

Blohm, Ivo; Wortmann, Felix; Legner, Christine; Köbler, Felix

Article — Published Version

Data products, data mesh, and data fabric

Business & Information Systems Engineering

Suggested Citation: Blohm, Ivo; Wortmann, Felix; Legner, Christine; Köbler, Felix (2024) : Data products, data mesh, and data fabric, Business & Information Systems Engineering, ISSN 1867-0202, Springer Fachmedien Wiesbaden, Wiesbaden, Vol. 66, Iss. 5, pp. 643-652, <https://doi.org/10.1007/s12599-024-00876-5>

This Version is available at:

<https://hdl.handle.net/10419/315766>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



CATCHWORD

Data products, data mesh, and data fabric

New paradigm(s) for data and analytics?

Ivo Blohm · Felix Wortmann · Christine Legner · Felix Köbler

Received: 19 June 2023 / Accepted: 19 April 2024 / Published online: 5 June 2024
© The Author(s) 2024

Keywords Data product · Data mesh · Data fabric · Data analytics · Data management · Data-driven innovation

1 Motivation

Turning data into action has been a driving theme for decades (Legner et al. 2020; Chen et al. 2012). However, despite the growing awareness that data-driven innovation is key to corporate success, enterprises still struggle to put data to effective use (Desai et al. 2022). Neither the technological advancements (Abbasi et al. 2016) nor improved data management practices (Legner et al. 2020) have resolved the fundamental issues and barriers in leveraging data and analytics at the enterprise scale. First, firms have traditionally focused on building data warehouses and

business intelligence tools, which are strongly governed, to provide curated, high-quality data to end-users (Watson 2002; Negash 2004). With data lakes and advanced analytics, this approach has failed to scale and meet the high demand for data from a growing number of analytics use cases. Not only have data lakes often turned into “swamps,” but data often resides in silos and analysts waste time finding and accessing relevant data sources (Giebler et al. 2009). Second, most organizations have assigned data responsibilities to a few experts in central groups who are in charge of providing data and supporting business users. These centralized data teams have become bottlenecks to scaling data-driven innovation across the whole enterprise (Someh et al. 2023). While they are highly specialized, they often lack business domain knowledge, which makes it difficult to embed data and analytics in all parts of the organization.

To overcome these “failure modes of data management” (Dehghani 2020), three concepts for using data more effectively and efficiently have recently emerged: *data product*, *data mesh*, and *data fabric*. These concepts are hotly debated as a paradigm shift in data and analytics practice. By defining socio-technical principles beyond the underlying technology stack, they aim to bring scale and standardization to meet the informational needs of an increasing number of internal or external data consumers. While each of the three concepts emphasizes specific aspects, they also share common themes such as providing an enterprise-wide focus on data and analytics, a focus on decentralized and agile data teams, as well as the effective usage of data. However, from an academic perspective, we do not really know whether and how these concepts differ from each other and whether they really constitute a fundamental paradigm shift in data and analytics or just reflect an evolution of existing concepts.

Accepted after one revision by Susanne Strahringer.

I. Blohm (✉)
Institute of Information Systems and Digital Business (IWI-HSG), University of St. Gallen, Müller-Friedberg-Str. 8,
9000 St. Gallen, Switzerland
e-mail: ivo.blohm@unisg.ch

F. Wortmann
Institute of Technology Management, University of St. Gallen,
Dufourstrasse 40a, 9000 St. Gallen, Switzerland
e-mail: felix.wortmann@unisg.ch

C. Legner
Faculty of Business and Economics (HEC), University of
Lausanne, 1015 Lausanne, Switzerland
e-mail: christine.legner@unil.ch

F. Köbler
Mindfuel GmbH, Römerstraße 18, 80801 Munich, Germany
e-mail: felix@mindfuel.ai

Against this backdrop, this catchword aims to demystify and contrast the three interrelated concepts and to integrate them into an overarching framework. Further, we propose a research agenda highlighting open questions for the Business and Information Systems Engineering community to address the underlying challenges of scaling data and analytics in enterprises. This is important for three reasons. First, information systems as well as data and analytics research has mostly focused on technology components and reference architectures (Abbasi et al. 2016), but has so far fallen short to properly conceptualize an enterprise-wide, decentralized, and usage-oriented perspective on data and analytics. Thus, providing an integrated framework and a related research agenda might advance our general understanding of the changing nature of data and analytics function in organizations, lead to more effective usage of data, and create data-driven innovation. Second, conceptual clarity is of utmost importance for scientific advancement (Podsakoff et al. 2016). Thus, highlighting the commonalities and differences among the three concepts contributes to a better understanding of the next wave of data and analytics and might help inform and aggregate future research in this domain. Finally, clarifying these concepts might help organizations to implement data and analytics solutions that fit their requirements more effectively.

2 Data products, data mesh, and data fabric

In this section, we discuss the three concepts of data products, data mesh, and data fabric. We provide definitions in Table 1 and more fine-grained conceptualizations below.

2.1 Data products

Although a product-centric view on data was already introduced by researchers from the Massachusetts Institute of Technology in the late 1990s (Wang et al. 1998), the

term has recently gained popularity with the increasing relevance of data and analytics. Following Hasan and Legner (2023), data products are referred to as “a managed artifact that satisfies recurring information needs and creates value through transforming and packaging relevant data elements into consumable form.” This definition and the emerging data product literature emphasize a consumption view on data, and that products are designed for meeting specific requirements and needs of a distinct group of users to achieve a certain goal or, more generally, to create value (Hasan and Legner 2023). Data products help to overcome the ad hoc nature of data usage and analytics in business contexts by standardization, scale, and ensuring domain-specific quality levels (Davis et al. 2020; Araújo Machado et al. 2022b). In an organizational context, data products refer to discoverable, understandable, high-quality, ready-to-use, and reusable data assets that people can apply to different business challenges (Bode et al. 2023; Dehghani 2020; Winter and Hackl 2023). They may range from primitive data products, i.e., the mere provision of data sets, to more intelligent and insightful products that include different types of descriptive, predictive, or prescriptive analytics (Davenport et al. 2022) and are tailored to various consumption archetypes such as digital apps, dashboards, or analytical information systems (Desai et al. 2022).

2.2 Data mesh

Bode et al. (2023) define the data mesh as “a socio-technical, decentralized, distributed concept for enterprise data management.” The concept was coined in a series of seminal blog posts by Dehghani (2019, 2020) who defines four socio-technical design principles to “move beyond a monolithic data lake:” *data-as-a-product*, *domain-orientation*, *self-service platforms*, and *federated computational governance*. These principles draw the attention to federated architecture, governance, and organizational design, which are neglected in the existing reference frameworks

Table 1 Definitions

Concept	Definition
Data Product	A data product is a managed artifact that satisfies recurring information needs and creates value through transforming and packaging relevant data elements into consumable form (Hasan and Legner 2023)
Data Mesh	Data mesh is a socio-technical, decentralized, distributed concept for enterprise data management (Bode et al. 2023). It is characterized by the four principles of data-as-a-product, domain orientation, self-service platforms, and federated computational governance (Dehghani 2020)
Data Fabric	Data fabric is a design concept for attaining reusable data integration services, data pipelines and semantics for flexible and integrated data delivery (Zaidi et al. 2019). It builds on the analysis, creation, and usage of metadata (Priebe et al. 2021; Liu et al. 2022) which are modelled as a knowledge graph

and in linear ways of conceptualizing data value chains (Abbasi et al. 2016). The data-as-a-product principle follows a paradigm in which data products are considered to be intentionally distributed and connected “mesh nodes” (Araújo Machado et al. 2022b). It considers data products as self-contained and consumption-ready data artifacts which are orchestrated and implemented within an enterprise-wide data ecosystem. Domain-orientation refers to the idea of decentralizing data ownership so that specific business domains/units own and leverage the data that they produce and that data is organized according to the prevailing logic of those domains by agile, autonomous, and product-oriented teams (Bode et al. 2023; Araújo Machado et al. 2022b). Self-service platforms reflect data platforms that allow for a technical abstraction of data domains and products (Bode et al. 2023; Wider et al. 2023) so that they serve as unified access points for finding, understanding, and using data products (Dehghani 2020). Such platforms ensure interoperability and connectivity as well as providing required tools and interfaces (Dehghani 2020). Thus, they contain a set of services that remove the complexity of building and managing data products from a lifecycle perspective and avoid replication of technical efforts (Araújo Machado et al. 2022a; Dehghani 2020; Bode et al. 2023). Finally, federated computational governance refers to the definition of global governance standards. It ensures the interoperability and combination of data products as well as policies for decentralized and self-sovereign governance by the distributed teams (Wider et al. 2023; Dehghani 2020; Bode et al. 2023). The execution of these policies is supported by computational approaches, e.g., specific access servers that manage access control for the different data products including sensitive data (Wider et al. 2023). Beyond these central provisions, the principle highlights self-sovereign governance by distributed teams, thus attempting to balance centralization and decentralization (Dehghani 2020).

2.3 Data fabric

The data fabric concept is concerned with the more effective integration of heterogeneous and isolated data sources so that data provision in organizations can be improved (Priebe et al. 2021; Liu et al. 2022; Ghiran and Buchmann 2019; Macías et al. 2024). The term originated in 2015 and has been strongly propagated by the consulting agency Gartner (Priebe et al. 2021) who defines it as “a design concept for attaining reusable data integration services, data pipelines, and semantics for flexible and integrated data delivery” (Zaidi et al. 2019). Pivotal to this concept are the analysis, creation, and usage of metadata (Priebe et al. 2021; Liu et al. 2022) which are modelled as a knowledge graph, i.e., graph-based data models depicting

key entities and their relationships across different data sources and products (Noy et al. 2019; Hogan et al. 2021). Data cataloging, i.e., the classification and identification of data assets and/or products including the description of data collection and processing (data lineage), reflect the foundation for creating the required metadata (Priebe et al. 2021). The knowledge graph itself is created by means of conceptual modelling (Ghiran and Buchmann 2019) or machine learning (Liu et al. 2022), i.e., active metadata according to the Gartner terminology (Beyer 2021; Priebe et al. 2021). Finally, the concept implies the creation of data pipelines for the automated and systematic ingestion, transformation, cleansing, and integration of data (Macías et al. 2024; Abu Rumman and Al-Abbadi 2023). Summing up, the data fabric concept introduces a semantic data virtualization layer that allows for automation in the processes of managing data products by connecting isolated data sources.

3 Towards an integrated framework

The three concepts of data products, data mesh, and data fabric share the common goal of helping organizations *facilitate data-driven innovation at scale* (see Fig. 1). In the following, we essentialize the three concepts’ core characteristics, clarify on the overlapping ideas between them, and highlight their inherent perspectives on affording data-driven innovation along a continuum from data consumption to provision. In so doing, we attempt to synthesize the current debate on how data can be put into more effective use in organizations and highlight that all concepts provide complimentary and required building blocks for the changing nature of the organizational data and analytics function. While all these concepts can in principle be implemented in isolation, we thus believe that their full potential can only be reaped in a combined fashion.

In the following, we discuss the framework in detail.

3.1 Data-driven innovation at scale

The implicit notion of all three concepts is that they allow to make data assets more standardized, combinable, and reusable. As a consequence, organizations might become more agile and effective in implementing data-driven innovation (e.g., novel decision-making and business processes or new products and services that are based on data) due to the logic of network effects (Desai et al. 2022; Wider et al. 2023; Dehghani 2019).

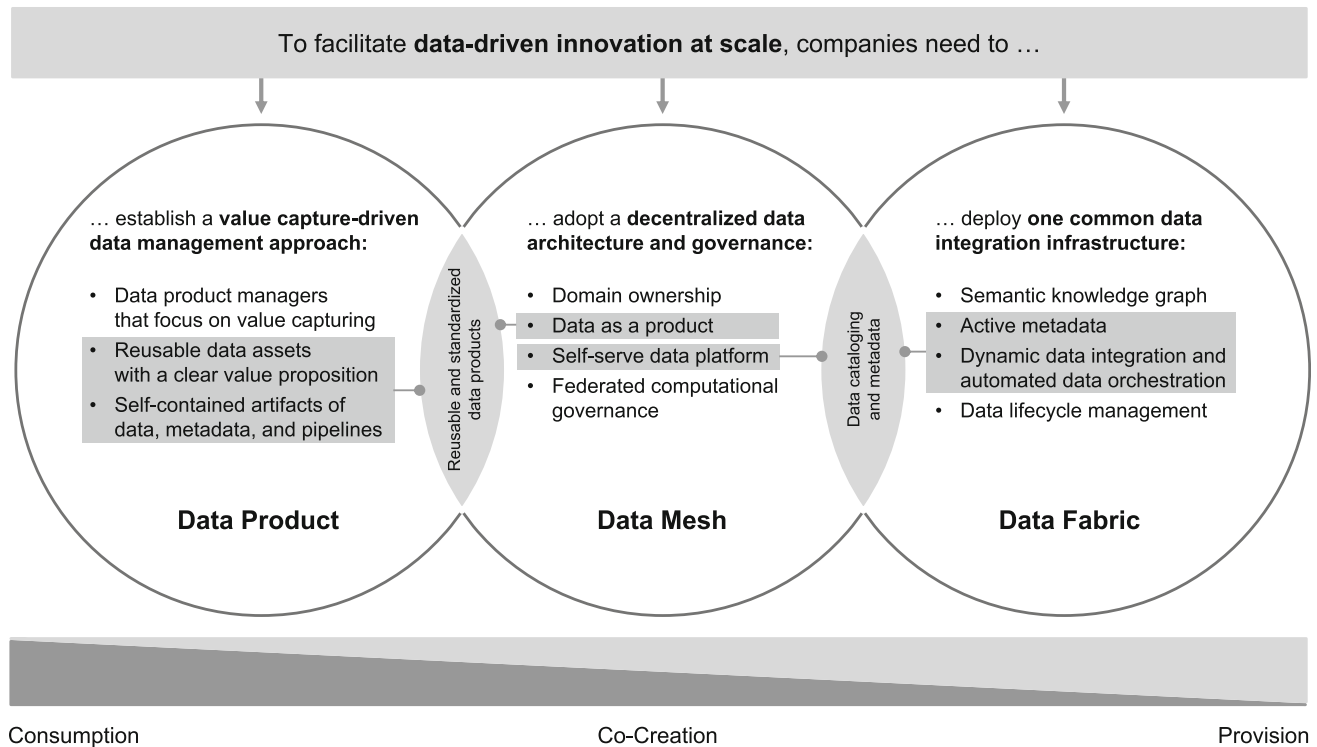


Fig. 1 Integrating the concepts of data product, data mesh, and data fabric

3.2 Value capture-driven data management – data product

The data product literature stresses the consumption of data and analytics by offering a clear value proposition to data consumers. The implicit notion is that a data product manager is in charge for ensuring that the value of data is captured (Davenport et al. 2022; Desai et al. 2022). The responsibilities of data product managers include developing a product strategy and a product roadmap to actively promote the usage and monetization of data. By contrast, data mesh and data fabric take a different perspective by focusing on data sharing and open data accessibility, without emphasizing the necessity to promote the usage of data (e.g., active selling of data products) and the way the value of data is captured (e.g., pricing, terms-of-use). Thus, data consumers are frequently considered as being paying customers.

3.3 Reusable and standardized data products – data mesh vs. data product overlap

The idea of data products is also encoded in one of the data mesh principles, that is, data-as-a-product. Both concepts consider data products as fundamental and self-contained logical units (Bode et al. 2023; Dehghani 2020; Winter and Hackl 2023). They contain data, metadata, code, policies,

and infrastructure dependencies to facilitate data sharing and usage (Wider et al. 2023). Thus, data products are considered as being vehicles for democratizing data sharing (Bode et al. 2023). In so doing, the common understanding implies the notion of reusability and standardization of data products so that they can be reused and combined by various organizational stakeholders (Desai et al. 2022; Dehghani 2020).

3.4 Decentralized and federated data architecture – data mesh

The core idea of the data mesh refers to the creation of decentralized and federated data architecture following data domains (Priebe et al. 2021). Data is owned and managed by decentralized organizational entities, e.g., autonomous teams, that are closest to the data (Bode et al. 2023; Dehghani 2020). These domain teams can use and manage data based on a self-service platform that relies on federated governance which is supported by automatic provisions (Woodie 2021; Randall 2021). This perspective of decentralized ownership and federated governance are much less pronounced in the data product and fabric concepts (Priebe et al. 2021; Hasan and Legner 2023).

3.5 Data cataloging and metadata – data mesh vs. data fabric overlap

The common goal of data mesh and data fabric refers to making data accessible and interoperable in organizations to break data silos (Bode et al. 2023). For doing so, both concepts emphasize data cataloging and the active creation, usage, and analysis of metadata (Woodie 2021; Randall 2021; Araújo Machado et al. 2022a, b; Bode et al. 2023; Beyer 2021). In the data mesh, data catalogs are referred to as cross-domain inventories of available data products (Priebe et al. 2021; Abu Rumman and Al-Abbadi 2023) and a central technology for implementing self-service platforms for which accurate and complete metadata is a core requirement (Joshi et al. 2021; Araújo Machado et al. 2022a). In the data fabric, data catalogs are considered as the starting point for the automatic creation of metadata which is required for creating a knowledge graph (Priebe et al. 2021). In a similar vein, both concepts introduce the notion of data lifecycle management including the creation, registration, updating, and deletion of data products in the data catalog and the knowledge graph.

3.6 Common infrastructure for data integration – data fabric

The data fabric basically describes a technology architecture and aims at creating a common infrastructure for integrating and orchestrating data (Bode et al. 2023). To overcome data silos and bottlenecks of centralized architecture, it introduces the concept of the knowledge graph as a single, virtualized semantic layer on top of distributed data sources. The promise is, that once the knowledge graph is created, new data sources can be easily integrated and linked to all other data asset and products, allowing the data to be orchestrated more dynamically (i.e., the overarching management of data integration pipelines) (Zaidi et al. 2019; Beyer 2021).

3.7 The continuum of data-driven innovation – consumption, co-creation, and provision

The three interrelated concepts imply different – but complementary – perspectives on how data-driven innovation is afforded at scale in organizations. Data products focus on facilitating the *consumption* of data and analytics. By contrast, the data fabric implies a *provision* perspective. It attempts to improve data integration as well as automate the creation, provision, and delivery of data and analytics. The data mesh concept refers to a set of socio-technical design principles for bridging the duality of data consumption and provision. Given its focus on data democratization (Bode et al. 2023) it refers to an architectural

paradigm for *co-creation* between data consumers and providers. For instance, the concept stresses the systematic interaction with data consumers, their integration into data product design, data sharing, and self-service platforms – all of which are central facets of co-creation.

4 Implications for business and information systems engineering

The management of data and analytics has been a core topic in the field of Business and Information Systems Engineering for decades with research streams related to data management (Legner et al. 2020) or business intelligence (Chen et al. 2012). Comparing and integrating the three concepts, we can infer a paradigm shift of data and analytics from the provision of data assets using centralized organizational and architectural approaches to facilitating the usage and consumption of data products at scale. This goes along with a decentralization of the data and analytics function and a distributed approach for data provision, analytics delivery, and the architecture of required systems (see Table 2).

However, it is important to consider the three concepts' combination and conceptualize the underlying changes on the continuum from data consumption to provision. Yet, it is too early to say whether they can overcome the existing “failure modes” of centralized approaches and lead to more effective, agile, and scalable data and analytics in organizations. We see three interesting avenues for research that may help to answer this question (see Table 3).

4.1 The consumption avenue – data product design, management, and value

While the idea of data products is gaining traction in practice and academia, operationalizing data products and instantiating systematic data product management is not yet fully understood. Little is known about the effective design and development of data products (Fruhworth et al. 2020) and their integration with already existing agile development practices (Vestues et al. 2022). The progression of (generative) artificial intelligence (AI) might change the nature of data products, that might turn into AI products, which may have differences in organizational, technical, and legal requirements, mechanisms of value creation/capture, or development approaches.

In the current academic discussion, the notion of data products is still focusing on the more effective usage of data. As data needs to be managed as products with their own lifecycle, a broader perspective is needed. This might involve studying the changes to existing operating models,

Table 2 Shifting perspectives in data and analytics

Dimensions		From...	To...
Consumption	Focal Orientation	Curation of high-quality data resources/assets	Usage of data in innovation and decision-making
	Analytics Delivery	Use-case-specific delivery of data sets, reports, dashboards, and analytical applications	Standardized data products built for scale across an enterprise-wide ecosystem
Co-Creation	Logic of Value Creation	Business enablement	Value co-creation
	Governance/Locus of Control	Centralized services by IT and/or data and analytics units	Decentralized and federated governance
Provision	Data Integration	Functional data processing and integration processes (e.g., Extract-Transform-Load (ETL))	Enterprise-wide semantic layer/knowledge graph
	DataArchitecture	Hub-and-spoke (e.g., Data Warehouse, Data Lake, Data Lakehouse)	Decentralized, domain-driven architecture that includes hub-and-spoke domains

Table 3 Research topics for business and information systems engineering

Domain	Research topics
Consumption: Data Product Design, Management, and Value	<p>Effective design and development of data products and their integration into agile development approaches</p> <p>Changing nature of data products with the proliferation of (generative) artificial intelligence</p> <p>Management of the data product lifecycle, including new operating models and roles at the intersection of IT, data, and business such as data product managers/domain owners</p> <p>Developing culture and literacy for using and creating data products</p> <p>Capturing the value of data products across different settings</p>
Co-Creation: Decentralization and Democratization of Data and Analytics	<p>Applicability, transferability, and effective implementation of decentralization principles to data and analytics</p> <p>Integrating data consumers in the design and development of data products, including open innovation and crowdsourcing</p> <p>Managing and governing decentralized value co-creation processes, including the definition of domains and responsibilities/roles at the domain level</p> <p>Design and deployment of self-service platforms for co-creation (e.g., low code/no code development platforms)</p>
Provision: Automating Data Management and Integration	<p>Data and integration architectures/models for knowledge graphs and enterprise-wide semantics</p> <p>Automatic creation of knowledge graphs and semantic models (e.g., based on data profiles and other data characteristics)</p> <p>Standards and rules for (automatically) governing data and data products according to FAIR principles</p> <p>Collaborative, decentralized, and agile data provision and management at the enterprise level</p>

development/deployment processes, roles and organizational structures, or leadership approaches. As of now, we have only started to gain an understanding of these changes. For instance, research is stressing the importance of roles such as data product managers/designers or data domain owners whose tasks, competencies, and responsibilities need to be defined in the context of cross-functional and agile teams (Vestues et al. 2022) and how end-to-end-responsibility and ownership of data products can be

established in organizations. When AI will become an integrative part of data products, data products and their development will become even more complex. For instance, existing development/deployment processes need to be adapted for being able to constantly deliver the promised customer value, e.g., robust prediction performance over time, or keeping pace with today's highly dynamic technological developments in the fields of data and AI. Thus, these approaches will have to be integrated with

existing concepts of machine learning operations (Kreuzberger et al. 2023).

It remains to be shown whether data products can better align business value, customer requirements, technical feasibility, and maintainability than traditional approaches to data consumption. Also, although it is a core idea of the concept, we need to better understand how the value of data products can be captured across different settings. For instance, today's most successful data products have been developed in the business-to-consumer domain where consumers do not directly pay for their usage, i.e., providing attention to advertising and/or personal data. Against that backdrop, it is much less clear how data products can be commercialized in internal arrangements or business-to-business settings.

Finally, the creation and usage of data products rely on a vivid data culture that is a rather novel phenomenon lacking conceptualization and evidence in the academic literature (Joshi et al. 2021). More specifically, it is still an open question of how organizations can create a culture and mindset to transform their data operations from managing data-related projects with given goals, timeframes, and budgets to managing data products that must be continuously improved to satisfy informational needs of target consumers following a lifecycle perspective. It is implicitly assumed that employees possess the required data literacy for being able to leverage data products. The proposed concepts reflect a socio-technical approach that refers to a fundamental (techno-) change management process about which we lack an in-depth understanding (Joshi et al. 2021; Markus 2004; Goedegebuure et al. 2023).

4.2 The co-creation avenue – decentralization and democratization of data and analytics

The core ideas for sharing interoperable and combinable data products on self-service platforms can be seen as the translation of well-adopted concepts from software development to the data and analytics domain. Similar principles have been discussed under the names of service-oriented architectures or microservices (Becker et al. 2011; Vom Brocke et al. 2009; Wider et al. 2023; Goedegebuure et al. 2023). These concepts are all based on fundamental socio-technical design principles such as componentization and modularization, domain-driven design, loose coupling, decentralized product ownership, as well as automated testing, integration, and delivery. They allow developers to effectively integrate IT-based services into overarching value propositions to co-create value with end-users; they have proven themselves over the last decade and are considered as being state-of-the-art in developing and running complex application landscapes. Hence, transferring these

principles into the domain of data and analytics seems reasonable and consequent.

However, can they easily be transferred? It is well known that the benefits of service orientation can come at a cost of performance, increased coordination efforts between domains, and less cross-domain collaboration (Becker et al. 2011; Feuerlicht 2006). Since cross-domain collaboration and data integration are key to data and analytics, more research is needed on how to structure the existing data landscapes and define data domains. Furthermore, effective data analytics often relies on substantial amounts of (historical) data, which calls for studying how complex data products should be designed and implemented in a domain-oriented approach. In comparison to the data exchanged by conventional microservices, data products often combine datasets from heterogeneous sources and can be expected to be much larger in size and scope. In fact, data products might be delivered via coarse-grained microservices. Unfortunately, it is well known from a microservice design that scoping in terms of the “right” granularity has strong implications for flexibility, scalability, reusability, and costs. Companies have just started to decentralize data and analytics, so that, at present, it is unclear whether the ideas can really be successfully applied to data and how they must be implemented within different organizational contingencies (Winter and Hackl 2023).

Bringing data and analytics closer to its end-users can be considered a driving theme behind the decentralization shift. This shift directly implies close interactions with consumers and end-users of data analytics to better understand their requirements and to create more effective data products (Dehghani 2020). For achieving that goal, value co-creation frequently goes beyond close interaction and strives for integrating end-users directly into problem-solving or development processes, e.g., open innovation or crowdsourcing. While research has shown that these ideas can be applied to complex data analytics problems effectively (Lakhani et al. 2013), the implications for data product development and the effective design of self-service platforms have yet to be explored. For instance, it must be answered whether the design and development of data products remains a task for professional developers and how end-users as “data and analytics laymen” could or should be integrated here. The emergence of low code/no code development platforms or the potentials of generative AI may also change the nature of self-service platforms. They might not only serve as a central repository for accessing data products but could also equip end-users with the means to create and publish their own data products.

With decentralizing data and analytics, organizations strive towards creating organization-wide ecosystems in which data can be easily shared, remixed, and augmented

to create value and foster innovation. While domain-orientation fosters autonomy and business focus, the coordination and governance of such decentralized structures for co-creating value with data and analytics is an open question (Someh et al. 2023). This does not only involve technical questions of federated computational (data) governance and the design of interoperability standards on self-service platforms but also impacts data-driven value creation as such. Future research needs to study network effects that may emerge due to the interoperability of data products (Wider et al. 2023) as well as coordination mechanisms to realize a joint value proposition across multiple domains and data products, including the alignment of motives and the creation of compatible incentives.

4.3 The provision avenue – automating data management and integration

While there is agreement on the fundamental aim of data provisioning, i.e., easy access to high quality data that is integrated across domains, core enablers of effectively automating the provision of integrated data are still to be investigated. There is a need for more research on the future of data and integration architectures (Wider et al. 2023; Goedegebuure et al. 2023; Araújo Machado et al. 2022a). While semantic integration and knowledge graphs can satisfy complex information needs and integrate diverse sources in heterogeneous formats, it is not clear whether an enterprise-wide semantic layer will complement or replace the traditional architectural paradigms that rely on strongly governed (conceptual, logical, and physical) data models and data flows. Also, the required level of granularity remains an open question from the provision perspective. The idea of the data fabric implies the creation of knowledge graphs on metadata while the original concept of knowledge graphs has been developed for storing complex relationships for single entities (e.g., product recommendations for individual customers) more effectively (Noy et al. 2019). Thus, it remains to be answered whether metadata is rich enough for creating meaningful semantic knowledge representations. In a similar vein, the latest data provisioning software comes with sophisticated data profiling capabilities to examine and analyze existing data assets that may afford novel opportunities. Also, the mesh structure might allow for better tracing of how data is used by which parts of the organization. Like the experiences in process mining, in which process models can be created from the trace data of using information systems, such novel technological affordances might help to close the gap between target data models and their actual implementation. Future research could further explore how knowledge graphs and semantic integration layers can be automatically created (Ye et al. 2022). Such approaches

would enable novel approaches to data integration and/or the development, delivery, and maintenance of data products. However, the implementation of knowledge graphs is not free of challenges. Open questions in the field of knowledge graphs, e.g., the management of changing knowledge or knowledge extraction from multiple structured and unstructured data sources (Noy et al. 2019), might warrant further research in the domain of data provision as well.

There is a need to develop infrastructures that define common rules and standards, not only for governing data but also for data products. For instance, several authors are calling for a standardized data product definition language with which data products can be systematically described such as web services (Wider et al. 2023; Goedegebuure et al. 2023). Further, future research needs to investigate how the ideas of federated computational governance can be implemented and how these approaches can be applied within volatile and ever-changing business requirements (Wider et al. 2023; Goedegebuure et al. 2023). In addition, increasing regulation, e.g., the General Data Protection Regulation and the AI Act of the European Commission, might require novel approaches and solutions to facilitate and automate the enforcement of data privacy, security, and compliance (Vestues et al. 2022). Extending this line of reasoning, self-service platforms may turn into a new generation of data catalogs that help make data FAIR (findable, accessible, interoperable, and reusable) and consumable by a wider number of users.

Finally, more insights are necessary on how collaborative data provisioning and management can be facilitated between a diverse set of stakeholders such as data engineers, data architects, data scientists, data analysts, data product managers, or data consumers. Researchers might address self-service platforms that go beyond today's data catalogs and offer novel services that bring the core innovations of collaborative software engineering into the domain of data management and integration. However, there is no thorough evidence on whether and how these tools enable effective data governance or a more collaborative or and agile development and maintenance of metadata, business glossaries, and other forms of data documentation.

5 Conclusion

Data products, data mesh, and data fabric reflect complementary concepts to foster data-driven innovation at scale in organizations. The concepts are promising to make data usage more effective, agile, and scalable by adopting an enterprise-wide perspective and emphasizing socio-technical principles. For the field of Business and Information

Systems Engineering this comes with the opportunity to rejuvenate one of our discipline's established research fields and to develop alternatives to centralized data management approaches and the technology-centric, linear conceptualization of data value chains.

Funding Open access funding provided by University of St.Gallen.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbasi A, Sarker S, Chiang RH (2016) Big data research in information systems: toward an inclusive research agenda. *J Assoc Inf Syst* 17(2):3. <https://doi.org/10.17705/1jais.00423>
- Abu Rumman A, Al-Abbadi L (2023) Structural equation modeling for impact of Data Fabric Framework on business decision-making and risk management. *Cogent Bus Manag* 10(2):2215060. <https://doi.org/10.1080/23311975.2023.2215060>
- Araújo Machado I, Costa C, Santos MY (2022a) Advancing data architectures with data mesh implementations. In: De Weerd J, Polyvyanyy A (eds) *Intelligent information systems*. CAiSE 2022. Springer, Cham, pp 10–18. https://doi.org/10.1007/978-3-031-07481-3_2
- Araújo Machado I, Costa C, Santos MY (2022b) Data mesh: concepts and principles of a paradigm shift in data architectures. *Procedia Comp Sci* 196:263–271. <https://doi.org/10.1016/j.procs.2021.12.013>
- Becker A, Widjaja T, Buxmann P (2011) Value potentials and challenges of service-oriented architectures. *Bus Inf Syst Eng* 3(4):199–210. <https://doi.org/10.1007/s12599-011-0167-3>
- Beyer MA (2021) Data fabric architecture is key to modernizing data management and integration. <https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration>. Accessed 18 Jan 2024
- Bode J, Kühl N, Kreuzberger D, Hirschl S (2023) Data mesh: motivational factors, challenges, and best practices. *arXiv preprint* <https://doi.org/10.48550/arXiv.2302.01713>
- Chen H, Chiang RH, Storey VC (2012) Business intelligence and analytics: from big data to big impact. *MIS Q* 36(4):1165–1188. <https://doi.org/10.2307/41703503>
- Davenport TH, Bean R, Jain S (2022) Why your company needs data-product managers. <https://hbr.org/2022/10/why-your-company-needs-data-product-managers>. Accessed 18 Jan 2024
- Davis J, Nussbaum D, Troyanos K (2020) Approach your data with a product mindset. <https://hbr.org/2020/05/approach-your-data-with-a-product-mindset>. Accessed 18 Jan 2024
- Dehghani Z (2019) How to move beyond a monolithic data lake to a distributed data mesh. <https://martinfowler.com/articles/data-monolith-to-mesh.html>. Accessed 18 Jan 2024
- Dehghani Z (2020) Data mesh principles and logical architecture. <https://martinfowler.com/articles/data-mesh-principles.html>. Accessed 18 Jan 2024
- Desai V, Fountaine T, Rowshankish K (2022) A better way to put your data to work. *Harv Bus Rev* 100(4):100–107
- Feuerlicht G (2006) Enterprise SOA: what are the benefits and challenges. In: *International conference systems integration*, Prague, pp 36–43
- Fruhwith M, Breiffuss G, Pammer-Schindler V (2020) The data product canvas: a visual collaborative tool for designing data-driven business models. In: *Bled eConference*, Bled, pp 515–528. <https://aisel.aisnet.org/bleed2020/8/>
- Ghiran A-M, Buchmann RA (2019) The model-driven enterprise data fabric: a proposal based on conceptual modelling and knowledge graphs. *Knowledge Science. Engineering and Management*. Springer, Cham, pp 572–583
- Giebler C, Gröger C, Hoos E, Schwarz H, Mitschang B (2009) Leveraging the data lake: current state and challenges. In: *21st international conference big data analytics and knowledge discovery*, Linz, pp 179–188. https://doi.org/10.1007/978-3-030-27520-4_13
- Goedegebuure A, Kumara I, Driessen SW, Nucci DD, Monsieur G, Heuvel W-Jvd, Tamburri DA (2023) Data mesh: a systematic gray literature review. *arXiv preprint* <https://doi.org/10.48550/arXiv.2304.01062>
- Hasan MR, Legner C (2023) Understanding data products: motivations, definition, and categories. In: *European conference on information systems*, Kristiansand. https://aisel.aisnet.org/ecis2023_rp/229
- Hogan A, Blomqvist E, Cochez M, D'amato C, Melo GD, Gutierrez C, Kirrane S, Gayo JEL, Navigli R, Neumaier S, Ngomo A-CN, Polleres A, Rashid SM, Rula A, Schmelzeisen L, Sequeda J, Staab S, Zimmermann A (2021) Knowledge graphs. *ACM Comput Surv* 54(4):1–37. <https://doi.org/10.1145/3447772>
- Joshi D, Pratik S, Rao M (2021) Data governance in data mesh infrastructures: the Saxo Bank case study. In: *International conference on electronic business*, Nanjing, pp 599–604. <https://aisel.aisnet.org/iceb2021/52>
- Kreuzberger D, Kühl N, Hirschl S (2023) Machine learning operations (MLOps): overview, definition, and architecture. *IEEE Access* 11:31866–31879. <https://doi.org/10.1109/ACCESS.2023.3262138>
- Lakhani KR, Boudreau KJ, Loh P-R, Backstrom L, Baldwin C, Lonstein E, Lydon M, MacCormack A, Arnaout RA, Guinan EC (2013) Prize-based contests can provide solutions to computational biology problems. *Nat Biotechnol* 31(7):108–111. <https://doi.org/10.1038/nbt.2495>
- Legner C, Pentek T, Otto B (2020) Accumulating design knowledge with reference models: insights from 12 years' research into data management. *J Assoc Inf Syst* 21(3):2. <https://doi.org/10.17705/1jais.00618>
- Liu K, Yang M, Li X, Zhang K, Xia X, Yan H (2022) M-data-fabric: a data fabric system based on metadata. In: *5th international conference on big data and artificial intelligence*, Fuzhou, pp 57–62. <https://doi.org/10.1109/BDAI56143.2022.9862807>
- Macías A, Muñoz D, Navarro E, González P (2024) Data fabric and digital twins: an integrated approach for data fusion design and evaluation of pervasive systems. *Inf Fusion* 103:102139. <https://doi.org/10.1016/j.inffus.2023.102139>
- Markus ML (2004) Technochange management: using IT to drive organizational change. *J Inf Technol* 19(1):4–20. <https://doi.org/10.1057/palgrave.jit.2000002>

- Negash S (2004) Business intelligence. *Commun Assoc Inf Syst*. <https://doi.org/10.17705/1CAIS.01315>
- Noy N, Gao Y, Jain A, Narayanan A, Patterson A, Taylor J (2019) Industry-scale knowledge graphs: lessons and challenges. *Commun ACM* 62(8):36–43. <https://doi.org/10.1145/3331166>
- Podsakoff PM, MacKenzie SB, Podsakoff NP (2016) Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Org Res Methods* 19(2):159–203. <https://doi.org/10.1177/1094428115624965>
- Priebe T, Neumeier S, Markus S (2021) Finding your way through the jungle of big data architectures. In: *IEEE international conference on big data*, Orlando, pp 5994–5996. <https://doi.org/10.1109/BigData52589.2021.9671862>
- Randall L (2021) Data fabric vs data mesh: 3 key differences, how they help and proven benefits. <https://www.informatica.com/blogs/data-fabric-vs-data-mesh-3-key-differences-how-they-help-and-proven-benefits.html>. Accessed 18 Jan 2024
- Someh I, Wixom B, Davern M, Shanks G (2023) Configuring relationships between analytics and business domain groups for knowledge integration. *J Assoc Inf Syst* 24(2):592–618. <https://doi.org/10.17705/1jais.00782>
- Vestues K, Hanssen GK, Mikalsen M, Buan TA, Conboy K (2022) Agile data management in NAV: a case study. In: Stray V et al (eds) *International conference on agile software development*, Copenhagen, Springer, pp 220–235. https://doi.org/10.1007/978-3-031-08169-9_14
- Vom Brocke J, Sonnenberg C, Simons A (2009) Value-oriented information systems design: the concept of potentials modeling and its application to service-oriented architectures. *Bus Inf Syst Eng* 1(3):223–233. <https://doi.org/10.1007/s12599-009-0046-3>
- Wang RY, Lee YW, Pipino LL, Strong DM (1998) Manage your information as a product. *Sloan Manag Rev* 39(4):95
- Watson H (2002) Recent developments in data warehousing. *Commun Assoc Inf Syst* 8:1. <https://doi.org/10.17705/1CAIS.00801>
- Wider A, Verma S, Akhtar A (2023) Decentralized data governance as part of a data mesh platform: concepts and approaches. In: *IEEE International conference on web services*, Chicago. <https://doi.org/10.1109/ICWS60048.2023.00101>
- Winter R, Hackl T (2023) Data mesh at scale. Exploration of current practices in large organizations. University of St. Gallen, Institute of Information Management, St. Gallen
- Woodie A (2021) Data mesh vs. data fabric: understanding the differences. <https://www.datanami.com/2021/10/25/data-mesh-vs-data-fabric-understanding-the-differences/>. Accessed 18 Jan 2024
- Ye H, Zhang N, Chen H, Chen H (2022) Generative knowledge graph construction: a review. In: *Proceedings of the conference on empirical methods in natural language processing*, Abu Dhabi. Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.2210.12714>
- Zaidi E, Thoo E, Simoni GD, Beyer M (2019) Data fabrics add augmented intelligence to modernize your data integration. <https://www.gartner.com/en/documents/3978267>. Accessed 18 Jan 2024