Datenbank-Spektrum (2023) 23:89–96
https://doi.org/10.1007/s13222-023-00445-2

SCHWERPUNKTBEITRAG

Data Catalogs in the Enterprise: Applications and Integration

Nils Jahnke[1] · **Boris Otto**[1,2]

Received: 3 March 2023 / Accepted: 31 May 2023 / Published online: 21 June 2023

Abstract
Despite investing heavily in data-related technology and human resources, enterprises are still struggling to derive value from data. To foster data value creation and move toward a data-driven enterprise, adequate data management and data governance practices are fundamental. To support these practices, organizations are building (meta)data management landscapes by combining different tools. Data catalogs are a central part of these landscapes as they enable an overview of available data assets and their characteristics. To deliver their highest value, data catalogs need to be integrated with existing data sources and other data management tools. However, enterprises struggle with data catalog integration because (a) not all data catalog application types foster enterprise-wide data management and data governance alike, and (b) several technical characteristics of data catalog integration remain unclear. These include the supported data sources, data catalog federation, and ways to provision data access. To tackle these challenges, this paper first develops a typology of data catalog applications in the enterprise context. Based on a review of the academic literature and an analysis of data catalog offerings, it identifies four enterprise-internal and three cross-enterprise classes of data catalog applications. Second, an in-depth analysis of 51 data catalog offerings that foster enterprise-wide metadata management examines key characteristics of the technical integration of data catalogs.

Keywords Data Catalog · Data Management · Metadata Management · Typology · Integration

1 Introduction

The relevance of data as an organizational asset with intrinsic value is widely accepted. More data are produced and stored by organizations every year [25]. However, value from data is only created once they are used for operational excellence, product innovation, improved business models, or monetized in the data economy. Therefore, data must be transformed, enriched, and contextualized to create actionable information [18].

To realize the promise of data and analytics for competitive advantage, organizations steadily increase their investments in technology and people. Yet, improvements in data culture, data value creation, and innovation capability remain limited [3]. Enterprises continue to struggle in areas such as data acquisition, data enablement, or data compliance [15, 16, 21, 26].

(Meta)data management and data governance are essential means to address these challenges and to improve data usage and thus firm performance [7, 24]. To implement and support these activities, data catalogs (DCs) play an important role as they (a) empower users to work with data; (b) make data-related issues visible; (c) reduce data preparation time and (d) promote compliant data handling and usage [1, 8, 28]. For a holistic metadata management approach, DCs need to be integrated into the existing enterprise data ecosystem [14]. This includes the integration with upstream data sources, downstream analytics applications, and further tools for data curation as part of a metadata management landscape.

However, implementing DCs as part of a holistic metadata management landscape is currently challenging [19]. First, practitioners are faced with a vast array of commercial and open-source DC offerings that focus on different application areas with different goals. For example, not all

✉ Nils Jahnke
nils.jahnke@isst.fraunhofer.de

Boris Otto
boris.otto@isst.fraunhofer.de

[1] Data Business, Fraunhofer Institute for Software and Systems Engineering, Speicherstraße 6, 44147 Dortmund, Germany

[2] Chair for Industrial Information Management, TU Dortmund University, Joseph-von-Fraunhofer-Str. 2–4, 44227 Dortmund, Germany

DC applications support enterprise-wide metadata management [29]. As the spectrum of DC applications and their capabilities remain undefined, it is demanding for practitioners to select the right tools to build such a tool landscape [8]. Second, the successful technical integration of DCs depends on several factors including automatic data source integration, DC federation, and data access provisioning [14]. Yet, lacking clarity about these characteristics hampers the usage of DCs as fundamental components of metadata management landscapes.

To address the challenge of DC integration, the paper first develops a typology of DC applications in the enterprise context. Typologies allow for a reduction of the plethora of entities into a lesser number of classes with key attributes. The theory bases on an extensive survey of DC offerings and is enriched by an analysis of the scientific state of the art. Second, the paper discusses relevant issues for integrating DCs into metadata management landscapes by analyzing 51 DC offerings fostering enterprise-wide metadata management in greater detail. The examined characteristics include (a) deployment types; (b) DC connectors and integrations; (c) DC federation; (d) means to provide data access, and (e) further potential modules of metadata management landscapes. The remainder of the paper is structured as follows: Sect. 2 introduces DCs and typologies. Sect. 3 describes the research methodology. In Sect. 4, the authors present the developed typology of DC applications, followed by an analysis of the current state of practice in DC integration in Sect. 5. Sect. 6 summarizes the main findings and presents further research directions.

2 Background

2.1 Data Catalogs

DCs represent a relatively new data management tool. Interest in DCs is rising due to their enabling function for metadata management, data governance, and data democratization. Yet, no widely accepted definition of DCs exists [19]. As synthesis of existing DC definitions in research and practice [8, 12, 29] this paper defines DCs as follows: DCs are metadata management tools, that support the curation of data by providing capabilities to inventorize and discover data on an integrated platform, thus connecting data supply and demand. This highlights the platform character of DCs and underlines the fundamental purpose of supporting data inventory and data discovery. In this paper, DC offerings are DCs that are made available for purchase or under an open source license by a third party. DC applications are contexts in which DCs are implemented. In addition to data inventory and discovery, Labadie et al. [19] outline other capability groups (and capabilities) of DCs. These include data assessment (e.g., data usage, data quality, data profiling), data governance (workflows, roles and responsibilities, rules and policies), data collaboration (tagging, sharing, commenting), and analytics metadata (data stories, data application repository). The functional capabilities are facilitated by supporting capabilities like visualization, administration, and automation, often including artificial intelligence features.

Fig. 1 depicts the positioning of DCs in the enterprise data ecosystem. DCs can automatically ingest metadata from data sources through built-in integrations. If the DC runs in a different environment than the data sources, so-called metadata harvesters can be used to collect and transfer metadata to the main component. DCs integrate with further data management tools that may be part of a data management platform [4], enterprise systems, or DCs to leverage their specialized capabilities and enable enterprise-wide metadata management. Integration with business intelligence and data analysis tools allows metadata to be provided for model creation and analytical models to be listed in the DC to describe data usage.

Due to their extensive capabilities, many positive outcomes along the data lifecycle have motivated the use of DCs on an enterprise-wide scale. These include gaining visibility about data assets and potential data issues, reducing time spent searching for and evaluating data, and enabling

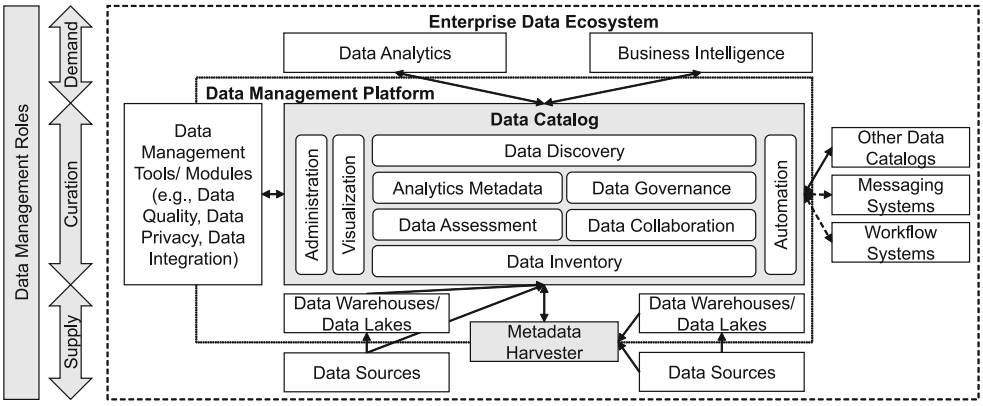

Fig. 1 DC functional model

more users to work with data while supporting governance and compliance efforts [17, 20, 27, 28].

While the general positioning and role of DCs in an integrated metadata management landscape is clear, organizations continue to face challenges in building concrete implementations. These challenges include mapping capabilities to the variety of metadata management tools, or integrating those tools with each other and with existing data sources [9, 14]. Accordingly, there is still a need for greater clarity about the types and core properties of DC applications in the enterprise sphere. In addition, more knowledge about the technical characteristics of DC integration is needed to enable better decisions when building metadata management landscapes.

2.2 Typologies

Typologies classify certain dimensions or characteristics of entities by abstracting the commonalities found in independent observations [11]. According to Nickerson et al. [22], the terms typology and taxonomy are often used interchangeably, with taxonomies usually referring to empirical and typologies to conceptually derived systems of classification. Since this paper addresses both the conceptual and the empirical perspectives, but essentially focuses on the types of DC applications, the resulting artifact will hereafter be referred to as typology. Typologies belong to the class of 'analytic theory' and therefore help to organize the body of knowledge and provide structure for the further analysis of a phenomenon [13]. They are especially valuable when describing a relatively new object of concern [11]. Accordingly, a typology of DC applications in the enterprise context can help to resolve the conceptual ambiguity surrounding this relatively new topic. The typology helps practitioners better understand the types and characteristics of DC applications, thereby aiding the planning and design process of metadata management landscapes.

3 Research Methodology

To develop a typology of DC applications this paper adapts the methodology of Nickerson et al. [22] in three iterations. First iteration consisted of an empirical analysis of DC offerings in the enterprise context (empirical to conceptual). DC offerings were identified based on a web search, the scanning of analysts reports, and the authors' experience in the DC context. Only the most comprehensive offer of a vendor was included as study subject in case multiple offers were available. Information about each offering was gathered from the vendor website, provided documentation, and tutorial videos. Second iteration consisted of an analysis of the state of the art in scientific DC literature (conceptual to empirical). Following search string was used to identify articles in the databases IEEE Xplore, AIS eLibrary, ACM Digital Library and SpringerLink:

> ("data catalog" OR "data catalogue" OR "metadata management solution" OR "metadata management tool") AND (enterprise OR business)

In the third iteration, conceptual and empirical findings were juxtaposed and the remaining knowledge gaps were filled (empirical to conceptual). In the end, 73 DC offerings[1] and 27 research papers were identified and analyzed. To answer the identified questions in the area of DC integration, 51 offerings fostering enterprise-wide metadata management were investigated in greater detail[2]. The survey results show only those observations that occurred more than once in the surveyed population to exclude outliers.

4 A Typology of Data Catalog Applications

Based on the research methodology described above, seven classes of DC applications were identified. These classes can be structured according to the following dimensions: (a) organizational area; (b) integration; (c) metadata management scope; (d) data management level, and (e) provider – consumer relationship as depicted in Table 1. The following section first describes the dimensions and their characteristics, and then discusses the identified classes of DC applications in greater detail.

Organizational area describes the subcontext of DC application. DCs can be applied for data curation within (intra-organizational) or across (inter-organizational) enterprises. In an intra-organizational setting, actors are usually represented by business users, whereas in an inter-organizational setting actors consist of organizations or principals acting on their behalf. Further, inter-organizational settings usually require stricter data protection regimes. *Integration* refers to the delivery of DC functionality to the respective environment. DCs can either be implemented as standalone solution or as module of a wider solution offering. For example, a DC can be seen as a modular part of a data marketplace [10]. The *scope* dimension describes the extent to which metadata management and data governance are supported by a DC application. Specific refers to the support of a specific environment (e.g., a cloud platform) or data application (e.g., business intelligence) by providing specifically fitted capabilities. Holistic refers to support for metadata management across all types, sources, and potential data applications in the organizational area. DC applica-

[1] For more details on the DC offerings included and the survey approach, see https://doi.org/10.24406/fordatis/257.

[2] See footnote 1.

Table 1 Typology of DC applications

Types	Dimensions				
	Organizational area	Integration	Metadata Management Scope	Data Management Level	Provider – Consumer Relationship
Enterprise Data Catalog	Intra-organizational	Stand-alone	Holistic	Metadata	Many-to-many
Context-specific Data Catalog	Intra-organizational	Stand-alone	Specific	Metadata	Many-to-many
Enterprise Data Management Platform	Intra-organizational	Module	Holistic	Data and Metadata	Many-to-many
Enterprise Data Marketplace	Intra-organizational	Module	Holistic	Metadata	Many-to-many
Data Spaces Data Catalog	Inter-organizational	Stand-alone	Holistic	Metadata	Many-to-many
Data Portal	Inter-organizational	Module	Holistic	Data and Metadata	One-to-many
Ecosystem Data Marketplace	Inter-organizational	Module	Specific	Both options possible	Many-to-many

tions can be further divided into those that primarily curate metadata and those that also have the ability to manage or deliver the actual data. This is characterized by the *data management level* dimension. Lastly, *provider-consumer relationship* refers to the amount of entities interacting with each other based on the DC application as a platform. Most applications foster many-to-many relationships of providers and consumers, while data portals are typically deployed by a single data provider to address the data needs of multiple consumers.

The first DC application class identified in the intra-organizational context are *Enterprise Data Catalogs*. They provide data cataloguing capabilities for all data-related roles in an organization and across departments or business units, enabling enterprise-wide data curation [19]. To this end, many data providers register the metadata of data assets from diverse systems, which can be leveraged by data consumers for different data applications. Enterprise Data Catalogs can be deployed as stand-alone solutions without the need to integrate with further data management tools.

In the *Context-specific Data Catalog* class, DCs only serve in a specific environment or for a specific data application. Examples of DCs that primarily serve a specific environment include AWS Glue or Cloudera Navigator. Both are limited to automated metadata ingestion from their respective cloud platform resources and focus on processes such as orchestration and ETL-processes. The survey also reveals DCs that provide data discovery capabilities only for a specific use case, such as data analytics (e.g., Tableau Catalog) or data privacy (e.g., Immuta Data Security Platform). While all of these offerings allow actors to leverage DC capabilities in a familiar environment, federation and interoperability are needed to avoid duplication of efforts and the creation of data silos.

The class of *Enterprise Data Marketplaces* was identified during the literature review phase. Researchers see the main function of Enterprise Data Marketplaces in providing data or data services brokerage features [10, 14]. To provide these capabilities, same researchers design Enterprise Data Marketplaces built on top of Enterprise Data Catalogs. However, this needs to be reconciled with the findings of Labadie et al. [19] who see brokerage functions such as data access requests as part of Enterprise Data Catalogs. Based on the analysis of real-world DC offerings, the authors of this paper argue that Enterprise Data Marketplaces are modular solutions that include an Enterprise Data Catalog module and an additional brokerage or marketplace component, which allows for the description and purchase of data products. Conversely, Enterprise Data Catalogs may support similar functionalities in a single module. Yet, this view is not represented in the overview of examined DC applications as no explicit commercial or open-source Enterprise Data Marketplace offering could be identified. However, an Enterprise Data Marketplace may be provided by implementing Enterprise Data Catalog and brokerage modules of Data Management Platform offerings.

Enterprise Data Management Platforms (EDMPs) support the management, storage, and distribution of data assets in the enterprise [4]. They are not tied to a specific context or use case. While the specific composition of EDMPs is up to each implementation, they typically consist of several modules including Enterprise Data Catalogs or Enterprise Data Marketplaces, to provide a listing of available data [14]. Data quality, data integration, or data privacy can also be modular components of EDMPs. EDMPs are deployed as an overarching layer that is agnostic to underlying databases, data lakes, or data warehouses. While they do access actual data for processes such as data integration, data quality, or data privacy assessments, they do not persist or replicate these data.

In the inter-organizational sphere, *Data Space Data Catalogs* enable the metadata-based inventory and discovery of data products to be shared between organizations in data spaces. Next to the functional semantic description of available data sources they allow for the definition and assessment of accessibility information and usage conditions for other organizations [6]. They do not hold the data itself as organizations want to preserve their data sovereignty and

therefore neglect the transfer of data to central platforms before the actual exchange. In general, Data Space Data Catalogs are agnostic to environments and use cases fostered by the data exchange.

Data Portals are leveraged to enable the reuse of data for societal or economic benefit. They are set up by the data providing entity and allow the discovery and access of data for multiple stakeholders including natural persons and enterprises. Data can be directly accessed or downloaded from the Data Portal. While most Data Portals base on CKAN, different modules are implemented as added benefit (e.g., data visualization) [23].

Lastly, *Ecosystem Data Marketplaces* match organizational data sellers and buyers and manage data exchanges and transactions [2]. In this sense, they can act as a trustee and manage access to data according to rules defined by the data seller. Metadata management, and thus the use of DC components, is seen as an essential module of Ecosystem Data Marketplaces. But, to fulfill the scope of data monetization, additional modules such as billing and invoicing are required. A comprehensive overview of Ecosystem Data Marketplaces is provided by Azcoitia et al. [2].

5 Insights into the architectural Integration of Data Catalogs

Having outlined the range of DC application classes suitable for enterprise-wide metadata management, this section examines how DCs falling into these classes are integrated into the enterprise data ecosystem from a technical perspective. Relevant questions for practitioners and research are investigated in seperate subsections[3].

5.1 Deployment Types

As depicted in Fig. 2, cloud, on-premises and hybrid deployments are supported by DCs. While cloud and on-premises deployments are the most provided options, each has its advantages and disadvantages, leading vendors to offer more than one option or advocate hybrid scenarios. On-premises deployments have drawbacks such as upfront infrastructure investment and user responsibility for updates, bug fixes, and backups. On the other hand, cloud deployments may not be able to access all on-premises systems because they are protected by restrictive firewalls or because of data security policies. Therefore, hybrid deployments are used as a solution to mitigate the disadvantages of both sides. Hybrid deployments use metadata harvesters to collect metadata from source systems and forward them

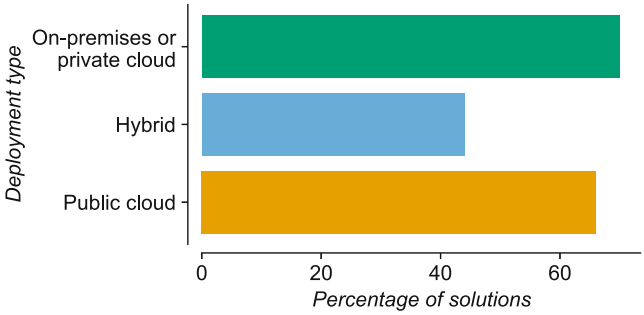

Fig. 2 Deployment Types of DCs

to the main cloud service. However, hybrid deployments are currently the least offered deployment type.

5.2 Connectors and Integrations

The ability to automatically ingest metadata from existing data systems into a DC is another important issue for DC integration [10]. To support holistic metadata management, batch as well as streaming systems need to be integrated [14]. Not only upstream data sources but also downstream data use in data analytics or business intelligence tools is of interest (see Fig. 3). In terms of upstream data sources, nearly all DC offerings support on-premises databases, cloud resources, and data lake or data warehousing systems to some degree. Metadata ingestion from data transformation or modelling platforms is often supported because it reveals important data lineage information. However, the support for metadata ingestion from streaming and messaging systems is currently limited. In addition, DC offerings currently lack comprehensive support for on-premises enterprise information systems such as ERP and CRM software, which are important parts of the data ecosystem in many enterprises. On the downstream

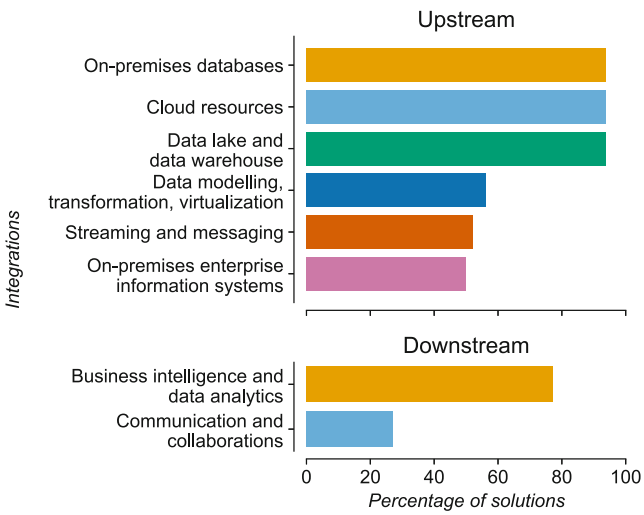

Fig. 3 Integrations of DCs

[3] For a comprehensive summary of the questions, approaches, and results for each subsection, see https://doi.org/10.24406/fordatis/257.

side, about 80 percent of DC offerings integrate with business intelligence tools. In addition, some DC offerings support the integration of communication and collaboration software.

5.3 Federation

Integrating information across multiple domains and organizations to enable data discovery is a challenging task [5]. In large enterprises, there may be multiple DCs in different domains, business units, or contexts that need to be integrated to avoid redundant maintenance [9]. Fig. 4 summarizes the findings regarding the potential for federating different DCs. More than half of the examined offerings provide the option to federate with at least one other DC offering. Currently, DC federation focuses on the integration of metadata from cloud environments by integrating with the respective context-specific DCs. Typically, DC application programming interfaces are leveraged for data exchange. Endeavours for metadata federation like Apache Atlas, Egeria, and Great Expectations are only supported by a some DC offerings.

5.4 Data Access

Currently, how to gain access through DCs is still an open research question [8]. Fig. 5 depicts the current status of data access via DCs. More than half of the surveyed offerings provide one of the data shopping functionalities mentioned by Eichler et al. [10], including service-access-management, transaction management, or subscription and order management. The actual process of data delivery depends on the DC offering. Following methods for data access provisioning were identified:

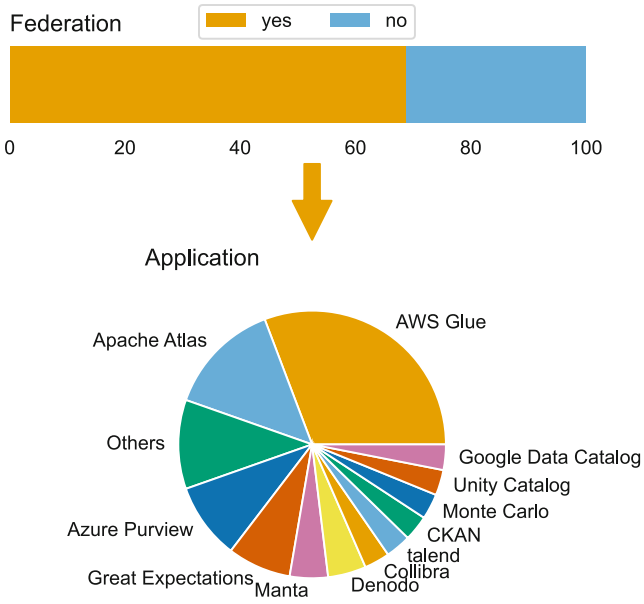

Fig. 4 Federation of DCs

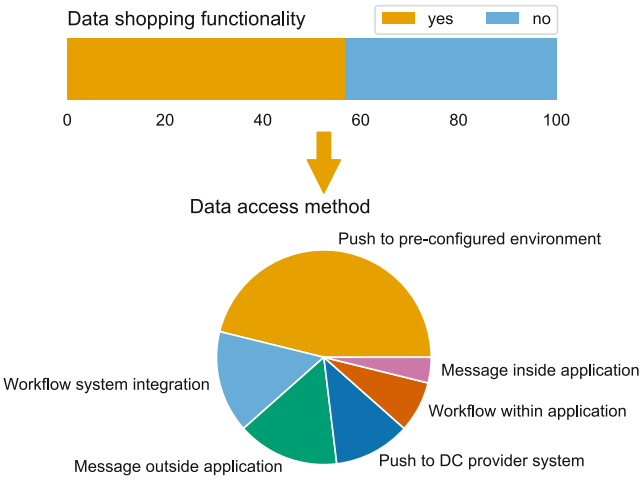

Fig. 5 Data Access in DCs

- *Push to pre-configured environment*: DC with access to actual data can push data to a previously set-up third-party environment with access for the user, e.g., a data warehouse.
- *Push to DC provider system*: DC can push data to an environment provided by the DC vendor.
- *Workflow system integration*: DC integrates with an external workflow or ticketing system and triggers a workflow, data access is provided manually, e.g., by creating a new account.
- *Workflow within application*: DC provides options to define and execute workflows internally, data access is provided manually.
- *Message outside application*: DC triggers a data access request message to a data owner or data steward in an integrated tool outside of the DC.
- *Message inside application*: DC generates and sends a data access request message to a data owner or data steward internally.

Conducting a push of data to pre-configured environments was identified as the most frequent mean for data provisioning. However, this way is mostly provided by DC offerings in the EDMP application class by leveraging additional modules. For metadata-only applications, integrating or providing workflow capabilities is the most common method of data provisioning.

5.5 Metadata Management Landscape

The last survey item aims to clarify the constituents of a comprehensive metadata management landscape [8]. Therefore, modules beyond the DC functionality (data discovery, data lineage, data governance and data collabo-

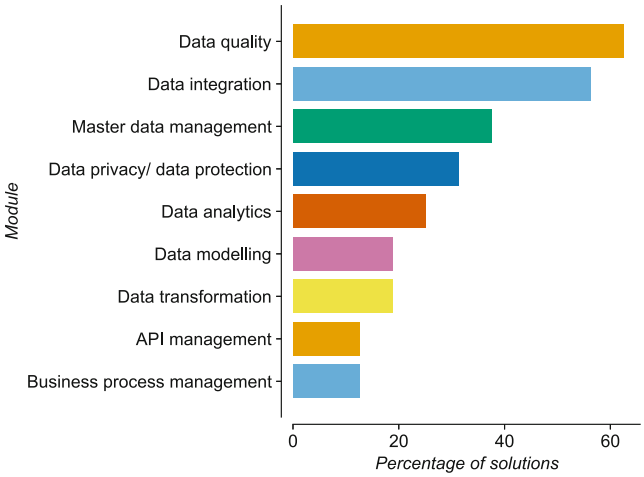

Fig. 6 EDMP modules

ration) of the 17 identified EDMPs were analyzed. The results in Fig. 6 show that data quality modules constitute the most common addition to DCs. Master data management modules are frequent parts of EDMPs as they enable to create so-called "golden records" of data objects used in many business processes. Data integration is another popular add-on module for making data available for analysis. At the metadata management level, data privacy and security modules complement DCs by ensuring that data are protected and handled in accordance with business and regulatory requirements. Some EDMPs also include modules for data analytics. In addition, data transformation and data modeling modules are often provided.

6 Conclusion

Despite ongoing high investments in data technologies and human resources, the promise of data-driven enterprises has yet to be realized. As a step toward improving data value creation and ultimately supporting data-driven businesses, DCs are being integrated with other metadata management tools into metadata management landscapes that support holistic metadata management and data governance across the enterprise. However, implementing DCs as part of such a metadata management landscape is challenging due to the variety of DC application classes and a general lack of understanding of DC integration.

To mitigate these challenges, this paper first develops a typology of DC applications in the enterprise context. Seven classes of DC applications could be identified and were structured along five dimensions. The typology helps practitioners to focus on the right DC application classes when building enterprise metadata management landscapes. It further supports future research by resolving the conceptual ambiguity around different classes of DC applications and their relationship. Additionally, important concerns for creating comprehensive metadata management landscapes were analyzed by a survey of 51 Enterprise Data Catalog and EDMP applications.

The study reveals several open challenges for research and practice. First, Enterprise Data Marketplaces seem to be a promising DC application class as they enable the description and provisioning of data in a more consumer-centric way. However, further conceptual and practical research is needed to clarify and demonstrate the capabilities and value-adds. Second, the current state of providing data access is unsatisfying as it demands high manual efforts on the data provider side. The automatic provisioning of data access based on pre-defined access conditions across all enterprise data sources therefore seems to be a promising research direction. Ultimately, more attention should be directed towards the development and implementation of methods and standards for the federation of metadata management tools as organizations move towards greater decentralization in data management.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Aikoh K, Isoda Y, Sugimoto K (2020) Data profiling method for metadata management. In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, pp 779–780 https://doi.org/10.1109/DSAA49011.2020.00113
2. Azcoitia SA, Laoutaris N (2022) A survey of data marketplaces and their business models. ACM Sigmod Rec 51(3):18–29. https://doi.org/10.1145/3572751.3572755
3. Bean R (2021) Why is it so hard to become a data-driven company? https://hbr.org/2021/02/why-is-it-so-hard-to-become-a-data-driven-company. Accessed 09.01.2023
4. Boch M, Gindl S, Barnett A et al (2022) A systematic review of data management platforms. In: Rocha A, Adeli H, Dzemyda G, al (eds) Information Systems and Technologies, Lecture Notes in Networks and Systems, vol 469. Springer, Cham, pp 15–24 https://doi.org/10.1007/978-3-031-04819-7_2
5. Bugbee K, Ramachandran R, Acharya A et al (2022) Selecting approaches for enabling enterprise data search: Nasa's science

mission directorate (smd) catalog. In: IGARSS 2022 – 2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Piscataway, pp 6836–6839 https://doi.org/10.1109/IGARSS46834.2022.9884711
6. Cirullies J, Schwede C (2021) On-demand shared digital twins – an information architectural model to create transparency in collaborative supply networks. In: Bui T (ed) Proceedings of the 54th Hawaii International Conference on System Sciences Hawaii International Conference on System Sciences, Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.202
7. Dinter B, Gluchowski P, Schieder C (2015) A stakeholder lens on metadata management in business intelligence and big data – results of an empirical investigation. Twenty-first Americas Conference on Information Systems.
8. Eichler R, Giebler C, Gröger C et al (2021) Enterprise-wide metadata management. Bus Inf Syst. https://doi.org/10.52825/bis.v1i.47
9. Eichler R, Gröger C, Hoos E et al (2022a) Data shopping — how an enterprise data marketplace supports data democratization in companies. In: de Weerdt J, Polyvyanyy A (eds) Intelligent Information Systems. Lecture Notes in Business Information Processing, vol 452. Springer, Cham, pp 19–26 https://doi.org/10.1007/978-3-031-07481-3_3
10. Eichler R, Gröger C, Hoos E et al (2022b) From data asset to data product – the role of the data provider in the enterprise data marketplace. In: Barzen J, Leymann F, Dustdar S (eds) Service-Oriented Computing. Communications in Computer and Information Science, vol 1603. Springer, Cham, pp 119–138 https://doi.org/10.1007/978-3-031-18304-1_7
11. Fawcett J, Downs FS (2016) The relationship of theory and research, 3rd edn. F. A. Davis Compagny, Philadelphia
12. Franklin M, Halevy A, Maier D (2005) From databases to dataspaces. ACM Sigmod Rec 34(4):27–33. https://doi.org/10.1145/1107499.1107502
13. Gregor S (2006) The nature of theory in information systems. MISQ 30(3):611. https://doi.org/10.2307/25148742
14. Gröger C (2021) There is no ai without data: Industry experiences on the data challenges of ai and call for a data ecosystem for industrial enterprises. Commun ACM 64(11):98–108. https://doi.org/10.1145/3448247
15. Habrat D (2020) Legal challenges of digitalization and automation in the context of industry 4.0. Procedia Manuf 51:938–942. https://doi.org/10.1016/j.promfg.2020.10.132
16. Harland T, Hocken C, Schröer T et al (2022) Towards a democratization of data in the context of industry 4.0. Sci 4(3):29. https://doi.org/10.3390/sci4030029
17. Hugh JW (2019) Update tutorial: big data analytics: concepts, technology, and applications. CAIS. https://doi.org/10.17705/1CAIS.04421
18. Koutroumpis P, Leiponen A, Thomas LDW (2020) Markets for data. Ind Corp Change 29(3):645–660. https://doi.org/10.1093/icc/dtaa002
19. Labadie C, Legner C, Eurich M et al (2020) Fair enough? enhancing the usage of enterprise data with data catalogs. In: Aier S, Guedria W (eds) 2020 IEEE 22nd Conference on Business Informatics. IEEE Computer Society, Conference Publishing Services, Los Alamitos, Washington, Tokyo, pp 201–210 https://doi.org/10.1109/CBI49978.2020.00029
20. Lefebvre H, Legner C, Fadler M (2021) Data democratization: toward a deeper understanding. ICIS 2021 Proceedings.
21. Lennerholt C, van Laere J, Söderström E (2018) Implementation challenges of self service business intelligence: A literature review. Proceedings of the 51st Hawaii International Conference on System Sciences, pp 5055–5063
22. Nickerson RC, Varshney U, Muntermann J (2013) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22(3):336–359. https://doi.org/10.1057/ejis.2012.26
23. Nikiforova A, McBride K (2021) Open government data portal usability: A user-centred usability analysis of 41 open government data portals. Telematics Inform 58:101–539. https://doi.org/10.1016/j.tele.2020.101539
24. Otto B (2011) A morphology of the organisation of data governance. ECIS 2011 Proceedings.
25. Reinsel D, Gantz J, Rydning J (2018) The digitization of the world: from edge to core
26. Roh Y, Heo G, Whang SE (2019) A survey on data collection for machine learning: a big data – ai integration perspective. IEEE Trans Knowl Data Eng 33(4):1328–1347. https://doi.org/10.1109/TKDE.2019.2946162
27. Samarasinghe S, Lokuge S (2022) Exploring the critical success factors for data democratization. ACIS 2022 Proceedings.
28. Schilling R, Aier S, Winter R et al (2020) Design dimensions for enterprise-wide data management: a chief data officer's journey. In: Bui T (ed) Proceedings of the 53rd Hawaii International Conference on System Sciences https://doi.org/10.24251/HICSS.2020.714
29. Zaidi E, de Simoni G, Edjlali R et al (2017) Data catalogs are the new black in data management and analytics