Towards Informational Self-determination: Data Portability Requests Based on GDPR by Providing Public Platforms for Authorised Minimal Invasive Privacy Protection

Dominik Schmelz(✉), Karl Pinter, Phillip Niemeier, and Thomas Grechenig

Industrial Software (INSO), Vienna University of Technology, 1040 Vienna, Austria
{dominik.schmelz,karl.pinter,phillip.niemeier,
thomas.grechenig}@inso.tuwien.ac.at

Abstract. The Universal Declaration of Human Rights (UDHR) defines that no human being should be subjected to arbitrary interference with his privacy. Yet last decade's digital platform progress has been legally widely unframed and untamed. Therefore, both collection and commercial use of personal data has become a widespread and profitable business model in which individuals currently practically have very little power. European Union's General Data Protection Regulation (GDPR) rebalances rights and obligations of data controllers processing personal data and data subjects whose data are being processed. Well-tailored and targeted use of blockchain technologies enables system transactions that strengthen individual regain of control over personal data and securely transfer it. The proposed system (PPAMIPP, public platform for authorised minimal invasive privacy protection) allows data subjects to claim the personal data processed by them and request their transfer in accordance with the GDPR by defining a respective novel process and supporting technical architecture. The proposed system is validated using a prototype implementation. In addition to demonstrating the feasibility of the system while maintaining confidentiality and integrity, the trade-offs between privacy and usability, as well as general problems of the defined process from legal and technical viewpoints, are highlighted.

1 Introduction

Technological advances allow for more data being recorded, transferred and processed in a shorter time [24]. Technologies such as Internet of Things (IoT) record profileable information about people, whereas Big Data applications allow data analysis in near real-time. These advances call for regulation as developments in information technology have raised concerns about information privacy, and its implications [2].

The European Union's regulation 2016/679, better know as General Data Protection Regulation (GDPR) has enabled on the one hand citizens of the

J. Prieto et al. (Eds.): BLOCKCHAIN 2021, LNNS 320, pp. 106–116, 2022.
https://doi.org/10.1007/978-3-030-86162-9_11

European Union to exercise given rights regarding their personal data [1, Art. 12–23], and on the other hand, forces companies, among other things, to secure data processing activities and enforce a certain level of transparency [1, Art. 24–43].

The GDPR grants a variety of rights to the data subject, including, among other things, the right of data subjects to request information about what data is being processed, for what purposes and by whom [1, Art. 15]. The GDPR also allows the data subject to apply for data portability [1, Art. 20]. These requests must be processed by the controller within one month of receipt [1, Art. 12 (3)]. Only in complex cases, the controller can extend the period to a maximum of three months [1, Art. 12 (3)].

Digital requests for data portability and answers to these request are not further defined in the GDPR on a technical level. Currently, most companies either implement a privacy web form to be filled out online or provide an email address for customers to contact.

This bares these three major tensions and problems:

- Easy, but secure transfer of data
- Secure authentication of the data subject, without further data disclosure
- Reproducible requests, without public disclosure of personal information

These problems lead to time-consuming processes, both for the controller and for the data subject. Most often, they also result in insecure processes that transfer personal data in an insecure manner and additionally lack proof of submission. The currently most implemented process for data portability is to write an email to a data protection officer who then sends the data via unencrypted email [13].

Blockchain technology, originally used in cryptocurrencies, is used in many fields such as logistics, identity management and insurances. The technology makes it possible to create a transparent, nearly immutable, decentralised record database appendable by multiple parties. In general, one advantage of blockchain can be seen when the parties involved share a distrust of each other but need to work together so that each party can rely on the data and applications stored on the blockchain rather than trusting each other [12].

The contribution of this publication is a structured analysis of the issue at hand and a solution proposal for the discussed problems. It introduces a technical concept and prototype implementation of a new solution that withstands the high expectations regarding data protection and privacy.

The paper is organised as follows. First, we present and analyse current research, solutions and background information (Sect. 2). Afterwards, we describe processes and roles relevant to the proposed solution (Sect. 3) and discuss the current state of the prototype implementation and architecture (Sect. 4). The data protection aspects of the prototype and decisions taken to implement privacy by design are discussed in Sect. 5. Next, we shed light on implications of the solution (Sect. 6). Finally, we reflect on the findings, draw the conclusion and discuss possible future work (Sect. 6).

2 Related Work

The research fields of this topic cover procedural, legal and technical areas, particularly e-government, blockchain technology, and data protection laws in the considered jurisdictions.

The prototype presented in this paper builds on previous work done by the authors and represents a further development in the direction of supporting individuals [13,14,18].

Identity Management is the process of ensuring that the right individuals have the proper access to resources. Governments and unions have been working on local and cross-country identification and authentication schemes for years. In 2014 the European Union started the process of cross-country identification, authentication, and trust services for electronic transactions (eIDAS) [11]. Sullivan et al. [20] have been working on blockchain and digital identity in the E-Government (E-Gov) environment. Ishmaev [5] researched Self-Sovereign Identity (SSI) solutions based on blockchain concerning sovereignty, privacy, and ethics. SSI developments on the basis of blockchain technology with compatibility to eIDAS are currently developed by the European Self-Sovereign Identity Framework (ESSIF) Lab[1] on the European Blockchain Services Infrastructure (EBSI).

In the field of E-Gov, there is a general trend toward digitising and optimising processes, focusing on data economy and data sovereignty [3,4]. The escalating collection of user and usage data has led to problems in terms of data protection and ethics, and laws such as GDPR are intended to counteract this. [19] therefore proposes a governance framework for data sovereignty. One problem is that governments around the world often use cloud service providers from other countries. This raises questions about data sovereignty, as [10] has explored.

Yuming et al. [25] dealt with the theory of data sovereignty. Fang et al. [4] worked on the topic of cyberspace sovereignty. Janssen et al. [6] clearly states that there is an urgent need for systems that supports data subjects in terms of their rights.

Data transfer via central data storages has been challenging. The ideal of end-to-end encrypted communication already existed in email communication (e.g. PGP) [9, p. 147ff] and also in the context of Blockchain [23]. File sharing platforms and cloud storages applied the concept in products such as Firefox Send. They either use a simple symmetric key transferred via a second channel or a private-public key pair. Tresorit[2] that added group access features with zero knowledge user authentication.

Blockchain has controversially presented solutions and problems regarding data protection at the same time. On the one hand, personal data processing operations on the blockchain have raised open legal questions with regard to GDPR [17] and on the other hand, consent management solutions, such as [15],

[1] https://essif-lab.eu.

[2] https://tresorit.com.

and [22], support the implementation and enforcement of the GDPR in the blockchain context.

The paper at hand uses the research mentioned above and combines it with a new approach to support individuals creating blockchain-based requests for data portability.

3 Privacy Enhanced Portability Process

A request for data portability is currently provided by means of an unstructured email or, at most, a proprietary portal. The authentication process is mainly performed manually. For lack of options, this is usually done by photographing or scanning the ID card or passport. This process has apparent shortcomings, such as the recipient being able to reuse the ID, and having no way to verify the ID. From a data protection perspective, the transmission of an ID card via a non-encrypted and non-integrity-protected channel is not recommended. Also, the requested data itself is oftentimes transferred via the same channels.

These factors lead to a process that is devastating from a data protection perspective. A system must be in place to help data subjects efficiently exercise their rights without putting their data at further risk.

The analysis of the GDPR Article 15 lead to the identification of the following roles, that were consequently used for the prototype:

- **Data Subject:** An identified or identifiable natural person to whom personal data relates.
- **Controller:** Is responsible for the means and purposes of the processing activities. Provides information in accordance with the GDPR.
- **Processor:** Processes personal data on behalf of the controller.
- **Data Protection Authority (DPA):** The DPA is a national supervisory authority for data protection in each country responsible for processing complaints.

According to the GDPR, data subjects have several rights concerning their personal data. These are, among other things, the right to information, access, rectification, withdrawal of consent, object, and the right to be forgotten. The authors developed a data request platform in the form of a prototype according to [1, Art. 20], namely the right of data portability. The data request platform uses a simple, standardised process for the data subject to create an inquiry faced to the controller or later in the process to the DPA if needed.

The standard process is illustrated in Fig. 1 and can proceed as follows: The data subject visits the data request platform, chooses a controller and submits info to request data including a customer ID or an equal identifier known to the controller. The data request platform generates a private and public key pair in the frontend and sends *only the public key* to the server (1). The data request platform client prints the private key without sending it to the server. The data request platform generates a Portable Document Format (PDF) with the request for data portability and link to the upload platform, including the public key of the data subject, the mail address of the data subject and a signature.

The data request platform asks the user to identify to and sign the PDF with an electronic IDentification, Authentication and trust Services (eIDAS) Identity Provider service (2). The data request platform puts the PDF in an email file and sends the email to the controller and data subject (4), and stores Simple Mail Transfer Protocol (SMTP) proof and a hash of the PDF on the blockchain using openTimeStamp (3). As a result, it can be proven later that the email with the provided content was sent to the data protection officer. This information may be provided later to the DPA in case of a dispute.

The email also contains a calendar entry file as a reminder to the data subject that a complaint can be lodged with the DPA after the expiry of the statutory deadline. The controller receives the request, clicks the link and uploads data (5). The data is automatically hybrid encrypted with the public key of the data subject. Hybrid encryption is used to increase the efficiency of large documents because in this method, the file is encrypted with a random symmetric session key, and this is asymmetrically encrypted with the recipient's public key. The data is only contained in clear text in the Controller User Interface (5a) and only sent encrypted to the Data Request Platform (5b). After the upload is finished, the controller receives proof that he or she uploaded the file using the openTimeStamp stored on the blockchain. The data request platform takes the email address out of the request, checks the signature and send a notification to the data subject.

The data subject receives an email with a link and opens it. The data subject enters the private key to retrieve the encrypted file (6a), which is decrypted in the Subject Interface (6b).

In case of a delay of more than three months (see Sect. 1), a complaint can be filed with the DPA, in which the proof is stored, and the DPA can carry out a review of the request (7). Included are the signature of the data subject, proof of the application including the time and proof of the delivery of the application to the data protection officer.

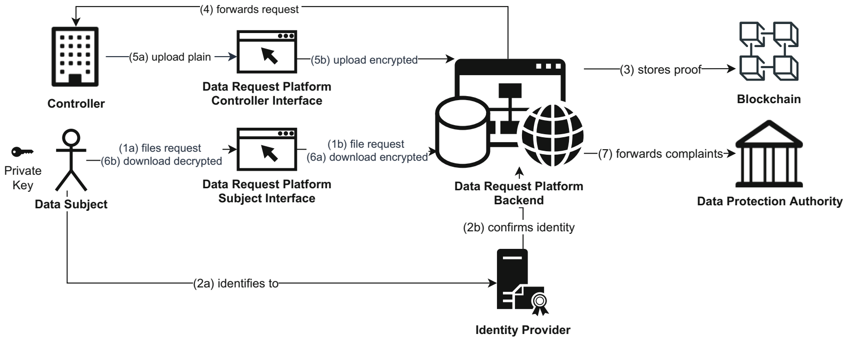

Fig. 1. Actors and processes in the context of the system.

With this mechanism, the following aims are achieved:

- The data subject is identified without disclosing much information to the data request platform or data controller
- The data is transferred with guaranteed confidentiality and integrity
- The data request platform has no access to the transferred files
- The request can be proven to authorities when filing a complaint with the authorities

Disclosure of less information is achieved through an identification system that does not reveal more information than needed for identification (eIDAS). eIDAS also provides for a legally binding authentication and signature on the request documents, and therefore prevents that a user can request data from a stranger. Confidentiality and integrity are obtained through the use of hybrid asymmetric encryption and blockchain stamping. This also provides the opportunity to prove the request to the authorities in case of escalation. The browser-based end-to-end encryption (E2EE) results in the property that the operator has no access to the transferred files.

4 Architecture and Prototype

The data request platform itself can be provided and managed by anyone (e.g. DPA, private companies, governments). It is not decentralised, but uses the decentralised blockchain to secure the data on it. It does not necessarily need a direct interface to the DPA nor to the controller. All national data protection agencies are contactable via email[3] while the contact details of each Data Protection Officer (DPO) must be published by the controller according to the GDPR, in many cases including an email address [1]. Therefore, the interfacing of the platform with these entities is executed via email.

Since the data request platform itself is processing personal data, it must comply with data protection regulations and be a role model regarding data protection implementation. It implements privacy by design and default and minimises the processing and storage of data. The platform does not store any information entered by the data subject but rather hashes the inputs as a filled-out PDF form or an email and deletes the original information after sending it to the controller and data subject via email. The filled-out form then is only saved by the data subject, not on the platform. For convenience reasons, a link containing the entered information is sent to the user, so he/she does not have to enter it again in case of a complaint (see Sect. 3).

The prototype that was created to demonstrate the potential of such a platform consists of a server-side application to provide basic functionalities regarding front-end content delivery, PDF generation, SMTP communication and interfacing with an Identity Provider, Notary Service and Object Storage (see Fig. 2).

[3] A list of all national data protection agencies of EU member states can be found on https://edpb.europa.eu/about-edpb/board/members_en.

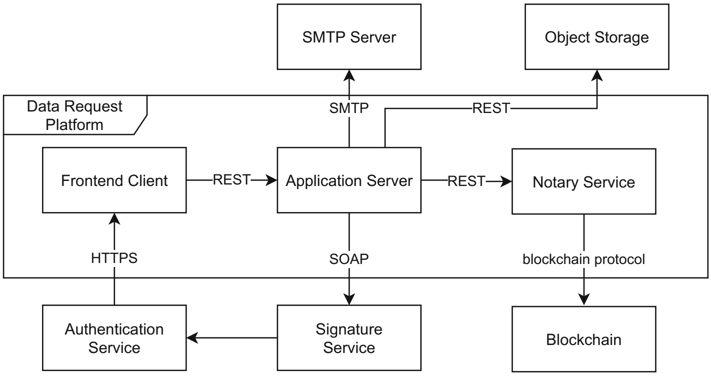

Fig. 2. Architectural overview of the system.

The Identity Provider provides an Authentication Service and a Signature Service. The prototype's front-end is realised in plain web technologies providing the user-facing forms and sending requests to the application server. In order to publish hashes on the Bitcoin Blockchain, the OpenTimestamps [21] protocol is used by a separate notary service. The OpenTimestamps protocol achieves scalability through aggregation. Hashes that are to be stored on the blockchain are collected on the OpenCalendar servers and committed in bulk as a Merkle tree root. This is done to minimise the transaction costs. This approach uses a public blockchain without additional transaction costs for more documents. Therefore a private blockchain is not used, nor recommended in this case.

As visualised in Fig. 2, the three components of the platform communicate via restful HTTP requests amongst themselves and with the authentication service, i.e. the respective provider of the eID. In the prototype the local eIDAS provider of Austria[4] was used. A further development could be the integration of an SSI such as eSSIF (see Sect. 2). The signature service provided by the authentication service is called via a SOAP interface. Requests to the data controllers are sent using the SMTP protocol. The communication with the blockchain is executed via the corresponding Blockchain protocol.

5 Data Protection

The data request platform processes personal data and therefore needs to be analysed regarding its processing activities. The platform was designed to implement 'Privacy by Default' and 'Privacy by Design' [8]. The data subject transmits the personal data to the platform, being a controller in the means of GDPR itself [1, Art. 4 (7)]. The data collection and processing happen on the basis of consent [1, Art. 6 (1)] of the data subject. The data is processed only for the purpose of creating the request and is deleted immediately after processing. No

[4] ID Austria https://eid.egiz.gv.at.

special categories of data are processed. The filled-out request is sent only to the controller. The responsible party uploads the data of the data subject to the data request platform in encrypted form. The operator or order processor of the data repository never has access to the data of the data subject because of the end-to-end encryption. The data at the order processor is deleted at the push of the "Delete Data" button by the data subject. The data subject has the right to view the processing operations at any time and to have his or her data restricted or deleted with immediate effect. The data stored on the blockchain only contains a hash and a timestamp of the requested and returned data. Neither does it contain any conclusions about the individual user. To exclude temporal conclusions, no IP addresses are stored in the logs of the server.

The blockchain service, therefore, does not store the document itself or any personal relatable information. It only receives a hash of the signed document and stores (as described in Sect. 4) the root of a Merkle tree of document hashes.

6 Implications and Conclusion

The current research shows blockchain technologies are used in several ways to achieve decentralised authority. Requests for data portability can be complicated, insecure and hard to prove to data protection authorities. The presented prototype shows that transparency and legal certainty can be created for all parties involved in a real-world environment. The process of data portability can be handled digitally and in a standardised way. Currently, there is no uniformity for such requests. The prototype presented solves this problem. With this prototype, the problems described in Sect. 1 can be solved. The request for data portability can be proven to authorities in the case of an escalation, and the data is transferred confidentially and with integrity while the data request platform has no access to the data. This mitigates a possible breach through an attack that results in a data disclosure.

Because of this, the data request platform can be hosted and run by governments, data protection authorities or private companies. The blockchain supports the data transfer in a way that the properties of the blockchain are partially inherited. The costs of the system are dependent on the usage and configuration. Data retention policies and automatic deletion were not implemented in the prototype but could lead to reduced hosting costs.

While the aforementioned features mainly concern the technical implications of the system, the research further revealed deficiencies in the current GDPR-relevant processes. The following issues were encountered during the process of the design and implementation of the prototype:

- Lack of definition of data protected authentication mechanisms;
- No defined communication processes, formats or interfaces;
- Lack of motivation for companies to make data easy to understand.

These can be overcome by a step by step migration to and introduction of a new platform but need legal backing in order to make them mandatory.

In designing the platform, design decisions had to be made regarding the privacy of the platform itself. In particular, there are trade-offs between privacy and usability that can make it challenging to introduce privacy features for widespread use of the system.

Specifically, the following trade-offs were found:

1. Local private key storage vs zero-knowledge platform;
2. Data transfer between process steps vs centralised storage of information;
3. Identification vs profiling of subjects.

These trade-offs were decided in favour of the most data-protected option during the design phase of the platform. In concrete terms, this meant that in the case of the first trade-off (1) it was decided that a locally stored private key is better from a data protection point of view than creating an account. It was also decided (2) that between the process steps, such as uploading the files by the controller, the central storage of the data subject's email address would be worse than transferring it via the corresponding link. The identification service (3) choice was the most challenging decision from a data protection point of view since attention had to be paid to the additional possibility of profiling by third parties. For this reason, the prototype was based on a connection to a centralised identity provider. A future extension in the direction of DID would be possible, should this prove to be sufficiently protected against profiling.

Compared to both the traditional process mentioned in Sect. 3 and proprietary, purely centralised systems, this method is able to transport data from the data processor to the data subject with authentication, confidentiality and integrity and therefore more data protected. Because of its centralised nature it does not have reliability features as blockchain-based decentralised communication applications [7,16].

The General Data Protection Regulation has been in force for several years now. Nevertheless, the exercise of rights is burdensome for individual citizens. The rights support platform presented in this paper is designed to help citizens access their rights and redress the data imbalance. It allows citizens to securely retrieve their data stored with a data controller whilst preserving data integrity and confidentiality. The prototype of this development showed problems in handling these processes and possible solutions with blockchain technology. Further developments regarding other legal issues that can be supported with blockchain technology could be elaborated in future research.

This supporting rights platform is intended to be a further step in the direction of restoring the balance between responsible companies and affected parties.

References

1. Council of European Union: Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation). Official Journal of the European Union (2016). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

2. Crossler, R.E.: Privacy in the digital age: a review of information privacy research in information systems. Manage. Inf. Syst. Q. 1017–1041 (2011)
3. Falk, S., Römmele, A., Silverman, M. (eds.): Digital Government. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-38795-6
4. Fang, B.: Cyberspace Sovereignty: Reflections on Building a Community of Common Future in Cyberspace. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-0320-3
5. Ishmaev, G.: Sovereignty, privacy, and ethics in blockchain-based identity management systems. Ethics Inf. Technol. (2020). https://doi.org/10.1007/s10676-020-09563-x
6. Janßen, C.: Towards a system for data transparency to support data subjects. In: Abramowicz, W., Corchuelo, R. (eds.) Business Information Systems Workshops. BIS 2019. LNBIP, vol. 373. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-36691-9_51
7. Khacef, K., Pujolle, G.: Secure peer-to-peer communication based on blockchain. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds.) Web, Artificial Intelligence and Network Applications. WAINA 2019. Advances in Intelligent Systems and Computing, vol. 927, pp. 662–672. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15035-8_64
8. Klitou, D.: Privacy-Invading Technologies and Privacy by Design. T.M.C. Asser Press, The Hague (2014)
9. Kościelny, C., Kurkowski, M., Srebrny, M.: PGP systems and TrueCrypt. In: Modern Cryptography Primer, pp. 147–173. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41386-5_6
10. Kushwaha, N., Roguski, P., Watson, B.W.: Up in the air: ensuring government data sovereignty in the cloud. In: 2020 12th International Conference on Cyber Conflict (CyCon), pp. 43–61 (2020)
11. Morgner, F., Bastian, P., Fischlin, M.: Securing transactions with the eIDAS protocols. In: Foresti, S., Lopez, J. (eds.) Information Security Theory and Practice. WISTP 2016. LNCS, vol. 9895, pp. 3–18. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45931-8_1
12. Peck, M.E.: Blockchain world - Do you need a blockchain? This chart will tell you if the technology can solve your problem. IEEE Spectrum 38–60 (2017)
13. Pinter, K., Schmelz, D., Grechenig, T.: Koordination der Informationspichten laut DSGVO mithilfe der Blockchain. In: IRIS 2020 Internationales Rechtsinformatik Symposion (2020)
14. Pinter, K., Schmelz, D., Lamber, R., Strobl, S., Grechenig, T.: Towards a multi-party, blockchain-based identity verification solution to implement clear name laws for online media platforms. In: Business Process Management: Blockchain and Central and Eastern Europe Forum, pp. 151–165. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30429-4_11
15. Rantos, K., Drosatos, G., Demertzis, K., Ilioudis, C., Papanikolaou, A.: Blockchain-based consents management for personal data processing in the IoT ecosystem. ICETE **2**, 738–743 (2018)
16. Sarıtekin, R.A., Karabacak, E., Durgay, Z., Karaarslan, E.: Blockchain based secure communication application proposal: cryptouch. In: 2018 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1–4. IEEE (2018)
17. Schmelz, D., Fischer, G., Niemeier, P., Zhu, L., Grechenig, T.: Towards using public blockchain in information-centric networks: challenges imposed by the European Union's general data protection regulation. In: 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), pp. 223–228 (2018)

18. Schmelz, D., Pinter, K., Brottrager, J., Niemeier, P., Lamber, R., Grechenig, T.: Securing the rights of data subjects with blockchain technology. In: 2020 3rd International Conference on Information and Computer Technologies (ICICT), pp. 284–288 (2020)
19. Singi, K., Choudhury, S.G., Kaulgud, V., Bose, R.J.C., Podder, S., Burden, A.P.: Data sovereignty governance framework. In: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, pp. 303–306. Association for Computing Machinery (2020)
20. Sullivan, C., Burger, E.: Blockchain, digital identity, e-government. In: Treiblmaier, H., Beck, R. (eds.) Business Transformation Through Blockchain, pp. 233–258. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-99058-3_9
21. Todd, P.: OpenTimestamps: scalable, trustless, distributed timestamping with bitcoin (2016). https://petertodd.org/2016/opentimestamps-announcement
22. Wirth, C., Kolain, M.: Privacy by blockchain design: a blockchain-enabled GDPR-compliant approach for handling personal data. In: Proceedings of 1st ERCIM Blockchain Workshop 2018 (2018)
23. Yakubov, A., Shbair, W., State, R.: BlockPGP: a blockchain-based framework for PGP key servers. In: 2018 Sixth International Symposium on Computing and Networking Workshops (CANDARW), pp. 316–322 (2018)
24. Yin, S., Kaynak, O.: Big data for modern industry: challenges and trends. In: Proceedings of the IEEE, pp. 143–146 (2015)
25. Yuming, L.: Data sovereignty theory. In: Sovereignty Blockchain 1.0, pp. 37–77. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0757-8_2