

DATA MESH - A CASE STUDY PERSPECTIVE ON BUILDING INDUSTRIAL DATA PLATFORMS

Completed Research Paper

Philipp Kernstock, Technical University of Munich, Germany, philipp.kernstock@tum.de

Konstantin Biermann, Technical University of Munich, Germany, k.biermann@tum.de

Sebastian Sartor, Technical University of Munich, Germany, sebastian.sartor@tum.de

Anna Wimbauer, Technical University of Munich, Germany, a.wimbauer@tum.de

Julian Bohnet, Technical University of Munich, Germany, julian.bohnet@tum.de

Timo Böttcher, Technical University of Munich, Germany, timo.boettcher@tum.de

Andreas Hein, University of St. Gallen, Switzerland, andreas.hein@unisg.ch

Abstract

In an era where data is the new currency, organizations are contending to harness vast data influxes for data-driven innovation. Centralized data architectures, with extensive data warehouses and lakes, are buckling under the pressure of contextually complex and distributed demands, leading to data bottlenecks and alignment issues within organizational structures. Data Mesh emerges as a decentralized paradigm promising to surmount these hurdles by advocating domain ownership and federated governance. This research delves into the practicalities of Data Mesh through a case study of Alpha, a German manufacturing firm, revealing a transformative shift towards a data-driven modus operandi. It details Alpha's journey from siloed data storage to a unified, transparent data platform that repositions IT as an innovation linchpin. Bridging the theoretical and practical, the paper underscores Data Mesh's potential to amalgamate decentralized data ownership with central oversight, offering a roadmap for organizations to navigate the data-centric future effectively.

Keywords: Data Management, Data Mesh, Data Platform, Data-Driven Innovation.

1 Introduction

Organizations today are navigating an era marked by an unprecedented surge in data generation and consumption. This data influx, sourced from enterprise systems, e-commerce, social media, and machine logs, is a key driver of innovation, particularly in artificial intelligence (AI) (Mohr and Hürtgen, 2018). For instance, in the financial sector, AI algorithms analyze transaction data in real time to detect fraud and safeguard customer transactions (Desai, 2023). However, actualizing this potential in industrial practice presents significant challenges: A recent study by PwC shows that, despite considerable investments, 76% of projects that use AI barely break even (PwC, 2020). Other estimates expect that up to 80% of AI projects fail eventually (Bojinov, 2023). These studies illustrate that data access and quality are still one of the main roadblocks to realizing value from AI (Gröger, 2021).

As data diversity and volume increase and AI applications become more complex, centralized data architectures, such as Data Warehouses (DW) or Data Lakes (DL), demonstrate their inadequacy in meeting evolving demands regarding data quality and access to enable data-driven innovation and increasing responsiveness through distributed development of AI solutions (Dehghani, 2022; Machado et al., 2022). These centralized systems are characterized by rigid, tightly coupled stages of data

processing like ingestion, cleansing, aggregation, and serving. The rigidity and tight coupling of these stages often create bottlenecks in data discovery and introducing new features, slowing down the delivery and scalability of AI solutions (Dehghani, 2019; Nargesian et al., 2019). Another problem is that centralized architectures tend to lead to data silos that reduce the findability, accessibility, interoperability, and reusability of data in organizations (Zhang et al., 2022). As a result, organizations find it increasingly difficult to utilize their data effectively (Gröger, 2021).

Extant research emphasizes the increasing decentralization of data management in modern organizations across various sectors (Lefebvre et al., 2021). This highlights a pivotal shift in how knowledge is produced and used within organizations to mitigate the known problems of centralized structures and the resulting data silos (Alaimo and Kallinikos, 2022). This shift can be illustrated by the evolution of centralized DW and DL to more integrated architectures like Data Lakehouses (DLH) (Armbrust et al., 2021) and socio-technical approaches like Data Mesh (Dehghani, 2019). Data Mesh is a decentralized approach characterized by principles of domain ownership and federated governance to overcome the challenges of centralized structures. Data Mesh transcends the existing technology-focused concepts of DW, DL, and DLH by providing a socio-technical approach to data architecture (Dehghani, 2019).

However, despite the growing adoption of Data Mesh in practice, the empirical understanding of this architecture remains limited. Extant research has explored aspects of Data Mesh, such as data ownership and data governance (Joshi et al., 2021). However, transitioning from centralized architectures to this new approach requires further exploration and understanding. In particular, we lack an understanding of the practical implementation process and the governance of change in the organizations (Machado et al., 2022). Furthermore, research offers little insight concerning solutions, boundary conditions, and tensions governing Data Mesh as a federated socio-technical data architecture approach (Bode et al., 2023; Machado et al., 2022). Understanding these points is crucial to successfully realizing decentralized data architectures and overcoming the limitations of centralized data architectures to enable data-driven innovation. Hence, our research seeks to bridge this knowledge gap by investigating *how organizations can effectively implement Data Mesh as a federated socio-technical data architecture approach to cultivate data products and enable data-driven innovation*.

We employ a single-case study of Alpha, a large German manufacturing company, to explore implementing a Data Mesh architecture. Alpha, undergoing a significant transformation in its business model and IT infrastructure, presents a case of an incumbent industrial company that has become an early adopter of Data Mesh to transition towards a data-driven approach. The research methodology involved conducting 25 semi-structured interviews over four months in 2023, involving managers and employees knowledgeable about Data Mesh and data management challenges within Alpha and external data experts. Complemented by internal and external documents, these interviews facilitated an in-depth understanding of Alpha's organizational and technological shifts. The data analysis was guided by an inductive approach, allowing key themes to emerge naturally, thus providing a nuanced view of the Data Mesh implementation in a real-world industrial setting.

The results describe Alpha's strategic shift towards a Data Mesh architecture from a process perspective (Burton-Jones et al., 2015). The multifaceted implementation involved restructuring data storage and access by breaking down existing silos and creating a unified platform for easier data integration and sharing across departments. Additionally, Alpha invested in developing transparency of data sources, facilitating easier access and a better understanding of data. To align with the new architecture, Alpha redefined the role of IT from a cost center to a key player in innovation and data-driven strategies. The study's findings draw theoretical and practical implications. Theoretically, this approach mirrors the balance of decentralization and centralization in managing multi-business organizations (Acemoglu et al., 2007) and federated IT governance (Williams and Karahanna, 2013). Practically, the research shows how a Data Mesh transformation can be driven through a blend of technical strategies, such as implementing data lineage, and organizational and cultural measures, such as empowering organizational change. The insights contribute to data management literature by highlighting a socio-technical approach that enables data-driven organizations to implement and benefit from a federated data management model.

2 Theoretical Background

2.1 Data management from an organizational perspective

The increasing significance of data across various industries, such as manufacturing, healthcare, retailing, and finance, has been driven by the integration of digital technologies. This integration marks a pivotal change in knowledge production and use within and across organizations. Data are not just tools for management but have become crucial for organizations to understand and respond to their environments, shifting from a supportive element to a core resource (Alaimo and Kallinikos, 2022). However, challenges arise when aligning organizational structures with logical data pathways, often resulting in data silos and unclear responsibilities (Janssen et al., 2020). To address these challenges, the implementation of robust data governance is critical. Data governance is essential for promoting effective data practices, minimizing risks, and fostering an organizational data-centric culture (Benfeldt et al., 2020). Moreover, creating reliable and accurate data is crucial for trustworthy AI and overall trust-building within organizations (Janssen et al., 2020). Data governance specifies a cross-functional framework to manage data as a strategic asset (Abraham et al., 2019). It encompasses decisions about data and accountability for these decisions (Khatri and Brown, 2010; Otto, 2011). Generally, governance concerns decision-making authority, while management focuses on executing these decisions (Khatri and Brown, 2010). Data governance defines and oversees policies, standards, and procedures to ensure data quality, consistency, usability, security, privacy, and availability (Abraham et al., 2019).

Data governance operates along a spectrum that ranges from centralization to decentralization, opening up various degrees of autonomy across distinct decision-making domains (Khatri & Brown, 2010). However, little research investigates this balance of data governance and how to establish it (Weber et al., 2009). Thus, we draw on parallels from federated IT governance, where equilibrium is sought between the autonomy of individual business units and the cohesive decision-making required at an enterprise level to prevent disjointed local solutions while fostering the individual interests of various divisions (Williams and Karahanna, 2013). Federated models are particularly relevant to organizations that thrive on innovation and adaptability in fast-paced environments, which is increasingly crucial in environments where data is ubiquitous (Barkey & Godart, 2013). In federated IT governance, aligning enterprise objectives with local needs involves enhancing the capabilities of shared IT service groups, fostering collaborative relationships between IT and business managers to build trust, and putting in place adequate mechanisms that guide well-informed decision-making regarding IT investments and the business value that originates from these shared services (Fonstad and Subramani, 2009).

The FAIR data principles, originating in the scientific community, advocate that data should be findable, accessible, interoperable, and reusable (Wilkinson et al., 2016). Findable refers to easily locating data. Accessible means data should be retrievable and available to users. Interoperable suggests data should be in standardized formats. Reusable stresses data should be well-documented to provide the context for effective utilization (Labadie et al., 2020). The principles complement industrial data governance and ensure data is managed effectively, enhancing its value and utility across domains and applications.

2.2 Data warehouses and data lakes

In the 1990s, DW gained popularity for their ability to conduct nightly ETL (Extract, Transform, Load) processes. These processes involved gathering, transforming, and consolidating data from multiple sources into a structured data cube equipped with analytical tools for Business Intelligence (BI) (Gardner, 1998). DW excelled in quickly processing complex and frequent analytical queries, thus providing valuable insights for business decision-making (Gardner, 1998). However, the emergence of big data posed challenges to traditional DW, particularly regarding scalability and flexibility. This led to the development of DL in the past two decades. DL address the storage needs for high-speed, hybrid, unstructured data. They aggregate and store data from various sources in large data systems (Miloslavskaya and Tolstoy, 2016). This data is then analyzed, often using DW, to yield insights for BI. Challenges arise when data becomes overwhelming and unmanageable due to processing delays, leading

to resource wastage and inaccurate performance evaluations in metadata management systems (Nargesian et al., 2019).

A recent development is the DLH concept, which aims to merge the strengths of DL and DW to address their limitations. The DLH architecture supports big data storage, analytics, and decision query processes, repurposing existing technology where possible (Armbrust et al., 2021; Inmon et al., 2021). The DLH development indicates a shift in data architecture to meet evolving data needs, transitioning from traditional warehousing to more flexible solutions like DL. Despite this evolution, the centralization of data and a monolithic data management architecture have been persistent. However, as data assumes a more pervasive role in organizations, leading to their decentering (Alaimo and Kallinikos, 2022), there is a parallel shift in data architecture, mirroring the change in software engineering from monolithic to microservice architectures. This leads to approaches like Data Mesh that move away from central data architectures (Machado et al., 2022).

2.3 Data Mesh

Practitioners propose that Data Mesh, a novel concept, holds the potential to overcome the limitations of centralized data architectures such as DW or DL, ushering in a new era of data-driven innovation at a broader organizational scale (Dehghani and Fowler, 2022). Table 1 contrasts the technology solutions with Data Mesh as a socio-technical data management solution. Data Mesh was introduced as a practical solution by Dehghani (2019) to address the challenges of scaling and efficiently managing data within enterprises and revolves around four core principles, leaving the concrete technical realization open:

Domain ownership: This principle emphasizes assigning data ownership to individual domains or business units within an organization. Instead of a centralized data team managing all aspects of data, each domain is responsible for its data products, including collection, storage, processing, and quality. The decentralization allows domain experts to better understand their data, improving data quality, context, and usability.

Data-as-a-Product: Data is treated as a valuable product delivered to users within the organization. Data products should have well-defined contracts, a Service Level Agreement, and documentation like software products. This approach encourages data providers to create high-quality, reusable data products that can be consumed by different teams, promoting collaboration and reducing redundancy.

Self-serve data platform: This principle advocates for providing self-serve tools and platforms to empower data producers and consumers. Data infrastructure should allow domain teams to ingest, process, and serve data without relying heavily on central data engineering teams. This self-serve capability accelerates the data delivery process, promotes agility, and reduces bottlenecks.

Federated computational governance: Data governance is distributed across various domain teams, each responsible for ensuring data quality, privacy, compliance, and security. This principle promotes a federated model where domain teams collaboratively define and enforce governance policies. Centralized policies and standards can still exist but are adapted to the specific needs of each domain.

Existing literature on Data Mesh is primarily practitioner-driven (Goedegebuure et al., 2023). However, research in computer science and software development domains has recently emerged, delving into logical architecture, motivation factors, challenges, privacy concerns, and real-world use cases (Joshi et al., 2021; Machado et al., 2021). Several studies have identified motivational factors for adopting Data Mesh, such as reducing bottlenecks and leveraging domain knowledge (Bode et al., 2023). Furthermore, Data Mesh has been recognized as a vital enabler for sharing clinical data across organizations (Falconi and Plebani, 2023), with grey literature highlighting benefits including enhanced scalability, agility, data quality, and governance (Goedegebuure et al., 2023). Pörtner et al. (2023) suggest that Data Mesh may serve as an agile and adaptable framework for data-centered services and enable data ecosystems. However, organizations may face challenges transitioning from centralized to federated governance, shifting responsibilities, ensuring metadata quality, and defining data products (Bode et al., 2023).

In addition to the conceptual exploration, several studies have delved into the practical aspects of Data Mesh, such as its logical architecture. Architectural frameworks tailored to specific use cases and

practical examples have been proposed by various researchers (Hooshmand et al., 2022; Machado et al., 2021). Some argue that the logical architecture, built on domain-driven design and microservice architecture, forms the cornerstone of a scalable and flexible data management platform for enterprises (Hooshmand et al., 2022). Real-world case studies emphasize the importance of federal governance for balanced control and productivity (Joshi et al., 2021). They advocate clear ownership, scalable governance, and empowering diverse teams to promote secure data-driven innovation and better outcomes. Moreover, Hasan and Legner (2023) delve into the motivations behind firms developing data products and outline essential traits of data products, emphasizing recurring needs, a well-defined consumer base, measurable value, clear ownership, and being easily consumable.

| | Data Warehouse (cf. Gardner, 1998) | Data Lake (cf. Miloslavskaya and Tolstoy, 2016) | Data Lakehouse (cf. Armbrust et al., 2021) | Data Mesh (cf. Dehghani, 2022) |
|-------------------|--|--|--|--|
| Type | Technology architecture solution | Technology architecture solution | Technology architecture solution | Socio-Technical data management solution |
| Definition | A centralized repository for assembling and managing structured, data from various sources, optimized for reporting. | A centralized repository for large volumes of raw, unstructured, and semi-structured data. | A hybrid model combining features of DL and DW, offering structured and unstructured data storage. | A decentralized approach to managing and accessing data, emphasizing domain-oriented ownership and self-serve data infrastructure. |
| Focus | Querying and analyzing structured data. | Storage and large-scale processing of raw data in its native format. | Combine the scalability of DL with the management features of DW. | Provide data management methods that align with technical solutions and facilitate a flexible and interoperable data architecture. |
| Use Cases | BI, reporting, and historical analysis. | Big data processing, ML, and exploratory data analysis. | A mix of BI, real-time analytics, and ML. | Large-scale organizations with distributed teams and diverse data needs. |

Table 1. Comparison of technology architecture solutions with Data Mesh.

However, research on the organizational adoption of Data Mesh, necessitating profound technical and organizational adjustments, is still in its infancy and remains relatively scarce (Joshi et al., 2021; Machado et al., 2022). The existing literature, given its origin in computer science and software engineering rather than information systems research, predominantly provides descriptive accounts of the concept and its technical implementation and lacks a comprehensive exploration of the critical elements, such as organizational approaches, boundary conditions, and tensions involved in implementing this federated socio-technical data management approach.

3 Research Design

To explore organizational approaches, boundary conditions, and tensions in implementing Data Mesh as a federated socio-technical data management approach, we employ an exploratory single case study of Alpha, a globally operating manufacturing firm. An in-depth single case study is appropriate to answer the research question of how organizations can effectively implement Data Mesh, as it enables us to gain rich insights into the transformation process toward a data-centric organization (Sigelkow, 2007; Yin, 2018). Our research offers a starting point to address the limited theoretical knowledge regarding the implementation process of Data Mesh and the related data-centric transformation (Sigelkow, 2007). We conducted 25 interviews with the IT organization and data domains, such as marketing or finance. This allowed us to create an encompassing picture of how data is created, governed, and used within Alpha and develop a process perspective (Burton-Jones et al., 2015).

3.1 Case study setting at Alpha

Alpha is a German manufacturer which operates globally and runs production sites in Europe and Asia. The company is decentrally organized with high autonomy of the different business lines. The company makes between 1-5 billion € in revenue. As part of their latest strategy, data and digital products are introduced. Selecting Alpha to explore Data Mesh gives us an insightful perspective, as Alpha, on the one hand, represents a typical incumbent industrial organization that is pressured by digital innovation's imminent impact on its business model. On the other hand, Alpha is an early adopter in the industrial sector of Data Mesh, making it a fruitful edge case for research (Siggelkow, 2007). The industrial context of Alpha as an incumbent organization that has to manage diverse data sources, originating from a legacy machine park and IoT data, with high complexity, while its business model is not directed towards a data-driven approach, distinguishes it from services or software companies. Opting for Alpha as the focal point of investigation and selecting a single-case study design allows us to immerse into the case and get deeply accustomed to its cultural and organizational intricacies, revealing the motivations and challenges that are inherent in an industrial setting that is undergoing significant changes (Yin, 2018). The brownfield metamorphosis at Alpha confronts existing processes and IT architecture with the overall goal of fundamentally transforming the business. A central ERP system dominates Alpha's legacy IT landscape. Additional systems are used for customer relationship management and operations. Analytical data usage, such as BI, is enabled through siloed DW. For example, the marketing domain uses one DW, and the finance department uses another DW. There are no DL for analytical use cases, but data pipelines are setup ad hoc for specific use cases. While implementing Data Mesh, Alpha created a new data platform handling data ingestion, processing, persistence, consumption, and governance.

3.2 Data collection and analysis

We collected data during four months in 2023 from multiple sources such as interviews and external and internal documents. We conducted 22 interviews with managers and employees possessing crucial insights into Alpha's Data Mesh implementation and data management challenges and three additional interviews with external data management experts. Interview durations ranged from 15 to 60 minutes, with a median duration of 45 minutes, contingent upon the interviewee's role and the depth required for their expertise. These discussions comprehensively covered topics spanning data platform architecture, IT Strategy, Data Mesh implementation, ERP Integration, data engineering and analytics, change processes, and team structures. The interviewed participants represented a spectrum of expertise, comprising technical specialists such as data platform developers, IT managers, the CIO, data engineers, data consumers like data analysts, finance and marketing managers, and organizational experts such as change managers and agile coaches. We employed a semi-structured interview guide to maintain a standardized approach while allowing flexibility for in-depth exploration. We ensured the validity and reliability of collected data by cross-referencing inputs from multiple interviews or corroborating information with secondary sources like internal reports or documents (Eisenhardt, 1989; Yin, 2018).

We followed an inductive coding approach to analyze the collected data from the interviews. Starting with open coding, common themes were identified through axial coding, allowing us to label and cluster the data into second-order themes and aggregate dimensions (Corbin and Strauss, 1990; Gioia et al., 2013). For instance, we combined the open codes of quotes like *"Does the strategy exist? No, it does not exist"*, *"Our IT strategy? Forced. I would almost say we don't really have one, from my point of view"*, and *"There is no data management strategy in a company-wide context"* as *IT strategy and data management issues* related second-order theme. We then combined this second-order theme with related ones, such as *Organizational Bottlenecks*, and created the aggregate dimension *Challenges*.

We constantly triangulated our interview insights with secondary data and compared them to existing literature during coding. We repeatedly analyzed the transcripts and our previous findings to restructure our results until they could be illustrated in a data structure that offers a comprehensible representation of the different concepts and their associations with each other (Gioia et al., 2013). The analysis allowed

us to cluster our findings into four dimensions: challenges igniting a transformation process toward the federated data architecture, how the four Data Mesh principles were implemented, resulting implementation challenges, and emerging organizational approaches. We then established structural links between the dimensions and included the Data Mesh principles as the fifth dimension in our process-theoretic framework to visualize how the principles are addressed (Burton-Jones et al., 2015).

4 Results

The results of our study describe Alpha's strategic shift towards a Data Mesh architecture, a decision to overcome data fragmentation and enhance data utilization across the organization, from a process perspective. After outlining the challenges leading to Data Mesh implementation, we show how the Data Mesh principles address these challenges and how implementing the principles leads to new challenges that are overcome through a set of best practices. To align with this new architecture, the company also initiated a cultural shift, redefining the role of IT from a cost center to a key player in innovation and data-driven strategies. This comprehensive approach aimed to solve immediate data integration issues and foster a long-term transformation towards a data-centric organizational model. Figure 1 gives an overview of the challenges addressed by the Data Mesh principles (see Chapter 2.3), leading to implementation challenges tackled through organizational approaches. We then outline how the organizational transformation is shaped by organizational legacy, operating model fit, conflicting IT cultures, and balancing decentralized and centralized efforts.

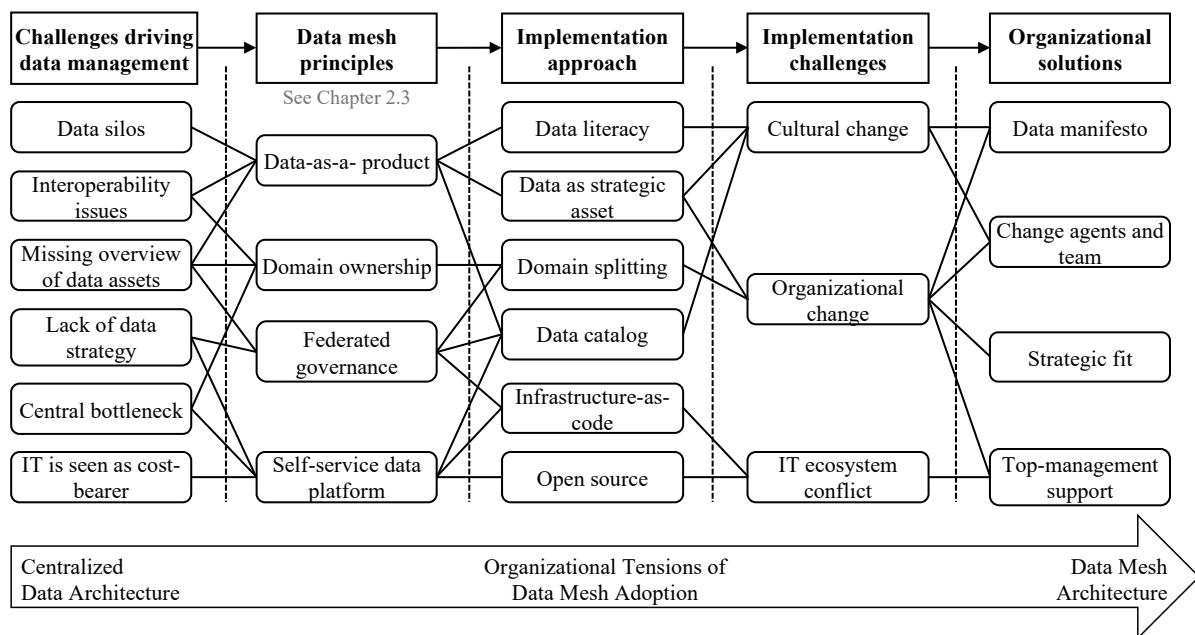


Figure 1. Alpha's Data Mesh transformation journey.

4.1 Challenges initiating the transformation process

“If you don't offer a digital twin for your products in a few years, the customer will simply go to another provider, and then the business will be completely lost.” – Finance Manager

Alpha's industrial business model is under increasing pressure due to the emergence of digital innovations. This has spurred the firm's ambitions to transform into a data-centric organization. Driven by a new strategy emphasizing the relevance of digital and data products, the current focus on developing and producing physical products is to be augmented with the development of digital services. However, the existing data architecture and management approach creates multiple challenges that hinder transformative efforts and led to the decision of Alpha to implement a federated data architecture. The challenges are sorted from technical to organizational to cultural ones.

Challenge 1: Data silos

“Today, if you have centralized silos, large data warehouse systems, you can't open them up to the entire organization overnight, and everyone can do what they want.” - Data Platform Team Lead

One of the prominent challenges of Alpha lies in the existence of data silos within various departments, as these silos generally complicate access and usage of that data for developing digital products, such as AI applications. However, integrating these silos across different departments and business units proves difficult. While data processing within individual departments may be relatively straightforward, leaving departmental or system boundaries creates issues regarding data findability, accessibility, interoperability, and reusability. For example, one customer might not have the same customer identity (ID) across different departments. Consequently, merging data necessitates significant manual effort and can impede the seamless flow of information across the organization.

Challenge 2: Interoperability issues

“There is some information on interfaces, but it doesn't scale because it's not known and it's too much” – IT Manager

The second challenge lies in the lack of an overview of IT architecture and interfaces, leading to the re-implementation of existing system interfaces and obstructing smooth data flow through the organization. Alpha's current IT architecture has grown historically and is thus very complex. Different systems are strongly interconnected, and a global overview of the IT architecture does not exist. For instance, the lack of central documentation of system interfaces led to the re-implementation of existing system interfaces, which is highly inefficient. The lack of well-defined interfaces between systems and applications can obstruct the smooth data flow across the organization. This interoperability gap hinders timely and accurate information exchange, affecting real-time insights and proactive decision-making.

Challenge 3: Missing overview of data assets

“We need a clear map. Where is the data? Which systems are accessible via which interface?” - CIO

The third challenge is the absence of a comprehensive overview of existing data, which poses a significant obstacle to Alpha's aspiration of becoming a data-driven enterprise. It hinders collaboration across the organization and prevents leveraging the potential of existing data sources. Insufficient transparency and documentation of data sources make it often unclear how data can be accessed as there is no central data access. Thus, data is often difficult to find and access and employees must either “just” know where to find and how to access the data or ask one of their colleagues. Without clear transparency of available data sources, departments struggle to leverage the full potential of data assets. This documentation deficit impedes informed decision-making and inhibits cross-functional initiatives.

Challenge 4: Lack of data strategy

“There is no data management strategy in a company-wide context.” - Finance Manager

The fourth challenge identified in Alpha's transformation process is the absence of a clearly articulated data management strategy, causing a lack of guidance on how data-driven innovations, such as AI solutions, should be enabled from an IT perspective. Without a unified vision for how data should be governed, accessed, and utilized, departments may unintentionally work towards different goals. This lack of strategic alignment can lead to inefficiencies, redundancies, and missed opportunities for deriving value from the organization's data assets.

Challenge 5: Central bottleneck for data initiatives

“The IT teams have been completely overloaded for years because they are being bombarded from all directions and can't cope with everything that's coming at them.” - Finance Manager

The fifth challenge is the bottleneck created by centralized resources relevant for data-driven projects, leading to delays and suboptimal outcomes of data initiatives. The IT organization's structure causes data innovation projects to depend on centralized specialists like data engineers and scientists. These experts are crucial for all data services but often lack specific domain knowledge due to the varied origins of use cases, leading to increased time for data preparation and bottlenecks in project workflows.

Challenge 6: IT is seen as a cost bearer

“IT must not be seen as a cost bearer but as an innovation partner” – Data Expert

The last but central challenge Alpha faces revolves around the perception of IT as a cost bearer rather than an innovation enabler, which is critical for fully embracing data-driven innovation. This discrepancy contradicts the fundamental objective of becoming a data-driven company. To fully leverage data for informed decision-making, IT must be recognized as a strategic partner facilitating innovation and providing the technological foundation for data initiatives. Shifting this perception is essential to harness the true potential of a data-driven transformation.

4.2 Implementation approach for federated data architecture

Current data silos arise from applications and data pipelines that are developed centrally for single-use cases but cannot be leveraged for other purposes. Thus, the current landscape is defined by distributed and monolithic data silos that are not accessible on an organizational scale. By implementing a federated data architecture, Alpha wants to liberate data from its silos and enable its employees to utilize the existing data to foster the transformation towards a data-centric organization. The decision was made that Data Mesh would be the appropriate concept to achieve these goals.

Alpha aimed at tackling challenges 1 and 2 through the Data Mesh principle of *data-as-a-product*. Alpha used the existing top-management perspective that regards data as a strategic asset and decided to develop and implement data principles in adherence with the FAIR data principles to improve data literacy across the organization. The setup of a data catalog helps implement the FAIR principles and facilitates finding and using data products. Further, product thinking is adopted to achieve higher product quality. Implementing the data product principles throughout the organization drives the creation of data products for internal and external use. However, when starting the implementation of the *data-as-a-product* principle, Alpha had to overcome cultural challenges regarding accepting data as an asset organization-wide and fostering investment decisions towards data.

To overcome challenges 2 and 3, Alpha began the implementation of the *domain ownership* principle. Alpha created an initial domain for the data platform to drive foundational topics through a centralized interdisciplinary team. Cutting domains was primarily driven by using functions already engaging a lot with analytical data. Secondarily, the existing data resources were leveraged to define domains. The goal of establishing decentralized data domains with centralized support was challenged by the difficulties of establishing clear boundaries between the domains and scaling the delegation of ownership. Ensuring support for this approach and promoting it across the organization proved difficult.

Further, to address challenges 3 and 4, the *federated computational governance* principle was implemented by introducing technology solutions for policing, observing, and cataloging data. To do so Alpha drew on the implementation approaches of domain splitting, infrastructure-as-code (IaC), and establishing a data catalog. The meta-data captured through the data catalog essentially acts as data product information for potential users, thus enabling the setup of a data marketplace, where users can then access the data products based on this information. The meta-data, semantically stored in a knowledge graph, is the base for the federated computational governance structure, which is further based on policies-as-a-code, such as data contracts. The approach of federated governance was welcomed, but creating the operating model and related business processes proved challenging, as, for instance, filling the data catalog requires extra effort that does not necessarily benefit the acting person.

Finally, building the *self-service data platform* addresses challenges 4, 5, and 6. Alpha leveraged IaC and open-source principles and built the data platform using technologies like Apache Iceberg, MinIO, and Trino. The platform uses data products as logical storage entities rather than implementing a data lake for physical data storage. The data platform provides IaC solutions to enable distributed building of new data products. As a centrally hosted tool, the data catalog functions as the discovery layer. The platform is open to introducing new technology, which is provided centrally to all domains. While implementing the data platform, conflicts of interest and interoperability arose between the established IT ecosystems, based on SAP and Microsoft technology, and the new open-source-based platform architecture, creating an additional challenge within the transformation process.

4.3 Cultural, organizational, and socio-technical implementation challenges

Alpha's transformation process towards a data-centric organization through the implementation of Data Mesh presents multifaceted challenges, underscored by the need for cultural and organizational shifts and IT ecosystem adjustments, which take a socio-technical perspective. The following chapter outlines the challenges during Alpha's Data Mesh implementation process.

Culturally, while Alpha's business strategy acknowledges the significance of data and digital products, this recognition has not thoroughly permeated all business areas. Strategic plans that include data harmonization are ineffective unless practically implemented, which is crucial for realizing their potential. Further, valuing data as a resource and actively engaging in its management to derive value is vital—a practice not yet widespread within Alpha. This is illustrated by the utility of data in decision-making, such as determining the phase-out of unique components during product discontinuation. However, establishing data literacy remains a significant hurdle, as evidenced by instances where departments fail to recognize the benefits of insights from initiatives like process mining.

“Data is the new gold. [...] But the thing is, a lot of people do not understand, there comes change with it.” - IT Manager

Organizational change is another challenge, particularly in establishing product thinking within the data domain. Data Mesh advocates for viewing data as products, which requires a mindset shift but promises to enhance data quality and the value-to-cost ratio. Defining what constitutes a data product—whether an attribute, a software component, or another entity—is vital for clarity and ownership, which can address data quality and support issues. Nevertheless, ownership can be contentious, with some data lacking clear custodianship and other data becoming the center of ownership disputes. The willingness to share data internally and externally represents a cultural and organizational challenge. While data is recognized as valuable, akin to gold, the change it brings must be more widely understood, impacting the willingness to share. Domains must commit resources, and the realization of value from data often depends on a network effect across the platform, which can be a complex dynamic to navigate.

Expanding the need for organizational change into a socio-technical dimension, the IT ecosystem at Alpha is at a crossroads with adopting a best-of-breed approach for Data Mesh. The best-of-breed approach leans heavily on open-source software, resonating with Alpha's core value of retaining vendor independence. This approach, however, brings new risks and has sparked debate. A competing proposition favors a standard-software-based data platform with the challenge of managing and integrating two distinct systems. The cultural and operational divide is particularly pronounced concerning the central ERP ecosystem, which maintains standardized processes across domains. This divergence signifies a broader challenge for Alpha in balancing innovation with established practices.

4.4 Organizational approaches to overcoming challenges

To overcome the emerging implementation challenges, Alpha derived strategic initiatives to navigate the transformation toward Data Mesh. These organizational approaches proved to be success factors in overcoming the cultural, organizational, and socio-technical challenges. Firstly, Alpha created an autonomous unit that has been critical in effectively implementing Data Mesh. Established as an interdisciplinary unit in 2018 to create a unified data structure for products, the team's scope has progressively broadened, spearheading Data Mesh's adoption across various work streams and application scenarios. Uniquely positioned under the R&D board area rather than the IT board area, the team encompasses technological and organizational transformations. Secondly, the autonomous unit crafted a manifesto outlining the objectives and intrinsic value of data within Alpha. This document acts as a strategic beacon, categorizing aims into organizational, semantic, and technical dimensions. A central tenet of the manifesto is the adherence to FAIR data principles.

Top management buy-in was achieved early on by implementing one highly relevant strategic use case. Maintaining this support by continuously delivering tangible results has been a success factor, as implementing Data Mesh means a business transformation, where top management is crucial. The Data

Mesh unit's overarching objective is to scale the approach throughout the organization, engaging in initiatives ranging from technical integration to governance and communication. Technical integration includes developing a data catalog, a centralized characteristic repository, a comprehensive data platform, and various data product use cases. A key initiative is constructing a self-service data platform founded on Open-Source technology and a Best-of-Breed paradigm. This platform integrates various technologies for data ingestion, persistence, processing, and consumption.

Furthermore, the team collaborates with various domains to develop and implement use cases. For instance, they have created a global product ID linked to a data hub that provides comprehensive datasets and insights about products from multiple sources. Beyond technical solutions, the Data Mesh unit is tasked with developing a data governance model for Alpha. This encompasses defining roles, requisite skills, and a domain concept to guide the communication strategy and rollout plan, thereby disseminating information about ongoing data and data management advancements. After initially prioritizing individual use cases, the unit is undergoing a strategic shift and now focuses on establishing a global governance framework. Their efforts now center on laying down foundational processes and technologies for the Data Mesh architecture. This strategic pivot reflects Alpha's commitment to embedding robust data governance and management practices, which are integral to realizing the full potential of Data Mesh in enhancing organizational agility and data democratization.

4.5 Organizational perspective on Data Mesh adoption

Alpha's IT organization is transforming to a decentralized operating model to better fit its global, multi-business structure. This change aims to harness domain-specific knowledge and operational processes across the decentralized model.

"We want to bring this decentralization, which we actually live as a company, into the world of technology, data, and analytics." - Data Platform Team Lead

Consequently, Alpha is moving away from a centralized data management approach less conducive to the decentralized organizational approach of the globally operating multi-business organization and its strategic direction. The shift from the organizational legacy to a Data Mesh-based operating model is driven by a use-case-based rollout and centralized governance efforts to enable the federated data architecture. A clash between the existing ERP ecosystem and the promoted open-source software approach creates tension within the transformation process, visualized by the arrow in Figure 1.

The move toward a decentralized model involves integrating data teams into individual domains to alleviate bottlenecks caused by centralized data engineering. The business lines are supported by the global IT department and an internal IT department within the business lines. This integration is essential for aligning the development of data products with the core of domain-specific expertise. The IT departments of the business lines are willing to do more and support new processes and tools within their domains. Further, introducing product thinking in the domains is emerging as a crucial factor in enhancing data quality and availability across Alpha.

However, this transformation is challenging. The existing operational framework for data analytics follows a centralized model in which a central team provides support and implementation of data pipelines and reports. Moving from a centralized to a decentralized data analytics framework necessitates a cultural shift, hiring new technical resources, and redefining roles and responsibilities. Further, adopting decentral approaches also raises the question of how to cut domains and how large a domain is, hence how many data products they are responsible for. Some business lines embrace these changes, while others show hesitation, often driven by the varying degrees of technical competencies within the business lines. A critical remark from within the organization highlights this tension, indicating that the redistribution of accountability is one of the most significant hurdles in this transition.

"People have to take responsibility, and others have to give up responsibility, which is always a difficult process." – Data Engineer

Complicating the transformation is the clash between the incumbent ERP-first IT culture and the new best-of-breed employed for the technical realization of Data Mesh, i.e., the data platform development, which heavily relies on open-source software components. Maintaining organizational independence

caters to one of Alpha's core values. However, the introduction of a best-of-breed ecosystem introduces risks, especially regarding the integration and interoperability with the existing ERP ecosystem. Moreover, the divergence in culture and operational models within these ecosystems is pertinent. The ERP ecosystem, distinguished by a centralized global team, is underpinned by a strong impetus to uphold standardized ERP processes across various business lines. This divergence in IT culture underscores the need for a careful balance between innovation and the maintenance of reliable, standardized processes. Addressing the complexity of the Data Mesh framework is crucial for its successful adoption. Upskilling employees and providing additional technical staff and resources are necessary to build and manage data products effectively.

Furthermore, a use case and self-service-driven rollout of the Data Mesh approach is being tailored to meet various requirements, such as real-time versus batch processing and differing data granularity needs. Alpha's requirements regarding a data platform originate from different use cases. These span from internal process innovation, providing data to external stakeholders, such as customers, vendors, or partners, and the training of AI solutions. To balance the transformation approach based on distributed use cases, Alpha also emphasized centralized governance efforts, mainly by introducing foundational processes and technologies that address the Data Mesh principles of federated computational governance and self-service data platform. Alpha's journey to a federated data management paradigm is marked by a concerted effort to align the operating model with organizational legacy, navigate IT ecosystems, and leverage the use-case-driven approaches to drive innovation while ensuring robust governance and support from top management.

5 Discussion

In our study, we examined Alpha's strategic shift toward a Data Mesh architecture, sparked by pressure on the business model through digital innovation and aiming to surmount the obstacles of data fragmentation and boost data utility across the enterprise. This evolution, fueled by a strategy valuing digital and data products, aims to augment physical product development with digital service creation. Initially identifying the catalysts for adopting Data Mesh, the study reveals how its principles address these challenges, leading to new obstacles navigated via established best practices. Accompanying the architectural shift is a cultural repositioning, transitioning IT's role from a cost center to a pivotal player in innovation and data-driven strategies. These initiatives encapsulate the organizational transformation shaped by legacy, operating model suitability, IT culture conflicts, and the balancing act between decentralized autonomy and centralized governance. This comprehensive strategy addresses immediate integration issues and cultivates a long-term shift towards a data-centric organizational ethos.

5.1 Data Mesh as federated data architecture

Data are not just tools for management but have become crucial mediums through which organizations understand and respond to their environments. Digital data are now integral to a broad range of operations, shifting from a supportive element to a core resource (Alaimo and Kallinikos, 2022). These theoretical perspectives align with the practical reality of Alpha's Data Mesh transformation, which seeks to shift from monolithic data management to a more integrated and operational use of data across domains and the product thinking approach, emphasizing the significance of data in adding value and driving organizational activities. Our results show that Data Mesh is a suitable approach to enable data-driven innovation at industrial organizations because it aligns with their operating models, preventing the misalignment prevalent with centralized data architectures (Janssen et al., 2020).

The technological advancements through AI have become relevant to all types of firms, from native software companies to manufacturers such as Alpha. As these advancements, such as generative AI, have also become faster and more easily accessible to organizations, the technological frontier has moved closer. Thus, the decentralized nature of Data Mesh also echoes Acemoglu et al.'s (2007) assertion that firms closer to the technological frontier are more likely to decentralize, given the agility and innovation such proximity necessitates. This notion resonates with the concept of federated IT governance, where federal setups are beneficial in changing environments (Williams and Karahanna,

2013). Data Mesh's federated nature allows Alpha more flexibility and context-specific solutions to address its environment's increasing variability and complexity. Firms operating in such environments benefit from the federated data architecture approach as it enables different domains or teams to manage their data independently, facilitating learning and adaptation to their specific circumstances. By enabling domain-specific data management, firms can foster rapid transformation and learning within varied operational contexts, aligning with insights into the decentralization of knowledge and data within firms (Alaimo and Kallinikos, 2022).

Further, Data Mesh addresses the trade-off between the information advantage of localized, decentralized teams and the potential loss of centralized control (Acemoglu et al., 2007). Similar to federated IT Governance, decisions close to a specific asset, e.g., regarding quality or access, are decentralized, whereas decisions concerning interoperability are driven by central teams, e.g., infrastructure, policies, standards (Fonstad and Subramani, 2009). The federal decision-making context at Alpha is focused on digital innovation rather than standardization, aligning with literature, which highlights the suitability of federal setups in the context of change (Williams and Karahanna, 2013). Alpha's Data Mesh implementation achieves this by empowering domain teams with the authority to act on their contextual knowledge while still adhering to a global framework of interoperability and governance. The setup of centralized measures, such as the data manifesto or the self-service data platform, and decentralized measures, such as domain splitting and use-case-driven data products, allows for effectively managing the decentralization of information and control within its organization.

5.2 Learnings for fostering the data-centric transformation of incumbents

Moving towards Data Mesh in organizations yields technical, organizational, and cultural challenges. Based on our findings and existing literature on data management, we derived learnings for establishing and adopting a federated data management approach in incumbent industrial organizations. The first three learnings address technical challenges. *Learning 1: Establish data lineage* – A strong and reliable data lineage that considers both business and technical elements can meet the data management needs of Data Mesh, improve data quality, and reduce technical debt. This aligns with our insights at Alpha and academic research (Joshi et al., 2021), highlighting its vital role in establishing a solid foundation for a top-quality data product inventory. *Learning 2: Implement a data catalog* – An established data catalog becomes crucial in enabling Data Mesh by enabling thorough data exploration for users (Labadie et al., 2020). The focus is on producing top-notch data items with conscientiously preserved metadata to entice and retain initial adopters while averting losses due to inferior data quality. *Learning 3: Automated checks and quality scores* – Implementing automatic checks and quality assessments is a promising route to guarantee data quality and helped Alpha achieve levels of data quality.

Driving organizational change, three learnings have emerged as helpful for addressing challenges in our research. *Learning 4: Work use-case driven* – An incremental and iterative approach presents an opportunity to direct organizational transformation, emphasizing prioritized and specific use cases. Emphasizing quick wins to attain and retain top management approval helped Alpha propel the embrace of Data Mesh across the organization. *Learning 5: Create a dedicated data governance team* – Establishing a specialized data governance team to uphold data principles and help balance the decentralized structure and diverse personnel displayed potential for a fruitful execution at Alpha. *Learning 6: Introduce monetization for data products* – Incentivizing the data productization process proved crucial for the data-centric transformation, and incorporating data monetization and cost transparency provides a value-driven perspective on data products (Wixom et al., 2021).

Finally, concerning the cultural aspect of the transformation, we developed three learnings from Alpha's case that facilitate a successful change. *Learning 7: Develop a communication strategy* – Achieving a cultural shift towards a data-driven mindset in an organization requires a communication strategy to express the vision. *Learning 8: Establish communities of practices* – Due to the decentralized and socio-technical character of the approach, it is vital to improve human communication and interaction. Given the innovative nature of the approach and its proximity to emerging technologies such as AI, it is crucial to establish communities of practice for various roles and topics to enable effective knowledge exchange

(Lefebvre and Legner, 2022). *Learning 9: Investments into IT skills* – It is essential to recognize that the decentralized approach requires significant investments in skills development to empower employees to adapt to the new methodologies.

The learnings show that implementing the Data Mesh paradigm is a multifaceted transformation involving technical, organizational, and cultural changes. The interplay between establishing data lineage (L1) and implementing a data catalog (L2) is foundational to this approach. Both are necessary for producing and maintaining high-quality data products, which are the currency of a Data Mesh, and are prerequisites for automated quality checks (L3) that maintain the integrity and reliability of the data ecosystem. These technical strategies underpin the organizational shifts towards a use-case driven approach (L4), the creation of a dedicated data governance team (L5), and the monetization of data products (L6) to highlight the tangible value of data. The cultural shift, in turn, is facilitated by a strategic communication plan (L7), the establishment of communities of practice (L8), and significant investments in IT skills (L9), ensuring that the workforce is prepared for and engaged in the data-centric transformation. Together, these learnings create a symbiotic framework where technical advancements bolster organizational and cultural evolution towards a successful Data Mesh implementation.

5.3 Implications

The study's findings on Alpha's adoption of Data Mesh draw theoretical and practical implications. From a theoretical perspective, we extend existing research on data management (Abraham et al., 2019) and the nascent research stream on Data Mesh (Machado et al., 2022) by providing rich empirical insights into the data architecture transformation of an incumbent industrial organization and developing a process perspective. We further show how Data Mesh overcomes the misalignment between the logical flow of data and the organizational structure by adjusting to the increasing decentralization of data generation and usage (Janssen et al., 2020). By developing structural links between the initial challenges and the Data Mesh implementation process, we develop insights into how organizations can leverage their increasing amounts of data to enable data-driven innovations like AI (Gröger, 2021). Lastly, we discuss how Data Mesh mirrors existing theories of federated IT governance to balance decentralization and centralization in data governance (Acemoglu et al., 2007; Williams and Karahanna, 2013).

From a practical perspective, we show how successful implementation in an industrial setting like Alpha involves a blend of technical strategies, such as developing a data catalog and implementing data lineage, organizational measures like empowering organizational change, and cultural measures focusing on people through robust communication and skill development. Younger firms, which may have yet to establish processes or legacy systems and are less constrained by entrenched central systems, can build their data infrastructure around decentralized principles from the outset. However, our results show that incumbents can also use Data Mesh to drive the transformation toward a data-centric organization. To aid practitioners in this transformation, we developed a set of learnings.

5.4 Limitations and future research

While providing depth and a starting point for future research in an area where limited theoretical insights exist (Siggelkow, 2007), the single-case study design on Alpha limits the generalizability of the findings. The unique position of Alpha as an early Data Mesh adopter and its specific organizational context means that the results may only partially apply to different corporate settings or industries. Focusing on one organization limits exposure to various Data Mesh implementation strategies and outcomes. Consequently, the complexity of balancing the decentralization of information and control within organizations calls for a nuanced understanding of the federated socio-technical data management that Data Mesh represents, necessitating further research into its practical implementation and governance within various organizational contexts. Multiple-case studies across industries could reveal additional factors influencing the successful adoption of Data Mesh, especially in different organizational cultures and structures. Longitudinal studies post-implementation would provide insights into the enduring impact of Data Mesh on organizational agility, data democratization, and innovation. Additionally, investigating relationships between technical implementation decisions and success in

promoting a data-centric organization through variance-based approaches could yield insights into the intersection of technology and management and extend our process perspective (Burton-Jones et al., 2015).

6 Conclusion

The study of Alpha's transition to a Data Mesh architecture provides a granular view of the complexities of shifting to a data-centric organizational model. Despite the cultural, organizational, and technological challenges, the case reveals how adopting a federated socio-technical data management approach can significantly enhance data utility and governance. Alpha's journey underscores the importance of balancing decentralized domain autonomy and centralized control in managing data assets. The results of our case study suggest that with the right strategies, even established industrial firms can successfully navigate digital transformation, leveraging Data Mesh to drive innovation and maintain competitiveness in an increasingly data-driven world.

References

- Abraham, R., J. Schneider and J. Vom Brocke (2019). "Data governance: A conceptual framework, structured review, and research agenda." *International Journal of Information Management* 49, 424-438.
- Acemoglu, D., P. Aghion, C. Lelarge, J. Van Reenen and F. Zilibotti (2007). "Technology, Information, and the Decentralization of the Firm." *The Quarterly Journal of Economics* 122 (4), 1759-1799.
- Alaimo, C. and J. Kallinikos (2022). "Organizations Decentered: Data Objects, Technology and Knowledge." *Organization Science* 33 (1), 19-37.
- Armbrust, M., A. Ghodsi, R. Xin and M. Zaharia (2021). "Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics." *Proceedings of CIDR* 8.
- Benfeldt, O., J. S. Persson and S. Madsen (2020). "Data Governance as a Collective Action Problem." *Information Systems Frontiers* 22 (2), 299-313.
- Bode, J., N. Kühl, D. Kreuzberger and S. Hirschl (2023). "Data Mesh: Motivational Factors, Challenges, and Best Practices." *arXiv preprint arXiv:2302.01713*.
- Bojinov, I. (2023). "Keep Your AI Projects on Track." *Harvard Business Review* (11/2023).
- Burton-Jones, A., E. R. McLean and E. Monod (2015). "Theoretical perspectives in IS research: from variance and process to conceptual latitude and conceptual fit." *European journal of information systems* 24 (6), 664-679.
- Corbin, J. M. and A. Strauss (1990). "Grounded theory research: Procedures, canons, and evaluative criteria." *Qualitative sociology* 13 (1), 3-21.
- Dehghani, Z. (2019). *How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh*. URL: <https://martinfowler.com/articles/data-monolith-to-mesh.html> (visited on 2023-09-21).
- Dehghani, Z. (2022). *Data Mesh*. 1st Edition.: O'Reilly Media.
- Dehghani, Z. and M. Fowler (2022). *Data mesh: delivering data-driven value at Scale*. 1st Edition. Sebastopol, CA: O'Reilly Media.
- Desai, M. A. (2023). "What the Finance Industry Tells Us About the Future of AI." *Harvard Business Review*.
- Eisenhardt, K. M. (1989). "Building theories from case study research." *Academy of management review* 14 (4), 532-550.
- Falconi, M. and P. Plebani (2023). "Adopting Data Mesh principles to Boost Data Sharing for Clinical Trials." *2023 IEEE International Conference on Digital Health (ICDH)*, 298-306.
- Fonstad, N. O. and M. Subramani (2009). "Building enterprise alignment: A case study." *MIS Quarterly Executive* 8 (1), 31-41.
- Gardner, S. R. (1998). "Building the data warehouse." *Communications of the ACM* 41 (9), 52-60.
- Gioia, D. A., K. G. Corley and A. L. Hamilton (2013). "Seeking Qualitative Rigor in Inductive Research." *Organizational Research Methods* 16 (1), 15-31.

- Goedegebuure, A., I. Kumara, S. Driessen, D. Di Nucci, G. Monsieur, W.-j. v. d. Heuvel and D. A. Tamburri (2023). "Data Mesh: a Systematic Gray Literature Review."
- Gröger, C. (2021). "There is no AI without data." *Communications of the ACM* 64 (11), 98–108.
- Hasan, M. R. and C. Legner (2023). "Understanding Data Products: Motivations, Definition, and Categories." *European Conference on Information Systems*.
- Hooshmand, Y., J. Resch, P. Wischniewski and P. Patil (2022). "From a Monolithic PLM Landscape to a Federated Domain and Data Mesh." *Proceedings of the Design Society* 2, 713-722.
- Inmon, W., M. Levins and R. Srivastava (2021). *Building the Data Lakehouse*. 2nd Edition. Basking Ridge, NJ: Technics Publications.
- Janssen, M., P. Brous, E. Estevez, L. S. Barbosa and T. Janowski (2020). "Data governance: Organizing data for trustworthy Artificial Intelligence." *Government Information Quarterly* 37 (3), 101493.
- Joshi, D., S. Pratik and M. P. Rao (2021). "Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study." *ICEB '21*.
- Khatri, V. and C. V. Brown (2010). "Designing data governance." *Communications of the ACM* 53 (1), 148-152.
- Labadie, C., C. Legner, M. Eurich and M. Fadler (2020). "FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs." *2020 IEEE 22nd Conference on Business Informatics (CBI)*.
- Lefebvre, H. and C. Legner (2022). "How communities of practice enable data democratization inside the enterprise." *European Conference on Information Systems*.
- Lefebvre, H., C. Legner and M. Fadler (2021). "Data democratization: toward a deeper understanding." *International Conference on Information Systems*
- Machado, I. A., C. Costa and M. Y. Santos (2021). "Data-Driven Information Systems: The Data Mesh Paradigm Shift." *Information Systems Development: Crossing Boundaries between Development and Operations (DevOps) in Information Systems (ISD2021 Proceedings)*.
- Machado, I. A., C. Costa and M. Y. Santos (2022). "Data mesh: concepts and principles of a paradigm shift in data architectures." *Procedia Computer Science* 196, 263-271.
- Miloslavskaya, N. and A. Tolstoy (2016). "Big data, fast data and data lake concepts." *Procedia Computer Science* 88, 300-305.
- Mohr, N. and H. Hürtgen. (2018). *Achieving business impact with data* McKinsey & Company. URL: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/achieving-business-impact-with-data> (visited on 2023-07-28).
- Nargesian, F., E. Zhu, R. J. Miller, K. Q. Pu and P. C. Arocena (2019). "Data lake management: challenges and opportunities." *Proceedings of the VLDB Endowment* 12 (12), 1986-1989.
- Otto, B. (2011). "Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers." *Communications of the AIS* 29, 45-66.
- Pörtner, L., R. Möske and A. Riel (2023). "Data Ecosystem for Industrial Product- Service Systems (IPS2) Based on a Decentralized Data Architecture." *Procedia CIRP* 119, 1228-1233.
- PwC (2020). "Why Most Organizations' Investments in AI Fall Flat." *Harvard Business Review*.
- Siggelkow, N. (2007). "Persuasion With Case Studies." *Academy of Management Journal* 50 (1), 20-24.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos and P. E. Bourne (2016). "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (1), 1-9.
- Williams, C. K. and E. Karahanna (2013). "Causal Explanation in the Coordinating Process: A Critical Realist Case Study of Federated IT Governance Structures." *MIS Quarterly* 37 (3), 933-964.
- Wixom, B. H., G. Piccoli and J. Rodriguez (2021). "Fast-Track Data Monetization With Strategic Data Assets." *MIT Sloan Management Review* 62 (4), 1-4.
- Yin, R. K. (2018). *Case study research and applications: Design and methods*. 6th Edition. Los Angeles; London; New Dehli; Singapore; Washington DC; Melbourne: SAGE.
- Zhang, Q., X. Sun and M. Zhang (2022). "Data Matters: A Strategic Action Framework for Data Governance." *Information & Management* 59 (4), 103642.