Issues in Informing Science + Information Technology

An Official Publication of the Informing Science Institute
InformingScience.org

IISIT.org

Volume 22, 2025

EXPLORING DATA MESH ADOPTION IN LARGE ORGANIZATIONS

Robert Winter*	University of St.Gallen, St.Gallen, Switzerland	Robert.Winter@unisg.ch
Tobias Hackl	University of St.Gallen, St.Gallen, Switzerland	Tobias.Hackl@unisg.ch

* Corresponding author

ABSTRACT

Aim/Purpose	This paper explores how large incumbent organizations adopt the newly proposed data management approach "Data Mesh". Particularly, this paper explores to which extent data ownership and data governance are shifting from a centralized to a decentralized approach and whether companies take different paths in this transition.
Background	Large incumbent organizations suffer from high complexity and often centralized infrastructure of data management that lead not only to high operating costs but also slow down innovation projects and make them more expensive. As a managerial complement of decentralized data management technology such as Data Fabric, Data Mesh was introduced as a management approach to address the organizational challenges of centralization and complexity.
Methodology	We studied how ten large Swiss incumbent organizations adapted Data Mesh. We interviewed positions such as chief information officers, chief data officers, head of information management, head of group IT. We developed a conceptual framework to position their data management approaches and how much they decentralized data governance and data ownership.
Findings and Contribution	All large incumbent companies adopt Data Mesh principles, but in different ways and to a different extent. While data governance continues to be increasingly centralized, the degree of data ownership centralization mostly remains unchanged, according to the degree of regulation and other contextual factors.
Recommendations for Practitioners	Data Mesh is regarded as beneficial by many organizations and thus a relevant option for corporate data management. But there is no "silver bullet" adoption process – instead the degree and way of adoption depends on a certain number of contextual factors, and several reference adoption models exist.

Editor: Eli Cohen | Received: January 25, 2025 | Revised: March 27, April 12, 2025 | Accepted: April 25, 2025

Cite as: Winter, R., & Hackl, T. (2025). Exploring data mesh adoption in large organizations. *Issues in Informing Science and Information Technology, 22,* Article 12. https://doi.org/10.28945/5509

Recommendations for Researchers and Future Research	The study proposes several hypotheses on Data Mesh adoption that need to be validated in other geographies, other types of organizations (non-incumbents, government, small companies) to better understand adoption paths, adoption barriers, and adoption economies.
Keywords	data management, data mesh, data governance, data ownership, data architecture

INTRODUCTION

In today's digital economy, organizations have access to unprecedented amounts of data, presenting unique challenges and opportunities in data management (DalleMule & Davenport, 2017). As data volumes increase, the need for scalable data management systems becomes critical, compelling organizations to optimize their data processing and provisioning for analytical purposes at an enterprise level (Dehghani, 2022). To address these challenges, different conceptual approaches have emerged, including Data Mesh and Data Fabric (Blohm et al., 2024). While Data Fabric represents a primarily technical approach, focusing on automated data integration through metadata analysis (Blohm et al., 2024), Data Mesh takes a sociotechnical perspective, decentralizing data ownership by assigning responsibility to individual business domains while enforcing federated governance via a central platform (Dehghani, 2022).

This study focuses on Data Mesh due to its growing practical relevance in combination with scarce academic research: extant literature provides contributions proposing conceptual frameworks, domain models, conceptual architectures, and explanations for its emergence (Araújo Machado et al., 2022). To a lesser extent, we see first empirical case studies on implementing Data Mesh in comparably "new" organizations such as Netflix (Cunningham, 2020) or Zalando (Schultze & Wider, 2020), as well as in the banking sector (Joshi et al., 2021). Although we find an increasing adoption in practice, Data Mesh adoption has not yet been comparatively investigated in large incumbent organizations.

It is no surprise that Data Mesh is often perceived as a catalyst for the next generation of data management - as most of the aforementioned challenges of current corporate data management practices can be addressed – in theory. To understand whether and how the concept is implemented in a real-world context, we conducted exploratory interviews in ten large incumbent companies on their current and planned approach to corporate data management. We focus on centralized and decentralized aspects of enterprise-wide data management - as Data Mesh poses a decentralized approach to data management that opposes most current (rather central) aspects to data management, in particular in financial services. We set out to answer the following two research questions:

(1) How do large, data-intensive incumbent companies combine and compose centralized and decentralized aspects of enterprise-wide data management?

(2) How has Data Mesh been adopted at scale to facilitate decentralization in enterprise-wide data management?

In Section 2 we present conceptual foundations such as pre-Data Mesh data management architecture and governance approaches, before we draw on interview data (Section 3) and finally present our results and discuss how large, complex organizations could evolve their data management towards Data Mesh.

CONCEPTUAL FOUNDATIONS

For this study, we assume an 'architectural' lens of analysis, with architecture being understood as "the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution" (ISO/IEC/IEEE,

2011). As a consequence, we differentiate between fundamental data management components and relationships on the one side and principles governing data management's design and evolution on the other.

Reviewing extant literature to identify representative architectural components in the context of social systems resulted in the identification of components such as decision rights (Khatri & Brown, 2010; Otto, 2011) and decision areas (Weber et al., 2009; Wende, 2007), accountabilities (Abraham et al., 2019; Khatri & Brown, 2010; Weber et al., 2009; Wende, 2007), guidelines, standards and rules (Weber et al., 2009; Wende, 2007).

Several of those components are also substantial elements of data governance as defined by (Otto, 2011): "A companywide framework for assigning decision-related rights and duties in order to be able to adequately handle data as a company asset" (p. 47). They also constitute substantial elements of data ownership which clarifies fundamental rights and responsibilities (Winter & Meyer, 2001). Hence, the fundamental difference between data governance and data ownership is that through data governance fundamental rights are assigned, and data ownership clarifies who is responsible for executing guidelines, standards, and rules (Winter & Meyer, 2001).

Each of these components can be implemented on a continuous scale from central to decentral, which eventually jointly reflects a company's approach to data governance and data ownership based on this scale. In extant literature we find three archetypes of organizing data management: centralized, decentralized (Wende & Otto, 2007) and federated (Weber et al., 2009). An archetypical centralized approach is determined by a central manifestation of data governance and a central manifestation of a data ownership approach which means, that a central IT or data team assigns fundamental rights to itself and is also responsible for executing guidelines, standards, and rules (Wende & Otto, 2007). Vice versa, in an archetypical decentralized approach, each decentral domain (also referred to as a business unit) sets up their own data governance and is responsible for execution (data ownership) (Wende & Otto, 2007). In a federated approach, data governance assigns decision rights from a central IT or data team and decentralized domains are responsible for executing set guidelines, standards, and rules (Weber et al., 2009).

According to the definition of enterprise-wide information logistics (Winter, 2008), the technical system refers to the storage and provision of cross-unit data. Therefore, most common data storage and provision approaches from literature fall into the scope of study, namely data warehouse, data lake, on-premise, cloud, and data platforms (Nambiar & Mundra, 2022). With regard to their tendency towards centrality and decentrality, data warehouses and data lakes can be considered as opposing approaches; similarly, on-premise and cloud approaches are contrary approaches (Nambiar & Mundra, 2022). A data warehouse is a centralized approach of storing data to generate a centralized and unified perspective on enterprise data (Kerschberg, 2001). A data warehouse as a central data repository (Venkatesh et al., 2007) requires the integration of data from various sources that comes with high initial investment cost (Inmon, 2000). A rather decentralized, opposing approach to data storage is a data lake, which was initially developed because data warehouses were assessed to be incapable of managing unstructured data (e.g., big data) (DalleMule & Davenport, 2017). A data lake can store virtually unlimited amounts of both, structured and unstructured data (DalleMule & Davenport, 2017). Despite a data lake also referring to a central data repository where data from several sources can be integrated, it is perceived as more decentral as it enables deploying data from users across the organization (Hagen & Hess, 2020) due to open access (Giebler et al., 2019; Hai et al., 2021). A data lake can be a low-cost approach in terms of integrating data from various users, but an open policy of data integration results in low quality data in various formats which increases costs for search and data transformation before usage (Hai et al., 2021). In addition to these extremes, several data storage approaches have been proposed as a compromise, such as federated data warehouses (Kerschberg, 2001) or a Data Lakehouse (Oreščanin & Hlupić, 2021).

An additional architectural aspect of technical systems is whether a component is running on-premise or cloud-based, respectively. Both environments build the basis for hosting software applications (Mell & Grance, 2011), including, for example, a data warehouse (Nambiar & Mundra, 2022). However, the underlying cost models are completely different: While on-premise requires high initial investments (Nambiar & Mundra, 2022), cloud computing is defined as a "model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell & Grance, 2011, p.2), usually following a pay-per-use cost logic (Lowe & Galhotra, 2018). Generally, cloud storage or processing can support a more decentralized approach compared to on-premise storage due to the benefit of scalable provisioning of computing resources (Mell & Grance, 2011).

As a further essential architectural component of a technical system for providing data on an enterprise level, we touch upon some form of a data platform as digital platforms in general have the ability to "facilitate the interaction and collaboration of different actors through highly-efficient resource matching" (Staub et al., 2021, p.6163). In addition, to enable data sharing and data providing, data platforms can be extended to enterprise analytics platforms (Fadler & Legner, 2020). However, due to their recent emergence, not much descriptive or prescriptive scientific knowledge about data platforms is currently available.

CORPORATE DATA MANAGEMENT BEFORE DATA MESH

In this section, we shed light on the evolution of data architecture approaches from a management perspective to better understand where we come from and why Data Mesh could be a potential next step in enterprise-wide data management. We begin our chronological overview in the 1980s with the concept of the data warehouse including its add-ons such as data marts and operational data stores (ODS). Thereafter, we elucidate how big data affected the data landscape and introduce resulting architectural approaches such as data lakes and data platforms.

The first phase of "information processing" in organizations focused on the automation of business transactions. In the 1980s, such "transaction systems" were extended by certain analytical functionalities, but still based on (siloed) transaction data (Jung & Winter, 2000). The increasing need to integrate data to create a subject (and not transaction) centric informational view, and the missing integration capabilities and performance of transaction systems to provide integrated, subject-oriented data, led to establishing a new information system component to bypass the transaction system bottleneck. New software systems designated as data warehouses, an enterprise-wide "single version of truth" was created where siloed transaction data were aggregated, integrated, and preserved over longer periods to enable subject-oriented data analytics. Given the example of a customer as a subject, data warehousing allows the tracking of a customer's purchases not only across channels, but also over time. The fundamental characteristics of data warehouse data collection as stated by Inmon (1996) are:

- Subject oriented: in contrast to the transactional view of data, a subject oriented view aggregates all transactions related to an information subject (for example: sales revenue for a customer over a longer period of time, across all channels and across all products).
- Integrated: as opposed to data from individual siloed information systems, data must be integrated from various sources to a central data warehouse to ensure a consistent holistic perspective on data and avoid multiple versions of truth and costly inconsistent data duplicates.
- Time-oriented: in a data warehouse, not only current transactions are covered, but data is stored that refer to a longer time period, often spanning many years. This enables longitudinal data analysis.

- Non-volatile: once data are integrated and cleansed, data is stored unchanged to preserve consistency between analyses at different points in time.

In theory, the "single version of truth" could be realized best by creating one single large data warehouse. However, large organizations faced the two main challenges and had to establish multiple integrated data warehouses. First, there was simply too much data for a single data warehouse which resulted in performance issues, and secondly, a high level of complexity in particular for integrating data from various sources via ETL processes led to data warehouses associated to individual functional divisions and eventually to a certain degree of "silo-ization." As functional divisions had ownership of data warehouses, they individually refined their data management by implementing various individual systems and tools, eventually leading to an increased company-wide level of data architecture complexity.

During the 1990s, further infrastructure components were added to the corporate data management landscape: to enable customer-oriented "near-real-time" applications such as a customer call center, where data of a specific customer must be not only accessible during the call, but also updated, so-called operational data stores (ODS) were developed (Jung & Winter, 2000). Smaller volumes of specific data are derived directly from transaction systems bypassing the complex data warehouse integration mechanisms, and data could also flow from the ODS back into transactional systems (Jung & Winter, 2000).

An additional component of a data architecture are data marts which can be described as an excerpt of the data warehouse solely containing data tailored for specific uses (Jung & Winter, 2000). In data warehouses, data exist in a normalized form to allow all possible queries and updates while preserving data consistency. In contrast, data marts could also contain data in denormalized or aggregated form, as user accessing a specific data mart show homogenous requirements on data (Jung & Winter, 2000). The data mart layer simplifies analytical exploitation of data, but also entails additional complexity of the corporate data management architecture.

However, with the advent of big data, enterprise-wide data management had to cope with torrents of data characterized by volume, variety, and velocity (McAfee & Brynjolfsson, 2012) as well as additional characteristics such as veracity, validity, volatility, and value (Khan et al., 2014). Based on individual data (warehouse) ownership, business divisions implemented additional tools and systems to handle (in particular) unstructured data, which consequentially resulted in an even more fragmented data architecture landscape.

When facing a continuously dispersing data architecture landscape containing various tools and systems, organizations developed the concept of a data lake in the early 2010s. Unlike data warehouses, data lakes can store a range of different data formats, including unstructured data. Moreover, data is uploaded in data lakes in raw format and not prepared "ready to use" on loading (like in a data warehouse). Rather, it is prepared for analysis when used (Nambiar & Mundra, 2022). Hence, a further advantage of a data lake over a data warehouse is that only data that is transformed, is actually used. However, uploading data in raw format into data lakes often results in low data quality and low discoverability by users (Nambiar & Mundra, 2022). In most large organizations, data lakes did not replace data warehouses, especially for structured data and regular, rather uniform use (e.g., for standard reporting). In these contexts, the concept of data warehouse appears to be superior to the "prepare on demand" principle of data lakes.

The rise of digital platforms (e.g., Amazon, Airbnb, Apple App Store), agile principles (e.g., Spotify's organizational model), and the concept of perceiving data as a product (rather than as a resource or asset) drove new developments in corporate data management. Digital platform models in data management emerged around 2015. While conceptualized differently, they offer advantages such as reusability through attachable modules (de Reuver et al., 2018; Yoo et al., 2010), reducing resource costs (Schmid et al., 2021). Digital platforms also lower transaction and search costs compared to traditional organizational approaches (Schmid et al., 2021) and foster innovation by leveraging participant

capabilities (Yoo et al., 2010). However, digital platforms pose challenges, including high initial investments and added complexity to corporate data architectures. This complexity shifts to a higher layer—for example, simplifying ETL processes among data warehouses while requiring only one integration into a central platform. Recent literature contributions provide first insights that data management platforms, as a subclass of digital platforms, yield the same benefits (e.g., architectural leverages) (Hackl et al., 2024). Since digital and data management platforms allow for diverse conceptualizations and governance choices based on yet unclear contingencies, more scientific research is needed to better understand potentials and challenges.

In review of the last approximately 40 years of data management approaches, we contend that, despite constant developments that predominantly eradicate drawbacks of existing approaches, corporate data management still appears to have much room for improvement. This is due to the fact that various approaches to data architecture exist simultaneously in diverse configurations and stages of maturity rather than fully replacing each other, finally leading to heterogeneous data and IT landscapes. The co-existence of various approaches and architectural components leads to an increased level of complexity. Retaining smooth interaction between system components of such complex data architectures comes at high maintenance cost as well as long and expensive change projects. Hence, effectiveness and efficiency in corporate data management is an issue for most organizations, especially large and mature 'incumbent' organizations.

For its potential to overcome some of the posed challenges, the following section introduces and discusses the concept of Data Mesh.

UNDERSTANDING DATA MESH

The term Data Mesh was first coined by Zhamak Dehghani (2019) and has continuously evolved due to the growing interest among data managers, data strategists, and enterprise architects. Dehghani (2019) characterizes Data Mesh as a decentralized "sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments – within or across organizations" (p. 57).

In line with traditional data management, data is divided into two main categories (Dehghani, 2022):

(1) operational data which is stored and processed in "databases that support running the business and keeping the current state of the business" (p. 16) and

(2) analytical data which is stored and processed in a "data warehouse or lake providing a historical, integrated and aggregate view of data created as a byproduct of running the business" (p. 16).

Operational data is cleansed, integrated and historized to create analytical data. Analytical data can be, among other uses, deployed to train machine learning (ML) models which in turn feed into the operational systems as intelligent services. Data Mesh is an approach to restructure an organization's analytical data. The main objective (Dehghani, 2022) of Data Mesh is to decentralize certain aspects of corporate data management for improving value creation from analytical data.

For the frictionless realization and implementation of a decentralized data architecture approach, Dehghani (2022) poses four foundational principles which are collectively necessary and sufficient to substantiate Data Mesh's logical architecture and applicable operating model. These so-called principles are explained in the following by providing a definition and underlying motivation.

PRINCIPLE 1: DOMAIN-ORIENTED OWNERSHIP

By decentralizing the ownership of analytical data to domains who are closest to their data in terms of being the data's main producer or user, domains should be responsible to manage the entire lifecycle of their data artefacts (data, code, metadata and policies) autonomously.

By doing so, scalability can be ensured as an organization grows with increasing data sharing and consumption. By decentralizing data ownership, an organization can simultaneously increase agility by minimizing cross-domain synchronizations and the opportunity for continuous change. Following this principle will also enhance data business trustworthiness, as the data is shared from the producer and main consumer itself. Lastly, by eliminating data pipelines, the resilience of analytics and ML products are expected to rise.

Principle 2: Data as a Product

This principle holds business domains responsible for sharing their data as a product with other data users e.g., data analysts or data scientists from other domains. Data as a product has the following usability characteristics: ensuring a certain level of data quality measured by accurately stated key performance indicators (KPIs); ensuring that the data product is easy to understand and use, for example, by providing easy access through various methods and tools; and making sure data products are interoperable and joinable to other domains' data products.

A data product is a unit of logical architecture which comprises and controls several structural components such as data, code, policies, and infrastructure dependencies with the overall goal to share data as a product in a frictionless and autonomous manner.

By complying with the principle of data as a product, organizations should be enabled to create high quality and trustworthy data which can be shared and used across domains (getting rid of domain-oriented data silos). When being able to access and share data, it is expected that organizations are enabled to create a data-driven innovation culture. Data as a product enables autonomy within domains. Moreover, clear contracts between the domains' data products ensure that adjustments of a data product will not interfere with data products from other domains.

Principle 3: Self-Serve Data Platform

Organizations must develop at least one self-serve data platform where domain-oriented teams are able to manage their Data Mesh of interconnected data products over an end-to-end life cycle. Within this platform, domain-oriented teams are also enabled to share their knowledge graph and lineage to enrich user experience for discovering, accessing, and using data products.

Hence, a domain's total cost of ownership of data and complexity management of data products along the life cycle is reduced. Consequently, by freeing (data) product owners from architectural responsibilities, organizations can deploy (more) generalist experts for data product development. However, developing such a self-serve data platform comes with certain initial investments and on an enterprise level, financial trade-offs between initial investment and reduced operating costs must be assessed. Technically, this self-serve data platform should be able to automatically ensure compliance policies are adhered to in a timely manner to facilitate the discovery, accessibility, building, and deployment of data products.

Principle 4: Federated Computational Governance

Federated computational governance describes a data governance operational model relying on a decentralized decision making and accountability structure consisting of domains gathering around the central data platform. Hence, a federated governance leaves domains enough autonomy to operate in an agile manner. However, simultaneously, domains must also respect global conformance to ensure their data products are interoperable and secure within the Data Mesh. As governance is executed automatically, policies can be established for each data product on a detailed level.

Following this governance, value from aggregating independent data products can be generated by compatible and connected data products. Enforcing governance requirements (e.g., security, privacy, legal compliance, etc.) automated via a data platform, organizational overhead can be minimized.

Figure 1 illustrates the relationship among the four foundational principles and how they complement each other. Data as a product prevents data siloing of domain ownership by demanding data

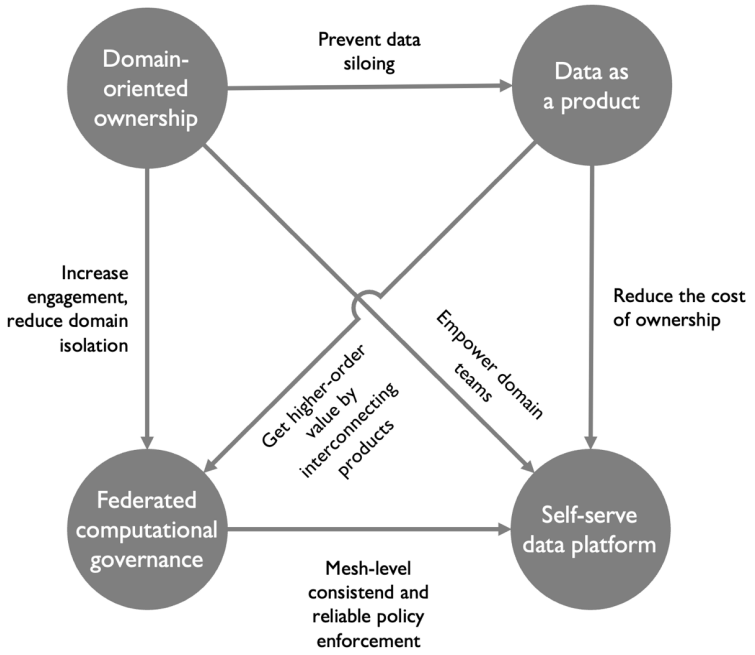

Figure 1. Four principles of Data Mesh and how one requires another (based on Dehghani, 2022, p.62)

product owners actively share and distribute their data product and thus foster synergies. The self-serve data platform empowers domain teams as they can autonomously exploit the data platform according to their needs. Moreover, the self-serve data platform reduces the cost of ownership of data as a product by offering a single shared infrastructure to consistently manage data products along the entire life cycle according to policies through federated computational governance. Federated computational governance connects and coordinates domain-oriented ownership and ensures interoperability of data as a product.

By conflating the four foundational principles of Data Mesh, the aspect of scalability is touched upon as follows: organizations pursue to scale their enterprise-wide data management approach to be able to include torrents of (big) data in various formats from various sources in their analysis to reveal novel patterns and eventually create value from data (Dehghani, 2022). For this purpose, decentralizing data ownership relegating responsibility to domains releases a central data team from lengthy data engineering jobs (e.g., ensuring data quality) and frees them for more valuable jobs requiring more specific knowledge. Considering data as a product which is managed and distributed actively by a product owner increases the products' reusability and simultaneously decreases time and cost to build similar products. A self-serve data platform from an architectural perspective can be more effective than point-to-point integration and hub-and-spoke integration, minimizing maintenance cost of ETL processes. Based on the self-serve data platform's technical capabilities, the federated computational governance enables scalability by giving decentralized domains enough flexibility to work effectively and efficiently as well as ensuring interoperability through standard setting.)

RESEARCH APPROACH

This study examines how large, data-intensive companies balance centralized and decentralized aspects of enterprise-wide data management (Research Question 1) and how Data Mesh at scale can support this balance (Research Question 2). To address these questions, we adopt an architectural perspective (ISO/IEC/IEEE, 2011), viewing data management as a system composed of technical and organizational components, their interrelations, and the principles governing their design and evolution.

Building on the sociotechnical systems perspective (Legner et al., 2020), we distinguish between technical components (e.g., storage, infrastructure) and social components (e.g., data governance, data ownership). This top-down approach focuses on the enterprise level rather than individual teams, ensuring an organization-wide perspective. Applying information logistics theory (Winter, 2008), we analyze how companies structure and integrate central and decentral aspects of data management within their architectures.

DATA COLLECTION

To address our research questions, we acquired data by semi-structured interviews with ten organizations. We purposefully selected large incumbent companies which, primarily due to their size, are expected to encounter a high level of complexity for managing data on the corporate level. Since we have access to senior executives in a number of those companies, we were able to gather in-depth insights into implementation practices by convenience sampling. Despite the relatively small sample size, interviewed companies are among Switzerland's largest, providing a sufficient base to explore our research questions in sufficient breath and, what is even more important, to ascertain an evolving, novel phenomenon, in sufficient depth.

The interviewed companies come from the following industries:

- Banking (medium-sized domestic bank, large domestic bank 1, large domestic bank 2, and a large Swiss universal bank)
- Insurance (large insurance company, medium-sized health insurance 1, and a medium-sized health insurance 2)
- Industrial services (mid-sized chemical company and a large electrical equipment company)
- National energy grid operator

In accordance with the proposed conceptual model, we gathered data on architectural components of the sociotechnical system and corresponding manifestations (i.e., centralized or decentralized) to obtain an overview the current composition of the centralized and decentralized aspects of data management's supply side. Furthermore, we asked for the respective developments in previous years as well as for planned data management strategic initiatives in the near future. Some additional comments on pressing challenges and data culture help to contextualize the architectural and strategic facts. Our survey was centered around the following questions:

1. What does your IT landscape for data management looks like? (How many data warehouses, data lakes, what other components; what is on-premise and what is in the cloud, on a very aggregate level?)
2. What types of data is managed centrally and what is managed decentrally?
3. If some data is managed decentrally, how do you coordinate that from an enterprise-level perspective?
4. Did the data management setup change in the last years or do you plan to change it significantly (e.g., data mesh, cloud migration, data products, etc.)?

5. How would you outline your data culture, i.e. what is the standing of data management compared to business and application management, what events or initiatives exist to push becoming data-driven; which company committees deal with data-orientation?
6. In general, to what extent do context and strategy of your company determine how you strategically understand and implement data management?

DATA ANALYSIS

Constituting a pioneering explorative analysis of large, information-intensive incumbent companies, we analyze the collected data as follows. First, from survey question 1 and 2, we identified technical components (e.g., data warehouse, data lake, data platform etc.) and social components (overall governance approach, ownership, responsibilities etc.). Second, based on survey questions 3 and 5, we analyze how certain technical components are managed with a focus on centralized and decentralized aspects. Third, based on our rich data set, in particular survey questions 4 and 6, we seek to provide an overview of how data management evolves over time.

We evaluate the characteristics of the dimensions of architectural components to be able to determine the composition of centralized and decentralized aspects of data management for each individual enterprise. This allows us to compare a company's data management requirements with the affordances of the Data Mesh concept.

RESULTS

In the following analysis, we proceed as follows. First, we position the interviewed companies on a two-dimensional grid where the y-axis represents the data governance approach (ranging from centralized to decentralized), while the x-axis represents the data ownership approach (also ranging from centralized to decentralized). The positioning reflects their as-is situation as shown in Figure 2.

The data governance dimension (y-axis) reflects where rules and standards are set; centralized data governance implies that e.g., a Chief Data Officer sets enterprise-wide rules and standards for data management and decentralized data governance entails that rules and standards are set on a lower hierarchical level (e.g., business unit level).

The data ownership dimension (x-axis) comprises the responsibility for realizing the given rules and standards. In a centralized data ownership setting, a Chief Data Officer (team) would execute the rules and standards and in a decentralized data ownership setting, e.g., business unit teams would implement rules and standards.

We derive the organization's individual positions from their directly surveyed approach to data governance and to a lesser extent also based on their architectural components (data warehouses, lakes, on-premise, and cloud). As stated above, we understand data warehouses and on-premise as rather centralized approaches, while we understand data lake and cloud as rather decentralized approaches. Hence, positioning companies in this two-dimensional grid provides insights how these organizations combine central and decentral aspects of enterprise-wide data management (Research Question 1).

Secondly, we discussed desired initiatives in response to major challenges of some organizations with regard to data governance and data ownership to understand their intentions. Both dimensions are not only fundamental to determine how centralized or decentralized companies organize their data management, but also to position them regarding their potential journey towards Data Mesh. Hence, we can see if Data Mesh at scale would potentially be an evolutionary next step and how it facilitates centralized and decentralized aspects of enterprise-wide data management (Research Question 2).

Regarding the as-is positioning of our company sample (see Figure 2), our insights are the following:

- Data Governance: In Figure 2, we see that the majority of our sample organizations follow a centralized approach to data governance. Regarding the company's industry and size, we see that insurance follows a rather centralized approach (only *medium-sized health insurance 2* is not strictly centralized). In the banking sector, we see centralized and decentralized data governance approaches. The fact that most organizations from financial services (banks and insurances) follow a rather centralized to balanced approach to data governance, is in line with extant literature: companies in highly regulated environments are generally anxious to keep control over their data (defensive data strategy) and to minimize downside risk (DalleMule & Davenport, 2017). Moreover, the tendency to a more centralized data governance might be in line with the idea to increase interoperability of data to foster cooperation and data sharing to eventually generate value from data (Gelhaar et al., 2021).
- Data Ownership: When we consider organization's data ownership positions, we find that organizations implemented a rather decentralized or balanced approach to data ownership, meaning that lower-level teams are responsible for implementing rules and standards. On the data ownership dimensions, we see a tendency that smaller banks (*medium-sized domestic bank*) follow a centralized approach whereas larger banks follow a decentralized approach (*Large domestic bank 1, Large domestic bank 2,* and *Large Swiss universal bank*). This also makes sense as larger banks decentralized their data ownership in order to increase scalability. By combining both dimensions, we see a tendency that a *medium-sized domestic bank* and *medium-sized health insurance 1* follow a centralized approach on both dimensions as they store data in a central system (core banking system and central data lake) and conduct analysis in a central analytics team. In contrast, larger banks tend to follow a decentralized approach on both dimensions. Whether the decentrality is by design or a consequence of evolutionary complexity buildup, needs to be further investigated.

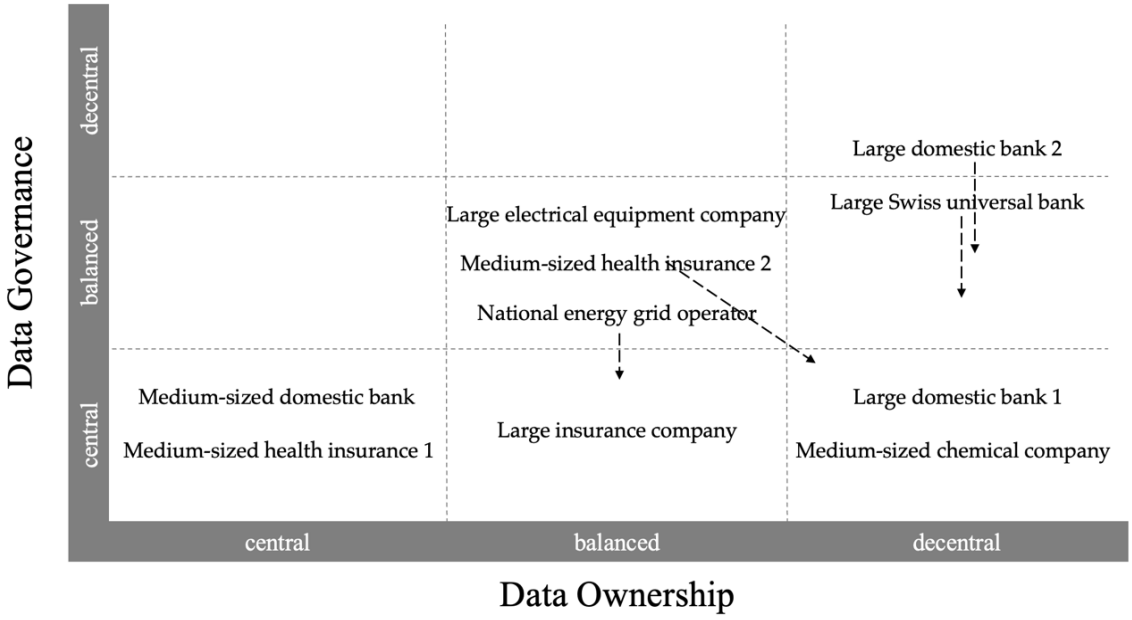

Figure 2. Results

With regard to storing and provisioning infrastructure, we see similar compositions across most companies. Data warehouses are used in nearly every business, especially in more centralized organizations – but also in larger organizations as these organizations tend to combine several infrastructure components. Data lakes are more deployed outside the financial services sector (*large electrical equipment company* and *medium-sized chemical company*), maybe because such companies deal with a larger variety and criticality of data, including technical/IoT data. We see a tendency towards cloud migration and accordingly less on-premise. However, most banks combine on-premise and cloud approaches. Finally, we observed a rise of data platforms as several companies commence conceptualizing and developing them. A trend towards 'platformization' has also been identified in previous literature (Fadler & Legner, 2020).

Regarding the evolution of corporate data management, we depict in Figure 2 planned strategic initiatives as dotted arrows in the direction of the to-be positioning. Obviously, a number of organizations aim at (further) centralizing data governance or even simultaneously decentralizing their approach to data ownership (*medium-sized health insurance 2*). *Large domestic bank 1* does not pursue a further repositioning on this two-dimensional grid. Please note that not all interviewed companies revealed their intended strategic initiatives, so that a missing to-be positioning should not be interpreted as an absence of such initiative.

Referring to the key principles of Data Mesh at scale (Section 3) and translating them into the two-dimensional grid in Figure 2 (data governance and data ownership), a to-the-book implementation of Data Mesh at scale would be positioned in the bottom right corner. Data Mesh pursues to set rules and standards in a central team which can also be composed of domain data owners and therefore qualifies for a central to balanced (federated) data governance. As the principle of data ownership underlies the concept of Data Mesh, on the data ownership dimension, Data Mesh must be located as decentral.

Summarizing the general direction of where the interviewed companies plan to evolve their corporate management approach, we can conclude a trend in the direction of Data Mesh at scale – or at least no trend in a completely opposite direction.

DISCUSSION AND CONCLUSIONS

Despite uncovering a general trend in data management towards Data Mesh at scale – or at least an indication for such trend in the initial wave of interviewed companies for this study – it is not yet clear how exactly the concept of Data Mesh can facilitate centralized and decentralized aspects of data management. To shed light on this issue, we now discuss Data Mesh's key principles and assess their theoretical impact on centralized and decentralized aspects of data management.

Based on the principle of domain-oriented ownership, Data Mesh requires a decentralized data ownership where individual business units are relatively free to cultivate their data regardless of the selection of architectural components (data warehouse, data lake etc.) within the scope of federated computational governance. Also, the high level of independence of data product managers underpins the high level of autonomy (and decentralized nature, respectively) in data ownership.

The other two key principles of Data Mesh rather relate to data governance – as federated computational governance sets the rules and standards for interoperability which must finally be enforced on a self-serve data platform. Then again, the self-serve data platform enables decentralized domains to retrieve data products independently and autonomously from other domains. In theory, Data Mesh at scale determines centralized and decentralized aspects on a higher level - but does not provide detailed guidelines for practical implementation. Therefore, and due to a lack of de facto implementations of Data Mesh in practice, a deeper analysis is needed.

Based on our findings, we conclude that we were able to uncover a trend towards, at least partly, adopting key principles of Data Mesh and thus moving towards this concept. This tendency can also

be ascertained in previous case descriptions of relatively "new" companies, e.g., Netflix (Cunningham, 2020) and Zalando (Schultze & Wider, 2020). Although we find an increasing adoption in practice, Data Mesh yet lacks a comprehensive scientific background, so that research in this field should be further pursued.

Based on this exploratory study, we propose the following plan for further research: In a first step, we suggest increasing the number of cases to confirm whether Data Mesh poses a viable approach for decentralized data management, and which are its favorable (or adverse) conditions. By underpinning our results with additional cases, our results gain further validity and generalizability. Positioning additional companies on the grid (see Figure 2), a more precise and more differentiated clustering will be possible. Moreover, examining additional organization's contingencies will afford to relate "data management archetypes" better to the different aspects of decentralization. Analyzing additional cases in-depth will also identify interesting outliers.

A better understanding of contingencies and archetypes will enable a better understanding of why organizations currently organize and reorganize their data management the way they do. These findings can then be linked to cost and value analyses to explore if value creation is distributed uniformly across organizations of various contingencies, or if there is evidence to link different data management approaches not only to context factors, but also to certain economic benefits. Finally, by grasping why organizations follow certain paths, we can contribute to practical knowledge by revealing patterns, certain "stages of maturity" including potential leap-frogging opportunities for organizational learning in data management.

Beyond that, the following broader questions regarding Data Mesh remain unclear. Future research should explore whether Data Mesh provides uniform benefits across industries and organizations of different sizes, identifying potential contingencies such as differences between large, complex enterprises and smaller firms or between service and production companies. Additionally, it remains unclear whether distinct archetypes of Data Mesh exist or if its characteristics are uniformly distributed across industries and companies. If archetypes do emerge, research should examine whether they share common benefits and challenges or represent different stages of maturity in adoption. Further investigation is also needed into how Data Mesh addresses data governance and ownership, particularly in balancing centralization and decentralization in large organizations. Lastly, research should assess the specific business implications of Data Mesh, including costs, operational tensions, and overall business value creation.

Last, but not least, the current developments around Data Mesh can and should be related with the emerging trend to transfer the digital platform concept (Tiwana & Konsynski, 2010) to data management. While these developments are 'technically' disjunct, they may have to be correlated from a corporate data management evolution perspective.

REFERENCES

Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management, 49*, 424-438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Araújo Machado, I., Costa, C., & Santos, M. Y. (2022). Data mesh: concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science, 196*, 263-271. https://doi.org/10.1016/j.procs.2021.12.013

Blohm, I., Wortmann, F., Legner, C., & Köbler, F. (2024). Data products, data mesh, and data fabric — New paradigm(s) for data and analytics? *Business & Information Systems Engineering* 66(5*),* 643–652. https://doi.org/10.1007/s12599-024-00876-5

Cunningham, J. (2020). *Netflix Data Mesh: Composable data processing.* https://www.youtube.com/watch?v=TO_IiN06jJ4

DalleMule, L., & Davenport, T. H. (2017). What's your data strategy? *Harvard Business Review, 95*(3), 112-121.

Dehghani, Z. (2019, May 20). *How to move beyond a monolithic data lake to a distributed data mesh*. Martin Fowler. https://martinfowler.com/articles/data-monolith-to-mesh.html

Dehghani, Z. (2022). *Data Mesh: Delivering data-driven value at scale*. O'Reilly Media, Inc.

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: a research agenda. *Journal of Information Technology, 33(*2), 124-135. https://doi.org/10.1057/s41265-016-0033-3

Fadler, M., & Legner, C. (2020). Building business intelligence & analytics capabilities-A work system perspective. *Proceedings of the 41th International Conference on Information Systems (ICIS 2020*), India https://aisel.aisnet.org/icis2020/governance_is/governance_is/14/

Gelhaar, J., Gürpinar, T., Henke, M., & Otto, B. (2021). Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. *Pacific Asia Conference on Information Systems (PACIS 2021)*, Dubai, UAE. https://aisel.aisnet.org/pacis2021/121/

Giebler, C., Gröger, C., Hoos, E., Schwarz, H., & Mitschang, B. (2019). Leveraging the data lake: current state and challenges. In *Big Data Analytics and Knowledge Discovery: 21st International Conference*, DaWaK 2019, Linz, Austria (pp.179-188). Springer International Publishing. https://doi.org/10.1007/978-3-030-27520-4_13

Hackl, T., Vetterling, D., & Winter, R. (2024). Leveraging corporate data platforms: an architectural perspective. *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS 2024)*, Honolulu, HI, USA. https://hdl.handle.net/10125/107124

Hagen, J. A., & Hess, T. (2020). Linking big data and business: Design parameters of data-driven organizations. *26th Americas Conference on Information Systems (AMCIS 2020).* https://aisel.aisnet.org/amcis2020/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/5/

Hai, R., Quix, C., & Jarke, M. (2021). Data lake concept and systems: a survey. arXiv. https://arxiv.org/pdf/2106.09592v1

Inmon, W. H. (1996). *Building the Data Warehouse* (2 ed.). Wiley Computer Publishing.

Inmon, W. H. (2000). Creating the Data Warehouse Data Model from the Corporate Data Model. *PRISM Tech Topics, 1*(2).

ISO/IEC/IEEE. (2011). Systems and software engineering — Architecture description (ISO/IEC/IEEE 42010:2011(E); Revision of ISO/IEC 42010:2007 and IEEE Std 1471-2000) https://doi.org/10.1109/IEEESTD.2011.6129467

Joshi, D., Pratik, S., & Podila, M. (2021). Data governance in data mesh infrastructures: the Saxo Bank case study. In *Proceedings of the International Conference on Electronic Business* (52), Nanjing, China. (pp. 599–604). https://aisel.aisnet.org/iceb2021/52/

Jung, R., & Winter, R. (2000). Data Warehousing: Nutzungsaspekte, Referenzarchitektur und Vorgehensmodell. In R. Jung & R. Winter (Eds.), *Data Warehousing Strategie* (pp. 3-20). Springer.

Kerschberg, L. (2001). Knowledge management in heterogeneous data warehouse environments. In: Kambayashi, Y., Winiwarter, W., Arikawa, M. (eds) *Data Warehousing and Knowledge Discovery*. DaWaK 2001. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44801-2_1

Khan, M. A.-u.-d., Uddin, M. F., & Gupta, N. (2014). Seven V's of Big Data understanding Big Data to extract value. In *Proceedings of the 2014 Conference of the American Society for Engineering Education IEEE*. Bridgeport, CT, USA. (pp. 1-5).

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148-152. https://doi.org/10.1145/1629175.1629210

Legner, C., Pentek, T., & Otto, B. (2020). Accumulating design knowledge with reference models: insights from 12 years' research into data management. *Journal of the Association for Information Systems, 21*(3), 735-770. https://doi.org/10.17705/1jais.00618

Lowe, D., & Galhotra, B. (2018). An overview of pricing models for using cloud services with analysis on pay-per-use model. *International Journal of Engineering & Technology, 7*(3.12), 248-254. https://doi.org/10.14419/ijet.v7i3.12.16035

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, 90(10), 60-68.

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing, special publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, https://doi.org/10.6028/NIST.SP.800-145

Nambiar, A., & Mundra, D. (2022). An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing, 6*(4). https://doi.org/10.3390/bdcc6040132

Oreščanin, D., & Hlupić, T. (2021). Data lakehouse - a novel step in analytics architecture. In *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* IEEE. Opatija, Croatia. (pp. 1242-1246).

Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information System, 29*(1). https://doi.org/10.17705/1cais.02903

Schmid, M., Haki, K., Tanriverdi, H., Aier, S., & Winter, R. (2021). Platform over Market - When is Joining a Platform Beneficial? *29th European Conference on Information Systems* (ECIS 2021). https://aisel.aisnet.org/ecis2021_rp/32/

Schultze, M., & Wider, A. (2020). data mesh in practice: How europe's leading online platform for fashion goes beyond the data lake databricks [Video]. YouTube. https://www.youtube.com/watch?v=eiUhV56uVUc

Staub, N., Haki, K., Aier, S., & Winter, R. (2021). Taxonomy of digital platforms: a business model perspective. *54th Hawaii International Conference on System Sciences* (HICSS 2021), Kauai, USA. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1220&context=wi2019

Tiwana, A., & Konsynski, B. (2010). Complementarities between organizational IT architecture and governance structure. *Information Systems Research, 21*(2), 288-304. https://doi.org/10.1287/isre.1080.0206

Venkatesh, V., Bala, H., Venkatraman, S., & Bates, J. (2007). Enterprise architecture maturity: the story of the veterans health administration. *MIS Quarterly Executive, 6*(2), 79-90.

Weber, K., Otto, B., & Oesterle, H. (2009). One size does not fit all-a contingency approach to data governance. *ACM Journal of Data and Information Quality, 1*(1), 1-27. https://doi.org/10.1145/1515693.1515696

Wende, K. (2007). A model for data governance–Organising accountabilities for data quality management. *18th Australasian Conference on Information Systems*, Toowoomba, Australia. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1079&context=acis2007

Wende, K., & Otto, B. (2007). A contingency approach to data governance. In M.A. Robert, R. O'Hare, M.L. Markus, & B. Klein (Eds.) *Proceedings of 12th international conference on information quality*, Cambridge, MA. USA. (pp. 163–176). https://www.alexandria.unisg.ch/213308/

Winter, R. (2008). Enterprise-wide information logistics: Conceptual foundations, technology enablers, and management challenges. In *ITI 2008-30th International Conference on Information Technology Interfaces*. IEEE. Zagreb, Croatia. (pp. 41-50). https://doi.org/10.1109/iti.2008.4588382

Winter, R., & Meyer, M. (2001). Organization of data warehousing in large service companies - a matrix approach based on data ownership and competence centers. *Proceedings of the Seventh Americas Conference on Information Systems* (AMCIS 2001). https://aisel.aisnet.org/amcis2001/65/

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The new organizing logic of digital innovation: an agenda for information systems research. *Information Systems Research, 21*(4), 724- 735. https://doi.org/10.1287/isre.1100.0322

AUTHORS

Robert Winter, University of St.Gallen (HSG), Switzerland is a full professor of business & information systems engineering and director of HSG's Institute for Information Systems and Digital Business. He is also founding director of HSG's Executive MBA program in Business Engineering. Having been vice editor-in-chief of the Business & Information Systems Engineering journal and senior editor of the European Journal of Information Systems, he currently serves on the editorial board of MIS Quarterly Executive. His research interests include design science research methodology and all aspects of enterprise-level IS research, such as enterprise architecture management, design and governance of digital platforms, corporate data management, and design and governance of enterprise transformation.

Tobias Hackl, University of St.Gallen (HSG), Switzerland is a PhD candidate and research assistant at HSG's Institute for Information Systems and Digital Business. In his research, he focuses on design, management and governance of data products in the context of Data Mesh.