BLOCKCHAIN
BUNDESVERBAND

BLOCKCHAIN
BUNDESVERBAND E.V.
VR 36105 B
27 / 661 / 52176

ADDRESS
c/o Jolocom
Waldemarstrasse 37a
10999 Berlin, Germany

info@bundesblock.de
www.bundesblock.de

Commerzbank Berlin
IBAN DE29 1004 0000
0228 8900 00
BIC COBADEFFXXX

Blockchain, data protection, and the GDPR

v1.0
25.05.2018

Contributors: Natalie Eichler, Silvan Jongerius, Greg McMullen, Oliver Naegele, Liz Steininger, Kai Wagner

Introduction

GDPR was created before Blockchain and is already outdated, since it doesn't account for decentralized technologies. We make recommendations for interpretation of the current law and highlight areas that can be improved in the future.

This document is the work of the Privacy and Data Protection subsection of the German Blockchain Association (Bundesblock). In this paper we make recommendations for the treatment of blockchain technology under the GDPR, aimed both at law makers and blockchain companies.

At the time the GDPR was conceived, we lived in a world of centralized cloud services and data collection business models that continue to persist as the main source of Internet-based revenue for companies. Since then, decentralized technology has developed rapidly, and may require adjustments to the GDPR framework. Applying the GDPR to decentralized technology like blockchains is complicated, as they complicate the distinction between server and user. Therefore, the enforcement of the GDPR presents a number of challenges that could threaten the adoption of decentralized technologies and Germany's place as a leader in blockchain development.

Germany has emerged as a home to world leading blockchain companies and fostered an innovative and supportive ecosystem. However, some blockchain companies have limited their activities and expressed concerns regarding the application of the GDPR to their businesses. A leading blockchain company based in Berlin decided to abandon efforts to create a commercial product that would provide "know your customer" (KYC) and anti-money laundering (AML) services to other blockchain companies, in accordance with German KYC-AML regulations, because of their fear of the impact of the GDPR. This decision was in part because of concerns that those services might not comply with the GDPR, despite their belief that their products would improve users' privacy. Another Berlin-based company withdrew support for a not-for-profit entity that was building a public blockchain database, citing concerns about GDPR compliance and the resulting potential exposure to liability.

In addition to the negative impact on existing German blockchain companies, the GDPR may increase the risk of liability for node or mining operators, in such a way that the balance of power on public blockchain networks shifts further away from Germany and the European Union and toward other countries, like China or Russia. This has more than an economic impact—enough control over the validating nodes could allow these countries to tamper with the integrity of the blockchains themselves, a worrying prospect as more critical applications, including governments and the financial sector, rely on these blockchains.

The intersection of blockchain and the GDPR

The GDPR sets out a broad range of obligations to anyone processing data, depending on the level of responsibility of the entity, and provides for differentiated rights for individuals whose personal data are processed. In this respect, the central questions are (1) if personal data are processed, (2) which stakeholders are deemed to be responsible for data protection, and (3) how rights of individuals may be guaranteed. In the application of blockchain technology, the answer to each question represents a major challenge and is far from being clear.

What constitutes personal data?

The provisions of the GDPR apply to personal data only. Thus, it is essential to know if data stored on and processed by means of blockchain technology are personal data.
The GDPR defines personal data as follows:

> *'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;* (Art. 4 no. 1 GDPR)

Recital 26 of the GDPR gives some guidance on how to interpret the concept of personal data:

> *"[...] To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. [4]To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. [...]"*

Essential to the concept of personal data is the linkability of information to an individual allowing his identification. Any information that allows for identification of natural person by reasonable means may constitute personal data.

Truly anonymous data do not constitute personal data, as stated in recital 26:

> *"[...]The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."*

It is important to note that pseudonymized data are personal data:

> *"[...] Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person."*

In blockchain environments, data processed on blockchains can be broken down to the following kinds of data to be considered in relation to the definition of personal data:

Public keys: Blockchains rely on public key cryptography, and data stored on blockchains are typically associated with a publicly visible public key. Whether (or when) those keys constitute personal data is a core question for blockchain technology and requires clarification. As soon as public keys can be associated to a natural person, it will constitute personal data under the GDPR. Since public keys are central to the operation of blockchain systems, it is important to identify cases in which public keys should **not** be considered personal data. We expect public keys will **not** be personal data in the following circumstances: When (1) the key does not belong to a natural person or is not created on behalf of a natural person; or (2) the key cannot be linked to a data subject by reasonable means and is therefore truly anonymous.

Other data stored on blockchains: Blockchains can store more than just financial transaction data. For example, Bitcoin transactions contain a notes field that allow any data, including personal data to be written along with a transaction, which could itself include personal data, depending on the sender or use of that transaction. Other blockchains have similar features.

Hashed data: Hashing functions are algorithms which accept any data of any size as input and generate a fixed length string as an output value. Running the hashing function again on the same input data will always generate the same output hash value. But if even a single bit of the input data is changed, the output hash value will be different as well. A hash value is typically smaller than the input data. There are three reasons to write hashed data to a blockchain: (1) to later validate data by comparing it to the hash, (2) to obscure plain text data, or (3) to overcome the limitations on the size of the data that can be written to a single transaction, by writing a hash of a larger block of data rather than the entire block of data. Hashed personal data has already been determined to be pseudonymous, not anonymous, by the Article 29 Working Group. However, if the data linking the hashed data to a data subject is kept off-chain and is later erased, the hashed data should once again be considered anonymous. We ask for clarification on this point, particularly on what steps would be required to make hashed personal data anonymous.

Encrypted data: There is a common misconception that encrypted personal data is not personal data and can be safely written to a public blockchain. The Article 29 Working Group has found encrypted data is pseudonymous, not anonymous. While it is reasonably unlikely for state-of-the-art encrypted personal data to be linked back to a natural person at the moment, in the future that may change and therefore pose a risk of future breaches. The Bundesblock takes the position that encrypted personal data should not be written to a public blockchain in the first place.

Zero Knowledge Proofs: Zero knowledge proofs allow someone who has data to show they have it without revealing the contents of the data. For example, Zcash offers an option to use a wallet that does not reveal any details of a transaction on the public blockchain (including sender, receiver, or amount), but only allows for a confirmation that the coins have been spent. Zero knowledge proofs could be applied in blockchains in other ways in the future.

Conclusion: It is clear that any data that is stored on a blockchain can constitute personal data. To minimize overlap and potential conflict with the GDPR, developers must limit the amount of personal data stored on blockchain, find new ways to anonymize data, and seek out GDPR-compliant off-chain data storage options.

Legal status of participants

The legal status of the different participants in blockchain networks and ecosystems such as nodes, miners, developers, users' wallets) and front-end or second layer services is not clear under the GDPR. The GDPR was designed in a world with a clear division of responsibilities between controllers and processors. For example, in the case of a smartphone app operated by a traditional company, users download an app to their smartphone and provide some personal data to the company. The company is a data controller and is responsible for their users' data. They can pass on these responsibilities to any third-party data processors with a data processing agreement.

The GDPR defines the following entities in its data protection model:

Controllers: *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; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law. (Art. 4 no. 7 GDPR)*

Processors: *A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. (Art. 4 no. 8 GDPR)*

Joint Controllers: *Two or more controllers who jointly determine the purposes and means of processing. (Art. 26 GDPR)*

In blockchain ecosystems, different stakeholders interacting with a blockchain by processing data may be singled out: (1) Nodes and miners, (2) Users interacting with blockchains by means of wallets or other front-end services and (3) application operators processing data to and from the blockchain when providing services by means of an application. Although they do not process data, developers of blockchain protocols also play a role in defining how the data is processed by way of the protocol.

Blockchains and other decentralized technologies do not fit cleanly in this model. The standard distribution of responsibilities maps differently to the parties involved, redistributing the previously centralized power over the data. In this section we propose ways in which the various actors in a blockchain ecosystem should be treated under the GDPR in the near term and point out areas where reforms may be necessary in the future.

We recommend the following handling of various entities in the blockchain ecosystem:

- For public permissionless blockchains (e.g. Bitcoin & Ethereum):
    - **Nodes** or **Miners** do not decide what data is written to the blockchain, beyond an analysis for general consensus according to the protocol as defined in the blockchain software. They do not determine the means and purposes of processing of personal data sent to the network by a third party. In our view, they cannot be data controllers, and they should be considered to be infrastructure, like routers or Internet backbones. Provided that they are considered data processors, this results in significant problems regarding necessary documentation under the GDPR. In a fully decentralized network with a large number of nodes, it would be very difficult for controllers to conclude data processing agreements with each of those nodes as the software allows them to participate without permission from a central party.
    - **Wallets** are software packages at the application level that allows companies or individual users to control their own private key and to interact with the blockchain network by sending transactions to the miners for processing. Wallets may pass personal data to miners, but only do so at the direction and control of the user, putting the user in control of processing of the personal data, or the application, which may or may not be operated by a central party which could be a data processor.
    - **Services and applications** that process users' personal data are likely data controllers, since they determine the purpose and means of data processing, especially if operated by a centralized party. This creates an obligation for them to enter into data processing agreements with any third-party data processors they use. But what if they want to store that data on a blockchain network? If every node is considered a processor, it would result in the obligation to conclude data processing agreements with each node participating in the blockchain—another reason to conclude that miners/nodes should be treated as infrastructure. This highlights a conflict between the GDPR and the nature of blockchain networks. It would not be reasonable or possible to contract with all miners/nodes, and there is no legal entity representing the blockchain network for a controller to contract with.
    - **Developers of blockchain protocols** or implementations should not be considered controllers or data processors. They do not collect or process data. They create tools, and it is up to the other participants in the system how the tools are used.
- For public permissioned blockchain (e.g. Sovrin):

  - A **governance body** like a foundation may be established to oversee the permissioned network, in particular by defining the rules for nodes and services to participate in the network. Under GDPR, it is thinkable that such governance body may play the role of a data processor, contracting with data controllers (e.g. application providers) who have collected personal data from individuals and serving a single point of legal contact with the network. Or the governance body could be a controller or a joint controller if it has influence over the purpose and means of processing.
  - **Nodes** may be most likely considered data processors, since they do not define the purpose of processing, but merely provide for the means of processing. Consequently, they are the entities to be contracted with to store personal data. This contract may be with services or with the governance body. One solution in a public permissioned chain would be to impose forks on the nodes to comply with GDPR requests. That would allow GDPR compliance to be enforced from above, while nodes with no control are not subject to liability.
  - **Services** can only be offered by those who are allowed to by the governance body. Services, while defining the purpose and means of data processing, are to be considered data controllers. Under certain circumstances, it is thinkable that the service can also be considered a joint controller together with the governance body. This depends on how the governance structure defines what data can be collected and for what purpose.

- Private permissioned blockchain (e.g. Ethereum Enterprise, Hyperledger, BigchainDB consortiums, Blockchain HELIX):
  - **Governance body** may be established to oversee the permissioned network. This governance body could play the role of a data processor, contracting with data controllers who have collected personal data from individuals and serving a single point of legal contact with the network. The governance body could also be a controller or a joint controller if it has influence over the purpose and means of processing.
  - **Nodes** or **validators** are data processors, contracted to store personal data. This contract may be with permissioned users or with the governance body.
  - **Permissioned users** while defining the purpose and means of data processing, are to be considered data controllers. Under certain circumstances the permissioned user can also be considered a joint controller together with the governance body.
- We recommend:
  - exploring the concept of “binding network rules” (inspired by the GDPR concept of binding corporate rules), allowing controllers to deal with an entire blockchain network as long as that network meets certain criteria.

- standard contractual clauses for blockchain ecosystems, allowing users to consent to the processing of personal data on a public permissioned blockchain networks.

Data Subject Rights

General principles: In general, data written to a public blockchain are permanent. When the data written is personal data, it may conflict with general principles of the the GDPR such as data minimization and the right to erasure (right to be forgotten). As pointed out above, anonymization should be kept in mind if blockchain technology is chosen to process personal data. We therefore recommend that any application writing data to a public blockchain keep Data Protection by Design principles in mind and consider the implications of the data being written to the public blockchain before it is written. This means personal data should never be written in plain text. In the view of the Bundesblock, storing hashed data should be acceptable in circumstances where it is ensured that the data cannot be reconstructed; storing encrypted data is too risky as the encryption may later be broken.

The treatment of public keys remains a challenge to this recommendation. Public keys are a central, unavoidable element to the operation of blockchain technology. The law must acknowledge a new, rights-compliant way to think about public keys.

Right to Erasure (Art. 17): For most categories of personal data processed on blockchains, it is safe to assume personal data cannot be erased or effectively blocked. However, any anonymization procedure should be considered to be an alternative way of erasing data. In particular, if hashed data remains available on a blockchain but the underlying unhashed data is deleted and not easily reconstructed, the right to erasure should be deemed to be met if the hashed data is not reasonably likely to be tied to a natural person.

As a result of the means of storage, erasing data written to a public blockchain is impossible, or involves “disproportionate effort” similar to that considered by section 35 BDSG neu in the context of non-automated processing. One possibility would be an exemption for automated processing that allows the blocking of data rather than erasure, similar to the exemption in that section. This obviously presents a risk of censorship, so the use of such blocking would need to be strictly supervised and applied as narrowly as possible.

Right to Restriction of Processing (Art. 18): Data written to a public blockchain in accordance with the general principles above should not be subject to restrictions on processing (unless as a alternative means to data erasure as set forth in section 35 BDSG). Restrictions on data processing may be required on the part of application developers who are controllers, but not on the part of nodes on a public blockchain network.

Right to Data Portability (Art. 20): Data written to a public blockchain is available to the general public and should be deemed to comply with data portability requirements. Services building on top of a public blockchain should consider the ability of users to download their data or move it to competing services.

Code of Conduct

The Bundesblock will initiate the creation of a code of conduct for blockchain technology in accordance with Art. 40 of the GDPR.