Blockchain: Research and Applications 4 (2023) 100129

Contents lists available at ScienceDirect

Blockchain: Research and Applications

journal homepage: www.journals.elsevier.com/blockchain-research-and-applications

Research Article

A systematic literature review of the tension between the GDPR and public blockchain systems

Rahime Belen-Saglam [a,*], Enes Altuncu [a], Yang Lu [b], Shujun Li [a,**]

[a] Institute of Cyber Security for Society (ICSS) & School of Computing, University of Kent, Keynes College, Canterbury, CT2 7NP, United Kingdom
[b] School of Science, Technology and Health York, St John University, Lord Mayor's Walk, York, YO31 7EX, United Kingdom

ARTICLE INFO

Keywords:
Blockchain
Distributed ledgers
Privacy
Data protection law
Legal compliance
GDPR
EU
EEA
UK

ABSTRACT

Blockchain technology has been rapidly growing since Bitcoin was invented in 2008. The most common type of blockchain system, public (permissionless) blockchain system, has some unique features that lead to a tension with the European Union's General Data Protection Regulation (GDPR) and other similar data protection laws. In this paper, we report the results of a systematic literature review (SLR) on 114 research papers discussing and/or addressing such a tension. To the best of our knowledge, our SLR is the most comprehensive review of this tension, leading to a more in-depth and broader analysis of related research work on this important topic. Our results revealed three main types of issues: (i) difficulties in exercising data subjects' rights such as the 'right to be forgotten' (RTBF) due to the immutable nature of public blockchains; (ii) difficulties in identifying roles and responsibilities in the public blockchain data processing ecosystem (particularly on the identification of data controllers and data processors); and (iii) ambiguities regarding the application of the relevant law(s) due to the distributed nature of blockchains. Our work also led to a better understanding of solutions for improving the GDPR compliance of public blockchain systems. It can help inform not only blockchain researchers and developers but also policymakers and law markers to consider how to reconcile the tension between public blockchain systems and data protection laws (the GDPR and beyond).

1. Introduction

Since Bitcoin was conceptualised in 2008, its underlying technology about blockchains (also known as distributed ledgers) has been considered as a breakthrough of secure computing without a centralised authority in an open environment. Its potential capabilities led many researchers and practitioners to consider that it is the next big revolutionizing technology after the Internet [1]. Its applications have boomed in many sectors for various purposes and many researchers also started conducting research on this emerging technology. Although the blockchain technology has some built-in security and privacy mechanism by design, it has also introduced new security and privacy concerns, one of which is the conflict between the immutable nature of data on blockchain and the "right to be forgotten" (RTBF) of data subjects introduced in new data protection laws such as the European Union (EU) General Data Protection Regulation (GDPR) introduced in 2016 [2]. Such new concerns let many researchers, practitioners, policymakers and blockchain users to debate about legal compliance of blockchain systems and to explore ways to make blockchain systems more legally compliant with such new data protection laws and regulations. This paper aims at providing a comprehensive review of such efforts in the research literature.

After being passed by the European Parliament in 2016, the GDPR entered into force on May 25, 2018 in all EU member states. In addition, law markers in three non-EU member states of the European Economic Area (EAA), Iceland, Liechtenstein and Norway, also decided to adopt the GDPR. For the UK, after it left the EU, its law markers decided to keep the GDPR in its national law, but made some necessary changes to reflect the new status of the UK as a non-EU/EEA country, which led to the so-called UK GDPR [3], a UK-specific version of the EU GDPR. In the rest of the paper, we will use the term GDPR in a broad sense to refer to the two different versions of the GDPR, since the differences are not essential for

* Corresponding author.
** Corresponding author.
E-mail addresses: R.Belen-Saglam-724@kent.ac.uk (R. Belen-Saglam), ea483@kent.ac.uk (E. Altuncu), y.lu@yorksj.ac.uk (Y. Lu), S.J.Li@kent.ac.uk (S. Li).

https://doi.org/10.1016/j.bcra.2023.100129
Received 11 October 2022; Received in revised form 22 December 2022; Accepted 5 January 2023

our discussions on the relationships between the blockchain technology and the relevant content defined in the GDPR.

In the context of the GDPR, legal compliance issues have been raised for a range of emerging technologies including Internet of Things (IoT), artificial intelligence (AI) and big data analytics, and also blockchains. One mostly discussed aspect of the tension between blockchains (especially public blockchain systems) and the GDPR is the following: the immutable nature of blockchains makes it impossible to delete personal information, therefore, it is not possible to exercise the RTBF (more formally known as the right to erasure) of data subjects as defined in the GDPR. Another aspect is about data sharing outside of the EU/EEA/UK: for a public blockchain system, it is normally the case that every node holds a full copy of all data, no matter where the node is physically located or even unknown.

Due to those GDPR-compliance challenges, many researchers looked at the tension between the GDPR and blockchains in recent years and some also attempted to propose solutions to address some of the challenges. In a 2018 report [4], the EU Blockchain Observatory & Forum stated that "Public, permissionless blockchains represent the greatest challenges in terms of GDPR compliance". Despite the active research on this very important topic, to date we have noticed only two systematic literature reviews (SLRs) covering related research progress, both published in 2021. In one of those SLRs, Haque et al. [5] identified 39 papers covering this topic by searching into two databases (IEEE and Scopus), and in the other one Suripeddi and Purandare [6] identified 41 papers for their review by searching into three databases (Science Direct, ACM and IEEE). Neither SLRs are sufficiently comprehensive due to the limited databases and keywords they used and the over-strict inclusion criteria. We also noticed another literature review paper following a different review technique (Levy and Ellis' narrative review of literature methodology), which used a forward and backward search technique to posit a framework for adopting a blockchain that follows the GDPR [7]. This non-systematic literature review also suffers from having a very limited number of papers covered—just 39.

For our SLR, we expanded the databases searched to Scopus, WoS (Web of Science) and Google Scholar, which allowed us to access gray literature as well. Our SLR therefore led to a much more comprehensive coverage with 114 research articles, making it possible to draw a much bigger picture of relevant research work. We also decided to limit our scope to public blockchains only considering the statement in the EU Blockchain Observatory & Forum's 2018 report [4]. This allowed us to focus on blockchain systems with more essential challenges in terms of the GDPR compliance.

Compared with past reviews on the same topic, our SLR makes a number of new contributions due to our larger coverage of related research papers and a more in-depth analysis of the included papers. First of all, we have considered different types of personal data that can be stored and processed on a blockchain and identified both challenges and proposed solutions for each data type. Our findings also cover limitations and consequences of proposed solutions as well as contradicting opinions that will allow our readers to get a better idea about the current state of the art. Secondly, we considered different roles and responsibilities in the blockchain data processing ecosystem, provided perspectives at the network and application levels, and categorised discussions in the research literature accordingly, all of which have been largely overlooked in other literature reviews. Finally, we reviewed the covered research papers by considering a broader scope of GDPR-related elements, which allowed a much more in-depth and precise representation of the literature.

For our SLR, we followed the PRISMA protocol widely used in many disciplines [8]. Our results revealed that the tension between the GDPR and public blockchains has been studied around three main issues: (i) difficulties in exercising data subjects' rights such as the RTBF due to the immutable nature of public blockchains; (ii) difficulties in identifying roles and responsibilities in the public blockchain data processing ecosystem (particularly on the identification of data controllers and data processors); (iii) ambiguities regarding the application of the relevant law(s) due to the distributed nature of blockchains. Our work also led to a better understanding of GDPR-compliance related solutions proposed in the literature, e.g., those around assuring the RTBF using hashing, and the use of smart contracts to manage consent. The results of our SLR can help inform blockchain researchers and developers, policymakers and law markers to consider how to reconcile the tension between public blockchain systems and the GDPR. Note that our results are not limited to the GDPR since many other data protection laws and regulations share similar data protection principles with the GDPR.

The rest of the paper is organised as follows. In Section 2, important background information about the blockchain technology and the GDPR is given. Section 3 explains our research methodology. Our detailed analysis of the covered papers is given in Section 4. Then, we summarise the results into three main areas (GDPR-compliance, proposed solutions, roles and responsibilities) in Section 5. The final section concludes the paper.

2. Background

2.1. Blockchain technology

From a technical perspective, a blockchain is a distributed database that is formed as a chain of data blocks and offers a solution through decentralising storage and processing of data. It was originally introduced for exchanging digital currency as its underlying technology, however, it has been used in many other areas, such as IoT [9], educational systems [10], and healthcare [11]. The main underlying concepts used to build a blockchain can be given as follows: cryptographic hash function, consensus mechanisms, network infrastructure and types of blockchain [12]. In this section, we will briefly explain those concepts to provide a basis for the rest of the paper.

Each block in a blockchain is normally composed of two parts: transactional data and metadata. The metadata typically contains, inter alia, a timestamp, a hash value of the block, and a hash value of the previous block [13]. All hash values are computed using a cryptographically hard one-way function. This allows blocks to be linked to each other to form a chronological database [14]. This very nature of the blockchains results in any modification of data to be detected by other participants of the network as the hash of the next block would not correspond to the data on the modified one [15]. This feature is called "immutability" and leads public blockchains to be regarded as tamper-proof.

As another important feature of public blockchains, a full copy of a distributed database is stored at each node that is part of the blockchain system. Since there are no central authorities, trust is achieved via the distributed storage (i.e., a distributed ledger) and a distributed consensus mechanism [13]. The latter is needed to ensure that different nodes will converge to the same distributed ledger, rather than all nodes produce different ledgers therefore leading to inconsistency in the system. The distributed consensus mechanism determines how new data blocks are added into a blockchain and how all nodes agree on which branch to follow if there are multiple chain branches. There are several consensus algorithms used by different blockchain systems, and Proof of Work (PoW) used by Bitcoin is so far the most widely used one, in which new blocks are added to the chain by nodes who compete against each other by solving a mathematical puzzle (normally defined by a cryptographic hashing function) [14]. The node who first solves the puzzle creates a new block and a longer chain for others to follow. Such nodes are called miners. Miners need to spend a lot of computational power on solving mathematical puzzles and are incentivised by being awarded coins for being the first puzzle solver. Those algorithms are used to confirm consensus of the current state of the ledger and to ensure that all nodes have the same copy.

The blockchain technology also utilises asymmetric cryptosystems, mainly for verifying authenticity of a transaction and its sender and

receiver. Each user in a blockchain network has their own private key and public key. The private key is used by a transaction sender to sign a transaction using a digital signature algorithm, which can then be verified by other users using the sender's public key.

Blockchain systems can be classified into three broad categories: public (permissionless) blockchains, consortium (permissioned) blockchains and private blockchains. Public blockchains are open to anyone and allow any participants to join the network and read, send, or receive data on the blockchain. In contrast, there are constraints on consortium blockchains, and normally write permissions are granted to a preselected set of participants only. When only one participant has such a privilege, we have a private blockchain system.

Finally, smart contracts are another associated technology based on the blockchain technology, which can fully automate self-enacting electronic contracts [16]. They allow a distributed protocol (such as a set of business rules) to be executed and enforced automatically.

2.2. The GDPR

In order to pursue the objective of protection of fundamental rights and to protect personal data of individuals, the GDPR strengthens the protection of individuals' personal data primarily by defining principles and the lawful bases for processing their personal data and also by specifying rights for individuals.

In this section, we will give the definition of a relevant subset of these elements which are important to understand the GDPR compliance issues of public blockchain systems.

2.2.1. Personal data and data subjects

Two concepts, personal data and data subjects, are at the core of the GDPR. The GDPR defines "personal data" in Article 4 as follows:

> "*any information relating to an identified or identifiable natural person ('data subject')*".

Here, the definition of "identifiable natural person" (i.e., "data subject") is given as:

> "*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*".

This definition is expanded under Article 4(1) and it is stated that it is possible to define information that in itself would not be considered personal data but, when combined with other information, can be considered personal data. Pseudonymised data can be given as an example here. The GDPR defines pseudonymisation as

> "*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*".

The ability to identify a person based on additional information in pseudonymisation technique leads pseudonymised data to be considered as personal data. This opinion is based on Recital 26, which states that

> "*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 the research literature, there have been different opinions on whether pseudonymisation could render data anonymous or not. One example of pseudonymous data is encrypted data [17], which is mentioned in Article 32 of the GDPR for ensuring the security of personal data. In its essence, encryption is a mathematical function which uses a secret value (the key) to encode data so that only users with access to that key can read the information. The holder of the key has the ability to re-identify individuals through decryption of that data. It is not denoted as a mandatory technique for the GDPR compliance but given as an essential data protection measure to mitigate the risk of data processing activities and a convenient way for data controllers to demonstrate compliance with the GDPR.

Unlike pseudonymised data, anonymised data is entirely excluded from the GDPR in its Recital 26. Regarding what constitutes anonymised data, Recital 26 defines anonymisation as follows:

> "*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 not an explicit definition, however, an opinion given by the Article 29 Data Protection Working Party [18] in relation to the EU DPD (Data Protection Directive) 1995 [19], the predecessor of the GDPR, is still widely used as a general guidance. The opinion sets a very high standard and requires that the identification must be prevented irreversibly.

2.2.2. Data controllers and data processors

In addition to personal data and data subjects, there are two other very important roles defined in the GDPR: data controllers and data processors. The definition of "data controller" is given in Article 4(7) as follows:

> "*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*".

The definition suggests that the controller is responsible for the processing of personal data, imposing several legal responsibilities for the controller. In Article 4(8), "data processor" is defined as

> "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*".

The personal data used in both definitions can be explained as any personal data related to a data subject.

The first part of the definition of the 'data controller' implies that no one, not even a natural person, is excluded from responsibility when it comes to the processing of personal data. The second part given as "that jointly or alone" deepens the definition to include joint responsibility for the processing of personal data. Finally, the third part given as "determines the purposes and means of the processing of personal data" exemplifies that the one who has the decision-making power, not the factual power over the processing, is considered as the controller.

The definition of the "data processor" clarifies that a controller must exist for a processor to exist. Furthermore, it should be a separate legal entity with regard to the controller to be classified as a (non-controlling) processor. Controllers delegate the task to processors who process data as separate legal entities within the means and purposes of the controller's own agenda.

2.2.3. Data protection principles

The GDPR sets specific criteria for data controllers and processors to assure that personal data is processed in a fair and lawful way. For this goal, its Article 5 sets seven key data protection principles: 1) *lawfulness, fairness and transparency*; 2) *purpose limitation*; 3) *data minimisation*; 4) *accuracy*; 5) *storage limitation*; 6) *integrity and confidentiality (security)*; and 7) *accountability*. For instance, gaining the data subjects' consent is an

example of *lawful processing* since it is a valid ground under Article 6(1)(a) of the GDPR for collecting and processing personal data. Using personal data in a *fair* way refers to not processing the data in a way that is unduly detrimental, unexpected or misleading to the data subject [20]. As its name suggests, *transparency* requires to be clear, open and honest to people from the beginning about how their personal data is being processed. The second principle, *purpose limitation*, requires that personal data be "collected for specified, explicit and legitimate purposes and not further processed". The principle of *data minimisation* is given in Clause 1(c) of Article 5 as

> "*personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed*".

Another related principle, *storage limitation*, requires that the period for which the personal data is stored is limited to a very strict minimum. It should not be longer than it is necessary for the purposes for which the personal data are processed, and data controllers must delete personal data when it is no longer needed. The principle of *accuracy* dictates that data must be kept up to date and inaccurate data must be deleted. The principle of *integrity and confidentiality* ensures that the personal data is processed and stored in a fashion that appropriate security measures are put in place to protect the personal data. Lastly, the GDPR requires a party to exist that is responsible under the principle of *accountability*. This party is expected to take the responsibility for what is done with personal data and to have appropriate measures and records in place to be able to demonstrate the compliance.

2.2.4. Lawful basis for processing

According to the GDPR, it is required to have a valid lawful basis in order to process personal data. Obtaining explicit consent from the data subject for the processing of any personal data is one of the most commonly used bases for lawful processing. Explicit consent implies freely given, specific, informed and unambiguous indication of the data subject's preferences about the processing of their personal data. Article 7 of the GDPR provides three fundamental principles or rules for obtaining consent from the data subjects: controllers are responsible for demonstrating consent that was given, a data subject has the right to withdraw consent at any time, and finally written requests for consent must be clear. Exercising the right to withdraw consent is expected to be as easy as giving consent. Article 22 also notes that the data subject has the right not to be subjected to automated decision making unless this kind of processing is based on the data subject's explicit consent. The controller/processor has to stop all automated processing of the data if an explicit consent is not gathered. However, they can continue such processing if they are able to demonstrate another compelling legitimate ground. Article 6(1)(f) gives controllers and processors a lawful basis for processing where interest of processing outweighs the data subject's rights and freedoms. It states that personal data can be processed without gathering explicit consent when

> "*processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child*".

2.2.5. Data protection by design and by default

The GDPR introduces explicit requirements for the protection of personal data concerning data protection by design and by default in Article 25. Data protection by default requires data controllers to implement technical and organisational measures that are designed to ensure that the processing of personal data meets the GDPR's requirements and otherwise to ensure protection of data subject's rights. Data protection by design requires data protection requirements to be considered in all phases of system development and appropriate measures to be fulfilled to built-in data protection measures.

2.2.6. Data subject's rights

There are a number of rights granted to the data subjects in the GDPR. Both controllers and processors must fulfil certain requirements and duties towards such rights of data subjects to comply with the GDPR.

2.2.6.1. Right to erasure (right to be forgotten—RTBF). The right to erasure (also known as the right to be forgotten—RTBF) mandates that controllers delete data in certain cases. According to Article 17, data subjects are granted with the right to request removing all related personal data. According to Article 6(1), when data is no longer necessary for the purposes for which it was collected, it must be erased. If the processing is based on consent and the data subject withdraws it, actions must be taken to erase the data as long as there is no other ground for processing (Article 7). It is also possible for a data subject to object to processing, and if there is no overriding reason to continue storing, it must be deleted (Article 21). Otherwise, as long as a lawful means for processing exists, the data can continue to be stored.

2.2.6.2. Right to rectification. Under Article 16 of the GDPR, data subjects have rights to make a request to have their inaccurate personal data rectified, or completed if it is incomplete. Here, rectification means that data is updated to be accurate. Thus, this right has close links to the accuracy principle of the GDPR explained before.

2.2.6.3. Right to be informed and right to access. Right to be informed requires data controllers to provide information to the data subjects regarding the processing and storage of their personal data. This right is expanded by the right of access, through which individuals can make access request to their personal data and gain in-depth information regarding the lawfulness of processing and how their personal data is handled.

2.2.6.4. Right to object and automated decision making. The right to object enables data subjects to object to the processing of their personal data in certain circumstances. The controller can continue such processing if and only if they are able to demonstrate a compelling legitimate ground and that their interests of processing outweigh the data subjects' rights and freedoms. Article 22 of the GDPR sets additional rules to protect individuals against automated decision-making that has legal or similarly significant effects on them. Automated decision-making means making a decision solely by automated means without human involvement. Under Article 22, data subjects have the right not to be subject to a decision solely based on automated processing.

2.2.6.5. Right to data portability. The right to data portability allows data subjects to access and move, copy or transfer their personal data easily from one electronic processing system to another in a safe and secure way, without affecting its usability. Under this right, data subjects have the right to request their personal data in a common and easy-to-read computer format or to request that a controller transmits this data directly to another controller.

2.2.6.6. Right to restrict processing. Data subjects have the right to request the restriction or suppression of their personal data in certain situations: if the data subjects contest the accuracy of their personal data; if the processing is unlawful; if the data subject needs them to establish, exercise or defend a legal claim; and finally if data subjects have objected to processing their data.

3. Research methodology

The overall aim of our research is to understand how researchers have studied the tension between public blockchain systems and the GDPR. To achieve this aim, we formulated the following research questions (RQs):

- **RQ1)** What issues public blockchain systems can lead to in relation to data subject's rights and data protection principles provided by the GDPR?
- **RQ2)** What solutions have been proposed in the research literature to address the tension between public blockchain systems and the GDPR?
- **RQ3)** How researchers have considered legal roles and responsibilities of different stakeholders of public blockchain systems, e.g., who should be considered as data controllers and processors in public blockchain systems?

3.1. Identifying data items

To conduct the SLR, we needed to first identify relevant data items – research papers for our study. To this end, we utilised the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) protocol widely used by researchers for SLRs in multiple disciplines [8]. The protocol involved a number of steps as shown in Fig. 1.

The first step is to select databases for searching for relevant research papers. We decided to use three scientific databases: Elsevier's Scopus,[1] Clarivate's Web of Science[2], and Google Scholar.[3] These databases were the most widely used databases with a very comprehensive coverage of research papers collectively. We did not use specific publisher's own databases because they are largely covered by the above three general databases. For all three databases, we used the same search query (note that all searches are case insensitive):

((blockchain* OR Bitcoin OR cryptocurrenc* OR "distributed ledger*") AND (GDPR OR "General Data Protection Regulation"))

For Scopus and WoS, we searched into the metadata, i.e., titles, abstracts and keywords. For Google Scholar, there were only two options for the searches: title and fulltext. When we attempted searching into fulltext, Google Scholar returned too many candidate data items, so we decided to search into titles. Because Google Scholar had a relatively simplistic search syntax, we split the above search query into four subqueries and then merged the results. All searches into the three databases were completed on 21st December 2021.

The initial set of data items returned from multiple searches were merged, which gave us 472 papers in total. Then, the results were deduplicated, leading to 413 papers. After that, we followed the following exclusion and inclusion criteria to screen all candidate papers.

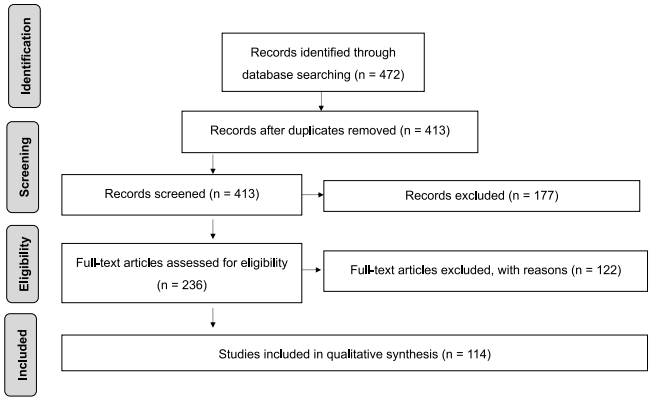


Fig. 1. The diagram of the SLR procedure we used following the PRISMA protocol and the results of different steps of data item identification.

Exclusion Criteria:

- Books, theses, book chapters and other data items that are not research papers were excluded because such items either have not been properly peer reviewed or their full texts are hard to obtain, and many parts of their content are often published as separate research papers.
- Non-English articles were excluded.
- The articles that were not peer reviewed were excluded.
- The articles covering private or consortium blockchains only were excluded.
- The articles in which the GDPR compatibility of public blockchains are covered only in the literature review sections with no further original discussions or research contributions relevant to our RQs were excluded.
- The articles that discuss blockchain or the GDPR in general and lack discussions regarding the GDPR compliance issues were excluded.

Inclusion Criteria:

- The articles that include some discussions on different GDPR-compliance issues of public blockchain systems were included.
- The articles that propose one or more methods to help manage the GDPR compliance of public blockchain systems were included.

The screening process was conducted by the first author, and it involved reading titles and abstracts to exclude papers (leading to 236 papers) and then reading fulltext to make the final selection (leading to 114 papers selected).

3.2. Encoding data items

After obtaining the relevant papers, the first author followed a thematic approach to qualitatively analyse all papers to develop an encoding theme. The encoding process was done using NVivo,[4] one of the most widely used software tools for qualitative analysis. During the qualitative analysis of all papers, the first author identified discussions related to one or more research questions identified for the SLR, and incrementally defined codes to capture such discussions. Generated codes have been reviewed regularly and adjusted where necessary. The encoding scheme was reviewed and validated by the second and third authors, each of whom reviewed 25 randomly selected papers and checked the encoding results. Their feedback was considered by the first author to finalise the encoding scheme and make necessary changes to the encoding results. The last author participated in the general discussion on the encoding scheme and reviewed the final version to approve it.

4. Results and findings

This section describes the results from the SLR.

4.1. General statistics

The distribution of the articles across years can be seen in Fig. 2. As displayed in the figure, the interest into the GDPR compliance issues of public blockchain systems gained pace in 2018 and received the most attention from researchers in 2019 and 2020. As the saturation point has been reached in 2020, a decline in the number of papers has been observed in 2021. There is also one paper published in 2022 because that paper became searchable in December 2021 but was included in a 2022 issue. A majority of the studies were conducted by researchers in computer science and related disciplines, but some were conducted by law

[1] https://www.scopus.com/.
[2] https://www.webofknowledge.com/.
[3] https://scholar.google.com/.
[4] https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/.

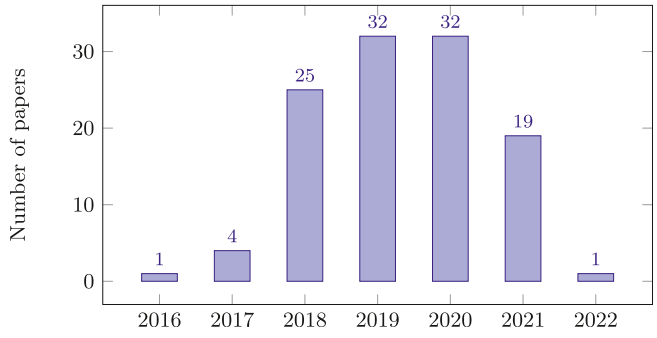


Fig. 2. Distribution of papers per year (2016–2022).

researchers (which is not surprising given the fact that the half of the topic is about data protection law).

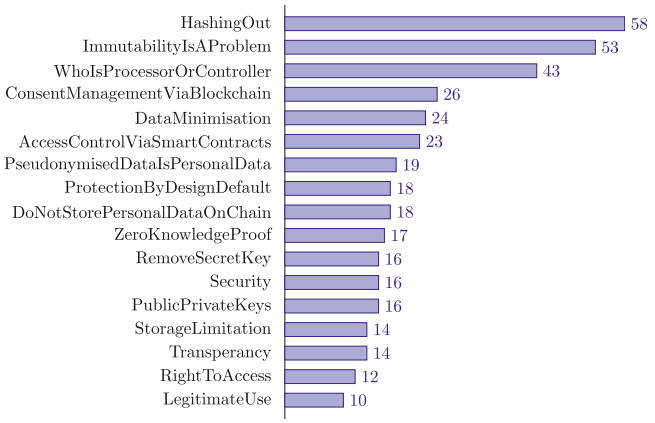


Fig. 3. Most popular codes.

4.2. Encoding scheme

Due to the broad scope of discussions in the literature, we ended up with several categories of codes as seen in the encoding scheme given in Table 1. The codes that received the most attention (the ones that were used in more than 10 articles) can be seen in Fig. 3. The codes are sufficiently self-explanatory so we do not include lengthy explanations to them in Fig. 3. More detailed discussions on all themes and codes are summarised in the following subsections: Section 4.3 covers personal data on blockchains, Section 4.4 covers GDPR-related roles and responsibilities in blockchains, Section 4.5 covers research papers proposing solutions to address the GDPR compliance issues of public blockchain systems, Sections 4.6–4.13 cover different data subject rights defined in the GDPR in the context of public blockchain systems, Section 4.14 covers the first data protection principle on lawfulness, fairness and transparency, Section 4.15 covers other data protection principles, and Section 4.16 covers two other topics: data protection by design and by default, and the territorial scope (i.e., data sharing beyond the EU/EEA/UK). The organisation of the Results and Findings section can be found in Fig. 4 where the themes summarised under each subsection are given to help an interested reader to navigate to the parts of their interest. Due to space limitation, themes are not given for the subsections which do not include any theme other than the ones already given in the titles.

Table 1
Encoding scheme.

Category/Theme	Codes
Personal data in blockchains Roles and responsibilities in blockchains Solutions for protection of personal data in blockchains	PublicPrivateKeys, PersonalDataOfOthers WhoIsProcessorOrController, DataSubjectIsDataController, WhoHasLegalResponsibility ZeroKnowledgeProof, ChamelonHash, RingSignatures, Salting, MerkleTrees, SecureMultiPartComputation, PseudonymisedDataIsPersonalData, PseudonymisedDataIsAnonymizedData, DoNotStorePersonalDataOnChain, RisksOfQuantumComputers, DoNotReusePublicKeys, AnonymizationIsGDPRCompliant, AnonymizationIsNotGDPRCompliant, AnonymizationIsIllegal, SensitiveDataStorageOnChain
Data subject's rights Lawfulness, fairness and transperancy Other principles Other topics	RTBF (ImmutabilityIsAProblem, ImmutabilityIsNotAProblem, HashingOut, RemoveSecretKey, ConsensusToDelete, DisableAccess, Prunning, MainChainSideChain), RightToRectification, RightToBeInformed, RightOfAccess, RightToObject, AutomatedDecisionMaking, RightToDataPortability, RightToRestictProcessing ConsentManagementViaBlockchain, AccessControlViaSmartContracts, LegitimateUse, UseCasesForLegitimateUse, Transperancy, Lawfulness, DataBreaches, DataBreachNotification DataMinimisation, StorageLimitation, Security ProtectionByDesignDefault, TerritorialScope

4.3. Personal data in public blockchain systems

In order to assess whether personal data may be processed legitimately on public blockchain systems, this subsection is dedicated to understanding different types of data on the blockchain and their key components.

4.3.1. Transactional data

Transactional data is the most common data type in all types of blockchain systems. Depending on the underlying use case, the content of a transaction tends to include personal data such as personal identifiers, financial or medical information relating directly or indirectly to data subjects. In case of public blockchain systems that cover smart contracts, executions of smart contract functions are also held in the transactions.

Transactional data can appear in three forms in blockchain systems: plain, encrypted, or hashed. Keeping data in plain text is problematic from a data protection perspective, especially for public blockchain systems. Therefore, it is often the case that some public blockchain systems choose to keep data in an encrypted or hashed form. However, encrypted or hashed data is still personal data, as it falls under the category of pseudonymised data defined by the GDPR as explained before. As mentioned before, the GDPR defines pseudonymisation as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information". Here, the important question is to define this additional information in the blockchain context and identify how the personal data can be revealed. When data is stored in an encrypted form, it can still be decrypted with the correct key, which makes the key the additional information to reveal the personal data.

It is discussed in the literature that there is a linkability risk when data is stored in a hashed form: the possibility to link a particular piece of data and a hash value can still be found, or a hash value might be used to infer personal information, when the same hash value is stored multiple times [21,22]. Therefore, it is not surprising that we observed a general consensus in the literature, which highlights that transactional data pseudonymised via encryption or hash functions should still be considered personal data [22,22–35].

However, even though not many, there are different opinions proposed in the literature. The GDPR makes it clear that pseudonymisation

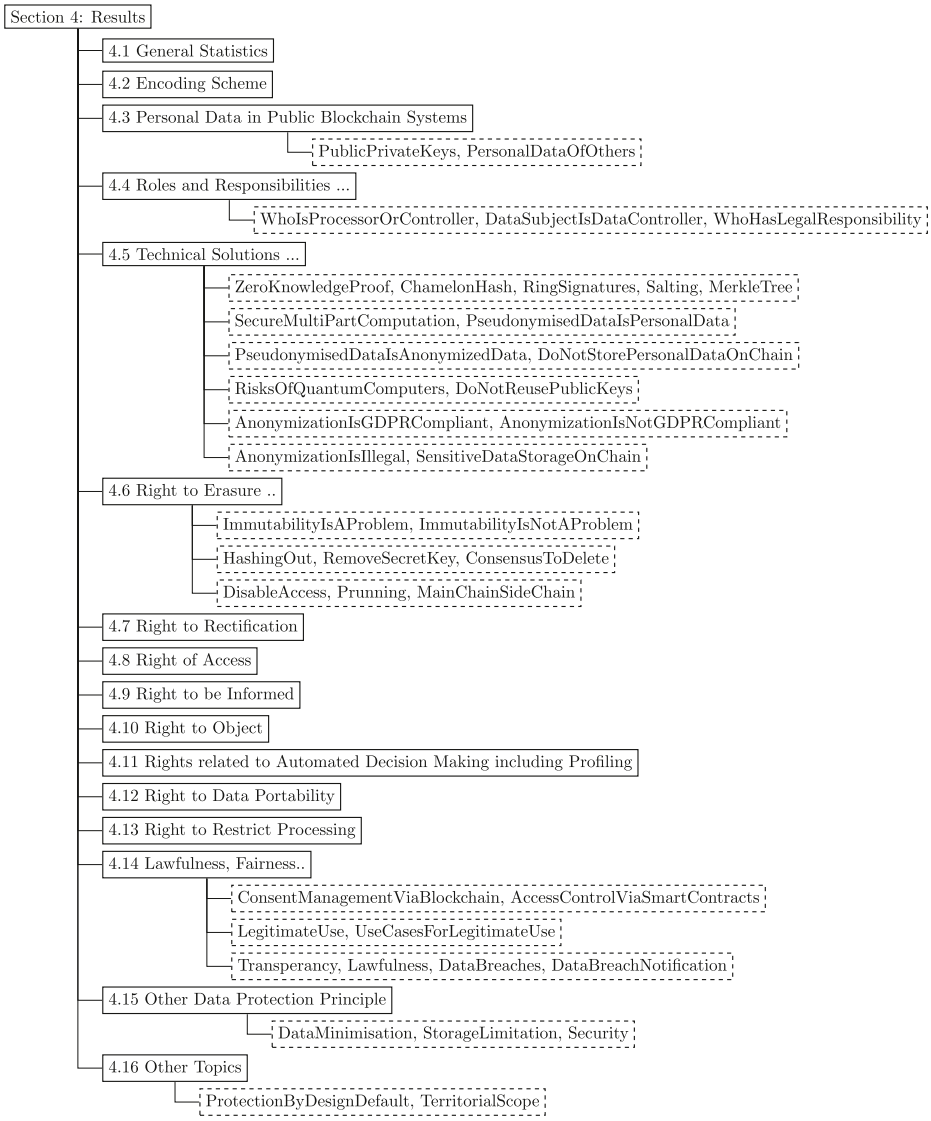


Fig. 4. Organisation of the results and findings.

of data does not equal to anonymisation,[5] however, it does not make the distinction between the two methods very clearly. Recital 26 of the GDPR specifies that data becomes anonymous if it is "reasonably likely" that no identification of the data subject can be derived, which led to different understandings among researchers and even different national data protection authorities. For example, according to the Irish data protection authority (DPA) Data Protection Commission (DPC), the data has to be rendered "irreversibly" anonymous, and the criterion of irreversibility is linked to the absence of reasonable likelihood of identifiability [36]. The French data protection authority CNIL (Commission nationale de l'informatique et des libertés) takes a similar position and acknowledges that anonymisation tends to make identifiability "practically impossible" [37]. However, Spanish DPA (AEPD, Agencia Española de Protección de Datos) provides a more absolute approach regarding hash functions and reported that whether to consider hashed data as anonymised or pseudonymised depends on a variety of factors ranging from the entities involved to the type of the data at hand [36]. Similarly, the UK's DPA Information Commissioner's Office (ICO) once advised on its website (on October 20, 2017) that[6]:

> "*Personal data that has been pseudonymised – eg key-coded – can fall within the scope of the GDPR depending on how difficult it is to attribute the pseudonym to a particular individual.*"

However, this moderate definition seems to have changed as now the ICO's website clearly states that[7]

> "*However, pseudonymisation is effectively only a security measure. It does not change the status of the data as personal data.*"

In the context of the GDPR's predecessor the EU DPD 1995, a similar opinion was provided by Article 29 Data Protection Working Party [18], which acknowledged that even though pseudonymisation reduces the linkability to original identity, it does not eliminate the risk of the data subject being identified, e.g., decrypting an encrypted piece of data via brute-force attack without the decryption key.

We observed various discussions on this topic in papers we covered in the SLR. For instance, it was noted by Erbguth [21] that when only the data subject has the key and nobody else can get hold of it, it is doubtful if the GDPR is meant to protect the data subject from the risk of decrypting

[5] The GDPR, supra note 291, Recital 26.

[6] https://ico.org.uk/media/for-organisations/data-protection-reform/overview-of-the-gdpr-1-13.pdf (Page 4).

[7] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/#pd4.

the data itself. Giordano [23] pointed to the same issue and emphasised that, since the key usually remains in the exclusive ownership of the user himself, there is no intermediary nor central body that has knowledge of the link between the key and its user. Therefore, he proposed that the nature of the information and of data flows affected by the blockchain technology is still far from being defined. In a similar study, Rampone [38] differentiated the two roles and proposed that key-coded data is personal data only for the owner of the list of correspondence that links the codes and the data subject's identities and for those who can reasonably gain possession of it. However, she did not consider key-coded data personal for those who do not have the list of correspondence and are not allowed to have access to it.

In a relevant study, Guggenmos et al. [39] conducted a participatory action based research and they held workshops to pinpoint blockchain systems' GDPR compatibility issues. Their study covered developers, regular stand-ups and management meetings. As a reflection from their legal analysis, they concluded that the prototype under analysis in the study did not comply with the GDPR as the use of an identifier made all data on the blockchain personal data. However, surprisingly, it was added that the legal opinion indicated that a pseudonymisation solution would resolve this problem. Perhaps more surprisingly, Stan and Miclea [40] argued that the health data, which is considered sensitive by the GDPR, could be stored as a hash on a block, without violating the GDPR as the data cannot be returned to its original state and it is therefore sufficiently anonymous. Similarly, Politou et al. [24] suggested to protect sensitive data in the long term by using symmetric algorithms with long key lengths. However, researchers also stated that such a choice would have a severe impact on the storage requirements of the designed blockchain systems. Eichler et al. [41] pointed to off-chain storage solutions in the same context and stated that if the data linking the hashed data to a data subject was kept off-chain and was later erased, the hashed data should once again be considered sufficiently anonymous.

Among those discussions in the literature, the main challenge is given on the fact that data is stored for an *unlimited* period of time in public blockchain systems. Thus, potential future technological development is frequently suggested to be considered when assessing the reliability of current techniques to protect privacy of personal data on blockchains. It is commonly emphasised in the literature that even though it is unlikely for the state-of-the-art methods to link encrypted personal data back to the data subject at the moment, the same cannot be guaranteed in the future due to the rapid advancement in the technology [21,25,27,41,42]. For instance, advancements in quantum computing were mentioned to pose risks to public-key cryptography as available quantum computers may soon be powerful enough to derive private keys used to encrypt personal data today [43].

To summarise, there have been different opinions regarding whether pseudonymisation techniques used by blockchain systems could render data sufficiently anonymous, and there is no observed consensus on what techniques are sufficient to anonymise personal data to the point where the resulting output can potentially be stored in a blockchain system in a GDPR-compliant way.

4.3.2. Metadata

Metadata is another set of data stored in blockchains that may qualify as personal data. Blockchain technologies rely on public-key cryptography where public keys are used for validating transactions. Those keys are essential elements of the metadata and must be publicly available on the blockchain to enable validation of transactions.

We observed a consensus in the literature that public keys serve as the type of identifiers mentioned in Recital 30 of the GDPR, since those keys are often used to identify the origin of transactions, and when associated with other information they constitute personal data [24,26,28,30,37,44–48].

The French DPA CNIL considered the risk of identification of individuals via use of additional information and noted that blockchain applications should implement solutions to ensure that any additional personal data is not stored on the blockchain in clear text [37]. This was stressed with highlighting the fact that public keys are essential to the blockchain's proper functioning and their retention periods are aligned with those of the blockchain's lifetime [37].

Finck [44] explained the potential of public keys to identify individuals with the following examples. 1) If someone uses the blockchain to transfer ownership of a house that will be public due to the nature of blockchain, then, if the person's neighbour would know that such a transfer took place, they could associate the public key to the transfer that was made and link the public key with the house owner. 2) Some users may prefer to share their public keys online intentionally to receive donations, which may link their address to their real-world identities. 3) Additional information that might be gathered in accordance with regulatory requirements, such as where cryptoasset exchanges perform "Know Your Customer" (KYC) and "Anti-Money Laundering" (AML) duties, can lead to disclosures of real-world identities behind the public keys (this scenario was also mentioned in Ref. [22]).

In addition to the above simple scenarios where identification of data subjects behind public keys can happen, more advanced pattern analysis is also given in the literature as a risk. It has been argued that patterns may emerge if the same public key is used by the same natural person in several transactions, which can be used to re-identify them [22,33].

Even though not many, there are researchers who expressed opposite opinions in the context of considering public keys as personal data. Rampone [38] argued that the definition of personal data, albeit in the form of pseudonymous data, given in the GDPR, does not apply to public keys used in blockchain systems. His argument is around public keys being used to solve a technical problem about facilitating trust in a peer-to-peer network and not actually being designed to allow for revealing personal identities. Therefore, he suggested them to be considered neither personal nor pseudonymous data even though they could be used to carry out advanced digital forensic searches to track down the identity of the private key holders. He noted that a public key is not always associated with a natural person and it may be used by a legal entity, which makes equating public keys and pseudonymous data wrong. He also added that, due to the lack of a correspondence list that maps public keys to personal IDs, and the fact that such a correspondence list cannot be easily obtained in normal conditions, public key is nothing but a piece of information indicating a certain credit availability. In payment situations where the debtor and the creditor know each other, it would be a contingent correspondence related only to a given transaction in progress, however, Rampone [38] noted that this could not be extended to other transactions.

With similar arguments, Eichler et al. [41] proposed that public keys were not expected to be personal data in two circumstances: when a key does not belong to a natural person or was not created on behalf of a natural person; and when a key could not be linked to a natural person by reasonable means and is therefore truly anonymous. As done by Rampone [38], highlighting that public keys are unavoidable component of the blockchain technology, Eichler et al. [41] noted that the law must acknowledge a new way to think about public keys.

A more moderate approach was followed by Koscina et al. [49] who recognised public keys as personal data, however, using them in blockchains is interpreted as the maximum minimisation of information (Article 5(c) of the GDPR). Similarly, Giannopoulou and Ferrari [27] also argued that combined with necessary privacy enhancing mechanisms, public keys could fulfil the data minimisation requirements of the GDPR.

In brief, to make a truly GDPR-compliant blockchain system, public keys are one of the biggest challenges as they are an essential component of the blockchain technology and cannot be moved to be off-chain like other data. However, there are techniques provided in the literature for anonymisation of public keys such as ring signatures and zero-knowledge proofs (ZKPs), which will be explained later in Section 4.5.

4.4. Roles and responsibilities in blockchain data processing

One of the most common debates in the literature in the context of GDPR compliance of public blockchain systems is around identification of data controllers and processors. As explained before, the GDPR identifies a data controller as an entity that jointly or alone determines the purposes and means for the processing of the personal data. The key component in this definition is being able to have the decision-making power over the processing. On the other hand, a processor is an entity who processes personal data on behalf of a controller.

For data controllers of public blockchain systems, there are three different opinions among researchers as discussed in the papers covered by our SLR: nodes (lightweight nodes, full nodes, miners, and all nodes as joint controllers), designers/developers, and users. In contrast, for data processors in public blockchain systems, there is more consensus among researchers: users and nodes on the blockchain network. In this subsection, we summarise such discussions and highlight conflicting interpretations.

The majority of the articles covered in our SLR provided a network perspective and are based on the assumption that data subjects are participants of the network and add personal data to the blockchain themselves. However, some other researchers considered the scenario where a user of a blockchain-based application adds personal data to the blockchain on behalf of a data subject. The identification of roles and responsibilities in these two types of scenarios differs greatly, as summarised in the following two subsections from two different perspectives (the blockchain network and application domains).

4.4.1. Perspective of blockchain network

As stated above, a majority of studies covered in our SLR highlighted different roles and responsibilities of different actors who contribute to the functioning of the blockchain network differently. A blockchain network can consist of full nodes (nodes that maintain a full local working copy of the whole blockchain) [50] and lightweight nodes (those who do not download the entire blockchain but only the block headers) [51]. Lightweight nodes rely on full nodes to access the full content of the network. Both lightweight and full nodes can request to create new transactions in the network by broadcasting such request to all nodes. Some full nodes are miners, who write to the blockchain network by investing processing power into solving a cryptographic puzzle so that they can create new blocks and get rewarded for each new block created. A blockchain system uses a distributed consensus protocol to allow miners to create new blocks and for full nodes to jointly decide which branch of a blockchain network will become the main chain. From the GDPR's perspective, this protocol determines the purposes of the blockchain system and means of data processing. The rules in a distributed consensus protocol and a blockchain system are created by developers who are another type of actors responsible for functioning of the network. In some studies, researchers did not differentiate the different types of nodes and stakeholders and argued that the owner of each node should be considered as a joint "controller" of the processing of personal data according to the GDPR [33,52]. However, other researchers recognised the different types of nodes and stakeholders. Their opinions are summarised below.

4.4.1.1. Lightweight nodes. According to Buocz et al. [45], lightweight nodes of the Bitcoin network cannot be considered controllers since they can only request to create transactions and determine only the input address, the output address, and the transferred amount. They argued that lightweight nodes' ability to request creation of transactions would be an excessive interpretation of the term "controller" since they only define "what" (the transactions) but not "why" and "how" of the processing like a real controller should do.

4.4.1.2. Full nodes and miners. Buocz et al. [45] noted the essential contributions of full nodes to the functioning of the blockchain network, however, as they cannot change the protocol by themselves or choose a different protocol within the respective software, they were not considered controllers. However, Jaccard and Tharin [48] argued that when a number of full nodes form more than 50% of all mining power, they should qualify as joint controllers. Considering the activity of full nodes similar to Internet hosting, they also claimed that every full node and perhaps every miner would qualify as a data processor under the GDPR. They further added that following a similar logic, lightweight nodes might also qualify as processors. Responsibilities of the miners were found insufficient as controllers by some researchers due to the lack of their power in determining the purposes or means of the processing [23,41,45,53,54]. Schellekens [54] added that miners unlikely qualify as processors because users neither know the miners nor do they have a contractual relation with them. On the contrary, miners were considered to be data processors by some other researchers [6,22].

It has also been acknowledged by some researchers that, in certain cases, nodes and miners could define their own purposes and set up their own means via accessing the public database stored on the blockchain to collect personal data for commercial purposes, or changing the rules of the blockchain-based platforms by creating a fork in the chain [22]. In such cases, one could argue that nodes and miners would become joint controllers.

Even though the lack of capacity to qualify as a controller dominates research papers covered in our SLR, we also identified some opposite opinions. For instance, Ibáñez et al. [55] argued that miners could be considered as controllers since they determine why and how their own local version of the block is processed. Some other researchers also agreed that every miner on a public blockchain network could qualify as a controller in theory [56,57].

4.4.1.3. All nodes as joint controllers. Holding all nodes responsible is an alternative opinion covered in many papers [25,26,34–36,45,58]. It has been argued that either all nodes collectively – as a partnership – are the controllers within the meaning of Article 4(7) of the GDPR or all individual nodes are joint controllers under Article 26 of the GDPR [25,45]. However, some researchers highlighted that joint controllership requires that controllers, by means of arrangements between them, determine (and thereby divide) their respective obligations, but this does not correspond to nodes in public blockchain systems [46,47]. Campanile et al. [59] added that lawful processing of personal data requires all nodes holding personal data to be known by the users as joint controllers, which is not possible for public blockchain systems.

4.4.1.4. Designers/developers. The discussions regarding responsibilities of designers and developers are even more diverse. While examining the responsibilities of the developers, Buocz et al. [45] started the discussions with the following question: who determines the content of the code ("governance of infrastructure")? This question is important since it defines the responsibility of the data controller(s) in a blockchain system. They highlighted the dynamic nature of open-source project teams where a dynamic group makes proposals and offers inputs to improve the code. They added that, despite this dynamic nature, development and maintenance of the code ultimately rely on a small number of core developers who play a key role in the design of the platform. Schellekens [54] had the view that even though the core developers determine the content of the protocol proposal, they cannot decide whether it will actually be the way that data will be processed within a blockchain network. Therefore, some researchers concluded that developers could not be considered as controllers [27,41,45,54]. Eichler et al. [41] added that the capacity of developers is limited to developing tools and it is up to participants of the blockchain system to decide how those developed tools are used. System administrators, for

instance, are mentioned as an important type of actors who decide whether to adopt the code or not.

Interestingly, Jaccard and Tharin [48] suggested that the first designers should be considered as the first data controllers and be held liable for certain damages and responsible for the respect of certain obligations. Keeping the data protection by design and by default principles in mind, they added that authorities and private parties could be tempted to hold those designers liable as data controllers.

We also observed discussions on the responsibilities of developers of smart contracts. Some researchers considered smart contract developers as processors since they process personal data on behalf of data controllers according to Article 28 of the GDPR [21,32,53]. Ramos and Silva [53] argued that the same applied to miners as they followed the data controller's instructions for checking whether a transaction meets the set technical criteria. Dutta et al. [32] noted that both smart contract developers and smart contracts themselves could also be considered as data processors.

4.4.1.5. Participants with special power. Some researchers claimed that [39,60], data on a blockchain network is pseudonymised, and it only qualifies as personal data to those participants who possess certain additional information that allows attribution of the data to a data subject. Based on that, they argued that only those participants who possess the additional information (e.g., decryption keys) required for attribution qualify as controllers. Related opinions were proposed in Refs. [34,61], where it was reported that in off-chain storage solutions, the party who controls the off-chain storage could be counted as a data controller as they have the power of determining the purposes and means of data processing.

4.4.1.6. Users. Another type of actors, users, were also discussed as controllers by some researchers since they decide what information is included in a transaction, and by this means, determine the details of processing [7,21,30,36,52,53,62–64]. This interpretation raises the interesting point that individual users can be both a data controller and a data subject, which makes many of the data processing requirements unnecessary. It also raises a difficult question [65]: might they be exempt from regulation considering that they are processing data in the course of a purely personal or household activity?

The strongest argument regarding the user's role is that if a user chooses to use a blockchain or blockchain-based application, this makes themself determine the "purposes" and "means" [22,54]. For instance, Duarte [22] claimed that when a user chooses to use a blockchain network even though there are different types of payment methods and different platforms, they determined that the "means" and making a transaction would mean determining the "purpose". On the other hand, Buocz et al. [45] argued exactly the opposite: although a user has the factual power over the processing as they could choose to connect to the blockchain and leave whenever they want, it is unclear if they have the power to determine the means and the purpose of the processing, and thus, they could not be labelled as a controller in a public blockchain system.

Schellekens [54] noted that it could be desirable to designate the user as a joint controller together with the administrators of nodes (and also with core developers of the blockchain system). He explained that this would create a clear addressing point for a data subject seeking to exercise their rights. However, considering the administrators as joint controllers together with the core developers was given as the strongest argument in the same study. Schellekens [54] also noted that even though a user could be considered as a data controller, they may not be able to fulfil the responsibilities of a controller, which would include making binding contracts with processors, exercising the necessary control over the full nodes and deleting data from the blockchain. In this context, Al-Abdullah et al. [7] recommended the use of a contract which would include the terms and conditions to be agreed upon whenever a user, a node or a miner first uses a blockchain system. This approach can help define the use case and then the role of a user as a data controller, a data processor or a data subject.

4.4.1.7. Enforcement of responsibilities. Some researchers also raised concerns on legal responsibilities of controllers. In summary, it was reported that problems of enforcement would remain unclear due to the distributed nature of blockchain systems, and the network lacks identifiable managing partners and clear and transparent allocation of responsibilities [31,44,45,48,66,67]. Holding a collective responsible that does not have any statutory representatives would cause ambiguities for legal enforcement bodies. It is hard to answer how the responsible parties should take care of blockchain security, for instance, in case of a data breach [31]. Teperdjian [31] also emphasised that data subjects require a contact person to exercise their rights such as the right of access, the right to object to the processing of data and to automated processing, however, decentralised and automated blockchain systems have no single point of contact to make these requests.

Buocz et al. [45] argued that legal responsibilities could be allocated to all peers in a blockchain network, however, he also added that identifying the individuals behind the network nodes could be very complicated in practice since they are constantly changing. Due to those difficulties, Tatar et al. [61] recommended imposing the obligation to identify a person or an entity as the representative of the whole users in a given blockchain network prior to joining the system. Wrigley [62] stated that even this would most likely require a significant amount of processor agreements in practice, and in theory, it is certainly not unfeasible to make joint agreements between all participants and parties that run the nodes.

Another concern was given regarding the governing law. Herian [57] stated that, to identify the jurisdiction whose law should be applied, we need to know where a data controller is physically located. This can be difficult to achieve for public blockchain systems.

4.4.1.8. Governmental positions. In addition to researcher's opinions, we observed in the research papers covered in our SLR that, it is useful to compare them with positions of relevant national authorities in the EU/EEA/UK. CNIL, the French DPA, noted in a 2018 report [68] that participants who have a right to write on the chain and who decide to submit data for validation by miners can be considered as data controllers. Therefore, CNIL considers any legal person who registers personal data, on behalf of a natural person, in a blockchain system as a data controller. However, natural persons, outside a professional or commercial activity, are not considered as data "controllers" due to the principle of domestic exception defined in Article 2 of the GDPR. CNIL did not evaluate developers or creators as a whole, but noted that designers of smart contract algorithms may qualify as processors or controllers, depending on their role in determining the purposes. They considered miners as processors and suggested creating a contract between miners and the controller, specifying the obligations of each party and incorporating the provisions of Article 28 of the GDPR. However, CNIL did not consider miners as controllers due to their lack of power on intervening in the purpose of the transactions.

On the other hand, the Hungarian NPA, National Authority for Data Protection and Freedom of Information, adopted a different position and considered each user who adds data to a blockchain as a data controller [69]. EU Blockchain Observatory and Forum, a semi-governmental body in the EU, also highlighted the ambiguities in a 2018 report [4], which stated that in situations where application developers or consortia act as intermediaries between individual users and a blockchain network, they would most likely be considered as data controllers. However, it was also covered in the report that there are cases where it was difficult, and perhaps impossible, to identify a data controller, particularly when blockchain transactions were written by data subjects themselves.

4.4.2. Perspective of application domains

The discussions given in Section 4.4.1 mainly depend on the assumption that the data subject is a participant of the blockchain system and puts personal data of their own (see Use Case 1 in Fig. 7). However, there is another use case, which is more common in some applications, where data subjects use an application where a blockchain system is used as a service. In the latter case, service and application providers that determine the purpose and means of personal data processing are argued to be data controllers [21,30,70]. It is noted that, only when nodes and miners have a more active role in processing the data for their purposes, they qualify as data controllers [7], not while processing data on behalf of user without any impact on the algorithm used.

4.5. Technical solutions for protection of personal data in public blockchain systems

Blockchain systems do not necessarily mean to keep personal data, however, the possibility of personal data being kept in a blockchain system cannot be ruled out. In this subsection, we first discuss data anonymisation methods for blockchains, and then, summarise technical solutions proposed in the literature to protect personal data in those systems so that the blockchain system can be more GDPR-compliant.

4.5.1. Anonymisation

Some researchers investigated the GDPR compliance issues of anonymised data in public blockchain systems. Anonymisation services were considered fully-compliant with the GDPR to protect privacy, however, there were concerns about de-anonymisation attacks and whether the GDPR's threshold for anonymisation is currently reachable on public blockchain systems [34,71,72]. On the other hand, Karasek-Wojciechowicz [34] claimed that if the data is anonymised, then the linkage of data on a public blockchain with a data subject would be impossible for any controller without the use of additional information possessed by that data subject. She considered it practically impossible for data controllers to find the user in possession of additional information needed to identify the data subject.

Another concern regarding anonymised data is the lack of legal certainties and untraceable payment transactions that contradict the KYC and AML laws [33]. Therefore, even though the use of data anonymisation services might solve issues with the GDPR, it will likely raise other legal issues.

It is important to underline that data anonymisation techniques are mainly applied to transactional data, however, as discussed before, personal data in public blockchain systems is not limited to transactions (e.g., public keys can be considered personal data). In the rest of this section, we summarise data protection techniques proposed for public blockchain systems, highlighting the type of personal data that they can protect.

4.5.2. Hashing out

The most common technique proposed for protecting personal data in transactions is the "hashing out" technique. It is achieved by storing hashes of data on-chain and keeping the actual data off-chain by using a local database, eliminating several concerns raised by the distributed nature of public blockchains. This also allows to store more data on the blockchain as the size of hashes is much smaller than that of actual data.

4.5.3. Zero-knowledge proof

Another proposed solution, which received the second most attention from research papers covered in our SLR, is zero-knowledge proof. ZKP is a cryptographic technique used to ensure privacy without damaging transparency [73]. It allows the entire blockchain network to agree on the validity of a transaction without revealing the content of the transaction and is recognised by several researchers as an effective privacy-enhancing technology that can lower the risk of liability for GDPR violations [6,21,26,30,32,41,48,64,74–76]. Schellinger et al. [42] recommended this technique if the verification of important information such as balances, coordinates, or signatures is required. Some researchers also added that this technique should be considered from the very beginning of the development cycle, i.e., it was recommended as a privacy-by-design solution [70,77–79]. This technique was also seen as a solution to comply with the RTBF [80]. Although it is a prominent solution used in many applications, its main drawback was reported as the high computational workload [81].

4.5.4. Merkle trees

Similar to ZKP, Merkle trees were recommended by some researchers to assure data integrity. Schellinger et al. [42] claimed that ZKPs or Merkle trees can be used to achieve a privacy-preserving record of data on the blockchain that does not fall within the scope of the GDPR. Merkle trees were also recommended in the scope of implementing privacy by design [33].

4.5.5. Ring signatures

Another type of cryptographic methods proposed for addressing the GDPR compliance issue is ring signatures. Ring signatures refer to digital signatures performed by a group, each member of which has a private key to sign a given transaction. While it is known that one of those members initiates the signing, it is not possible to know which member it is. Thus, this provides a strong protection to personal data. Therefore, some researchers adopted or recommended the use of ring signatures for GDPR compliance [7,32,48,64,82]. Giannopoulou [79], however, noted that ring signatures are not yet subject to standardisation processes by neither the developer communities nor any formal standardisation bodies. In addition, it also remains unclear if they reach the GDPR required anonymisation threshold.

Like ZKPs, ring signatures also rely on advanced cryptography, which makes it harder to integrate into blockchain protocols [81]. Moreover, the issues on ZKPs regarding high computational workload are also valid for ring signatures [81].

4.5.6. Secure multi-party computation

Secure multi-party computation (SMPC) is another technical solution proposed [48], which aims to provide privacy by allowing multiple parties to perform computations over encrypted data without revealing their input to each other. This enables hiding content of a transaction while still allowing validation of the content. In SMPC, during the processing of personal data, each user's input is split into multiple pieces and distributed randomly to other users. Therefore, each user can only see some meaningless portion of the original data. However, it does not seem to be possible to implement SMPC at scale since it requires a vast amount of system resources like ZKPs and ring signatures [24].

4.5.7. Other solutions

In addition to the solutions summarised above, stealth addresses, one-time keys and adding noise to data are other technical solutions mentioned to hide addresses and transactions [7]. It was also recommended to use a salt in the hash function as a means to reduce the probability of obtaining the input value [26]. Third-party mixing services of public blockchain transactions were also proposed to help users mitigate risks of re-identification [64]. This technique helps mitigate the risk of identification preventing "linkage attacks" to find out connections between transaction inputs and outputs to identify users [64].

Avoiding reusing of the public key was another (less technical) solution suggested as it becomes more difficult to de-anonymise a data subject when a unique public key is used for every transaction [43,48]. This was also suggested in the context of smart contracts [25].

4.5.8. Avoiding blockchains

Despite the variety of solutions proposed, some researchers claimed that storing personal data in blockchains conflicts with the GDPR and should simply be avoided [21,30,39,83–87]. Off-chain solutions are

encouraged by some researchers [39,83–85]. For instance, to achieve the GDPR compatibility, Alessi et al. [35] proposed to develop modules that allow to store personal data in a centralised cloud environment and preserve only business logic in a blockchain network.

There are some domain-specific discussions as well. Kolan et al. [28] argued that personal medical data should not be stored directly on blockchains. Zheng et al. [88] also preferred not to store health information in blockchains in their proposed solution. Similarly, Ma et al. [89] focused on personal data collected and processed by banking systems and provided a data privacy classification for data storage. For such systems, they proposed to allow only public information on chain without any restriction. Most sensitive information is not put on chain as a default setting. For sensitive information owned by customers, they are allowed to put them on chain if they so wish. Finally, sensitive information owned by banks, which is mainly confidential information required for the operation of banks, is given as bank's decision to put on the chain or not.

4.6. Right to erasure (right to be forgotten—RTBF)

Due to the nature of data immutability, public blockchain systems provide a high standard regarding data integrity. However, as mentioned before, this feature is in direct conflict with the data subject's "right to erasure" under the GDPR. This conflict is the most discussed topic in 53 papers (nearly half of all papers covered in this SLR), e.g., in Refs. [21,23,38,39,53,61,70,74,77,81,90–92]. Highlighting the natural outcome that any attempt to change or manipulate data stored in a block would distort the whole blockchain's consistency, some researchers noted that once a system based on the blockchain technology fulfils the request of data erasure, this would be accomplished at the expense of blockchain consistency, which would be detrimental to reliability and trust [61,70].

An ideal scenario given in the papers covered by this SLR is the consensus of all participants of a blockchain on the joint execution of requests to delete personal data from the decentralised ledgers [24,25,70,93,94]. It requires the agreement of a majority of all nodes. For instance, in an opaque blockchain whereby one party has the power over 50% of all nodes, the erasure of data can be feasible: the majority of nodes would erase the data, and all other nodes would subsequently erase the data as well. Similarly, the alteration of personal data can be resolved by changing the stored data for a majority of all nodes. In the case that data processors (not controllers) run nodes, it can be solved by making joint agreements between all participants and parties that run the nodes.

However, it was noted that this scheme adds significant performance overhead [24]. In addition, in public blockchains, no single node can efficiently eliminate a set of personal data requested for erasure or inform the network about such a request. Considering the difficulties in reaching a consensus to delete personal information in public blockchains, researchers proposed several other techniques to overcome this problem. In the following, we summarise those techniques and their consequences according to papers that this SLR covers.

4.6.1. Hashing out

Hashing out or off-chain storage is the most discussed solution for GDPR-compliant processing of personal data in blockchain with the aim of implementing the RTBF. As explained before, it falls under a new class of personal data created by the GDPR, which is pseudonymised data. In common understanding, pseudonymisation does not prevent personal data from being personal, but it gives the organisation more leeway for its processing as the corresponding risks are lower. Researchers proposed to use this technique in two ways to implement RTBF.

Some of the studies suggested erasing any off-chain data once a request to erasure is received, e.g., as proposed in Refs. [26,39,41,43,48,60,63,70,75,77,78,83,91,95–103]. In this type of solutions, off-chain repositories are used to store personal data of users and in the blockchains with only a hash value pointing to the storage location of the personal data stored on the off-chain repository. In this way, once a data subject requests deletion of their personal data, the personal data on the off-chain repository can be deleted, which makes the immutable hashed data pointer stored in blockchain become null and void, and thereby, the system becomes GDPR compliant. A main advantage of this approach is that the information on the blockchain can be used to validate data stored in local repositories [21,76], which helps ensure integrity. Barati et al. [104] proposed to combine this technique with smart contracts and to keep a Boolean value in a smart contract that determines whether the service provided by an actor enables users to erase their data at any time. In some studies, researchers enriched this solution by specifying the type of information that should be stored off-chain. For instance, Walters [64] suggested to store all personally identifiable information off-chain. In another study, Zheng et al. [88] proposed a solution where they preferred off-chain repositories for continuous dynamic data to make it easier to update data over time. They also preferred to hash personal demographic data such as name, address, person identifier, etc. Similarly, Ferrari et al. [30] argued that only transactional data should be recorded in blockchains, and any credential or identifying information should be stored off-chain.

In addition to the first type of solution, some researchers proposed to use the hashing-out technique with encryption to implement RTBF. In this way, a private key is transmitted to the data subject to encrypt the hashed value, and deleting the key is argued to amount to erasure for the purposes of the GDPR. Key destruction allows the service provider to erase the 'linkability' of the blockchain hash pointer to the data located in distributed off-chain repositories [21,39,60,66,105,106]. Unlike outright erasure, the encrypted hash values will still exist on-chain but can only be accessed by the data subject through the exclusive control of the private key. With a single point of storage, it is possible to delete the link between the blockchain and the data storage. Moreover, it gets difficult to apply reverse engineering to restore data from the encrypted hash. It is only important to ensure that the key information required to link hashed data to off-chain data can be shared securely, and deletion can be done reliably [60].

Another approach similar to disabling access to hashed data is storing the encrypted data in blockchains and deleting encryption keys once a data erasure request is received [30,40,48,59,61,106,107]. However, French data protection authority CNIL explained that this approach is not an actual erasure according to the GDPR [68], which was also confirmed by other researchers [34,59]. Karasek-Wojciechowicz [34] noted that if, after key destruction, the controller is still able to link the data stored on the ledger with the natural person, which could be achieved via analysing the content of public ledgers and external data that the controller could access, then, this process could not be considered as erasure under the GDPR. It was also noted in the literature that today's encryption algorithms might no longer be considered secure in the future, which might make it possible to decrypt the data without the knowledge of the original encryption key [41,55]. Another difficulty of this solution was given as managing the decryption keys among many parties that need access to the data [24]. Pagallo et al. [78] also noted that the destruction of data or keys did not eliminate the possibility to re-identify individuals.

Even though hashing out is the most discussed solution to implement RTBF, it was also considered as a "betrayal" to the decentralisation principle of blockchains since a certain degree of control of data remains in the hands of a single centralised party [55]. This leads to reintroduction of a trusted third-party, which would contradict to the motivation behind using blockchains [44,61]. Researchers provide potential solutions to design a GDPR-compliant off-chain storage solution that does not need a trusted third-party. However, off-chain solutions also introduce new vulnerabilities, reversing the benefits of storing data in a blockchain database in an immutable, tamper-proof, secure, and transparent way [61]. Data stored in off-chain repositories is prone to be compromised, and de la Cruz [82] argued duplicating the off-chain data with two different hashes so that if one set of data is compromised, the personal data does not become lost. However, this technique was also argued to introduce complexity and delays [24] and does not solve the

problems completely as transactional metadata saved on the chain may still be considered personal data and so subject to the GDPR [7,24,90]. For instance, when combined with the other information, hashed data may reveal sensitive personal information and can become a target to dictionary attacks [24]. Dissimilar to those opinions, however, Stan and Miclea [40] argued that if all information about a data item is stored only as a hash on chain, the hashes do not violate the GDPR as they are sufficiently anonymous.

4.6.2. Pruning techniques

Pruning is another technique used or suggested to overcome conflicts with the RTBF [44,70,108]. Particularly, old transactions and blocks are deleted after a predefined amount of time, whereas old block headers containing the hashed version of the removed block data are maintained, which ensures the integrity and security of the data [24]. Therefore, pruning techniques were considered by some researchers to serve regulatory requirements, allowing the old transactions to be forgotten from the network [108,109]. In addition, it was argued that it can also offer an increased level of user privacy since old transactions might not be locatable [24].

However, it was also argued that blockchain pruning meets scalability and privacy requirements at the expense of security [44]. Politou et al. [24] added that pruning might add an expensive overhead, leading further inconsistencies and scalability issues when the blockchain's state is verified. On the other hand, there is no guarantee that all nodes would choose not to store the full chain in public blockchains [24]. Dutta et al. [32] focused on state tree pruning and smart contract self-destruct in Ethereum in particular and added that data removal did not depend on participant's demands. The only way to remove code from the blockchain was reported as a contract at that address performing the "self-destruct" operation which leads the storage and code to be removed from the state. However, after this operation, it is still part of the history of the blockchain, and therefore, this process cannot be considered the same as deleting data from a hard disk [32].

4.6.3. Chameleon hashing

Another solution proposed by Ateniese et al. [110] is to use chameleon-hashed blockchains that allow a trusted authority to rectify, amend or overwrite the content of the blockchain. Chameleon-based hash functions work like any other hash functions with the difference that they have a trapdoor that can be used to generate collisions allowing to alter a data item without changing the corresponding hash value of the data, and therefore, being able to maintain the connection to its successor in a blockchain [110]. The knowledge of the trapdoor key allows to find collisions and thus to replace the content of a given block. So, the users among whom the trapdoor key is secretly shared or a centralised authority holding the key can redact the blockchain content in specific and exceptional circumstances. This functionality was listed a potential solution for implementing RTBF by some researchers [24,32,70,78,102].

Al-Abdullah et al. [7] criticised the approach as it defeats the purpose of blockchain by requiring third-parties and/or a centralised authority. Ibáñez et al. [55] also highlighted that adding redactability to an existing blockchain is not possible because the decision for this concept has to be made before a network is set up. Besides, old copies of blockchain would still contain the redacted data [44,110], which makes it highly questionable for this technique to be considered as deletion of personal data under the GDPR. Cutting the immutability of blockchains was also seen as a security risk as it opens an additional door for hackers [102].

4.6.4. Truncated hashing

Lee et al. [111] proposed to use truncated hash values to address the RTBF, which allow to modify transactions by making truncated hash values of modified versions equal to their original target values. A hierarchical multi-blockchain model was used to improve the efficiency of such transaction modifications. It was reported that the method did not sacrifice any of the core benefits of the blockchain technology including transparency, security, and traceability. This method, however, has not been sufficiently scrutinised by other researchers.

4.7. Right to rectification

While many research papers covered in this SLR discussed the RTBF, a much smaller number of papers looked at another highly related right, the right to rectification. Among those, a majority of the studies concluded that technical characteristics of public blockchains are in direct conflict with the right to rectification [5,22,33,112,113]. The main reason for this is that, similar to the issues with deletion, information in public blockchain systems cannot be corrected but changed by adding a new block to the chain. In addition to the immutable nature of blockchains, Al-Abdullah et al. [7] reported one more technical barrier for exercising this right: the difficulty or impossibility of addressing all the full nodes of the network to make necessary updates. In this sense, Duarte [22] added that even if it is possible for a data subject to identify all the nodes or to identify enough nodes (over 50%) to rectify their personal data, coordination among them would be extremely difficult to ensure.

There are a couple of solutions proposed regarding rectification. Al-Abdullah et al. [7] assumed that this right could be granted by providing a supplementary statement, which explains the fact that a transaction could be amended by publishing a new transaction with the new or correct data without the need to delete the previous one completely. Dejanovic et al. [106] designed a process where as the hash value of a new data block is added to the blockchain, the encrypted key for the old data block is deleted and therefore made inaccessible.

4.8. Right of access

As mentioned above, right of access gives data subjects the right to access their personal data held by any service provider subject to comply with the GDPR. Dissimilar to other rights such as the RTBF or the right to rectification, the right of access has been reported to be entirely compatible with the blockchain technology by some studies since the data is available to all members of the network [26,28,33,37,80].

On the other hand, Al-Abdullah et al. [7] argued that since the controllers in a blockchain system only handle the encrypted or hashed versions of data but not the actual form of it, in order to comply with this right, policies and user agreements should be provided to data subjects that explain technical details of how the network functions. Similarly, Duarte [22] stated that it is difficult for nodes to know exactly which data is stored on a blockchain so that they can provide data subjects with information concerning the processing of their personal data.

Providing another perspective, the EU Blockchain Observatory and Forum pointed out that, since the data subjects are not provided with a contact person from whom they could request whether their data were being processed and for what purpose, etc., it is problematic to enforce the right of access in public blockchain systems [4]. Riva [67] was in complete agreement with this opinion. Moreover, Giordano [23] added that even if it is possible to identify the specific node as the data controller, that node might not have the requested information.

4.9. Right to be informed

Even though not many, there are also some concerns regarding the right to be informed which requires data controllers to provide the data subject with information on the period for which their personal data will be stored. Due to the immutable nature of blockchains, this requirement was stated to be difficult to fulfil [113]. However, facilitating the exercise of this right was asserted not to represent a serious issue to organisations by some researchers [28,82]. One solution suggested by de la Cruz [82] is having privacy information on chain and enforcing it to be verified by all participants as an acknowledgement of having read it. She added that this information should include how data subjects could exercise their rights, and who is responsible for dealing with such a request when the occasion arises.

4.10. Right to object

Right to object is one of the GDPR requirements which is not easy to meet for public blockchains due to their permanent nature. However, this right is overlooked in majority of the discussions in the literature, and issues related to immutability are mainly centred around the RTBF.

One study that covers difficulties in exercising the right to object highlighted that knowing the controller of the blockchain is a prerequisite, and therefore, this right might be difficult to comply with [80]. In their proposed study, Daudén-Esmel et al. [103], however, claimed that smart contracts could be used to exercise the right to object as they could be implemented to allow to revoke consent.

4.11. Rights related to automated decision making including profiling

Another theme we identified is about concerns related to automated decision making. The GDPR requires explicit consents of the data subject to allow automated processing of their personal data. Herian [57] asserted that smart contracts engender risks contravening the fundamental rights of the data subject under Article 22 of the GDPR. The EU Blockchain Observatory and Forum also noted that the difficulties in identifying data controller prevented data subjects to oppose to automated decision making [4].

Highlighting the lack of a specific definition of what must be considered to be a decision, Riva [67] argued that it is not clear whether automated data processing management through blockchains falls into the GDPR regime or not. She hypothesised that the solution might be to allow the data subject to make the decision in real time, however, she also noted that this would undermine the benefits of automating the whole set of data processing. This was also agreed by Poelman and Iqbal [80]. Considering data subjects' set-up options as the "human intervention" required by the law was given as a potential solution by Riva [67], however, it was also stated that this might not satisfy the requirement as the human intervention had to occur at the end of the process and could not be general and preventive [67]. Based on this, Riva [67] claimed that some parts of the GDPR are neither up to date nor able to correctly tackle current socio-technological needs.

Unlike other researchers, Ferrari et al. [30] claimed that the protection from automated decision making is not a challenge for blockchain-based automation of transactions as smart contacts generally provided greater auditability and transparency compared to other methods of executing algorithmic-based transactions.

4.12. Right to data portability

Data portability is one of the GDPR requirements that aims to give data subjects more control over data, allowing them to obtain and reuse their personal data for their own purposes across different services [114].

Discussions similar to those related to the right to be informed and the right of access were also found regarding the right to data portability. Researchers considered the exercise of this right as compatible with the technical properties of the blockchain based on the fact that data written to a public blockchain is available to the general public [28,30,33,41]. CNIL agreed that there is no problem with exercising data portability in blockchain systems [68].

However, there are opposite opinions stated. For instance, some researchers noted that this right requires the interoperability between different blockchains, and this does not exist in practice yet due to the lack of standardisation of blockchain systems [48,115]. Giordanengo [116] provided another perspective and asserted that moving data between providers would imply the deletion of data held by the old provider, which was not possible in public blockchains. Finally, the lack of precise and identified data controllers was seen as another barrier for this right as it disallows data subjects to forward their data portability request [67].

4.13. Right to restrict processing

Smart contracts are the only technique identified in the limited number of papers that cover the right to restrict processing. According to Poelman and Iqbal [80], the right to restrict processing could be executed by implementing smart contracts to limit the use of data when necessary. It was noted that the first step for this solution is to establish which nodes have access to personal data. Such a solution was implemented by Daudén-Esmel et al. [103] where the restriction was performed by the consent smart contract, which could limit the personal data that could be collected.

4.14. Lawfulness, fairness and transparency

The decentralised nature of blockchain networks challenges the principle of lawfulness of processing, in particular for the use cases where personal data is processed based on consent [92]. Hereby, there is a considerable amount of research conducted on how to manage consent in public blockchain systems in a lawful way. In this subsection, we provide summary of such studies and another lawful basis "legitimate interest" which was identified as another theme for lawful processing. Due to the very nature of public blockchains, unsurprisingly, transparency is a third theme that emerged as a closely related topic. We put all the three themes together and more discussions are summarised at the end of this subsection.

4.14.1. Consent management

Smart contracts have been proposed by many researchers as an ideal solution to manage consents of the data subjects [21,25,26,78,81,81,95–97,117–122]. The idea behind using smart contracts for consent management is based on the power of translating privacy preferences into automated rules in smart contracts. Then, those contracts can check the validity of a data access request by a third party and allow individuals to verify who can access what part of their personal data [28,43,67,75,123].

This solution has been proposed in different contexts or applications, e.g., for protecting healthcare data [28,121] and financial data [89], two special categories of personal data according to the GDPR. Education is another context identified where contracts can be used to transfer personal data related to educational and professional personal records among educational stakeholders [124]. Consent management problems in online social networks are the focus of another study, where smart contracts were proposed as the essential component of a solution [47].

Even though it has been utilised in different contexts, the strategy behind using smart contracts to manage consents of individuals in a GDPR-compliant way is pretty much the same in all the studies. The aim of all such studies is to prevent undesired data access or to provide proofs for privacy violations. Neisse et al. [125] provided a preventive mechanism, which disallows undesired behaviours of data controllers including misuse or exchange with third-parties. They identified three possible models with different contracts in relation to the number of data subjects and controllers: data subject contract for a specific controller; data subject contract for a specific data item; and controller contract for multiple data subjects. In the first model, privacy preferences (usage control policies) of data subjects are embedded in specific smart contracts deployed in the blockchain for each controller or processor receiving their data. In the second model, smart contracts are created for each data item to be shared with multiple data controllers, allowing control at data item level. Finally, in the third model, each controller expresses their privacy conditions in a smart contract with an interface allowing users to join (give consent) or leave the contract (withdraw consent). The first and the second models were implemented in a follow-up study [119], and after running some experiments, it was reported that for more sensitive data with less frequent exchanges (e.g., medical data), the first model is more adequate. On the other hand, the third model was reported to be ideal for more dynamic data with more frequent exchanges and strict scalability and performance requirements [119].

Barati et al. [104] used smart contracts in a broader context, and in addition to consent management, they translated different GDPR rules into contracts including encryption for preventing unauthorised access to sensitive data, erasure of data upon request and restriction of personal data to be transferred outside EU/EEA/UK. The first contract developed by the researchers is a GDPR-compliant contract and it allows users to identify what operations could be executed on their personal data. Secondly, user consent contract was developed to enable users to give a vote as a consent or negation for the execution of operations already claimed through a GDPR-compliant contract. Via this contract, users can retrieve the corresponding blockchain records and accept or reject the execution of each operation. Two more contracts were provided in the study for submission and verification purposes. Verification contracts aim to identify data privacy violations when an executed operation of an actor does not get user consent or when some personal data processed by the operation is different from those already claimed by an actor via the GDPR-compliant contract [104].

Some other GDPR elements were considered by Barati et al. [126], in another study of them, where compliance with three GDPR obligations for cloud providers, namely data protection, data minimisation, data transfer and data storage, were implemented via smart contracts. In a similar study, Barati and Rana [127] used the blockchain technology again to provide the audit trail of IoT devices under GDPR rules. As done in Ref. [104], those rules are translated into smart contracts to protect personal data in a transparent and automatic way and to facilitate the automatic verification of smart objects whose roles are a data controller or a data processor. The abstract model and business processes proposed in the study were reported to show how the integration of GDPR and blockchain could appear in the design patterns of IoT devices to achieve a greater transparency of privacy [127]. Those solutions were improved in another study where Barati et al. [128] formally examined the verification of GDPR rules on IoT devices at the design time prior to the usage or manipulation of users' personal data. Finally, in their recent study, Barati and Rana [129] provided a reactive mechanism so that the cloud providers who have violated the GDPR rules can be detected via developed smart contracts. Highlighting the increasing number of cloud providers, they noted that even though some of the providers might be directly visible to a user, some others might not be, which was reported to raise data privacy concerns violating the transparency of the data processing. They proposed an architecture based on blockchains and smart contracts to address this requirement.

The privacy of users in the IoT ecosystem was also focused in a study conducted by Rantos et al. [117]. The user-centric solution they proposed allows data subjects to manage their consents regarding their personal data in the IoT ecosystem and to exercise their rights defined by the GDPR. Additionally, the blockchain technology is used to support the consent integrity, non-repudiation and versioning in a publicly verifiable manner. Another personal data sharing management system for IoT was developed by Alessi et al. [35], where smart contracts are proposed for the same purpose. The application was designed to delete data referenced by it from the cloud storage when the user withdraws a consent.

A similar system architecture was proposed by Marikyan et al. [130] to manage agreement between data subjects and data controllers (cloud service providers) before service delivery and any data usage. The verification process managed by the smart contracts involved ensuring that user consent has been obtained and that sharing of data with external cloud providers has been undertaken in a transparent way.

In addition to smart contracts developed to verify consent of the data subjects, Daudén-Esmel et al. [103] provided a purpose smart contract. In their design, once the consent is validated via consent smart contracts, the data controller creates a new purpose smart contract which allows data subjects to decide whether to agree on the processing purpose. The data processor is allowed to request data subject's personal data to process it if this contract is also validated.

Heiss et al. [131] focused on detecting consent violations in a publicly verifiable way and reporting them. Their smart contract based solution was designed to support service providers to fulfil three particular obligations: establishing an auditable archive of consent policies, providing an appropriate technical measure to ensure and to be able to demonstrate the legally valid processing of personally identifiable information, and finally reporting of consent violations to the supervisory authority. Similarly, in the architecture they proposed, de Sousa and Pinto [132] suggested that evidence of the data subject's consent would be stored in the blockchain to enable the regulator to traverse the blockchain whenever a consent needs verifying. In another paper published by the same authors, storing proofs of consent was reported to assure the integrity of persisted consents and evaluations [133]. The advantage of recording user consents and updates in blockchain immutably was covered in another study [134], where making the consents sticky to the data in order to empower data privacy was given as the main contribution of the study. Comparing their solution with traditional methods in text mode, it was noted that their proposal translates consents into switchable right buttons, making it very easy for users to express their consents by switching on/off the button preference [134]. Wirth and Kolain [25] followed a slightly different approach where smart contracts had access to a securely hosted decryption function. In the proposed framework, the data subject is the single source who has the key used for decryption, which enables the data subject to be notified whenever their personal data is accessed.

4.14.2. Legitimate use

Legitimate interests are one of the lawful bases for processing personal data, which gives flexibility to data controllers and processors if personal data is used in expected ways, with a minimal privacy impact, or where there is a compelling justification for the processing [135]. There is a considerable amount of research on legitimate use of a blockchain system, which was generally proposed as a counter-argument of the need to meet the RTBF. As mentioned above, the RTBF is not an absolute request and only applies when data is no longer necessary for the purposes for which it was collected. In the literature, some researchers proposed counter-arguments based on this point. For instance, Daoui et al. [113] argued that when deciding to participate in the blockchain, the data subject would know that their personal data would be processed for the duration of the blockchain, which would mean until the last server is destroyed. They noted that as long as data subjects are informed that the duration period would be infinite and that personal data of each member of the blockchain is necessary for the purpose of data processing, the RTBF would not apply. A very similar opinion was given by Mannan et al. [77], who argued that once users give their consent to have their data permanent on the blockchain, the irreversibility of opt-in mechanism does not conflict with the GDPR.

Another perspective focused on the flexible definition of deletion under the GDPR and it was argued that, considering the functioning principle of blockchains, data stored in blockchains is still necessary for the processing purpose, as those systems are immutable by design [77, 91]. According to these studies, such arguments give blockchains a legal ground for processing personal data permanently.

In the same context, Giordano [23] focused on Clause 2 in Article 17 of the GDPR, which provides the following:

> "*Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.*"

Based on this definition, he argued that since the erasure of personal data is not technically feasible or would require disproportionate efforts, data controllers should not be obliged to a result they could not realistically guarantee. However, he also added that data subjects would still be entitled to receive compensation for any damage they received.

Similarly, Walters [64] argued that a single data subject's privacy request should be overridden by the legitimate interests of all independent users – to ensure network functionality and the sanctity of information. Of course, it was highlighted that this argument could only stand if the design of the public blockchain is such that certain privacy standards are met. Zemler [90], however, highlighted that due to lack of judgements on this field, these legal arguments should be treated with prudent care as these approaches could be considered illegal by a law court in the future.

Protecting public interest and protecting legal interest were argued to be considered to overcome the conflict with the RTBF in another study [136]. With a similar motivation, different use cases were recommended in the literature where permanent storage would have a legal basis. For instance, de la Cruz [82] argued that the use of blockchain-based systems would be convenient where existence of that personal information needs to be proved. Similarly, Dutta et al. [32] recommended the use of blockchain systems for data governance in the areas of transparency and data provenance. Erbguth [21] suggested providing a permanent justification to use a blockchain-based system and provided the publication of election results as an example.

In another paper, Subramanian et al. [137] summarised results from a panel discussion, which include a recommended decision path to determine when to use blockchain technologies considering the privacy issues raised by the GDPR, derived from Pedersen et al.'s work in 2019 [138]. According to the proposed decision path, the following 11 questions should be considered to decide if a blockchain system and which type should be used: 1) Need for a shared common database? 2) Multiple parties involved? 3) Involved parties have conflicting incentives/trust issues? 4) Participants can/want to avoid a trusted third party? 5) Governing rules vary between participants? 6) Rules of transactions remain predominantly constant? 7) Need for an immutable log? 8) Governance allows public network access? 9) Are transactions to be public? 10) Where is consensus determined? and 11) Is off-chain data storage required?

4.14.3. Transparency

The GDPR requires data processing to be clear, open and honest with data subjects from the start about identity of the data controllers, and how and why their personal data is being used. Data processing is transparent in public blockchains, as data is stored in the chain and all participants have the information of the system transactions. Therefore, several researchers claimed that this principle is met by blockchains by design [26,107,112,139]. In addition, according to some of the respondents interviewed in a study conducted by Poelman and Iqbal [80], due to the transparency of the data in public blockchain systems, this technology gives individuals a strong position in regard to the unlawful exchange of data as they provide footprints which could be linked to given consents.

4.15. Other data protection principles

4.15.1. Data minimisation and purpose limitation

Data minimisation is another requirement of the GDPR which could be argued to contradict the nature of blockchains [21,30,33,44,58,140]. While one reason for this is that replication of the data across all nodes in a public blockchain prevents a system to store minimum amount of data [33,112], another reason was given as that the append-only nature of the blockchain systems prevents data to be deleted even though data minimisation principle requires data not to be further processed in a manner that is incompatible with legitimate purposes [44,46,80,99,107].

Even though not many, there are some arguments for the possibility of adapting data minimisation in blockchains. Off-chain solutions were recommended by several researchers as they allow performing modifications and minimisation [7,44,46,77]. In this context, Finck [44] highlighted the difficulty of storing pseudonymous public keys off-chain as they could not be retroactively removed from the ledger and argued that existence of public keys could be a challenge for data minimisation. Unlike Finck, Koscina et al. [49] interpreted storing public keys in blockchains as the maximum minimisation of information and argued that public keys could fulfil the data minimisation requirements of the GDPR.

As another possible solution, Barati et al. [128] drew attention to adapting smart contracts and proposed a solution in which a contract verifies the data minimisation via checking the scope of personal data processed by the system. Pruning is another technique suggested in the literature as it allows removing unused blocks from a blockchain [70]. In addition to smart contracts and pruning techniques, storing data hashes was given as another solution to minimise the processing and storage of personal data [141]. Finally, implementation of anonymisation procedures was suggested to ensure that personal data is stored when it is strictly necessary, and therefore, to comply with data minimisation [82].

There are also some debates in the literature regarding the legitimate use in discussions around data minimisation. For instance, Tatar et al. [61] stated that it is the choice of users what to share in blockchain systems, and therefore, they argued that users could limit the data share to the minimum scope necessary to achieve the specific purpose. Walters [64] provided another counter-argument emphasising that the principle of data minimisation depends on what the "purpose" of the information stored is. He asserted that when the purpose of storing information is to ensure the security and accuracy of the data, the immutable nature of public blockchains might not violate the data minimisation principle [64]. A government land registry was given as a possible valid use case for this purpose.

4.15.2. Storage limitation

Storage limitation requires not to keep personal data longer than it is needed. As expected, this principle stands in tension with immutability, which has been confirmed by multiple researchers [6,23,64,90,92,107,112,140]. The techniques suggested for complying with storage limitation are the same as those for data minimisation. For instance, pruning is one of those approaches, proposed by Moerel [70]. Implementing smart contracts that can keep the total time taken for data processing and the period of time during which personal data will be kept on the storage of an user is another technique followed in another study [104]. Finally, off-chain storage was also given as a potential solution as it allows erasing of personal data when necessary or at a specified time period [99].

4.15.3. Integrity and confidentiality (security)

Despite the variety of concerns reported in the literature regarding conflicts between characteristics of blockchains and different GDPR elements, for security principle, there is almost a consensus among researchers that blockchains offer multiple opportunities as a tool for enhanced security [80,99,140]. First of all, assuring the integrity of personal data was highlighted in several studies as the blockchain technology is generally noted for its resiliency and the absence of a single point of failure [5,59,80,140]. It was highlighted that distributed management and storage of those systems prevent a single point of failure and attacks on such a single point [99]. In addition, cryptography used in blockchains is considered to help to prevent unauthorised parties to alter, delete or steal data [5,59,80,99]. Thus, blockchain-based systems are argued to be preferably used in a GDPR-compliant environment in terms of implementing the security principle.

However, even though not many, we observed a couple of concerns raised in the literature regarding the security principle. Security of the encryption and decryption keys is one of them as highlighted by Erbguth [21]: due to the immutable nature of public blockchains, changing the passwords for the encrypted data on such systems is not possible. Considering the fact that those passwords need to be changeable when there is a risk that a password has been compromised, he argued that putting encrypted sensitive data on a blockchain might violate information security standards.

Another point highlighted in the literature was ambiguities in terms of responsible units for enforcing security policies, the use of appropriate encryption methods and training for users when required [46]. Hereby, it was noted that there might be an increase in the risk of personal data breaches. Foreseeing similar risks, Duarte [22] suggested off-chain repository solutions, arguing that data breaches could be detected more easily since any unauthorised alteration made could be easily spotted. Additionally, Hasselgren et al. [74] focused on data breach notifications required by the GDPR and suggested the use of smart contracts to ensure automated and imitated notifications to relevant authorities upon data breach. This was given as especially beneficial to avoid severe penalties.

Similarly, the confidentiality principle was found not aligned with the nature of public blockchains according to some studies [115,140]. Kusber et al. [115] asserted that confidentiality could only be ensured with off-chain storage of affected data. Besides, they argued that the use of hash algorithms, which can become weak, also deteriorated integrity. However, it is noteworthy that there are also opposite opinions in the literature regarding confidentiality. Particularly, cryptography used in blockchain systems was seen as a tool to support confidentiality by some other researchers [5,87,99,101].

4.16. Other topics

4.16.1. Data protection by design and by default

The GDPR defines requirements for data protection by design and data protection by default in an abstract way, which led researchers to interpret those requirements differently and follow different approaches to meet them. Even though there is a clear understanding that data protection requirements should be considered from the very beginning of the development cycle, a variety of techniques have been argued to be a solution to be inline with data protection by design and by default due to those different concerns and understandings.

Researchers focusing on privacy of personal data mainly recommended cryptographic techniques like zero-knowledge proofs [42,70,77–79]. Off-chain solutions [26] and not storing personal data on-chain were also recommended for the same purpose [70,80]. Another concern is regarding the immutable and distributed nature of blockchains. In this context, solutions were recommended like pruning [70] or limiting ledger storage by storing the entire ledger on one or a few instances only to enable deletion [70]. In addition, decentralised control and distributed storage were seen as a major conflict with privacy by design by some others [64].

There are also studies that argued that public blockchains satisfy those requirements due to its very nature. For instance, immutability and decentralised data storage are stated to ensure the integrity and accuracy of the data, and hence data protection by design and by default [5,61]. It was highlighted that the individuals receive the highest level of control over their personal data via blockchains acquiring the ability to selectively share to any service provider of their choice [112]. Based on this idea, privacy by design and by default was argued to be fulfilled [112]. Encryption techniques and hash functions were also considered to meet those requirements [5,26,35,79,91,141]. However, using hash values and public key cryptography alone was not seen to be able to guarantee privacy by design by some researchers [81].

4.16.2. Territorial scope

According to the territorial scope defined in Article 3 of the GDPR, the regulation applies to personal data processing of any data subjects if the data controller or processor has an establishment in EU/EEA/UK or if the data processing is for delivering a good or service or behavioural monitoring while a subject is in the EU/EEA/UK. Regarding transfer of personal data to outside of EU/EEA/UK, the GDPR requires the receiving country or organisation to provide an adequate level of data protection or the data controllers/processors to provide appropriate safeguard.

The distributed nature of public blockchains allows anyone to join the network as a node and create transactions. Therefore, there is no geographical border to its network, which could pose the issues regarding the localisation and the application of the law. It is almost impossible for data controllers to be aware of where the participants of the blockchain are located and to ensure compliance [142]. Given that, several studies that investigated the GDPR compliance of public blockchain systems reported this as a major conflict [22,30,32–34,57,113,142]. This was also mentioned by CNIL, where the constraints regarding territorial scope were found difficult to implement and permissioned blockchains were recommended instead of public ones [68]. EU Blockchain Observatory and Forum provided an opinion inline with CNIL's, where international transfers of personal data were found problematic for the GDPR compliance.

Even though solutions to overcome this issue are limited compared to other conflicts reported in the literature, a couple of studies proposed some strategies. Al-Abdullah et al. [7] suggested having a detailed data security and protection agreement prepared among participants of the network as a way to address the problem, following mechanisms such as the standard contractual clauses allowed by the European Commission. In some other studies, technical solutions based on smart contracts were provided that controll the country of each actor receiving personal data, and if it is a country that is outside EU/EEA/UK, the algorithm sets the GDPR compliance to false [104,127,129].

5. Further discussions

In this section, we summarise and discuss our key findings for our research questions, together with limitations of past research and open issues identified via the SLR. We also provide our recommendations regarding directions for future research.

5.1. GDPR compliance issues of public blockchain systems (RQ1)

Our first research question (RQ1) aims to investigate the issues that the public blockchain technology can lead to in relation to data subjects' rights and data protection principles provided by the GDPR. Our findings demonstrate that difficulties in exercising the right to erasure (right to be forgotten, RTBF) are the most commonly cited conflict in the literature. It is followed by the right to rectification and the data minimisation principle, which are both at odds with the immutable nature of public blockchains. Finally, territorial scope is another major theme identified.

It should be noticed that the conflicts regarding rights granted to data subjects by the GDPR or the data processing principles are tightly connected to the debates regarding anonymisation and allocation of responsibilities on blockchains. As discussed before, different techniques, including encryption, deletion of encryption keys, and off-chain storage, have been argued to correspond to anonymisation of data even though opponent opinions are more common in the literature. Those discussions are important as once such an assumption could be made where a technique is considered to satisfy the threshold set by the GDPR for anonymisation, the personal data storage in public blockchains via those techniques becomes GDPR-compliant by design. However, we did not identify any consensus regarding the existence of such a technique, and it remains to be seen as the potential advancements in technology make it impossible to claim the current approach to be 100% robust. Advancements in the future cannot be ruled out in those discussions since the data is stored in public blockchains permanently.

In addition, the question about allocation of responsibilities on blockchains is crucial in determining who should comply with the obligations, be held responsible for any violation and to whom the data subjects could reach out for their requests. Difficulties in identifying data controllers and processors are the second most commonly discussed conflict (see Section 5.3). However, it is noteworthy to briefly mention here that this issue was not only discussed in isolation but also the difficulties it causes in exercising rights guaranteed by the GDPR, such as the RTBF, the right to be informed, the right to access, or the right to data portability, were also highlighted by many researchers. Although it is of

critical importance, it is not possible to report a dominant opinion regarding identification of data controllers. Therefore, we can speculate that techniques that will be provided for GDPR compliance of public blockchains will remain insufficient as long as the rules and definitions are not updated in the GDPR to make them clearer for blockchains and other similar decentralised solutions.

As stated at the beginning of this subsection, data minimisation and purpose limitation are the third most common conflict covered in the literature. Two main challenges were highlighted in those discussions: replication of the data across all nodes in a public blockchain; and the append-only nature of the blockchain systems that prevents data to be deleted. Surprisingly, only a few studies discussed public keys in the context of data minimisation, and two opponent opinions were reported, which can be interpreted as an important open issue in the literature. In one of those studies, public keys were considered a challenge for data minimisation [44]. However, some others interpreted the use of public keys as the maximum minimisation of information and argued that they could fulfil the data minimisation requirements of the GDPR [27,49].

Finally, territorial scope is another major theme identified in the literature in the sense of the GDPR compliance issues of public blockchains. The constraints regarding territorial scope defined by the GDPR were found difficult to implement by several researchers due to the lack of geographical border of public decentralised systems. We believe that, similar to the definitions made for data processors and controllers by the GDPR, this is another area that should be handled in a clearer way in the regulations considering the inevitable increase in globalisation of information-centric businesses. Further studies could consider the value in exploring necessary updates for the GDPR to make it more future proof and sustain its value and applicability in to the future.

5.2. Solutions proposed for GDPR compliance (RQ2)

As a response to our second research question (RQ2), our results revealed three major research areas in the literature: techniques proposed to overcome problems regarding deletion of personal data; management of consent and privacy preferences; and techniques to secure data and ensure data privacy.

Regarding the first research area, hashing out is so far the most common technique proposed in the literature with the aim of implementing the RTBF. In this approach, personal data is stored in off-chain repositories and only the hash data pointing to offline storage is stored in the distributed ledgers. This allows off-chain data to be deleted once a request to erasure is received. In the second approach, a private key is transmitted to the data subject which is used to link the blockchain hash pointer to the data located in distributed off-chain repositories, and erasure is accomplished via key destruction. Even though both of the techniques were claimed to be effective in means of assuring RTBF in many studies, they fall short in maintaining the decentralisation principle of blockchains and reintroduce several security problems such as trusted third parties.

In Fig. 5, we summarised major solutions proposed in the literature for addressing the RTBF problem, where red hexagons represent limitations argued for each solution. Here, it is important to underline that none of those solutions can be used to delete metadata in blockchains even though it is recognised as personal data by a majority of the studies, and therefore, falls in the scope of the GDPR. We believe that more research needs to be done to address this open issue.

One interesting detail for this figure is perhaps the case where data subject is data controller. Once this assumption is made, many of the data processing requirements including the RTBF become unnecessary, and therefore, this is added as a solution in Fig. 5.

In the second research area, smart contracts were heavily discussed as a solution to managing consent and controlling privacy preferences. They mainly enable to translate privacy preferences into automated rules which can check the validity of a data access request by a third party and handle it transparently for all parties involved. This approach has two main use cases: preventing unauthorised data access and providing proofs for privacy violations. As this process is fully automated, it allows detection of consent violations in a publicly verifiable way and can demonstrate legally valid processing of personal information. The immutable nature of blockchains also enables to assure the consent integrity, non-repudiation and versioning in a publicly verifiable manner. However, it was also noted in the literature that in an ideal sense, data subjects should be able to obtain human interventions and challenge the decision after the smart contract has been executed, which is not possible via smart contracts [37]. We identified one study in the literature that recognised this problem and proposed to use smart contracts through a securely hosted decryption function that enables the data subject who is the single source with the key used for decryption to be notified whenever their personal data is accessed [25]. Therefore, designing blockchain solutions that can support human interventions could be reported as a potential research area for future studies.

Regarding techniques to assure data security and privacy, there are a variety of techniques proposed with two main goals: assuring data integrity and assuring data privacy or data confidentiality. ZKP is the most commonly cited technique that has been proposed to assure data confidentiality. Even though not receiving as much attention as ZKP, Merkle trees have also been proposed in the literature. Ring signatures, SMPC and tumbling techniques are the other emerging themes identified, all of which can be used to assure data integrity or confidentiality of personal data on the chain. A full list of solutions and their consequences reported in the literature can be seen in Fig. 6.

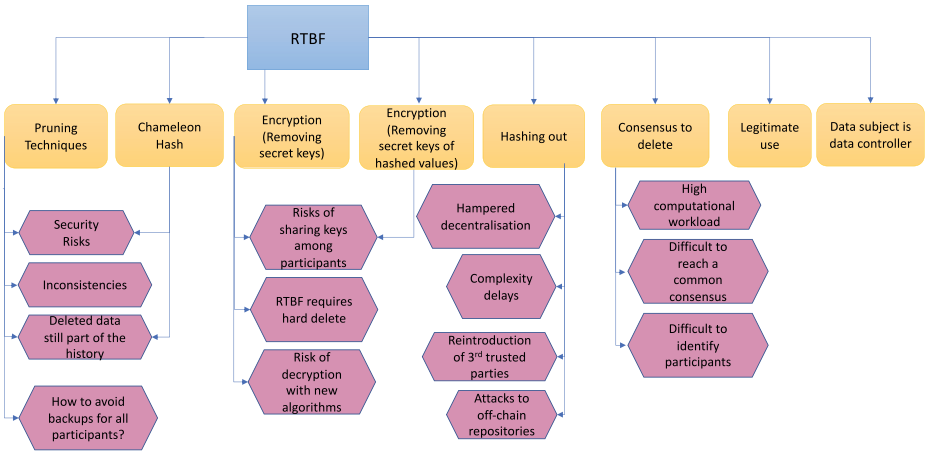


Fig. 5. Solutions proposed for addressing the RTBF.

Similar to the limitations in approaches regarding implementing the RTBF, one research gap in the literature regarding security and privacy of data on blockchains is the low number of solutions that focus on public keys. A majority of the solutions were proposed to assure the GDPR compliance of personal data in transactions, and they tend to overlook public keys, which need attention of researchers in future studies. In Fig. 6, we have categorised the solutions proposed, where yellow boxes represent the solutions for transaction data and green ones represent the techniques that consider the privacy of both transaction data and public keys. Similar to Fig. 5, red hexagons indicate problems raised in the literature for the proposed solutions.

We consider the categorisation presented here a major contribution of this study. It differentiates the techniques based on their focus (transactional data, metadata or both) and demonstrates both the strengths and limitations of the current progress in this research area.

5.3. Legal roles and responsibilities (RQ3)

The third research question (RQ3) aims to investigate how legal roles and responsibilities of different stakeholders of public blockchain networks have been considered in the literature, e.g., who are considered data controllers and processors. This is one of the most common debates in the literature in the context of GDPR compliance of public blockchains. However, it is not possible to report a consensus among the researchers in the sense of identification of data controllers. The ambiguities in this context even yield some researchers to claim that there is no controller in public blockchains, and therefore, the privacy and data protection law is not applicable [143].

However, a majority of the studies covered in our SLR discussed several different entities and their roles and responsibilities. In total, three main categories are discussed as controllers in the literature: nodes (lightweight nodes, full nodes, miners and all nodes as joint controllers), developers, and users. However, none of the categories dominate the discussions. On the other hand, regarding processors, users and nodes on the blockchain are argued to be processors in almost all of the studies.

It is important to highlight here that a vast amount of past research made a over-simplified assumption: personal data is added to the system by their owners (see Use Case 1 in Fig. 7). Past research generally overlooked cases where personal data is added to the system by others on behalf of the data subject, as shown in Use Case 2 in Fig. 7. This largely neglected use case impacts on several aspects including the identification of data controller and the ability of data subjects to exercise their rights given by the GDPR. CNIL slightly covered this use case in one of their reports [68] and explained that a natural person who sells or buys Bitcoin on their own account is not considered as a data controller since he/she is entering personal data outside a professional or commercial activity (in accordance with the principle of domestic exception provided in Article 2 of the GDPR). On the other hand, if they carry out these transactions on behalf of other natural persons inside their professional or commercial activity, the CNIL report states that they may be considered controllers. This definition is unhelpful to understand the use case where a natural person enters information of others in the course of a purely personal or household activity, with no connection to a professional or commercial activity. The second use case helps indicate that researchers should be more comprehensive when considering the GDPR compliance issue of public blockchains since they can change many aspects drastically. In addition to the above two use cases, more could be constructed, especially for hybrid blockchain systems where multiple blockchain systems and non-blockchain systems intermingle, leading to even more complicated relationships and behaviours of different entities in the larger ecosystem. Therefore, we encourage further studies that are more comprehensive to handle different use cases.

Ibáñez et al. [55] put forward the closest approach to cover this variation and discussed GDPR compliance of blockchains in three most common scenarios of how a data subject interacts with a blockchain. They evaluated the role assignment and enumerated applicable strategies for data minimisation and the RTBF and amendments for each scenario individually. Their scenarios are given as follows: an individual interacting directly with a permissionless blockchain; applications that use permissionless blockchains as the backend; and permissioned blockchains. We do not cover the third scenario in this SLR since permissioned blockchains are out of the scope of our work. For the first scenario, Ibáñez et al. [55] argued that it is not possible to identify a controller and the onus of compliance might need to be put on users themselves. For the second scenario, the owners of the intermediary application were argued as data controllers. While this separation is valuable, it falls short of addressing the situation where people put data of others into the blockchain system. This use case reflects a process that completely predates the person being able to give or withdraw consent and exercise other rights. Considering the immutable nature of blockchains, such a use case raises significant privacy concerns. More interestingly, we can also speculate that if a person can or should make irrevocable decisions on their future self's behalf, the importance of the RTBF is echoed if the initial decision turns out to be wrong in the future. We believe that impossibility of exercising the right to withdraw consent or the RTBF leads to major technical, philosophical, legal, ethical, moral and societal concerns and challenges. Limiting the types of data (e.g., disallowing sensitive personal data or illegal materials) or developing solutions that are not harmful for kids, younger people and other vulnerable people could be argued as responsibilities of service providers, who then should assume the role of data controllers. This is another important future direction for interested researchers.

6. Conclusions

We conducted an SLR of 114 research papers to produce a more in-depth and up-to-date understanding of the current research literature on GDPR compatibility of public blockchain systems. Due to a variety of

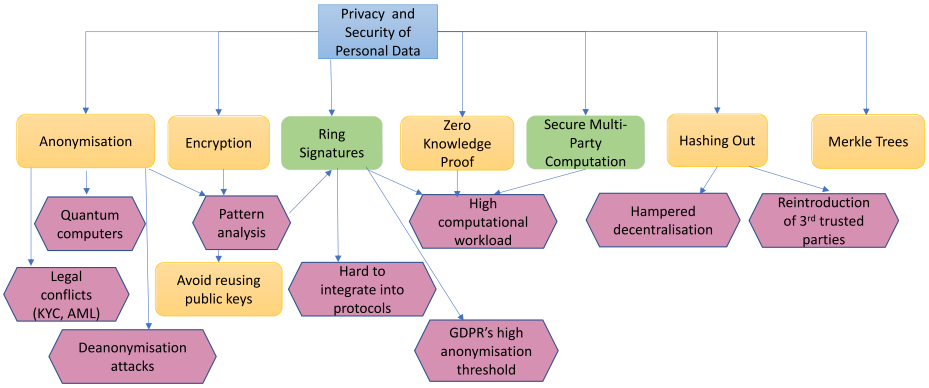
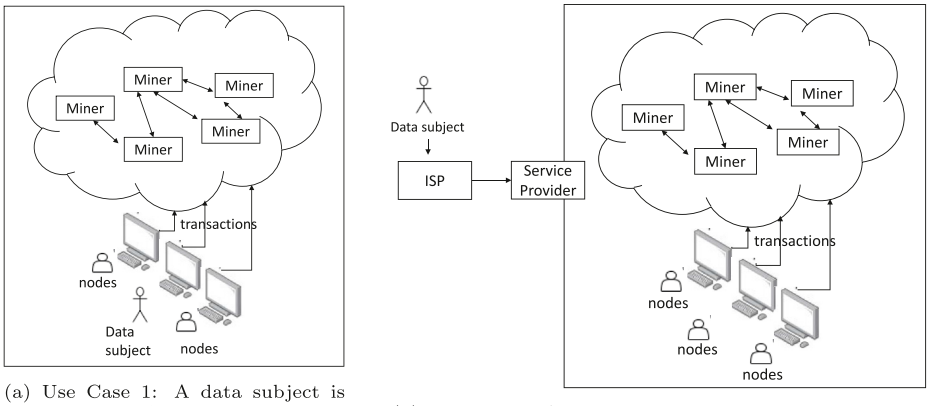


Fig. 6. Solutions proposed to assure data security and privacy.



(a) Use Case 1: A data subject is a participant of the blockchain as a data-adding node and adds their own personal data on chain

(b) Use Case 2: A data subject uses an application where the blockchain system is used as a service and adds their own personal data via the application as a service

Fig. 7. Two different use cases of blockchains in terms of roles and responsibilities in the GDPR context.

challenges discussed and solutions proposed for such challenges, as well as the different interpretations regarding complying issues and overlapping concepts, it is challenging to categorise discussions in the literature in a systematic way. However, our study recommends six broad categories to consider while evaluating the GDPR compatibility of public blockchain systems: categories of personal data in blockchains; roles and responsibilities in blockchains; technical solutions proposed for protection of personal data; discussions regarding how to exercise data subject rights guaranteed by the GDPR; discussions regarding lawfulness, fairness and transparency; and discussions regarding GDPR principles and other related elements.

Our contributions are four-fold. Firstly, the variety of personal data on a blockchain has been recognised in our study, which is generally overlooked in similar past studies, and we identified the challenges and proposed solutions for each category accordingly. Secondly, we identified and reported limitations and consequences of solutions proposed in the literature as well as contradicting opinions, allowing to get a better idea about the current status of the research in this area. Thirdly, we provided perspectives from both the network and application domains and categorised discussions accordingly, which have not been covered comprehensively in past studies. Finally, we include a broad scope of GDPR elements in our study and provided a much deeper and concise representation of the literature, identifying not only the conflicts (negative aspects) but also the compliance (positive) aspects between public blockchains and the GDPR.

While our study was comprehensive in means of identifying studies regarding GDPR compatibility aspects of public blockchain systems, we would like to acknowledge some limitations. Due to limited search facilities, Google Scholar can provide the high number of candidate data items if searched into fulltext, however, due to the large number of items and the low hit rate, we decided to search into titles only, which might have led us to miss some relevant studies. Despite this limitation, as we also used two additional scientific databases (Scopus and WoS), two most widely used databases for SLRs with a very comprehensive coverage of research papers, we believe that the research papers we covered still represent a very good representation of the current research. Secondly, as stated before, we set our scope to be limited to public blockchains only because it is relatively straightforward to comply with the GDPR for private and consortium blockchains. Therefore, our SLR does not cover solutions proposed for GDPR compatibility that are based on or for private and consortium blockchains only. To address the second limitation, a new SLR can be conducted to focus on private and consortium blockchains, and compare the results with what are reported in this paper for public blockchains.

CRediT authorship contribution statement

Rahime Belen-Saglam: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Data curation, Writing – original draft, Writing – review & editing. **Enes Altuncu**: Validation. **Yang Lu**: Validation. **Shujun Li**: Conceptualization, Methodology, Supervision, Funding acquisition, Project administration, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors' work was partly supported by the research project, PRIvacy-aware personal data management and Value Enhancement for Leisure Travellers (PriVELT, https://privelt.ac.uk/), funded by the EPSRC (Engineering and Physical Sciences Research Council, part of the UKRI – UK Research and Innovation), under the grant number EP/R033749/1.

References

[1] D. Puthal, N. Malik, S.P. Mohanty, E. Kougianos, C. Yang, The blockchain as a decentralized security framework [future directions], IEEE Consumer Electronics Magazine 7 (2) (2018) 18–21, https://doi.org/10.1109/MCE.2017.2776459.

[2] European Parliament, Regulation (EU) (2016) 2016/679 of the European Parliament and of the Council of 27 April 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), Off. J. Eur. Union 59 (2016). L 119), URL, https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[3] Information Commissioner's Office (ICO), UK, the UK GDPR, URL, https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/, ????.

[4] T. Lyons, L. Courcelas, K. Timsit, Blockchain and the GDPR: a Thematic Report Prepared by the European Union Blockchain Observatory and Forum, Thematic Report, European Union Blockchain Observatory and Forum, 2018. URL, https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf.

[5] A.B. Haque, A.K.M.N. Islam, S. Hyrynsalmi, B. Naqvi, K. Smolander, GDPR compliant blockchains–A systematic literature review, IEEE Access 9 (2021) 50593–50606, https://doi.org/10.1109/ACCESS.2021.3069877.

[6] M.K.S. Suripeddi, P. Purandare, Blockchain and GDPR – a study on compatibility issues of the distributed ledger technology with GDPR data processing, J. Phys. Conf. 4 (1964), https://doi.org/10.1088/1742-6596/1964/4/042005.

[7] M. Al-Abdullah, I. Alsmadi, R. AlAbdullah, B. Farkas, Designing privacy-friendly data repositories: a framework for a blockchain that follows the GDPR, Digital

Policy, Regulation and Governance 22 (5/6) (2020) 389–411, https://doi.org/10.1108/DPRG-04-2020-0050.
[8] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, J. Clin. Epidemiol. 62 (10), doi:10.1136/bmj.b2700.
[9] S. Huh, S. Cho, S. Kim, Managing IoT devices using blockchain platform, in: Proceedings of the 2017 19th International Conference on Advanced Communication Technology, IEEE, 2017, pp. 464–467, https://doi.org/10.23919/ICACT.2017.7890132.
[10] M. Turkanović, M. Hölbl, K. Košič, M. Heričko, A. Kamišalić, EduCTX: a blockchain-based higher education credit platform, IEEE Access 6 (2018) 5112–5127, https://doi.org/10.1109/ACCESS.2018.2789929.
[11] A. Hasselgren, J.-A. H. Rensaa, K. Kralevska, D. Gligoroski, A. Faxvaag, Blockchain for increased trust in virtual health care: proof-of-concept study, J. Med. Internet Res. 23 (7), doi:10.2196/28496.
[12] M. Raikwar, D. Gligoroski, K. Kralevska, SoK of used cryptography in blockchain, IEEE Access 7 (2019) 148550–148575, https://doi.org/10.1109/ACCESS.2019.2946983.
[13] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang, An overview of blockchain technology: architecture, consensus, and future trends, in: Proceedings of the 2017 IEEE International Congress on Big Data, IEEE, 2017, pp. 557–564, https://doi.org/10.1109/BigDataCongress.2017.85.
[14] A.M. Antonopoulos, Mastering Bitcoin: Unlocking Digital Cryptocurrencies, O'Reilly Media, Inc., 2014. URL, https://www.oreilly.com/library/view/mastering-bitcoin/9781491902639/.
[15] S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system, online document, URL, 2008, https://bitcoin.org/en/bitcoin-paper.
[16] V. Buterin, Ethereum: a next-generation smart contract and decentralized application platform, URL, 2014, https://ethereum.org/en/whitepaper/.
[17] U.K. ICO, What is encryption?, URL, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/encryption/what-is-encryption/, ????.
[18] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, Official Opinion, European Commission, 2014. URL, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.
[19] European Parliament, 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, Official Journal of the European Union No L 281/31 (2015). URL, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31995L0046.
[20] ICO, UK, Principle (a): lawfulness, fairness and transparency, URL, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/ (????).
[21] J. Erbguth, Five ways to GDPR-compliant use of blockchains, Eur. Data Protect. Law Rev. 5 (3) (2019) 427–433, https://doi.org/10.21552/edpl/2019/3/19.
[22] D. G. Duarte, An Introduction to Blockchain Technology from a Legal Perspective and its Tensions with the GDPR, Cyberlaw Journal of the Cyberlaw Research Centre of the University of Lisbon School of Law doi:10.2139/ssrn.3545331.
[23] M.T. Giordano, Blockchain and the GDPR: new challenges for privacy and security, in: Blockchain, Law and Governance, Springer, 2021, pp. 275–286, https://doi.org/10.1007/978-3-030-52722-8_20.
[24] E. Politou, F. Casino, E. Alepis, C. Patsakis, Blockchain mutability: challenges and proposed solutions, IEEE Trans. Emerg. Topic. Comput. 9 (4), doi:10.1109/TETC.2019.2949510.
[25] C. Wirth, M. Kolain, Privacy by blockchain design: a blockchain-enabled GDPR-compliant approach for handling personal data, in: Proceedings of 1st ERCIM Blockchain Workshop 2018, EUSSET, 2018, https://doi.org/10.18420/blockchain2018_03.
[26] F. Molina, G. Betarte, C. Luna, Design principles for constructing GDPR-compliant blockchain solutions, in: Proceedings of the 2021 4th IEEE/ACM International Workshop on Emerging Trends in Software Engineering for Blockchain, IEEE, 2021, https://doi.org/10.1109/WETSEB52558.2021.00008.
[27] A. Giannopoulou, V. Ferrari, Distributed data protection and liability on blockchains, in: Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, Revised Selected Papers, Springer, 2018, pp. 203–211, https://doi.org/10.1007/978-3-030-17705-8_17.
[28] A. Kolan, S. Tjoa, P. Kieseberg, Medical blockchains and privacy in Austria - technical and legal aspects, in: Proceedings of the 2020 International Conference on Software Security and Assurance, IEEE, 2020, https://doi.org/10.1109/ICSSA51305.2020.00009.
[29] S.H. Wilford, N. McBride, L. Brooks, D.O. Eke, S. Akintoye, A. Owoseni, T. Leach, C. Flick, F. Malcolm, M. Stacey, The digital network of networks: regulatory risk and policy challenges of vaccine passports, Eur. J. Risk Regulat. 12 (2) (2021) 393–403, https://doi.org/10.1017/err.2021.35.
[30] V. Ferrari, J.P. Quintais, A. Giannopoulou, B. Bodo, EU Blockchain Observatory and Forum Workshop on GDPR, Data Policy and Compliance, Research Nodes 2018/1, Blockchain & Society Policy Research Lab, Institute for Information Law, University of Amsterdam, 2018. URL, https://hdl.handle.net/11245.1/2c8359fb-8ced-4c17-8168-78e18bc0db53.
[31] R. Teperdjian, The puzzle of squaring blockchain with the general data protection regulation, Jurimetrics J. 60 (3) (2020) 253–313. URL, https://www.americanbar.org/digital-asset-abstract.html/content/dam/aba/publications/Jurimetrics/spring2020/teperdjian.pdf.
[32] R. Dutta, A. Das, A. Dey, S. Bhattacharya, Blockchain vs GDPR in collaborative data governance, in: Cooperative Design, Visualization, and Engineering: 17th International Conference, CDVE 2020, Bangkok, Thailand, Proceedings, Springer, 2020, pp. 81–92, https://doi.org/10.1007/978-3-030-60816-3. October 25–28, 2020.
[33] D. Schmelz, G. Fischer, P. Niemeier, L. Zhu, T. Grechenig, Towards using public blockchain in information-centric networks: challenges imposed by the European Union's general data protection regulation, in: IEEE International Conference on Hot Information-Centric Networking, IEEE, 2018, pp. 223–228, https://doi.org/10.1109/HOTICN.2018.8606000.
[34] I. Karasek-Wojciechowicz, Reconciliation of anti-money laundering instruments and European data protection requirements in permissionless blockchain spaces, J. Cybersecurity. 7 (1), doi:10.1093/cybsec/tyab004.
[35] M. Alessi, A. Camillò, E. Giangreco, M. Matera, S. Pino, D. Storelli, A decentralized personal data store based on Ethereum: towards GDPR compliance, J. Commun. Software.Syst. 15 (2) (2019) 79–88, https://doi.org/10.24138/jcomss.v15i2.696.
[36] A. Giannopoulou, Data protection compliance challenges for self-sovereign identity, in: Blockchain and Applications: 2nd International Congress, Springer, 2020, pp. 91–100, https://doi.org/10.1007/978-3-030-52535-4_10.
[37] F. Martin-Bariteau, Blockchain and the European Union General Data Protection Regulation: the CNIL's Perspective, Working Paper, Blckchn.ca, 2018, https://doi.org/10.2139/ssrn.3275783.
[38] F. Rampone, Data protection in the blockchain environment: GDPR is not a hurdle to permissionless DLT solutions, Ciberspazio Dirit. 19 (61) (2018) 457–478. URL, https://www.mucchieditore.it/index.php?option=com_virtuemart&view=productdetails&virtuemart_product_id=2794.
[39] F. Guggenmos, J. Lockl, A. Rieger, A. Wenninger, G. Fridgen, How to develop a GDPR-compliant blockchain solution for cross-organizational workflow management: evidence from the German asylum procedure, in: Proceedings of the 53rd Hawaii International Conference on System Sciences, University of Hawai'i at M[[TnqNmdEntities]]amacr;[[/TnqNmdEntities]]noa, USA, 2020, pp. 4023–4032. URL, http://hdl.handle.net/10125/64234.
[40] O.P. Stan, L. Miclea, New era for technology in healthcare powered by GDPR and blockchain, in: 6th International Conference on Advancements of Medicine and Health Care through Technology, Springer, Cluj-Napoca, Romania, 2019, pp. 311–317, https://doi.org/10.1007/978-981-13-6207-1_49.
[41] N. Eichler, S. Jongerius, G. McMullen, O. Naegele, L. Steininger, K. Wagner, Blockchain, Data Protection, and the GDPR, Technical Report VR 36105 B 27/661/52176, German Blockchain Association (Bundesblock), 2018. URL, https://www.crowdfundinsider.com/wp-content/uploads/2018/06/GDPR_Position_Paper_v1.0.pdf.
[42] B. Schellinger, N. Urbach, F. Völter, J. Sedlmeir, Yes, I. Do, Marrying blockchain applications with GDPR, in: Proceedings of the 55th Hawaii International Conference on System Sciences, University of Hawai'i at M[[TnqNmdEntities]]amacr;[[/TnqNmdEntities]]noa, USA, 2022, pp. 4631–4640. URL, http://hdl.handle.net/10125/79900.
[43] A. Shahaab, R. Maude, C. Hewage, I. Khan, Managing gender change information on immutable blockchain in context of GDPR, J. Br. Blockchain Assoc. 3 (1) (2020) 23–28, https://doi.org/10.31585/jbba-3-1-(3)2020.
[44] M. Finck, Blockchains and data protection in the European union, Eur. Data Protect. Law Rev. 4 (1) (2018) 17–35, https://doi.org/10.21552/edpl/2018/1/6.
[45] T. Buocz, T. Ehrke-Rabel, E. Hödl, I. Eisenberger, Bitcoin and the GDPR: allocating responsibility in distributed networks, Comput. Law Secur. Rep. 35 (2) (2019) 182–198, https://doi.org/10.1016/j.clsr.2018.12.003.
[46] A.E. Dekhuijzen, Call for action on the EDPB to provide guidance concerning GDPR and blockchain: is public blockchain sustainable under the GDPR? Comp. Law Rev. Int. 20 (2) (2019) 33–36, https://doi.org/10.9785/cri-2019-200202.
[47] J. Ahmed, S. Yildirim, M. Nowostawski, M. Abomhara, R. Ramachandra, O. Elezaj, Towards blockchain-based GDPR-compliant online social networks: challenges, opportunities and way forward, in: Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC) vol. 1, Springer, 2020, pp. 113–129, https://doi.org/10.1007/978-3-030-39445-5_10.
[48] G. Jaccard, A. Tharin, GDPR & blockchain: the Swiss take, jusletter IT, URL, https://lawded.ch/wp-content/uploads/2019/07/Jusletter-IT_gdpr-blockchain-t_5aebbf8be4_en.pdf.
[49] M. Koscina, M. Lombard-Platet, C. Negri-Ribalta, A blockchain-based marketplace platform for circular economy, in: Proceedings of the 36th Annual ACM Symposium on Applied Computing, ACM, 2021, pp. 1746–1749.
[50] Bitcoin Wiki, Full node, URL, https://en.bitcoin.it/wiki/Full_node (????).
[51] Bitcoin Wiki, Lightweight node, URL, 2018, https://en.bitcoin.it/wiki/Lightweight_node.
[52] N. Fabiano, The Internet of Things ecosystem: the blockchain and privacy issues. The challenge for a global privacy standard, in: Proceedings of the 2017 International Conference on Internet of Things for the Global Community, IEEE, 2017, https://doi.org/10.1109/IoTGC.2017.8008970.
[53] L.F.M. Ramos, J.M.C. Silva, Privacy and data protection concerns regarding the use of blockchains in smart cities, in: Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, ACM, 2019, pp. 342–347, https://doi.org/10.1145/3326365.3326410.
[54] M. Schellekens, Conceptualizations of the controller in permissionless blockchains, J. Intellectual Property, Info. Technol. E-Comm. Law. 11 (2) (2020) 215–227. https://www.jipitec.eu/issues/jipitec-11-2-2020/5099.

[55] L.-D. Ibáñez, K. O'Hara, E. Simperl, On blockchains and the general data protection regulation, project report, 2018. https://eprints.soton.ac.uk/422879/.

[56] D. Hofman, V.L. Lemieux, A. Joo, D.A. Batista, The margin between the edge of the world and infinite possibility": blockchain, GDPR and information governance, Record Manag. J. 29 (1/2) (2019) 240–257, https://doi.org/10.1108/RMJ-12-2018-0045.

[57] R. Herian, Regulating disruption: blockchain, GDPR, and questions of data sovereignty, J. Internet Law 22 (2) (2018) 8–16. URL, https://www.proquest.com/docview/2089334432.

[58] N. Fabiano, Internet of Things and blockchain: legal issues and privacy. The challenge for a privacy standard, in: Proceedings of the 2017 IEEE International Conference on Internet of Things, IEEE Green Computing and Communications, IEEE Cyber, Physical and Social Computing, IEEE Smart Data, IEEE, 2017, pp. 727–734, https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.112.

[59] L. Campanile, M. Iacono, A.H. Levis, F. Marulli, M. Mastroianni, Privacy regulations, smart roads, blockchain, and liability insurance: putting technologies to work, IEEE Secur. Privacy. 19 (1) (2020) 34–43, https://doi.org/10.1109/MSEC.2020.3012059.

[60] L. Campanile, P. Cantiello, M. Iacono, F. Marulli, M. Mastroianni, Risk analysis of a GDPR-compliant deletion technique for consortium blockchains based on pseudonymization, in: Computational Science and its Applications – ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part VIII, Springer, 2021, pp. 3–14.

[61] U. Tatar, Y. Gokce, B. Nussbaum, Law versus technology: blockchain, GDPR, and tough tradeoffs, Comput. Law Secur. Rep. 38 (2020), 105454, https://doi.org/10.1016/j.clsr.2020.105454.

[62] S. Wrigley, When people just click": addressing the difficulties of controller/processor agreements online, in: Legal Tech, Smart Contracts and Blockchain, Springer, 2019, pp. 221–252, https://doi.org/10.1007/978-981-13-6086-2_9.

[63] G. Kondova, J. Erbguth, Self-sovereign identity on public blockchains and the GDPR, in: Proceedings of the 35th Annual ACM Symposium on Applied Computing, ACM, 2020, pp. 342–345, https://doi.org/10.1145/3341105.3374066.

[64] N. Walters, Privacy law issues in blockchains: an analysis of PIPEDA, the GDPR, and proposals for compliance, Can. J. Law Technol. 17 (2) (2019) 276–305. https://digitalcommons.schulichlaw.dal.ca/cjlt/vol17/iss2/5/.

[65] C. Millard, Blockchain and law: incompatible codes? Comput. Law Secur. Rep. 34 (4) (2018) 843–846, https://doi.org/10.1016/j.clsr.2018.06.006.

[66] C. Lima, Blockchain GDPR privacy by design: how decentralized blockchain Internet will comply with GDPR data privacy, Tech. Rep. (2018). URL, https://blockchain.ieee.org/images/files/pdf/blockchain-gdpr-privacy-by-design.pdf.

[67] G. M. Riva, What happens in blockchain stays in blockchain. A legal solution to conflicts between digital ledgers and privacy rights, Front. Blockchain. doi: 10.3389/fbloc.2020.00036.

[68] CNIL, Solutions for a responsible use of the blockchain in the context of personal data, Technical Report, URL, 2018, https://www.cnil.fr/en/blockchain-and-gdpr-solutions-responsible-use-blockchain-context-personal-data.

[69] Hungarian National Authority for Data Protection for Data Protection and Freedom of Information, The opinion of the Hungarian national authority for data protection and freedom of information on blockchain technology in the context of data protection, URL, 2018, https://www.naih.hu/files/Blockchain-Opinion-2018-01-29.pdf.

[70] L. Moerel, Blockchain & data protection …and why they are not on a collision course, Eur. Rev. Priv. Law 26 (6) (2018) 825–851, https://doi.org/10.54648/erpl2018057.

[71] A. Biryukov, D. Khovratovich, I. Pustogarov, Deanonymisation of clients in Bitcoin P2P network, in: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2014, pp. 15–29, https://doi.org/10.1145/2660267.2660379.

[72] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G.M. Voelker, S. Savage, A fistful of bitcoins: characterizing payments among men with No names, in: Proceedings of the 2013 Internet Measurement Conference, ACM, 2013, pp. 127–140, https://doi.org/10.1145/2504730.2504747.

[73] M. Quiniou, Blockchain: the Advent of Disintermediation, ISTE Ltd, 2019, https://doi.org/10.1002/9781119629573.

[74] A. Hasselgren, P.K. Wan, M. Horn, K. Kralevska, D. Gligoroski, A. Faxvaag, GDPR Compliance for Blockchain Applications in Healthcare, 2020, https://doi.org/10.48550/arxiv.2009.12913 arXiv:2009.12913 [cs.CR].

[75] F. Molina, G. Betarte, C. Luna, A Blockchain Based and GDPR-Compliant Design of a System for Digital Education Certificates, 2020, https://doi.org/10.48550/arxiv.2010.12980 arXiv:2010.12980 [cs.SE].

[76] C. Damian, I. Lazăr, D.-G. Vişoiu, S. Romanescu, L. Alboaie, Applying blockchain technologies in funding of electrical engineering industry applications, in: Proceedings of the 2019 International Conference on Electromechanical and Energy Systems, IEEE, 2019, https://doi.org/10.1109/SIELMEN.2019.8905896.

[77] R. Mannan, R. Sethuram, L. Younge, Practitioner's corner • GDPR and blockchain: a compliance approach, Eur. Data Protect. Law Rev. 5 (3) (2019) 421–426, https://doi.org/10.21552/edpl/2019/3/18.

[78] U. Pagallo, E. Bassi, M. Crepaldi, M. Durante, Chronicle of a clash foretold: blockchains and the GDPR's right to erasure, in: Legal Knowledge and Information Systems, IOS Press, 2018, pp. 81–90, https://doi.org/10.3233/978-1-61499-935-5-81.

[79] A. Giannopoulou, Putting data protection by design on the blockchain, Eur. Data Protect. Law Rev. 7 (3) (2021) 388–399, https://doi.org/10.21552/edpl/2021/3/7.

[80] M. Poelman, S. Iqbal, Investigating the compliance of the GDPR: processing personal data on a blockchain, in: Proceedings of the IEEE 5th International Conference on Cryptography, Security and Privacy, IEEE, 2021, pp. 38–44, https://doi.org/10.1109/CSP51677.2021.9357590.

[81] S. Schwerin, Blockchain and privacy protection in the case of the European general data protection regulation (GDPR): a delphi study, J. Br. Blockchain Assoc. 1 (1), doi:10.31585/jbba-1-1-(4)2018.

[82] R. de la Cruz, Privacy laws in the blockchain environment, Ann. Emerg. Technol. Comput.(AETiC). 3 (5) (2019) 34–44, https://doi.org/10.33166/AETiC.2019.05.005.

[83] N. D. Sarier, Comments on biometric-based non-transferable credentials and their application in blockchain-based identity management, Comput. Secur. 105, doi:10.1016/j.cose.2021.102243.

[84] N. D. Sarier, Efficient biometric-based identity management on the Blockchain for smart industrial applications, Pervasive Mob. Comput. 71, doi:10.1016/j.pmcj.2020.101322.

[85] J. Peel, The GDPR: the biggest threat to the implementation of blockchain technology in global supply chains, UMKC Law Rev. 88 (2) (2019) 497–517. URL, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf.

[86] M. Zichichi, S. Ferretti, G. D'Angelo, On the efficiency of decentralized file storage for personal information management systems, in: Proceedings of the 2020 IEEE Symposium on Computers and Communications, IEEE, 2020.

[87] J.B. Bernabe, J.L. Canovas, J.L. Hernandez-Ramos, R.T. Moreno, A. Skarmeta, Privacy-preserving solutions for blockchain: review and challenges, IEEE Access 7 (2019) 164908–164940, https://doi.org/10.1109/ACCESS.2019.2950872.

[88] X. Zheng, R.R. Mukkamala, R. Vatrapu, J. Ordieres-Mere, Blockchain-based personal health data sharing system using cloud storage, in: Proceedings of the 2018 IEEE 20th International Conference on E-Health Networking, Applications and Services, IEEE, 2018, https://doi.org/10.1109/HealthCom.2018.8531125.

[89] S. Ma, C. Guo, H. Wang, H. Xiao, B. Xu, H.-N. Dai, S. Cheng, R. Yi, T. Wang, Nudging data privacy management of open banking based on blockchain, in: Proceedings of the 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks, IEEE, 2018, pp. 72–79, https://doi.org/10.1109/I-SPAN.2018.00021.

[90] F. Zemler, Concepts for GDPR-compliant processing of personal data on blockchain: a literature review, Anwendungen und Konzepte der Wirtschaftsinformatik 10 (2019) 96–107. URL, https://ojs-hslu.ch/ojs3211/index.php/akwi/issue/view/10.

[91] M. Berberich, M. Steiner, Practitioner's corner • blockchain technology and the GDPR – how to reconcile privacy and distributed ledgers? Eur. Data Prot. L. Rev. 2 (3) (2016) 422–426, https://doi.org/10.21552/EDPL/2016/3/21.

[92] P. Van Eecke, A.-G. Haie, Practitioner's corner • blockchain and the GDPR: the EU blockchain observatory report, Eur. Data Protect. Law Rev. 4 (4) (2018) 531–534, https://doi.org/10.21552/edpl/2018/4/18.

[93] E. Kadena, P. Holicza, Security issues in the blockchain(ed) world, in: Proceedings of the 18th IEEE International Symposium on Computational Intelligence and Informatics, IEEE, 2018, pp. 211–216, https://doi.org/10.1109/CINTI.2018.8928212.

[94] P. Hillmann, M. Knüpfer, E. Heiland, A. Karcher, Selective deletion in a blockchain, in: Proceedings of the 40th IEEE International Conference on Distributed Computing Systems, IEEE, 2020, pp. 1249–1256, https://doi.org/10.1109/ICDCS47774.2020.00160.

[95] B. Faber, G.C. Michelet, N. Weidmann, R.R. Mukkamala, R. Vatrapu, BPDIMS: a blockchain-based personal data and identity management system, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, University of Hawai'i at M[[TnqNmdEntities]]amacr;[[/TnqNmdEntities]]noa, 2019, pp. 6855–6864. USA, http://hdl.handle.net/10125/60121.

[96] I. Makhdoom, I. Zhou, M. Abolhasan, J. Lipman, W. Ni, PrivySharing: a blockchain-based framework for privacy-preserving and secure data sharing in smart cities, Comput. Secur. 88, doi:10.1016/j.cose.2019.101653.

[97] B. Moslavac, Consent by GDPR vs. Blockchain, revista acadêmica escola superior do ministério público do, Ceará 12 (1) (2020) 149–166. URL, https://www.bib.irb.hr/1076576/download/1076576.ARTIGO-149-166.pdf.

[98] N. Chavez, S. Kendzierskyj, H. Jahankhani, A. Hosseinian, Securing transparency and governance of organ supply chain through blockchain, in: Policing in the Era of AI and Smart Societies, Springer, 2020, pp. 97–118, https://doi.org/10.1007/978-3-030-50613-1_4.

[99] N. Naik, P. Jenkins, Your identity is yours: take back control of Your identity using GDPR compatible self-sovereign identity, in: Proceedings of the 2020 7th International Conference on Behavioural and Social Computing, IEEE, 2020, https://doi.org/10.1109/BESC51023.2020.9348298.

[100] F. Casino, E. Politou, E. Alepis, C. Patsakis, Immutability and decentralized storage: an analysis of emerging threats, IEEE Access 8 (2019) 4737–4744, https://doi.org/10.1109/ACCESS.2019.2962017.

[101] D. Rotondi, M. Saltarella, G. Giordano, F. Pellecchia, Distributed ledger technology and European Union General Data Protection Regulation compliance in a flexible working context, Internet Technol. Lett. 2 (5), doi:10.1002/itl2.127.

[102] C.M. Wulff, The right to be forgotten in post-google Spain case law: an example of legal interpretivism in action? Comparative Law Rev. 26 (1) (2020) 255–279, https://doi.org/10.12775/CLR.2020.010.

[103] C. Daudén-Esmel, J. Castellà-Roca, A. Viejo, J. Domingo-Ferrer, Lightweight blockchain-based platform for GDPR-compliant personal data management, in:

Proceedings of the 2021 IEEE 5th International Conference on Cryptography, Security and Privacy, IEEE, 2021, pp. 68–73, https://doi.org/10.1109/CSP51677.2021.9357602.
[104] M. Barati, I. Petri, O.F. Rana, Developing GDPR compliant user data policies for Internet of Things, in: Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing, ACM, 2019, pp. 133–141, https://doi.org/10.1145/3344341.3368812.
[105] V. Zieglmeier, G.L. Daiqui, GDPR-compliant use of blockchain for secure usage logs, in: Proceedings of EASE 2021: Evaluation and Assessment in Software Engineering, ACM, 2021, pp. 313–320, https://doi.org/10.1145/3463274.3463349.
[106] S. Dejanovic, J. Marjanovic, I. Lendak, A. Erdeljan, Using blockchain to decentralize and protect user privacy in compliance with GDPR, in: ICIST 2019 Proceedings, Information Society of Serbia, 2019, pp. 171–173. URL, http://www.eventiotic.com/eventiotic/files/Papers/URL/a26839e2-6bcd-4194-b481-5cc3e09d892a.pdf.
[107] R. El-Gazzar, K. Stendal, Examining how GDPR challenges emerging technologies, J. Info. Pol. 10 (1) (2020) 237–275, https://doi.org/10.5325/jinfopoli.10.2020.0237.
[108] S. Farshid, A. Reitz, P. Roßbach, Design of a forgetting blockchain: a possible way to accomplish GDPR compatibility, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, University of Hawai'i at M [[TnqNmdEntities]]amacr;[[/TnqNmdEntities]]noa, USA, 2019, pp. 7087–7095. URL, https://hdl.handle.net/10125/60145.
[109] R. Géraud, D. Naccache, R. Roşie, Twisting lattice and graph techniques to compress transactional ledgers, in: Security and Privacy in Communication Systems: 13th International Conference, SecureComm 2017, Niagara Falls, ON, Canada, October 22–25, 2017, Proceedings, Springer, 2017, pp. 108–127, https://doi.org/10.1007/978-3-319-78813-5_6.
[110] G. Ateniese, B. Magri, D. Venturi, E. Andrade, Redactable blockchain – or – rewriting history in Bitcoin and friends, in: Proceedings of the 2017 IEEE European Symposium on Security and Privacy, IEEE, 2017, pp. 111–126, https://doi.org/10.1109/EuroSP.2017.37.
[111] N.-Y. Lee, J. Yang, M.M.H. Onik, C.-S. Kim, Modifiable public blockchains using truncated hashing and sidechains, IEEE Access 7 (2019) 173571–173582, https://doi.org/10.1109/ACCESS.2019.2956628.
[112] W.L. Sim, H.N. Chua, M. Tahir, Blockchain for identity management: the implications to personal data protection, in: Proceedings of the 2019 IEEE Conference on Application, Information and Network Security, IEEE, 2019, pp. 30–35, https://doi.org/10.1109/AINS47559.2019.8968708.
[113] S. Daoui, T. Fleinert-Jensen, M. Lemperiere, GDPR, blockchain and the French data protection authority: many answers but some remaining questions, Stanford J. Blockchain Law. Pol. 2 (2019) 240–251. URL, https://stanford-jblp.pubpub.org/pub/gdpr-blockchain-france.
[114] ICO, UK, Right to data portability, URL https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/, (????).
[115] T. Kusber, S. Schwalm, K. Shamburger, U. Korte, Criteria for trustworthy digital transactions - blockchain/DLT between eI-DAS, GDPR, data and evidence preservation, in: Open Identity Summit 2020, Gesellschaft für Informatik e.V., 2020, pp. 49–60, https://doi.org/10.18420/ois2020_04.
[116] A. Giordanengo, Possible usages of smart contracts (blockchain) in healthcare and why No one is using them, in: MEDINFO 2019: Health and Wellbeing E-Networks for All, IOS Press, 2019, pp. 596–600, https://doi.org/10.3233/SHTI190292.
[117] K. Rantos, G. Drosatos, K. Demertzis, C. Ilioudis, A. Papanikolaou, Blockchain-based consents management for personal data processing in the IoT ecosystem, in: Proceedings of the 15th International Joint Conference on E-Business and Telecommunications, SciTePress, 2018, pp. 572–577, https://doi.org/10.5220/0006911005720577.
[118] C. Kaiser, M. Steger, A. Dorri, A. Festl, A. Stocker, M. Fellmann, S. Kanhere, Towards a privacy-preserving way of vehicle data sharing – a case for blockchain technology?, in: Advanced Microsystems for Automotive Applications 2018: Smart Systems for Clean, Safe and Shared Road Vehicles (Proceedings of the AMAA 2018 Conference) Springer, 2018, pp. 111–122, https://doi.org/10.1007/978-3-319-99762-9_10.
[119] R. Neisse, G. Steri, I.N. Fovino, Blockchain-based identity management and data usage control (extended abstract), in: Privacy and Identity Management. The Smart Revolution: 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, Revised Selected Papers, Springer, 2018, pp. 237–239, https://doi.org/10.1007/978-3-319-92925-5_15. September 4-8, 2017.
[120] B. Sharma, R. Halder, J. Singh, Blockchain-based interoperable healthcare using zero-knowledge proofs and proxy Re-encryption, in: Proceedings of the 2020 12th International Conference on Communication Systems & Networks, IEEE, 2020, https://doi.org/10.1109/COMSNETS48256.2020.9027413.
[121] J. Bowles, T. Webber, E. Blackledge, A. Vermeulen, A blockchain-based healthcare platform for secure personalised data sharing, in: Public Health and Informatics: Proceedings of MIE 2021 vol. 281, IOS Press, 2021, pp. 208–212, https://doi.org/10.3233/SHTI210150.
[122] F. Loukil, C. Ghedira-Guegan, A.-N. Benharkat, PATRIoT: a data sharing platform for IoT using a service-oriented approach based on blockchain, in: Service-Oriented Computing: 18th International Conference, ICSOC 2020, Dubai, United Arab Emirates, December 14–17, 2020, Proceedings, Springer, 2020, pp. 121–129, https://doi.org/10.1007/978-3-030-65310-1_10.
[123] W. Silva, A.C.B. Garcia, Where is our data? A blockchain-based information chain of custody model for privacy improvement, in: Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, IEEE, 2021, pp. 329–334, https://doi.org/10.1109/CSCWD49262.2021.9437727.
[124] A. Alkouz, A. HaiYasien, A. Alarabeyyat, K. Samara, M. Al-Saleh, EPPR: using blockchain for sharing educational records, in: 2019 Sixth HCT Information Technology Trends (ITT), IEEE, 2019, pp. 234–239, https://doi.org/10.1109/ITT48889.2019.9075126.
[125] R. Neisse, G. Steri, I. Nai-Fovino, A blockchain-based approach for data accountability and provenance tracking, in: Proceedings of the 12th International Conference on Availability, Reliability and Security, ACM, 2017, https://doi.org/10.1145/3098954.3098958.
[126] M. Barati, G.S. Aujla, J.T. Llanos, K.A. Duodu, O.F. Rana, M. Carr, R. Rajan, Privacy-aware cloud auditing for GDPR compliance verification in online healthcare, IEEE Trans. Ind. Inf. 18 (7) (2021) 4808–4819, https://doi.org/10.1109/TII.2021.3100152.
[127] M. Barati, O. Rana, Enhancing user privacy in IoT: integration of GDPR and blockchain, in: Blockchain and Trustworthy Systems: First International Conference, BlockSys 2019, Guangzhou, China, December 7–8, 2019, Proceedings, Springer, 2019, pp. 322–335, https://doi.org/10.1007/978-981-15-2777-7_26.
[128] M. Barati, O. Rana, I. Petri, G. Theodorakopoulos, GDPR compliance verification in Internet of Things, IEEE Access 8 (2020) 119697–119709, https://doi.org/10.1109/ACCESS.2020.3005509.
[129] M. Barati, O. Rana, Privacy-aware cloud ecosystems: architecture and performance, Concurrency Comput. Pract. Ex. 33 (23), doi:10.1002/cpe.5852.
[130] D. Marikyan, J. Llanos, M. Barati, G. Aujla, Y. Li, K. Adu-Duodu, S. Tahir, O. Rana, S. Papagiannidis, R. Ranjan, et al., Privacy & cloud services: are we there yet?, in: Proceedings of the 2021 IEEE International Conference on Service-Oriented System Engineering IEEE, 2021, pp. 11–19, https://doi.org/10.1109/SOSE52839.2021.00006.
[131] J. Heiss, M.-R. Ulbricht, J. Eberhardt, Put Your money where Your mouth is – towards blockchain-based consent violation detection, in: Proceedings of the 2020 International Conference on Blockchain and Cryptocurrency, IEEE, 2020, https://doi.org/10.1109/ICBC48266.2020.9169455.
[132] H.R. de Sousa, A. Pinto, On the feasibility of blockchain for online surveys with reputation and informed consent support, in: Ambient Intelligence – Software and Applications –, 9th International Symposium on Ambient Intelligence, Springer, 2018, pp. 314–322, https://doi.org/10.1007/978-3-030-01746-0_37.
[133] H.R. de Sousa, A. Pinto, Blockchain based informed consent with reputation support, in: Blockchain and Applications: International Congress, Springer, 2019, pp. 54–61, https://doi.org/10.1007/978-3-030-23813-1_7.
[134] X. Pei, X. Li, X. Wu, L. Sun, Y. Cao, UDPP: blockchain based open platform as a privacy enabler, in: Proceedings of the 2020 10th Annual Computing and Communication Workshop and Conference, IEEE, 2020, pp. 500–505, https://doi.org/10.1109/CCWC47524.2020.9031142.
[135] ICO, UK, Legitimate interests, URL https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/, (????).
[136] A. Labancz, The conflict of blockchain and the EU general data protection regulation in the area of business law, in: Law 4.0 – Challenges of the Digital Age, Széchenyi István University, 2019, pp. 76–81.
[137] H. Subramanian, K. Cousins, L. B. Bouyad, A. Sheth, D. Conway, Blockchain regulations and decentralized applications: panel report from AMCIS 2018, Commun. Assoc. Inf. Syst. 47 (1), doi:10.17705/1CAIS.04709.
[138] A.B. Pedersen, M. Risius, R. Beck, A ten-step decision path to determine when to use blockchain technologies, MIS Q. Exec. 18 (2) (2019) 99–115. URL, https://aisel.aisnet.org/misqe/vol18/iss2/3/.
[139] N. D'Agostini, GDPR and Blockchain: What does this Mean for In-house Counsel?, Int. In-house Counsel J. 11 (45), URL https://www.iicj.net/subscribersonly/18october/iicj4oct-privacy-NataliaDAgostini-icare.-australia.pdf.
[140] M. Arisi, P. Guarda, Blockchain and eHealth: seeking compliance with the general data protection regulation, BioLaw J.: Rivista di BioDiritto (2) (2020) 477–496, https://doi.org/10.15168/2284-4503-673.
[141] D. Schmelz, K. Pinter, J. Brottrager, P. Niemeier, R. Lamber, T. Grechenig, Securing the rights of data subjects with blockchain technology, in: Proceedings of the 2020 3rd International Conference on Information and Computer Technologies, IEEE, 2020, pp. 284–288, https://doi.org/10.1109/ICICT50521.2020.00050.
[142] F. Zemler, M. Westner, Blockchain and GDPR: application scenarios and compliance requirements, in: Proceedings of the 2019 Portland International Conference on Management of Engineering and Technology, IEEE, 2019, https://doi.org/10.23919/PICMET.2019.8893923.
[143] N. Fabiano, Blockchain and data protection: the value of personal data, J. System. Cybernet. Informat. 6 (6) (2018) 112–115. URL, http://www.iiisci.org/journal/sci/FullText.asp?var=&id=ZA165NO18.