CONCEPTS FOR GDPR-COMPLIANT PROCESSING OF PERSONAL DATA ON BLOCKCHAIN: A LITERATURE REVIEW

Florian Zemler

Faculty of Computer Sciences and Mathematics

OTH Regensburg
Prüfeninger Straße 58
93049 Regensburg, Germany

ABSTRACT

Blockchain is an emerging technology that is currently highly discussed in academia and practice. It offers a new approach of sharing data with participants in a network without the need to fully trust them. Among other reasons, this can be achieved because data once stored on a Blockchain is immutable. The recently introduced General Data Protection Regulation (GDPR) standardizes the European privacy regulations and brings major changes regarding how to handle personal data. Literature agrees that processing personal data on Blockchain is not compliant with the requirements of the GDPR. The present paper conducts a structured literature review and identifies four possible concepts for potentially GDPR-compliant data processing using Blockchain: Encryption & Key Destruction, Legal Argumentation, Off-Chain Storage, Redactable Blockchain. Each concept is described and analyzed in view of GDPR's requirements. It becomes clear that the concepts Legal Argumentation and Encryption & Key Destruction might at the moment not be totally compliant with the GDPR.

KEYWORDS

Blockchain, GDPR, Personal Data, Compliance

INTRODUCTION

Blockchain is an emerging technology and many different use-cases have been identified by scholars and practitioners. One of the first applications building upon that technology was Bitcoin, a peer-to-peer cash system (Eberhardt and Tai 2017). Besides the use of Blockchain for cryptographic currencies, the idea of an immutable and distributed ledger is getting more interesting for a large range of different industries and business sectors. Therefore, it is reasonable that applications are and will be designed which process personal data on Blockchain. Especially since late May 2018, when the General Data Protection Regulation (GDPR) came to effect, companies are sensitized how personal data is processed because unlawful handling of it can be heavily prosecuted. Especially when personal data is stored on Blockchain, the literature, e.g., Finck (2018) and Marnau (2017), already identified a conflict with GDPR's requirements. The goal of the paper at hand is to identify concepts and solutions for GDPR-compliant processing of personal data on Blockchain. Thus, it poses the following research question:

RQ: What are possible concepts to enable GDPR-compliant processing of personal data on Blockchain?

To answer this research question, the paper conducts a structured literature review, because it is an appropriate method and creates a "firm foundation for advancing knowledge" (Webster and Watson 2002). In addition, this review helps later researchers to analyze what is already known in this research area.
The paper first provides an overview of the Blockchain technology and the GDPR. Then, the conflict between both is outlined which occurs when personal data is processed using Blockchain. The next part describes the research methodology to allow replicability of this review and provides a descriptive analysis of the identified literature. The next chapter presents the findings and evaluates the results. It starts with a concept-centric matrix of the literature. Then, all identified concepts for GDPR-compliant processing of personal data on Blockchain are presented. Each concept is described according to the literature and is analyzed in view of the GDPR's requirements. In the last section, the concepts are evaluated, and the advantages and disadvantages are compiled. Finally, the results of the literature review are summarized and a proposal for further research based on the results of the present paper is provided.

The paper at hand contributes to current research by summarizing the existing knowledge and providing a first in-depth analysis of potentially GDPR-conform concepts for processing personal data on Blockchain in light of GDPR's requirements. To the author's best knowledge, this is the first structured literature review addressing this aspect.

BACKGROUND

The following chapter provides a basic overview of the Blockchain technology and the GDPR. The first section briefly presents the concept of Blockchain. The second part provides an overview of the scope of the GDPR. The last part outlines the conflict between both.

Blockchain

Blockchain was first introduced by Nakamoto in 2008 in form of the peer-to-peer cash system Bitcoin (Marnau 2017; Azaria et al. 2016; Eberhardt and Tai 2017).
With Blockchain, data is distributed among all participants (nodes) in the network (Viriyasitavat and Hoonsopon 2018). Data is stored as a transaction and several transactions are summarized to a single block (Dinh et al. 2018). Every block holds a reference to its precursor. Blocks can only be added to the end of Blockchain which leads to an append-only data structure (Dinh et al. 2018).
The connection between two blocks is established by the cryptographic hash of a block which is stored in its successor (Zheng et al. 2017). The following Figure 1 provides an overview of this procedure.

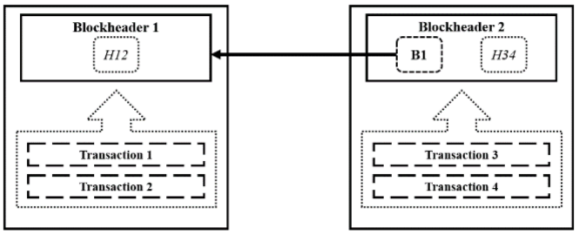

Figure 1: Simplified representation of two blocks in a Blockchain. In dependence on Drescher 2017.

The fact that each block is connected to its precursor by its cryptographic hash leads to the immutability of a Blockchain. Every change of a transaction, which is already stored on-chain, would modify the hash of every block after it and thereby the hash values of the whole Blockchain.
The special feature of Blockchain is that the participants in the network do not have to fully trust each other, because the technology ensures that every participant agrees on a common consensus (Dinh et al. 2018). This is accomplished by so-called consensus algorithms (Dinh et al. 2018).
The Blockchain technology can be divided into three main types:

- **Public Blockchain**: Completely decentralized and open to everyone (Zheng et al. 2017; Viriyasitavat and Hoonsopon 2018)
- **Private Blockchain**: Completely centralized and only selected participants can read and create transactions (Zheng et al. 2017; Vo et al.)
- **Consortium Blockchain**: Partly decentralized and managed by several organizations (Zheng et al. 2017; Ateniese et al. 2017)

GDPR

The GDPR came to effect on May 25th, 2018 and replaced the existing Data Protection Directive from 1995. The GDPR is relevant for every automated processing of personal data in the EU or from subjects located within the EU. In Art. 4 (1) GDPR defines personal as:

any information relating to an identified or identifiable natural person […]; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name [or] an identification number […].

The processing of personal data must meet certain requirements. Art. 5 GDPR prescribes that every processing must obey the "principles relating to processing of personal data". These principles state, among others, that personal data must be correct, up to date, and that incorrect data must be erased immediately. Moreover, data should only be stored as long as it is necessary.
The GDPR grants the data subject, i.e., the person whose personal data is processed, several additional rights regarding the processing of its data. Art. 16 GDPR allows the data subject to have incorrect personal data corrected. The right to be forgotten of Art. 17 GDPR grants the right that under certain conditions the data subject can have its personal data deleted. Additional rights are, e.g., the right to object (Art. 21 GDPR) or the right of access of the data subject by Art. 15 GDPR.
Security and privacy of the processed personal data is also an important part of the GDPR. In Art. 25 GDPR demands appropriate technical and organizational measures to guarantee the principles of processing personal data as well as procedures to only process data of a data subject that is necessary for a certain scenario. Art. 32 GDPR deals with the security of personal data processing. It requires the implementation of technical and organizational measures to secure the processing. These measures are, among others, the use of encryption and pseudonymization as well as methods of ensuring the confidentiality, integrity, and availability of personal data.
Finally, Art. 44 GDPR manages the transfer of personal data to third countries. The transfer of these kinds of data is, in general, only permitted when the receiving country ensures a proper level of protection. If this is not the case, there are some additional conditions which however permit the transfer. According to Art. 49 GDPR this can be for example the explicit consent of the data subject.

Conflict between Blockchain and GDPR

After presenting the basic concept of the Blockchain technology and an overview of the GDPR-requirements of processing personal data, it becomes clear that conflict situations may arise when personal data is processed on Blockchain.
Distributing data among all participants in a network and making changes impossible are basic concepts of the

Blockchain technology. But in view of privacy regulations, this leads to several problems. As mentioned before the processing of personal data must follow certain principles according to Art. 5 GDPR.

Marnau (2017) compared the requirements of Art. 5 with the Blockchain technology and came to the conclusion that especially the principles of Accuracy (Art. 5 (1) lit. d) and Storage limitation (Art. 5 (1) lit. e) cannot be easily fulfilled. Due to its immutability, personal data stored on Blockchain cannot be updated or deleted if necessary. This also contradicts the right to be forgotten and the right of rectification of the data subject.

Therefore, it must be avoided to store personal data on Public Blockchain because, first, it cannot be guaranteed that the data is used in proper ways (Marnau 2017) and, second, it might be possible that the data is transferred to countries which do not have a certain level of protection according to Art. 45 (1) GDPR.

It seems obvious that with the basic concept of Blockchain, it is not possible to fulfill these requirements for lawful processing of personal data of the GDPR.

METHODOLOGY

The following chapter provides an overview of the methodology applied in this paper.

As recommended by Webster and Watson (2002) the review starts with collecting relevant peer-reviewed literature by searching in commonly known databases of high-quality journals and conferences.

The analysis starts with a keyword search consisting of the terms “Blockchain” AND “GDPR” AND “Personal Data” in all fields in the current top 10 Information System and Management journals[1]. This approach generated no results. A second research in the commonly known conference proceedings of the AMCIS, ECIS, HICSS, ICIS, and PACIS lead to the same result of no records.

Because of the lack of high-quality literature, the search scope was extended to peer-reviewed articles in further journals and conferences as well as non-peer-reviewed and practitioners’ sources. For this approach, the same keywords were used in commonly known academic databases. The following Table 1 provides an overview of the databases that were consulted and the corresponding number of results.

Table 1: List of databases and number of results for the keyword search

Nr.	Database	Results
1	Google Scholar	867
2	ScienceDirect	32
3	IEEE Xplore	5
4	EBSCOhost	2
5	ProQuest	0
		906

In the first step, the articles were reviewed for relevancy by reading their titles. Then, the abstracts of the remaining 60 articles were analyzed for relevancy. The remaining 40 articles were reviewed for relevancy by skimming and scanning their full texts. In the end, a total number of 16 relevant sources remained. Then, as suggested by Webster and Watson (2002) a forward and backward search was performed to identify additional relevant literature. This procedure lead to three more articles. To identify concepts from practitioner and further sources a Google search with the same keywords was performed and the first 100 results were inspected. This procedure revealed an additional amount of 9 articles and papers for the literature review. In summary, the keyword search returned 28 sources. Figure 2 provides an overview of the literature selection process.

The following Figure 3 shows the distribution of the identified literature over time as well as their appearance

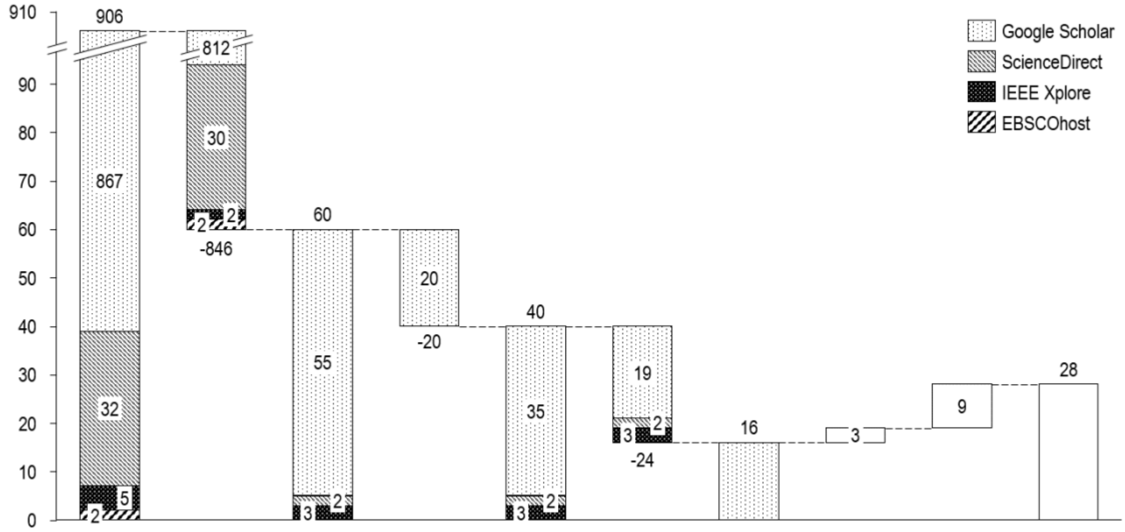

Figure 2: Waterfall diagram representing the literature selection process. Own figure.

[1] https://www.scimagojr.com (last access on 28.12.2018). Journals sorted by h-index.

in a journal, conference, or other sources like internet documents.

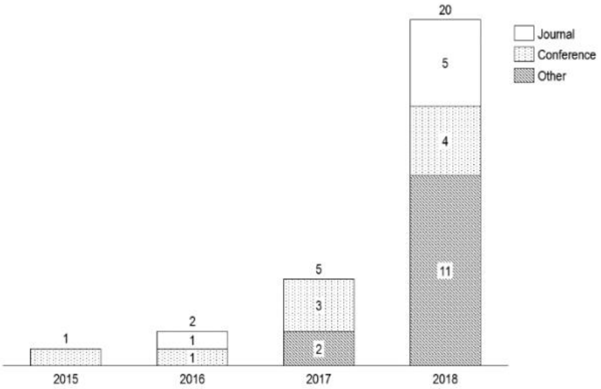

Figure 3: Overview of the distribution of the publications by year and type. Own figure.

It seems to be reasonable that most of the relevant literature (20; 71%) was published in 2018 when the GDPR came into effect. Many of the papers (13; 46%) come from non-peer-reviewed sources like internet documents or government (related) facilities' reports. This shows the need for more scholarly literature in this specific research area.

As recommended by Webster and Watson (2002) the results were structured concept-centric. The review of the literature identified two main concepts:

1. Off-Chain Storage
2. Redactable Blockchain

Beside these two concepts above, two others, the Legal Argumentations, and the Encryption & Key Destruction are mentioned in the literature. Each of these concepts is described according to the reviewed literature and checked against the requirements of the GDPR. In the end, the concepts are evaluated, and their advantages and disadvantages are outlined.

FINDINGS

This chapter presents the findings of the literature review. First, an overview of the analyzed literature and the corresponding concepts is presented. Then, the identified concepts from the literature are introduced and described. Each concept is analyzed in view of the GDPR. In the end, a final evaluation of the concepts is presented, which outlines their individual advantages and disadvantages.

Overview of the analyzed literature

The review of the papers and practitioner reports revealed two main concepts: The Off-Chain Storage and the Redactable Blockchain. In addition, further concepts were identified which are summarized under the item "Other". Each concept is explained in detail below.

The following Table 2 provides an overview of the analyzed articles in alphabetical order and the concepts they are representing.

Table 2: Overview of the analyzed papers and their corresponding concepts

Author(s)	Off-Chain Storage	Redactable Blockchain	Other	Peer-reviewed?
Ateniese et al. (2017)		✓		✓
Azaria et al. (2016)	✓			✓
Berberich and Steiner (2016)	✓		✓	✓
Cichosz et al. (2018)	✓			✓
CNIL (2018)	✓		✓	✗
Eberhardt and Tai (2017)	✓			✓
Eichler et al. (2018)	✓			✗
Esposito et al. (2018)	✓			✓
Fan et al. (2018)	✓			✓
Finck (2018)	✓	✓		✓
Gräther et al. (2018)	✓			✓
Ibáñez et al. (2018)	✓	✓	✓	✗
Jensen (2018)	✓		✓	✗
Jussila (2018)	✓	✓	✓	✗
Katuwal et al. (2018)	✓			✗
Kuner et al. (2018)		✓		✓
Lima (2018)	✓		✓	✗
Michels (2018)	✓		✓	✗
Moser (2017)	✓			✗
Pagallo et al. (2018)	✓	✓	✓	✓
Salmensuu (2018)			✓	✗
Steichen et al. (2018)	✓			✓
The European Union Blockchain Observatory & Forum (2018)	✓		✓	✗
van Geelkerken and Konings (2017)			✓	✓
van Humbeeck (2017)	✓			✗
Wirth and Kolain (2018)		✓		✓
Zhangy et al. (2018)	✓			✗
Zyskind et al. (2015)	✓			✓
	23	7	11	15

Table 2 shows that most of the available literature deals with Off-Chain Storage as a possible solution for the Blockchain and personal data conflict. It is followed by

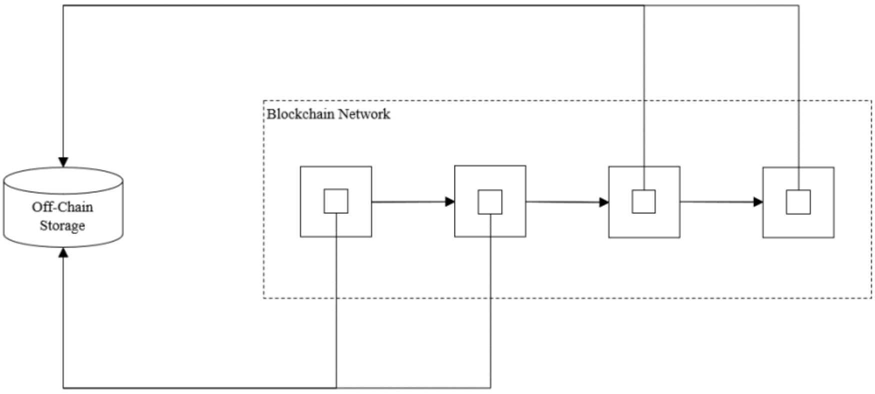

Figure 4: Off-Chain Storage architecture. Own figure.

the Redactable Blockchain as a sole concept and two additional concepts summarized under the term “Other”.

Identified concepts

This section details the identified concepts for storing personal data using Blockchain. It starts with Off-Chain Storage, which is the most discussed concept. In the next step, the Redactable Blockchain concept is introduced. In the last section, two more concepts are presented.

Off-Chain Storage

Off-Chain Storage is the most discussed concept in the reviewed literature for GDPR-compliant processing of personal data on Blockchain. Storing data "off-chain" means, that the data, in this case, the personal data or in general the payload, is not kept inside the Blockchain network, but stored outside, e.g., in a traditional database (Esposito et al. 2018; Ibáñez et al. 2018). Only a reference (for example, a hash value) to the outside storage location where the actual data is stored, is saved on the Blockchain (Katuwal et al. 2018; Zyskind et al. 2015; Steichen et al. 2018; Pagallo et al. 2018; van Humbeeck 2017). Figure 4 provides a simplified overview of an Off-Chain Storage architecture.

In general, storing larger data sets outside of a Blockchain network is highly recommended, because storage capacity on a Blockchain is relatively cost-intensive (Zhangy et al. 2018) and currently not very performant (Jensen 2018). Omaar (2017) identified the cost of storing one Gigabyte of data on the Etherum Blockchain to be approximately 17.500 Ethereum per Gigabyte which was worth approximately two million euros in March 2019

As Figure 4 illustrates, off-chain data is stored outside of the Blockchain network. That often implies the reintroduction of a trusted third party (TTP) which guarantees the confidentiality and integrity of the data. This provides a certain degree of control to a centralized party that seems to be a violation of the principles of Blockchain (Ibáñez et al. 2018). For this reason, Eberhardt and Tai (2017) recommend the use of a content addressable storage for off-chain location, which stores the files not by their names, but by their hash values. This has the advantage that it is now possible to trustless outsource data because a change of the data would lead

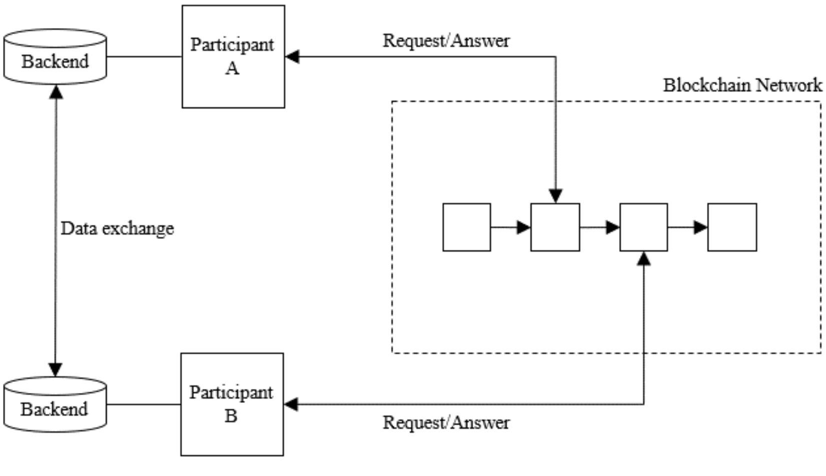

Figure 5: Off-Chain Storage concept by van Humbeeck (2017). Own figure.

to an adjustment of its hash value and therefore to its storage location (Eberhardt and Tai 2017).

In the reviewed literature Zyskind et al. (2015) came up first with a solution to use Blockchain in combination with a distributed file system and to only store the reference on-chain to avoid a centralized storage location (Zyskind et al. 2015).

In practice, Steichen et al. (2018) and Gräther et al. (2018) make use of this technique in their applications. They both use the decentralized Interplanetary File System as their content addressable storage system. That means, that no centralized location is needed, and trust is ensured by the way the data is stored on the file system.

Van Humbeeck (2017) presents a slightly different concept. In his solution, data is also stored off-chain, but not in a centralized location or in a content addressable storage, instead in the backend system of each participant of the Blockchain network. The Blockchain itself only contains links to the off-chain locations and the hash values of the requested data (van Humbeeck 2017). If a participant has the privilege and the need to access a certain data set, s/he receives the reference (for example, in form of an access token or database string) to where the data is stored and the corresponding hash value. Then, the requestor can directly fetch the data from the backend system where it is stored. Figure 5 provides an overview of the architectural design.

In the context of the GDPR, storing personal data off-chain brings, at first sight, many benefits.

Some argue that with this procedure no personal data stays on the Blockchain and therefore the requirements of the GDPR can be met (Steichen et al. 2018; Katuwal et al. 2018; Ibáñez et al. 2018). However, in most cases, the hash value of the personal data is still kept on-chain and works as a reference or as proof-of-correctness. At the moment it cannot be said with absolute certainty that hash values of personal data can be regarded as anonymous data or should more likely be treated as pseudonymous data (Eichler et al. 2018). This objection is based on a report by the Article 29 Data Protection Working Party (2014), a former advisory body of the European Commission, which clearly states that hashing techniques must be considered as pseudonymization. The fact that hashed personal data should be considered as pseudonymous data is heavily discussed at the moment (The European Union Blockchain Observatory & Forum 2018) but seems to be widely accepted in the reviewed literature (Finck 2018; Ibáñez et al. 2018; Jensen 2018; Jussila 2018). The proposition, whether the hash value of personal data should be considered as pseudonymized personal data depends on the linkability, i.e. the possibility to connect the hashed data with the original data, between them. Article 29 Data Protection Working Party (2014) argues, that even if a hash cannot be reversed, it can simply be recalculated if the range of input values and the hash function are known. The risk of hash recalculation can be minimized by adding additional information to the dataset, for example, a secret key (Article 29 Data Protection Working Party 2014; The European Union Blockchain Observatory & Forum 2018).

The French National Commission on Informatics and Liberty, CNIL (2018), published a report on the responsible use of Blockchain and personal data. They recommend to only store personal data on a Blockchain as a cryptographic commitment[2]. When that is not

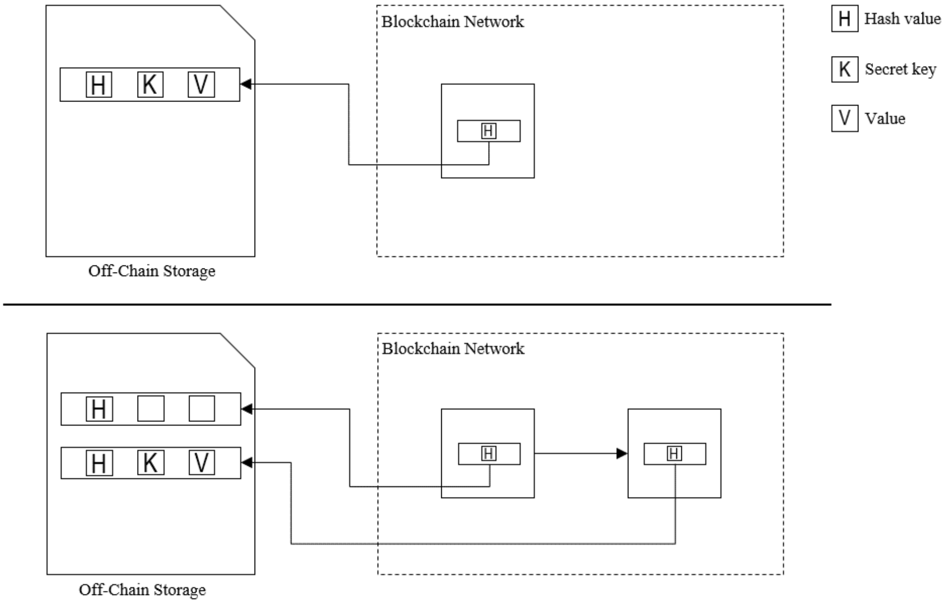

Figure 6: Example of rectification of data with Off-Chain Storage. Own figure

[2] Describes a method where it is possible to commit to a certain value but keeping this information hidden to others. Later, it is possible to reveal this information. The procedure seems to be firstly described by Blum1983 in form of the “coin flipping by telephone” problem.

possible the personal data should be stored as a keyed-hash value on Blockchain. And only if that again is not possible, state of the art encryption algorithms should be applied.

Especially when looking at the data subjects' right to be forgotten and the right of rectification, it is technically not possible to fulfill them. The CNIL (2018) however argues that with the use of state-of-the-art keys and algorithms it is possible to move closer to the desired effects. The erasure of personal data can be performed by deleting the data off-chain and the corresponding key, used for generating the hash value, which is stored on-chain. In this case, it is not possible to prove or verify which data has been hashed (CNIL 2018) and the data staying on-chain could be considered as anonymous data (Eichler et al. 2018). In view of the right of rectification, the old dataset can be deleted as described before and a new transaction containing the corrected data can be submitted to the Blockchain. Figure 6 above provides an overview of this technique.

In summary, it can be said, that at the moment there is no legal guarantee that working with cryptographic references based on personal data on-chain is a GDPR-compliant concept for personal data processing using Blockchain. Even referenced data may still be perceived as pseudonymized personal data from a legal perspective and erasure or rectification is technically not possible. Of course, these circumstances can change over time with changing legal perspectives on this procedure.

Redactable Blockchain

Redactable Blockchain was firstly mentioned in the reviewed literature by Ateniese et al. (2017) and is a relatively new concept. By "redactable" the authors mean to rewrite one or more blocks that were already written on Blockchain, to compress any number of already existing blocks to a smaller number and to insert one or more blocks to the existing chain (Ateniese et al. 2017). At first, this seems to contradict the immutability of Blockchain, one of its basic concepts. However, Ateniese et al. (2017) argue that immutability may not be appropriate for all new applications based on Blockchain technology. This can, i.e., be the storage of files or the management of personal health records. This data should be able to be deleted if they contain errors or it is required by law (Ateniese et al. 2017). This law could be the GDPR.

The immutability of a Blockchain comes from the collision resistance of the hash values that connect each block to its precursor. To make Blockchain mutable the concept makes use of a special form of a so-called "chameleon hash function" (Ateniese et al. 2017). A chameleon hash function works like any other hash function with the difference that it has something like a trapdoor which can be used to generate collisions (Ateniese et al. 2017). These collisions can, e.g., be used to alter transactional data without changing the corresponding hash value of the block and by this maintain the connection to its successor. Ateniese et al. (2017) describe the procedure like adding a lock to the connection between two blocks which can be opened with the right key.

Figure 7 presents the three different phases of a Redactable Blockchain. In (1) the Redactable Blockchain behaves like any other Blockchain and no modification is possible because all locks are secured. At (2) the connection between B2 and B3 is opened with the secret key and modification is possible. In (3) the modifications at B2 are finished and result in form of the block B2'. The connection between B2' and B3 is locked again and no more modification is possible.

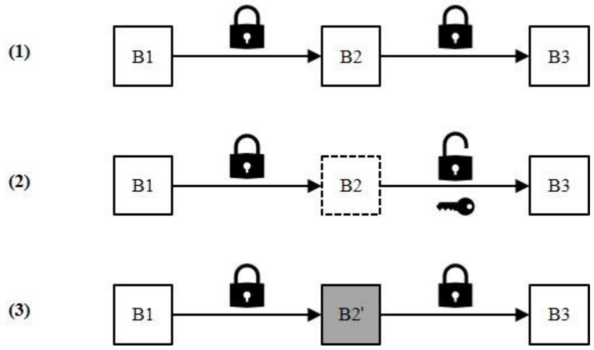

Figure 7: Principles of the Redactable Blockchain. Own figure. Adapted from Ateniese et al. 2017

It is important to notice that when the key to the lock of the hash function is lost or gets destroyed it is not possible to modify the blocks and the Blockchain is immutable again (Ateniese et al. 2017). Therefore, the management of the trapdoor key is an essential part of this concept. Ateniese et al. (2017) describe that, e.g., in a Private Blockchain network the key could be given to the central authority or in a Consortium Blockchain it could be shared among all the participants of the network.

In the analyzed literature only a few authors, like Finck (2018), Ibáñez et al. (2018), or Pagallo et al. (2018), identified the potential use of the Redactable Blockchain concept in the context of the Blockchain and GDPR conflict. A real-world application with a Redactable Blockchain could not be identified in the reviewed literature.

The Redactable Blockchain could be an interesting solution for the described conflict (Jussila 2018). The idea of directly removing blocks containing personal data would solve many problems. However, the concept also faces some problems. First, adding redactability to an existing Blockchain is not possible, that means that the decision for this concept must be made before the network is set up (Ibáñez et al. 2018). Secondly, old copies of Blockchain would still contain the redacted data (Ateniese et al. 2017; Finck 2018) but a compliant Blockchain node will accept the redacted data (Ibáñez et al. 2018) and delete the old copies (Ateniese et al. 2017). Finally, there is always the risk that a party redacts the Blockchain to its favor (Ibáñez et al. 2018).

In the context of the GDPR, this concept seems to solve the problems of the conflict between storing personal data on Blockchain and the data privacy regulation. The ability to delete or alter data after they have been stored on-chain is a huge benefit. Of course, it must be ensured that all the participants in the Blockchain network are operating in compliance with the GDPR that means that redactions to the Blockchain are performed immediately and old copies are deleted trustworthy. This procedure can, for example, be audited by an external party. Critics may argue that giving the possibility to redact data on the Blockchain to a single entity is a violation of the basic principles of the technology (Pagallo et al. 2018). Even the general opportunity of altering data stored on the Blockchain contradicts the basic idea of Blockchain. Ateniese et al. (2017) argue that the immutability of Blockchain should be reconsidered as applications, based on this technology, develop from simple lab experiments to real-world deployments.

Other

The review of the literature also identified two additional proposals for solutions which could not be assigned to one of the two previously described concepts. This section briefly describes these concepts and discusses them in the context of GDPR.

Legal Argumentation

Legal Argumentation deals with the imprecise definition of some parts of this regulation. Due to the immutability of the Blockchain technology, it is a real challenge to delete personal data once stored on Blockchain. From a legal point of view, the literature review identified two different ways of arguing against this proceeding.

Ibáñez et al. (2018) bring forward that the right to be forgotten is not an absolute right and the concept of erasure leaves room for interpretation. In their example, they present that erasure or amendment is performed by adding a new transaction to Blockchain which contains a reference to the obsolete entry and invalidates it semantically (Ibáñez et al. 2018).

Another solution is presented by Berberich and Steiner (2016) who illustrate that with the help of Article 17 (1) lit. b it might be possible to argue that personal data stored on-chain is required for processing because Blockchain needs a persistent chain to function correctly. Therefore, the right to be forgotten of the data subject could not be applied here.

All these legal argumentations should currently be treated with prudent care. Because of the lack of judgments on this field, this concept operates in a grey area. It is very likely that these procedures could be considered illegal by a court in the future. Then, it is not possible to remove the personal data form an existing Blockchain, even if required by law.

Encryption & Key Erasure

The last concept identified in the course of the literature review deals with encrypting data on a Blockchain to reach GDPR compliance. Some studies suggest encrypting the data that is stored on a Blockchain and, when it must be deleted, simply destroy the encryption key (van Geelkerken and Konings 2017; Jussila 2018; Pagallo et al. 2018; Michels 2018; Jensen 2018; Ibáñez et al. 2018). This concept assumes that with the use of state-of-the-art encryption techniques data gets inaccessible when the encryption key is not available anymore. Following their argumentation, this procedure comes close to the erasure of data.

In view of the GDPR, this concept must be seen critically. First, it is possible that today's encryption algorithms are no longer considered secure in the future so that it might be possible to decrypt the data without the knowledge of the original encryption key (Michels 2018; Ibáñez et al. 2018; Eichler et al. 2018). Second, Article 29 Data Protection Working Party (2014) clearly states that encryption must be considered as a form of pseudonymization and it does not automatically anonymize the data. It is important to notice that encryption only guarantees confidentiality over a certain time period but anonymization should last indefinitely (Article 29 Data Protection Working Party 2014).

In general, it can be said that the use of Encryption & Key destruction should not be considered as the main technique for a GDPR-compliant concept.

Evaluation

In the previous part, the different concepts from the literature were described and critically reflected. This section summarizes the results and provides a final evaluation.

The literature review revealed two main concepts of processing personal data on Blockchain in a GDPR-compliant way. Two additional, but not as strongly represented concepts, where also among the findings. Each of these concepts is for itself considered more or less suitable to fulfill the requirements of the GDPR.

The following Table 3 provides an overview of the concepts discussed in the paper at hand and summarizes the advantages and disadvantages of each concept.

It becomes clear that every concept described below has its own advantages and disadvantages. Currently, it seems that the Redactable Blockchain concept can fulfill the requirements of the GDPR because it enables the change of data directly on a Blockchain. On the downside, this contradicts the immutability of Blockchain, and it could be abused by participants in the network.

Off-Chain Storage may require the reintroduction of a TTP, which, in some aspects, contradicts the basic idea of the Blockchain technology. There are ways of avoiding a central storage location and thus a TTP, for example, by distributing the data among the participants in a Blockchain. However, this cannot always be feasible depending on the use-case. In view of privacy regulations, it seems to be possible to process personal data on a Blockchain with the Off-Chain Storage concepts when no personal data is stored on the blocks.

Table 3: Advantages and disadvantages of the identified concepts

Concept	Advantages	Disadvantages
Off-Chain Storage	– Personal data is stored off-chain – Only a reference to the off-chain storage location and the hash of the data is kept on-chain – Enables the processing of large data	– May require the reintroduction of a TTP – Depending on the way personal data is stored off-chain, the hash on-chain must be handled as personal data – May require major technical modifications depending where the data is held
Redactable Blockchain	– Provides a way of altering and deleting personal data directly on Blockchain – Does not require a change in the way Blockchain is used	– Requires technical adaptions – Redactability cannot be added to an existing Blockchain – If the secret key is lost, the Blockchain remains immutable – Nodes are required to delete old copies of the redacted blocks – Nodes can abuse this feature to their favor – May require off-chain storage for large data
Legal Argumentations	– No changes of the Blockchain technology required – Data which is obsolete can simply be invalidated with a new transaction	– Concept must be handled very carefully because it operates in legal grey area – Judgment can declare this concept as not compliant with the GDPR – May require off-chain storage for large data
Encryption & Key Erasure	– No changes of the Blockchain technology required – Data on-chain is stored encrypted – When data needs to be removed from Blockchain the encryption key is simply destroyed	– Should not be used as the main concept for GDPR-compliance – Encrypted personal data must be handled as pseudonymized data – It might be able to decrypt data in the future with the use of up-to-date technology – Judgment can declare this concept as not compliant with the GDPR – May require off-chain storage for large data

Besides privacy concerns, keeping data outside of the Blockchain network is often reasonable, otherwise, there might be a loss of performance when the datasets get larger. Therefore, Redactable Blockchain concepts might reach their limits quickly depending on the use case. In such situations, a combination of redactability and off-chain storage could make sense.

The concepts, summarized under the term "other", are discussed to a lesser extent. The Legal Argumentations and the Encryption & Key Destruction concept both make use of interpretations of the GDPR which may be declared invalid by a court of law. Therefore, it seems reasonable to wait for judgments in this field before investing further research in this specific topic with the risk of a court ruling against these procedures.

CONCLUSION

The paper at hand conducted a structured literature review and identified four potential concepts for GDPR-compliant processing of personal data using Blockchain in academic literature and practitioner sources. A first analysis of the literature revealed a lack of scholarly research papers in this specific research field. Many of the sources (13; 46%) for this review did not undergo a peer-review process. A reason for this might be the rapid development of Blockchain technology which is mainly documented in internet sources as well as the recent legal validity of the GDPR.

The concept that is discussed the most in the reviewed literature is Off-Chain Storage. With this concept, personal data is stored outside of the Blockchain network and only references the off-chain location on Blockchain. The problem is that in most cases a "proof-of-correctness" in form of the hash of the data is kept on-chain and depending on the hashed data this could as well be personal data. This concept is already applied in practice.

The second identified concept, Redactable Blockchain, makes use of a special hashing algorithm, which allows data altering of a block without changing its corresponding hash. This enables altering and deleting data on Blockchain and is in view of the requirements of the GDPR a very interesting solution. On the downside, this concept is often criticized as a violation of the immutability of Blockchain, one of its fundamental characteristics.

The next concept, Legal Argumentations, makes use of interpretations of the GDPR. On the one hand, it can be argued that the right to be forgotten is not an absolute right and data on Blockchain could be altered or deleted by adding a new transaction to the block which declares the old one invalid. On the other hand, it could be possible to claim that consistent data processing is required for Blockchain to function properly. Therefore, no data can be altered or deleted. This argumentation can be based on Art. 17 (1) GDPR.

The last concept, Encryption & Key Erasure, describes a technique in which the personal data is encrypted before storing the data on Blockchain and the key is stored apart. If the data is not required anymore, the encryption key is simply deleted, and the data is thereby not accessible anymore. The main critic of this method is that in the near future it might be possible to decrypt the data without the need of the decryption key and the data gets accessible again.

In general, it is currently strongly recommended to not store personal data on a Public Blockchain. If personal data was stored on a Public Blockchain, it could not be guaranteed that the data stays within the territorial scope of the GDPR and it might violate the principles of transferring personal data to third countries. The Legal Argumentation concept acts in grey area and so it is always possible that this method is declared illegal. The situation is similar to the concept of Encryption & Key Erasure. In this concept, the personal data stays encrypted on Blockchain, as long as the Blockchain exists. Even if the decryption key is destroyed, the data currently counts as personal data.

It seems that only the Off-chain Storage and the Redactable Blockchain concepts can guarantee compliant processing under data protection regulations. When using the Off-Chain Storage concept it should be paid attention that no recalculating of the hash value, which often stays on-chain as a "proof-of-correctness", is possible. If the Redactable Blockchain concept should be applied, it is important to manage the secret key very securely. If the key is lost, it is not possible to redact data stored on Blockchain.

Referring to the research question, the review revealed four possible concepts for GDPR-compliant processing of personal data in the literature. Based on the analysis in view of the GDPR, only the Off-Chain Storage and Redactable Blockchain concept can fulfill the data privacy requirements at the moment.

Based on the results of the present paper it is now possible for further researchers to develop Blockchain applications which allow processing of personal data in compliance with the GDPR.

Limiting factors of this literature review are that nearly half of the sources are not peer-reviewed and their significance might not meet all academic standards. As mentioned before the Blockchain technology is in rapid development and most of this knowledge is shared on internet platforms. Due to the research methodology, the paper at hand focused on research papers and less on online articles from developers or independent researchers. In addition, the paper did not examine the practical applicability of the identified concepts. Hence, it is possible that a concept cannot be realized with current state of the art technology.

REFERENCES

Article 29 Data Protection Working Party (2014): Opinion 05/2014 on anonymisation techniques. 0829/14/EN WP216. Edited by Article 29 Data Protection Working Party. Available online at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, checked on 12/30/2018.

Ateniese, Giuseppe; Magri, Bernardo; Venturi, Daniele; Andrade, Ewerton (2017): Redactable Blockchain – or – Rewriting history in Bitcoin and friends. In : 2017 IEEE European Symposium on Security and Privacy (EuroS&P). 2017 IEEE European Symposium on Security and Privacy (EuroS&P). Paris, France, 26.04.2017 - 28.04.2017: IEEE, pp. 111–126.

Azaria, Asaph; Ekblaw, Ariel; Vieira, Thiago; Lippman, Andrew (2016): MedRec: Using Blockchain for medical data access and permission management. In : 2016 2nd International Conference on Open and Big Data (OBD). 2016 2nd International Conference on Open and Big Data (OBD). Vienna, Austria, 22.08.2016 - 24.08.2016: IEEE, pp. 25–30.

Berberich, Matthias; Steiner, Malgorzata (2016): Blockchain technology and the GDPR – How to reconcile privacy and distributed ledgers? In *European Data Protection Law Review* 2 (3), pp. 422–426. DOI: 10.21552/EDPL/2016/3/21.

Blum, Manuel (1983): Coin flipping by telephone a protocol for solving impossible problems. In *ACM SIGACT News* 15 (1), pp. 23–27.

Cichosz, Simon Lebech; Stausholm, Mads Nibe; Kronborg, Thomas; Vestergaard, Peter; Hejlesen, Ole (2018): How to use Blockchain for diabetes health care data and access management: An operational concept. In *Journal of diabetes science and technology*. DOI: 10.1177/1932296818790281.

CNIL (2018): Blockchain. Solutions for a responsible use of the Blockchain in the context of personal data. Available online at https://www.cnil.fr/sites/default/files/atoms/files/blockchain.pdf, checked on 12/28/2018.

Dinh, Tien Tuan Anh; Liu, Rui; Zhang, Meihui; Chen, Gang; Ooi, Beng Chin; Wang, Ji (2018): Untangling Blockchain: A data processing view of blockchain systems. In *IEEE Trans. Knowl. Data Eng.* 30 (7), pp. 1366–1385. DOI: 10.1109/TKDE.2017.2781227.

Drescher, Daniel (2017): Blockchain basics. A non-technical introduction in 25 steps. Place of publication not identified, New York, NY: Apress; Distributed to the book trade worldwide by Springer Science+Business Media New York. Available online at http://proquest.tech.safaribooksonline.de/9781484226049.

Eberhardt, Jacob; Tai, Stefan (2017): On or off the Blockchain? Insights on off-chaining computation and data. In Flavio de Paoli, Stefan Schulte, Einar Broch Johnsen (Eds.): Service-oriented and cloud computing, vol. 10465. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 3–15.

Eichler, Natalie.; Jongerius, Silvan; McMullen, Greg; Naegele, Oliver; Steininger, Liz; Wagner, Kai (2018): Blockchain, data protection, and the GDPR.

Blockchain Bundesverband. Available online at http://www.bundesblock.de/wp-content/uploads/2018/05/GDPR_Position_Paper_v1.0.pdf.

Esposito, Christian; Santis, Alfredo de; Tortora, Genny; Chang, Henry; Choo, Kim-Kwang Raymond (2018): Blockchain: A Panacea for healthcare cloud-based data security and privacy? In *IEEE Cloud Comput.* 5 (1), pp. 31–37. DOI: 10.1109/MCC.2018.011791712.

Fan, Kai; Wang, Shangyang; Ren, Yanhui; Li, Hui; Yang, Yintang (2018): MedBlock: Efficient and secure medical data sharing via Blockchain. In *Journal of medical systems* 42 (8), p. 136. DOI: 10.1007/s10916-018-0993-7.

Finck, Michèle (2018): Blockchains and data protection in the European Union. In *European Data Protection Law Review* 4 (1), pp. 17–35. DOI: 10.21552/edpl/2018/1/6.

Gräther, Wolfgang; Kolvenbach, Sabine; Ruland, Rudolf; Schütte, Julian; Torres, Christof; Wendland, Florian (2018): Blockchain for education: Lifelong learning passport. In W. Prinz & P. Hoschka (Ed.): Proceedings of the 1st ERCIM Blockchain Workshop 2018, Reports of the European Society for Socially Embedded Technologies.

Ibáñez, Luis-Daniel; O'Hara, Kieron; Simperl, Elena (2018): On Blockchains and the General Data Protection Regulation. Available online at https://eprints.soton.ac.uk/422879/1/BLockchains_GDPR_4.pdf.

Jensen, Greg (2018): Reconciling GDPR rights to erasure and rectification of personal data with Blockchain. Available online at https://blogs.oracle.com/cloudsecurity/reconciling-gdpr-rights-to-erasure-and-rectification-of-personal-data-with-blockchain, checked on 12/28/2018.

Jussila, Jani-Pekka (2018): Reconciling the conflict between the 'immutability' of public and permissionless Blockchain technology and the right to erasure under Article 17 of the General Data Protection Regulation.

Katuwal, Gajendra J.; Pandey, Sandip; Hennessey, Mark; Lamichhane, Bishal (2018): Applications of Blockchain in healthcare: Current landscape & challenges. Available online at http://arxiv.org/pdf/1812.02776v1.

Kuner, Christopher; Cate, Fred; Lynskey, Orla; Millard, Christopher; Ni Loideain, Nora; Svantesson, Dan (2018): Blockchain versus data protection. In *International Data Privacy Law* 8 (2), pp. 103–104. DOI: 10.1093/idpl/ipy009.

Lima, Claudio (2018): Blockchain-GDPR privacy by design. How decentralized Blockchain Internet will comply with GDPR data privacy. Blockchain Engineering Council. Available online at https://blockchain.ieee.org/images/files/pdf/blockchain-gdpr-privacy-by-design.pdf, checked on 12/28/2018.

Marnau, Ninja (2017): Die Blockchain im Spannungsfeld der Grundsätze der Datenschutzgrundverordnung. In Maximilian Eibl, Martin Gaedke (Eds.): Informatik 2017 - Bände I-III. Tagung vom 25.-29. September 2017 in Chemnitz. Bonn: Gesellschaft für Informatik (GI-Edition Proceedings, 275), pp. 1025–1036.

Michels, Dave (2018): Can Blockchain operators comply with EU data protection law? Available online at https://journal.binarydistrict.com/can-blockchain-operators-comply-with-eu-data-protection-law/, checked on 12/28/2018.

Moser, Jana (2017): The application and impact of the European General Data Protection Regulation on Blockchains. R3. Available online at https://www.r3.com/wp-content/uploads/2018/04/GDPR_Blockchains_R3.pdf, checked on 12/28/2018.

Nakamoto, Satoshi (2008): Bitcoin: A peer-to-peer electronic cash system. Available online at https://bitcoin.org/bitcoin.pdf, updated on 1/10/2019.

Omaar, Jamila (2017): Forever isn't free: The cost of storage on a Blockchain database. With assistance of Simon Schwerin, McMullen Greg. Available online at https://medium.com/ipdb-blog/forever-isnt-free-the-cost-of-storage-on-a-blockchain-database-59003f63e01, checked on 2/28/2019.

Pagallo, Ugo; Bassi, Eleonora; Crepaldi, Marco; Durante, Massimo (2018): Chronicle of a clash foretold: Blockchains and the GDPR's right to erasure. In M. Palmirani (Ed.): Legal Knowledge and Information Systems: JURIX 2018: The Thirty-first Annual Conference: IOS Press, pp. 81–90.

Salmensuu, Cagala (2018): The General Data Protection Regulation and Blockchains. University of Helsinki Law Faculty. Available online at https://ssrn.com/abstract=3143992.

Steichen, Mathis; Fiz, Beltran; Norvill, Robert; Shbair, Wazen; State, Radu (2018): Blockchain-based, decentralized access control for IPFS. In : 2018 IEEE International Conference on Blockchain. Halifax, Canada.

The European Union Blockchain Observatory & Forum (2018): Blockchain and the GDPR. The European Union Blockchain Observatory & Forum. Available online at https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf, checked on 12/28/2018.

van Geelkerken, F.W.J; Konings, K. (2017): Using Blockchain to strengthen the rights granted through the GDPR. In : 7th International youth science forum «Litteris et Artibus». Lviv, Ukraine, pp. 458–461.

van Humbeeck, Andries (2017): The Blockchain-GDPR paradox. Available online at https://medium.com/wearetheledger/the-blockchain-gdpr-paradox-fc51e663d047, checked on 12/28/2018.

Viriyasitavat, Wattana; Hoonsopon, Danupol (2018): Blockchain characteristics and consensus in modern business processes. In *Journal of Industrial*

Information Integration. DOI: 10.1016/j.jii.2018.07.004.

Vo, Hoang Tam; Kundu, Ashish; Mohania, Mukesh: Research Directions in Blockchain Data Management and Analytics.

Webster, Jane; Watson, Richard T. (2002): Analyzing the past to prepare for the future: Writing a literature review. In *MIS quarterly*, pp. xiii–xxiii.

Wirth, Christian; Kolain, Michael (2018): Privacy by BlockChain design: A Blockchain-enabled GDPR-compliant approach for handling personal data. In W. Prinz & P. Hoschka (Ed.): Proceedings of the 1st ERCIM Blockchain Workshop 2018, Reports of the European Society for Socially Embedded Technologies.

Zhangy, Shifa; Kim, Anne; Liu, Dianbo; Nuckchadyy, Sandeep C.; Huangy, Lauren; Masurkary, Aditya et al. (2018): Genie: A secure, transparent sharing and services platform for genetic and health data. Available online at http://arxiv.org/pdf/1811.01431v1.

Zheng, Zibin; Xie, Shaoan; Dai, Hongning; Chen, Xiangping; Wang, Huaimin (2017): An overview of Blockchain technology: Architecture, consensus, and future trends. In : 2017 IEEE International Congress on Big Data (BigData Congress). 2017 IEEE International Congress on Big Data (BigData Congress). Honolulu, HI, USA, 25.06.2017 - 30.06.2017: IEEE, pp. 557–564.

Zyskind, Guy; Nathan, Oz; Pentland, Alex 'Sandy' (2015): Decentralizing privacy: Using Blockchain to protect personal data. In : 2015 IEEE Security and Privacy Workshops. 2015 IEEE Security and Privacy Workshops (SPW). San Jose, CA, 21.05.2015 - 22.05.2015: IEEE, pp. 180–184.