



University of Stuttgart - Institute for Parallel and Distributed Systems / AS

Enterprise-Wide Metadata Management

An Industry Case on the Current State and Challenges


Rebecca Eichler, Corinna Giebler, Christoph Gröger, Eva Hoos, Holger Schwarz and
Bernhard Mitschang

In: Proceedings of the 24th International Conference on Business Information
Systems (BIS 2021)

BIBTEX:

```
@inproceedings{Eichler2021,  
author = { Eichler, Rebecca and Giebler, Corinna and Gröger, Christoph and  
Hoos, Eva and Schwarz, Holger and Mitschang, Bernhard},  
booktitle = { Proceedings of the 24th International Conference on Business  
Information Systems (BIS 2021)},  
title = {{Enterprise-Wide Metadata Management - An Industry Case on the Current  
State and Challenges}},  
year = {2021},  
doi = {https://doi.org/10.52825/bis.v1i.47}  
}
```

Published with Technische Informationsbibliothek (TIB) in open access model

 Creative Commons Attribution 4.0 International (CC BY 4.0)

This is the author's version of the work. The final authenticated version is available
online at: <https://doi.org/10.52825/bis.v1i.47>

Enterprise-Wide Metadata Management

An Industry Case on the Current State and Challenges

Rebecca Eichler¹, Corinna Giebler¹, Christoph Gröger², Eva Hoos², Holger Schwarz¹, and Bernhard Mitschang¹

¹ University of Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany
{Firstname.Lastname}@ipvs.uni-stuttgart.de

² Robert Bosch GmbH, Borsigstraße 4, 70469 Stuttgart, Germany
{Firstname.Lastname}@de.bosch.com

Abstract. Metadata management is a crucial success factor for companies today, as for example, it enables exploiting data value fully or enables legal compliance. With the emergence of new concepts, such as the data lake, and new objectives, such as the enterprise-wide sharing of data, metadata management has evolved and now poses a renewed challenge for companies. In this context, we interviewed a globally active manufacturer to reveal how metadata management is implemented in practice today and what challenges companies are faced with and whether these constitute research gaps. As an outcome, we present the company's metadata management goals and their corresponding solution approaches and challenges. An evaluation of the challenges through a literature and tool review yields three research gaps, which are concerned with the topics: (1) metadata management for data lakes, (2) categorizations and compositions of metadata management tools for comprehensive metadata management, and (3) the use of data marketplaces as metadata-driven exchange platforms within an enterprise. The gaps lay the groundwork for further research activities in the field of metadata management and the industry case represents a starting point for research to realign with real-world industry needs.

Keywords: Metadata Management, Data Sharing, Data Transparency, Data Catalog, Data Marketplace

Introduction

In recent years, metadata management has regained focus in the scientific field and has once more become a topic of discussion in enterprises. Metadata management is important as it constitutes the activities to administrate an organization's data assets through metadata [1]. Without metadata, an organization does not know, for instance, what data it has collected, what it represents or whether it is confidential. Consequently, topics like legal compliance cannot be guaranteed without, e.g., information on confidentiality or the data's value cannot be fully leveraged when its meaning is unclear.

Data value in the form of new insights can be extracted through methods such as data analytics and is of great significance in enterprises as it can provide a competitive advantage [2]. In order to maximise the utilization of data and the extraction of its value, it needs to be made available to a wide range of users. In order to make data available throughout the enterprise, that is, beyond individual systems like data lakes or enterprise resource planning (ERP) systems, enterprise-wide metadata management is needed. Enterprise-wide metadata management encompasses and integrates metadata management initiatives of both analytical systems such as data lakes and operational systems such as ERP systems and enables the access to and usage of metadata across the enterprise [3], for instance, in the form of a data-asset-inventory across various source systems. Yet, recent research, such as [3],[5],[6],[7],[8], mainly deals with metadata management explicit to data lakes, in which it is a central aspect [9]. Apart from research, there are a number of metadata management tools like data catalogs on the market, designed to solve specific metadata

management tasks [3]. However, how to conduct enterprise-wide metadata management, which tasks are involved and which tools are suited, remains unclear.

We have conducted interviews with a globally active manufacturer, to gain insights into the current metadata management strategies and tools used in industry. Based on the case of the global manufacturer we examine metadata management challenges in practice and investigate whether these are solved by scientific research or existing tools. As a result this paper delivers four main contributions: (1) We present metadata management goals and their solution-oriented approaches in practice, based on the case of a large industrial enterprise; (2) We provide insight into current metadata management challenges; (3) We evaluate whether these challenges are sufficiently addressed in scientific literature or by tools from industry or research, and (4) based on this evaluation, we investigate research gaps in metadata management.

The remainder of this paper is structured as follows: The subsequent section illustrates the manufacturer’s metadata management goals and the section thereafter presents their approaches for addressing these. Within the next section the manufacturer’s challenges in metadata management are highlighted together with associated literature and tool coverage. The second to last section presents research gaps in metadata management and the last section concludes this paper.

The Industry Case and Metadata Management Goals

Metadata management is generally conducted to support data management. Hence, the data management goal needs to be clear in order to set up metadata management. The manufacturer is engaged in various sectors, such as the mobility or industrial sector and operates a global manufacturing network for mass and individual production. In this context, a lot of data is collected and stored, e.g., through the internet of things devices, enterprise resource planning systems and manufacturing execution systems. The manufacturer is pursuing the business strategy to become a data-driven industry 4.0 company. In the context of becoming more data-driven, the manufacturer has implemented novel technologies and concepts such as data lakes, storage repositories for data at scale and analytical purposes [2], and aims to establish an environment in which data can be shared freely and efficiently within the enterprise. With this goal the manufacturer aims to drive innovative data utilization and leverage more data value. For example, the ability to perform more data analysis supports realizing industry 4.0 use cases like predictive maintenance or real-time manufacturing quality analysis [10].

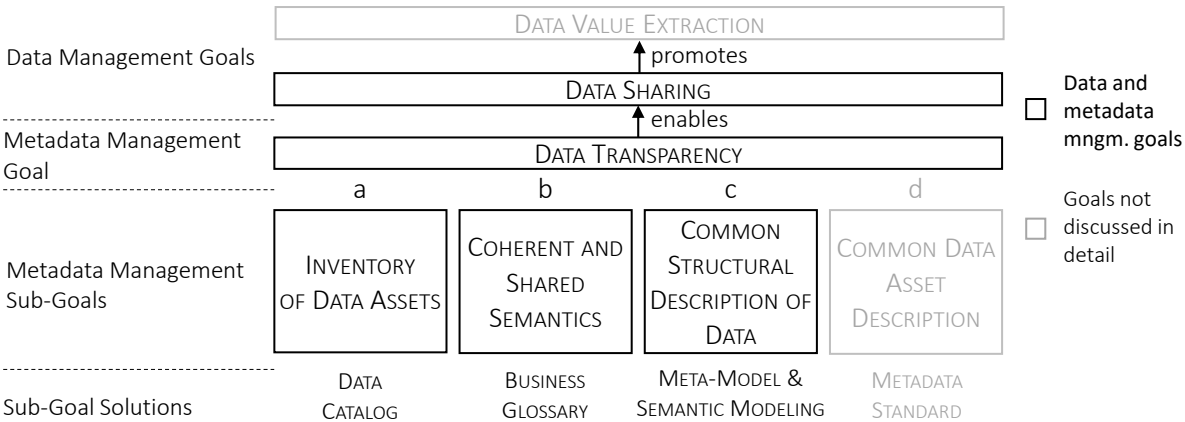


Figure 1. The manufacturer’s data management and metadata management goals with sub-goals and corresponding solutions.

Data sharing enables a broader use of data and thus also promotes extracting the data's value. The sharing of data entails one party provisioning data and others accessing and using it. Sharing data 'freely' means the data will be made available to many of its employees and all types of users. To support all user types, the provisioning, access and usage of data should be enabled through self-service functionality without requiring the involvement of IT specialists. Doing so 'efficiently' signifies the process of sharing data should involve little effort. Nonetheless, the data's compliance and trust must be retained.

One prerequisite for sharing data freely and efficiently is *data transparency*. Zhu defines information transparency as the degree to which information is visible and accessible [11]. By our understanding, data transparency involves the ability to find, understand and access the data, which substantially aids data sharing. Data transparency can be ensured through metadata management by, e.g., acquiring sufficient documentation. Hence, the main metadata management goal is to establish data transparency, for which the manufacturer compiled four metadata management sub-goals, as illustrated in Figure 1. In the following we only discuss sub-goals (a) to (c), as the manufacturer has not yet dealt with sub-goal (d) in detail.

- Sub-goal (a) involves taking *inventory of data assets* across the enterprise. To exploit data value, it must first be known what data are available. As there is a multitude of storage systems, like data lakes and ERP systems, it is infeasible for single employees to retain an overview of existent data. For instance, a data scientist searching for customer data would have to know all systems which contain such data. Therefore, an enterprise-wide inventory of data assets is required.
- The second sub-goal, (b), entails the introduction of *coherent and shared semantics* throughout the entire enterprise. By ensuring that people use and understand the same concepts, data can be described and understood without misunderstandings. For example, the concepts 'customer' and 'consumer' are often used interchangeably, although a customer purchases goods and a consumer is the end user of goods, so they do not share the same semantics. Some also have several meanings, like the concept 'right', which can mean 'correct' or refer to a direction. Hence, coherent and shared semantics are needed to clarify data's meaning and avoid misunderstandings.
- Sub-goal (c) aims at establishing a *common structural documentation of data assets*, i.e., a modeling standard. There is a multitude of data models within the enterprise. These are modeled with varying tools, abstraction levels and documentation standards, for instance, as an entity relationship diagram created with tools similar to Visual Paradigm¹. This makes it difficult to access, understand, integrate and reuse these data models. Therefore, a modeling standard based on, e.g., a common meta-model, must be established to foster model access, understanding, and sharing.

Data findability and consequently accessibility are improved through an inventory. Coherent and shared semantics together with a modeling standard and a standardized data asset description facilitate the understanding required for data transparency. Together these four sub-goals provide the data transparency required for data sharing.

Practical Approaches for Addressing the Metadata Management Sub-Goals

Having discussed the metadata management goals, this section illustrates the manufacturer's approach for reaching the sub-goals (a) to (c) with the solutions listed in Figure 1.

A Data Catalog for Taking Inventory of Data Assets. Sub-goal (a) is attained through the introduction of a commercial *data catalog*. A data catalog is a metadata management

¹ <https://www.visual-paradigm.com/>

tool, which is essentially a data inventory with documentation on registered data sources and data assets [12]. Alation² or the Collibra Data Catalog³ are examples of such tools. The documentation ranges from business metadata such as the content description ‘customer purchases’, over technical metadata on, e.g., the data type ‘String’, to operational metadata describing for instance, the data’s access history. The catalog provides a single interface for an enterprise-wide search of data. Beyond that, it provides functionality like enrichment, collaboration and governance features through, e.g., tags, commenting and user roles [12]. With search, documentation and other features, the catalog enables findability and understandability. The data scientist, for instance, can find customer data through the search and ascertain whether it fits their needs through the documentation.

Coherent and Shared Semantics by Means of a Business Glossary. Coherent and shared semantics, which constitute sub-goal (b), are established by compiling a *business glossary*. A business glossary specifies business terms and their definitions together with term relations for all business relevant concepts, such as a customer to product relation [1]. A business glossary tool such as erwin Data Literacy⁴ can be used and embedded into the overall application landscape, so the terms do not merely serve as documentation but can be reused in other applications such as an enterprise-knowledge graph [13] or data models. For instance, a data model with a customer entity would refer to the corresponding glossary entry, thereby clarifying what exactly the entity represents. To establish enterprise wide acceptance and trust in the terms, business term management is executed. This entails governing the terms with, e.g., responsibilities and defining term maturity for signifying standardization levels. Hence, introducing a business glossary and conducting business term management is one step to gaining enterprise-wide shared semantics and thus, a basis for shared understanding of data assets.

A Meta-Model and Semantic Modeling for Shared Structural Descriptions. Sub-goal (c) involves creating a modeling standard and is done by introducing a *meta-model* for standardizing the modeling approach, and a tool set for *semantic modeling* [14]. The meta-model differentiates between model abstraction layers like a business-object layer and domain layer. The more abstract business-object layer may, for instance, describe a ‘machine’ object and the domain layer a ‘bench drill’. The higher abstraction layers facilitate the integration of models, which results in an enterprise-wide knowledge graph and support insights into cross-functional usage of business objects. The integration is done through semantic modeling, by connecting more specific model’s instances to the more abstract instances. For this, a semantic modelling tool set has to be set up, for example with Protégé⁵, and integrated with the glossary, enabling users to search, explore and access semantic models based on business terms. Through the meta-model, semantic modeling tool set and integration with the business glossary, both enterprise-wide model understandability and findability are improved. The data scientist can now, for example, understand and find all data models which contain a customer entity.

With the introduction and integration of the listed tools a metadata management tool-landscape is created, which facilitates data transparency.

Challenges in Practice: A Literature and Tool Review

The manufacturer encountered several metadata-related challenges in the process of achieving their metadata management goal. We focus on three key challenges which we consider to be most relevant for research in metadata management. These pertain to: (1) *metadata management for data lakes*, (2) the *selection and composition of metadata management tool types* and (3) the implementation of *easy data provisioning, access and*

² <https://www.alation.com/>

³ <https://www.collibra.com/data-catalog>

⁴ <https://erwin.com/products/erwin-data-literacy/>

⁵ <https://protege.stanford.edu/>

use through data marketplaces. As depicted in Figure 2, challenge one occurred irrespective of the sub-goals and challenge two throughout all the sub-goals. While implementing the sub-goals it was determined that the aspects of data access, usage and provisioning necessary for data sharing and value extraction are not yet sufficiently supported and therefore, this was identified as challenge three and as a missing sub-goal. For each challenge, related literature and tools are examined to find either solutions or research gaps. Although the challenges originate from the manufacturer, we evaluate these independent of the company, so the results are representative and not company-specific.

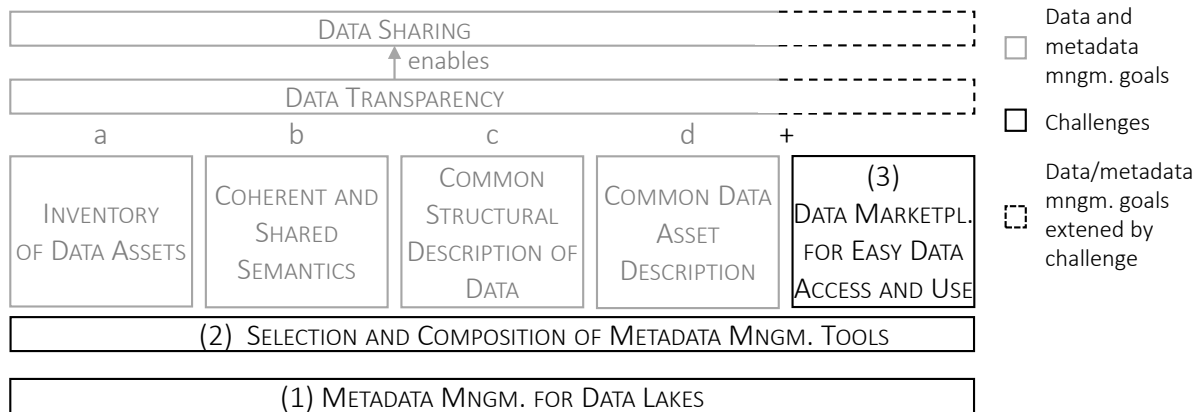


Figure 2. Extract of Challenges Arisen While Implementing the Data Sharing Goal

Challenge 1: Metadata Management for Data Lakes

A data lake is a storage repository for data at scale, which can incorporate data from heterogeneous sources, in its raw format, in various structures and without an overall schema [2], [6]. Data lakes can support the goal of making data available, e.g., by removing data silos [6] and exploiting its value. Metadata management is a critical success factor in data lakes, as it prevents these from turning into a data swamp, an inoperable data lake holding data that is not fit for use [5].

Challenges in Conducting Metadata Management for Data Lakes

Although it is known that metadata management is required in data lakes, it is not clear which combination of metadata tasks are sufficient to prevent the transition to a data swamp. There is a wide range of tasks which are supported by metadata management, to name a few: implementing governance specifications [4], data quality management and query processing [6]. For some of these tasks there is no sufficient insight as to how the corresponding metadata management needs to be implemented for a data lake. This concerns questions such as what metadata needs to be collected, what tools, protocols, and standards are needed for this task and do these exist, what do the processes look like and who is responsible, or also, how is the collected metadata integrated into the enterprise-wide landscape, but especially, how are these aspects different in data lakes. For example, these questions arise for metadata management to support data quality management within data lakes. Lack of clarity about essential tasks and how some of these have to be conducted in a data lake also leads to the challenge of designing and implementing a metadata management system or system landscape for managing the data lake.

Current State in Metadata Management for Data Lakes

A variety of scientific articles address the issue of data swamps and the required metadata management. The data lake system Constance and metadata management system GEMMS counteract a data swamp by collecting semantic and structural metadata, e.g., as annotations and schemata [6], [7]. The authors of Constance state that metadata management is essential for data reasoning, query processing and data quality management [6]. Yet it is not explained what metadata is collected or used for within data

quality management. This is also not clarified in the system CLAMS, designed to bring quality to data lakes [15]. This is an example of a specific metadata management task which implementation is unclear and in addition is neglected by other systems such as GEMMS, which in contrast to Constance does not address data quality management. Sawadogo and Darmont define six mandatory tasks for metadata management systems ranging from semantic enrichment and data indexing over link generations and data polymorphism to data versioning and usage tracking [5]. However, a system containing all of these has not been implemented and hence not tested yet, and they do not seem to include the same scope of structural metadata, i.e. schemata, as previously mentioned systems nor topics such as data quality. According to Gröger and Hoos, metadata management must support self-service and governance in the lake [4]. There are multiple systems for managing data lakes like GOODS [16], Ground [17] and CoreKG [18] which implement a variety of data and metadata management tasks. Due to the divergences it remains unclear which tasks are strictly necessary to prevent a data swamp, how some of the metadata tasks differ when conducted in data lakes, which tasks are best suited to be integrated in an overall metadata management system for data lakes and which should be outsourced to a specialized system, like a system for data quality management.

Challenge 2: The Multitude of Metadata Management Tool Types

As described in the section on practical approaches the manufacturer is building a metadata management tool-landscape for achieving data transparency and data sharing. However, the selection and combination of tools has become increasingly difficult. In the following, we focus on tool types so the examinations are not dependent on individual tools and their specific characteristics and, therefore, more general observations and statements can be made.

Challenges in Differentiating and Combining Metadata Management Tool Types

There are many different tool types, such as business glossaries, data catalogs, and data marketplaces. The scope of their functionality is unclear and overlaps, for example, data marketplaces, which are platforms for trading data [19], also contain cataloging functionality [20]. Furthermore, the commercial tools are evolving by integrating new functionality, which is sometimes common for other tool types. For example, some data catalog products like the Informatica Catalog⁶ have added data preparation features which is also a central aspect in data marketplaces [19]. Moreover, some vendors have rebranded their metadata management tools to, e.g., data catalogs, to capitalize customer interest [12]. There are also new tool types, which might simply be a synonym of another tool type, for instance, the emerging tool type called data hub might be a data marketplace. To identify the suitable tool types for comprehensive metadata management, a categorization of these tools is in order. In this context, information on their functional scope, characteristics, synonyms and subtypes are of interest.

As the tool types' functionality overlap, it is not clear which types to combine so they complement each other. Hence, the manufacturer needs an overview of the tool types' functional building blocks. For instance, data catalogs often contain a business glossary and data marketplaces contain data catalogs. Insights into compositions of these building blocks and how these work together are needed as a guideline for the erection of a compatible and comprehensive metadata management tool-landscape.

In closing, the manufacturer is struggling to select a set of tools, which enable comprehensive metadata management as these are not clearly differentiated, it is not clear which building blocks they contain, which are required and how these work together.

⁶ <https://www.informatica.com/de/products/data-catalog.html>

Current State of Metadata Management Tool Types

There are variously detailed definitions and lists of functionality per tool type in literature. For instance, Zaidi et al. provide a definition on data catalogs [12] and Meisel and Spiekermann supply one for data marketplaces with a list of their functionality [19]. Some sources also specify sub-types of tools, such as Zaidi et al. on data catalogs [12], or Lange et al. [21] and Meisel and Spiekermann [19] for data marketplaces. However, the information on the tools has to be assembled from a multitude scientific articles, white papers and tool webpages and then be compared and evaluated. For example, Bhardwaj et al. define a tool called data hub, which is suspiciously similar to a data marketplace [22].

There are very few comparisons of metadata management tool types in scientific literature. Gröger and Hoos, differentiate the data dictionary, data catalog and data lake management platform through a high-level description [4]. There are various consulting blog articles on tools such as [23] which differentiates business glossaries from data dictionaries. Gartner published a list of metadata management tools by various vendors and presents the vendors' strengths and cautions, not, however, those of the tools [3]. Therefore, there is no comprehensive categorization and differentiation of metadata management tools to help in the tool selection.

In terms of functional building blocks, Zaidi et al. mention that a glossary can be contained in a catalog [12], Wells' framework of a data marketplace contains a data catalog [20] and based on Gröger and Hoos the data lake management platform also has a data catalog [4]. Hence, there is information spread throughout scientific articles from which one can laboriously deduce which tools possibly contain or complement each other. Some vendors' tool suites such as IBM's InfoSphere platform⁷ can be used as a reference for combining tools, but these often focus on other topics like data management and not specifically metadata management. We have not found a comprehensive overview of tool types or their building blocks. It follows that there is no proposed building block assembly for comprehensive metadata management.

Challenge 3: Implementing Easy Data Provisioning, Access and Use

Currently, the manufacturer's metadata management tools support finding and understanding, not however, provisioning and accessing data. For this, additional metadata is required, such as metadata on the data owner or technical metadata to build an automated pipeline for provisioning. Having found the data, a user must at present contact the data owner and organize data access. If not available, they must set up the required environment to use it, e.g., for analysis. This process is time-consuming and challenging, especially for non-technical users. Therefore, the manufacturer needs additional metadata-driven tooling as depicted in Figure 2. To enable efficient data sharing, they need a platform that builds on the established metadata management tools and through which data provisioning and access is offered compliantly, via self-service with, e.g., data preparation functionality. Data marketplaces, metadata-driven platforms for sharing data, partially offer such functionality [19]. Optimally, an analysis environment can be obtained with courses, e.g., to learn analytics. For instance, a virtual machine with tools like Tableau⁸ could be offered with the data. This would save technical users time and strongly support non-technical users, such as marketing specialists.

⁷ <https://www.ibm.com/de-de/analytics/information-server>

⁸ <https://www.tableau.com/>

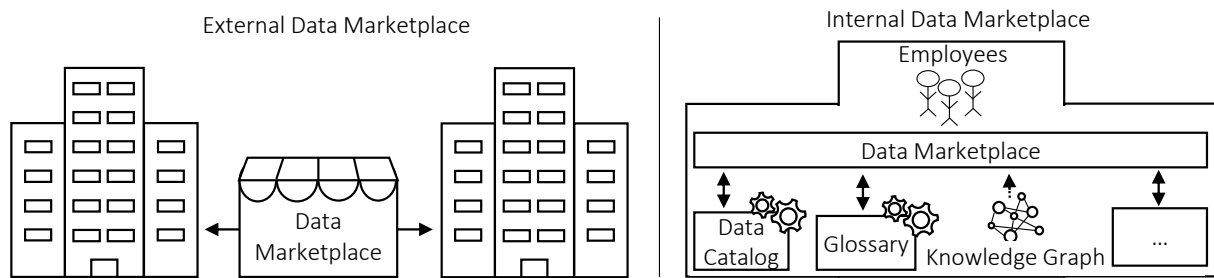


Figure 3. Data Marketplaces for Exchanging Data Between or Within an Enterprise

Challenges in Realizing Efficient Data Sharing Through Data Marketplaces

As shown in Figure 3, the data marketplace concept is initially designed for exchanging data between and not within an enterprise [19]. For the internal use, the marketplace should integrate with the existing tool landscape, tools such as data catalogs so it reuses collected metadata and does not duplicate functionality. But the marketplace tools are designed as standalone solutions and contain tools such as a data catalog. Besides, a marketplace for internal use should offer most of the enterprises data, as opposed to only specifically uploaded or selected data like the external marketplaces. In this context it should not only be able to store the data but also to refer to it, e.g., in the data lake. Furthermore, in the externally used marketplaces, monetary aspects serve as the main motivation to supply data [24]. Internally, the employees need other functionality driving them to supply their data, as monetarization promotes market competition, which is not desired within the enterprise. Recapitulatory, the manufacturer is challenged by extending its tool-landscape to support easy data provisioning, access and usability via self-service functionality through internal data marketplaces.

Current State in Applying Data Marketplaces for Enterprise-Internal use

In terms of acquiring data via self-service, Wells suggests that data marketplace receive a data storefront for shopping experiences like on Amazon [20]. To support data analytics, features for data preparation, curation etc. are envisioned [19], [20]. Moreover, Meisel and Spiekermann define external service providers as users of data marketplaces and suggest that these can provide analytical and infrastructure services as well as courses and services to support non-technical users [19]. However, the conceptual and implementation details of these analytics related features are not described in detail.

Using data marketplaces for an enterprise-internal exchange of data is suggested by Wells [20] and by Tata Consultancy Services [25]. Both provide a framework with functionality, but do not elaborate in detail. Using Commercial marketplace products like the Data Intelligence Hub⁹ is difficult, as most are explicitly built as exchange platforms between companies, are often hosted as external platforms, and are therefore unsuited for storing or connecting all company data. Some marketplaces like Chordant¹⁰ can be set up for private use, but Chordant is specialized for specific data on, e.g., autonomous mobility and is therefore not suited for making all company data available. Furthermore, commercial marketplaces have their own tool ecosystem. They double functionality and are not designed to fit into an existing metadata management landscape. Also, marketplaces like Chordant store the data and can not necessarily refer to data in other systems as required. In literature architectural aspects of data marketplaces are addressed in contexts like the trading of cloud of things resources [26], marketplaces in the IoT ecosystem [27] or trusted data marketplaces [28], but not for internal marketplaces. Finally, a multitude of monetization models have been discussed in various sources such as [29], but have not been verified in the context of internal marketplaces. Concluding, commercial marketplaces are so far

⁹ <https://dih.telekom.net/>

¹⁰ <https://www.chordant.io/solutions>

unsuitable for the internal use and there are no detailed concepts, solutions with architectural proposals and detailed functional scopes for an internal data marketplace in literature.

Recapitulatory, the manufacturer had difficulties in implementing metadata management for data lakes, differentiating and combining metadata management tools and finally, enabling easy data provisioning, access and use of data through data marketplaces. For all the challenges, literature and tools from industry and research were reviewed and scanned for solutions, but none were found that fully solve the issues.

Identified Research Gaps

Based on the literature and tool review conducted in the previous section, we have identified three research gaps. As described in the section on challenge 1, it remains unclear which metadata management tasks are strictly necessary to prevent a data swamp, how to execute some of these tasks and how these differ when conducted in data lakes. It is also unclear which of these tasks need to be integrated into a data lake specific metadata management system. Hence, the topic *metadata management for data lakes* contains a research gap.

Second, as shown in the section on challenge 2, there is no work on the metadata management tool's building blocks and how these can be assembled to a comprehensive metadata management landscape. As more and more companies are faced with the challenge of building such a tool landscape, this is a significant topic. Hence, *the categorization and composition of metadata management tools* also constitutes a relevant research gap.

Lastly, the use of *data marketplaces within an enterprise* to foster the internal exchange of data and for these to serve enterprise-internal needs, has only been suggested, but not sufficiently explored, and therefore constitutes research gap three.

Conclusion

Metadata management has evolved in recent years and now presents a new challenge for companies. In this context, interviews were conducted with a globally active manufacturer to find out how metadata management is implemented in practice today, what challenges companies are faced with and whether these present research gaps.

It was established, that the manufacturer's overall goal is to share its data freely and efficiently throughout the enterprise. To achieve this, the manufacturer defines data transparency as a prerequisite and derives four metadata management sub-goals. These are namely, taking inventory of data assets, the creation of coherent and shared semantics, the introduction of a common structural and common data asset description. While implementing these sub-goals, the manufacturer encountered several challenges, three of which, we identified as research gaps. The research gaps include: metadata management for data lakes, lacking categorizations and information on compositions of metadata management tools for comprehensive metadata management, and finally, absent research on the use of data marketplaces within an enterprise. Having defined research gaps, the groundwork is laid for further scientific research on metadata management, which will later enable exploiting data value fully and foster innovative data utilization while remaining compliant and within a legal framework.

References

- [1] DAMA International, *DAMA-DMBOK: Data Management Body of Knowledge*. Technics Publications, 2017.
- [2] C. Giebler, C. Gröger, E. Hoos, H. Schwarz, and B. Mitschang, "Leveraging the Data Lake: Current State and Challenges," in *Proc. of the 21st International Conference on*

Big Data Analytics and Knowledge Discovery (DaWaK), Springer, Cham, 2019.
https://doi.org/10.1007/978-3-030-27520-4_13

- [3] G. De Simoni, M. Beyer, and A. Jain, "Magic Quadrant for Metadata Management Solutions," *Gartner*, 2019.
- [4] C. Gröger and E. Hoos, "Ganzheitliches Metadatenmanagement im Data Lake: Anforderungen, IT-Werkzeuge und Herausforderungen in der Praxis," in *Proc. of the 18. Fachtagung für Datenbanksysteme für Business, Technologie und Web (BTW)*, 2019. <https://doi.org/10.18420/btw2019-26>
- [5] P. Sawadogo and J. Darmont, "On data lake architectures and metadata management," *J. Intell. Inf. Syst.*, 2020. <https://doi.org/10.1007/s10844-020-00608-7>
- [6] R. Hai, S. Geisler, and C. Quix, "Constance: An intelligent data lake system," in *Proc. of the 2016 International Conference on Management of Data (SIGMOD)*, 2016. <https://doi.org/10.1145/2882903.2899389>
- [7] C. Quix, R. Hai, and I. Vatov, "Metadata Extraction and Management in Data Lakes With GEMMS," *Complex Syst. Informatics Model. Q.*, no. 9, 2016. <https://doi.org/10.7250/csimq.2016-9.04>
- [8] R. Eichler, C. Giebler, C. Gröger, H. Schwarz, and B. Mitschang, "Handle - a generic metadata model for data lakes," in *Proc. of the 22nd International Conference on Big Data Analytics and Knowledge Discovery (DaWaK)*, 2020. https://doi.org/10.1007/978-3-030-59065-9_7
- [9] C. Giebler, C. Gröger, E. Hoos, R. Eichler, H. Schwarz, and B. Mitschang, "The Data Lake Architecture Framework: A Foundation for Building a Comprehensive Data Lake Architecture," in *Proc. der 19. Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW)*, 2021.
- [10] C. Gröger, "Building an Industry 4.0 Analytics Platform," *Datenbank-Spektrum*, vol. 18, no. 1, 2018. <https://doi.org/10.1007/s13222-018-0273-1>
- [11] K. Zhu, "Information Transparency in Electronic Marketplaces: Why Data Transparency May Hinder the Adoption of B2B Exchanges," *Electron. Mark.*, vol. 12, no. 2, 2002. <https://doi.org/10.1080/10196780252844535>
- [12] E. Zaidi, G. De Simoni, R. Edjlali, and A. D. Duncan, "Data Catalogs Are the New Black in Data Management and Analytics," *Gartner*, 2017.
- [13] L. Ehrlinger and W. Wöß, "Towards a definition of knowledge graphs," *Semant. (Posters, Demos, SuCCESS)*, vol. 48, 2016.
- [14] Y. Svetashova, S. Schmid, and A. Harth, "Towards semantic model extensibility in interoperable IoT data exchange platforms," in *Proc. of the 2018 Global Internet of Things Summit (GloTS)*, 2018. <https://doi.org/10.1109/GIOTS.2018.8534561>
- [15] M. Farid, A. Roatis, I. F. Ilyas, H. F. Hoffmann, and X. Chu, "CLAMS: Bringing quality to data lakes," in *Proc. of the 2016 International Conference on Management of Data (SIGMOD)*, 2016. <https://doi.org/10.1145/2882903.2899391>
- [16] A. Halevy *et al.*, "Goods : Organizing Google 's Datasets," in *Proc. of the 2016 International Conference on Management of Data (SIGMOD)*, 2016.
- [17] J. M. Hellerstein *et al.*, "Ground : A Data Context Service," in *Proc. of the 8th Biennial Conference on Innovative Data Systems Research (CIDR)*, 2017.
- [18] A. Beheshti, B. Benatallah, R. Nouri, and A. Tabebordbar, "CoreKG: A Knowledge Lake Service," *Proc. VLDB Endow.*, vol. 11, no. 12, Aug. 2018. <https://doi.org/10.14778/3229863.3236230>
- [19] L. Meisel and M. Spiekermann, "Datenmarktplätze - Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy," *Fraunhofer ISST*, 2019.

- [20] D. Wells, "The Rise of the Data Marketplace: Data as a Service," *Eckerson Gr.*, 2017.
- [21] J. Lange, F. Stahl, and G. Vossen, "Datenmarktplätze in verschiedenen Forschungsdisziplinen: Eine Übersicht," *Informatik-Spektrum*, vol. 41, no. 3, 2018. <https://doi.org/10.1007/s00287-017-1044-3>
- [22] A. Bhardwaj *et al.*, "Collaborative data analytics with DataHub," *Proc. VLDB Endowment*, vol. 8, no. 12, 2015. <https://doi.org/10.14778/2824032.2824100>
- [23] R. Lutton, "Data Management 20/20: Business Glossary Best Practices – TDAN.com," 2019. [Online]. Available: <https://tdan.com/data-management-2020-business-glossary-best-practices/25216>. [Accessed: 02-Nov-2020].
- [24] R. C. Fernandez, P. Subramaniam, and M. J. Franklin, "Data market platforms: Trading data assets to solve data problems," *Proc. VLDB Endow.*, vol. 13, no. 12, 2020.
- [25] S. Saxena, "Enterprise Data Marketplace: Democratizing Data within Organizations," *Tata Consult. Serv.*, 2018.
- [26] A. S. Alrawahi, K. Lee, and A. Lotfi, "AMACoT: A Marketplace Architecture for Trading Cloud of Things Resources," *IEEE Internet Things J.*, vol. 7, no. 3, 2019. <https://doi.org/10.1109/JIOT.2019.2957441>
- [27] S. Schmid *et al.*, "An architecture for interoperable IoT Ecosystems," in *Proc. of the 2nd International Workshop on Interoperability and Open-Source Solutions for the Internet of Things (InterOSS-IoT)*, 2016. https://doi.org/10.1007/978-3-319-56877-5_3
- [28] D. Roman and G. Stefano, "Towards a reference architecture for trusted data marketplaces: The credit scoring perspective," in *Proc. of the 2nd International Conference on Open and Big Data (OBD)*, 2016. <https://doi.org/10.1109/OBD.2016.21>
- [29] M. Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods," *Intereconomics*, vol. 54, no. 4, 2019. <https://doi.org/10.1007/s10272-019-0826-z>