BLOCKCHAIN TECHNOLOGY – A THREAT OR A SOLUTION FOR DATA PROTECTION?

Aleksander Wiatrowski

University of Lapland, Faculty of Law, Institute for Law and Informatics
Yliopistonkatu 8, 96300 Rovaniemi, FI
aleksander.wiatrowski@ulapland.fi; http://www.ulapland.fi/EN/Units/Institute-for-Law-and-Informatics

Keywords: ***Blockchain, privacy, personal data, data protection, right to be forgotten***

Abstract: *In this paper, I raise the question of data protection and the application of the right to be forgotten within blockchain technology. Blockchain technology may stand against the GDPR when it comes to collecting and storing data. There is a risk it can prevent the application of the right to be forgotten. However, there are studies showing that blockchain can be a solution for data protection. Additionally, there is an issue of personal data. Nowadays this term is hard to define, it is almost elusive. With the rapid technology development, much more data can be recognized as personal data.*

1. Introduction

Conflict between law and technology is not something new. Even the best law could become helpless or even useless in confrontation with new technology. The right to be forgotten, an old concept, but a new addition to law, has already been met with some criticism, but now it may face another issue – blockchain technology. In some cases, blockchain technology prevents data from being erased by design. That includes personal data.

2. Blockchain

Blockchain technology – a cryptographically secured chain of blocks – was first described in 1991 by Stuart Haber and W. Scott Stornetta.[1] Today, blockchain technology is widely recognized as the technology behind the Bitcoin cryptocurrency.[2] Blockchain is like a large ledger within which every bit of data of every single entry is saved. When new data is added to a blockchain, peers in the network check the data to ensure that it is valid for addition, to avoid fraud by rogue nodes. The data that the peers need to check, needs to be stored transparently in the blockchain.

It is very important to underline that blockchain can be used for various purposes. For handling money, it is a highly anonymous public blockchain. In case of storing user data, it can be a true distributed database that allows for editing and deleting records. At least in some cases, but there are already some examples.[3] Blockchain is not always anonymous, Bitcoin is. It very much depends on what blockchain technology is used for.[4] If a blockchain is anonymous, then it is theoretically excluded from of General Data Protection Regulation.[5] All that being said, blockchains might not necessarily be bad for privacy.

---

1 Haber, Stuart/Stornetta, Scott, How to time-stamp a digital document, Journal of Cryptology. 3 (2), 1991, p. 99–111.

2 «Blockchain, The next big thing – Or is it?», The Economist Online, 2015, http://www.economist.com/news/special-report/21650295-or-it-next-big-thing (all websites last visited in January 2018); Pilkington, Marc, Blockchain Technology – Principles and Applications, p. 225, in: Olleros, Zhegu (ed.), Research Handbook on Digital Transformations, Cheltenman, 2016.

3 See below.

4 Lucas, Matt, The difference between Bitcoin and blockchain for business, IBM, 2017, https://www.ibm.com/blogs/blockchain/2017/05/the-difference-between-bitcoin-and-blockchain-for-business/.

5 Anonymous data that definitely does not allow to identify the data subjects are excluded from the scope of the GDPR.

3. Personal Data of the Future

The Data Protection Directive gives us a definition of personal data – «*‹Personal data› shall mean any information relating to an identified or identifiable natural person (‹data subject›); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity; ...*»[6]

The General Data Protection Regulation (GDPR) modifies this definition by making it a little more detailed – «*‹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; ...*»[7]

The GDPR will give individuals substantial control over their data by providing them with more information as to how their data is processed, which must be presented in a clear and understandable way. It gives the right to know when their data has been hacked, adds data protection safeguards and privacy-friendly default settings and strong enforcement of GDPR violations. Additionally, the GDPR contains a clear right to be forgotten provision in Article 17. This is designed to help people better manage data protection risks online and allow individuals to delete their data if there are no legitimate grounds for the information to remain public.

However, technology, as usual, complicates the situation. The new personal data definition includes online identifiers as potential revealing factors. It is difficult to say, at this point, which of those online identifiers can be found in relation to blockchain technology. To make it clearer, I will give an example: The TOR Project Internet Browser. This browser prevents somebody from watching your Internet connection and knowing what sites you visit, it prevents the sites you visit from accessing your physical location and it lets you access sites which are blocked.[8] To be able to use the full functionality of it there are certain, strict rules. Some of them are extremely unusual, but give the idea of what can be an online identifier. From the TOR browser guide, we can learn the following:[9]

We should not resize our browser windows – if we do, our browser instance has a potential to have a unique viewport size and hence, there is a probability we can be tracked. Whatever size window TOR Browser opens, do not resize it. Removing a menu bar or using full screen in TOR Browser is recommended against. The latter is known to modify the screen size, which is bad for the web fingerprint. The point of this advice is that if your browser is full screen, it will be the same size always and therefore it is potentially trackable between sessions. If a custom browser size is associated with a one-time-use anonymous session then it should not be a problem except that it leaks information about what window sizes are possible on our system, and our custom resize is very unlikely to be random, but rather probably proportional to our actual screen size.

6 Article 2 of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

7 Article 4 of 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 (General Data Protection Regulation).

8 https://www.torproject.org/projects/torbrowser.html.en.

9 https://www.torproject.org/docs/faq.html.en.

Knowing something as small and as simple as a window size, let us have a look at blockchains. Blockchains can hold vast amounts of data and depending on the application of the blockchain some of this data may be classed as personal data under General Data Protection Regulation. Depending on how the blockchain operates, the pseudonymized public addresses which are recorded in every transaction and hashed onto the blockchain may also constitute personal data. If personal data is being recorded onto the blockchain, then every node on the network will be a data processor as soon as it receives a new block of data for updating its own copy of the ledger. This poses difficulties ensuring compliance with the GDPR.

Additionally, every blockchain contains transaction data. That data needs to be designed so that it is not disclosive in and of itself, which may be a tricky balance as that data might also be necessary to assess whether the transaction is valid and therefore to prevent fraud or errors. Transactions should also be designed so that they cannot be used to add comments that might include personal data.

In some cases, personal data is required to verify validity of a transaction on the blockchain. For a node to check a Bitcoin transaction, it must have access to all previous transactions and be able to check that the person giving the Bitcoins has some at his or her disposal. It must be possible to reconstruct the full financial history of every person exchanging Bitcoins: How many Bitcoins they have, where they got those Bitcoins from, whom they spend their Bitcoins with – this is personal data. Pseudo anonymity[10] of the Bitcoin address can help, but it can easily be breached if the address is associated with a donate button. Therefore, it is advised to hold several Bitcoin addresses and not to transfer Bitcoins between those accounts to avoid others linking them together.[11]

It is important to remember that blockchains do not have to expose personal data directly to reveal private information about people. A blockchain recording used by health practitioners does not need to include the entirety of someone's health records to reveal information about them. Metadata may be sufficient to reveal personal details. Thus leads to the following question: Considering the main characteristic of the blockchain technology, that is immutability, how can we apply the right to be forgotten?

4. Right to be Forgotten within the Blockchain

The right to be forgotten can be described as the right of individuals to have their data no longer processed and deleted when they are no longer needed for legitimate purposes.[12] It has clear limits and rules given in art. 17 General Data Protection Regulation (right to erasure, «right to be forgotten»). Contrary to popular believe, it is not absolute because of clear legal limitations. There are some possible limitations from new technologies (i.e. Blockchain).

The key benefit of blockchains is the immutability[13] of data – all data being recorded and maintained in the chain from the beginning of the blockchain are an undisputable record in terms of verification purposes. Once data has been written to a blockchain, no one, not even a system administrator, can change it. If this data is made up of personal data, then erasure or rectification of the personal data would theoretically be impossible. This is likely to pose a greater issue as data subjects have a right to require data controllers to rectify and erase their data.[14]

---

10 Bitcoin is pseudonymous. Sending and receiving Bitcoins is like using a pseudonym. If an author's pseudonym is ever linked to their identity, everything they ever wrote using that pseudonym will now be linked to them. «Bitcoin Anonymity – Is Bitcoin Anonymous?», https://www.buybitcoinworldwide.com/anonymity/.

11 Tennison Jeni, What is the impact of blockchains on privacy?, open data institute, 2015, https://theodi.org/blog/impact-of-blockchains-on-privacy.

12 Communication from the Commission to the European Parliament, the Council, the economic and social Committee and the Committee of the Regions. A comprehensive approach on personal data protection in the European Union, COM(2010) 609 final, p. 8.

13 Immutability is relative and relates to how hard something is to change.

14 Russel, Luke, Blockchains: The legal landscape, Blake Morgan, 2016, https://www.blakemorgan.co.uk/training-knowledge/features-and-articles/blockchains-legal-landscape/.

Due to the unalterable character of the blockchain, it is impossible to erase data once it has been added. It seems that blockchains and the right to be forgotten aren't compatible. Inalterability and decentralization means that the register is made up of indelible data and that this register is shared with all users in the network. Applying the right to be forgotten goes against the very principle of inalterability, which lies at the core of the blockchain technology.[15]

On the other hand, blockchain technology could also be beneficial for the protection of personal data by encoding permissions, conditions and restrictions for its use. It could enable data portability and provide an easily auditable trail with evidence of consent.[16]

A blockchain stores a series of transactions, which can be data of any sort, in blocks, which get added to the blockchain one after the other. Blockchains are what is known as an append-only data store.[17] That means you can only add data to the store, you cannot take it away. Blockchains are maintained by a peer network of nodes in which every node has a copy of the blockchain and has equal authority to add to it. Every node publishes that data for other nodes to pick up and use. One of the unique selling points of blockchains is that once data is embedded in the blockchain it cannot be altered without that change being detected and rejected by the other nodes in the network. This is useful for data that people need to trust because it provides a guarantee that the data in the blockchain has not been changed since it was put there.

Greg McMullen[18] gave a rather pessimistic image regarding the possibility to apply the right to be forgotten within blockchain: «*Assuming personal information is encrypted before it is written to a blockchain, destroying the key renders the data unreadable. But is this enough to comply with the right to be forgotten, if the data is technically still there? Regulators should accept the destruction of a key as an erasure for the purposes of the GDPR, so long as the destruction is done in accordance with best practices and in an auditable way*.»[19]

However, there are possible ways to make changes in the blockchain, which potentially gives a chance to apply the right to be forgotten.

To clear out the data, over half the nodes would have to work together to rebuild the blockchain in the state that existed before that data was added. This process is like rebuilding from a backup: while being rebuilt, the blockchain would be rewound to a previous state, days or weeks or even more out of date. During this time, the data would not be up to date. This might also be a time when unwanted changes to data that was trustworthy could get in.

The second idea is that a court could try to compel the entire set of nodes to be shut down. Putting aside that nodes may reside in different legal jurisdictions, that would have huge practical implications. It would mean removing all the rest of the data held in the blockchain as well as the target of the order. Unfortunately, blockchains are usually holding many types of data and are supporting many types of applications. Because of that, there is a real risk that bad data simply must continue to exist to prevent massive disruption of the provision of good data for other applications. Therefore, even if this solution were possible, it might be too risky for the blockchain.[20]

---

15 «When the right to be forgotten becomes possible on the Ethereum blockchain», NewsBTC, https://www.newsbtc.com/press-releases/bcdiploma-right-to-be-forgotten-ethereum-blockchain/.

16 Lumb, Richard/Treat, David/Jelf, Owen, Editing the uneditable Blockchain. Why distributed ledger technology must adapt to an imperfect world, accenture, 2016, p. 6, https://www.accenture.com/t00010101T000000__w__/it-it/_acnmedia/PDF-33/Accenture-Editing-Uneditable-Blockchain.pdf.

17 Jeni (note 11).

18 Founder and Executive Director of IPDB Foundation, https://ipdb.io/.

19 McMullen, Greg/Glatz, Florian, Blockchain & Law in 2017: Finally friends or still foes?, 2017, https://medium.com/ipdb-blog/blockchain-and-law-in-2017-f535cb0e06c4.

20 Jeni (note 11).

Other solutions include controlling what becomes public within a peer-to-peer network of trusted nodes, therefore hiding data in the blockchain that should not be shared in the first place.[21]

But, there is another solution. The issue of the right to be forgotten is seemingly resolvable in a permissioned system which would allow the controlling party to use a blockchain editor tool, like the one Accenture[22] has recently filed a patent for.

This solution should offer new room to manoeuvre, not only in financial services but across industries. The invention modifies existing blockchain technology to allow designated authorities to edit, rewrite or remove previous blocks of information without breaking the chain. Its main features include:

- It is compatible with current blockchain designs,
- can be implemented now,
- requires only minimal changes to current application software.

This invention enables blockchain editing by using a new variation of the so-called chameleon hash function, which can recreate matching algorithms using secure private keys. After a change has been made to a block, the original blockchain remains fully intact and there is no need to rebuild subsequent blocks. That means flawed smart contracts could be updated at the time the contract was issued and the changes would apply to subsequent smart contracts in the chain. Even if edits to one block impact subsequent blocks, the fix would be far easier than a hard fork. The editable blockchain invention provides the means to build a virtual padlock for the link connecting two blocks.

Redacting the blockchain is simple: The chameleon hash key is used to unlock the link between the block that must be changed and its successor. Thanks to the key, it is possible to substitute the block with a new one without breaking the hash chain. The invention is designed to preserve the virtues of immutability as well. The editable blockchain invention is designed for permissioned systems, which have a designated administrator who manages the systems and grants permission to use it.[23]

It seems that this solution allows the application of the right to be forgotten. However, it is important to remember that all immutable unapproved systems are very likely to not be compliant. If Accenture's invention is truly effective, it might lead to the situation that it is regarded as a standard blockchain in terms of the GDPR regime.

Finally, there is a blockchain that follows the GDPR, including the right to be forgotten. It is Ethereum[24] and it allows to store diplomas and personal data. To accomplish compliance with the GDPR, the data is encrypted and secured using a set of three keys:

1. Graduate Key – This is the property of the graduate and is integrated into the diploma's URL.
2. Persistent Key – It is kept by the educational establishment. When the graduate wishes to exercise his or her right to be forgotten, he or she only has to destroy this key.
3. School Permanent Key – This is kept by the educational establishment.

---

21 Farmer, Steven, Blockchain technologies and the EU «right to be forgotten» – an insurmountable tension?, International Business Times, 2017, http://www.ibtimes.co.uk/blockchain-technologies-eu-right-be-forgotten-insurmountable-tension-1580166.

22 Lumb/Treat/Jelf (note 16).

23 Lumb/Treat/Jelf (note 16), p. 7.

24 https://www.ethereum.org/.

There is an algorithm allowing total security of the diploma's keys. It is not stored and can be generated only by assembling three keys through a derivation process.[25]

In conclusion, solutions for the described problems exist and are ready for deployment. It seems that «old» blockchains may pose some problems. However, every new database built on blockchain technology should be designed to comply with GDPR rules. Whether it will be using given examples or there will be some other ways, is a question for the coming future.

5. Blockchain Used for Data Protection

From a data protection perspective, blockchain technology is particularly interesting because it theoretically allows transactions between parties without them having to disclose their identity. Anonymity and pseudonymity are also addressed as data protection law instruments. If a transaction cannot be traced back to the individuals, their fundamental right to self-determination is not affected.[26]

Can blockchain technology be an opportunity for personal data protection? To a certain extent, yes. Blockchains are decentralized and distributed. Currently, various trusted third parties process personal data. These entities are centralized and, therefore, often constitute single points of failure. Leaks of unimaginable amounts of data due to cybercrime often occur in the form of an attack on a single entity, such as a hospital, email service provider, etc.[27]

Blockchains are public and transparent. We do not currently have any effective control over who processes our personal data and how. In fact, the data subject is in control of its personal data only to a restricted degree. Upon a transfer of that data, the subject loses control over how it is subsequently used.

Blockchains are very safe. Using cryptography (digital signatures, encryption, time-stamping) and systemically embedded economic incentives for network maintaining entities, blockchains provide a secure way of storing and managing information, including personal data.

When all nodes in a network need to be in sync, they all need to have the same version of reality which means they all need the same data. How does this fit in with the current regulatory environment around data privacy and with the new GDPR coming in? I would argue that the GDPR and blockchain advocates point to the same thing: The need to fundamentally change the way in which personal data is managed.

Most significantly, blockchain technology may enable individual control of one's personal data. According to Martin Ruubel, president of Amsterdam-based GuardTime, people will be in control of data, will be able to share it with whomever they want to, and will be paid for it.[28] In the future, the widespread adoption of blockchain technology can remove the need for large companies to maintain data and provide individuals with complete control over their personal data.[29]

clear answer to the question whether

---

25 The Right to Be Forgotten Becomes Possible on the Blockchain», Cryptotimes, 2017, http://cryptotimes.org/blockchain/right-forgotten-becomes-possible-blockchain/.

26 Wilke, Stefan/Krings, Dennis, Blockchain from a perspective of data protection law. A brief introduction to data protection ramifications, DeLoitte, https://www2.deloitte.com/dl/en/pages/legal/articles/blockchain-datenschutzrecht.html.

27 Czarnecki, Jacek, Blockchains and Personal Data Protection Regulations Explained, coindesk, 2017, https://www.coindesk.com/blockchains-personal-data-protection-regulations-explained/.

28 Shin, Laura, The Top 10 Blockchain Takeaways From Europe's Trustech Conference, Forbes, 2016, https://www.forbes.com/sites/laurashin/2016/12/05/the-top-10-blockchain-takeaways-from-europes-trustech-conference/#6bb7a0e97ba6.

29 Berms, Jeffrey K., Blockchains Can Assist EU Regulatory Fight for Personal Data Protection, Berns Weiss, 2016, https://www.law111.com/blockchains-can-assist-eu-regulatory-fight-for-personal-data-protection.

Blockchain technology can better address the privacy concerns to which the GDPR and EU regulators are responding. Zyskind et al.[30] for example, call into question the current centralized model of protecting personal data through trusted third parties and describe a more secure, unhackable and decentralized peer-to-peer personal data management system using a blockchain. The authors' proposed system focuses on mobile platforms and ensures that individuals own and control their personal data. Individuals decide with whom they share their personal data through delegated permissions.

Another example is Civic[31]. It is a digital platform that uses Bitcoin's public blockchain for identity management.

1. A user signs up to the Civic app which collects various identifying information for them.
2. All of that is passed through to either a government agency or a third-party identification verification service depending on the country.
3. Once verified, Civic takes a cryptographic hash of all the information, inserts the hash into the public blockchain, and then erases the personal data from their servers.
4. Then, when you want to authenticate to use another service, you share whatever information they ask of you and they can send the information through Civic's special algorithm to check it against the hash on the blockchain.

Once authenticated, the service using Civic no longer needs to store your information for identification or authentication purposes.

Finally, it is possible to encrypt data stored within the blockchain. The main problem with this approach is that if the decryption key for encrypted data is ever made public, the encrypted content is readable by anyone with that key; there is no way of encrypting the data with a different key once it is embedded within the blockchain. Also, if the key is ever lost, the data cannot be read. And there is the problem of sharing the key for the data amongst all those who legitimately need to be able to read it.

It must also be mentioned that the anonymity in the blockchain is far from being perfect. It is possible to associate public keys with each other, and with external identifying information. Appropriate tools allow to observe the activity of known users in detail. Additionally, an interested party can potentially deploy marked Bitcoins and collaborate with other users to discover even more information. Large centralized services such as the exchanges and wallet services can identify and track user activity.[32]

6. Conclusion

As said in the beginning of this paper, blockchains are not necessarily bad for privacy. It all depends on how they are designed. Anyone experimenting in the area should be thinking about implications. How the right to be forgotten plays out in the context of blockchain does, of course, remain to be seen. For example, it could be argued that there is a legitimate reason for retaining transaction blocks. Precisely how EU regulators and courts would look to police this right, considering jurisdictional hurdles, are just two key questions which come to mind. In the future, the widespread adoption of blockchain technology can remove the need for large companies to retain data and provide individuals with complete control over their personal data. Moreover, laws and regulations could be programmed into the blockchain itself, so that they are enforced automatically.

[30] Zyskind, Guy/Nathan, Oz/Pentland, Alex, Decentralizing Privacy: Using Blockchain to Protect Personal Data, IEEE Security and Privacy Workshops, 2015, http://ieee-security.org/TC/SPW2015/IWPE/5.pdf.

[31] https://www.civic.com/intel.

[32] Reid, Fergal/Harrigan, Martin, An Analysis of Anonymity in the Bitcoin System, 2011, p. 26, https://arxiv.org/abs/1107.4524.

There is no clear answer to the question whether blockchain technology is a threat or a solution to data protection. However, a lot depends on the design of the blockchain. The «perfect» one may in some case be a serious threat, immutable and at the same time not perfectly anonymous. Fortunately, it is possible to have blockchain respecting data protection by design. Corresponding examples were given in this paper, and I am sure that with the rapidly growing popularity of the blockchain technology new ideas will emerge.

I am also sure that as blockchain technology will become widely adopted new issues will arise. Mostly, I am concerned about personal data and every detail within blockchain databases that may lead to disclosing this personal data.

There is no doubt that blockchain technology is both a challenge for programmers and lawyers, but also a chance to protect privacy, despite being the potential threat to privacy itself.