was achieved by integrating exclusively open specifications and standards into the XKatalog, which is set to become an open standard itself in 2026. Furthermore, overarching public service governance is implemented via the central IT-Governance steering group that coordinated the standardization effort across all interoperability layers. The steering group further coordinated the standard's enterprise architecture resulting from the application of FISAD and ensured utmost architectural interoperability with legacy and target processes, applications and systems. The flexible and iterative requirements engineering approach maintained agile responsiveness to an ever-changing environment of moving targets across all layers of interoperability and architecture. SeMoX has proven to be exceptionally viable in capturing semantic knowledge from domain experts that have no technical expertise while simultaneously simplifying the XKatalog's technical implementation and tooling for technical experts.

Although HoDS has offered conceptual benefits, notable barriers to its adoption for XKatalog include the reliance on a pre-established expert network, securing political funding and commitment, and stakeholder engagement across all phases of the standardization process. Nevertheless, the proposed HoDS methodology has demonstrated to be very robust towards the ongoing legal development of EN17015-2 in CEN TC 440, the mediation of requirements between domain and technical experts, and changes to the XKatalog's architecture and specification throughout the whole standardization process. As a result, the XKatalog mitigates data inconsistencies by migrating structured data from the pre-award phase to the post-award phase. Moreover, the standard facilitates the transformation of this data into downstream documents such as electronic orders and invoices by binding to the same syntax-family (UBL). Thus, administrative costs can be reduced through automatic validation and transformation capabilities. Consequently, HoDS has proven to be both feasible and successful in practice.

7. Discussion & Conclusion

We used Design Science Research (Hevner, 2007; Hevner et al., 2004) to harmonize existing methodological components towards a holistic data standardization methodology in public e-procurement (HoDS). To achieve this research objective, two research questions are discussed. In **RQ1**, suitable frameworks and methodologies for data standardization in public e-procurement are identified by analyzing rigorous and relevant foundations (Section 4). In the DSR's rigor cycle, a structured literature review is carried out in line with Carrera-Rivera et al. (2022) and Webster & Watson (2002). Synthesizing these findings with practical insights from workshops with data standardization and domain experts within the DSR's relevance cycle (Ørngreen & Levinsen, 2017) has enabled the identification of four methodological components: the EIF (European Commission, 2017), FISAD (Schmitz & Wimmer, 2023), the requirements engineering process by (Klier et al., 2017), and SeMoX (Kottmann et al., 2024; Schmitz et al., 2025). **RQ2** harmonizes these findings within the design cycle of DSR by incrementally developing and evaluating the proposed HoDS methodology with standardization experts from practice (Section 5). The **EIF** provides overarching interoperability across multiple layers and principles that all further methodological components adhere to. **FISAD** provides baseline and target architectural considerations to ensure rigorous and relevant grounding of the architectural environment in which the standard should operate, thereby preventing blindness towards required processes, applications and networks. A dynamic **requirements engineering** process is utilized alongside a central IT-Governance steering group to systematically translate these considerations into actionable and measurable specifications while being flexible towards changing demands at any point in the data standardization process. Based on the requirements, the data standard's specification is written in XML via **SeMoX**, whose tooling enables the automatic generation of human-readable specifications in PDF format. Hence, SeMoX enables domain experts to iteratively review and validate the semantic content of the data standard while simultaneously retaining full machine readability for further transformation and validation. Because the four methodological components are defined as modular artifacts with clear and close references to each other, the proposed methodology is highly agile and adaptive to changes at any point in the data standardization process. Instead of considering the data standardization process as a sequential waterfall, HoDS considers standardization as several parallel process instances that may influence each other at any given time. Consequently, as outlined in the case of the XKatalog as a German e-catalogue standard (Section 6), resulting data standards align with the overarching vision of interoperability, can be validated against their consolidated requirements, and adjusted to meet continually changing requirements over time.

In consideration of these findings, three limitations arise from the scope of this research. First, this paper has only addressed the barriers of data standardization in public e-procurement to a limited extent. Hence, it is recommended that future research develops a comprehensive overview of challenges for data standardization in the public sector. Second, this paper does not provide a systematic review of existing interoperability frameworks and methodologies regarding their strengths and weaknesses for data standardization in the public sector, which should be addressed in subsequent research. Third, the proposed methodology (HoDS) has only been validated in one use-case thus far. Therefore, future research should evaluate its feasibility in further data standardization scenarios, encompassing both public e-procurement and a broader range of e-government domains. One potential scenario could be the application of HoDS to standardize data exchange transactions between data trusts and data spaces in the context of Germany.

8. References

Aasa, A., Petersson, K., & Synek, D. (1988). Concrete syntax for data objects in functional languages. 1988 *ACM Conf. on LISP and Funct. Program.*, 96–105.

Ali, M., Papageorgiou, G., Aziz, A., Loukis, E., Charalabidis, Y., & Lopez Pellicer, F. J. (2024). Towards the Development of Interoperable Open Data Ecosystems: Harnessing the Technical, Semantic, Legal, and Organizational (TSLO) Interoperability Framework. *Proc Int Conf Dig Gov Res*, 909–919.

Amar, F., April, A., & Abran, A. (2024). Electronic health record and semantic issues using fast healthcare interoperability resources: systematic mapping review. *J. of Med. Internet Res.*, *26*, e45209.

Avbelj, M. (2011). Supremacy or Primacy of EU Law—(Why) Does it Matter? *Eur. Law J*, *17*(6), 744–763.

Babica, V., Sceulovs, D., & Rustenova, E. (2019). Digitalization of public procurement: Barriers for innovation. *WMSCI 2019 Proc.*, *3*, 7 – 12.

Bajaj, A. (2004). The effect of the number of concepts on the readability of schemas: an empirical study with data models. *Requirements Engineering*, *9*, 261–270.

Bobowski, S., & Gola, J. (2018). E-procurement in the European Union. *Asia-Pac. J. EU Studies, Materiał Powielony Złożony Do Druku*, *17*(1–2018), 23–35.

Bowers, S., & Ludäscher, B. (2004). An ontology-driven framework f data transformation in scientific workflows. *Int. WS on Data Integr in Life Sci*, 1–16.

Campmas, A., Iacob, N., & Simonelli, F. (2022). How can interoperability stimulate the use of digital public services? An analysis of national interoperability frameworks and e-Government in the European Union. *Data & Policy*, *4*, e19.

Carrera-Rivera, A., Ochoa, W., Larrinaga, F., & Lasa, G. (2022). How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX*, *9*, 101895.

Dang, D. D., & Pekkola, S. (2017). Systematic literature review on enterprise architecture in the public sector. *Electron. J. of E-Government*, *15*(2), 57–154.

DeNardis, L. (2010). E-Governance Policies for Interoperability and Open Standards. *Policy & Internet*, *2*(3), 129–164.

Dzierzawa, F., Petrik, D., Stuber, K., Merz, S., Jaensch, L., & Herzwurm, G. (2025). Bad Data Quality Eats Ecosystem for Breakfast. Proc. of *58th HICSS*.

Egorova, M., Andreeva, L., Andreev, V., Tsindeliani, I., & Kikavets, V. (2021). Digitalization of public procurement in the Russian Federation: Case study. *NISPAcee J. Public Admin. Policy*, *14*(1), 87–106.

Eichelberger, H., & Schmid, K. (2009). Guidelines on the aesthetic quality of UML class diagrams. *Inf. and Software Technology*, *51*(12), 1686–1698.

European Commission. (2017). New European Interoperability Framework: Promoting seamless services and data flows for European public administrations. In *Publ. Office of the Eur. Union*.

European Commission. (2024). *European Interoperability Reference Architecture (EIRA)*. https://interoperable-europe.ec.europa.eu/collection/european-interoperability-reference-architecture-eira/solution/eira/release/610

European Parliament. (2014). Directive 2014/24/EU of 26 February 2014 on public procurement and repealing Directive 2004/18/EC. *J. of the European Union*.

Ferreira, I., & Amaral, L. A. (2016). Public e-procurement: Advantages, limitations and technological "pitfalls." *ACM Int. Conf. Proc. Series*, *01-03-March-2016*, 9 – 12. https://doi.org/10.1145/2910019.2910089

Folmer, E., Oude Luttighuis, P., & van Hillegersberg, J. (2011). Do semantic standards lack quality? A survey among 34 semantic standards. *EM*, *21*, 99–111.

Gal, M. S., & Rubinfeld, D. L. (2019). Data standardization. *NYUL Rev.*, *94*, 737.

Halpin, T., & Bloesch, A. (1999). Data modeling in UML and ORM: a comparison. *J of DB Mg*, *10*(4), 4–13.
Hartati, S., Sugiharto, I., Fakhri, J., Siswadi, Wekke, I. S., Azhari, & Roslina. (2020). Implementation of government goods and services procurement using the e-purchasing method in the tegal religious court: Mediating role of supply chain in Indonesia. *Int. J. of Supply Chain Management*, *9*(2), 843 – 852.
Henning, F. (2018). A theoretical framework on the determinants of organisational adoption of interoperability standards in Government Information Networks. *GIQ*, *35*(4), S61–S67.
Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scand. J. of Inf. Sys*, *19*(2), 87–92.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.
International Organization for Standardization. (1996). Standardization and related activities—General vocabulary. *Geneva: Int Electrotech Commission*.
International Organization for Standardization. (2022). *Software, systems and enterprise - Architecture description (ISO/IEC/IEEE 42010:2022)*. https://www.iso.org/standard/74393.html
Jämsä-Jounela, S.-L. (2007). Future trends in process automation. *Ann. Rev. in Control*, *31*(2), 211–220.
Janner, T., Lampathaki, F., Hoyer, V., Mouzakitis, S., Charalabidis, Y., & Schroth, C. (2008). A core component-based modelling approach for achieving e-business semantics interoperability. *J. of Theor. Appl. Electronic Commerce Research*, *3*(3), 1–16.
Klier, J., Klier, M., & Muschter, S. (2017). How to manage IS requirements in complex public sector structures: toward an action design research approach. *Requirements Engineering*, *22*, 419–432.
Kottmann, R., Pauken, C., & Schmitz, A. (2024). Simple Semantic Data Modeling in XML (SeMoX). *XML Prague*, 231–249.
Ma, Z. M., & Yan, L. (2007). Fuzzy XML data modeling with the UML and relational data models. *Data & Knowledge Engineering*, *63*(3), 972–996.
Margariti, V., Stamati, T., Anagnostopoulos, D., Nikolaidou, M., & Papastilianou, A. (2022). A holistic model for assessing organizational interoperability in public administration. *GIQ*, *39*(3), 101712.
Mavidis, A., & Folinas, D. (2022). From Public E-Procurement 3.0 to E-Procurement 4.0; A Critical Literature Review. *Sustainability*, *14*(18).
Moe, C. E., Newman, M., & Sein, M. K. (2017). The public procurement of information systems: dialectics in requirements specification. *E.J.I.S.*, *26*(2), 143–163.
OECD. (2023). *Size of Public Procurement - Government at a glance indicators*. https://data-viewer.oecd.org/?chartId=2a6845b9-40e3-40e0-ab8f-0e91aa7954d1
Olivares, M., Saban, D., Weintraub, G. Y., Lara, E., Zanocco, P., & Moreno, P. (2024). Saving Millions in Government Procurement Through Data Science and Market Design. *Informs J. on applied Analytics*. https://doi.org/10.1287/inte.2023.0002
OpenPEPPOL AISBL. (2023). *PEPPOL Business Interoperability Specification Pre-award guide: Notification & Open Procedure* . https://docs.peppol.eu/pracc/files/Peppol-BIS-pre-award-guide_Notification-and-Open-Procedure-v1.0.pdf
Ørngreen, R., & Levinsen, K. T. (2017). Workshops as a research methodology. *Electr. J. E-Learn.*, *15*(1), 70.
Paľová, D., & Vejačka, M. (2022). Identifying the challenges in e-procurement standardization. *IDIMT-2022*.
Pandey, D., Suman, U., & Ramani, A. K. (2010). An effective requirement engineering process model for software development and requirements management. *2010 Int. Conf. on Adv. in Rec. Technol. in Commun. and Comput.*, 287–291.
Pardo, T. A., Nam, T., & Burke, G. B. (2012). E-government interoperability: Interaction of policy, management, and technology dimensions. *Social Science Computer Review*, *30*(1), 7–23.
Peckham, J., & Maryanski, F. (1988). Semantic data models. *ACM Comp. Surv. (CSUR)*, *20*(3), 153–189.
Pekolj, N., Hodoscek, K., Valjavec, L., & Ferk, P. (2019). Digital Transformation of Public Procurement as an Opportunity for the Economy. *LEXONOMICA*, *11*(1), 15–42.
Prałat, E. (2019). Public e-procurement tools in European Union. *Nierówności Społeczne a Wzrost Gospodarczy*, *58*, 188–197.
Prier, E., & McCue, C. P. (2008). The implications of a muddled Definition of public procurement. *Journal of Public Procurement*, *3*(3), 64–82.
Schmitz, A., Pauken, C., Kottmann, R., & Wimmer, M. A. (2025). Architecture Interoperability through Semantic Modelling: Domain-driven data specification in public procurement. *ECIS 2025 Proc*.
Schmitz, A., Siapera, M., Prentza, A., & Wimmer, M. A. (2023). Harmonization in eProcurement: Design of a Holistic Solution Model for Pre-award Procedures. *Int. Conf. on E-Gov.*, 18–33.
Schmitz, A., & Wimmer, M. A. (2023). Framework for interoperable service architecture development. *GIQ*, *40*(4). https://doi.org/10.1016/j.giq.2023.101869
Siapera, M., Schmitz, A., Wimmer, M. A., & Prentza, A. (2024). Closing the gap: Leveraging data for seamless integration between pre-award and post-award in public procurement. *Proc. of 57th HICSS.*
Tatsis, V., Mena, C., Van Wassenhove, L. N., & Whicker, L. (2006). E-procurement in the Greek food and drink industry: drivers and impediments. *Journal of Purchasing and Supply Management*, *12*(2), 63–74.
Uyarra, E., & Flanagan, K. (2010). Understanding the innovation impacts of public procurement. *European Planning Studies*, *18*(1), 123–143.
Vaidya, K., & Campbell, J. (2016). Multidisciplinary approach to defining public e-procurement and evaluating its impact on procurement efficiency. *Inform Syst frontiers*, *18*(2), 333–348.
Varney, M. (2011). E-Procurement—current law and future challenges. *ERA Forum*, *12*(2), 185–204.
Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii.
Wood, J. (1985). What's in a link? Readings in Knowledge Representation.