Data Product Metadata Management: An Industrial Perspective

Stefan Driessen[1(✉)], Geert Monsieur[2], and Willem-Jan van den Heuvel[1]

[1] Tilburg University, JADS, 's-Hertogenbosch 5211DA, The Netherlands
s.w.driessen@jads.nl

[2] Technical University of Eindhoven, JADS, 's-Hertogenbosch 5211DA, The Netherlands

Abstract. Decentralised data exchanges are promising alternatives to monolithic data lakes and warehouses which are typically emerging around complex service solutions. In theory, this removes some of the bottlenecks of traditional data management solutions. In practice, the road towards achieving such goal is a long way ahead. In this work, we provide an industry perspective on the implications for such work, with a focus on metadata management; the work in question draws from an in-vivo action research approach we enacted at a major German automotive company that is transitioning to an internal decentral data market. Our results provide insight into an industry perspective on the requirements for metadata management. Additionally, we propose and validate a solution design for metadata management in decentralised data exchanges based on semantic web service technology.

Keywords: Data mesh · Data market · Data product · Metadata · Semantic web

1 Introduction

Despite the promises of big data to revolutionise the way companies do business, many organisations still grossly fail to fully capitalise on the data they are generating[1]. For example, several surveys and market analyses show that 60% to 85% of data analytic and data science initiatives never make it to production [18]. Critics have blamed this inability to make full use of data inside an organisation or company on the monolithic approaches typically exploiting such data—e.g., data lakes and warehouses—that are nowadays the standard architecture approach for storing and exchanging data [11,15]. The main downside of these monolithic approaches is that they fail to scale with the number of data sources on the one hand and data science and analytics use cases on the other [15,25].

To address these shortcomings, grey and white literature is showing increasing interest in decentralised data exchanges, such as data markets [6], data

[1] https://www.sisense.com/blog/why-businesses-fail-to-capitalize-on-their-data/.

J. Troya et al. (Eds.): ICSOC 2022 Workshops, LNCS 13821, pp. 237–248, 2023.
https://doi.org/10.1007/978-3-031-26507-5_19

meshes [9], and data spaces [19]. Despite some minor differences in how these platforms approach the sharing and exchange of data, their approaches all focus on offering data as a product/service and are heavily inspired by the microservice paradigm [16,17]. Whereas monolithic approaches rely on a central data office to facilitate data management, in decentralised data exchanges, data providers[2] are responsible for taking (operational) data from their domain and providing it in a manner that is fully optimised for consumption by data consumers from across the organisation.

Despite the theoretical promises of decentralised data exchange platforms, many challenges currently impede their effective implementation and migration. As far as we know, no company or organisation claims to have successfully organised its data exchange in accordance with any proposed theoretical framework. Currently academic studies [9,11,15], and grey literature [23,25] focus on high-level architectural concerns, as well as categorising the challenges and solutions associated with migration and design [3,6]. In this paper we explore one such challenge, namely metadata management in a decentral data exchange. In particular we emphasise the challenge of achieving interoperability (i.e. relating disparate data sources), which has historically always been addressed by adding a central component to the data platform architecture. To supplement existing academic work with practical concerns, we take the industrial perspective of a large German automotive manufacturer who is in the process of transitioning from monolithic architectures towards a decentralised data exchange.

The rest of this paper is organised as follows: In the next section we discuss relevant works on metadata management for decentralised data exchanges and introduce the industrial context provided by the company where our research took place. Section 3 describes how we leveraged design science research to establish goals, problems, requirements and a potential solution. Then, in Sect. 4 we present the results of our research. Finally, in Sect. 5 we discuss the threats to validity in our research approach, the implications of our work and suggestions for follow-up research.

2 Background and Related Literature

We observe that metadata management for internal data exchange platforms is still very much in its infancy. Indeed, Eichler et. al. discuss state-of-the-art metadata management and conclude that there is a research gap in metadata management, especially for internal decentral data exchanges [7]. Some works do exist that focus on the exchange of data *between* companies and organisations. For example, Roman et. al. present an ontology and show how it can be used to harmonise data from different organisations using well-established ontology development methods [20]. Similarly, Spiekermann et. al. present a metadata model for data products in the context of commercial data markets [21].

[2] Alternatively called data owners, (data) product providers, or (data) product developers [6].

When it comes to *internal* data exchanges, proposed solutions for metadata management tend to focus on modelling (meta-)data in knowledge graphs using semantic web technology [26]. For example, Hooshmand et. al. emphasise the power of semantic web technology to capture and combine domain knowledge on business objects and technical information on data assets. They propose a transition towards a decentral data mesh for managing data in the product lifecycle management landscape and discuss how different domains can have separate knowledge graphs that can be mapped to achieve interoperability; however, they do not discuss explicitly what metadata management should look like [11]. Other relevant solutions for metadata management have been proposed in the context of centralised architectures. Stach et. al. note the advantages of semantic web technology in terms of ease of use for data consumers who are not experts at data modelling and propose a method for describing desired data products [22]. Similarly, Dibowski and Schmid introduce a full ontology for describing data assets on (internal) data lakes and explain how this improves the discoverability and reusability of data [4].

For our investigation of interoperable data products, we engaged the IT division of a major German automotive manufacturer, which we refer to as the Data Market Implementation Team (DMIT). The company was experiencing challenges in effectively sharing data, and the DMIT was investigating new ways to tackle these challenges with an internal data market. The automotive manufacturer operates with a multi-billion euro revenue in a global market, is organised in several organisational units across multiple continents, employs more than 100.000 employees, and has numerous partner companies in its business ecosystem. Importantly, the operations of this company are not limited to manufacturing but extend to different post-sales services as well. This collaboration allowed us to approach the problem in an industrial setting and get direct input from real-world data providers, data consumers and infrastructure providers.

3 Research Methodology

We employ the design science research approach, which focuses on creating and evaluating artefacts to simultaneously address industry-relevant problems and contribute new knowledge to the scientific community [10]. As shown in Fig. 1, our method consists of three steps in the design cycle: problem evaluation, treatment design, and treatment validation [27]. Below, we describe each of these steps individually, after which Sect. 4 describes the results of each step and how these results informed the consecutive steps.

Problem Evaluation. During the problem evaluation, we started by investigating existing literature on decentral data exchanges in conjunction with repeated interviews with experts from the DMIT, to establish who would be the main stakeholders affected by the implementation of the internal data market. Literature was gathered by snowballing on two existing structured literature reviews (SLRs) for data market- and data mesh design [6,9].

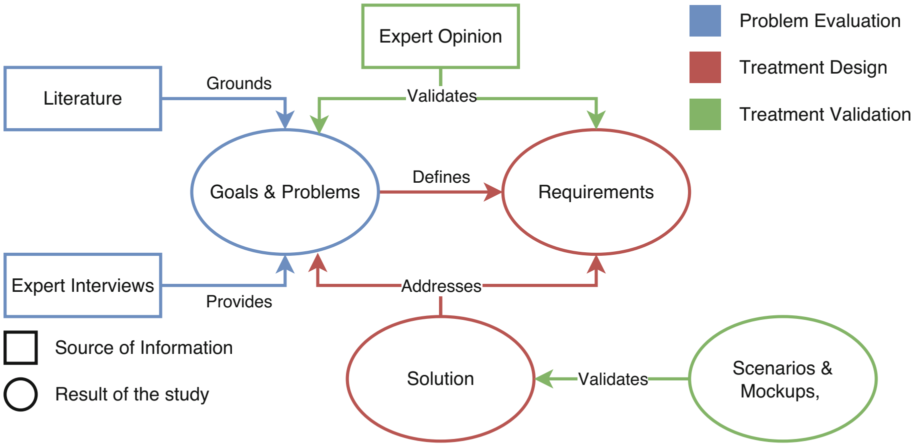

Fig. 1. A figure showing how our methodology relates to the results of the study. The rectangles show external sources of information, and the ellipses show the results presented in this paper. Additionally, the steps of the design cycle [27] are grouped by colour.

These efforts yielded three central stakeholders. 1) The data provider, who is ultimately responsible for the data product. 2) The data consumer, who is the intended user of data products. 3) The platform provider, who is responsible for the IT infrastructure of the decentral data exchange, including the metadata management. Afterwards, we selected several experts from across the company with perspectives for each type of stakeholder, who were then interviewed to establish the different stakeholder goals and problems.

Treatment Design. Based on the goals identified in the interviews and the context provided by the DMIT, we specified the requirements for our treatment and established how satisfying those requirements would contribute to the stakeholder goals. We then considered existing literature on metadata management for effective data sharing and found that an approach based on semantic web technology had the potential to address all the identified requirements.

Treatment Validation. In order to validate our proposed solution, we relied on expert opinion whereby various stakeholders were asked to evaluate a potential solution by coming up with potential problems and benefits [2,14,27]. To help validate that our solution addresses the requirements, we created and described several scenarios for creating, updating, combining and reusing data products and metadata. Furthermore, one or more mock-ups were created for each scenario to illustrate how our solution addressed the corresponding requirements. Afterwards, the scenarios and mock-ups were presented to experts from various domains. This included the interviewees from the problem evaluation step and members of the DMIT, who were asked for feedback. Finally, a workshop was organised where our findings were presented to a large audience of over 50 stakeholders at the company. At this point, feedback was solicited again from these stakeholders.

4 Results

In this section, we discuss the results of our research. As mentioned in Sect. 3, these insights led us to propose a high-level solution and validate it through scenarios and mockups. In order to preserve space, in this paper, we choose to emphasize the results of the interviews, which lead to the formulation of eight goals, eight problems, and five requirements for metadata management in a decentral data exchange. These problems, goals and requirements can guide practitioners and academics alike in their efforts toward creating metadata management solutions for decentral data exchanges. An overview of our proposed solution, as well as the scenarios and mockups used to validate these, can be found online at https://anonymous.4open.science/r/Data-Product-Interoperability-E633.

4.1 Problem Evaluation

In order to accurately identify the requirements for interoperable data product metadata management, we performed interviews with ten expert stakeholders across the company. The experts were selected in consultation with the DMIT to represent the perspectives of the different stakeholders on data exchange. Additionally, as reflected by Table 1, we consciously tried to interview people from various departments with varying levels of expertise and seniority. However, one challenge that arose was that there were no real data providers yet. This seems likely to occur in many organisations looking to transition towards a decentralised data exchange architecture. As mentioned in Sect. 2, in the existing landscape of centralised architectures such as data warehouses and data lakes, onboarding data is the responsibility of a central team of platform providers. Part of moving towards a decentral data exchange entails moving these responsibilities to domains' expert data providers who work directly with the operational data [3]. In order to still get the data provider's perspective, we selected interviewees that had all been involved in previous initiatives to improve the existing data exchange infrastructure. Consequently, they were platform providers that had explicitly considered the challenge of onboarding new data assets, if not data products.

Table 1 shows the characteristics of the interviewed experts. On the one hand, the interviewees' experience with the existing data exchange infrastructure and its limitations made them ideal candidates for our investigation because they had already considered the goals and problems for their own initiatives. On the other hand, the emphasis on different perspectives (both domains and roles) ensured that we would not end up with a subset of relevant problems, goals and requirements. The interviews themselves were semi-structured, focusing on existing processes and desired processes for data exchange and the planned internal data exchange. The goals and problems that resulted from these interviews are described in Tables 2 and 3 respectively and are discussed below per stakeholder.

Data Provider. The data provider has three broad goals in the data exchange, the first of which (**G1**) is to decide which data assets would make the most *valuable* data products. However, as captured by **P1**, whenever the data consumer and the data provider come from different domains, it becomes challenging to assess what data products are relevant in the context of the data consumers domain [8,13]. Lack of data engineering expertise and the costs associated with creating and maintaining data products (**P2**) reinforce the need to prioritise when creating data products.

After identifying which data assets to turn into products, the second goal for data providers in an internal data exchange is to create and maintain these data products (**G2**). The literature describing a transition towards a decentralised data exchange platform emphasises that one of the greatest challenges in this transition is the organisational "left shift" whereby domains have to take on the new functional responsibilities of providing data [3]. Our interviews confirm that it is especially challenging to onboard new data providers, particularly if the process of creating and maintaining data products is perceived to be costly (**P2**).

Ensuring that data products are valuable by prioritising is crucial to the success of an internal data exchange. However, this is not enough for the data provider, as they must also convey *why* and *how* the data product is valuable to other actors (**G3**). First among these actors are the data consumers, who might not understand the value, leading to an unused data product. However, many interviewees also explicitly expressed the need to convince colleagues and

Table 1. Overview of the interviewed experts. Providers (P) gave their perspectives on both the platform providing and the data providing. Consumers (C) provided insights for the consumption of data.

Role	Job title	Work experience	Department
P	Enterprise Data Architect	23 years	Enterprise Architecture
P	Manager Enterprise Architecture	7 years	Enterprise Architecture
P	IT-Consultant/IT-Manager	23 years	IT Digital Services
C & P	Data Manager	17 years	Product Digitalisation
C	Manager Data Governance Office	25 years	Finance & Controlling
P	Technical Lead/Methods & Tools	12 years	IT Product Engineering
C	IT Project Manager	10 years	Big Data & AI
P	BI & Analytics Architect	8 years	Technical Architecture Finance Analytics
C & P	Manager Enterprise Business Architecture	29 years	Enterprise Architecture
P	Manager Technology Strategy	15 years	External Consultant

Table 2. The different goals and corresponding problems for each stakeholder. The frequency column shows how many of the interviewed experts (out of nine) mentioned each of the goals.

Stakeholder	Goal	Problems	Freq.
Data provider	**G1** Prioritise data assets to turn into data products	**P1 P2**	7
	G2 Create and maintain data products in a cost-effective manner	**P2**	8
	G3 Convincingly express the value of data products	**P1 P3 P4**	8
Data consumer	**G4** Discover and understand relevant data products	**P1 P3 P4**	7
	G5 Consume & combine data products	**P5**	8
	G6 Incentivise the creation of relevant new data products	**P1 P2 P6**	5
Platform provider	**G7** Create and maintain metadata tools as part of easy-to-use self-serve infrastructure	(**P1**–**P6**) **P7**	8
	G8 Extend internal platform to external data exchanges	**P8**	2

Table 3. There are seven problems that the interviewees expect to run into, which describe the obstacles faced by the three main roles. Proper metadata management in an internal data exchange should try to address these problems.

Problem	Description	Freq.
P1	There is a gap between the domain knowledge of data providers and data consumers	7
P2	It is costly to (learn how to) create and maintain data products	8
P3	Similar or identical business objects can lead to significantly different data products	4
P4	It is challenging to understand data product semantics	7
P5	Combining data from different sources is technically challenging	8
P6	Sometimes data is not available but still desired	5
P7	End users lack data engineering expertise	8
P8	External organisations might use different standards	2

management from the provider's domain of the necessity of spending resources on creating and maintaining data products. This is because without the support from these colleagues and managers, providing data is not a sustainable activity. The gap in knowledge between data providers and consumers (**P1**) makes it hard for data providers to convey this value. In contrast, the overall difficulty of understanding the semantics of data products (**P4**) hinders the consumers' efforts to recognise it. Finally, several interviewees reinforced an idea proposed in literature [12] that it is quite difficult to figure out which data is relevant for them whenever multiple data sources are available that describe the same business object from the perspective of different domains (**P3**).

Data Consumer. The data consumer also has three goals: the first one is to find the *right* data for their use case, which requires a *semantic* understanding of the data product (**G4**). This understanding includes the context of the data, the meaning of the different values and attributes, but also the relevant policies and service level agreements. This goal is similar to **G3**, only viewed from the consumer's side. The problems that can be addressed by metadata management are also the same. In particular, the fact that it is challenging to find and understand relevant data products (**P4**) is reinforced by the gap in knowledge between data providers and consumers (**P1**). Additionally, as noted above, differentiating between data products from different domains can be especially challenging (**P3**).

Once the data consumer understands the data product, their next goal is to integrate the data in their use cases, either directly or by combining it with other data (**G5**). Even if the consumer fully understands the data provider's offering, there are still technical challenges associated with consuming and combining data products. In particular, even if it is clear what each attribute in the data product means semantically, this does not automatically lead to a way to connect it to other data (e.g. through schema matching) (**P5**) [8].

Finally, our interviews revealed that data consumers want the ability to discover what data assets exist in operational systems and to incentivise the creation of data products offering data perceived as useful (**G6**). This goal directly tries to address the problem that sometimes data is not available but still desired (**P6**). Additionally, understanding what data exists in operational systems is challenging when these systems exist in domains that are separate from the consumer (**P1**). Furthermore, Incentivisation is hindered by the perceived costs that the data consumer incurs when creating a data product (**P2**).

Platform Providers. The platform providers are charged with creating the internal data exchange on which the data products are exchanged. As such, their main concern is to provide the infrastructure that allows the data providers and data consumers to achieve their goals and overcome their problems (**P1**–**P6**). However, as data providers and consumers are generally not data engineering experts (**P7**), this infrastructure should come through the creation of an easy-to-use self-serve infrastructure layer [3] **(G7)**.

In addition to addressing the goals and problems of the data providers and data consumers, however, the interviewees from the DMIT also expressed the wish to expand their internal data exchange in the long term to work with external platforms such as the Catena-X initiative [1] which connects automotive data platforms (**G8**). The problem that we foresee with this goal is that standards and metadata management initiatives that are developed internally for the automotive company might not extend easily to other platforms (**P8**).

4.2 Treatment Design

Based on the goals and problems described above, we formulate five requirements that *any* approach for metadata management in a decentral data exchange should try to meet (see Table 4).

Table 4. Five requirements were identified to help the actors reach their goals and overcome their respective problems. For each requirement, the goals and problems addressed by that requirement are shown in the final column.

Req.	Description	Addresses
R1.	The metadata management tools should allow data providers to capture domain expertise (i.e. semantic knowledge) as well as technical expertise (i.e. data schemas and statistics) in their models	**G3, G4, G5, P1, P3, P4, P5**
R2.	The resulting models should allow data products to be connected on a data level, even when crossing domain or organisational boundaries	**G5**, **G8**, **P5**, **P8**
R3.	The resulting models should relate data products semantically, even when crossing domain or organisation boundaries	**G3**, **G4**, **G5**, **P1**, **P3**, **P4**
R4.	Metadata should be created autonomously by data providers and this should be as easy as possible	**G2**, **G7**, **P2**, **P7**
R5.	The metadata management tools should allow data consumers to express data product requirements	**G1**, **G6** **P1**, **P6**

The first requirement (**R1**) is a direct consequence of the same goals and problems that motivate a transition towards decentralised data exchanges. Understanding the semantics of the data (e.g. which business processes are involved, how it is collected and for what purpose) is essential for discovering and understanding data (**G4**). Central data offices are not as familiar with these aspects of the data as data providers from the domain, who are better suited to capturing this information in the metadata as the number of sources increases (**G3**). At the same time, technical information, such as the data schema and describing statistics, is still necessary to consume it effectively for a use case (**G5**). Therefore, metadata management in decentral exchanges should allow data providers to explain both types of properties in a human-readable (and possibly machine-readable) manner within the same environment.

The second requirement (**R2**) has always been a main requirement for, and indeed focus of, centralised (meta-)data management. Data warehouses, in particular, address this problem by tightly coupling schemas to a global mediated schema [5]. Meanwhile, for data lakes, interoperability is usually addressed on a use case basis by the members of a central data office. These approaches allow data consumers to easily consume and combine data from different sources (**G5**). However, their reliance on a central bottleneck make them unsuited for decentral data exchanges. Even if data products cannot be tightly coupled to a single cross-domain schema, the metadata management tools should make it easy for data providers and consumers alike to connect (the schemas of) different data products.

The next requirement (**R3** shows that, for the semantic information to be truly effective in helping data consumers find, understand and consume data products (**G4**, **G5**), it needs to relate to their domain knowledge. This helps the data consumer understand the differences, similarities and nuances between business processes that often involve similar business objects (e.g. cars) but generate vastly different data and can greatly reduce the time and efforts required to decide if- and how to use that data. Moreover, if the data provider succeeds in relating their domain knowledge to that of the data consumer, it is more likely that they can convincingly express the value of their data product (**G3**).

The fourth requirement (**R4**) takes into consideration the previously noted organisational "left shift" of responsibilities from the central data office to the data provider that accompanies the transition toward a decentral data exchange (**G2**). Reducing cognitive load for platform users is a problem from cognitive science that has been well-researched for IT artefacts [24]. Therefore tools and standards should be made available that enable data providers to create and manage metadata autonomously, without direct interference from the platform providers (G7). In this sense, data products mirror microservices, which are designed to be self-contained.

The final requirement (**R5**) is not as prominent in academic literature. Still, it becomes apparent when realising that innovation in (meta)data management in most industrial settings is driven by data consumers, who feel existing shortcomings most acutely. Allowing data consumers to express and incentivise the creation of new data products (**G6**) allows faster data product development and more accurate prioritisation by the data provider (**G1**).

5 Discussion and Conclusion

The requirements discussed in Sect. 4 were mostly consistent with those mentioned in existing literature, but the goals and problems that underlie them had not yet been discussed in detail before. Moreover, we found that an important requirement for practitioners that has been mostly overlooked in academia is the need to assign priorities to data assets that need to be transformed into data products. This prioritisation has two sides: the data providers want to validate their efforts, effectively ensuring that their created data products will be consumed. Similarly, however, data consumers want to express their needs for new data products. Literature on decentral data exchanges seems to have mostly overlooked these goals; only Stach et. al. have investigated this problem in the context of data lakes [22].

Our second finding is that creating proper data providers is a major challenge for organisations trying to transition to decentral data exchanges. Although this may not be a novel insight, the implications this challenge has on metadata management are. Ease of use has already been mentioned in academic literature. However, we find that metadata management tools should also make it easy for data providers to use existing resources (e.g. ontologies or data products) as a template. At the same time, the use of these templates should not be enforced too rigorously, and it should be easy for data providers to deviate from them whenever their ground truth demands it. We believe a metadata management approach based on Semantic Web Technology (SWT) to be a promising way to address these requirements. SWT emphasises the ability of decentral actors to create their own (domain-based) ontologies and standards which can then be related to other domains. These relations can be created either directly between domains as envisioned by Roman et. al. [20], or through a higher-level and more abstract company-wide ontology such as the one proposed by Hooshmand et. al. [11]. Additionally, such an approach can combine semantic and technical metadata into a single entity and present information in a human-readable manner. Moreover, SWT is machine-readable, which opens the door for the development of automatic interoperability tools in the future.

We acknowledge the several threats to the validity of our experiments and conclusions. First, concerning the internal validity, we note that no true data providers existed yet, in the sense that data was provided autonomously by domain teams for one or more external data consumers. We addressed this concern by interviewing extra platform providers and focusing our efforts on those who worked directly with domain teams. A threat to our findings' external validity is that they are founded on investigations inside a single company. To address this concern, we ground the findings in academic literature wherever possible. Additionally, we intend to follow up on our findings with a survey with participants across many organisations. Finally, we note that our research has lead the automotive company to take the first steps towards implementing a semantic-web based approach for metadata management based on our recommendations and validations.

References

1. Catena-X: Automotive Network (2021)
2. Alexander, I.F., Beus-Dukic, L.: Discovering Requirements: How to Specify Products and Services. Wiley, Chichester (2009)
3. Dehghani, Z.: Data Mesh: Delivering Data-Driven Value at Scale, 1st edn. O'Reilly (2022)
4. Dibowski, H., Schmid, S.: Using knowledge graphs to manage a data lake. In: INFORMAITK 2020, Lecture Notes in Informatics (LNI), pp. 41–50 (2021)
5. Doan, A., Halevy, A., Ives, Z.: Principles of Data Integration, 1st edn. Elsevier, Waltham, MA (2012)
6. Driessen, S., Monsieur, G., Van Den Heuvel, W.: Data market design: a systematic literature review. IEEE Access **10**, 33123–33153 (2022). https://doi.org/10.1109/access.2022.3161478
7. Eichler, R., Giebler, C., Gröger, C., Hoos, E., Schwarz, H., Mitschang, B.: Enterprise-wide metadata management: an industry case on the current state and challenges. In: Business Information Systems (July), pp. 269–279 (2021). https://doi.org/10.52825/bis.v1i.47
8. Fernandez, R.C., Subramaniam, P., Franklin, M.J.: Data market platforms: trading data assets to solve data problems. Proc. VLDB Endow. **13**(12), 2150–8097 (2020)
9. Goedgebuure, A.: Data mesh: systematic gray literature study, reference architecture, and cloud-based instantiation at ASML (2022). https://stefan-driessen.github.io/publication/data-mesh-systematic-grey-literature-study/
10. Hevner, A., Chatterjee, S.: Design Research in Information Systems: Theory and Practice, vol. 28. Springer, NY (2010). https://doi.org/10.1007/978-1-4419-5653-8

11. Hooshmand, Y., Resch, J., Wischnewski, P., Patil, P.: From a monolithic PLM landscape to a federated domain and data mesh. Proc. Design Soc. **2**, 713–722 (2022)
12. Koutroumpis, P., Leiponen, A., Thomas, L.: The (unfulfilled) potential of data marketplaces. ETLA Working Papers 2420(53) (2017). http://pub.etla.fi/ETLA-Working-Papers-53.pdf%0Apub.etla.fi/ETLA-Working-Papers-53.pd
13. Koutroumpis, P., Leiponen, A., Thomas, L.D.W.: Markets for data. Ind. Corp. Chang. **29**(3), 645–660 (2020). https://doi.org/10.1093/icc/dtaa002
14. Lauesen, S.: Software Requirements-Styles and Techniques. Pearson Education (2002)
15. Loukiala, A., Joutsenlahti, J.-P., Raatikainen, M., Mikkonen, T., Lehtonen, T.: Migrating from a centralized data warehouse to a decentralized data platform architecture. In: Ardito, L., Jedlitschka, A., Morisio, M., Torchiano, M. (eds.) PROFES 2021. LNCS, vol. 13126, pp. 36–48. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-91452-3_3
16. Narayan, S.: Products over projects (2018). https://martinfowler.com/articles/products-over-projects.html
17. Newman, S.: Monolith to Microservices: Evolutionary Patterns to Transform Your Monolith. O'Reilly (2020). https://www.oreilly.com/library/view/monolith-to-microservices/9781492047834/
18. O'Neil, B.T.: Failure rates for analytics, AI, and big data projects = 85% - yikes! (2019)
19. Otto, B., Steinbuß, S., Teuscher, A., Lohmann, S.: IDSA reference architecture model. International Data Spaces Association (April) (2019). https://internationaldataspaces.org/download/16630/
20. Roman, D., et al.: The euBusinessGraph ontology: a lightweight ontology for harmonizing basic company information. Semantic Web **13**(1), 41–68 (2021). https://doi.org/10.3233/sw-210424
21. Spiekermann, M., Tebernum, D., Wenzel, S., Otto, B.: A metadata model for data goods. In: MKWI 2018 - Multikonferenz Wirtschaftsinformatik 2018-March, pp. 326–337 (2018)
22. Stach, C., Bräcker, J., Eichler, R., Giebler, C., Mitschang, B.: Demand-driven data provisioning in data lakes. In: Association for Computing Machinery, vol. 1 (2021). https://doi.org/10.1145/3487664.3487784
23. Strengholt, P.: ABN AMRO's data and integration mesh (2020). https://www.linkedin.com/pulse/abn-amros-data-integration-mesh-piethein-strengholt/
24. Sweller, J.: Cognitive load during problem solving: effects on learning. Cogn. Sci. **12**(2), 257–285 (1988). https://doi.org/10.1016/0364-0213(88)90023-7
25. Dehghani, Z.: How to move beyond a monothilitic data lake to a distributed data mesh (2019). https://martinfowler.com/articles/data-monolith-to-mesh.html
26. W3C: Semantic web - leading the web to its full potential (2015)
27. Wieringa, R.J.: Design Science Methodology for Information Systems and Software Engineering. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43839-8