Solutions to Data Protection Challenges in Distributed Ledger and Blockchain Technologies: A Combined Legal and Technical Approach*

Danaja Fabčič Povše[1], Alfredo Favenza[2], Davide Frey[3], Zoltán Ádám Mann[4], Angel Palomares[5], Lorenzo Piatti[6], and Jessica Schroers[7]

[1] Vrije Universiteit Brussel, Belgium - Danaja.Fabcic.Povse@vub.be
[2] LINKS Foundation, Italy - alfredo.favenza@linksfoundation.com
[3] Univ Rennes, Inria, CNRS, IRISA, France - davide.frey@inria.fr
[4] University of Amsterdam, Netherlands - z.a.mann@uva.nl
[5] Atos, Spain - angel.palomares@atos.net
[6] InfoCert, Italy - lorenzo.piatti@infocert.it
[7] KU Leuven, Belgium - Jessica.Schroers@law.kuleuven.be

Abstract. Blockchains and distributed ledgers have attracted increasing attention since the introduction of the Bitcoin blockchain. The ability to run decentralized computations on open networks, on Bitcoin and on the Ethereum Virtual machine has led practitioners and researchers to investigate the use of blockchains and distributed ledgers for a variety of applications that involve the management of personal data.
However, the very characteristics of such distributed ledger technologies (DLTs)—immutability, decentralization, and automation—appear at odds with data protection legislation like the European Union's General Data Protection Regulation (GDPR). This poses significant challenges when designing applications involving personal data. This chapter provides an analysis of possible solutions to these challenges, including results from the literature, proposals for new solutions, and a discussion of challenges that remain despite these solutions. In all cases, solutions require a combination of legal and technical contributions. For example, legal interpretations must take into account the decentralized and general-purpose nature of DLTs, while solutions like mutable ledgers and geographically aware storage may provide answers to some legal concerns.

Keywords: Distributed ledger; DLT; Blockchain; Data protection; Privacy; GDPR

1 Introduction

Since their introduction in 2008 with the Bitcoin cryptocurrency [1], blockchains have found widespread use in a variety of online applications, often dealing

* This chapter was published in: N. El Madhoun, I. Dionysiou, E. Bertin (editors): *Building Cybersecurity Applications with Blockchain and Smart Contracts*, pp. 153-181, Springer, 2024. https://doi.org/10.1007/978-3-031-50733-5_7

with personal data [2, 3]. However, as observed in our companion Chapter [4], the use of blockchain, blockchain-like data structures and, more generally, distributed ledger technologies (DLTs) often comes at odds with the requirements set by data-protection regulations like the General Data Protection Regulation (GDPR) of the European Union (EU) [5].

In our companion chapter [4] we focused on issues and challenges emerging from the application of the GDPR to DLTs. We found that significant challenges arise from three main properties of DLTs: immutability, decentralization, and automation. Immutability directly clashes with principles of the GDPR, like the "right to be forgotten". Decentralization makes it difficult to identify key legal roles such as a data controller, or data processor. Automation clashes with data subject rights, such as the right not to be subject to solely automated processing. Some of these challenges may be more or less critical depending on the type of DLT being considered. For example, identifying the data controller becomes easier in the case of a permissioned DLT managed by a company or set of companies. But many of the challenges remain regardless of the type of DLT.

Given the importance and difficulty of data protection in DLTs, this area has attracted considerable research [6]. In particular, several technical solutions have been proposed to improve data privacy, such as zero knowledge proofs, mixing services, and ring signatures. While these techniques can indeed contribute to better data privacy in certain situations, they do not fully alleviate the legal concerns in relation to data protection regulations. There is a general misalignment between the computer science literature, which proposes sophisticated technical solutions that nevertheless fail to ensure compliance with data protection regulations, and the legal literature, which analyzes the compliance with such regulations in different setups, but without in-depth coverage of the technical possibilities.

This chapter thoroughly discusses the solutions proposed in the literature and provides hints for future research both at a legal and at a technical level. Like in our companion chapter [4], we focus our attention on the GDPR and cover all its relevant provisions, and on the different types of DLTs (public vs. private, permissioned vs. permissionless DLTs), while identifying the solutions to the challenges identified therein. Lastly, this chapter collects some of the most mature and active projects that are trying to leverage DLTs to deliver data processing and transfer use cases, applying some of the solutions we outline on the next pages.

We start by recalling the challenges we identified in our companion chapter [4] in Section 2. Then, Section 3 analyzes potential solutions. Section 4 showcases some specific projects on data protection in DLTs, Section 5 discusses related work, and Section 6 concludes the chapter.

2 Summary of Challenges

We start by summarizing the findings described in our companion chapter [4]. Like there, we structure our discussion based on the three key properties that

we identified as posing challenges for data protection: **immutability**, **decentralization**, and **automation**.

2.1 Challenges resulting from the immutability of DLTs

Immutability is often advertised as what provides the reliability and security of DLTs. Indeed, the ability to maintain tamper-proof information allows DLTs to serve as a reliable log-book for a variety of applications. However, problems start when the information being permanently stored on the ledger turns out to be personal data. Data subject rights like the **right to rectification** (art. 16 GDPR) or the **right to erasure** (art. 17 GDPR) become impossible to guarantee for data that is stored on a ledger. Moreover, it is currently unclear whether some data types like hash values effectively constitute personal data. But if this is the case, their continuous processing during the system's operation raises issues related with the rights *to restriction of processing* (art. 18 GDPR) and **to object** (art. 21 GDPR) as well with principles such as **accuracy** (art. 5 (d) GDPR), **storage limitation** (art. 5 (e) GDPR), **purpose limitation** (art. 5 (b)) and **data minimization**. Finally, immutability also makes it impossible for data subjects to withdraw consent if consent information is stored on the ledger.

2.2 Challenges resulting from the decentralization of DLTs

The GDPR was conceived with a client-server model in mind. It is therefore not surprising that the decentralized nature of DLTs comes at odds with requirements of GDPR such as the identification of a data controller (**principle of accountability** of art. 5(2) and **general responsibility** under art. 24), or the need to identify in which country the data is being stored and processed. In addition, like for immutability, decentralization, and the difficulty in identifying a data controller also interferes with the ability to obtain and manage consent.

2.3 Challenges resulting from the automation in DLTs

With respect to automation, we identified three potential reasons for which DLTs may clash with the GPDR. First, the GPDR stipulates the right not to be subject to fully automated decision-making. Second, automation tools like smart contracts may themselves be used for unlawful purposes. Finally, automated tools like smart contracts may cause the violation of several GDPR principles. These include for example the principle of *lawfulness, fairness, and transparency*, and that of *confidentiality*. Automated tools may hide some of the data from end users, while malicious or erroneous implementations may result in data leaks. Other requirements, like the need for the controller to record processing activities (art. 30 GDPR) or to notify of data breaches (art. 33) may also be vulnerable to malicious and erroneous code.

3 Possible solutions to the challenges

This section discusses how the data protection challenges identified in our previous chapter [4] may be addressed in the context of DLTs. We follow the same structure as in that chapter [4], investigating possible solutions for challenges stemming from immutability, decentralization, and automation in DLTs. We discuss both legal and technical measures for addressing the challenges, paying also attention to the different types of DLTs as well as more complicated constructs based on DLTs, such as smart contracts or Decentralized Autonomous Organizations (DAOs). Table 1 summarizes our findings.

3.1 Immutability

As explained in the previous chapter's section on immutability [4], the immutability of DLTs leads to difficulties with regard to various data protection principles (accuracy, storage limitation, purpose limitation and data minimization) and data subject rights (right to rectification, erasure, restriction of processing, to object). Furthermore, immutability makes it difficult to support the withdrawal of consent and to comply with data protection by design and by default. All these issues result from the fact that, normally, in a DLT it is not possible to change or delete data of past transactions.

To address this challenge, different solutions have been proposed: to keep the personal data off the chain, to use private or mutable DLTs, or to interpret the legislation differently. These different solutions are analyzed below.

3.1.1 Keeping personal data off the chain The usually mentioned solution is to simply keep the personal data off the blockchain. Different variations of this approach have been proposed [7, 8], from not using DLT for personal data, over storing the personal data somewhere else and only including hash pointers on the DLT, to storing the data in encrypted form on the DLT and deleting the encryption keys if necessary. The advantages, disadvantages, and open questions with regard to these approaches are discussed next.

Not using DLT when personal data is processed. Not using DLT to process personal data would on principle solve the issue of compliance with data protection legislation since, as no personal data are processed, data protection legislation does not apply. The open question that remains is whether it is possible to completely refrain from processing personal data with DLT.

Keeping the data "off-chain" in a database, with hash pointers on the DLT. Another possible solution is to segregate personal data in a separate "off-chain" database and keep only a hash of this data on the blockchain as a pointer to the personal data [9]. This makes it possible to erase data in the external storage when someone exercises the right to be forgotten, making the referral information on the blockchain useless. DLT transactions would only contain information needed to access the personal data in the separate database. In this manner, it would be possible to confine personal data to the off-chain

Table 1: Summary of challenges and solutions.

Challenges		Solutions						
		Not using DLT	**Keeping data 'off-chain' in a database, with hash pointers**	**Encryption of data, deletion of encryption keys**	**Using private DLTs**	**Using mutable blockchain-like data structures**	**Using legal scope of interpretation**	**Other technical solutions**
Immutability	**Right to rectification**	☑	~ Unclear status of on-chain hash	~ Data only made inaccessible	~ Requires loose interpretation	☑	~ If accepted by courts	✖
	Right to erasure	☑	☑	☑	~ Requires loose interpretation	☑	~ If accepted by courts	✖
	Right to object	☑	☑	☑	~ Requires loose interpretation	☑	~ If accepted by courts	✖
	Right to withdraw consent	☑	☑	☑	~ Requires loose interpretation	☑	~ If accepted by courts	✖
Decentralization	**Identification of the data controller**	☑	✖	✖	~ Requires loose interpretation	✖	~ If accepted by courts	✖
	Identification of the data transfers	☑	✖	✖	~ Requires loose interpretation	✖	~ If accepted by courts	Sharding
	Consent management	☑	✖	✖	✖	✖	✖	SSI
Automation	**Right not to be subject to fully automated decisions**	☑	✖	✖	✖	✖	~ If accepted by courts	Auditing

☑ Solution solves the challenge

✖ Solution does not help

~ Solution solves the challenge with the specified caveat

storage and avoid storing such data on the DLT [10]. Off-chain storage would enable the modification and erasure of personal data stored off-chain in appropriate databases in line with Articles 16 and 17 GDPR. Yet, as observed by [11] and the CNIL [8], hash pointers are pseudonymous. Moreover, deletion of an item may turn out to be impossible if a non-compliant third party makes a copy of the data being stored before this is deleted.

This solution would resolve the issues with regard to the right to erasure, right to restriction of processing and right to object. It would also allow compliance with the principles of storage limitation, purpose limitation and data minimization.

The right to rectification and with it the principle of accuracy might remain difficult. If a hash of the data is included on the DLT and the data in the database is rectified, the hash will change and will not be the same as on the DLT anymore. This would usually defy the purpose of including the hash in the DLT. Furthermore, as long as the hash is connected to the personal data, it is considered to be personal data itself. on principle, this should be no problem, if it loses this status as soon as the connected personal data are deleted. However, the status of the on-chain hash, after the data has been erased, has not been explicitly stated yet [7, p. 97].

Encrypting the data, deleting the encryption keys. Another proposed solution is to encrypt all the personal data on the blockchain with a key that allows only the associated data subject, and possibly a few authorized parties, to access the data. If a data subject requests their blockchain data to be deleted, the key is deleted, making the information inaccessible. Simple hashing is not sufficient to make data no longer identifiable. Encryption can make data anonymous if the encryption is strong enough and the encryption key is deleted and cannot be restored by anyone. Not deleting the encryption key would mean the data could be decrypted, and hence the individual would be identifiable. Some authors [8, 12] have proposed storing only encrypted personal data on a ledger, and handling erasure requests by throwing away the decryption key. Clearly, this key should not be stored on the DLT, or its deletion would become impossible.

3.1.2 Using private DLTs The use of private or enterprise blockchains represents another possibility for compliance with GDPR's requirements. Private blockchains could limit the dissemination of personal data to just one company or a limited number of companies in a predefined consortium [11]. This would in turn limit the access to sensitive information to only a few individuals, thereby significantly reducing the possibility of data breaches. A relatively loose interpretation of the GDPR may already permit the storage of personal data in permissioned blockchains without risking violations. A first possible such interpretation could consist in allowing "erasure" to consist in restricting access rights to a data subject's personal data so that only the data subject can view. A second interpretation could consist in classifying hashed personal data as anonymized data, even though this goes against current opinions [8, 11]. With these interpretations, private blockchains that store personal data via a hash

function and/or by enabling adequate access restrictions, would not violate the GDPR [13]. The European Blockchain Service Infrastructure (EBSI) is following a permissioned approach leveraging a private, Ethereum-based implementation (Hyperledger Besu)[1].

3.1.3 Using mutable blockchain-like data structures Another potential solution that has been suggested in the literature is the use of editable (aka mutable) blockchains. Given the definition we gave in this paper, the term "editable blockchain" may sound like an oxymoron: a blockchain is an "append-only data structure", thus making it editable goes against its definition. But apart from this philosophical remark, the idea of having data structures that share some properties of the blockchain but are mutable remains interesting.

The concept of a mutable blockchain-like data structure is inherently tied to that of private, consortium blockchains. As observed by Politou et al. [11, 14], permissionless DLTs use immutability as a means to establish trust. Since trust in a third party already exists in private (permissioned) DLTs makes immutability less critical and thus makes it easier to design mutable blockchain-like data structures. The first mutable blockchain was proposed by Ateniese et al. [15] and was the subject of an Accenture patent[2]. It uses Chameleon hashes [16], which support a trapdoor that, once known, makes it possible to identify arbitrary collisions, thereby rendering the hash non-unique. However, the use of a trapdoor requires either the presence of a trusted third party that will decide when to use that trapdoor or a secret-sharing mechanism among a set of nodes. The authors warn that sharing the trapdoor key among a large set of nodes (> 200) may lead to performance issues. As a result, they recommend sharing the trapdoor key among a preselected committee. However, this causes the permissionless setting to fall back onto a permissioned one, at least for the sake of editing the blockchain.

A more recent system, μChain [17], proposes a solution in which the blockchain maintains several encrypted versions of each transaction. This makes it possible to revert a transaction by switching the active version. This is done by hiding the key corresponding to the old active version and revealing the key corresponding to the new version. In this case, the decision to switch versions can be managed by the entire set of miners, making it possible to really support permissionless systems. However, it is not clear what happens to transaction versions that are hidden after a change, as the corresponding key may already be out in the wild.

Deuber et al. [18] highlight a further problem with μChain: the fact that a malicious user may not include alternative versions and is free to prevent others from reverting his transaction. They thus propose a scheme that addresses this issue while improving scalability by not relying on multiparty computation. Their approach effectively allows nodes to remove all information that is edited out of

[1] https://ec.europa.eu/digital-building-blocks/wikis/display/EBSIDOC/Ledger+API
[2] https://cointelegraph.com/news/accenture-secures-patent-for-its-editable-blockchain-technology

the blockchain without relying on encryption, but only on majority voting. As a result, it requires (like other blockchain properties) a majority of honest miners.

The scheme proposed by Deuber appears to provide enough flexibility to enable blockchain applications to remove data when needed, without significantly hampering the security properties of the blockchain. This should make it possible to address most, if not all, of the issues associated with immutability, with the possible exception of public keys (see the discussion in our previous chapter's section on private keys [4]. It remains to be seen whether this technology or its future evolutions will become prominent in the blockchain ecosystem.

3.1.4 Using legal scope of interpretation Ambiguities in the interpretation of the legislation could in some cases be used to argue that certain issues could be solved by a different interpretation of legislative provisions. This is, however, only possible to a certain degree and will depend in the end upon the acceptance of the interpretation by supervisory authorities as well as the courts as the final legal instance.

Right to rectification. If the data has been stored on the blockchain itself, it would not be possible to adjust the data itself, but only to add new data which corrects it [19, p. 24] [8, p. 10]. If courts accept this solution, then this addendum could also be considered as informing the other nodes of the rectification. Therefore, this would also possibly satisfy article 19 GDPR, which requires the data controller to also inform other parties who are using the data.

Right to erasure. If the personal data to be erased is stored in the blockchain, it cannot be simply erased. According to Finck, the law stipulates that controllers should take account of available technology and the cost of implementation, but this only relates to the obligation to inform other controllers who are processing the personal data, and therefore it is not clear whether this exception could be taken into account in the case of DLTs [19, p. 24].

Another possible approach is that the notion of 'erasure' is not defined, and could therefore possibly also be interpreted in a more lenient way, for example only making data inaccessible or limiting the processing, if deletion is not possible [19, p. 25]. The ICO, the British data protection authority, refers to this as *"putting data beyond use"* [20]. Essentially, this means that data cannot be used to inform any decision in respect of any individual or in a manner that affects the individual in any way, that no other organization can access the data, that proper technical and organizational security measures are put in place, and that the controller commits to permanent deletion of the information when or if that becomes possible. However, since eventual deletion is not possible on DLTs, it is not certain that other data protection authorities would consider this method sufficient for compliance [19].

Another argumentation is based on art. 17 (1) (a) GDPR, which provides that the right to erasure can apply if the personal data are no longer necessary in relation to the purposes for which they were collected or processed. The argument in that case could be that the processing of the data as it is included in the DLT is still necessary either for the operation of the DLT or for other purposes to bez

evaluated on a case-by-case basis. Similarly, Art. 17 (1) (b) GDPR provides that the right to erasure applies if the data subject withdraws consent. In that case, it could potentially be argued that the core functioning principle of a technology is a legal ground for the processing [9, p. 426]. However, these arguments would still need a careful analysis before the personal data is added to the DLT, and would probably in many cases not be accepted.

Right to object, right to restriction of processing. In some cases (see the section on immutability in [4]), the issue is that the regular verification of the whole blockchain could be considered processing of personal data, and such processing might not be allowed if the data subject invoked the right to object or the right to restriction of processing. For the right to restriction of processing, it seems reasonable that the inadvertent processing of the data without any other effect could be considered to fall under the exception of storage (art. 18 (2) GDPR). However, different from the right to restriction of processing, the right to object does not include a storage exception. It might be possible to invoke the legitimate interest of the controller to continue using the DLT, assuming that the balancing exercise is carried out appropriately.

3.2 Decentralization

The decentralization of actors conflicts with GDPR's notion of a controller as a central processing entity. The designation of the controller will depend on the type of ledger used: in the case of permissioned blockchain, whoever operates and controls the blockchain deployment may be the controller. However, the situation is less clear for permissionless ledgers, as there is no entity that "controls" the blockchain. The situation is exacerbated by the global nature of the DLTs, as nodes may be located anywhere in the world.

3.2.1 Determining the controller The data controller is the entity that determines the purposes and means of the processing of personal data, alone or jointly with other entities (art. 4(7)). Ascertaining who determines the means and purposes of processing can be carried out in three ways under the applicable law.

Option 1: applying the household exemption. The household exemption is specified in art. 2(2)(c), stating that the GDPR does not apply to the processing of personal data by a natural person in the course of a purely personal or household activity. This exception must however be interpreted restrictively, and for example does not include filing systems intended to be used by other persons, as decided in the 2018 Jehovah's witnesses case [21].

In a blockchain context, lightweight nodes only perform operations by interacting with full nodes. As a result, they do not store transactions and can be considered to fall under the household exemption, especially when their operations only involve their own personal data [9]. Likewise, users who only submit their own personal data to the blockchain, might not be considered as falling

under the scope of the regulation as a result of the exemption [22]. For example, this is the case in Self-Sovereign-Identity solutions, where users (acting as Holders) only process their own information.

Option 2: existing case-law on joint controllership – the "effective means" test. Recently, the opinions of the EDBP and the case-law of the CJEU have focused on the "effective means" as the factor in determining the controller [7, 23].

Two recent decisions of the Court of Justice of the EU (CJEU) have dealt with the notion of (joint) controllership. In its ruling C-210/16, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH, (hereafter: the Wirtschaftakademie decision), the Wirtschaftsakademie was running a Facebook fan page without notifying visitors that their information was being collected through cookies. The court, following the Advocate General's opinion, ruled that the determination of a controller is factual rather than formal and must be interpreted broadly in order to ensure the effective and complete protection of data subjects.

In the second decision, C-40/17, Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV (hereafter: the Fashion ID decision), the online retailer Fashion ID embedded a Facebook Like button on its website, through which personal data were transferred to Facebook without the data subject's knowledge. When a consumer protection organization brought proceedings against it, the retailer claimed it had no control over data transmission to Facebook nor how the latter would use those personal data. The court ruled that the embedding of a Facebook Like button points to a decisive influence of Fashion ID on the data transfer to Facebook, which would not have occurred without the plugin. In other words, "the fact that Fashion ID does not have access to the data collected and transmitted to Facebook did not change CJEU's conclusion" [24] that Facebook and Fashion ID are joint controllers.

Keeping the "effective means" test in mind, that means that in DLTs there are the following options for a controller:

- Each full node/miner separately is a controller. While lightweight nodes probably fall under the household exemption (see Option 1 above), full nodes, who create and store transfers, do not as they provide essential contribution. Nevertheless, it remains difficult to see full nodes as controllers, since they themselves cannot determine the means and purposes of processing. Similarly, miners, who collect data into blocks and validate it, make essential contributions but are unable to determine the means and purposes of processing on their own [25].
- Another option is to consider the collective of full nodes, who hold the economic power, and miners, who can provide processing power. Together, they can modify or break the consensus protocol by making changes called forks. Soft and hard forks in the code are users who have the power to adopt new rules, and thus have the power to determine the why and how of processing, thus potentially falling under the definition of the controller [25].

– Network users who submit personal data as part of their business activities might be considered controllers insofar as they are not transferring their own personal data [22].

All three of these possibilities are characterized by either collective controllership (the entire network as one controller) or joint controllers. However, enforcing the law in an environment with potentially hundreds or thousands of controllers would be impossible—instead, a third argument can be made that identifying a controller should be based on a contextual assessment.

Option 3: contextual assessment rather than upfront designation of a controller. Since an upfront designation of a controller does not seem feasible in many cases, DLTs can instead be seen as a general-purpose technology. As an analogy, a general controller cannot be determined for the internet as a general type of technology. Instead, controllers are defined for specific applications (e.g., website owners are controllers for their website). The same argument could be applied to DLTs [26]. Thus, the entity running the blockchain (i.e., providing means of processing) as the responsible party for implementing measures such as privacy by design, could, by implication, be considered a controller. This approach, however, requires identifying the "entity running the blockchain".

To do so, parallels can be drawn to the entities running the Internet as another general-purpose technology:

– Nodes & internet intermediaries (e.g., ISPs, cloud service providers) can act as either data controllers or data processors, depending on whether they process data for their own ends [23, 27]. Nodes that process data on the blockchain for users' purposes and not their own, could likewise be considered processors and not controllers.
– Hosting data on (external) servers: following the EDPB's argument in its guidelines on controllership that, while purely hosting personal data corresponds to being a processor, if the host can do anything further with the data, then it should be considered a data controller [27].

3.2.2 Transfer of data outside the EU As analyzed in the section on challenges of decentralization in [4], decentralization leads to further challenges when controllers are based in non-EEA jurisdictions. While private permissioned and public permissioned DLTs can carefully select nodes and/or clients based on their geographical location, public permissionless DLTs remain the problem due to the lack of such control. In [4] we described the four options under the GDPR for transfers to non-EEA jurisdictions: a) territory with an adequacy decision, b) transfer under additional safeguards, c) binding corporate rules, and d) exception for non-repetitive transfers, if explicit consent is given.

We identify three ways to tackle compliance of global DLTs. The first two have a legal flavor: either finding a legal basis for international transfers under the GDPR, or political action which could lead to regulatory convergence. The third relies on technical advancements that are currently being investigated.

Solutions under the GDPR (de lege lata). Potential solutions depend on the type of DLT:

- *Private permissioned DLT*: the company running the blockchain has control over the location of nodes as well as clients. Since the company is likely to be considered the controller, it will need to ensure adherence to the GDPR rules on potential international transfers. Clearly, it can opt for either option a) or b). This implies careful selection of the nodes, taking into account their geographical location, similarly to choosing a data processor. Following art. 28(1), the controller can only use those processors that provide sufficient guarantees to implement appropriate technical and organizational measures in such a manner that processing will comply with the rules of the GDPR and ensure the exercise of data subject rights. If the company running the blockchain belongs to a group of undertakings, binding corporate rules are also an option.
- *Public permissioned DLT*: Similarly to the private permissioned case, the consortium as the data controller can choose either option a) or b). Depending on the status of the clients as either controllers or processors, additional agreements must be concluded: an arrangement with division of responsibilities among the joint controllers, especially on the exercise of data subjects' rights (art. 26 of the GDPR); or a controller-processor agreement under art. 28(3). Binding corporate rules under c) can only be an option insofar the clients and the company/consortium can be considered to belong to the same group. However, it is currently technically impossible to control where nodes are located, even if the chain's protocol could in theory be amended to restrict processing to EEA-based nodes [19].
- *Public permissionless DLT*: this case does not seem to fit any of the four options.

Option a) could only be relevant if all jurisdictions in the world had an adequacy decision, which is politically unlikely to happen.

Adopting appropriate safeguards under option b) is potentially relevant, but could be difficult in practice. The first step in ensuring compliance of international transfers is mapping the personal data and knowing where it is located. This can be extremely challenging if a multitude of controllers and processors are involved. While the EDPB does not mention DLTs explicitly, they are definitely an example of such a complexity [28].

Option c), binding corporate rules, is discussed by Renieris and Greenwood [29]. Transfers might be possible insofar the network rules are legally binding, offer data subjects a mechanism to enforce their rights, and ensure the minimal data protection standards foreseen in art. 47(2). However, it is not clear how this would work in practice, considering how difficult it is to enforce data subject rights in public permissionless environments in general. Nor is it clear whether a public permissionless blockchain falls within the meaning of "group of undertakings".

The last remaining possibility to lawfully transfer data is the derogation contained in art. 49. Art. 49(1) allows for an exemption from options a), b) and c) insofar any of these options is not possible, but the transfer is only permissible if it meets restrictive criteria. First, the data subject has given explicit consent after

being informed about the possible risks. Second, the transfer is not repetitive, it concerns only a limited number of data subjects, and it is necessary for the purposes of compelling legitimate interests pursued by the controller which are not overridden by the interests or rights and freedoms of the data subject. Finally, the controller has assessed all the circumstances surrounding the data transfer and has on the basis of that assessment provided suitable safeguards with regard to the protection of personal data. However, art. 49 is a derogation, which must be interpreted restrictively lest it become the rule [28].

It is difficult to tell whether transactions in a public permissionless blockchain meet all these criteria under option d). While obtaining explicit consent to third-country transfer could be realistic, the question remains whether blockchain transactions fall under the non-repetitiveness criterion. The EDPB explains that non-repetitive transfers "may happen more than once, but not regularly" and this happens "outside the regular course of actions, for example, under random, unknown circumstances and within arbitrary time intervals" [28]. DLTs are persistent and decentralized, resulting in regular, constant transactions of personal data, which renders transfers based on the explicit consent derogation unlikely to be lawful.

No data protection authority has yet declared a public permissionless blockchain unlawful, despite the difficulties in finding a solid legal basis for international transfers. However, this does not mean that enforcement will never occur, resulting in a difficult situation for operators.

Regulatory convergence – potential legal solutions (de lege ferenda). Global regulatory harmonization could help mitigate risks relating to market functioning of DLTs [30]. However, that depends largely on political consensus: if DLTs are seen by policy-makers as the equivalent of the next Internet, convergence of regulatory guidelines is likely to follow, at least regarding permissioned blockchains (both public and private) [31]. On the other hand, pseudo data location requirements are not seen as desirable by policy-makers such as the European Commission. Nor are such requirements realistic, since DLTs (and the Internet) are global, and putting data location requirements in place could be a potentially large barrier for European companies who are active on a global scale [32].

Technological solutions. Instead of trying to understand how to apply the law to a hard-to-control system, a possibility consists in building systems that naturally and by their very nature abide by the law. In the context of DLTs, the PriCLeSS project[3] is currently exploring whether sharding solutions can improve ledgers not only with respect to scalability but also by making them more legally compliant. In layperson terms, sharding consists in the decomposition of a ledger into multiple sub-ledgers (the shards) each responsible for a subset of the data and a subset of the transactions. It is therefore natural to see sharding as an opportunity to store data only on devices that are legally allowed to do so. For example, one could imagine sharding a ledger so that data from a company is only stored on devices managed by partner companies and not competitors.

[3] `https://project.inria.fr/pricless/`

Similarly, in an international context, one could imagine that data about EU citizens will remain on shards managed by nodes located within the EU.

Although this may sound like a simple and straightforward solution, the approach clashes with one of the principles that ensure the safety and security of distributed ledger: the unpredictability of storage locations. Sharding solutions leverage randomness in the assignment of nodes to shard to limit the possibility of an attacker to gain control of an entire shard. The challenge being addressed in PriCLeSS lies in balancing these two contrasting needs.

3.2.3 Consent management in a decentralized environment According to the GDPR, and as stated at the end of the section on the legal background of DLTs in [4], consent management involves three steps: Collection, Update, and Revocation. Collection refers to consent being collected from the Data Subject and stored in compliance with the accountability principle. The Update step occurs each time the Privacy-Notice use-consent collection is modified. When this happens, consent must be collected again. The Revocation step happens when the data subject revokes their consent. This may happen at any time.

In a decentralized environment, and as far as a Data Controller is identified (according to Section 3.2.1), consent management may be an issue (as described at the end of the section on decentralization in [4]), but it can be managed thanks to different entry points: applications built on a Blockchain may leverage smart front-ends that can interact with the the Data Subject in order to collect consent.

In the case of online Wallets for example, the company providing the wallet—e.g. an exchange or gaming platform—appers the most likely data controller as discussed in Option 3 in Section 3.2.1. The company can clearly leverage the wallet for collecting consent: either by placing a special transaction on the wallet, or simply by requesting a confirmation on the website before accessing the service. However, this may be a good approach for solving the Collection part of the problem, but update and revocation remain difficult if the user can interact with the wallet outside the web-based platform.

A potential solution, which would however require some research, could be to leverage the feature of Self-Sovereign Identity platforms. For example, we could imagine the data subject acting as an issuer of a credential that grants consent to a data controller, acting as a credential holder. Legal authorities would then take the role of verifiers and verify that the credentials held by the data controller are valid. This framework would for example allow the data subject to revoke an issued credential at any time. Optionally, smart contracts could further automate this process, by enforcing checks on the consent credentials held by data controllers.

3.3 Automation

We now discuss the solution to the challenges posed by automation. We divide our discussion into three parts. First, we consider the legal solutions, then tech-

nical solutions, and then apply them to the examples we gave at the end of the automation section in [4].

Legal Solutions. As discussed in [4] section on automation, solely automated data processing can pose significant challenges when it comes to a valid legal basis of data processing. Possible legal solutions can come from the specific exceptions to the general prohibition of solely automated data processing set in Article 22 of the GDPR.

The most common exception refers to the collection of consent, in the form of a compliant privacy notice. The controller should obtain explicit consent from the data subject for any automatic data treatment or decision based solely on automated processing that produces legal effects on the data subject. Consent can be obtained in a decentralized setting using the solutions outlined in Section 3.2.3. Another exception involves processing that is necessary for entering into, or executing, a contract between the data subject and a data controller. Finally, the third exception involves processing that legally authorized by the Member State law to which the controller is subject.[4].

Technical Solutions. From a technical perspective, a promising solution to the hurdles of automated decision-making comes from auditing. Auditing a smart contract—or a DAO, a form of a complex smart contract—can contribute to understanding and outlining it specific behavior. As a result, audited smart contracts allow all the stakeholders to fully understand the logic behind the automation, enabling a transparent data treatment. While auditing can be achieved manually, this is labour-intensive and error-prone. Automated auditing tools, albeit still in their infancy, offer a promising alternative.

In particular, a number of authors have proposed the use of formal-verification techniques for auditing smart contracts. Murray and Anisi [33] propose a survey of existing approaches to the formal verification of smart contracts. In particular, they consider several contributions in the area of formal verification of smart contracts. One of the earliest such contributions [34] proposes rewriting smart contract using the F^* [35] language, to enable formal verification through a combination of SMT solving and manual proofs. However, their SOLIDITY to F* compiler (SOLIDITY*) cannot support many of the features of SOLIDITY, which limits its applicability. The paper also proposes EVM*, a decompiler which can translate Ethereum byte code into F*, making it possible to analyze smart contract with no publicly available source code. Similarly, KEVM [36] provides an executable formalization of the Ethereum Virtual Machined based on the K Framework [37]. Another approach consist in using model checking. For example, the NuSMV [38] symbolic model checker was successfully used [39] to model the Ethereum blockchain, its smart-contract execution environments and some smart contracts to demonstrate the feasibility of this approach.

Automated, or semi-automated, auditing is not only an interesting research topic, but also used by a number of companies. For example, OpenZeppelin[5]

[4] For more details, please refer to Article 22(2) of the GDPR

[5] `https://www.openzeppelin.com/security-audits`

appears to use a semi-automated auditing technique with security experts that exploit tools to analyze smart contract code. Solidified[6] also performs semi-automated auditing thanks to a panel of experts that include cryptographers, distributed-systems researchers and economists. They also offer code reviews and penetration testing. Diligence[7] is a product by Consensys that offer automatic scans as well as manual code reviews by security experts. They also advertise the ability to monitor the code for vulnerabilities while it's being maintained and changed. Finally, Solidity Finance [8] advertise a combination of automated tools, including static analysis, and manual code reviews.

Putting solutions into practice. Let us now reconsider the examples we presented in the automation section of the previous chapter [4] in light of the solutions we just considered.

In the first example, a vendor uses a smart contract enriched with the Data Subject's personal data: applying what was set out just above, the Data Controller will need to comply with using one of the legal bases-or exceptions-under the GDPR and, above all, subjecting its code to audit, perhaps by a third party.

In the case of a DAO, the solutions outlined above need to be implemented with even greater care. It might be difficult to apply one of the exceptions, given the amount and variety of use of the tool, having to resort to consent. It remains, even in this example, necessary to audit the code, which might be more complex than the simple, single smart contract, mentioned in the first example.

In the third example, the applicable solutions remain the same, although things get more complicated: while the use of different DLTs does not change the approach to the legal basis, it will certainly be more complicated to perform a full audit. Indeed, each chain has its own computational logic that could increase the attack surface: in this case, it might be useful to perform more than one audit to ensure compliance.

3.4 Applying the proposed solutions to the examples

3.4.1 Public Keys As we discussed in the section on public keys in [4], it is often hard, if not impossible, to determine if a given public key should be considered personal data. Moreover, even when a public key can be determined to be personal data, the information that links it to an identified or identifiable natural person may not be directly available. From a data protection point of view, it may be best to react to this doubt by acting as if public keys consist of personal data. This would ensure protection of personal data and compliance with the legislation. However, from a technical perspective, it appears to be overkill and above all impractical to consider all the public keys stored on a blockchain as personal data. A more sensible approach would consist in considering the information concerning the link between a public key and a person

[6] `https://www.solidified.io/#what-we-do`
[7] `https://consensys.net/diligence/`
[8] `https://solidity.finance/`

as personal data. We argue that this measure could be sufficient to protect the privacy of individuals with respect to their public keys. In all cases, since perfect compliance with the GDPR does not appear to be possible in the case of public keys on the blockchain, a data subject using a blockchain should be aware of the potential consequences and reduced possibility to delete them later.

The difficulty to identify controllers, also applies in the context of public keys, and actually the solutions we describe in Section 3.2.1 apply here. For example, if a data subject is using the blockchain and their public keys in the scope of their purely personal or household activities, the household exemption would apply. If, on the other hand, a person or company uses public keys of a data subject in a commercial transaction, then they would become the controller. Like for immutability, solutions that hide the link between a public key and the associated natural person, if any, appear to be the most promising.

When considering the risks associated with data transfer outside the EU, the solutions we considered in Section 3.2.2 also apply to public keys. In this respect, it is also worth noting that, since the transfer of public keys tends to be regular, it is unlikely that the derogation of art. 49 GDPR can be used.

With respect to the problem of linkability highlighted in the public keys section of [4], several technical solutions exist. One of them consists of Deterministic wallets, which make it possible to derive many keys from a single "seed." The most advanced form of deterministic wallets is the HD wallet defined by the BIP-32 specification[9]. HD wallets contain keys derived in a tree structure, such that a parent key can derive a sequence of children keys, each of which can derive a sequence of grandchildren keys, and so on, to an infinite depth. Hierarchical deterministic wallets have also been proposed as a solution to privacy in private distributed ledgers. An HD wallet uses algorithms to create a new public-private key pair for each transaction or piece of a larger trade, generating a virtually infinite stream of keys from a single master seed. This can provide unlinkability, and make the user's identity difficult to trace, provided that the user correctly refrains from reusing previous keys.

Another solution providing unlinkability consists in the use of ring signatures. A ring signature combines a group of individual signatures to produce a unique signature that can be used to trigger a transaction. One of the most widely known examples of a blockchain using ring signatures by default to protect privacy is Monero[10]. To make it hard to identify a transaction's sender, ring signatures combine his/her identity with that of other users, making it computationally infeasible to determine which one originally generated the transaction.

These solutions do not prevent public keys from being personal data, but they at least solve the linkability problem.

3.4.2 Wallet solutions To address the data protection challenges faced by DLT wallets, several possible solutions have been proposed. In order to secure the private key management, hardware wallets can significantly reduce the risk

[9] `https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki`
[10] `https://www.getmonero.org/`

of unauthorized access or theft [40] Additionally, multi-signature wallets require multiple keys to authorize a transaction, providing an extra layer of security. The integration of privacy-enhancing technologies, such as zero-knowledge proofs, into blockchain wallets can help protect user anonymity and maintain transaction privacy while preserving the transparency of the underlying blockchain [41] Another countermeasure involve the use of strong authentication methods, such as two-factor authentication (2FA) and biometric authentication, which can help protect user accounts and minimize unauthorized access [42].

3.4.3 Self-Sovereign Identity In the section on self-sovereign identity in [4], we outlined the problems that may arise due to the use of a blockchain in SSI systems. In particular, we highlighted that even the storage of public DIDs may still raise issues with respect to the principle of data minimization and the right to erasure. The problem results from the immutable and persistent nature of the data that is stored on a blockchain. Once a DID has been published there, it cannot be unpublished.

The solution to this lies in assessing the actual requirements of DID storage in SSIs. It is true that public DIDs do need to remain publicly accessible. But they only need to do so for as long as they are active. Thus, a blockchain may not be the wisest choice to implement a DID registry. Projects like SOTERIA[11] (Section 4.5) are currently working on blockchain-less solutions for Self-Sovereign Identity Systems, for example.

Another issue that was mentioned in the section on self-sovereign identity in [4] relates to the use of persistent identifiers in eIDAS and eIDAS2[12]. But, as already pointed out, this is not strictly an SSI problem, but it is associated with the way the eIDAS regulation interprets the concept of identity. From an SSI viewpoint, persistent identifiers are not a requirement.

4 Projects

This section gives some examples of recent projects using DLTs in conjunction with personal data processing.

4.1 DIZME

DIZME[13] is a Trust Framework based on the SOVRIN Foundation framework and ledger. The project is based on the Self Sovereign Identity (SSI) paradigm, which aims to give back to the identity owner (the "Owner") the management of his/her digital identity. The solution is technically based on ZKP and data minimization oriented tools. Most of the pillars of the GDPR are also embraced by the SSI guiding principles:

[11] `https://www.soteria-h2020.eu/`

[12] `https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation`

[13] The name is a contraction of "this is me". An early version of the website can be found under `https://dizme.io/`

- Control: users must control their identities
- Access: users must have access to their own data
- Transparency: systems and algorithms must be transparent
- Portability: information and services about identity must be transportable
- Consent: users must agree to the use of their identity
- Minimization: disclosure of claims must be minimized

The SOVRIN ecosystem is based on two W3C standards: Decentralized IDentifier ("DID") and Verifiable Credentials ("VC" or simply "Credential").

This approach provides for three main players:

- Issuer: the legal/natural person who issues the Credential
- Verifier: the legal/natural person that asks for and verifies the Credential
- Holder: the legal/natural person to whom the Credential is related

Once the Issuer has given the Credential to the Holder, the Holder can spend it in front of the Verifier for any needed purpose, without involving the Issuer. This approach helps to protect personal data because on one hand the Issuer does not know where the Holder is spending his/her credential, and on the other hand it allows the Holder to have – by design – most of the rights given by the GDPR (see the section on data subjects' rights in the legal introduction of [4].

DIZME is a Domain Specific Governance Framework (DSGF) for trusted identity and it comes with a specific mobile application (DIZME wallet) that enables the control of the Owner over his/her digital identity. In the DIZME framework, Issuer, Verifier and Holder interact easily with a QR code system that allows to ask for a specific Credential, which is seamlessly spent through the wallet itself. The wallet is bound to the Holder through a secure onboarding procedure: from this starting point, the user can start to ask Credential issuing, according to the level of assurance he/she needs. The user can choose three different levels of identification: level 1 (L1) is a self-assessed identity, level 2 (L2) is L1 corroborated by some ID checks, and level 3 is L2 corroborated with specific remote onboarding procedure compliance. This multi-level approach is in accordance with the data minimization principle. DIZME is also a bridge between the SSI and the Qualified Trust Services given in the eIDAS Regulation[14].

4.2 KRAKEN

The KRAKEN project aims to enable the sharing, brokerage, and trading of personal data including sensitive data by returning control to both data subjects and data providers throughout the entire data lifecycle. The project is providing a data marketplace which allows the sharing of personal data and its usage for research and business purposes, by using privacy-preserving cryptographic tools.

[14] "Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC"

To achieve this goal, KRAKEN is developing an advanced platform to share certified information between users and organizations by leveraging DLT, promoting the vision of Self Sovereign Identity solutions, preserving security, privacy, and the protection of personal data in compliance with EU regulations. The development carried out by the project is set out in three main pillars:

- Developing an SSI solution to provide a decentralized user-centric approach to personal data sharing.
- Implementing a set of analytic techniques based on advanced cryptographic tools to permit privacy-preserving data analysis.
- Integrating the above techniques into a data marketplace, allowing the sharing of personal data and privacy-preserving data analytics.

The cryptographic tools provided by KRAKEN are based on Secure Multi-Party Computation (SMPC). SMPC allows a group of nodes to compute on secret inputs jointly, without disclosing their respective inputs to the other nodes or any other party. Even if one of the nodes is malicious, SMPC guarantees that this malicious node cannot infer anything about the input of the other nodes of the network. Furthermore, the correctness of the computation can be guaranteed as long as a sufficiently large fraction of the nodes behave honestly.

In addition, KRAKEN implements an Identity and Access Management (IAM) approach, based on blockchain, for managing the identity of end users in order to empower data subjects to control their data. This SSI approach provides a solution where the end user has the whole control of their identity information in their own mobile device. In KRAKEN, the development of an SSI solution is going one step further by addressing one of the biggest challenges of SSI: what happens if the end user loses his/her mobile phone or has different devices where they want to use their identity information. KRAKEN implements a backup service of the different secrets and identity information necessary for the use of an SSI solution using cryptographic proxy re-encryption techniques that ensures that the secrets and identity information are never disclosed outside the end user's mobile device. KRAKEN also provides the functionality to obtain eIDAS compliance credentials and adheres to the recommendations and suggestions proposed by the EBSI and eSSIF bodies.

All these techniques and functionalities are applied by a marketplace, leveraging the privacy preservation of personal or even sensitive data. The marketplace is piloted in two different domains: eHealth and Education.

4.3 European Blockchain Service Infrastructure – EBSI

The European Blockchain Service Infrastructure (EBSI) is a blockchain-based initiative launched by the European Union in 2019 to create a standardized and interoperable infrastructure for the delivery of cross-border public services. The EBSI is built on top of existing national blockchain networks and leverages open-source technology to enable secure and transparent data exchange across European borders. EBSI leverages leading standards such as the W3C

Decentralized Identifiers (DIDs), Verifiable Credentials (VCs), and Verifiable Presentations (VPs), as well as OpenID for Verifiable Credentials, the General Data Protection Regulation (GDPR), the electronic Identification, Authentication, and Trust Services (eIDAS) regulation, and other relevant EU regulations. By building on these standards, EBSI aims to establish a generic profile for the entire life-cycle of self-sovereign identity (SSI), from credential issuance to verification and presentation. This approach enables individuals to retain control over their personal data and facilitates the secure and privacy-preserving exchange of verified information across different services and applications.

4.4 PriCLeSS

PriCLeSS (Privacy Conscious Legally Sound blockchain Storage[15]) is a French project funded by the CominLabs LabEx[16]. PriCLeSS consists of a partnership between computer-science and law researchers and aims at bridging the gap between GDPR and blockchain-based storage applications. The project addresses three major challenges. The first consists in leveraging the characteristics of distributed ledgers as a tool to automate the auditing of operations on personal data. The second consists in providing novel ledger designs that can take into account the requirements of GDPR like the enforcement of European borders in the context of data transfer. The third involves the design of an ecosystem of tools that can complete the blockchain with other resilient distributed data structures that can offer features that are currently absent in the blockchain context, like mutability, the ability to erase data, and access control.

4.5 SOTERIA

SOTERIA[17] is an H2020 Innovation Action led by IDNow, a leading identity-proofing provider in Europe. SOTERIA aims to build a decentralized platform for the management and storage of personal data, combined with advanced identity management tools. This platform will be tested in three use cases involving citizens. But beyond this high TRL endeavor SOTERIA is also exploring the design of novel blockchain-less decentralized identity management solutions.

5 Related work

Previous work on data protection aspects of DLTs consists of papers from a technical point of view (reviewed in Section 5.1) and papers from a legal point of view (reviewed in Section 5.2).

[15] `https://project.inria.fr/pricless/`
[16] `https://cominlabs.inria.fr/`
[17] `https://www.soteria-h2020.eu/`

5.1 Technical literature

Halpin and Piekarska provide a high-level overview of the security and privacy challenges of the blockchain [43]. The identified challenges include a lack of formally stated privacy and security properties, the difficulty of upgrading the cryptographic primitives used in a blockchain system, and the limited privacy and anonymity offered by blockchain. For the latter challenge, also some potential solutions—like mixing services or Succinct Non-interactive ARguments of Knowledge (SNARKs)—are mentioned.

Bayle et al. address the problem of how the blockchain may be used in such a way that it complies with one of the provisions of the GDPR, the right to be forgotten, despite the immutability of the blockchain [44]. They suggest keeping any personal data off-chain, and only use the blockchain for keeping track of actions affecting the data. This way, personal data can be erased if necessary.

Li et al. survey security challenges of blockchain systems, with a particular focus on cryptocurrencies [45]. They identify 9 main security risks, review several real-world attacks on blockchain systems, as well as proposals for enhancing the security of blockchain systems. They do not specifically consider data protection. However, they mention "transaction privacy leakage", i.e., the possibility for an attacker to (partially) re-identify pseudonymized transaction data, as one of the risks.

Feng et al. present a survey of privacy challenges in blockchain systems, also with a focus on cryptocurrencies [46]. They identify two main challenges: identity privacy (i.e., no leakage of identity information about the participants) and transaction privacy (i.e., no leakage of transaction data, such as the transferred amount). After presenting possible attacks on privacy, they review techniques that have been proposed for privacy protection in blockchain: centralized and decentralized mixing services, as well as different cryptographic approaches including ring signatures, non-interactive zero-knowledge proofs (NIZK), zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARK), and confidential transactions.

Bernabe et al. present a comprehensive survey of privacy challenges and privacy-preserving solutions for blockchain, with a particular focus on identity management [6]. They identify eleven challenges, one of which is compliance with data protection regulations like the GDPR. They review several privacy-preserving techniques and group them into four categories: smart contract / key management, identity data anonymization, transaction data anonymization, and on-chain data protection.

A common shortcoming of these papers is that they do not address compliance with the GDPR as a whole. Existing technical solutions may enhance privacy, but this does not guarantee compliance with the GDPR.

5.2 Legal literature

The literature surveyed often points to the conflicted relationship between data protection law and distributed ledger technologies. Nevertheless, authors often

clarify that DLTs in themselves are not necessarily always incompatible with the law. This is because blockchain, like the internet, is a general purpose technology, and thus in order to assess legal challenges and compliance we need to look at the context in which the blockchain is being used and keep in mind that in itself using the blockchain is not data processing [22, 26].

The main clashes occur in the area of scope of application (both material and territorial), responsibility and accountability, including the exercise of data subject rights; and implementing compliance with the obligations of the GDPR. Legal consequences also differ depending on what type of DLT is used (private/public, permissioned/permissionless), as well as which technological features are at stake [47].

Blockchains operate globally, while the GDPR only applies within predetermined territories; it could possibly apply to blockchain actors established outside the EU under the territoriality principle under its art. 3. However, this could make enforcement difficult, since only the EU authorities are authorized to enforce the compliance with the GDPR. [9, 19] Material scope of application refers to the concept of personal data in the DLT context. It is unclear to which data exactly GDPR could apply. Most authors consider that information such as public keys, user credentials, their copies, and revocations could be considered personal data [9, 48]; some question whether private keys should also fall under this definition [8, 19, 49].

The notion of controller is replaced by system architecture and code, which makes the DLTs more reliable but less flexible and less accountable [50]. It is unclear whether one actor or the entire network should carry the brunt of the responsibility to comply with the law and apply technical and organizational measures; especially in the case of public networks where either no node, or every node where data are technically processed is responsible [9,25]. The 'accountability gap' means that if there is no controller, the data subject rights lose effectiveness [9]. It is important to understand which actors or entities have the responsibility, otherwise it could lead to further problems with enforcement, enforcing data subject rights, and determining who should implement measures [25]. It has been suggested to adopt a micro vs. macro perspective and that users of blockchain platforms are likely to be deemed controllers, while the miners and nodes act as processors [10]. In the case of public blockchains, their "gatekeepers" are likely to be targeted by regulation [9], thus leading to an assumption that they are to be considered responsible.

DLT in itself is not necessarily always against data protection law, and DLT technologies can function as tools that can facilitate data protection compliance: cryptography, data portability and integrity are some of the features that can serve this objective. This leads to a curious dichotomy: privacy by design (and compliance) is either implemented in the DLT by default, or there is an intrinsic clash between them [9].

While many authors have suggested solutions to the challenges posed by compliance, those solutions are partial and rarely take into account both tech-

nological and legal possibilities, or differentiate the solutions based on the type of DLT used.

Berberich and Steiner [9] considered the challenges posed by privacy by design, the right to be forgotten, territorial scope of application and whether data on the blockchain can be considered personal data. However, they only consider private and public DLTs, without a distinction between permissioned and permissionless. Technical solutions are discussed in abstracto, and legal solutions are limited to the context of privacy by design under art. 25 of the GDPR. Bacon et al. focus only on distributed versus centralized approaches [47] without suggesting viable solutions to bridge the gap.

Some technical solutions for compliance already exist, such as mechanisms that allow data deletion—even though a link would remain on the block, it might suffice to comply with a data erasure request. Likewise, deleting all instances of a private key could be considered sufficient. Technical advances in blockchain deletion might prove to be useful in implementing privacy by design obligation. [10] Other specific solutions could be, e.g., explicit consent under art. 49(1)(a), off-chain storage of PD to comply with the data minimization principle, and anonymization or shredding provide possible means of escape from the GDPR's scope [19, 51].

6 Conclusions and outlook

This chapter presented a survey of the potential solutions for data protection in DLTs, together with some examples and projects that studying this problem. In our companion chapter [4], we observed how the properties of immutability, decentralization, and automation make it difficult to operate DLT-based applications that comply with the GDPR. Here we analyze existing and potential solutions both from a legal and a technical perspective.

The considered solutions cover most of the problems in the case of private and permissioned DLTs, but some challenges remain, particularly in the permissionless case, where it is often difficult to identify a data controller. Ongoing projects are proposing novel technical solutions that can combine the benefits of distributed ledger with more flexible interfaces. At the same time, legal researchers and officials will play a crucial role in understanding how to apply existing regulations or develop new ones in the context of a dynamic technological environment.

References

1. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. `https://git.dhimmel.com/bitcoin-whitepaper/` (2008)
2. Panetta, R., Cristofaro, L.: A closer look at the EU-funded My Health My Data project. Digital Health Legal pp. 10–11 (Nov 2017). `https://doi.org/10.5281/zenodo.1048999`, `https://doi.org/10.5281/zenodo.1048999`

3. Zyskind, G., Nathan, O., Pentland, A.S.: Decentralizing privacy: Using blockchain to protect personal data. In: 2015 IEEE Security and Privacy Workshops. pp. 180–184 (2015). https://doi.org/10.1109/SPW.2015.27
4. Fabčič Povše, D., Favenza, A., Frey, D., Mann, Z.Á., Palomares, A., Piatti, L., Schroers, J.: Data protection challenges in distributed ledger and blockchain technologies: A combined legal and technical analysis. In: El Madhoun, N., Dionysiou, I., Bertin, E. (eds.) Building Cybersecurity Applications with Blockchain Technology and Smart Contracts. Springer (2024)
5. Timan, T., Mann, Z.: Data protection in the era of artificial intelligence: trends, existing solutions and recommendations for privacy-preserving technologies. In: The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem, pp. 153–175. Springer (2021)
6. Bernabe, J.B., Canovas, J.L., Hernandez-Ramos, J.L., Moreno, R.T., Skarmeta, A.: Privacy-preserving solutions for blockchain: Review and challenges. IEEE Access **7**, 164908–164940 (2019)
7. European Parliamentary Research Service: Blockchain and the General Data Protection Regulation. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf (2019)
8. Commission Nationale Informatique & Libertés: Blockchain – solutions for a responsible use of the blockchain in the context of personal data. https://www.cnil.fr/sites/default/files/atoms/files/blockchain_en.pdf (2018)
9. Berberich, M., Steiner, M.: Blockchain technology and the GDPR – how to reconcile privacy and distributed ledgers. European Data Protection Law Review **2**(3), 422–426 (2016)
10. Bacon, J., Michels, J.D., Millard, C., Singh, J.: Blockchain demystified: a technical and legal introduction to distributed and centralized ledgers. Richmond Journal of Law & Technology **25**, 1 (2018)
11. Politou, E., Casino, F., Alepis, E., Patsakis, C.: Blockchain mutability: Challenges and proposed solutions. IEEE Transactions on Emerging Topics in Computing **9**(4), 1972–1986 (2021)
12. Jensen, G.: Reconciling GDPR rights to erasure and rectification of personal data with blockchain. Oracle Cloud Security, https://blogs.oracle.com/cloudsecurity/reconciling-gdpr-rights-to-erasure-and-rectification-of-personal-data-with-blockchain (2018)
13. Mirchandani, A.: The GDPR-blockchain paradox: exempting permissioned blockchains from the GDPR. Fordham Intellectual Property, Media and Entertainment Law Journal **29**(4), 1201–1241 (2019)
14. Politou, E., Alepis, E., Virvou, M., Patsakis, C.: Privacy and Data Protection Challenges in the Distributed Era. Springer (2022)
15. Ateniese, G., Magri, B., Venturi, D., Andrade, E.: Redactable blockchain–or–rewriting history in bitcoin and friends. In: 2017 IEEE European Symposium on Security and Privacy (EuroS&P). pp. 111–126. IEEE (2017)
16. Krawczyk, H., Rabin, T.: Chameleon signatures. In: Proceedings of the Network and Distributed Systems Security Symposium. pp. 143–154 (2000)
17. Puddu, I., Dmitrienko, A., Capkun, S.: μChain: How to forget without hard forks. IACR Cryptology ePrint Archive, 2017/106 (2017)
18. Deuber, D., Magri, B., Thyagarajan, S.A.K.: Redactable blockchain in the permissionless setting. In: IEEE Symposium on Security and Privacy (SP). pp. 124–138 (2019)
19. Finck, M.: Blockchain and data protection in the European Union. Max Planck Institute for Innovation & Competition Research Paper No. 18-01 (2017)

20. Information Commissioner's Office: Deleting personal data, guidance, 26.02.2014. https://ico.org.uk/media/for-organisations/documents/1475/deleting_personal_data.pdf (2014)
21. CJEU: Tietosuojavaltuutettu vs. Jehovan todistajat — uskonnollinen yhdyskunta. ECLI:EU:C:2018:551 / C-25/17, https://curia.europa.eu/juris/document/document.jsf?docid=203822 (2018)
22. European Union blockchain observatory & forum: Blockchain and the GDPR. https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf (2018)
23. Article 29 Data Protection Working Party: Opinion 1/2010 on the concepts of "controller" and "processor". https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf (2010)
24. CJEU: Fashion ID GmbH & Co. KG vs. Verbraucherzentrale NRW eV. ECLI:EU:C:2019:629/ C-40/17, https://curia.europa.eu/juris/liste.jsf?num=C-40/17 (2019)
25. Buocz, T., Ehrke-Rabel, T., Hödl, E., Eisenberger, I.: Bitcoin and the GDPR: allocating responsibility in distributed networks. Computer Law & Security Review **35**(2), 182–198 (2019)
26. Moerel, L.: Blockchain & data protection... and why they are not on a collision course. European Review of Private Law **26**(6), 825–851 (2018)
27. European Data Protection Board: Guidelines 07/2020 on the concepts of controller and processor in the GDPR, version 2.0. https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf (2021)
28. European Data Protection Board: Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data. https://edpb.europa.eu/sites/default/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf (2020)
29. Renieris, E., Greenwood, D.: Unblocking blockchain data flows in the wake of Schrems II. MIT Computational Law Report (2020), https://law.mit.edu/pub/unblockingblockchaindataflowsinthewakeofschremsii
30. European Union blockchain observatory & forum: Legal and regulatory framework of blockchains and smart contracts. https://www.eublockchainforum.eu/sites/default/files/reports/report_legal_v1.0.pdf (2019)
31. Melin, K.: The GDPR compliance of blockchain: A qualitative study on regulating innovative technology. Thesis, University of Uppsala (2019)
32. Christakis, T.: After Schrems II: Uncertainties on the legal basis for data transfers and constitutional implications for Europe. European Law Blog, https://europeanlawblog.eu/2020/07/21/after-schrems-ii-uncertainties-on-the-legal-basis-for-data-transfers-and-constitutional-implications-for-europe/ (2020)
33. Murray, Y., Anisi, D.A.: Survey of formal verification methods for smart contracts on blockchain. In: 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). pp. 1–6 (2019). https://doi.org/10.1109/NTMS.2019.8763832
34. Bhargavan, K., Delignat-Lavaud, A., Fournet, C., Gollamudi, A., Gonthier, G., Kobeissi, N., Kulatova, N., Rastogi, A., Sibut-Pinote, T., Swamy, N., Zanella-Béguelin, S.: Formal verification of smart contracts: Short paper. In: Proceedings of the 2016 ACM Workshop on Programming Languages and Analysis for Security. p. 91–96 (2016)

35. Swamy, N., Hriţcu, C., Keller, C., Rastogi, A., Delignat-Lavaud, A., Forest, S., Bhargavan, K., Fournet, C., Strub, P.Y., Kohlweiss, M., Zinzindohoue, J.K., Zanella-Béguelin, S.: Dependent types and multi-monadic effects in f*. SIGPLAN Not. **51**(1), 256–270 (jan 2016). https://doi.org/10.1145/2914770.2837655, https://doi.org/10.1145/2914770.2837655
36. Hildenbrandt, E., Saxena, M., Rodrigues, N., Zhu, X., Daian, P., Guth, D., Moore, B., Park, D., Zhang, Y., Stefanescu, A., Rosu, G.: Kevm: A complete formal semantics of the ethereum virtual machine. In: 2018 IEEE 31st Computer Security Foundations Symposium (CSF). pp. 204–217. IEEE Computer Society, Los Alamitos, CA, USA (jul 2018). https://doi.org/10.1109/CSF.2018.00022, https://doi.ieeecomputersociety.org/10.1109/CSF.2018.00022
37. Roşu, G., Şerbănuţă, T.F.: An overview of the k semantic framework. The Journal of Logic and Algebraic Programming **79**(6), 397–434 (2010). https://doi.org/https://doi.org/10.1016/j.jlap.2010.03.012, https://www.sciencedirect.com/science/article/pii/S1567832610000160, membrane computing and programming
38. Cimatti, A., Clarke, E., Giunchiglia, E., Giunchiglia, F., Pistore, M., Roveri, M., Sebastiani, R., Tacchella, A.: NuSMV Version 2: An OpenSource Tool for Symbolic Model Checking. In: Proc. International Conference on Computer-Aided Verification (CAV 2002). LNCS, vol. 2404. Springer, Copenhagen, Denmark (July 2002)
39. Nehaï, Z., Piriou, P.Y., Daumas, F.: Model-checking of smart contracts. In: 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). pp. 980–987 (2018). https://doi.org/10.1109/Cybermatics_2018.2018.00185
40. Lehto, N., Halunen, K., Latvala, O.M., Karinsalo, A., Salonen, J.: CryptoVault - a secure hardware wallet for decentralized key management. IEEE International Conference on Omni-Layer Intelligent Systems (COINS) pp. 1–4 (2021)
41. Babel, S.: Bringing data minimization to digital wallets at scale with general-purpose zero-knowledge proofs. arXiv:2301.00823 (2023)
42. Zhang, Li, L.: Distributed blockchain-based data protection framework for modern power systems against cyber-physical attacks. IEEE Transactions on Smart Grid **11**(4), 3130–3142 (2020)
43. Halpin, H., Piekarska, M.: Introduction to security and privacy on the blockchain. In: 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). IEEE (2017)
44. Bayle, A., Koscina, M., Manset, D., Perez-Kempner, O.: When blockchain meets the right to be forgotten: technology versus law in the healthcare industry. In: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). pp. 788–792. IEEE (2018)
45. Li, X., Jiang, P., Chen, T., Luo, X., Wen, Q.: A survey on the security of blockchain systems. Future Generation Computer Systems **107**, 841–853 (2020)
46. Feng, Q., He, D., Zeadally, S., Khan, M.K., Kumar, N.: A survey on privacy protection in blockchain system. Journal of Network and Computer Applications **126**, 45–58 (2019)
47. Bacon, J., Michels, J.D., Millard, C., Singh, J.: Blockchain demystified. Queen Mary University of London, School of Law Legal Studies Research Paper no. 268 (2017)
48. Kondova, G., Erbguth, J.: Self-sovereign identity on public blockchains and the GDPR. In: Proceedings of the 35th Annual ACM Symposium on Applied Computing. pp. 342–345 (2020)

49. Manteghi, M.: Blockchain and the European Union's General Data Protection Regulation: from conflict to "peaceful" coexistence? http://dx.doi.org/10.2139/ssrn.3805647 (2021)
50. Tatar, U., Gokce, Y., Nussbaum, B.: Law versus technology: Blockchain, GDPR, and tough tradeoffs. Computer Law & Security Review **38**, Art. 105454 (2020)
51. Finck, M.: Blockchain and data protection in the European Union. European Data Protection Law Review **4**(1), 17–35 (2018)