Validated Data Quality Assessment with "Skin in the Game": A Smart Contract Approach

Stefan W. Driessen[1(✉)], Geert Monsieur[1], Willem-Jan van den Heuvel[1], and Damian A. Tamburri[2]

[1] Jheronimus Academy of Data Science - Tilburg University, Sint Janssingel 92, 5211 'S-Hertogenbosch, DA, The Netherlands
s.w.driessen@jads.nl

[2] Jheronimus Academy of Data Science - Eindhoven University of Technology, Sint Janssingel 92, 5211 'S-Hertogenbosch, DA, The Netherlands

Abstract. Data Markets are becoming increasingly popular but are very challenging to deploy and maintain successfully. We discuss some of the challenges related to the success of data markets, focusing particularly on the diverse challenge of assessing data quality. We introduce a novel, holistic approach whereby a blockchain-based smart contract called a *Quality Assessment contract* allows an actor called the *quality assessor* to assess the quality of a data asset, provide immutable proof of their efforts on the blockchain, and get rewarded for their efforts proportionally to the value of their quality assessment efforts. We discuss how such an approach could be used in practice to assess the quality of different data assets and discuss some architectural considerations for using a quality assessment contract.

Keywords: Data markets · Quality assessment · Blockchain · Smart contracts

1 Introduction

There has been an explosive rise in the availability of data [12,24,32], an accompanying increase in analytical tools [27,29] and, above all, the growing interest in data as a tradeable, valuable good which can be leveraged to improve or enable business processes [3,38]. Consequently, data markets have become both heavily investigated by the academic community as well as industry [30]. The potential value of leveraging big data has been estimated to grow from $138.9 billion in 2020 to $229.4 billion by 2025 [13]. Anticipating this, initiatives have been put forward to build the infrastructure necessary for complex data markets [6,19,21].

It seems, however, that data markets are very challenging to deploy and operate successfully, as demonstrated by the large number of data markets that have come and gone in the last decade alone [28,31,33]. Existing literature has

J. Barzen (Ed.): SummerSOC 2021, CCIS 1429, pp. 119–130, 2021.
https://doi.org/10.1007/978-3-030-87568-8_8

identified a multitude of challenges for data markets: On the one hand there are economic and societal challenges, such as a lack of willingness amongst customers to pay for data [10,22,33] and the inability of legal codes and law enforcement to cover the intellectual rights or address privacy concerns [11,20]. On the flip side there are challenges that concern the design and mechanisms of data markets, which are of a more technical nature. Based on our understanding of existing literature in data markets, these challenges come in four general categories:

1. Maintaining *data sovereignty*: Data is, by its very nature, easy to duplicate and manipulate. This makes it hard to verify how, where and by whom the data, once it has left the data seller's control, will be used. A lack of data sovereignty can mean violation of privacy or legal requirements, unauthorised reselling of a data product or unintended use of the data [7,20]. This challenge is also tied to data security, as the more data is shared, the more vulnerable it becomes to being stolen.
2. *Data quality* assessment: Data is an *experience good*, meaning that its value is highly dependent on the user, the application and their context [11]. Because of this, we view data quality in this context as more than simply a measure of how complete, feature-rich or frequently updated a data asset is. Instead we informally define data quality as a measure for how well-suited the data is for its intended purpose. This definition is in line with previous work that has noted that "*Data characteristics have no judgemental value. Without considering them in a specific context ..., they are neither good nor bad and describe only characteristic properties.*" [25]. Taking this approach to data quality allows us to consider different aspects of data quality appropriately such as: on the one hand, a real-time streaming data asset for financial trading applications, which prioritises speedy access over completeness and, on the other hand, a large, feature-rich data set on household income in a nation used for finding economic trends, which might be only updated once every year.
3. *Product recommendation and querying*: As in any (digital) market, potential data buyers need a means to browse and select from amongst all possibilities those data assets that best suit their needs. Because of their prevalence outside the scope of data and data markets, recommender systems are perhaps the best understood, and certainly the most researched of all challenges, see e.g. [26].
4. *Price determination algorithms*: The challenge of accurately assessing data quality, as well as a lack of precedent on pricing comparable products make it hard to set a price for data assets [11].

In this paper we aim to address the challenge of data quality assessment, which we deem to be crucial for both price determination and developing an optimal querying mechanism. After all, in an ideal scenario the price of a data asset is highly dependent on its quality and any recommender system should prefer high-quality data assets over lower quality data assets. We find current literature on quality assessment for data markets to be lacking. With this paper, we hope to take a careful first step towards changing that.

We introduce a novel, holistic, approach, whereby a blockchain-based smart contract, called a *Quality Assessment (QA) contract*, allows an actor called the *quality assessor* to assess the quality of a data asset, provide immutable proof of their efforts on the blockchain and get rewarded for their efforts proportionally to the value of their quality assessment efforts. We believe that the QA contract should be easy to incorporate in any data market platform which leverages blockchain and smart contract technology as part of its design.

One of the major strengths of our approach is that the contract poses few restraints on quality assessment methods: We believe there does not exist a single data quality assessment approach that works well for all types of data. Instead we allocate the task of quality assessment to the quality assessor, who can prepare a variety of test plans. This makes our approach capable of addressing the many facets of data quality, doing so in a transparent, cheap-to-implement manner.

We explain our approach further in the next sections of this paper: in particular, Sect. 2.1 discusses the background behind blockchain technology in data market design, focusing on the motivations behind their use, as well as discussing the most relevant state-of-the-art work. Section 2.2 introduces the idea of staking, which we use to ensure that the quality assessor puts real effort into providing valuable insights on the data quality. In Sect. 3 we illustrate the types of quality assurance tests that our proposal enables. Next, in Sect. 4 we propose the a smart contract, called the *curation contract*, that logs the quality assessment efforts and rewards quality assessors proportionally to the value they add to the data market ecosystem. In Sect. 5 we draw a conclusion and sketch a short road-map for further research.

2 Background

In this section we explain key concepts of blockchain that are relevant for data markets as well as the concept of staking, both of which play an important role in our proposal. Smart contracts allow stakers to prove quality assessment efforts and facilitate rewarding good quality assessment, whereas staking demonstrates that the quality assessor is convinced of the quality of the data asset they have assessed.

2.1 Blockchain Technology and Data Markets

Blockchain technology has become popular in recent years as researchers and practitioners alike are starting to discover scenarios in which it makes sense to leverage its qualities (and when its downsides outweigh the pros). In essence, a blockchain is a ledger which stores information in transactions, which are added in batches called *blocks*. The key principle of blockchains is that they are maintained and manipulated in a *distributed* and *decentralised* manner. Distributed in this context means that the information in the blockchain is stored in multiple physical locations and decentralised means that no single party is in charge of the ledger, which instead is governed by a *consensus protocol*, which all participants adhere to [2]. Most public blockchains also have a *cryptocurrency* associated with them (e.g., Bitcoin [14]) which is a digital token that represents real-world value and whose ownership is tracked on the blockchain. Some blockchains, most prominently Ethereum [37], allow developers to store and manipulate pieces of code called *smart contracts*. Smart contracts are particularly interesting because the distributed and decentralised nature of the blockchain allow them to be both completely transparent and autonomous in their execution.

It can be no surprise that blockchain technology and smart contracts are actively being investigated in the context of data markets, see e.g. [1,9,12,23, 34,35]. As motivated by these researchers, blockchain- and smart contract technology bring several advantages to motivate their use for data markets. We discuss some of these advantages below:

1. The transactional nature of blockchains is well suited for implementing payments and provides an immutable record of all transactions for verification.
2. The transparency of blockchain technology adds to the transactional nature by allowing for verification of both the behaviour of buyers and sellers, as well as the inner workings and logic of those aspects of the data market that are coded in smart contracts on the blockchain.
3. Because smart contracts execute autonomously and their logic is fully transparent, they can take on the role of a trusted third party in exchange, where participants might not be inclined to trust each other.
4. Blockchain infrastructure operates similarly to cloud infrastructure and this makes it easy to deploy new smart contracts or disable old smart contracts on the fly. Automation, combined with convenient deployment can also lead to a reduction of the costs associated with operating a data market platform.

2.2 Staking on Blockchains

Staking is an activity, whereby an owner or *staker* of some cryptocurrency communicates to the blockchain network that they are locking their cryptocurrency tokens, effectively making it impossible for them to spend those tokens. Staking is usually tied to some process on the blockchain and is done to signal that the staker is committed to a positive outcome of the process, either because they will lose something if the outcome is bad, or because they stand to gain something if the outcome is good. The most prevalent example of staking is the Proof-of-Stake (PoS) consensus protocol [36] but other endeavours have been proposed to leverage staking for quality assessment, such as token-curated registries [8] (TCRs). TCRs are lists, which are curated by members (stakers) who have all staked some tokens. Members can vote on which items should be on the list, and their vote is proportional to the amount of tokens they've staked. In order to become a member, an applicant has to both buy tokens and be voted in by existing members. TCRs have been proposed to curate lists of data sources [23]; if a curated list is held in high regard, owners of data assets will want to put it on the list, in order to propose their asset they will want to become members, which means buying tokens. Thus, if the quality assessment efforts are valuable, demand for the tokens will increase leading to an increase in the value of the staked tokens.

Another example may be drawn from the Ocean Protocol [15], which is a data market protocol with a blockchain back-end for registering, browsing and purchasing data assets. Ocean users can stake tokens on specific data assets, signalling that they believe that it is of high quality and likely to be purchased[1]. Every time a data asset is purchased, the purchaser pays a small fee which is divided amongst the stakers proportionally.

Based on these examples, we have obtained some initial evidence that the staking holds the promise to add significant value to a data market by demonstrating "investedness" in the value of quality assessment efforts, i.e., stakers who provide useful quality assessment will stand to gain from this whereas stakers who do not provide useful quality assessment might lose money. Our proposal expands upon this idea, because it enables quality assessors to not only signal investedness, but also prove that the staking is tied to some real quality assessment efforts.

3 Quality Assessment Tests

As mentioned above, one of the most important features of our suggested approach is the notion that different data assets require different types of quality assessment. Before discussing the QA contract, we illustrate some types of quality assessment methods we envision that are supported by our approach.

We believe quality assessment is best achieved when tests, captured in a set of instructions, are run against a data asset and record a result that demonstrates the quality of that data asset. We call these tests *Quality assessment (QA) tests.* When deciding on appropriate QA tests, the quality assessor considers tests that are, on the one hand, *appropriate* to demonstrate the value of the data asset and, on the other hand, *respect data sovereignty* as explained in Sect. 1.

In line with our view of quality assessment, deciding which QA tests are appropriate depends on the envisioned use of the data asset. For example, if a data asset is promoted for training machine learning algorithms, it makes sense to test one or more common algorithms on the data set and record both which algorithms were tested, as well as their resulting performance (e.g., accuracy, recall, etc.). On the other hand if a data asset provides pay-per-query real-time global weather information, a good test could simply be to request information on several different locations, compare the results to other available sources, and log the places queried, as well as the comparison on the blockchain.

Respecting data sovereignty means that not only the test results but also the tests themselves can be safely shared with potential buyers without infringing on the rights of data providers. This can be a major limitation whenever a QA test requires additional data sources: either the additional data itself should be

[1] In reality, staking in the Ocean data marketplace serves an additional purpose: price determination through the use of automated market makers. A full discussion of price determination is beyond the scope of this paper.

shareable (e.g., because it is in the public domain), or the data market should be able to combine multiple data assets. We illustrate this by giving some examples.

Purchasing an Entire Data Set
A very straightforward example would be a data set which, after purchase, can be downloaded as a whole and can be used, but not shared, by the data buyer. Examples of such a data set would be mailing lists (e.g., the ones offered on Data and Sons [17]). In this case, the best test would be for a quality assessor to buy the data set and manually check whether it is complete and if the metadata describing the data asset is accurate. In this case the "result" of the test would be a confirmation of the purchase, which *can* be logged on the blockchain. In this scenario the act of staking is an important signal of the quality of the data: despite the fact that the test result seemingly does not convey much information about the data asset, the quality assessor, who has demonstrably purchased the data, indicates that they believe the data set to be likely to be purchased.

Pay-Per-Query
Some data assets, such as the public data sets offered through Google BigQuery [18], can be accessed through a "pay-per-query" payment model. For such a data set, tests can be designed to assess the metadata of the data asset through aggregate results, such as number of rows, dimension of the data or the number of missing values. Additionally, if the marketplace supports it, queries can attempt to join different data assets and record the success or failure of the join operator. The test results can then validate and extend the metadata already available on the marketplace, as well as demonstrate the integrability of the offered data assets.

Real-Time-Data
We, once again, consider the example from the introduction of a real-time streaming data asset for financial applications, such as the Binance API for cryptocurrency trading [16]. In this case, test cases might be designed to assess the timeliness of the response and the test results could be timestamps, indicating the time between receiving the request and the response. Since the data quickly loses its value after publication, the data provider might not even mind if some actual query responses are shown on the blockchain which would demonstrate that the data asset works as intended.

Machine Learning with Compute-to-Data
Some data markets (e.g., [12,15,21]) aim to deliver a secure computation environment so that data never has to leave the control of the data provider. Instead the algorithmic computation is brought to the data and only aggregated results are returned to the data buyer. In this scenario our approach would be especially easy to implement, as the aggregate results can generally be considered to be the property of the data quality assessor who can share them as they please. An example for such a scenario would be to train some well-known algorithms on a data product (e.g., K-nearest neighbours, support-vector machines or random

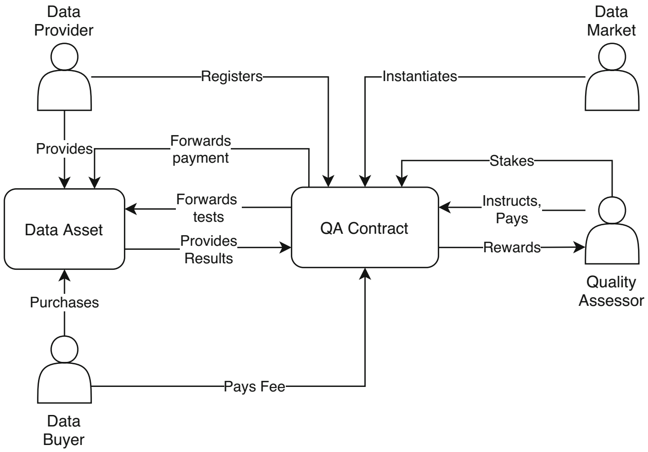

Fig. 1. The relation between the different actors, the data asset and the QA contract in the data market ecosystem

forests in the case of classification) and have test results consisting of performance metrics such as accuracy, recall, etc. The data product could even be enriched with public data sets which can be safely recorded as part of the tests.

4 The Quality Assessment Contract

In this section we introduce the QA contract, which is the key element that enables validated QA testing. Our approach has been designed to leverage the blockchain advantages mentioned in Sect. 2.1: the transactional nature facilitates, in principle, easy payment to quality assessors and makes it relatively simple to find and identify the quality assessment efforts. The transparency and trust allow us to demonstrate that quality assessment was actually conducted, and the easy deployment allows data sellers to easily leverage our solution.

We start out Sect. 4.1 by describing the relevant interactions that the different actors (e.g. data provider, data buyer, etc.) and artefacts (i.e. the data asset and QA contract) have with each other in the process of using the QA contract. The actors, artefacts and interactions are shown in Fig. 1. After outlining the interactions, we discuss some architectural considerations in Sect. 4.2, specifically we explain which parts occur on the blockchain and how these interact with the off-chain parts.

4.1 Functionality of the QA Contract

The QA contract is a smart contract which is deployed by the data market for the benefit of data providers, data quality assessors and data buyers. The choice of whether to use a QA contract lies with the data provider, who has to register their asset in the QA contract after making it available on the data marketplace. As discussed in Sect. 3, the data asset can be a wide variety of offerings, ranging from a query-able database to a data set which can be purchased as a whole or even a subscription to a data service.

When a data asset is registered, the smart contract assigns it a unique identifier, which it stores for future communication. The data provider makes sure that the data asset can receive instructions (such as purchase requests) from transactions that are send to the QA contract (Sect. 4.2 explains this more in-depth). Finally, the data market ensures that the "front-end" (i.e., the environment where potential buyers can view information about the data asset) is updated so data quality assessments can more easily be found.

After the data asset has been registered, data quality assessors can come up with QA tests that are both appropriate and respect data sovereignty. After deciding on such QA tests, they send instructions for executing the tests, along with payment for accessing the data asset, to the QA contract in a transaction on the blockchain. As discussed in Sect. 3, how these QA tests are executed and what the test results look like are highly dependent both on the type of data asset, as well as the data market, but we can assume that some appropriate result, is returned which can be logged on the blockchain.

The QA contract then forwards the payment to the data provider and the instructions for the tests to the data asset. After the tests are completed, the results are sent, in a transaction, to the QA contract[2]. Since the purchase, the test instructions and the test results are all communicated through the blockchain, they are logged in transactions to- and the smart contract which are transparent and constitute immutable proof of the purchase, the test and the test results.

After seeing the results of their tests, the staker decides whether they believe the data asset has a high quality. If they do believe the data asset has a high quality, they can decide to stake some cryptocurrency on it, through the QA contract. The QA contract verifies that the staker indeed performed some tests before accepting the stake and keeps track of which quality assessor has staked how much on each data asset.

Data buyers who are interested in the data asset can view the QA tests that have been applied, as well as the corresponding results. Since these are logged in blockchain transactions, the data buyer will know that the tests actually occurred. The QA tests can assist the data buyer in their decision on whether or not to buy the data asset. If they do decide to purchase the data asset, a small fee is charged, which is distributed amongst all quality assessors who have run tests and have staked on the data asset. The exact distribution key is an interesting problem in-and-of-itself. This will, most likely, depend partially on the size of the stake of the quality assessor.

[2] Of course, sending this transaction costs cryptocurrency, we presume that the quality assessor compensates the data asset for this transaction, but different implementations are also possible depending on the data market design.

4.2 Architectural Considerations

When implementing the QA contract it is important to consider which elements are on- and which are off the blockchain. The positive qualities we introduced in Sect. 2.1 come at the cost of increased cost of operation [2] and increased cost for debugging and maintenance [4]. We therefore end our discussion of the QA contract by discussing some of the architectural considerations of implementing the QA contract in a data market environment. Note, that we believe that any good data market should also offer a front-end (e.g., by using a website), which can be used to interact with the blockchain-based elements that act as a back-end. In order to further illustrate the QA contract, we provide a mock-up implementation in Solidity in our online appendix[3].

In order for the QA contract to properly function, it is important that the *payment* and the *communication* of the *QA tests* and *results* are done through the blockchain. Because of this, *at the very least*, the data asset and the quality assessor need cryptocurrency wallets on the blockchain. The quality assessor needs to be able to provably pay for the data asset and receive rewards whenever a fee is paid by a future data buyer and the data asset needs to provably receive payment and return transactions with the test results. Ideally, the data buyer would have a cryptocurrency wallet as well, so they can purchase the data asset through the blockchain, but the data market platform might offer the option to pay in fiat currency and take care of the curation fee "behind the scenes".

It was already alluded to in Sect. 4.1 that the data asset has to be able to receive instructions from the QA contract. The easiest way to achieve this, would be to have the QA contract forward a transaction to the wallet associated with the data asset each time it receives test instructions from the quality assessor. The data asset can then (automatically) search for the blockchain for the most recent transactions to the QA contract that mention its unique identifier and extract the instructions from there, for example in Ethereum, the quality assessor could include the instructions in the "data" field of its transaction to the QA contract [5].

Finally, it is important to note the data buyer can access the QA tests and results directly from the blockchain without needing their own account. Since it is not desirable potential data buyers scrape the blockchain for relevant transactions, the data market should provide links (and possibly recaps) of the relevant transactions in the overview of the data asset on the front-end of the market.

5 Conclusion

In this paper we discussed the problem of data quality assessment, taking particular notice of the multi-faceted nature of data quality. We suggested some simple quality assessment tests and showed how such tests, in combination with our QA contract have the potential to add real value to a data market ecosystem by providing provable insights into the quality of data assets.

[3] https://github.com/Pindapinda/QA-Contract.

In future work, we intend to evaluate the concept of a QA contract in a real-world setting by seeking out industrial partnerships. We envision that the development of appropriate and data sovereignty respecting QA tests will be particularly relevant for future data market-related research.

References

1. Bajoudah, S., Dong, C., Missier, P.: Toward a decentralized, trust-less marketplace for brokered IoT data trading using blockchain. In: Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019, pp. 339–346 (2019). https://doi.org/10.1109/Blockchain.2019.00053
2. Butijn, B.J., Tamburri, D.A., Heuvel, W.J.V.D.: Blockchains: a Systematic Multivocal Literature Review (2019). http://arxiv.org/abs/1911.11770
3. Demchenko, Y., Cushing, R., Los, W., Grosso, P., De Laat, C., Gommans, L.: Open data market architecture and functional components. In: 2019 International Conference on High Performance Computing and Simulation. HPCS 2019, pp. 1017–1021 (2019). https://doi.org/10.1109/HPCS48598.2019.9188195
4. Driessen, S., Nucci, D.D., Monsieur, G., Tamburri, D.A., Heuvel, W.J.V.D.: AGSolT : a Tool for Automated Test-Case Generation for Solidity Smart Contracts, pp. 1–15 (2021). https://arxiv.org/abs/2102.08864
5. Eiki (2019). https://medium.com/@eiki1212/ethereum-transaction-structure-explained-aa5a94182ad6
6. Exner, J.P.: The ESPRESSO - Project - A European approach for Smart City Standards (2011)
7. Filippi, P.D., Mccarthy, S.: Cloud computing?: centralization and data sovereignty. Eur. J. Law Technol. **3**(2), 1–18 (2012)
8. Goldin, M.: Token-Curated Registries 1.0 (2021). https://docs.google.com/document/d/1BWWC__-Kmso9b7yCI_R7ysoGFIT9D_sfjH3axQsmB6E/edit
9. Gupta, P., Kanhere, S.S., Jurdak, R.: A decentralized IoT data marketplace (2019)
10. Hayashi, T., Ohsawa, Y.: Preliminary case study on value determination of datasets and cross-disciplinary data collaboration using data jackets. Procedia Comput. Sci. **112**, 2175–2184 (2017)
11. Koutroumpis, P., Leiponen, A., Thomas, L.D.W.: Markets for data. Ind. Corporate Change **29**(3), 645–660 (2020). https://doi.org/10.1093/icc/dtaa002
12. Koutsos, V., Papadopoulos, D., Chatzopoulos, D., Tarkoma, S., Hui, P.: Agora: a privacy-aware data marketplace. In: Proceedings - International Conference on Distributed Computing Systems 2020-Novem, pp. 1211–1212 (2020). https://doi.org/10.1109/ICDCS47774.2020.00156
13. Market Reports World: Big Data Market by Component, Deployment Mode, Organization Size, Business Function (Operations, Finance, and Marketing and Sales), Industry Vertical (BFSI, Manufacturing, and Healthcare and Life Sciences), and Region - Global Forecast to 2025. Technical Report (2020). https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html
14. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system (2009)
15. Ocean Protocol Foundation: The Ocean Protocol (2019). https://oceanprotocol.com/tech-whitepaper.pdf#h.uvec5xf3qzn1
16. Online: Binance API. https://binance-docs.github.io/apidocs/spot/en/#introduction
17. Online: Data and Sons. https://www.dataandsons.com/

18. Online: Google BigQuery Public Data Sets. https://cloud.google.com/bigquery/public-data/
19. Online: GAIA-X (2021). https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html
20. van Ooijen, I., Vrabec, H.U.: Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. J. Consum. Policy **42**(1), 91–107 (2019). https://doi.org/10.1007/s10603-018-9399-7
21. Otto, B., et al.: Industrial Data Space - digitale Souveränität über Daten. Whitepaper (2016). https://www.fraunhofer.de/de/forschung/fraunhofer-initiativen/international-data-spaces.html
22. Potoglou, D., Patil, S., Palacios, J.F., Feijóo, C., Gijón, C.: The value of personal information online: results from three stated preference discrete choice experiments in the UK. In: ECIS 2013 - Proceedings of the 21st European Conference on Information Systems, pp. 1–12 (2013)
23. Ramachandran, G.S., Radhakrishnan, R., Krishnamachari, B.: Towards a decentralized data marketplace for smart cities. In: 2018 IEEE International Smart Cities Conference. ISC2 2018, pp. 1–8 (2019). https://doi.org/10.1109/ISC2.2018.8656952
24. Reinsel, D., Gantz, J., Rydning, J.: Data Age 2025: The Digitization of the World From Edge to Core. Seagate, IDC (November), vol. 28 (2018). https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf
25. Reiter, M., Breitenbucher, U., Dustdar, S., Karastoyanova, D., Leymann, F., Truong, H.L.: A novel framework for monitoring and analyzing quality of data in simulation workflows. In: 2011 IEEE Seventh International Conference on eScience, pp. 105–112. IEEE (2011)
26. Ricci, F., Rokach, L., Shapira, B.: Recommender Systems Handbook, 2 edn. Springer, Boston (2015). https://doi.org/10.1007/978-1-4899-7637-6, https://link.springer.com/book/10.1007%2F978-1-4899-7637-6#about
27. Rodríguez-Mazahua, L., Rodríguez-Enríquez, C.A., Sánchez-Cervantes, J.L., Cervantes, J., García-Alcaraz, J.L., Alor-Hernández, G.: A general perspective of big data: applications, tools, challenges and trends. J. Supercomputing **72**(8), 3073–3113 (2016). https://doi.org/10.1007/s11227-015-1501-1
28. Schomm, F., Stahl, F., Vossen, G.: Marketplaces for data: an initial survey. SIGMOD Rec. **42**(1), 15–26 (2013). https://doi.org/10.1145/2481528.2481532
29. Sharma, S.: Rise of big data and related issues. In: 12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3–C3). INDICON 2015, pp. 1–6 (2016). https://doi.org/10.1109/INDICON.2015.7443346
30. Spiekermann, M.: Data marketplaces: trends and monetisation of data goods. Intereconomics **54**(4), 208–216 (2019). https://doi.org/10.1007/s10272-019-0826-z
31. Stahl, F., Schomm, F., Vossen, G.: Data marketplaces: an emerging species. Front. Artif. Intell. Appl. Databases **2013**, 145–158 (2014). https://doi.org/10.3233/978-1-61499-458-9-145
32. Stahl, F., Schomm, F., Vossen, G., Vomfell, L.: A classification framework for data marketplaces. Vietnam J. Comput. Sci. **3**(3), 137–143 (2016). https://doi.org/10.1007/s40595-016-0064-2
33. Stahl, F., Schomm, F., Vossen, G., Vomfell, L.: A classification framework for data marketplaces. Vietnam J. Comput. Sci. **3**(3), 137–143 (2016). https://doi.org/10.1007/s40595-016-0064-2

34. Travizano, M., Ajzenman, G., Sarraute, C., Minnoni, M.: Wibson: A Decentralized Data Marketplace, pp. 1–6 (2018)
35. Uriarte, R.B., De Nicola, R.: Blockchain-based decentralised cloud/fog solutions?: challenges, opportunities and standards. IEEE Commun. Standards Mag. (September) (2018). https://doi.org/10.1109/MCOMSTD.2018.1800020
36. Li, W., Andreina, S., Bohli, J.-M., Karame, G.: Securing proof-of-stake blockchain protocols. In: Garcia-Alfaro, J., Navarro-Arribas, G., Hartenstein, H., Herrera-Joancomartí, J. (eds.) ESORICS/DPM/CBT-2017. LNCS, vol. 10436, pp. 297–315. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67816-0_17
37. Wood, G.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper, pp. 1–32 (2014)
38. Zech, H.: Data as a tradeable commodity-implications for contract law. In: Proceedings of the 18th EIPIN Congress: The New Data Economy between Data Ownership, Privacy and Safeguarding Competition (2017)