

# Data Trading and Monetization: Challenges and Open Research Directions

Qusai Ramadan  
Institute of Software Engineering  
University of Koblenz  
Germany  
qramadan@uni-koblenz.de

Zeyd Boukhers  
Fraunhofer Institute for Applied  
Information Technology FIT, and  
Uniklinik Köln  
Germany  
zeyd.boukhers@fit.fraunhofer.de

Muath AlShaikh  
College of Computing and Informatics  
Saudi Electronic University  
Kingdom of Saudi Arabia  
m.alshaikh@seu.edu.sa

Christoph Lange  
Fraunhofer Institute for Applied  
Information Technology FIT, and  
RWTH Aachen University  
Germany  
christoph.lange-  
bever@fit.fraunhofer.de

Jan Jürjens  
University of Koblenz, and  
Fraunhofer-Institute for Software and  
Systems Engineering ISST  
Germany  
juerjens@uni-koblenz.de

## ABSTRACT

Traditional data monetization approaches face challenges related to data protection and logistics. In response, digital data marketplaces have emerged as intermediaries simplifying data transactions. Despite the growing establishment and acceptance of digital data marketplaces, significant challenges hinder efficient data trading. As a result, few companies can derive tangible value from their data, leading to missed opportunities in understanding customers, pricing decisions, and fraud prevention. In this paper, we explore both technical and organizational challenges affecting data monetization. Moreover, we identify areas in need of further research, aiming to expand the boundaries of current knowledge by emphasizing where research is currently limited or lacking.

## KEYWORDS

Data Marketplaces, Data Trading, Data Monetization, Data Exchange, Data Sharing

### ACM Reference Format:

Qusai Ramadan, Zeyd Boukhers, Muath AlShaikh, Christoph Lange, and Jan Jürjens. 2023. Data Trading and Monetization: Challenges and Open Research Directions. In *The International Conference on Future Networks and Distributed Systems (ICFNDS '23)*, December 21–22, 2023, Dubai, United Arab Emirates. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3644713.3644758>

## 1 INTRODUCTION

The amounts of data being created, captured, and replicated are continuously increasing at a fast pace. According to a recent study, the projected growth of annual data on a global scale indicates

a substantial rise from 33 zettabytes in 2018 to an estimated 175 zettabytes by the year 2025 [32]. The availability of data and advanced technologies offer substantial potential for creating value. Therefore, within the business landscape, data has transitioned from being merely an enabler of products to becoming valuable products in and of themselves, serving as strategic resources for companies [37]. For example, it is predicted that the European Union's data economy will attain a value of €829 billion, accompanied by a notable expansion in the workforce of data professionals, expected to reach 10.9 million by 2025 [32].

In situations where organizations lack efficient amounts and high-quality data to effectively execute or enhance their processes and services (which is the case in many SMEs) they may seek to acquire data from other organizations through purchasing arrangements. Traditionally, data monetization involves selling or licensing data directly to third parties. However, this approach often raised privacy concerns and required transferring large volumes of data, which could be time-consuming and costly [19]. As a result, in recent times, a multitude of digital data marketplaces and platforms have emerged, serving as intermediaries between data providers and consumers. Data marketplaces represent a recent development in the market, with recent instances emerging as a means to facilitate industrial data trading and monetization.

Despite the growing establishment and acceptance of data marketplaces in industry (with more than 70% of enterprises in the EU having already acquired external data), substantial obstacles persist that hinder data trading [19]. In this paper, we shed light on the current technical and organizational challenges that affect data monetization. We also identify open research directions to push the boundaries of existing knowledge by highlighting areas where research is limited or lacking.

This paper is organized as follows: in Section 2 we provide the necessary background. Section 3 provides the state of the art. In Section 4 we explore the challenges that affect data monetization, and we propose future research directions toward addressing the identified challenges. In Section 5 we discuss the limitations of our research findings. Finally, Section 6 concludes this paper.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICFNDS '23, December 21–22, 2023, Dubai, United Arab Emirates  
© 2023 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0903-6/23/12  
<https://doi.org/10.1145/3644713.3644758>

## 2 BACKGROUND

Online data marketplaces have evolved through different phases and in different categories. Initially, data catalogues focused on open data, providing search and browse mechanisms for metadata and links to data sets. Then, government and industrial marketplaces facilitated secure data exchange and data sovereignty. In later phases, personal data markets and open science marketplaces emerged<sup>1</sup>, along with Sensor Data Marketplaces for IoT data<sup>2</sup>. Three distinct directions have been deliberated and implemented concerning the architecture of data marketplaces:

**Centralized data marketplaces.** The centralized marketplace design acts as a standard multi-sided platform, attracting various participants from the data ecosystem and providing facilitation services through a technological platform, enabling the platform operator to benefit from positive participation [40].

A centralized design prioritizes larger data suppliers as costs decrease with scale. However, creating and maintaining diverse raw data records is challenging. Ideally, the marketplace should possess metadata for all data records to enable future trades [30]. In addition to the expensive maintenance of this type of platform, the distribution of exchange volumes and profits will heavily favour the larger entities that provide high volumes of data at a fixed cost of metadata [37]. A centralized platform can enforce specific entrance policies and fees, which creates a clear boundary. A centralized design without boundaries and enforcement is more suitable for non-private data. However, for private data, there is a risk of misuse after the exchange.

**Decentralized data marketplaces.** Distributed Ledger Technologies (DLTs) and blockchains have sparked interest in decentralized marketplace designs. Initially applied in virtual currency markets [14]. In the decentralized design, market clearing, and transparency are improved, resulting in a shift of taxonomic tagging responsibilities from the centralized marketplace to individual data suppliers [7, 41]. Each data provider must maintain a data taxonomy that annotates records with common standards for searching, aggregating, analyzing, and reselling data.

A decentralized marketplace has similar features to a centralized marketplace but overcomes some of its drawbacks. The key advantage is that trades can be executed and verified directly by participants. Additionally, proof of provenance is decentralized and can be verified independently [18, 38]. In contrast to centralized approaches, a decentralized marketplace tackles three crucial challenges: Firstly, it addresses the issue of data privacy by recording trades in a public ledger and authenticating transactions, eliminating privacy obstacles in trading [51]. Secondly, it distinguishes between data asset holders and original creators, granting creators the right to control future access and usage if desired, thereby ensuring ownership and control over data assets [18]. Lastly, it enables an open and transparent exchange environment, allowing for easy search based on unique characteristics [3].

**Federated data marketplaces.** Federated data marketplaces allow organizations to collaborate on data analysis without sharing the underlying raw data [47]. Federated data marketplaces differ from decentralized data marketplaces regarding data governance and

control. In decentralized data marketplaces, data is typically stored and managed using distributed ledger technology (e.g., blockchain), where data owners have direct control over their data and can grant access to specific parties based on predefined rules and smart contracts. In contrast, federated data marketplaces rely on federated learning [21] or other privacy-preserving techniques [15] to enable collaborative data analysis. Data remains in the possession of the original owners, and analysis is performed on local data sets. Aggregated results or model updates are then shared among the participants while preserving data privacy. This way, organizations can unlock the value of their data assets while maintaining control over data privacy and compliance.

## 3 STATE OF THE ART

We conducted desk research to provide an overview of the current status of data marketplaces. In this section, we discuss the most related industrial-oriented research projects, initiatives, and standards. Our research deliberately focuses on recent developments in the field of data trading within the European Union (EU). We have limited the scope of our study to this specific region due to the EU's prominent role in shaping policies and frameworks for secure data exchange. For instance, following the introduction of the EU's 2030 vision, which seeks to establish a unified European Data Market facilitating seamless data exchange across sectors and member states, numerous research projects have been funded by the EU Commission towards realizing the EU vision.

### 3.1 Projects and Initiatives

In the following, we list the most related industrial-oriented projects and initiatives that have been developed within the EU towards facilitating data trading and monetization.

**AI4EU.** This initiative constructed an on-demand platform for AI within Europe, with the objective of reducing barriers to innovation, fostering technology transfer, and facilitating the growth of startups and SMEs across all sectors<sup>3</sup>. Its goals included promoting shared research, ensuring long-term sustainability, expanding the availability of AI assets within its ecosystem, and ensuring interoperability with existing AI and data components. Additionally, AI4EU established the Ethical Observatory to guarantee adherence to human-centric AI values, European data privacy regulations, and AI systems' transparency.

**AEQUITAS.** This project, funded by the European Union in 2022, has set out to establish a controlled experimentation environment with the aim of assisting AI producers in enhancing their awareness of the discrimination that could result from AI systems<sup>4</sup>. Specifically, AEQUITAS proposes the design of a controlled experimentation environment that caters to both developers and users, enabling them to conduct controlled experiments for various purposes. These include detecting and mitigating bias in AI systems. This project provides an experimentation environment to generate synthetic data sets with diverse characteristics that influence fairness, enabling testing in laboratory settings.

<sup>1</sup><https://marketplace.eosc-portal.eu/>, Accessed: (14 September 2023)

<sup>2</sup><https://rubygarage.org/blog/big-data-marketplaces>, Accessed: (14 September 2023)

<sup>3</sup><https://www.ai4europe.eu/>, accessed: 14/09/2023

<sup>4</sup><https://www.aequitas-project.eu/consortium>, accessed: 14/09/2023

**EUH4D.** This project, initiated in 2020 by the European Union, aimed to establish a federation of Big Data Digital Innovation Hubs (DIHs) across Europe<sup>5</sup>. The project's main objective is to develop a European catalog of data sources and federated data-driven services, make this offer accessible at the regional level for European SMEs and start-ups, and foster cross-border data-driven experimentation through data sharing and interoperability. To achieve this objective, the project established an initial ecosystem of 12 Big Data DIHs from the four EU poles. Furthermore, the project aimed to engage SMEs, start-ups, and web entrepreneurs, conducting cross-border experiments and involving a significant number of companies in the data-driven innovation program. Starting with 12 EU regions in nine countries, EUHubs4Data planned to expand its reach to over 20 regions in 14 countries, establishing an ecosystem for data-driven innovation in Europe.

**DATAPorts.** This project, initiated in 2020 and funded by the European Union, aimed to transition European seaports from connected and digital entities to smart and cognitive ones<sup>6</sup>. It provided a secure environment for aggregating and integrating data from diverse sources within digital ports, owned by different stakeholders. This was achieved through the design, implementation, and operation of the Cognitive Ports Data Platform, which interconnected existing digital infrastructures within seaports, established trusted data sharing and trading policies, and offered advanced data analytic services. The project involved deploying the platform in two European seaports, addressing local constraints, and conducting a global use case to target inter-port objectives. The broader implications of the DataPorts project extend to seaports across Europe, enhancing the trustworthiness, reliability, and efficiency of business operations and reinforcing the European Single Market.

**MUSKETEER.** An EU-funded project initiated in 2018, aimed to create a validated, federated machine learning platform capable of addressing privacy preservation, scalability, and efficiency challenges for real-world applications in the field of smart manufacturing and healthcare<sup>7</sup>. The key objectives of MUSKETEER include creating machine learning models suitable for privacy-preserving scenarios, establishing robust security measures against external and internal threats, developing a standardized and expandable architecture, and introducing a rewarding model to monetize data sets based on their real value.

**TRUSTS.** This EU-funded project, running from 2020 to 2022, aimed to establish trust and confidence in data markets by developing an independent and federated platform based on insights gained from previous national projects<sup>8</sup>. A primary focus of the project was the thorough examination of legal and ethical aspects throughout the entire data valorization chain. With the objective of creating a GDPR-compliant data marketplace for both personal and non-personal data, TRUSTS enhanced existing marketplaces and showcased the potential of the TRUSTS Platform through various use cases in the corporate business data sector.

## 3.2 Standards

In the following, we discuss the most related reference architectures and standards that have been developed within the EU to facilitate data trading.

**International Data Spaces Reference Architecture Model (IDS-RAM).** The International Data Spaces Association (IDSA) has defined the IDS Reference Architecture Model (IDS-RAM) and agreements to establish trusted virtual data spaces [28]. The central component for secure data exchange is the IDS connector, which acts as a trusted security gateway. It enables direct transmission of data from certified data spaces, allowing the original data provider to maintain control and set usage conditions [29]. The connector encapsulates data within a virtual container, ensuring compliance with mutually agreed-upon terms.

It is noteworthy that IDS operates as an architecture, not a platform, enabling seamless integration with other systems to universally implement the concept of data spaces across various domains<sup>9</sup>. To establish formal trust within this architecture, The IDSA has developed a certification concept for IDS components and participants' operational environments. Independent evaluation facilities ensure transparent and trust mechanisms. Optional components, such as the Clearing House, Vocabulary Provider, Broker, and App Store, prepare the system for utilization as a data marketplace<sup>10</sup>.

**Gaia-X.** The Gaia-X initiative was initiated in Europe and is now supported and led by representatives from politics, business, and science in France, Germany, and other European partners<sup>11</sup>. It aims to promote interoperability and establish data and infrastructure ecosystems aligned with European values and standards. Its primary goal is to enhance the development of a federated, trusted, and user-friendly digital ecosystem. The project addresses various challenges, including decentralized processing locations, diverse technology stacks, limited transparency and sovereignty over stored and processed data and infrastructure, ambiguity regarding applicable jurisdiction, sector-specific data spaces and ontology gaps, insufficient accessibility of widely accessible application programming interfaces (APIs), multiple stakeholders, and limited accessibility to existing data and infrastructure services [5, 25].

Gaia-X does not aim to create a competing product but rather seeks to network different elements through open interfaces and standards, linking data and fostering innovation. By leveraging existing approaches, Gaia-X aims to offer an alternative data ecosystem approach to the dominant global platforms [39]. It establishes its own data sharing and storage services, ensuring digital sovereignty for data owners while providing a foundation for smart services and innovative business processes.

## 4 CHALLENGES AND OPEN RESEARCH DIRECTIONS

This section enumerates and engages in a thorough discourse regarding the contemporary challenges that necessitate additional research and development efforts. In order to advance the state

<sup>5</sup><https://euhubs4data.eu/overview/>, accessed: 14/09/2023)

<sup>6</sup><https://dataports-project.eu/>, accessed: 14/09/2023)

<sup>7</sup><https://musketeer.eu/>, accessed: 14/09/2023)

<sup>8</sup><https://www.trusts-data.eu/>, accessed: 14/09/2023)

<sup>9</sup><https://www.fraunhofer.de/content/dam/zv/en/fields-of-research/industrial-data-space/whitepaper-industrial-data-space-eng.pdf>, accessed: 14/09/2023)

<sup>10</sup><https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/>, accessed: 14/09/2023)

<sup>11</sup><https://www.gaia-x.eu/>, accessed: 19/09/2023)

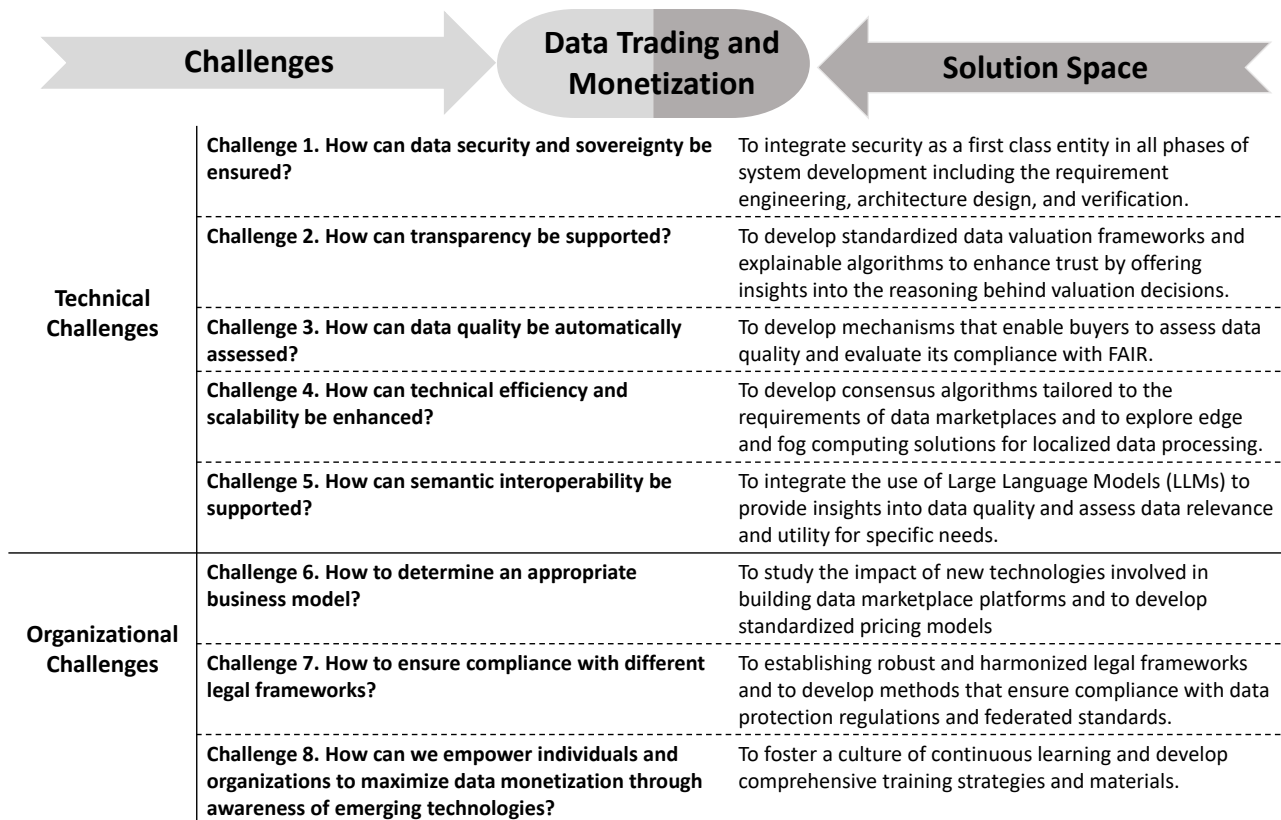


Figure 1: Overview of Challenges and Solutions in the Data Trading and Monetization Landscape.

of knowledge in the field of data trading and monetization, it is imperative to recognize the areas where current limitations persist and where novel solutions and investigations are required.

Fig. 1 provides an overview of the identified challenges and solutions in the data trading and monetization landscape. We classify the identified challenges into two categories: *technical challenges*, which pertain to issues related to data infrastructure, security, and analytics, and *organizational challenges*, which encompass hurdles associated with data governance, workforce training, and business models. We have framed the challenges as open research questions, endeavoring to provide prospective solutions for each, in order to facilitate the work of fellow scholars and practitioners in the domain. Detailed descriptions of the technical challenges and organizational challenges are provided in Section 4.1 and Section 4.2, respectively.

#### 4.1 Technical Challenges

This section delves into the challenges encountered in the realm of data marketplaces, which are obstacles to data monetization.

**Challenge 1. How can data security and sovereignty be ensured?** Similar to any traditional information system that stores and transmits sensitive data, data marketplaces are potential targets for various cybersecurity threats. Therefore, data marketplaces

must implement technical and organizational measures to protect data privacy, ensure data integrity, prevent data breaches, address identity and authentication issues, and establish secure data transfer protocols. Additionally, data marketplaces must navigate regulatory compliance requirements, such as data protection laws, to maintain the confidentiality and trustworthiness of the data being exchanged. However, maintaining data security and privacy in data marketplaces is more challenging compared to traditional information systems due to the following factors:

- **Data sharing complexity:** Data marketplaces involve the sharing and exchange of data among multiple parties. This complex network of data sharing increases the risk of data exposure, unauthorized access, and breaches during data transmission [34, 35].
- **Trust and anonymity concerns:** Data marketplaces often require a level of trust and anonymity to encourage participation from various stakeholders. Striking the right balance between maintaining trust and protecting user privacy requires careful design and implementation [44]. More specifically, the trade of personal data can result in unintended disclosure of personal information, particularly when individuals act as data providers and large corporations or governments serve as data buyers. This dynamic can create an imbalance

of power, increasing the risk of personal information being exposed without the explicit consent or awareness of the individuals involved [8, 20].

- *Complexity of controlling data monetization.* The complexity of controlling data monetization poses challenges for data providers in effectively tracking and ensuring compliance with data-sharing agreements. This lack of visibility and control raises concerns among data providers, as they are unable to monitor how their data is being used and whether it aligns with the agreed-upon terms. The fear of competitors benefiting from their data in unforeseen ways further amplifies the apprehension [37]. This situation not only creates uncertainties surrounding the intended usage of the data but also introduces potential privacy risks [19], as the extent of data utilization and the resulting implications on individual privacy become uncertain.

**Toward addressing challenge 1.** Future research should develop comprehensive and dynamic strategies to ensure the consistent safeguarding of data security. Specifically, in the software engineering process of building a data marketplace, it is crucial to give importance to security and privacy through automation support to address several key aspects: Firstly, it involves the elicitation and specification of requirements that effectively capture the essential security and privacy requirements [1, 35]. Secondly, it requires careful architectural design, ensuring adherence to data protection concerns [31, 50]. Thirdly, rigorous validation and verification processes are essential to ensure that the resulting data marketplace meets the specified security and privacy requirements [2], especially when the data marketplace evolves.

**Challenge 2. How can transparency be supported?** Ensuring transparency is a significant challenge in data marketplaces, particularly concerning the relationship between data providers and data brokers. There is often a lack of clear and open communication, leading to limited visibility into how data brokers evaluate the value of data and the methodologies employed in this assessment [27]. Data providers often find themselves unaware of the specific criteria or fair processes used by data brokers to determine the worth of their data. This lack of transparency hampers the ability of data providers to make informed decisions about their data and understand the fairness and accuracy of the valuation process [12]. As a result, data providers may feel disadvantaged and uncertain about the fairness and legitimacy of the transactions taking place in the data marketplace ecosystem.

**Toward addressing challenge 2.** To overcome the transparency challenge, future research should involve the development of a standardized data valuation framework. The valuation framework should encompass the following components: First, valuation metrics to provide a common language for assessing data and making it easier for both data providers and consumers to comprehend how data is priced [11]. Second, explainable algorithms to enhance trust by offering insights into the reasoning behind valuation decisions, fostering transparency and accountability [33]. Third, blockchain technology for immutable records to ensure the integrity of data transactions [49].

**Challenge 3. How can data quality be assessed in an automated way?** Data marketplaces often struggle with issues related

to data accuracy, completeness, and consistency, which are crucial for effective decision-making within data marketplaces [19]. Additionally, the FAIR principles, advocating for data to be findable, accessible, interoperable, and reusable, present another hurdle [17]. However, one specific challenge lies in the lack of mechanisms that support buyers in assessing data quality or evaluating its adherence to FAIR principles before purchasing. This absence makes it difficult for buyers to determine the accuracy of the data they acquire and compromises their ability to make informed decisions.

**Toward addressing challenge 3.** To overcome the challenge of ensuring data quality, efforts are needed to establish mechanisms that enable buyers to assess data quality and evaluate its compliance with FAIR principles before making a purchase [17]. Transparent metadata management systems, standardized data quality assessments, and clear documentation of adherence to FAIR principles can significantly enhance buyer confidence. Implementing such mechanisms would not only facilitate informed decision-making but also promote trust and reliability in market data places, ultimately supporting successful data monetization strategies.

**Challenge 4. How can technical efficiency and scalability be enhanced?** Data marketplaces utilizing distributed ledger technology or decentralized architectures face increased communication costs due to the constant synchronization and communication required between multiple nodes [22]. The more nodes involved, the higher the communication overhead, which can negatively impact the performance and scalability of the marketplace [16]. Additionally, the computational complexity of operations such as data verification, encryption, and indexing contributes to substantial computation costs. Meeting these computational demands in a timely and cost-effective manner is essential for ensuring efficiency and scalability in data marketplaces.

**Toward addressing challenge 4.** To address the challenge of technical efficiency and scalability, a promising research direction involves optimizing distributed computing. This involves a multifaceted approach, including the investigation of customized consensus algorithms tailored to the unique requirements of data marketplaces such algorithms can enhance data transaction speeds, reduce latency, and ensure data integrity, particularly in large-scale data trading scenarios [9, 24]. Additionally, exploring edge and fog computing solutions for localized data processing can alleviate the strain on centralized servers [43], improving overall system efficiency.

**Challenge 5. How can semantic interoperability be supported?** As previously articulated in [4], the cornerstone of effective data exchange is the mutual understanding of the shared data, encapsulated by the concept “Semantic Interoperability”. However, achieving this understanding is often a hurdle, contingent upon the data consumer’s (i.e., data Buyer) grasp of the vocabulary and metadata from the data provider (i.e., data Seller). This necessitates the accessibility of metadata to decode the data content, adhering to the FAIR principles [46], and the utilization of a common vocabulary. In instances where a common vocabulary is absent, a mapping between the vocabularies of the data consumer and provider is imperative. Illustratively, in a Digital Advertising Dataset, the attribute “Engagement” may have different interpretations between the data provider and consumer. For instance, does “Engagement” denote a viewer watching an ad video for a duration exceeding a

predefined number of seconds? Or, hovering their mouse over an ad for a specific period? Such discrepancies underline the necessity for clear vocabulary mapping to foster effective data exchange and in return data trading and monetization.

**Toward addressing challenge 5.** Web technologies have historically tackled the issues of semantic interoperability. However, with the recent surge in machine learning advancements, there are renewed directions to explore and address these challenges more comprehensively [4]. Notably, the transformative capabilities of transformer architectures, especially Large Language Models (LLMs), uncovered new possibilities. These models can act as a canal, enhancing understanding between data providers and consumers. Such capabilities not only assist in resolving semantic interoperability hurdles but also empower data providers to enhance the trade and monetization potential of their data. Simultaneously, it provides data consumers with insights into data quality, allowing them to assess its relevance and utility for their specific needs.

## 4.2 Organizational Challenges

Beyond the technical challenges discussed in the previous section, it is essential to recognize that substantial organizational obstacles exist for data monetization. Significant non-technical challenges for data monetization are:

### Challenge 6. How to determine an appropriate business model?

Organizations recognize the value of their data assets but struggle to determine the most effective and profitable strategies for monetization [37, 48]. Without established guidelines, pricing strategies, revenue models, and value propositions become ambiguous, hindering the alignment of data monetization efforts with business objectives and leading to missed opportunities for revenue generation. Additionally, the absence of standardized guidelines hampers trust and transparency in data transactions [6, 37].

However, determining the appropriate price for data products and services is complex due to the intangible nature of data and the absence of standardized pricing models [45]. Additionally, there is reluctance among data consumers to pay for data, as organizations struggle to perceive its direct value and return on investment [26]. Furthermore, data providers encounter difficulties in formulating an optimal profit strategy that strikes a balance between maximizing revenue and managing the cost structure associated with data acquisition [23]. This challenge arises from the complexity of assessing the potential returns on data investments and the uncertainty surrounding the monetization potential of their data assets.

**Toward addressing challenge 6.** To address the lack of clear business models, future work should study the impact of new technologies involved in building data marketplace platforms such as blockchain and IDS connectors on the business models. This will enable the development of well-defined business model guidelines to foster trust, ensure fair transactions, and provide clarity and transparency in data monetization, enabling organizations to maximize the value of their data assets [6, 42]. In addition, bridging the gap in price acceptance requires efforts to educate about the value of data, raise awareness of its potential impact, and develop standardized pricing models [10, 23, 26].

**Challenge 7. How to ensure compliance with different legal frameworks?** Data is subject to various legal considerations, including privacy, security, intellectual property, and data protection laws. However, the evolving nature of data and rapid technological advancements often outpace the development of comprehensive legal guidelines [37]. This leads to ambiguity and uncertainty regarding data rights, ownership, and consent [36]. For example, data ownership and consent become more intricate in data marketplaces where data is sourced from multiple providers. Specifically, some scholars suggest that data (especially personal data) should be owned by individuals. However, alternative viewpoints from legal scholars argue that data ownership may be unfeasible or conceptually flawed in addressing the complex societal and economic challenges presented by the data economy [13]. Therefore, the absence of clear legal frameworks hinders organizations from leveraging data and raises risks of legal disputes and compliance issues.

**Toward addressing challenge 7.** To address the challenge of the lack of legal frameworks, future research should concentrate on: First, Establishing robust and harmonized legal frameworks to provide clarity, protect rights, ensure privacy, and facilitate fair and secure data monetization practices. Second, developing methods to ensure compliance or (at least) alignment with (i) data protection regulations such as GDPR to protect individuals' privacy, (ii) federated standards for data infrastructure such as Gaia-X to promote interoperability and data sovereignty, and (iii) reference architecture models such as the IDS-RAM to establish standardized data management practices.

### Challenge 8. How can we empower individuals and organizations to maximize data monetization through awareness of emerging technologies?

The rapid advancements in technology have introduced innovative tools and platforms that can enhance data monetization efforts, but the shortage of accessible and up-to-date training resources limits individuals' ability to acquire the knowledge and competencies required for success. This skills gap hampers organizations' capacity to fully leverage data monetization and capitalize on the opportunities presented by emerging technologies. Addressing this challenge is crucial to bridge the skills gap and empower workers in the field of data monetization.

**Toward addressing challenge 8.** To address the skills gap organizations and companies should focus on developing comprehensive training strategies and materials, so organizations can provide individuals with the necessary guidance and expertise to navigate and leverage emerging technologies within data marketplaces effectively. Additionally, fostering a culture of continuous learning and providing ongoing training opportunities can help workers stay updated with the latest advancements and adapt to evolving technologies.

## 5 RESEARCH LIMITATIONS

The findings of our research are subject to several threats. A threat to *external validity* is that we discuss the current challenges in the field of data trading and monetization solely from academic perspectives, without the direct involvement of industrial experts. Although our knowledge closely aligns with the experiences of industrial experts, this approach introduces a potential limitation in

terms of the practical applicability of our suggestions for addressing the identified challenges in real-world settings. To mitigate this concern, future work should involve consultation with industry experts and compare their insights with our findings.

A threat to *internal validity* is that we have limited the scope of the state-of-the-art to recent developments in the field of data trading within the European Union (EU). However, this focused approach poses challenges in terms of generalizability. For example, ensuring compliance with specific standards within the EU, such as Gaia-X and IDS, may not be a requirement for data trading practices in other regions or on a global scale. As a consequence, the generalizability of our findings to broader contexts may be compromised. To address this limitation, a systematic literature review is required to investigate data trading developments in various regions, facilitating a more comprehensive understanding of the challenges within this dynamic field.

## 6 CONCLUSIONS

In this paper, we have examined the landscape of data monetization, highlighting both technical and non-technical challenges that data marketplaces face. The rapid growth in data generation and technological advancements has opened up opportunities for data monetization, but it has also introduced complex hurdles that need to be addressed for successful implementation.

On the technical side, challenges related to data security, privacy, transparency, data quality, and technical efficiency have been discussed. These challenges underscore the need for robust technological solutions and standards to ensure the integrity and effectiveness of data marketplaces. Addressing these issues will be crucial in building trust among data providers and consumers.

On the organizational side, challenges such as the lack of clear business models and legal frameworks, have also been identified as significant barriers to data monetization. The lack of attention given to these organizational challenges hinders the full realization of the potential of data monetization. Therefore, it is imperative to expand the focus of research and development efforts to encompass these organizational challenges and devise comprehensive strategies that ensure the successful implementation and sustainable growth of data monetization initiatives.

## ACKNOWLEDGMENTS

This research was partially supported by the EU's Horizon Digital, Industry, and Space program under grant agreement ID 101092989-DATAMITE.

The contributions of Zeyd Boukhers and Christoph Lange to this work were funded by the FAIR Data Spaces project of the German Federal Ministry of Education and Research (BMBF) under grant numbers FAIRDS05 and FAIRDS15.

## REFERENCES

- [1] Rene Abraham, Johannes Schneider, and Jan vom Brocke. 2023. A taxonomy of data governance decision domains in data marketplaces. *Electronic Markets* 33, 1 (2023), 22.
- [2] Amir Shayan Ahmadian, Jan Jürjens, and Daniel Strüder. 2018. Extending model-based privacy analysis for the industrial data space by exploiting privacy level agreements. In *Proceedings of the 33rd annual ACM symposium on applied computing*. 1142–1149.
- [3] Sangam Bhujel and Yogachandran Rahulmathavan. 2022. A survey: Security, transparency, and scalability issues of nft's and its marketplaces. *Sensors* 22, 22 (2022), 8833.
- [4] Zeyd Boukhers, Christoph Lange, and Oya Beyan. 2023. Enhancing Data Space Semantic Interoperability through Machine Learning: a Visionary Perspective. In *Companion Proceedings of the ACM Web Conference 2023*. 1462–1467.
- [5] Arnaud Braud, Gaël Fromentoux, Benoit Radier, and Olivier Le Grand. 2021. The road to European digital sovereignty with Gaia-X and IDSA. *IEEE network* 35, 2 (2021), 4–5.
- [6] Cristina Georgiana Calancea and Lenuța Alboae. 2021. Techniques to improve b2b data governance using fair principles. *Mathematics* 9, 9 (2021), 1059.
- [7] Christian Catalini and Joshua S Gans. 2020. Some simple economics of the blockchain. *Commun. ACM* 63, 7 (2020), 80–90.
- [8] Vassilis Charitsis, Detlev Zwick, and Alan Bradshaw. 2018. Creating worlds that create audiences: Theorising personal data markets in the age of communicative capitalism. (2018).
- [9] Natalia Chaudhry and Muhammad Murtaza Yousaf. 2018. Consensus algorithms in blockchain: Comparative analysis, challenges and opportunities. In *2018 12th International Conference on Open Source Systems and Technologies (ICOSST)*. IEEE, 54–63.
- [10] Lingjiao Chen, Hongyi Wang, Leshang Chen, Paraschos Koutris, and Arun Kumar. 2019. Demonstration of Nimbus: Model-based Pricing for Machine Learning in a Data Marketplace. In *Proceedings of the 2019 International Conference on Management of Data*. 1885–1888.
- [11] Michele Colli, Jonas Nygaard Uhrenholt, Ole Madsen, and Brian Vejrum Waehrens. 2021. Translating transparency into value: an approach to design IoT solutions. *Journal of Manufacturing Technology Management* 32, 8 (2021), 1515–1532.
- [12] Matthew Crain. 2018. The limits of transparency: Data brokers and commodification. *new media & society* 20, 1 (2018), 88–104.
- [13] Barbara J Evans. 2011. Much ado about data ownership. *Harv. JL & Tech.* 25 (2011), 69.
- [14] David S Evans. 2014. Economic aspects of Bitcoin and other decentralized public-ledger currency platforms. *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper* 685 (2014).
- [15] Nick Hynes, David Dao, David Yan, Raymond Cheng, and Dawn Song. 2018. A demonstration of sterling: a privacy-preserving data marketplace. *Proceedings of the VLDB Endowment* 11, 12 (2018), 2086–2089.
- [16] Georgy Ishmaev. 2020. The ethical limits of blockchain-enabled markets for private IoT data. *Philosophy & Technology* 33 (2020), 411–432.
- [17] Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, Mercè Crosas, Michel Dumontier, Chris T Evelo, et al. 2020. FAIR principles: interpretations and implementation considerations. , 10–29 pages.
- [18] Pantelis Koutroumpis, Aija Leiponen, and Llewellyn DW Thomas. 2017. *The (unfulfilled) potential of data marketplaces*. Technical Report. ETLA working papers.
- [19] Pantelis Koutroumpis, Aija Leiponen, and Llewellyn DW Thomas. 2020. Markets for data. *Industrial and Corporate Change* 29, 3 (2020), 645–660.
- [20] Sebastian Lawrenz, Priyanka Sharma, and Andreas Rausch. 2019. Blockchain technology as an approach for data marketplaces. In *Proceedings of the 2019 international conference on blockchain technology*. 55–59.
- [21] Yuzheng Li, Chuan Chen, Nan Liu, Huawei Huang, Zibin Zheng, and Qiang Yan. 2020. A blockchain-based decentralized federated learning framework with committee consensus. *IEEE Network* 35, 1 (2020), 234–241.
- [22] Kang Liu, Wuhui Chen, Zibin Zheng, Zhenli Li, and Wei Liang. 2019. A novel debt-credit mechanism for blockchain-based data-trading in internet of vehicles. *IEEE Internet of Things Journal* 6, 5 (2019), 9098–9111.
- [23] Weichao Mao, Zhenzhe Zheng, and Fan Wu. 2019. Pricing for revenue maximization in iot data markets: An information design perspective. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 1837–1845.
- [24] Du Mingxiao, Ma Xiaofeng, Zhang Zhe, Wang Xiangwei, and Chen Qijun. 2017. A review on consensus algorithm of blockchain. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2567–2572.
- [25] Crispin Niebel, Abel Reiberg, and Peter Kraemer. 2022. *Gaia-X for SMEs*. White Paper. Gaia-X Hub Germany.
- [26] Chaoyue Niu, Zhenzhe Zheng, Fan Wu, Shaojie Tang, and Guihai Chen. 2020. Online pricing with reserve price constraint for personal data markets. *IEEE Transactions on Knowledge and Data Engineering* 34, 4 (2020), 1928–1943.
- [27] Hyeontaek Oh, Sangdon Park, Gyu Myoung Lee, Hwanjo Heo, and Jun Kyun Choi. 2019. Personal data trading scheme for data brokers in IoT data marketplaces. *IEEE Access* 7 (2019), 40120–40132.
- [28] Boris Otto, Michael ten Hompel, and Stefan Wrobel. 2019. International Data Spaces: Reference architecture for the digitization of industries. *Digital transformation* (2019), 109–128.
- [29] Heinrich Pettenpohl, Markus Spiekermann, and Jan Ruben Both. 2022. International data spaces in a nutshell. *Designing Data Spaces; Springer: Cham, Switzerland* (2022), 29–40.

- [30] Gowri Sankar Ramachandran, Rahul Radhakrishnan, and Bhaskar Krishnamachari. 2018. Towards a decentralized data marketplace for smart cities. In *2018 IEEE International Smart Cities Conference (ISC2)*. IEEE, 1–8.
- [31] Dumitru Roman and Gatti Stefano. 2016. Towards a reference architecture for trusted data marketplaces: The credit scoring perspective. In *2016 2nd International Conference on Open and Big Data (OBD)*. IEEE, 95–101.
- [32] David Reinsel-John Gantz-John Rydning, John Reinsel, and John Gantz. 2018. The digitization of the world from edge to core. *Framingham: International Data Corporation* 16 (2018), 1–28.
- [33] Wojciech Samek and Klaus-Robert Müller. 2019. Towards explainable artificial intelligence. *Explainable AI: interpreting, explaining and visualizing deep learning* (2019), 5–22.
- [34] Ghaidaa Shaabany and Reiner Anderl. 2018. Security by design as an approach to design a secure industry 4.0-capable machine enabling online-trading of technology data. In *2018 International Conference on System Science and Engineering (ICSSE)*. IEEE, 1–5.
- [35] Ghaidaa Shaabany, Marco Grimm, and Reiner Anderl. 2016. Secure information model for data marketplaces enabling global distributed manufacturing. *Procedia CIRP* 50 (2016), 360–365.
- [36] Jan-Terje Sørli and Jörn Altmann. 2019. Sensing as a service revisited: A property rights enforcement and pricing model for IIoT data marketplaces. In *Economics of Grids, Clouds, Systems, and Services: 16th International Conference, GECON 2019, Leeds, UK, September 17–19, 2019, Proceedings 16*. Springer, 127–139.
- [37] Markus Spiekermann. 2019. Data marketplaces: Trends and monetisation of data goods. *Intereconomics* 54, 4 (2019), 208–216.
- [38] Melanie Swan. 2015. *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc."
- [39] Hubert Tardieu. 2022. Role of Gaia-X in the European Data Space Ecosystem. In *Designing Data Spaces: The Ecosystem Approach To Competitive Advantage*. Springer International Publishing Cham, 41–59.
- [40] Llewellyn DW Thomas and Aija Leiponen. 2016. Big data commercialization. *IEEE Engineering Management Review* 44, 2 (2016), 74–90.
- [41] Arvind Upadhyay, Sumona Mukhuty, Vikas Kumar, and Yigit Kazancoglu. 2021. Blockchain technology and the circular economy: Implications for sustainability and social responsibility. *Journal of cleaner production* 293 (2021), 126130.
- [42] Montijn van de Ven, Antragama Ewa Abbas, Zenlin Kwee, and Mark de Reuver. 2021. Creating a taxonomy of business models for data marketplaces. In *34th Bled EConference: Digital Support from Crisis to Progressive Change*. University of Maribor Press, 313–325.
- [43] Blesson Varghese, Nan Wang, Dimitrios S Nikolopoulos, and Rajkumar Buyya. 2020. Feasibility of fog computing. *Handbook of Integration of Cloud Computing, Cyber Physical Systems and Internet of Things* (2020), 127–146.
- [44] Shefali Virkar, Gabriela Viale Pereira, and Michela Vignoli. 2019. Investigating the social, political, economic and cultural implications of data trading. In *Electronic Government: 18th IFIP WG 8.5 International Conference, EGOV 2019, San Benedetto Del Tronto, Italy, September 2–4, 2019, Proceedings 18*. Springer, 215–229.
- [45] Lara Vomfell, Florian Stahl, Fabian Schomm, and Gottfried Vossen. 2015. *A classification framework for data marketplaces*. Technical Report. Ercis working paper.
- [46] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.
- [47] Ronghua Xu and Yu Chen. 2021. Fed-ddm: A federated ledgers based framework for hierarchical decentralized data marketplaces. In *2021 international conference on computer communications and networks (ICCCN)*. IEEE, 1–8.
- [48] Mengxiao Zhang, Fernando Beltrán, and Jiamou Liu. 2023. A Survey of Data Pricing for Data Marketplaces. *IEEE Transactions on Big Data* (2023).
- [49] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. 2017. An overview of blockchain technology: Architecture, consensus, and future trends. In *2017 IEEE international congress on big data (BigData congress)*. Ieee, 557–564.
- [50] Mirko Zichichi, Contu Michele, Stefano Ferretti, Rodríguez-Doncel Víctor, et al. 2020. Ensuring personal data anonymity in data marketplaces through sensing-as-a-service and distributed ledger. In *CEUR Workshop Proceedings*, Vol. 2580. Sun SITE Central Europe/RWTH Aachen University, 1–16.
- [51] Guy Zyskind, Oz Nathan, et al. 2015. Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE security and privacy workshops*. IEEE, 180–184.