July 2025

SMEs aiming to operate as Data Trusts: A Reference Architecture

Florian Rieder
University of Koblenz, rieder@uni-koblenz.de

Timon T. Aldenhoff
University of Koblenz, timonaldenhoff@uni-koblenz.de

David Acev
University of Koblenz, dacev@uni-koblenz.de

Sankalp Biyani
sbiyani@uni-koblenz.de

Dennis M. Riehle
University of Koblenz, riehle@uni-koblenz.de

See next page for additional authors

Presenter Information
Florian Rieder, Timon T. Aldenhoff, David Acev, Sankalp Biyani, Dennis M. Riehle, and Maria A. Wimmer

SMEs aiming to operate as Data Trusts: A Reference Architecture

Completed Research Paper

Florian Rieder
University of Koblenz
Universitätsstraße 1, 56070 Koblenz
rieder@uni-koblenz.de

Timon T. Aldenhoff
University of Koblenz
Universitätsstraße 1, 56070 Koblenz
timonaldenhoff@uni-koblenz.de

David Acev
University of Koblenz
Universitätsstraße 1, 56070 Koblenz
dacev@uni-koblenz.de

Sankalp Biyani
University of Koblenz
Universitätsstraße 1, 56070 Koblenz
sbiyani@uni-koblenz.de

Dennis M. Riehle
University of Koblenz
Universitätsstraße 1, 56070 Koblenz
riehle@uni-koblenz.de

Maria A. Wimmer
University of Koblenz
Universitätsstraße 1, 56070 Koblenz
wimmer@uni-koblenz.de

Abstract

As data has become a critical asset in today's digital economies, the sharing and reuse of data has received high attention from researchers. New concepts of data intermediaries have evolved, addressing the strategic goal of greater digital sovereignty while ensuring compliance with European legal frameworks. This paper introduces an initial concept for a reference architecture for Data Trusts, which act as intermediaries facilitating secure and trustworthy data stewardship and sharing within relevant ecosystems. Through a structured analysis of existing literature, we identify and evaluate four distinct Data Trust architectures using qualitative and quantitative criteria. Our analysis reveals notable diversity in application domains, core components, and candidate technologies. Building on this, we derive a foundational and overarching schema for developing a holistic, integrated reference architecture. As our research particularly targets SMEs aspiring to establish themselves as Data Trusts, we spot specific challenges and requirements they face in fulfilling this role.

Keywords: Data Trusts, Digital Ecosystems, Reference Architecture, Interoperability

Introduction

Sharing data while ensuring data sovereignty has become increasingly important for fostering innovation and economic growth. Along this, data intermediaries have evolved that enable secure access to valuable data while respecting the rights and control of data owners (European Commission, 2020). Data intermediaries have the potential to drive innovation, improve decision-making, and solve complex challenges by providing access to diverse and valuable insights across industries, while the evolving role of data – ranging from operational excellence and product offerings to fostering business innovation and supporting long-term economic sustainability – requires businesses to rethink how they manage and share

data with internal and external partners (Otto, 2022). However, barriers such as lack of trust, legal restrictions, security concerns, and insufficient technological capabilities prevent private sector organizations – particularly small and medium-sized enterprises (SMEs) – from making their data accessible (Fassnacht et al., 2023; Jussen et al., 2024). This results in a significant amount of potentially valuable data remaining untapped, thereby limiting opportunities for innovation and collaboration.

The growing importance of data sovereignty, alongside the need to leverage the value of data, has brought to light various challenges surrounding data governance, privacy and access. As organizations face increasing pressure to comply with privacy regulations, ensure data sovereignty, and address concerns around security and trust, *data intermediaries* are crucial facilitators in addressing these challenges. They support organizations in managing and safeguarding data, facilitating secure data exchange, and ensuring compliance with legal frameworks such as the Data Governance Act (DGA) (EU Regulation 2022/868). By offering neutral, secure, and compliant platforms, data intermediaries reduce barriers to data sharing and reuse, enabling organizations to unlock the potential value of data while respecting legal and ethical boundaries.

Various types of data intermediaries exist, differing in their objectives and architectures (Janssen & Singh, 2022). Among these, *Data Trusts* represent a specific form of data intermediary, characterized by their emphasis on transparent governance and fiduciary responsibilities to safeguard the interests of data providers and data owners. Their primary goal is to foster trust and fairness within data ecosystems, facilitate both voluntary and mandated data sharing, and safeguard the interests of key stakeholders, such as data owners and providers (Feth & Rauch, 2024; Paprica et al., 2023; Stachon et al., 2023).

This study investigates the architectural concept of Data Trusts by analyzing and comparing various concepts presented in the literature. The research aims to identify the essential components of a Data Trust architecture and to propose a foundational Reference Architecture (RA). Furthermore, it highlights the challenges and requirements for SMEs aiming to operate as Data Trusts. The research is guided by the following research objective (RO) and research questions (RQs):

RO: To identify the key components of a reference architecture for Data Trusts.

RQ1: What (reference) architectures for Data Trusts exist in the literature, and how do they differ in terms of their core components and level of detail?

RQ2: What key elements should be included in a reference architecture for Data Trusts?

RQ3: What challenges and requirements arise for SMEs that aspire to operate as Data Trusts?

The paper is structured as follows: The *Theoretical Background* section provides an overview of the foundational concepts, including data intermediaries, Data Trusts and reference architectures, which form the theoretical basis for the study. This is followed by the *Methodological Foundations* section, which outlines the approach employed to address the RQs. The *Findings from the Literature Analysis* section presents the outcomes of the study, organized into three subsections. The first subsection on *Literature Review on Data Trust Reference Architectures* addresses **RQ1** by comparing various architectures for Data Trusts identified in the literature. Building on this analysis, the second subsection on *Key Components of a Data Trust Reference Architecture* tackles **RQ2** by extracting key components from the reviewed architectures and proposing an initial high-level model for a reference architecture for Data Trusts. The third subsection responds to **RQ3** by examining the challenges and requirements faced by organizations – particularly SMEs – aspiring to operate as Data Trusts. The *Discussion and Conclusion* section highlights the key findings, and the paper concludes with *Limitations and Outlook*, which reflect on the study's limitations and suggested directions for future research.

Theoretical Background

This section briefly outlines the foundational concepts of data intermediaries, focusing on their role in facilitating secure and efficient data exchange. It introduces various types of intermediaries and provides a deeper understanding of their function and significance. Additionally, the section outlines the concept of Data Trusts, emphasizing their fiduciary responsibility, and discusses the importance of reference architectures in ensuring interoperability and scalability within Data Trust ecosystems.

Data Intermediaries

Data intermediaries are neutral entities, organizations or systems designed to mediate and facilitate connections between multiple parties, such as data providers and consumers, while addressing barriers related to legal compliance, trust, and technical challenges (Janssen & Singh, 2022; Schweihoff et al., 2024). They facilitate secure and efficient data sharing by offering services across several categories, including technical infrastructure, data management, transaction facilitation, governance and compliance, and ensuring data sovereignty (Schweihoff et al., 2024). Their role, scope, and implementation vary depending on the context, ranging from centralized to decentralized or hybrid models (Janssen & Singh, 2022; Schinke & Roßmann, 2024). Distinct types of data intermediaries have emerged, each tailored to specific needs and contexts (Janssen & Singh, 2022):

(1) *Data marketplaces* serve as platforms for trading and monetizing data, connecting data providers with consumers while reducing transaction costs.
(2) *Data commons* involve pooling data by a group or community to support shared goals or societal benefits through collective ownership and governance.
(3) *Data Trusts* act as fiduciaries, securely and ethically managing data on behalf of data owners (those controlling the data) or data providers (those supplying it), with a focus on aligning data use with the interests of data subjects (the individuals to whom the data pertains).
(4) *Data cooperatives* are member-controlled organizations that enable participants to pool and share data resources democratically.
(5) *Data collaboratives* bring together stakeholders from different sectors – such as private companies, research institutions, and public entities – to share data for solving societal challenges or advancing public goods.
(6) *Personal Information Management Systems* (PIMS) empower individuals to manage consent and control how their personal data is shared and utilized, leveraging technology to enforce data rights.

Data intermediaries must comply with legal requirements, including the DGA, which defines key requirements for how they are permitted to offer their services. These include ensuring neutrality, interoperability, offering transparent and non-discriminatory services, and implementing robust data security measures (Carovano & Finck, 2023; EU Regulation 2022/868; Richter, 2023).

Data Trusts

Data Trusts represent a specialized form of data intermediary. They are structured to steward data on behalf of data providers, ensuring that the data is managed securely and in alignment with predefined governance frameworks and fiduciary duties (Janssen & Singh, 2022). The Open Data Institute (ODI) defines a Data Trust as a "*legal structure that provides independent stewardship of data*" (Hardinges, 2018), and the fiduciary duties associated with it (Hardinges, 2020). Moreover, the ODI refers to data stewardship as "*collecting, maintaining and sharing*" data along the management of access rights and usage conditions (Hardinges, 2020). However, a unified and broadly accepted definition of Data Trusts has not yet been established in the literature, owing to various interpretations and applications of the concept (Stalla-Bourdillon et al., 2020).

Building upon the ODI's definition, a Data Trust can be regarded as an organization or entity that acts as a neutral and independent (trusted) third party, focusing on fiduciary stewardship of data (without any commercial interest in the data itself). Furthermore, it may specialize in specific industries, types of data, security and transparency standards, or value-added services (Stachon et al., 2023). This understanding has been widely adopted in scholarly discourse, serving as a foundation upon which different definitions of Data Trusts are frequently built (Gomer & Simperl, 2020; O'Hara, 2019; Stalla-Bourdillon et al., 2020).

Data Trusts are also investigated as components or nodes within a broader distributed network of interconnected yet independent actors (Otto et al., 2022). The International Data Spaces Association (IDSA) is actively involved in establishing secure and trustworthy data spaces with a decentralized, federated structure. In this context, Data Trusts may operate as integral parts of a larger data ecosystem grounded in shared standards and collaboration.

This paper focuses on standalone and independent operating entities or organizations with their own business models and trust mechanisms. However, the concept of the International Data Spaces (IDS) and

its Reference Architecture Model (IDS-RAM) (International Data Spaces Association, 2023) encompasses some important considerations for possible architectures of Data Trusts and other data intermediaries. Consequently, an architecture for Data Trusts can be built on the concepts and components proposed in the IDS-RAM (Schinke et al., 2023).

In summary, Data Trusts differ from other data intermediaries, such as data marketplaces, data commons, and data cooperatives, primarily due to their fiduciary obligations. Unlike data marketplaces that prioritize economic incentives, Data Trusts are legally bound to act in the best interests of data providers and owners, ensuring that data use aligns with ethical and regulatory principles (Hardinges, 2020).

Reference Architecture

The development and utilization of reference architectures has a long-standing tradition within the Information Systems (IS) discipline. Over the past three decades, numerous definitions for RAs have been proposed, reflecting an evolving understanding of their nature and purpose (Aldenhoff et al., 2024; Angelov et al., 2012; Bass et al., 2003; Garcés et al., 2021; Nakagawa et al., 2014; The Open Group, 2022). While some definitions focus on the mapping of reference models to software elements and data flows, others adopt a broader perspective, encompassing business and technology aspects (Angelov et al., 2012; The Open Group, 2019). Recognizing this diversity, our research seeks to adopt a more abstract conceptualization of RAs, one that extends beyond software to encompass business and technology considerations. Drawing from this understanding, a RA is characterized as a generic architecture, which provides guidelines for developing more specific architectures and implementing solutions (The Open Group, 2019). It encompasses proven building blocks and principles designed to address common requirements, offering an abstract, reusable foundation for developing consistent, interoperable, and high-quality solution architectures (Reidt et al., 2018; The Open Group, 2022). Within the data economy, RAs can play a crucial role in establishing common frameworks for data management, exchange, and utilization. By providing a shared language and conceptual foundation, RAs facilitate interoperability, reduce complexity, and accelerate the development of data-driven solutions. Moreover, RAs can serve as blueprints for designing and implementing architectures that align with organizational goals and industry best practices (Angelov et al., 2012; The Open Group, 2022).

Methodological Foundations

For the study, we applied the methodology of a structured, topic-centric literature review (Rowe, 2014; vom Brocke et al., 2015; Webster & Watson, 2002) of peer-reviewed scientific literature to achieve the research objectives. This approach enabled a systematic and comprehensive examination of existing scientific literature, ensuring that relevant perspectives and findings from a wide range of sources were considered. Figure 1 illustrates the slightly modified PRISMA scheme for our search process, as proposed by Page et al. (2021). The search included the following keyword combination: ("data trust" OR "data trustee" OR "data intermediary" OR "data ecosystem" OR "data platform") AND ("reference architecture" OR "architecture requirements"). While "reference architecture" and "architecture requirements" are closely related, the latter term was included to capture studies that discuss key elements and structural needs for architectures beyond formalized RAs. This methodology allows us to identify not only established studies but also lesser-known works that could offer valuable contributions to the research questions, particularly those that reflect newer, innovative approaches in the field.

We carried out an enhanced initial search using the databases ACM Digital Library, IEEE, ProQuest, Scopus, Springer, and Web of Science, which hold an abundance of peer-reviewed scientific papers. Only English language sources were included. The initial search yielded a total of 416 papers. Since the search embodied several databases, duplicates of the papers were subsequently eliminated, which reduced the number of papers to 328. The screening step involved a title search for the remaining papers. We selected the papers that were applicable to our research questions, resulting in a total of 118 papers. The inclusion criteria for the literature review focused on selecting peer-reviewed journal articles and conference papers that explicitly discuss Data Trust architectures. Additionally, studies addressing security, governance, and interoperability in data exchange were prioritized to ensure comprehensive analysis. Papers that were unrelated to data intermediaries or trust-based data sharing were excluded from consideration. Articles that focused solely on data marketplaces without incorporating governance aspects were also dropped. Furthermore, non-peer-reviewed sources such as blog posts and white papers were not included to

maintain the academic rigor of the study. A further 38 papers were discarded as we did not have access to them. This decreased the total number of relevant papers to 80. The fifth step involved a thorough examination of the abstract and a rigorous review of all the publications to see if the findings were applicable to our area of interest. This scrutiny identified 46 relevant papers for our study. The sixth and final step applied backward and forward search to find more relevant papers that were not identified with our initial search (vom Brocke et al., 2009). We identified 35 additional papers, increasing the total number of papers for our study to 81. The final selection of four (reference) architectures was based on their recurrence in peer-reviewed literature, diversity of design and implementation approaches (ranging from domain-specific to general-purpose), and alignment with current industry and policy trends, including compatibility with frameworks like IDS-RAM and emphasis on user-centric governance. Since Data Trusts represent a fairly new area of research, the number of comprehensive architectural proposals found in peer-reviewed sources is still limited. As such, the selected architecture models represent some of the few concrete and sufficiently detailed models currently available for analysis.

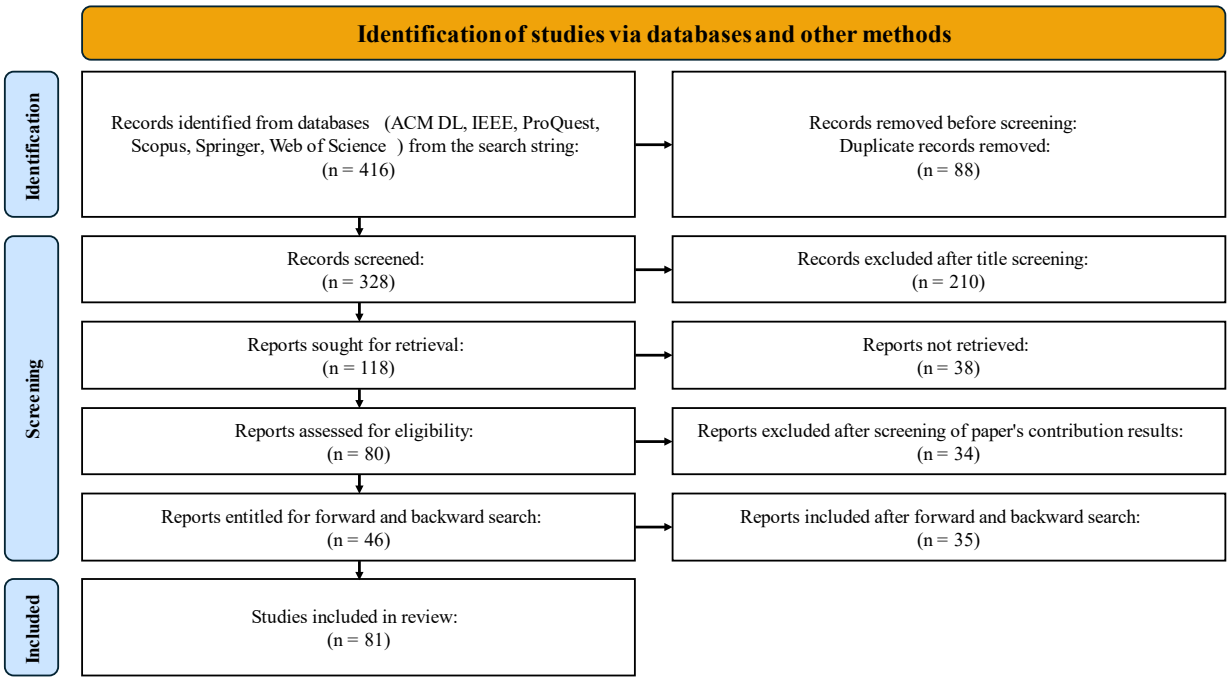

Figure 1. Systematic Literature Review based on Page et al. (2021).

Findings from the Literature Analysis

In this section, we present the results of our analysis on Data Trust reference architectures. Our examination focused on four key papers that propose different architectural designs for Data Trusts, a critical aspect of ensuring trustworthy and sovereign data stewardship. In addition, we analyzed the key components essential to building effective Data Trust architectures and explored the challenges and requirements, particularly for SMEs in adopting such frameworks. The following subsections provide a detailed presentation of the findings and insights derived from this analysis.

Literature Review on Data Trust Reference Architectures

Along the analysis of the 81 papers, we identified four architecture models for Data Trusts that represent the attributes we strive to convey. While our primary focus in this analysis was on the papers that explicitly propose architectural designs for Data Trusts, the remaining papers played a crucial role in shaping our understanding of the foundational concepts, essential components, and the broader context of Data Trusts. The four architecture models we analyzed and compared in detail present different approaches for the

design and implementation of Data Trust architectures to fulfill specific requirements in various fields of application:

(1) Alsaad et al. (2019) suggest an Institutional Repository— an archive designed to collect, preserve, and share digital copies of a research institution's intellectual output – as a candidate technology for a Data Trust infrastructure. The key idea is to build on the capabilities of an Institutional Repository to facilitate trustworthy data sharing among multiple parties in data-intensive research engagements (Alsaad et al., 2019).
(2) Lomotey et al. (2022) introduce Data Trusts as a Service platform, a cloud-based multi-data sharing environment. The Data Trusts as a Service architecture aims to connect and facilitate interactions between data subjects, data collectors, and data consumers, providing data subjects with greater control over the use of their data (Lomotey et al., 2022).
(3) O'Hara (2019) proposes an architecture for a Data Trust Portal that serves as an ethical, transparent, and accountable platform for trustworthy data sharing. Rather than storing data itself, the DTP acts as a platform, where data controllers can list their datasets and establish terms for data access and usage. The author also proposes the adoption of the Web Observatory as a candidate experimental technology for implementation (O'Hara, 2019).
(4) Schinke et al. (2023) introduce an architecture for a Data Trust in the forestry sector to enable trustful and sovereign data sharing among various stakeholders. This architecture addresses specific challenges of the forestry sector, such as unreliable internet connectivity, a highly heterogeneous landscape of systems and data formats, and the need to protect sensitive business and personal data while facilitating data sharing and collaboration (Schinke et al., 2023).

Further research explores various design options to construct a general solution space (Lauf et al., 2023; Stachon et al., 2023) or introduce architectures for specific aspects of Data Trusts. For instance, Ayappane et al. (2024) introduce a comprehensive reference architecture for a policy-based consent management service. However, while detailed examinations of individual components can provide valuable insights, they fall outside the scope of this comparative analysis, which aims to synthesize a complete picture of Data Trust architectures. This approach ensures that the analysis remains comprehensive and holistic, capturing a broad spectrum of elements that are crucial for the effective implementation and operation of Data Trusts. It also helps in identifying common patterns, best practices, and potential gaps within the broader landscape of Data Trust architectures.

The four architecture models identified and outlined above were compared based on qualitative attributes as presented in Table 1. The attributes used for the comparison were informed by the taxonomy for designing Data Trusts presented in Stachon (2023). This taxonomy served as a guiding framework rather than being directly replicated. The attributes considered include:

(1) *Model*: What is the underlying model and implementation strategy of the architecture?
(2) *Field of application*: For which specific domains is the architecture developed (e.g., healthcare, research, forestry, or finance)?
(3) *Data storage*: Where is the managed data stored?
(4) *Data type*: What types and structures of data can be managed by the Data Trust?
(5) *Core application components*: What are the key application components of the proposed architecture?
(6) *Special functionalities & capabilities*: What central functionalities and capabilities does the proposed architecture offer (e.g. for the data governance, the findability or the access to data)?

Among the four analyzed architectures, the following notable distinctions are worth mentioning: **1)** The architectures vary in their levels of abstraction: e.g. while (O'Hara, 2019) introduces a conceptual high-level architecture and required capabilities, the architecture developed by Lomotey et al. (2022) is more detailed and implementation-oriented. **2)** The architectures proposed by O'Hara (2019) and by Lomotey et al. (2022) are widely applicable across various domains. In contrast, Schinke et al. (2023) and Alsaad et al. (2019) both introduce architectures intended for application within specific domains. **3)** The manner of data storage varies from merely decentralized data storage in O'Hara (2019) to flexible storage options enabling both centralized and decentralized data storing simultaneously (Schinke et al., 2023), to a hybrid combination of centralized storage and decentralized blockchain (Lomotey et al., 2022). **4)** Finally, the architectures differ in their components and candidate technologies aimed at fostering trustful and sovereign data stewardship. Alsaad et al. (2019) propose the Institutional Repository as candidate

technology for the implementation of Data Trusts. Schinke et al. (2023) build their architecture on the concepts and components provided by the IDS-RAM (International Data Spaces Association, 2023) to enable seamless integration and interoperability with data spaces. Lomotey et al. (2022) utilize cloud and blockchain technology for scalability and accessibility while also enforcing data usage policies, tracking provenance, and maintaining auditable records of all data sharing activities. O'Hara (2019) suggests a generic architecture that is not built on any specific existing technology but discusses the Web Observatory as a meaningful candidate experimental technology for Data Trust implementation.

Table 1. Comparison of four Data Trust architectures based on qualitative attributes.

Literature / Attribute	Alsaad et al. (2019)	Lomotey et al. (2022)	O'Hara (2019)	Schinke et al. (2023)
Model	Data Trust based on Institutional Repository	Cloud-based digital platform: Data Trusts as a Service	Data Trust Portal	IDS-compliant Data Trust
Field of application	Multi-partner research engagements	Data sharing among multi-parties, generic and applicable to various domains	General, not specific to any domain	Specifically designed for the forestry sector
Data Storage	1) Within the Data Trust repository itself, 2) in networked storage within the data-owning organization, or 3) in external cloud services	Hybrid storage approach: central data storing and processing, decentral blockchain to record data provenance and access logs	Decentral: the Data Trust Portal does not store data itself, but stores and manages metadata and provenance information	Flexible approach: decentral or central (optional)
Data Type	Innovation- and research-based data with support for a wide variety of data formats	Various kinds of personal, organizational, and research data	Very generic, no limitations to any type of data	Various types of human- and machine-generated data in the forestry sector, such as sensor, inventory, harvester production, and personal data
Core Application Components	Storage Management, User Management, Computation	Middleware Controller, Data Lake, Blockchain & Smart Contracts, Visualization Tools	Data Trust Portal, Data Catalogue & Metadata Store, Visualization Tools	Interfaces (Data Trust Connector, Data Service Interface, User Interface), Clearing House, Data Storage (optional), Apps (optional)
Special Functionalities & Capabilities	Quality Assurance, Resource Discovery, Community-evaluated Apps with bespoke functionality	Usage Policy Specification & Enforcement (Smart Contracts), Event Monitoring and Visualization, Trust and Reputation Management	Ethical Code & Membership Model	Usage Control & Policy Enforcement, User-friendly Interfaces, IDS-RAM compatibility, offline functionality

To gain a deeper understanding of both the scope and the level of detail, a second, more quantitative comparison was conducted. This involved evaluating the completeness and granularity of the architectures by assessing the extent to which they incorporate relevant components outlined in The Open Group Architecture Framework (TOGAF) (The Open Group, 2022). We selected TOGAF as a reference due to its structured, modular approach to enterprise architecture, which defines key artifacts produced during the architecture development method (ADM). TOGAF has been widely adopted across industries as a methodology for designing scalable and interoperable architectures. While other architecture frameworks, such as Zachman or ArchiMate, provide structured classification and modeling techniques, TOGAF offers a more comprehensive development lifecycle, ensuring that architectural components align with strategic business needs (The Open Group, 2022). Given that Data Trusts operate at the intersection of legal, technical, and organizational domains, TOGAF's emphasis on holistic governance and stakeholder alignment makes it a suitable choice. Table 2 presents the artifacts and their coverage in the analyzed Data Trust architectures. An "x" indicates that the artifact is proposed and elaborated in more detail, while a "(x)" implies that it is proposed but not elaborated in detail.

Table 2. Comparison of Data Trust architectural artifacts along TOGAF's domain architectures.

	Architectural Artifact	Alsaad et al. (2019)	Lomotey et al. (2022)	O'Hara (2019)	Schinke et al. (2023)
Business	Business Actors & Roles	(x)	x	x	x
	Business Goals & Objectives	x	x	x	x
	Business Functions & Services	x	x	x	x
	Business Use-Cases	(x)	x	x	x
	Business Processes		(x)		(x)
	Business Requirements	x	x	x	x
Information Systems	Conceptual Data Model				
	Logical Data Model				
	Data Exchange Model	(x)	(x)		(x)
	Data Catalogue	(x)	(x)	(x)	(x)
	Application Use-Cases	(x)	x		
	Application Components & Services	x	x	(x)	(x)
	Interfaces		x		(x)
Technology	Technology Standards		x		(x)
	Technology Components & Services	(x)	x	(x)	(x)

The analysis reveals that the level of detail and specific technical components differ among the architecture models in terms of addressing and developing components at various architectural levels. The architecture proposed by Lomotey et al. (2022) stands out as the most comprehensive and implementation-oriented. It is also noteworthy that certain artifacts are missing from all the analyzed architectures. For example, the data structures of the supported data types and formats are not depicted through a conceptual or logical data model. Additionally, detailed process models, which could provide insights into internal and external data exchange processes, are not included. However, Lomotey et al. (2022) and Schinke et al. (2023) offer

more extensive textual descriptions. Interfaces are only addressed in these two publications, with Lomotey et al. (2022) providing more detailed specifications.

The absence of conceptual and logical data models in all analyzed architectures raises concerns about interoperability and semantic alignment, which are critical in multi-stakeholder environments. Similarly, the lack of explicit process flows limits the understanding of operational governance, data handling practices, compliance planning, and system integration. This also makes it more difficult for SMEs to plan implementations, assign operational roles, and develop standardized procedures.

Key Components of a Data Trust Reference Architecture

Drawing upon the insights gained from the comparative analysis of Data Trust architectures in the previous subsection, we extracted the basic components for our Data Trust reference architecture by synthesizing findings from the four papers. The process involved identifying key components explicitly outlined in the four papers and then validating and refining these components by cross-referencing them with the foundational insights and concepts derived from the remaining 77 papers. This approach ensured that the components were both grounded in existing architecture models and broadly representative of wider literature. Along this, we aimed to ensure that the reference architecture aligns with the technical, semantic, organizational, and legal requirements for fostering trustworthy, ethical, and sovereign data stewardship (Alsaad et al., 2019; Lomotey et al., 2022; O'Hara, 2019; Schinke et al., 2023). Special components that offer additional bespoke functionalities or are related to implementation have not been considered. Table 3 shows the abstract components, which are included in at least three or more of the analyzed architectures. Again, "x" indicates that the artifact is proposed and elaborated in more detail, while "(x)" implies that it is proposed but not elaborated in detail. Based on the identified components, we developed an initial abstract model of a reference architecture for Data Trusts with different vertical and horizontal layers. This layered approach, shown in Figure 2, provides a structured framework that facilitates the design, implementation and management of Data Trust architectures for SMEs. The vertical and horizontal layers are not mutually exclusive but work together to create a comprehensive architecture. The individual layers are explained below and linked to the architecture components to show where these components can be found in a reference architecture. The numbers in parentheses refer to the corresponding architecture components listed in Table 3.

Table 3. Overview of key components for Data Trusts present in the compared architectures.

Architecture Components	Alsaad et al. (2019)	Lomotey et al. (2022)	O'Hara (2019)	Schinke et al. (2023)
1. User Management	x	(x)	(x)	x
2. Consent Management		x	x	(x)
3. Compliance Reporting	x	x	x	x
4. Identity & Access Management	x	x	x	x
5. Interfaces	x	x		x
6. Data Visualization	x	x	x	
7. Data Processing	x	x	x	x
8. Data Catalogue, including Metadata Management	x	x	x	x
9. Data Storage	x	x	x	x
10. Monitoring & Logging	x	x	x	x

The vertical layers encapsulate the core functions and principles that play a role in all horizontal layers. The *Protection Layer* (cf. Table 3, component 4) focuses on the security and privacy of data. It includes mechanisms for encryption, anonymization, pseudonymization, and usage control. These measures ensure that data remains confidential, integrity is maintained, and access is restricted to authorized parties. The *Analysis Layer* (components 3, 6, and 10 in Table 3) enables data-driven insights important to each horizontal layer through visualization, real-time monitoring, reporting, and logging. These tools enable organizations to understand data usage patterns, detect anomalies, and ensure regulatory compliance. The *Trust Layer* (component 3) embodies the principles of transparency, ethics, and certification. It includes mechanisms that promote trust between data providers and consumers. These mechanisms include clear communication of data usage policies, adherence to ethical guidelines, and independent verification of data practices.

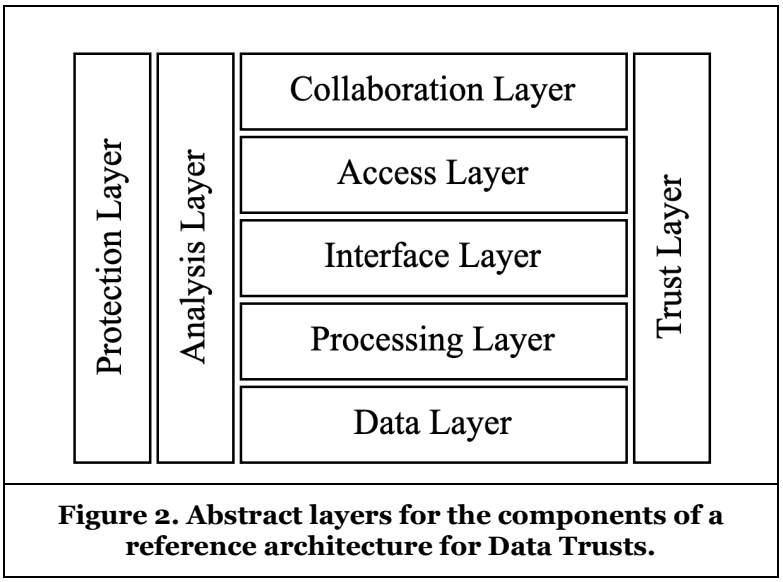

Figure 2. Abstract layers for the components of a reference architecture for Data Trusts.

The horizontal layers are enclosed by the vertical layers to show that the vertical layers influence all horizontal layers. The *Collaboration Layer* (component 2) facilitates data sharing, acquisition, and consent management. It may include platforms, marketplaces, or connectors that enable organizations to collaborate on data projects and exchange data securely. The *Access Layer* (components 1 and 4) manages user authentication and authorization. It encompasses components such as user management, and identity and access management, ensuring that only authorized users can access and interact with the data. The *Interface Layer* (component 5) provides interfaces for users and machines (UI and API) for accessing and interacting with data. This layer ensures interoperability between different components and systems, enabling seamless data exchange and integration. The *Processing Layer* (component 7) handles data transformation and enrichment. It includes functions for data aggregation, cleansing, validation, and contextualization, ensuring that the data is accurate, reliable, and ready for analysis. Finally, the *Data Layer* (components 8 and 9) is the foundation of the architecture, responsible for data storage, metadata management, and the data catalog.

Challenges and Requirements for Small and Medium-Sized Enterprises

Small and medium-sized enterprises (SMEs) face a variety of challenges in attempting to establish themselves as Data Trusts. These barriers span financial, technological, organizational, legal, and trust-related dimensions, each of which affects their ability to manage data securely, transparently, and in compliance with regulatory frameworks.

Financial Barriers: Establishing the necessary infrastructure for a Data Trust requires substantial upfront investment as well as long-term operational funding. Unlike larger firms, SMEs often lack access to external funding sources such as public grants or private capital, making it difficult to sustain data trust operations over time (Radosevic et al., 2023). The Open Data Institute (2020) argues that securing

adequate financial resources for both substantial initial investments and ongoing operational costs is particularly challenging. To overcome these barriers, SMEs must develop sustainable business models capable of covering both the initial setup and long-term operational expenses.

Technological Barriers: Many SMEs rely on outdated or fragmented IT systems that are not designed for modern data-sharing environments. These legacy systems often lack essential features such as interoperability to integrate with modern data-sharing platforms or decentralized frameworks (Alqoud et al., 2022). Furthermore, they also face challenges in implementing advanced features such as consent management, data provenance tracking, and secure access controls. Without the technical foundation to support these capabilities, SMEs are at a disadvantage when attempting to fulfill the functional requirements of a Data Trust.

Organizational Capacity: Beyond technical infrastructure, SMEs frequently face internal capacity challenges related to skilled personnel and expertise in emerging technologies (Zimmermann & Thomä, 2016). There is often a shortage of skilled professionals with knowledge in areas such as secure data architecture, privacy-preserving technologies, and compliance with data governance regulations. This shortage makes it difficult to design, implement, and manage a trustworthy Data Trust platform. Furthermore, the absence of dedicated legal and technical staff exacerbates the burden of regulatory compliance and ongoing system maintenance.

Legal and Regulatory Complexity: The legal environment surrounding data stewardship is complex and rapidly evolving. SMEs must comply with strict requirements set forth by regulations such as the General Data Protection Regulation (GDPR) and the Data Governance Act (DGA), which demand transparency, neutrality, legal compliance, legal unbundling, non-discrimination, security, privacy, and interoperability with other data intermediation services (Carovano & Finck, 2023; EU Regulation 2022/868; Richter, 2023; von Ditfurth & Lienemann, 2022). Navigating these obligations requires significant legal and administrative resources. For SMEs, meeting these obligations often entails restructuring internal processes and making additional investments in compliance mechanisms.

Trust and Stakeholder Management: Trust is a fundamental enabler for the success of any Data Trust. SMEs, however, often face difficulties in demonstrating their ability to meet fiduciary responsibilities, which are essential for building trust with data providers and consumers. Building trust requires effective management of credibility and institutional reputation. Establishing trust involves more than technical security; it requires transparent governance mechanisms, well-defined policies and processes for fiduciary data stewardship, and proactive stakeholder engagement (Paprica et al., 2020). Active communication with stakeholders is critical, particularly with data owners who may have concerns about the risks and benefits of data sharing (Fassnacht et al., 2023; Jussen et al., 2024). Following this, SMEs must also compete with larger organizations that benefit from more established brands and a greater marketing reach. Consequently, SMEs must actively raise awareness about their services and cultivate ongoing engagement with the public and relevant stakeholders (Paprica et al., 2020; Radosevic et al., 2023).

Discussion and Conclusion

This paper presents a literature review and comparative analysis of Data Trust architectures. We propose a layered Reference Architecture model based on key components, which fosters secure, ethical, and transparent data stewardship along with trustworthy data sharing.

Through a structured review of the literature, we identified four distinct architectural models, each varying in terms of their abstraction, implementation strategies, and domains of application. Our comparative analysis highlights significant variations in how Data Trust architectures address core challenges. While O'Hara (2019) provides a conceptually sound governance model emphasizing ethics and transparency, it lacks concrete implementation mechanisms. Lomotey et al. (2022), in contrast, offer a more technically detailed and implementation-oriented architecture, yet it assumes a level of technological readiness that may not be feasible for all organizations. Schinke et al. (2023) contribute valuable insights into sector-specific implementations, but their approach may not be generalizable well across industries. While we derived a generic reference architecture for Data Trusts, future research should synthesize the strengths of these models, creating a more adaptable and operationalizable reference architecture that accommodates diverse organizational needs.

The diversity in these architectures underscores the need for a flexible yet comprehensive reference architecture capable of adapting to various organizational contexts and technological requirements. Based on the comparative analysis and synthesis of existing architectural concepts, the proposed model integrates ten identified key components of Data Trusts into horizontal and vertical layers, which serve as a foundation for the development of a more comprehensive reference architecture to address common challenges and to foster trust among stakeholders, including SMEs aiming to operate a Data Trust. The development should be guided by a comprehensive framework, such as TOGAF. It provides a structured and methodological approach to develop holistic and integrated architectures. Its approach ensures the creation of a consistent and systematic reference architecture for Data Trusts that adequately addresses all relevant architecture layers.

The study also highlights the specific challenges SMEs encounter when aspiring to operate as Data Trusts. SMEs often face barriers such as limited resources and competencies, legal uncertainties, and technological constraints, which hinder their ability to securely manage and share data. Our research aims to address these challenges by developing an initial model for a reference architecture that supports SMEs in establishing themselves as neutral and trustworthy Data Trusts. This facilitates compliance with regulatory frameworks, such as the Data Governance Act, while fostering stakeholder trust through transparent and accountable processes. The key components of the reference architecture model enable SMEs to meet legal and ethical standards, thereby promoting the creation of sustainable and trustworthy data ecosystems.

Limitations and Outlook

This study presents a high-level model for a reference architecture for Data Trusts, organized into vertical and horizontal layers. However, it does not provide a fully detailed reference architecture. Future research will address this limitation by expanding and refining the model to incorporate detailed components, such as conceptual and logical data models, comprehensive process flows, and integrative representations utilizing architecture description languages like ArchiMate (The Open Group, 2023). Collaboration with industry stakeholders across diverse sectors can further validate the model in practical applications and ensure its adaptability to real-world contexts.

Furthermore, while the exclusion of implementation-specific components allowed us to maintain a high level of abstraction and generalizability, this choice may limit the immediate practical applicability of the model for practitioners seeking direct implementation guidance. Future work will address this by extending the architecture with more detailed implementation blueprints and technology stacks, potentially tailored to specific SME use cases or industry contexts.

Lastly, this study limits its systematic comparison to papers that specifically propose architectures for Data Trusts. While this focus is essential, future research should also consider insights from architectures developed for other types of data intermediaries, such as multi-sided platforms (Aulkemeier et al., 2019; Fürstenau et al., 2019; Otto & Jarke, 2019) or data spaces (Cuno et al., 2019; Falcão et al., 2023; Scheider et al., 2023). Despite their differing primary objectives, these intermediaries face shared challenges, including compliance with the Data Governance Act, robust implementation of security and privacy mechanisms, governance frameworks, and interoperability standards. Drawing from these related architectures can help identify shared best practices, uncover innovative solutions to common challenges, and ensure that the proposed reference architecture for Data Trusts aligns with broader trends in data governance and stewardship.

Acknowledgements

This research was supported by the German Ministry of Education and Research (BMBF) under the Funding No. 16DTM218 as part of the NextGenerationEU program of the European Union. The authors are responsible for the content of this publication.

References

Aldenhoff, T. T., Arz von Straussenburg, A. F., & Riehle, D. M. (2024, July 2). Designing for High Availability – A Reference Architecture for IoT Data Platforms. *PACIS 2024 Proceedings*. PACIS 2024, Ho Chi Minh City, Vietnam.

Alqoud, A., Schaefer, D., & Milisavljevic-Syed, J. (2022). Industry 4.0: A systematic review of legacy manufacturing system digital retrofitting. *Manufacturing Review*, *9*, 32. **https://doi.org/10.1051/mfreview/2022031**

Alsaad, A., O'Hara, K., & Carr, L. (2019). Institutional Repositories as a Data Trust Infrastructure. *Companion Publication of the 10th ACM Conference on Web Science*, 1–4. **https://doi.org/10.1145/3328413.3329402**

Angelov, S., Grefen, P., & Greefhorst, D. (2012). A framework for analysis and design of software reference architectures. Information and Software Technology, 54(4), 417–431. **https://doi.org/10.1016/j.infsof.2011.11.009**

Aulkemeier, F., Iacob, M.-E., & van Hillegersberg, J. (2019). Platform-based collaboration in digital ecosystems. Electronic Markets, 29(4), 597–608. **https://doi.org/10.1007/s12525-019-00341-2**

Ayappane, B., Vaidyanathan, R., Srinivasa, S., Upadhyaya, S. K., & Vivek, S. (2024). Consent Service Architecture for Policy-Based Consent Management in Data Trusts. ACM Int. Conf. Proc. Ser., 155–163. Scopus. **https://doi.org/10.1145/3632410.3632415**

Bass, L., Clements, P., & Kazman, R. (2003). Software Architecture in Practice. Addison-Wesley Professional.

Carovano, G., & Finck, M. (2023). Regulating data intermediaries: The impact of the Data Governance Act on the EU's data economy. Computer Law & Security Review, 50, 105830. **https://doi.org/10.1016/j.clsr.2023.105830**

Cuno, S., Bruns, L., Tcholtchev, N., Lämmel, P., & Schieferdecker, I. (2019). Data governance and sovereignty in urban data spaces based on standardized ICT reference architectures. Data, 4(1), 16. **https://doi.org/10.3390/data4010016**

EU Regulation 2022/868, European Commission, Regulation 2022/868 on European data governance (Data Governance Act) (2022). **https://eur-lex.europa.eu/eli/reg/2022/868/oj**

European Commission. (2020, February 19). A European strategy for data. COM(2020) 66 final. **https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066**

Falcão, R., Matar, R., Rauch, B., Elberzhager, F., & Koch, M. (2023). A Reference Architecture for Enabling Interoperability and Data Sovereignty in the Agricultural Data Space. Information, 14(3), Article 3. **https://doi.org/10.3390/info14030197**

Fassnacht, M. K., Benz, C., Heinz, D., Leimstoll, J., & Satzger, G. (2023). Barriers to Data Sharing among Private Sector Organizations. Hawaii International Conference on Systems Sciences (HICSS-56), 3695–3704. **https://doi.org/10.24251/HICSS.2023.453**

Feth, D., & Rauch, B. (2024). Datentreuhänder in der Praxis. Datenschutz und Datensicherheit - DuD, 48(2), 103–109. **https://doi.org/10.1007/s11623-023-1889-3**

Fürstenau, D., Auschra, C., Klein, S., & Gersch, M. (2019). A process perspective on platform design and management: Evidence from a digital platform in health care. Electronic Markets, 29(4), 581–596. **https://doi.org/10.1007/s12525-018-0323-4**

Garcés, L., Martínez-Fernández, S., Oliveira, L., Valle, P., Ayala, C., Franch, X., & Nakagawa, E. Y. (2021). Three decades of software reference architectures: A systematic mapping study. Journal of Systems and Software, 179, 111004. **https://doi.org/10.1016/j.jss.2021.111004**

Gomer, R. C., & Simperl, E. (2020). Trusts, co-ops, and crowd workers: Could we include crowd data workers as stakeholders in data trust design? Data and Policy, 2(433). Scopus. **https://doi.org/10.1017/dap.2020.21**

Hardinges, J. (2018, October 19). Defining a "data trust." The Open Data Institute. **https://theodi.org/insights/explainers/defining-a-data-trust/**

Hardinges, J. (2020, March 17). Data trusts in 2020. The Open Data Institute. **https://theodi.org/news-and-events/blog/data-trusts-in-2020/**

International Data Spaces Association. (2023, February 27). IDS-RAM 4 | IDS Knowledge Base. **https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/**

Janssen, H., & Singh, J. (2022). Data intermediary. Internet Policy Review, 11(1), 1–9. Scopus. **https://doi.org/10.14763/2022.1.1644**

Jussen, I., Möller, F., Schweihoff, J., Gieß, A., Giussani, G., & Otto, B. (2024). Issues in inter-organizational data sharing: Findings from practice and research challenges. Data & Knowledge Engineering, 150, 102280. **https://doi.org/10.1016/j.datak.2024.102280**

Lauf, F., Scheider, S., Friese, J., Kilz, S., Radic, M., & Burmann, A. (2023). Exploring Design Characteristics of Data Trustees in Healthcare—Taxonomy and Archetypes.

Lomotey, R. K., Kumi, S., & Deters, R. (2022). Data Trusts as a Service: Providing a platform for multi-party data sharing. International Journal of Information Management Data Insights, 2, 100075. **https://doi.org/10.1016/j.jjimei.2022.100075**

Nakagawa, E. Y., Guessi, M., Maldonado, J. C., Feitosa, D., & Oquendo, F. (2014). Consolidating a Process for the Design, Representation, and Evaluation of Reference Architectures. 2014 IEEE/IFIP Conference on Software Architecture, 143–152. **https://doi.org/10.1109/WICSA.2014.25**

O'Hara, K. (2019). Data trusts: Ethics, architecture and governance for trustworthy data stewardship. **http://dx.doi.org/10.5258/SOTON/WSI-WP001**

Open Data Institute. (2020, April 28). Designing sustainable data institutions. **https://theodi.org/insights/reports/designing-sustainable-data-institutions-paper/**

Otto, B. (2022). The Evolution of Data Spaces. In B. Otto, M. ten Hompel, & S. Wrobel (Eds.), Designing Data Spaces: The Ecosystem Approach to Competitive Advantage (pp. 3–15). Springer International Publishing. **https://doi.org/10.1007/978-3-030-93975-5_1**

Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the International Data Spaces case. Electronic Markets, 29(4), 561–580. **https://doi.org/10.1007/s12525-019-00362-x**

Otto, B., Ten Hompel, M., & Wrobel, S. (Eds.). (2022). Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer International Publishing. **https://doi.org/10.1007/978-3-030-93975-5**

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, n71. **https://doi.org/10.1136/bmj.n71**

Paprica, P. A., Crichlow, M., Maillet, D. C., Kesselring, S., Pow, C., Scarnecchia, T. P., Schull, M. J., Cartagena, R. G., Cumyn, A., Dostmohammad, S., Elliston, K. O., Greiver, M., Nelson, A. H., Hill, S. L., Isaranuwatchai, W., Loukipoudis, E., McDonald, J. T., McLaughlin, J. R., Rabinowitz, A., … McGrail, K. (2023). Essential requirements for the governance and management of data trusts, data repositories, and other data collaborations. International Journal of Population Data Science, 8(4). Scopus. **https://doi.org/10.23889/ijpds.v8i4.2142**

Paprica, P. A., Sutherland, E., Smith, A., Brudno, M., Cartagena, R. G., Crichlow, M., Courtney, B. K., Loken, C., McGrail, K. M., Ryan, A., Schull, M. J., Thorogood, A., Virtanen, C., & Yang, K. (2020). Essential requirements for establishing and operating data trusts: Practical guidance co-developed by representatives from fifteen canadian organizations and initiatives. Int. J. Popul. Data Sci., 5(1). Scopus. **https://doi.org/10.23889/IJPDS.V5I1.1353**

Radosevic, N., Duckham, M., Saiedur Rahaman, M., Ho, S., Williams, K., Hashem, T., & Tao, Y. (2023). Spatial data trusts: An emerging governance framework for sharing spatial data. International Journal of Digital Earth, 16(1), 1607–1639. **https://doi.org/10.1080/17538947.2023.2200042**

Reidt, A., Pfaff, M., & Krcmar, H. (2018). Der Referenzarchitekturbegriff im Wandel der Zeit. HMD Praxis der Wirtschaftsinformatik, 55(5), 893–906. **https://doi.org/10.1365/s40702-018-00448-8**

Richter, H. (2023). Looking at the Data Governance Act and Beyond: How to Better Integrate Data Intermediaries in the Market Order for Data Sharing. GRUR International, 72(5), 458–470. Scopus. **https://doi.org/10.1093/grurint/ikad014**

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. European Journal of Information Systems, 23(3), 241–255. **https://doi.org/10.1057/ejis.2014.7**

Scheider, S., Lauf, F., Möller, F., & Otto, B. (2023). A Reference System Architecture with Data Sovereignty for Human-Centric Data Ecosystems. Business & Information Systems Engineering, 65(5), 577–595. **https://doi.org/10.1007/s12599-023-00816-9**

Schinke, L., Hoppen, M., Atanasyan, A., Gong, X., Heinze, F., Stollenwerk, K., & Roßmann, J. (2023). Trustful Data Sharing in the Forest-based Sector—Opportunities and Challenges for a Data Trustee. 49th International Conference on Very Large Data Bases (VLDBW'23), Vancouver, Canada.

Schinke, L., & Roßmann, J. (2024). Enabling Trustful Data Sharing in Industry 4.0—A Comparison between Data Spaces, Data Markets and Other Related Concepts. TechRxiv. **https://doi.org/10.36227/techrxiv.172651024.46145536/v1**

Schweihoff, J., Lipovetskaja, A., Jussen-Lengersdorf, I., & Möller, F. (2024). Stuck in the middle with you: Conceptualizing data intermediaries and data intermediation services. Electronic Markets, 34(1), 48. **https://doi.org/10.1007/s12525-024-00729-9**

Stachon, M., Möller, F., Guggenberger, T., Tomczyk, M., & Henning, J.-L. (2023). Understanding Data Trusts. ECIS. Thirty-first European Conference on Information Systems (ECIS 2023), Kristiansand, Norway.

Stalla-Bourdillon, S., Thuermer, G., Walker, J., Carmichael, L., & Simperl, E. (2020). Data protection by design: Building the foundations of trustworthy data sharing. Data & Policy, 2, e4. **https://doi.org/10.1017/dap.2020.1**

The Open Group. (2019). Reference Architectures and Open Group Standards for the Internet of Things. **http://www.opengroup.org/iot/wp-refarchs/p2.htm**

The Open Group. (2022). The TOGAF Standard. **https://pubs.opengroup.org/togaf-standard/architecture-content/index.html**

The Open Group. (2023). ArchiMate® 3.2 Specification. **https://pubs.opengroup.org/architecture/archimate3-doc/**

vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. **https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1145&context=ecis2009**

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Communications of the Association for Information Systems, 37. **https://doi.org/10.17705/1CAIS.03709**

von Ditfurth, L., & Lienemann, G. (2022). The Data Governance Act: – Promoting or Restricting Data Intermediaries? Competition and Regulation in Network Industries, 23(4), 270–295. Scopus. **https://doi.org/10.1177/17835917221141324**

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26(2), xiii–xxiii. **http://www.jstor.org/stable/4132319**

Zimmermann, V., & Thomä, J. (2016). SMEs face a wide range of barriers to innovation—Support policy needs to be broad-based. KfW Research. Focus on Economics, No 130.