



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

---

# Data Shopping — How an Enterprise Data Marketplace Supports Data Democratization in Companies

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

---

© by Springer Nature

This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this contribution will be published in the conference proceedings of the:

**34th International Conference on Advanced Information Systems Engineering (CAiSE) Forum 2022**

The final authenticated version is available online at:  
[https://doi.org/10.1007/978-3-031-07481-3\\_3](https://doi.org/10.1007/978-3-031-07481-3_3)

# Data Shopping — How an Enterprise Data Marketplace Supports Data Democratization in Companies

Rebecca Eichler<sup>1</sup>, Christoph Gröger<sup>2</sup>, Eva Hoos<sup>2</sup>, Holger Schwarz<sup>1</sup>, and Bernhard Mitschang<sup>1</sup>

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

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

**Abstract.** To exploit the company’s data value, employees must be able to find, understand and access it. The process of making corporate data available to the majority of the company’s employees is referred to as data democratization. In this work, we present the current state and challenges of data democratization in companies, derived from a comprehensive literature study and expert interviews we conducted with a manufacturer. In this context a data consumer’s journey is presented that reflects the required steps, tool types and roles for finding, understanding and accessing data in addition to revealing three data democratization challenges. To address these challenges we propose the use of an enterprise data marketplace, a novel type of information system for sharing data within the company. We developed a prototype based on which a suitability assessment of a data marketplace yields an improved consumer journey and demonstrates that the marketplace addresses the data democratization challenges and consequently, shows that the marketplace is suited for realizing data democratization.

**Keywords:** Data Marketplace, Data Democratization, Data Sharing.

## 1 Introduction

The potential of a company’s data can only be exploited if its employees can find, access and use it for their respective use cases. However, it has been reported that 60-73% of data in the enterprises remains unused [1]. To address this issue, data democratization has become increasingly relevant [2].

The objective of data democratization is to empower and motivate the majority of company employees to find, understand, access, use and share data across the enterprise, in a secure and compliant way [2–4]. In this sense, Lefebvre et al. [2] define four data democratization dimensions, the first signifies enabling broader access to data and tools, the second, developing data-related and data-analytic skills, e.g., for data cleansing. The third entails the collaboration and knowledge sharing amongst employees, and the fourth comprises the promotion of data value. To assess the current state of data democratization based on these dimensions and investigate the democratization challenges, we conducted a literature study and expert interviews with a

globally active manufacturer. In this work, we propose the use of data marketplaces as a platform for addressing the data democratization challenges.

Data Marketplaces are information systems for trading data and data-related services [5, 6]. Their main features include searching, buying and selling of data [6]. These services are offered through a storefront interface where users can shop for data [7]. So far, there is little research on enterprise data marketplaces (EDMP) which enable the exchange of corporate data between employees [7, 8] and thus, directly contribute to the first democratization dimension. However, how an enterprise data marketplace addresses data democratization has not been investigated in detail.

In this respect, we present the following contributions: Based on the literature study and interviews conducted with the manufacturer, we developed *a data consumer's journey* which reflects the steps, the involved roles and current tool types for finding, understanding and accessing data within a company, in Section 2. From this journey, we derive *current challenges of data democratization*, also provided in Section 2, and present data marketplaces as a possible solution approach. Lastly, we implemented a prototype based on which we assess *the suitability of enterprise data marketplaces* for addressing the data democratization challenges in Section 3. Section 4 covers related work and Section 5 concludes this paper.

## 2 Data Democratization - Current State & Challenges

In order to ascertain the current state and challenges of data democratization we conducted a literature study. Some of the research introduces definitions and dimensions of data democratization, yet lack detail on how it is implemented and on the workflows surrounding data [2–4, 9]. Labadie et. al [10] propose data catalogs, data asset inventories with discovery, description and organization functionality, as tools for data democratization, yet it remains unclear how employees gain access to data. Other research provides a rough insight into the current workflows for handling data [3, 11, 12]. To incorporate a practical perspective on workflows for data democratization, we conducted over ten expert interviews with employees of a global manufacturer in various data-related roles, such as data scientists as well as enterprise and solution architects. The manufacturer is active in a variety of sectors like the mobility or industrial sector, and operates a global manufacturing network.

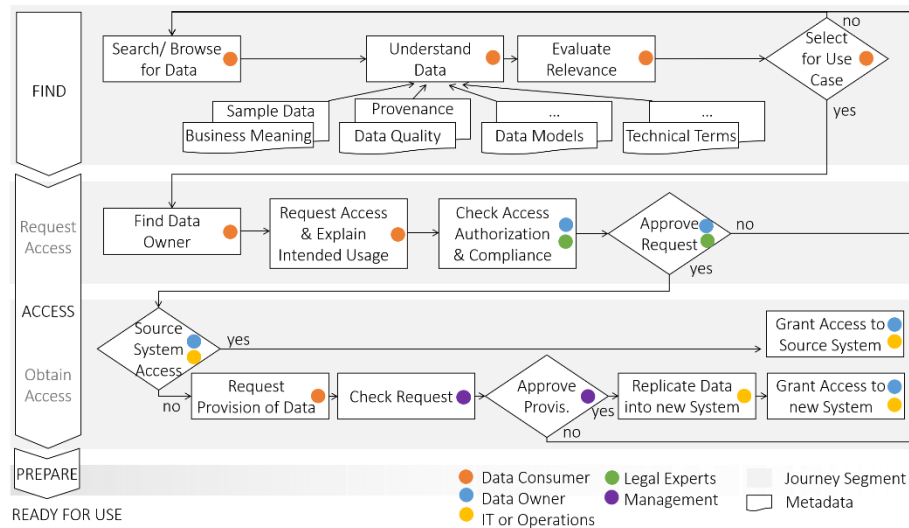
On the basis of the literature study and the expert interviews, we derived a representative consumer journey for industrial enterprises describing how consumers obtain relevant data for their use case. The consumer journey presented in Section 2.1 provides insights into the required steps, the involved roles and which tool types are used for finding, understanding and accessing data. Based on this, current challenges in data democratization for data consumers are deduced in Section 2.2.

### 2.1 The Data Consumer Journey

The representative data consumer journey for industrial enterprises, illustrated in Fig. 1, consists of four segments, *finding*, *requesting* and *obtaining access to*, and

*preparing* data. Due to lack of space, the data preparation is not discussed here. To illustrate the journey, we exemplify it through a manufacturing engineer who needs maintenance data from production lines, to create a machine maintenance dashboard.

To begin with, data consumers must find relevant data for their use case which involves *searching* for data, *understanding* it and *evaluating* its suitability. In order to use a search function all data must be inventoried through, e.g., a data catalog tool [10]. For example, the engineer searches for “sensor data production line P1” in the catalog to find registered sensor data. The consumer then evaluates the data’s relevance based on explanations or metadata. These are provided through metadata management tools like data catalogs, data quality platforms or business glossaries, or by the data owner or domain expert. Amongst others, the metadata includes the business meaning, its provenance, quality and modelling information. Having determined the data’s suitability, the consumer continues searching or pursues data access. In continuation of our example, the engineer consults the metadata offered through the data catalog, e.g., explaining that it is data on machine temperatures and thus is relevant for the maintenance dashboard. The engineer gathers additional information on the quality from a data quality platform and term definitions from a business glossary.



**Fig. 1.** Data Consumer Journey in Industrial Enterprises with Involved Roles

Gaining access to data entails requesting and obtaining access. Within the first step, the consumer must *find the data owner* and send a *request for access* including the intended usage. This is relevant for compliance to legal regulations such as the General Data Protection Regulation (GDPR). The request is either sent manually, e.g., via email, or through an authorization tool if the tool registers the desired data. The owner or a legal entity then *checks the consumers authorization* based on the data’s confidentiality and the employee’s clearance level. If either is insufficient the request is declined and the consumer resumes the search. In our example the engineer wants

to access data which cannot be requested through the authorization tool, however, the owner is registered in the data catalog and the engineer can send a request through other channels, e.g., email. As the sensor data is not personal data, the engineer may process it for the intended use case and hence, the request is approved by the owner.

Next, the owner or IT specialists determine whether to grant *access to the source system* based on factors such as the system's capacity. If this is not possible the data must be provisioned to another system like a data lake. The provisioning may have to be approved through management, as it can be costly. At this stage, the consumers either resume the search or have gained access to the data. In our example, the engineer is granted access to the source system and can now begin using the data.

## 2.2 Data Democratization Challenges for Data Consumers

Several data democratization challenges can be derived from the data consumer journey as described above. Firstly, *the process involves several parties (C1)* ranging from the data consumer over the data owner, IT specialists, legal experts and management. With each additional party, the process becomes more complex and time-consuming, as the responsible people have to be found and contacted. In our example scenario, the manufacturing engineer has to locate various persons such as the data owner or domain expert, if not defined in the catalog. Each of the contacted people have to process the request and report back. The second challenge lies therein that the *metadata for understanding the data is spread across a variety of tools (C2)*. It is inconvenient and challenging to maintain an overview of both the tools and metadata viewed on each data asset. The manufacturing engineer, for instance, had to collect metadata from at least the data catalog, the business glossary, data quality platform. Lastly, *the tools are not integrated across the access process (C3)*, so several tools are required and not all data that can be found in the metadata management tools like the catalog can be requested in the authorization tool. The engineer, for example, had to request access of the data owner through channels like email.

As the consumer journey demonstrates that sharing data, i.e., making data available and accessing it, is a key aspect in data democratization, and as data marketplaces are platforms for sharing data [13], we examine in the following to what extent these are suited to support data democratization and how they address the challenges. In the ensuing sections, the term (data) marketplace refers to an enterprise data marketplace.

## 3 Assessing the Suitability of Enterprise Data Marketplaces for Data Democratization Based on a Marketplace Prototype

In this section, we demonstrate to which extent the EDMP as a platform for trading data within an enterprise [7], supports the consumer journey and solves the challenges (C1-3). To do this, we have developed an EDMP prototype which is built on a micro-services architecture and is implemented with the Spring framework<sup>1</sup>. It has a metada-

---

<sup>1</sup> <https://spring.io/>

ta repository realized with a Neo4J<sup>2</sup> database and the metadata is modeled according to our metadata model HANDLE [14]. The enterprise landscape is simulated by a variety of databases and a data lake which are registered in the data catalog Apache Atlas<sup>3</sup>. With regard to functionality, it offers browse and search options with a detailed view for data discovery, service access- and subscription management, service registration for publishing data and license management for data governance.

### 3.1 The Consumer Journey with an Enterprise Data Marketplace

We conducted a user study in which we recorded the consumers' journeys while using the EDMP prototype. Fig. 2 shows the workflow with the marketplace in accordance to the consumer journey as presented in Fig. 1.



Fig. 2. Workflow of the Data Consumer Journey with an Enterprise Data Marketplace

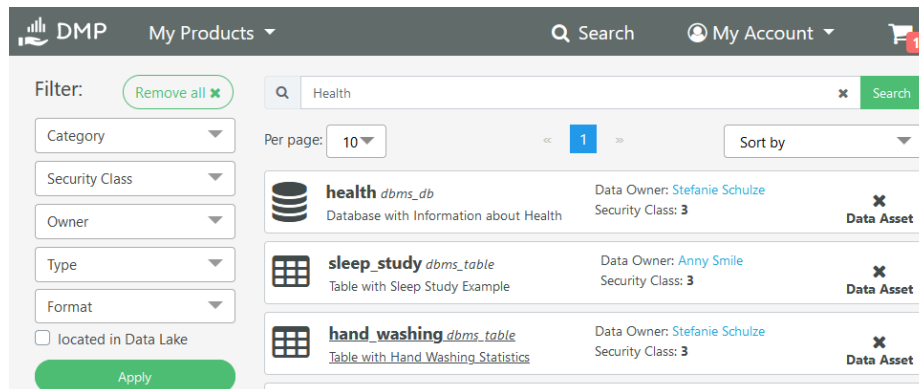


Fig. 3. Enterprise Data Marketplace Prototype – Search Results

In step 1, the consumer can find and browse through search results as shown in Fig. 3. Clicking on a search result opens a detailed description in step 2, which contains the available metadata on the data to help the consumer understand and evaluate its usefulness. This metadata and the search results are collected from the data catalog. If fitting, the consumer selects the data by adding it to the shopping cart in step 3. In the shopping cart, the intended usage and the desired provisioning are specified and selected for each data set, as displayed in Fig. 4. Requesting access is step 5, by submit-

<sup>2</sup> <https://neo4j.com/>

<sup>3</sup> <https://atlas.apache.org/>

ting the order in the shopping cart. The request is forwarded automatically to the according data owner, with information on the intended usage, provisioning and information on the consumer such as the clearance level, department etc. The data owner either declines or approves the request in step 6. If not present, the owner is prompted to supply product specific metadata in step 7, such as the license and permitted usage. If necessary the owner must then initiate the process of replicating the data into the data lake or another system in step 8. Meanwhile the consumer can track the current state of the request to see, e.g., if it is being approved, enriched with metadata, replicated into a new system etc. Lastly, the consumer receives access details to the data.

Therefore, the EDMP supports the consumer in the first two journey segments, finding and requesting access. The third segment, obtaining access, is partially supported. Only the request to provision and the realization thereof are not part of the marketplace. It can be seen that most of the consumer journey within industrial enterprises can be carried out through the EDMP and it is therefore improved. How the marketplace supports the journeys in other enterprise types remains to be investigated.

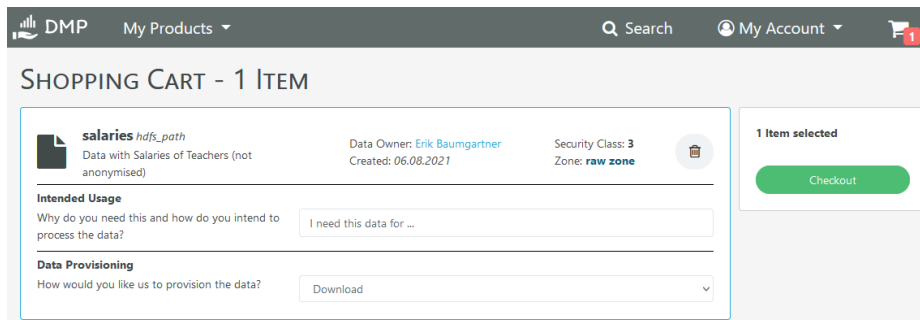


Fig. 4. Enterprise Data Marketplace Prototype - Shopping Cart View

### 3.2 Addressing of the Data Democratization Challenges

Challenge C1 refers to the number of people that need to be found and contacted in the consumer process. As explained in [8], companies are building data catalogs, and are assigning data owners to datasets. Our prototype integrates with a catalog, and extracts metadata like the owner, so the consumer does not have to find the relevant people manually. Furthermore, communications such as the access request are sent to the relevant people automatically and the marketplace interface provides participants with a workflow and overview of new, active and expired requests and subscriptions.

The second challenge concerns the distribution of metadata in different tools. The EDMP prototype addresses C2 by integrating these metadata in the detailed description which is accessible through the search mask and references the respective tools.

Lacking integration in the toolchain across the access process is challenge C3. As shown in Section 3.1, an EDMP provides support through most of the consumer journeys workflow. Finding data and requesting access are supported in the EDMP. Replicating data into another system, however, is exempt from its capabilities and not

supported through an integrated toolchain. Summing up, it has been shown that an enterprise data marketplace addresses the three democratization challenges.

## 4 Related Work on Data Marketplaces

We have searched for literature on relevant platforms for this topic such as dblp<sup>4</sup> and found out that data marketplaces are studied in both general and specialised contexts. For instance, data marketplace characteristics [15–17], emerging markets, trends [16, 17] and challenges [17] as well as current research fields [5] are identified in research articles. More specialized research focuses on, e.g., data marketplaces for IoT data [18, 19], or the use of specific technologies such as the distributed ledger technology [19, 20]. The listed research provides insights into a wide variety of data marketplace topics, however, not into the specifics of enterprise data marketplaces.

There are few research articles that address enterprise data marketplaces. Among these is research by Fernandez et al. [12] who differentiate internal marketplaces, which are synonymous to enterprise data marketplaces or a report by Wells [7] highlighting characteristics, components, services and involved technologies of such marketplaces and our work [8] in which we underline the need for enterprise data marketplaces. This research does not, however, discuss the enterprise marketplace in the context of data democratization. In contrast, Gröger [3] who presents the marketplace as a central element in the data ecosystem as well as Tata Consultancy Services [21] argue that enterprise data marketplaces are relevant for democratizing data. Yet both do not provide details on how the marketplace supports data democratization. Hence, this article covered this gap in existing literature by examining the democratization challenges and how these are addressed by an enterprise data marketplace.

## 5 Conclusion

Data democratization with its objective of making data available within the company has become increasingly important. Based on a representative data consumer journey within industrial enterprises, we presented the current state and challenges of data democratization. In this work we proposed the use of an enterprise data marketplace to support data democratization. Our marketplace prototype yields a consumer journey in which the democratization challenges are addressed. Consequently, we have shown that enterprise data marketplaces are a suited approach to foster data democratization in industrial enterprises. In future, we intend to investigate the data provider perspective and further implementation aspects of enterprise data marketplaces.

## 6 References

1. Gualtieri, M. et al.: The Forrester Wave: Big Data Hadoop Distributions, Q1 2016.

---

<sup>4</sup> <https://dblp.org/>



- Forrester Res. (2016).
2. Lefebvre, H., Legner, C., Fadler, M.: Data democratization: toward a deeper understanding. In: Proc. of the 42nd International Conference on Information Systems (ICIS) (2021).
  3. Gröger, C.: There is no AI without data. *Commun. ACM.* 64, 98–108 (2021).
  4. Awasthi, P., George, J.: A case for data democratization. In: Proc. of the 26th Americas Conference on Information Systems (AMCIS) (2020).
  5. Lange, J., Stahl, F., Vossen, G.: Datenmarktplätze in verschiedenen Forschungsdisziplinen: Eine Übersicht. *Informatik-Spektrum.* 41, 170–180 (2018).
  6. Meisel, L., Spiekermann, M.: Datenmarktplätze - Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy. Fraunhofer ISST. (2019).
  7. Wells, D.: The Rise of the Data Marketplace: Data as a Service. Eckerson Gr. (2017).
  8. Eichler, R. et al.: Enterprise-Wide Metadata Management: An Industry Case on the Current State and Challenges. In: Proc. of the 24th International Conference on Business Information Systems (BIS). pp. 269–279 (2021).
  9. Zeng, J., Glaister, K.W.: Value creation from big data: Looking inside the black box, (2018).
  10. Labadie, C. et al.: FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. In: Proc. of the IEEE 22nd Conference on Business Informatics (CBI). pp. 201–210 (2020).
  11. Gröger, C., Hoos, E.: Ganzheitliches Metadatenmanagement im Data Lake: Anforderungen, IT-Werkzeuge und Herausforderungen in der Praxis. In: 18. Fachtagung für Datenbanksysteme für Business, Technologie und Web (BTW) (2019).
  12. Fernandez, R.C., Subramaniam, P., Franklin, M.J.: Data Market Platforms: Trading Data Assets to Solve Data Problems. *Proc. VLDB Endow.* 13, 1933–1947 (2020).
  13. Spiekermann, M. et al.: A metadata model for data goods. In: Multikonferenz Wirtschaftsinformatik (MKWI). pp. 326–337 (2018).
  14. Eichler, R. et al.: Modeling metadata in data lakes—A generic model. *Data Knowl. Eng.* 136, (2021).
  15. Schomm, F., Stahl, F., Vossen, G.: Marketplaces for data: An initial survey. In: ACM SIGMOD Record. pp. 15–26 (2013).
  16. Stahl, F. et al.: Marketplaces for Digital Data: Quo Vadis? *Comput. Inf. Sci.* 10, 22 (2017).
  17. Spiekermann, M.: Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics.* 54, 208–216 (2019).
  18. Zheng, Z. et al.: Challenges and opportunities in IoT data markets. In: Proc. of the 4th International Workshop on Social Sensing (SocialSense). pp. 1–2 (2019).
  19. Ramachandran, G.S., Radhakrishnan, R., Krishnamachari, B.: Towards a Decentralized Data Marketplace for Smart Cities. In: Proc. of the IEEE International Smart Cities Conference (ISC2). pp. 1–8. IEEE (2018).
  20. Roman, D., Stefano, G.: 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). pp. 95–101. IEEE (2016).
  21. Saxena, S.: Enterprise Data Marketplace: Democratizing Data within Organizations. Tata Consult. Serv. (2018).