# Towards Federated Machine Learning and Distributed Ledger Technology-based Data Monetization

Timo Himmelsbach[1,*], Yongli Mou[2], Stefan Decker[2] and Armin Heinzl[1]

[1]*University of Mannheim, Mannheim, Germany*
[2]*RWTH Aachen University, Aachen, Germany*

**Abstract**

Data sharing and monetization provides organizations with new sources of revenue and value creation. However, an accurate and scalable approach to data sharing and monetization for organizations is still lacking in practice. Due to the lack of effective mechanisms for control and enforcing governance as well as accurate valuation and pricing mechanisms, organizations are hesitant to share data. As a result, a large share of the economic value-creation potential of data is foregone. We propose a distributed-ledger-technology-based approach for decentralized data valuation incorporating federated machine learning to enable decentralized data-enabled learning and data valuation in a collaborative manner. We evaluate the proposed concept model with empirical evidence from expert interviews and single out the predictive maintenance context for future prototype development and testing.

**Keywords**

Usage-based Reverse Data Valuation, Distributed Ledger Technology, Federated Machine Learning, Decentralized Data Platform Ecosystems

## 1. Introduction

In practice, an accepted and scalable approach to data sharing and monetization for organizations is still lacking [1]. In today's interconnected digital economies, data is a key enabler for economic value creation and innovation [2]. Driven by the ongoing digital transformation, organizations produce, exchange, and consume data in every aspect of their business operations [3]. The surging amounts of available data provides organizations with abundant avenues for new business value creation such as improving operational efficiency, enhancing decision-making, and innovating with emerging technologies [4]. Moreover, data can provide firms with new revenue streams, when being monetized externally [4]. Despite this promising outlook, data exchange and monetization beyond organizational boundaries is still scarce [5]. Today, most data transactions are still completed through offline negotiations between individual data sellers and data buyers sold for an (arbitrarily) fixed price set by the data owner [1] or are exchanged via centralized third-party data platforms. Centralized data platforms, such as Snowflake and Advaneo, offer companies a pathway to exchange their data by providing a common infrastructure, complementary resources (e.g., APIs, SDKs), and governance policies for its participants (complementors) to monetize data. However, a major drawback of centralized platforms, characterized by having a central actor (data platform owner), is the centralization of control and decision rights (e.g., platform access and data sharing rules). As such, the platform owner accumulates large power over the platform complementors, allowing it to capture large fractions of the economic value created within the platform ecosystem. This disincentivizes companies to participate in the data exchange. As a result, a large share of the economic value-creation potential of data remains unused.

In the emerging literature strand of data monetization, various challenges have been identified hindering the systematic monetization of data. Besides general issues relating to the provisioning of data to third parties such as security and privacy, data management, or organizational challenges [6, 4], the ownership, control, and governance [3] as well as the valuation and pricing of data [4, 1] have been singled out as key challenges preventing data producers from sharing and monetizing their data effectively. [3] state that data ownership, control, trust, and the enforcement of inter-organizational governance mechanisms are central to effective and sovereign data exchange. Furthermore, [1] postulates that extant research tends to focus on the value creation aspect of data sharing, while the question of a fair compensation for data sharing is largely avoided. An emerging literature stream on centralized platforms focuses on antitrust and regulation, addressing major drawbacks such as market power and anti-competitive practices that reduce innovation and consumer welfare [7, 8]. Decentralized platforms, fueled by Distributed-Ledger-Technology (DLT) and fed-

*Joint Workshops at 49th International Conference on Very Large Data Bases (VLDBW'23) — Workshop on Data Ecosystems (DEco'23), August 28 - September 1, 2023, Vancouver, Canada*
*Corresponding author.
himmelsbach@uni-mannheim.de (T. Himmelsbach); mou@dbis.rwth-aachen.de (Y. Mou); decker@dbis.rwth-aachen.de (S. Decker); heinzl@uni-mannheim.de (A. Heinzl)

erated machine learning capabilities have the potential to overcome the drawbacks of centralized control, market power, valuation, privacy, and security concerns by providing a decentralized infrastructure that enables both, secure data-enabled learning, and data valuation without an intermediary [9, 10, 11]. In our quest to facilitate a secure and scalable approach to data sharing and monetization, we set out to answer the following research question:

*RQ: How can Distributed-Ledger-Technology and federated machine learning foster data monetization in decentralized data ecosystems?*

Due to the nascent nature of DLT and federated machine learning-based data monetization, we propose a blockchain-based artefact including federated learning capabilities and pre-evaluate it in the context of an Industrial Internet of Things (IIoT) enabled predictive maintenance use case. In IIoT, the value of information exchange as well as the challenges thereof have been discussed manifold [12]. At the current stage, this research-in-progress paper presents a conceptual model for an artifact that is to be developed and evaluated in the near future. The envisioned artefact draws on and extends the solution proposed by [1]. The presented conceptual model addresses the two fundamental issues of centralized platforms —disproportional control and value capturing— as well as data privacy and data security by proposing an approach for DLT-based data valuation and federated data-enabled learning. The suggested approach combines the scalable concept of compute-to-data (edge computing) with a public, permissioned blockchain (Ethereum) for secure and transparent data access, valuation, and monetization. We contribute to the growing body of Information Systems literature on data monetization and data ecosystems, as well as Computer Science literature on federated machine learning in data ecosystems and provide an innovative data monetization approach in form of a conceptual model to practitioners.

# 2. Background

## 2.1. Data Valuation and Data Monetization

Data valuation and pricing are key obstacles to data sharing and monetization. Data is a "non-rivalrous experience good" [4]. Non-rivalrous means that, once created, data can be exploited repeatedly by multiple parties without deteriorating in value [4]. Experience means that data must be used to realize its value [4]. Consequently, the value of the same data set varies significantly depending on the use case, context, and time. Even in adequate contextual conditions, there are only very few objective measures and limited methodologies available to accurately determine the value of data. As such, the data quality dimensions (e.g., completeness, accuracy, timeliness) or the relative position along the data value chain (e.g., collecting, pre-processing, analyzing, using) help organizations to better gauge the value of their data [13, 14].
Extant literature proposes mainly four overarching, methodological data valuation concepts [13, 15, 1]: cost-oriented, market-price-oriented, risk-oriented, and usage-oriented data valuation. Cost-oriented approaches build their valuation on all costs that arise throughout the data value chain such as data storage or data analytics. However, this approach is limited as the actual value created by the data is entirely neglected. Market-price-oriented concepts assume that data assets are traded on markets where their prices as an approximation of value are determined. This approach is limited by the availability of comparable idiosyncratic use case-data combinations [4] and the assumed homogeneity among data buyers in their willingness-to-pay [16]. Risk-oriented valuation approaches consider potential business risks (e.g., measured as a function of probability and business cost of contingent outcomes) that may arise for a company from loss or misusage of data assets [13, 15]. Given that both, the probability and the business costs of adversarial scenarios are extremely difficult to quantify and forecast, the practicality of this approach is arguably low. Finally, usage-oriented valuation refers to the contribution that a data asset can generate to the company performance. [1] propose an approach for the usage-oriented data valuation suggesting that data valuation and pricing should comprise a combination of both, forward-looking expected value anticipation and ex-post value measuring, which depends on the actual value contribution of a data asset.
Data monetization provides organizations with an incentive to share their data with external parties and participate in data ecosystems [4]. Data monetization describes the usage of data to achieve "quantifiable economic benefit" [17]. In a broader sense data monetization refers to both, indirect efforts aiming at improvement of business processes and decision making, as well as external efforts aiming at outright selling data assets (via a data broker or independently) or data-based products and services [18, 19, 20]. While extant research examines the value creation potentials of data sharing manifold, the question of a fair valuation in data sharing arrangements is largely avoided [1].

## 2.2. Decentralized Data Platform Ecosystems

Centralized digital platforms can be defined as "the extensible codebase of a software-based system that provides core functionality shared by apps that interoperate with it, and the interfaces through which they interoperate" [21]. They enable multisided transactions and innovations between different market players and create value through network effects [22, 23, 24]. A centralized data platform refers to a technical environment for recording, storing, analyzing, and presenting (big) data [25, 26]. Ecosystems are described by a "group of interacting firms that depend on each other's activities." [27], e.g., developers on Google Android depending on software updates provided by the platform owner or data providers on a data platform such as Snowflake depending on certain standards and requirements for data storage.

Decentralized data platform ecosystems can be understood as a subtype of data platform ecosystems. However, literature still lacks a generally adopted and recognized definition of decentralized data platform ecosystems. In management and organizational theory decentralization is mostly referring to decision making and authority [28, 29]. Correspondingly, in the platform context, decentralization regularly refers to governance, the mechanisms employed by platform owners aiming to orchestrate and influence ecosystem outcomes to foster value co-creation [30]. On a more technical level decentralization in data platform ecosystems can also refer to the data infrastructure [31]. Within this paper decentralized data platforms are understood as platforms with decentralized data infrastructure and decentralized governance building on DLT [32, 33]. With that, our understanding of decentralized data platforms follows [34], who define a decentralized data marketplace as lacking both a central authority and a central data repository.

DLT serves as an umbrella term for multiparty systems operating in an environment without a central authority or operator [35]. Blockchain technology is frequently regarded as a certain subgroup of DLT using a specific decentralized data structure building on a chain of hash-linked data blocks, representing transactions that are distributed and consistent among the network participants, the so-called nodes [35, 36]. Public blockchains allow all nodes to read the transactions logs while private blockchains only permit the reading of transactions only to authorized nodes. Permissioned blockchains restrict transaction validation, i.e., the participation in the consensus mechanisms, to chosen nodes while in permissionless blockchains all nodes validate transactions [35, 37, 38, 31]. Due to their distributed nature and the peer-to-peer validation of transactions, DLT-based data platform ecosystems eliminate the need for a central platform owner [9, 39, 32].

Central aspects of decentralized DLT-based data platform ecosystems are smart contracts and tokens. Smart contracts can be understood as "systems which automatically move digital assets according to arbitrary pre-specified rules" [40]. Thus, smart contracts are algorithms that comprise a-priori specified business logics (e.g., ownership, access-, and control rights), automatically execute transactions accordingly, and record all transactions to the blockchain [31]. In combination with blockchain infrastructure, smart contracts provide a "reliable, secure, and convenient approach to specifying an agreement, which is essential for data sharing" [1], enhancing the transparency and traceability of transactions within the system.

Tokenization refers to the "abstract representation of physical assets in the form of blockchain tokens" [41]. There are different token types, each with different characteristics and taking central roles in the governance and accessibility of decentralized DLT-based platform ecosystems. A high-level categorization distinguishes between three token types: Payment tokens, security tokens, and utility tokens [42]. In this paper, we focus on two subtypes of utility tokens. Utility tokens are required for accessing the functionality of DLT token platforms. Without ownership of such tokens neither the platform's services can be used, nor any transactions can be executed. The first utility token subtype, non-fungible token (NFT), is based on the ERC721 standard [43]. Non-fungible means that while it can be transferred between participants within the ecosystem, the token is unique. Thus, NFTs certify ownership and tradeable rights to a digital asset. The second utility token subtype, fungible tokens, are classified by the ERC20 standard [43]. Fungible tokens are identical and interchangeable and represent access rights to digital assets. These access rights can be traded and managed much like any other good. In sum, smart contracts and tokenization provide an infrastructure for a decentralized data platform ecosystem, enabling data sovereignty and trust and, thus, eliminating the need for a central platform owner. Further central aspects of decentralized data platform ecosystems that enhance data sharing and monetization are federated machine learning and digital twin capabilities.

### 2.3. Digital Twins and Federated Machine Learning

Digital twins refer to a virtual representation or digital replication of a physical object, system, or process [44]. It is a digital counterpart that mirrors the characteristics, behavior, and attributes of its real-world counterpart in real-time or near real-time and allows for diagnostics by using data captured from connected sensors. This data can be further utilized to optimize the operation and performance of the physical counterpart or predict future states [45]. Digital twins have been widely adopted in various fields, such as manufacturing, healthcare and transportation, due to their ability to offer real-time monitoring and simulation, performance optimization, and fault prediction [46]. The emergence of Internet of Things (IoT) technology and machine learning has significantly accelerated the implementation and application of Digital twins. IoT allows real-time data collection from various sensors placed on the physical twin, and machine learning enables the processing of this vast amount of data. Machine learning, as a subset of Artificial Intelligence (AI), has been central in enhancing the capabilities of digital twins. Machine learning provides the necessary algorithms and methods to analyze the data and create models capable of learning from this data, identifying patterns, and making predictions.

Currently, machine learning process models are majorly centralized. The process involved collecting data from various sources and aggregating it at a central point (e.g., server or cloud), where a machine learning model would then be trained [47]. However, this process raises several concerns [48]. Firstly, the transmission of data to a central repository results in security risks. Data can be intercepted during transmission, and the central repository itself can be a target for cyberattacks. Secondly, the aggregation of data at a central point raises privacy concerns and ethical issues. In many cases, the data used for machine learning contains sensitive information about individuals or organizations. Even if anonymized, the risk of re-identification through data linkage remains. Thirdly, the centralized approach requires significant computational resources and is less efficient. Large volumes of data have to be moved, requiring substantial bandwidth and storage. The latency associated with moving data to a central point can also slow down the learning process.

To address these limitations, the concept of federated (machine) learning was introduced [49]. Instead of requiring data centralization like other conventional approaches, federated (machine) learning describes a distributed machine learning approach that allows for training a global model on decentralized data sources while data remains on its original [11]. It is designed to address privacy concerns and data localization requirements, particularly in scenarios where data cannot be easily centralized due to privacy regulations or data ownership considerations. The process typically involves the repetition of the following steps, namely, client selection, global model distribution, local model training, model verification, model aggregation and global model update. Federated Learning provides a methodology to build machine learning models using data located across different devices or servers while ensuring data privacy and reducing the requirements for data transmission. This strategy is particularly beneficial in scenarios where data privacy is critical, or where devices have limited connectivity or resources.

# 3. Proposed conceptual Model

## 3.1. Concept Model Overview - Predictive Maintenance as Exemplary Application Context

Figure 1 provides an economic-centered overview of the proposed conceptual model of DLT-based data monetization, focusing on the use case of predictive maintenance. The exchange of information as base for innovative machine learning models that enhance the accuracy of predictive maintenance and prevent machine downtime has proven to be high [12], thus providing a suitable basis for our DLT-based data monetization concept model.

Currently, the value of datasets is estimated pre-acquisition, at the moment of sale based on cost, risk, market, or usage-oriented calculations. Post-acquisition value creation, not a-priori considered currently is not reflected in the data valuation, leading to high uncertainty for data sellers, especially regarding competitive data sets. The proposed reverse data monetization logic is a step-by-step, post-acquisition valuation and pricing approach based on data usage, actual costs, and generated business impact, aiming to consider future value creation in the determination of data value over the course of data usage. Data valuation and pricing is associated with the achieved business outcome by the data buyer after purchasing a data set. In the predictive maintenance use case, the business impact of a data set is determined by potential prevented machine downtime costs and production losses it can reduce. For instance, a dataset on rare frequencies of industrial pumps leading to breakdowns, ultimately leading to production stops of industry goods, would be of high value for competitors running similar pump systems in case machine downtimes could be prevented based on predictive maintenance precautions.

Potential machine downtime costs are calculated by the amount of produced items per hour with a certain profit per item equaling total costs of production losses. Furthermore, the value of a dataset is comprised of costs, such as data usage costs (curation, storage, monitoring, analysis) and (predictive) maintenance costs (employees, operating resources). Finally, a negotiable profit margin for the data seller completes the data valuation determination. The proposed reversed data monetization approach further comprises upfront and post-acquisition compensation. Upfront compensation is based on pre-acquisition valuation using traditional cost- and risk-oriented pricing models to mitigate any costs associated with the data

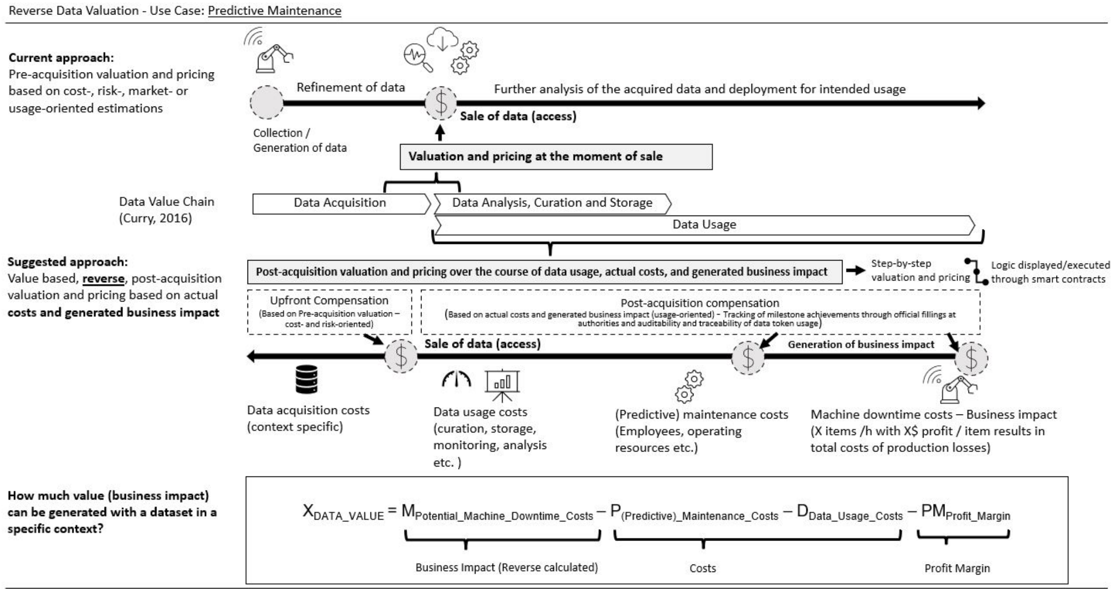


**Figure 1:** Reverse Data Monetization Concept Model

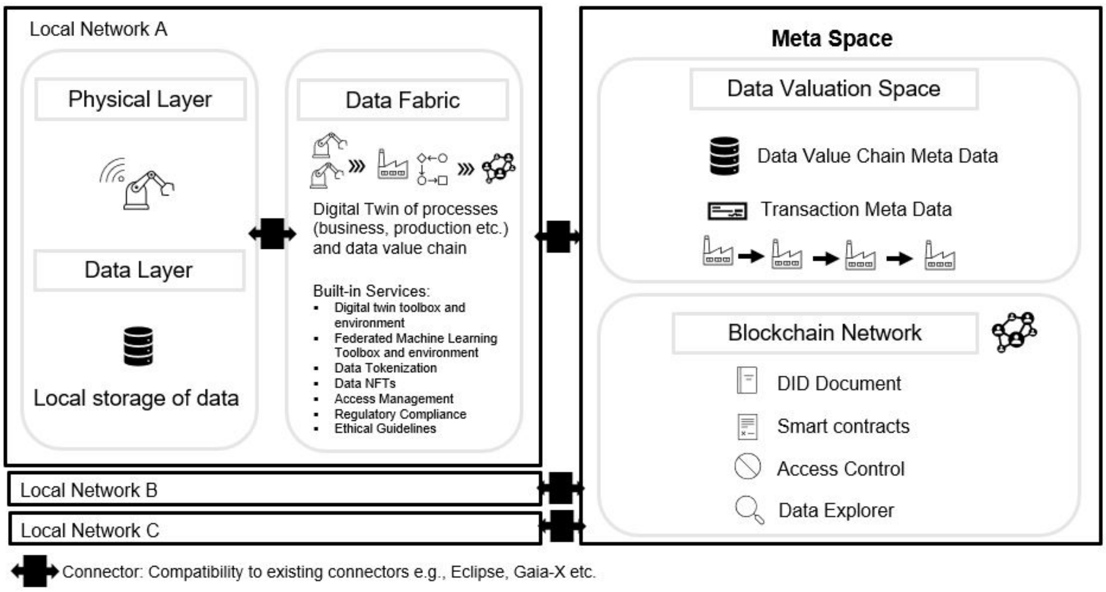


**Figure 2:** Federated Learning enabled DLT-based Data Monetization Architecture

buyer side. Post-acquisition compensation is determined by the actual costs and generated business impact along a pre-acquisition defined and negotiable time period. The value of data is determined at certain pre-acquisition defined milestones over the course of data usage, actual costs, and generated business impact.

Figure 2 provides a technology-centric overview of the proposed DLT-based reverse data monetization concept model enabling a more accurate data valuation of data monetized between different actors in a digital ecosystem: The proposed model comprises a decentralized data fabric with built in services such as a digital twin visualization service, tokenization, federated machine learning capabilities, and an access and authority management, as well as a DLT-enabled data valuation meta space. In the following, the key components and functions of the proposed model are elaborated.

### 3.2. Blockchain-enabled Data Valuation Meta Space

The Blockchain-based infrastructure provides the foundation for a decentralized platform ecosystem and the reversed data monetization logic. The decentralized platform ecosystem, we refer to as meta space, consists of two sub-dimensions: Data valuation space and Blockchain network. The Blockchain network comprises the key infrastructure elements of a public, permissioned blockchain (Ethereum), smart contracts, access control, DIDs, and a data explorer. The data valuation space contains and visualizes data transactions through digital twins along the data value chain from different actors within the ecosystem. It aims to provide a full traceability of data transactions to enable the reverse data valuation logic over the course of data usage. While the actual data sets are stored off-chain in a local organizational network, the access control of meta data assets in form of tokens is stored on-chain [34]. The Blockchain network comprises smart contracts, that allow for a-priori specified business logics (e.g., ownership, access-, and control rights), automatically execute transactions accordingly, (e.g., micropayments for data transactions) and record all transactions as meta data to the blockchain [50, 31]. Therefore, the data infrastructure is designed to function without a central platform owner shifting the control and governance over the data assets to the data owners. Moreover, the underlying blockchain provides the data owner with full traceability of the actual data usage of third parties, enabling a usage-based pricing approach.

### 3.3. Data Fabric

To better understand the data value chain, the proposed model comprises a data fabric, a decentralized framework, that enables organizations to seamlessly and efficiently manage, integrate, and distribute data across diverse and distributed systems. The data fabric allows for a unified and consistent view of data, and facilitates data sharing, accessibility, and analysis by providing built in services such as a digital twin toolbox. The visualization of business and production processes in a digital manner allows for replication of the data value chain, including data transactions, forming the basis for reverse data valuation. Connectors (E.g., Eclipse, Gaia-X) provide a secure and trustful exchange of data within a local network of an organization as well as the exchange of meta data between the meta space and the local network of organizations.

The data fabric further comprises federated learning capabilities, enabling the data sellers to retain their data on-premise while allowing third parties access to a data set. Federated machine learning enables access to data assets distributed across multiple devices without revealing sensitive information to a central cloud server [11]. As such, the decentralized data storage and federated learning builds the foundation for enhanced control and data sovereignty for the data owner. Previously, the data seller had no control over the data set, once it was sold to a third party. A further built-in service of the data fabric, tokenization, enables transparent and auditable access to data assets. The ERC721 NFT certifies ownership and full rights to a digital asset [51, 52]. By acquiring an ERC20 data token that contains access rights to a certain data set, a data buyer can access and use the data assets as pre-specified by smart contracts. Access can either mean accessibility to a full data asset or parts of a data asset, or the deployment of approved algorithms by the data seller on the data asset. Furthermore, access to a data asset might be “perpetual”, “time-bound” or “one-time” [53]. By enabling traceability and control, tokenization contributes to data sovereignty of the data owner.

## 4. Pre-Evaluation

To test and reach a better understanding of the suggested concept of DLT-based reverse data monetization, a first pre-evaluation with experienced industry experts was conducted. For this purpose, the data monetization logic was precisely described to receive feedback. The drawn conclusions and the feedback were incorporated in the proposed framework. Two key challenges need to be overcome. First, it might be difficult to persuade data sellers to take part in this monetization scheme as risk is shifted to the data sellers. High costs may accrue for collecting and preparing the data prior to a monetization, efforts which ask for some a-priori reward. High compensation in case of success might increase data seller’s willingness to enter expenditure in advance, however a certain upfront payment might also become necessary. Second, evaluating the step-by-step implementation as introduced within this paper one expert notes that it might be a challenge to achieve consensus between data buyer and data seller about when a particular milestone is reached and to what extent the data set was used for achieving this outcome. This paper aims at mitigating this second challenge by unambiguously defining the monetization milestones, ideally in conjunction with external third parties, representing it in smart contracts and tracking data usage using tokenization and a permissioned blockchain. Especially, a clear definition of milestones and corresponding to this triggering of the following actions like for example payments, implemented by smart contracts, are emphasized in that respect.

## 5. Conclusion and Future Work

This paper presents a conceptual model for a prototype that is to be developed and further evaluated in the near future. The proposed concept combines the scalable concept of federated machine learning (compute-to-data) with a public, permissioned blockchain, a reverse valuation logic and addresses the two fundamental issues of centralized platforms, disproportional control, and value capturing, as well as key issues of data sharing, trust, data privacy, and ultimately, data valuation and pricing. The evaluation of the approach was only carried out in an initial step. The prototype is yet to be developed and tested in practice. Consequently, the technical specifications such as automation, security, and low transaction costs – while we recognize their importance for the prototype development [1] – were not central to the proposed conceptual logic in its current stage. Future work on this artefact will focus on an extensive evaluation with industry experts in order to complete and extend the proposed concept model. We recognize the difficulty of evaluating the actual value generated by using a data asset as the final value generation often cannot be measured and the value creation also depends on external factors. To mitigate that we aim to test our concept model in the predicitve maintenance use case, in which the actual data usage can effectively be measured ex-post. We aim to contribute to the growing body of Information Systems literature on data monetization and data ecosystems as well as literature in the field of Computer Science especially, federated machine learning and data ecosystems, and offer an initial innovative data monetization approach that may reduce the hesitancy of firms to monetize their data and actively contribute to the value creation within their data ecosystems.

## References


[1] Z. Wang, Z. Zheng, W. Jiang, S. Tang, Blockchain-enabled data sharing in supply chains: Model, operationalization, and tutorial, Production and Operations Management 30 (2021) 1965–1985.

[2] J. Gelhaar, T. Groß, B. Otto, A taxonomy for data ecosystems, in: Proceedings of the 54th Hawaii International Conference on System Sciences, 2021, pp. 6113–6122. URL: https://hdl.handle.net/10125/71359.

[3] D. Lis, B. Otto, Data governance in data ecosystems – insights from organizations, in: Proceedings of the Americas Conference on Information Systems, volume 12, 2020.

[4] P. Parvinen, Advancing data monetization and the creation of data-based business models, Communications of the Association for Information Systems 47 (2020) 25–49. URL: https://doi.org/10.17705/1CAIS.04702.

[5] J. Gelhaar, B. Otto, Challenges in the emergence of data ecosystems, in: Proceedings of the Twenty-Third Pacific Asia Conference on Information Systems, 2020, p. 175. URL: https://aisel.aisnet.org/pacis2020/175.

[6] S. Mirbagherimarvili, H. Ghanbari, M. Rossi, Data monetization challenges in established organizations: A systematic literature, in: 13th Scandinavian Conference on Information Systems, volume 3, 2022.

[7] G. Parker, G. Petropoulos, M. W. Van Alstyne, Digital platforms and antitrust, SSRN (2020). URL: http://dx.doi.org/10.2139/ssrn.3608397.

[8] J. Tirole, Economics for the Common Good, Princeton University Press, 2017. URL: https://press.princeton.edu/books/hardcover/9780691175164/economics-for-the-common-good.

[9] P. Bossauer, D. Schreiber, T. Neifer, C. Pakusch, G. Stevens, Dezentralisierung der sharing economy-potentiale blockchain-basierter sharing-plattformen., in: Wirtschaftsinformatik (Zentrale Tracks), 2020, pp. 158–173.

[10] T. Kollmann, S. Hensellek, K. de Cruppe, A. Sirges, Toward a renaissance of cooperatives fostered by blockchain on electronic marketplaces: a theory-driven case study approach, Electronic Markets 30 (2020) 273–284.

[11] Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: Concept and applications, ACM Transactions on Intelligent Systems and Technology (TIST) 10 (2019) 1–19.

[12] M. Platenius-Mohr, S. Malakuti, S. Grüner, J. Schmitt, T. Goldschmidt, File-and api-based interoperability of digital twins by model transformation: An iiot case study using asset administration shell, Future generation computer systems 113 (2020) 94–105.

[13] E. Curry, Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems, Springer, 2016. URL: https://doi.org/10.1007/978-3-030-29665-0.

[14] V. Grover, R. H. Chiang, T.-P. Liang, D. Zhang, Creating strategic business value from big data analyt-

ics: A research framework, Journal of management information systems 35 (2018) 388–423.
[15] F. Bodendorf, K. Dehmel, J. Franke, Scientific approaches and methodology to determine the value of data as an asset and use case in the automotive industry (2022) 5618–5627. URL: https://hdl.handle.net/10125/80023.
[16] S. Mehta, M. Dawande, G. Janakiraman, V. Mookerjee, How to sell a data set? pricing policies for data monetization, Information Systems Research 32 (2021) 1281–1297.
[17] Gartner glossary, 2022. URL: https://www.gartner.com/en/information-technology/glossary/data-monetization, accessed: 16 November 2022.
[18] J. Baecker, M. Engert, M. Pfaff, H. Krcmar, Business strategies for data monetization: Deriving insights from practice., in: Wirtschaftsinformatik (Zentrale Tracks), 2020, pp. 972–987.
[19] M. S. Najjar, W. J. Kettinger, Data monetization: Lessons from a retailer's journey., MIS Quarterly Executive 12 (2013).
[20] B. H. Wixom, J. W. Ross, How to monetize your data, MIT Sloan Management Review 58 (2017) 9–13. URL: https://search.proquest.com/docview/1885859901?accountid=14570.
[21] A. Tiwana, Platform Ecosystems: Aligning Architecture, Governance, and Strategy, Morgan Kaufmann, 2014. URL: https://doi.org/10.1016/C2012-0-06625-2.
[22] A. Hagiu, J. Wright, Multi-sided platforms, International Journal of Industrial Organization 43 (2015) 162–174.
[23] M. A. Cusumano, D. B. Yoffie, A. Gawer, The future of platforms, MIT Sloan Management Review 61 (2020) 26–34.
[24] Y. Yoo, O. Henfridsson, K. Lyytinen, The new organizing logic of digital innovation: An agenda for information systems research, Information Systems Research 21 (2010) 724–735. URL: https://doi.org/10.1287/isre.1100.0322.
[25] G. Mijsters, A Focus Area Maturity Model for Data as a Platform, Master's thesis, 2019.
[26] P. Kramberg, A. Heinzl, Data platform ecosystems, HMD Praxis der Wirtschaftsinformatik 58 (2021) 477–493.
[27] M. G. Jacobides, C. Cennamo, A. Gawer, Towards a theory of ecosystems, Strategic management journal 39 (2018) 2255–2276.
[28] V. Govindarajan, Decentralization, strategy, and effectiveness of strategic business units in multi-business organizations, Academy of Management Review 11 (1986) 844–856.
[29] J. Zabojnik, Centralized and decentralized decision making in organizations, Journal of Labor Economics 20 (2002). URL: https://doi.org/10.1086/323929.
[30] A. Halckenhaeusser, J. Foerderer, A. Heinzl, Platform governance mechanisms: An integrated literature review and research directions, in: Proceedings of the 28th European Conference on Information Systems (ECIS), 2020, pp. 15–17.
[31] A. Zutshi, A. Grilo, T. Nodehi, The value proposition of blockchain technologies and its impact on digital platforms, Computers & Industrial Engineering 155 (2021). URL: https://doi.org/10.1016/j.cie.2021.107187.
[32] J. Pereira, M. M. Tavalaei, H. Ozalp, Blockchain-based platforms: Decentralized infrastructures and its boundary conditions, Technology Forecasting & Social Change 146 (2019) 94–102. URL: https://doi.org/10.1016/j.techfore.2019.04.030.
[33] C. Catalini, J. S. Gans, Some simple economics of the blockchain, Communications of the ACM 63 (2020) 80–90.
[34] M. Travizano, C. Sarraute, M. Dolata, A. M. French, H. Treiblmaier, Wibson: A case study of a decentralized, privacy-preserving data marketplace, in: H. Treiblmaier, T. Clohessy (Eds.), Blockchain and Distributed Ledger Technology Use Cases, Springer, Cham, 2020, pp. 149–170. URL: http://dx.doi.org/10.1007/978-3-030-44337-5_8.
[35] F. Glaser, F. Hawlitschek, B. Notheisen, Blockchain as a platform, Business Transformation through Blockchain: Volume I (2019) 121–143.
[36] M. Rauchs, A. Glidden, B. Gordon, G. C. Pieters, M. Recanatini, F. Rostand, K. Vagneur, B. Z. Zhang, Distributed Ledger Technology Systems – A Conceptual Framework, Technical Report, Cambridge Centre for Alternative Finance, 2018.
[37] G. W. Peters, E. Panayi, Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money, in: Banking Beyond Banks and Money, 2016, pp. 239–278.
[38] D. Trabucchi, A. Moretto, T. Buganza, A. MacCormack, Disrupting the disruptors or enhancing them? how blockchain reshapes two-sided platforms, Journal of Production Innovation Management 37 (2020) 552–574. URL: https://doi.org/10.1111/jpim.12557.
[39] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Decentralized business review (2008).
[40] V. Buterin, et al., A next-generation smart contract and decentralized application platform, white paper 3 (2014) 2–1.
[41] O. Ross, J. R. Jensen, T. Asheim, Assets under tokenization, Available at SSRN 3488344 (2019).
[42] L. Oliveira, L. Zavolokina, I. Bauer, G. Schwabe, To token or not to token: Tools for understanding blockchain tokens, in: International Conference on

Information Systems (ICIS), 2018.
[43] W. Entriken, D. Shirley, J. Evans, N. Sachs, Eip-721: Erc-721 non-fungible token standard, Ethereum Improvement Proposals 721 (2018).
[44] D. Jones, C. Snider, A. Nassehi, J. Yon, B. Hicks, Characterising the digital twin: A systematic literature review, CIRP Journal of Manufacturing Science and Technology 29 (2020) 36–52.
[45] A. Fuller, Z. Fan, C. Day, C. Barlow, Digital twin: Enabling technologies, challenges and open research, IEEE access 8 (2020) 108952–108971.
[46] Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, M. J. Deen, A novel cloud-based framework for the elderly healthcare services using digital twin, IEEE access 7 (2019) 49088–49101.
[47] S. Studer, T. B. Bui, C. Drescher, A. Hanuschkin, L. Winkler, S. Peters, K.-R. Müller, Towards crisp-ml (q): a machine learning process model with quality assurance methodology, Machine learning and knowledge extraction 3 (2021) 392–413.
[48] S. Welten, Y. Mou, L. Neumann, M. Jaberansary, Y. Y. Ucer, T. Kirsten, S. Decker, O. Beyan, A privacy-preserving distributed analytics platform for health care data, Methods of information in medicine 61 (2022) e1–e11.
[49] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial intelligence and statistics, PMLR, 2017, pp. 1273–1282.
[50] J. Schmeiss, K. Hoelzle, R. P. Tech, Designing governance mechanisms in platform ecosystems: Addressing the paradox of openness through blockchain technology, California Management Review 62 (2019) 121–143.
[51] J. J. Jones, A dataset for the study of identity at scale: Annual prevalence of american twitter users with specified token in their profile bio 2015–2020, PLoS One 16 (2021) e0260185.
[52] M. Dowling, Is non-fungible token pricing driven by cryptocurrencies?, Finance Research Letters 44 (2022) 102097.
[53] T. McConaghy, Ocean protocol: Tools for the web3 data economy, in: Handbook on Blockchain, Springer, 2022, pp. 505–539.