Distributed Data Protection And Liability On Blockchains

Giannopoulou, A.; Ferrari, V.

DOI
10.1007/978-3-030-17705-8_17

Publication date
2019

Document Version
Author accepted manuscript

Published in
Internet Science

Accepted manuscript
INSCI 2018 Proceedings- Lecture Notes in Computer Science (LNCS), Springer

Distributed Data Protection And Liability On Blockchains

Alexandra Giannopoulou[1] and Valeria Ferrari[2]

[1] Blockchain and Society Policy Lab, IViR, University of Amsterdam, 1018WV The Netherlands
[2] Blockchain and Society Policy Lab, IViR, University of Amsterdam, 1018WV The Netherlands
a.giannopoulou@uva.nl ; v.ferrari@uva.nl

Abstract

Blockchains and the GDPR pursue similar objectives where they seek to grant users greater control over their personal data. While the latter pursues this goal by imposing duties of care to centralised controllers and collectors of data, blockchains go a step beyond by trying to eliminate these stakeholders and the need to trust them. Nevertheless, the rules set out by the GDPR apply whenever personal data are at stake, and various actors of the blockchain ecosystem risk liability for controlling of processing data in violation of privacy requirements. A possible solution is to re-contextualise the concepts of data controlling and responsibility, as framed by the GDPR, in light of blockchains' enhanced individual autonomy. In this paper, we set the framework for a further inquiry on the role of users as both data subjects and data controllers of distributed ledgers.

Keywords Blockchain, Decentralization, Data Protection.

1 Introduction

The development of decentralized technologies at scale is the holy grail for the reorganization of social structures. The variant degrees of decentralization as well as the different structures that are created around it, aim to reinforce individuals and achieve collective social empowerment [10-11]. Blockchain technology represents the latest technological solution to decentralise the problem of trust in key societal and economic interactions. The growing interest in blockchains reflects a renovated call for reorganisation of power and for the elimination of unnecessary and untrustworthy intermediaries.

Blockchain is a combination of pre-existing technologies [6-7] which results in a digital medium with disruptive potential. It is an append-only distributed database that connects a decentralized network of nodes using a range of cryptographic methods. The participants of the network coordinate with each other based on algorithmic consensus rules encoded in the blockchain protocol and they continuously update the database. The applications of the technology promise to eliminate the need of trust, of trusted intermediaries and of trusted institutions from a number of human activities.[1] The protocol ensures full transparency; the mathematical verifiability of the

[1] According to the creator of Bitcoin, Satoshi Nakamoto, 'what is needed is an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party.'

transactions managed through the digital infrastructure offers an alternative to centrally organized institutions and intermediaries [3]. Blockchain technology provides thus a mechanism to guarantee security in the recorded transactions among parties that do not know or trust each other.

The distributed technological architecture of the blockchain requires also the articulation of governance principles because the continuously growing involvement of new actors creates a need for more organised decision-making processes. The extent to which these new actors control and process data is partially defined by their activity in the governance of the project, but mainly it is defined by the architecture of the blockchain technical infrastructure at the various layers of its stack. The combination of these qualities will also determine the legal qualification of their role as well as their legal liability.

In light of these considerations, it becomes apparent that technological and governance choices of decentralized systems have an impact on legal compliance [5]. If blockchain technology presents itself as an alternative to centralized models of information and value management, it is nonetheless subject to traditional law enforcement within the respective legal frameworks. In that regard, territoriality, privacy, data protection and liability pose a legal challenge when applied in blockchain technological applications.

The recent reforms of data protection rules in the European Union dictate that the technological design of web services and applications must take into account data controlling and processing rights and obligations dictated by the legal framework . The General Data Protection Regulation[2] - which came into effect on May 25, 2018 - requires the restructuring of most of the systems and processes that handle data collection and processing services in order to implement user rights and actor obligations. Decentralization is a priori not incompatible with data protection rules. However, the centralized model of data processing and control that the GDPR implies - which presupposes traditional single providers of computing processes - makes these rules hard to satisfy in the context of decentralized blockchains. In principle, both blockchain technology and the GDPR aspire to increase transparency, user agency, and trust. The GDPR, on the one hand, does so by identifying central actors who have increased control of the data processing and by assigning responsibilities to specific parties. On the other hand, the distributed and decentralized architecture of the blockchain ensures trust and transparency not by relying on few central actors but by incentivizing the use of the processing power of its distributed user base. Thus, if the goals of blockchains and the GDPR are similar, their approaches diverge on a fundamental level.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Hereinafter GDPR.

Nevertheless, as a distributed database running worldwide, blockchains are required to comply with data protection laws. Compliance with the GDPR will be necessary for all blockchain applications whose activities include "*processing of personal data in the context of the activities of an establishment of a controller or processor in the European Union, regardless of whether the processing takes place in the Union or not*". What is worth investigating is which data will fall under the scope of application of the GDPR obligations and whether legal compliance can be imposed on or ensured by actors that emerge in various blockchain applications. If the GDPR introduces the concept of privacy by design, it is interesting to examine how blockchain's technological design can be crafted as to accommodate data protection rules.

2 Defining Personal Data On The Blockchain

Data storage and processing is at the core of any blockchain. The data protection principles of the GDPR will only be applied to data that are qualified as personal[3] according to the definitions introduced in the Regulation. Complementary, the analysis issued by the Article 29 Working Party on the concept of personal data [2] serve as interpretation guidelines for the Courts. The obligation to ensure compliance will firstly depend on the qualification of the data stored and processed in the blockchain. While storage and processing of personal data fall into the material scope of data protection regulations, data that are not qualified as personal are not subject to the same regime. If the qualification of personal data applies, the data protection rules established by the GPPR generally require the consent of the data subject for the collection, storing and processing of that data. Moreover, such qualification implies mandatory duties of care for personal data protection and creates accountability for the actors involved in the personal data processing.

At the current state of its technological development - and according to the CJEU's case law interpreting the concept of personal data -, blockchains operating as databases that process data worldwide are likely to fall under the scope of application of the GDPR. The determination of the data stored on the blockchain as personal data or not is, however, one of the core issues that need to be addressed in order to identify the regulatory requirements the technology must be compliant with. The fluid, ever-expanding, and contextual approach to qualifying data as personal, coupled with rapid technological progress in data aggregation and analytics, is progressively transforming the GDPR into "the law of everything". For example, it is not uncommon for data to be "anonymous at the time of collection, but turn into personal later, just sitting there, simply by virtue of technological progress." [8]

Data stored and processed on the blockchain could potentially qualify as personal data if they refer to identifiable natural persons. For example, there are the data which

[3] According to article 4(1) GDPR, personal data are defined as 'any information relating to an identified or identifiable natural person'.

identify or which are associated with transactions occurring between the users. These can include fragments of plain data, but most of the data in question are stored in encrypted and hashed form on the blockchain. Because of the type of the information that they usually convey, these data are called transactional data. Also, the public keys of the users that participated in the transactions, which serve to refer to the sums of funds each account owns and can spend over the network could be qualified as personal data. The qualification of these data as anonymous[4] would make them fall outside of the scope of application of the GDPR.

Besides the plain data - which are only very rarely stored on the blockchain for efficiency purposes -, transactional data in encrypted and hashed form, as well as public keys, are considered pseudonymous data in the context of data protection regulation. Namely, encrypted data are vulnerable to decryption techniques that revert the data to its original state and thus revealing information related to an identified or identifiable person. Similarly, hashes can be linked to the data they have been derived from and can lead to the identification of the data subject. Hence, -while widely used to ensure security in transactions-, these techniques do not guarantee anonymity but merely pseudonymity.

According to article 4(5) of the GDPR, pseudonymisation is defined as "*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable person*". Techniques of pseudonymisation do not prevent personal data from being attributed to the data subject [9] and from falling under the scope of the GDPR. Therefore, in absence of other legitimate purposes for processing as provided by the GDPR, the storage and processing of these kinds of data requires the consent of the data subjects, in compliance with obligations and rights that the regulation ascribes to users and actors. This consent is hardly proved or expressed in blockchain transactions.

Both hashing and encryption techniques have already been identified by the Article 29 Working Party [1] as pseudonymization (hence, not as anonymization) techniques, since the re-identification of the data subject may be considered difficult but it is not irreversibly prevented. However, even if the re-identification of pseudonymous data on the blockchain can be successful, the effort required to achieve it remains dependent on the methods used for encrypting or hashing the data and on the prerogatives of data storage on the blockchain. Thus, the difficulty in re-identification depends on various variables that can be more or less related to the technology used.

[4] According to the Article 29 Working Party's Opinion on Anonymisation Techniques, data are considered anonymous only when their processing irreversibly prevents identification.

In general, pseudonymous data qualify as such every time re-identification can be achieved within a reasonable amount of time and effort.

Even if qualified *a priori* pseudonymous, public keys could, in certain circumstances that relate to the technical and architectural choices on the blockchain, not lead to the identification of a natural person applying re-identification technical efforts within the constraints of reasonable means as described in the GDPR. In general, practice has shown that public keys can be linked to a natural person in a variety of ways, but there can be instances where such identification becomes less easy.[5] At the same time, data that is at a given point considered anonymous are susceptible to lead to identification because of the availability of technological advancements at an environment where data are destined to be stored for an undetermined amount of time. In fact, the qualification of anonymous data is tied to the absence of re-identification processes. However, the qualification of unequivocal anonymity becomes more difficult for data stored on blockchains, because - given the lack of time restrictions that storage on blockchains implies - the technological evaluation of anonymity becomes more fluid. Thus, anonymous data approach the qualification of pseudonymous every time the probability of de-anonymization is increased in light of new technological breakthroughs. The legal significance of such shift in qualification from anonymous to pseudonymous data is that the data in question falls again into the scope of application of the GDPR.

Stretching the meaning of adequacy and necessity in the GDPR's principle of minimisation of personal data processing has been put forward as a possible workaround for the pseudonymous data that are stored on and off-chain [4]. According to Article 25(1) GDPR, the implementation of "*appropriate technical and organizational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimization, in an effective manner*" is necessary. The prerogatives of the compliance with such requirements for blockchain systems' architects remain uncertain, because the application of the GDPR to the technology is unclear. Respecting the obligation of data minimization through the architectural design choices of the blockchains could quickly demonstrate the limits of the proposed solution. For example, it is unclear which of the various applications of encryption methods would better respond to the data minimization obligation since the technological progress on this domain is in constant flux.

As the GDPR hints, the technological design can be set-up as to achieve legal compliance and provide to users the tools they need to invoke their data protection rights; thus, users could exercise their rights by being involved in the technological

[5] For instance, some end-users applications allow the generation of a new public key for each transaction, so that it becomes harder to link a set of transactions to an identifiable user. Moreover, cryptocurrencies like Monero deploy sophisticated techniques such as ring signature and Ring Confidential Transactions that prevent to link transactions and funds to public keys.

design choice process by selecting technology that are more compliant with privacy rules. The limits of user empowerment are challenged when determining if 'appropriate technical and organizational measures' were adopted by the blockchain architecture, such as for example pseudonymization using designated secure technologies or personal data storing technological choices.

The interplay between qualifications of personal and non personal data and the consequent legal requirements that apply to blockchain-based projects will, therefore, be a determining factor in the technological decisions during the life of the project in question. Finally, the immutability of blockchains poses a particular challenge in front of the user control over their personal data introduced by the GDPR. The use of technological solutions when facing legal challenges could end up reconciling these two characteristics representing transparency on the one hand and privacy on the other. According to the state of the accepted technological standards at the time regarding encryption or data storage and processing, the redeployment of a blockchain by the application of forking of the integrality of the chain could be perceived as a tool that ensures more appropriate data protection systems. However, the unforeseeable consequences of these technological choices could end up harming the core structure and ideology that justified the creation of blockchain technology.

3 Architecture-Based Liability Of Actors

Architectural choices in the design of blockchain technology are determined by the objectives of ensuring transparency, enhancing security as well as privacy and, finally, empowering users while ensuring scalability. More specifically, the interplay between decentralization and privacy aims at enhancing user control. The choices that define the interactions between the technology and the actors involved proceed to demonstrate the prioritization between the diverse goals and interests that finally produce what is an optimal balance between them. The final design reflects these choices, as the degree of decentralization of each system will define the degree of control attributed to different actors. The compliance of blockchains with the rules of the GDPR looks, as of today, improbable, as some fundamental technical features of decentralised technologies are in direct conflict with the latter. However, the malleability of the technical design, the variety of possible governance schemes and the interests vested in the development of the technology open possibilities to the construction of GDPR-compliant blockchains. [4]. Which choices such a construction would require is yet to be understood.

In building decentralised ledger technologies, the conflicting objectives of transparency and data protection impose a balancing process that depends on, and ultimately influence, the roles and responsibilities of various actors involved in the creation and maintenance of the network.. The legal obligations and responsibilities enshrined in the GDPR, apply - in presence of data qualified as personal - to those

actors that perform controlling and processing functions over personal data, as defined by the legal instrument itself. The liability of actors for storing or processing data through a blockchain network depends, firstly, on the qualification of the data stored on blockchains as personal data. Then, the question that needs to be tackled is: which actors are, in the context of blockchains, susceptible to be qualified as data controllers and data processors as referred to by the European Regulation?

In centralised data storage infrastructures, a single legal entity is generally responsible for a given server or cloud. In blockchains, instead, the data storage and processing is sparse among a network of uneasily identifiable computers, and the design of the technology tells us very little about the actual use of data from the actors involved.

According to the GDPR, *"'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data*". In a blockchain scenario, this definition could, at first sight, be best attributed to developers, as they "set[s]-up the code design and (the facto) govern[s] the distributed ledgers" [12]. Developers have, in public blockchains, the technical capability to determine how data are validated, stored and processed by the nodes of the networks. However, they are - at least in theory - bound to the network consensus considering that, for them to successfully upgrade or modify the protocol, the network of validating nodes must demonstrate their agreement with the stated choice by upgrading their functioning accordingly. Therefore, as long as developers are merely executing what the majority of the consensus group agrees on, they cannot be considered actual stakeholders in the blockchain's governance system nor can they be considered a source of independent determination over the processing of data [4].

The GDPR gives individuals control over their personal data; but it also assumes that clearly identified (or identifiable) actors have control over storing and processing of such data and are therefore accountable for such control. Potential liability under the GDPR could, therefore, apply to those parties that emerge as centralised sources of power within the blockchain ecosystem, as they mediate the interactions between users, and between users and the digital ledger. These comprehend a varied range of entities such as platforms (e.g. Ethereum, Filecoin, Dash etc.), service providers (cryptocurrencies exchanges; wallet providers), and companies that build all sorts of applications on top the blockchain protocol.

Consider, for example, a company crowdfunding itself by issuing a token on the Ethereum platform. Upon the receival of the funds, the company will collect and eventually analyse all the public keys - and associated data - of the users (eventually pseudonymised) which participated to the token sale; these pseudonymous data, however, will also be publicly accessible on the Ethereum blockchain, sacrificing the privacy of the company's clients in favor of the necessary transparency of the ledger.

Can, therefore, the company be held liable with regards to the users' data protection rights? Or should, instead, the Ethereum platform - which builds the technical infrastructure and defines the modalities of data processing - be considered the data controller?

This question certainly does not have a clear-cut answer. The use of the data and the ability to determine the means of data processing vary significantly based on the governance and technological design of each blockchain protocol and platforms built on top of it. The roles of actors of the blockchain space do not directly correspond to the definitions as provided in the GDPR. Moreover, the hierarchy of the blockchain stack adds complexity to the identification of roles and responsibilities with regard to the processing of data.

The technical design of blockchains imposes that the entire networks share and validate the ledger of information. Hence, no single data processor -which, according to the GDPR definition, is "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller"*- can be identified. Rather, such qualification could apply to all the nodes running the blockchain in order to validate transactions such as for example miners or qualified nodes.[6] However, in these examples the GDPR fails again to apply in a sound and clearcut manner. Each individual node, or miner, is in fact unable to determine the validation process, as its functioning is dependent upon the rules embedded in the protocol and on the cooperation of the consensus group.[7] Moreover, only certain kinds of nodes (i.e. "full nodes", "super nodes") download the entire blockchain and contribute to validate and support the shared ledger, and the level of involvement of nodes in the processing of data varies depending on the consensus algorithm and on the level of openness of the blockchain setup.

As long as the data stored on the ledger is considered personal data, blockchain's technical design is highly problematic for GDPR compliance. Distributed storage of information, transparency and immutability are, in fact, fundamental conditions of blockchain technologies, as they are functional to the purpose of creating a trusted, tamperproof repository of information for a network of users that do not necessarily trust each other. Such design, therefore, can be seen as something more than the product of interests or choices of few controlling parties. It represents, ideally, the

[6] Note that the present paper refers generally to so called "public", "permissionless" blockchains, in which any user can access the data, enter the network as validator or record transactions with no attribute-based or geographical restriction. As recognised by the CNIL opinion in the issue, "private" blockchains to do not pose specific problems concerning the attribution of liability for GDPR compliance. In fact, as private blockchains are developed and maintained by one or more identified actors, they perform as traditional databases whose storage is distributed but centrally controlled.

[7] Miners are, for instance, unable to individually influence changes in the protocol. They cannot alter or modify the data. They don't get to choose which data are stored on the blockchain nor the criteria based on which data get stored.

solution to a collective problem: that of cooperation in geographically sparse peer-to-peer networks of anonymous users.

Users adhering to a blockchain network do not commit to a unilateral transfer of their own data to a controlling party. Rather, they store information in a system in which they are both subjects and controllers, given that all users - as well as companies, platforms and other potential key-players - share the same information and are subject to common, consensus-based, processing rules.

Each individual is, in permissionless blockchains, entitled to become not only a user but also an active participant in the storage and processing mechanism set out by the blockchain protocol. This is where the GDPR mismatches the blockchain technological setting: a clear distinction between data controllers and data subjects, as spelled by the legal instruments, cannot be identified in DLTs. Further inquiry on the re-contextualisation of the concepts of data controlling and responsibility, as framed by the GDPR, in light of blockchain's enhanced individual autonomy is awaited, and it could solve compliance issues in decentralized technologies.

4. Conclusions

The greatest involvement of users in data processing that occurs in blockchain networks reflects, somehow, the GDPR attempt to grant users a higher degree of control over their own data. However, as long as the processed data qualifies as personal data as defined by the GDPR, the dictatum of the European legal instrument does not stop to exert its effects because of the ideological purposes underlying distributed technologies. The present work has addressed possible interpretative ways of fitting GDPR notions – namely, the definition of personal data, of data controllers and data processors – to distributed ledgers technological contexts. In this regard, the paper points out that – while they can certainly be reconciled at a conceptual level – data protection requirements and the organisation of information through blockchains present several points of conflicts. The highlighted problems are not exhaustive, and further issues, such as the enforceability of the right to erasure and the right to rectification in blockchain technical scenarios, deserve focus as well. This work, therefore, calls for broader researches into viable solutions to make GDPR and blockchains a sound match.

References

1. Article 29 Working Party, Opinion 04/2014 on Anonymisation Techniques, 0829/14/EN

2. Article 29 Working Party, Opinion 4/2007 on the Concept of Personal Data, 01248/07/EN WP136

3. Dingle S (2018), In Math We Trust: Bitcoin, Cryptocurrency and the Journey To Being Your Own Bank. Tracey McDonald Publishers

4. Finck M (2018), Blockchains and Data Protection in the European Union, EDPL, 4(1):17-35, https://doi.org/10.21552/edpl/2018/1/6

5. Ibanez LD, O'Hara K, Simperl E (2018) On Blockchains and the General Data Protection Regulation, https://eprints.soton.ac.uk/id/eprint/422879

6. Nakamoto S (2009), Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

7. Narayanan A, Clark J (2017), Bitcoin's academic pedigree, Communications of the ACM, 60(12):36–45. doi:10.1145/3132259

8. Purtova N (2018), The law of everything. Broad concept of personal data and future of EU data protection law, Law, Innovation and Technology, 10(1):40-81, doi 10.1080/17579961.2018.1452176

9. Schmelz D, Fischer G, Niemeier P, Zhu L, Grechenig T (2018), Towards Using Public Blockchain in Information-Centric Networks: Challenges Imposed by the European Union's General Data Protection Regulation, In: Proceedings of 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN 2018), pp. 223-228

10. Wright A, De Filippi P (2015), Decentralized Blockchain Technology and the Rise of Lex Cryptographia. https://doi.org/10.2139/ssrn.2580664

11. Wright, A, De Filippi P (2018), Blockchain and the Law: The Rule of Code, Cambridge, MA: Harvard University Press

12. Zetzsche DA Buckley RP, Arner DW (2017), The Distributed Liability of Distributed Ledgers: Legal Risks of Blockchain, University of New South Wales Law Research Series. Law Working Paper Series, Number 2017-007, http://dx.doi.org/10.2139/ssrn.3018214