Yes, I Do: Marrying Blockchain Applications with GDPR

Benjamin Schellinger
FIM Research Center, University of Bayreuth
benjamin.schellinger@fim-rc.de

Fabiane Völter
Project Group Business & Information Systems Engineering of the Fraunhofer FIT
fabiane.voelter@fit.fraunhofer.de

Nils Urbach
Frankfurt UAS & Fraunhofer FIT
nils.urbach@fb3.fra-uas.de

Johannes Sedlmeir
University of Bayreuth
johannes.sedlmeir@uni-bayreuth.de

Abstract

Due to blockchains' intrinsic transparency and immutability, blockchain-based applications are challenged by privacy regulations, such as the EU General Data Protection Regulation. Hence, scaling blockchain use cases to production often fails to owe to a lack of compliance with legal constraints. As current research mainly focuses on specific use cases, we aim to offer comprehensive guidance regarding the development of blockchain solutions that comply with privacy regulations. Following the action design research method, we contribute a generic framework and design principles to the research domain. In this context, we also emphasize the need for distinguishing between applications based on blockchains' data integrity and computational integrity guarantees.

1. Introduction

Since its invention in 2009 by Nakamoto [1], blockchain technology has been widely attested large potentials in facilitating or improving inter-organizational digital workflows [2]. This is mainly due to the transparency established between various actors: Instead of placing trust in one distinguished entity, every party within an ecosystem is given access to the underlying data and can engage in its verification. The immutability of the underlying data within a blockchain is often a desirable feature, since this facilitates auditability even among mutually distrusting parties. However, the benefits of information exposure for transparency must be carefully traded off with related challenges. In specific, using blockchain technology for data processing often violates requirements derived from privacy regulation. For example, the European Union (EU)'s General Data Protection Regulation (GDPR) restricts access to personal data to legitimate parties for clearly defined goals only ("purpose limitation"). Furthermore, the GDPR conflicts with blockchains' immutability since it gives subjects a right to rectification (Art. 16) and erasure (Art. 17) of their data, meaning they can withdraw their consent to the processing of their data at any time [3]. These challenges of blockchain technology in the context of privacy regulation negatively affect its adoption. For example, GDPR compliance was identified as a pressing issue for blockchain projects in the EU Blockchain Observatory & Forum report [4]. However, the neglection of regulatory aspects during prototyping often hinders later scaling to productive systems. Sağlam et al. [5] demonstrate that more than two thirds of blockchain projects do not communicate about GDPR at all. Consequently, Haque et al. [6] point out the necessity of further investigating how blockchain solutions can comply with privacy by design.

So far, several approaches have been proposed to meet regulatory compliance in blockchain-based applications. For example, some researchers recommend to build upon private permissioned blockchain implementations that limit data visibility to registered participants. However, these solutions are criticized for failing to deliver advantages in comparison to conventional, centralized information systems [7]. Moreover, while making personal data available to only a few selected nodes certainly mitigates privacy issues to some extent, conflicts with GDPR's purpose limitation and right to erasure remain. To address the right to erasure, backdoors have been proposed to make blockchains redactable [8, 9]; however, it has been argued that such approaches can severely affect the security of blockchain implementations [10].

Against this backdrop, Zemler and Westner [11] call for a comprehensive framework for the development of GDPR-compliant blockchain solutions. Meanwhile, cryptographic tools that allow to meet blockchain

URI: https://hdl.handle.net/10125/79900
978-0-9981331-5-7

technology's original aims of independent verification among various stakeholders while addressing privacy requirements are increasingly used. For example, storing hashes of data facilitates immutability proofs without storing sensitive information, and Zero-Knowledge Proofs (ZKPs) allow for the on-chain verification of off-chain computations [12]. This approach allows to prove computational integrity while data itself does not need to be disclosed. However, so far, research and practice lack knowledge on how these mechanisms can address GDPR-related requirements. Thus, we ask the research question: *How can blockchain technology and data protection under the GDPR be reconciled to promote the adoption of blockchain applications?*

We address this research question by proposing a framework for the development of GDPR-compliant blockchain-based solutions. We follow an Action Design Research (ADR) approach and develop our framework based on both insights gained from a systematic literature review and from focus groups with practitioners. In specific, we derive our insights from the context of the energy sector. The underlying use case aims to apply blockchain technology for enabling the verifiability of asset logging information and the authenticity of guarantees of origin for electricity with the help of a blockchain. Integrating insights from both literature and practice enables us to derive design principles that help reconcile data protection regulation with blockchain-based solutions by design.

The remainder of this paper is structured as follows. In Section 2, we introduce blockchain technology, selected privacy-related cryptographic methods, and the GDPR. We present our methods in Section 3. Section 4 outlines the development process of our framework during the design cycles. We the present and describe our framework in Section 5 and discuss our findings and conclude in Section 6.

2. Background

2.1. Blockchain

Blockchain technology initially drew attention with its first application, Bitcoin, a concept for a decentralized digital currency system [1]. The underlying technology of Bitcoin solves the immanent problem of preventing the double-spending of digital assets in a decentralized system by applying concepts formerly proposed by Chaum [13] and Back [14] and, thus, allows for trust-free cryptographic transactions in a network without any distinguished central entity [15, 16]. As such, a blockchain can be described as a publicly available immutable registry that stores data in a linked, append-only list of blocks [16, 17]. Since no central entity is involved, the network needs to agree on which blocks to include. For this purpose, blockchain systems use so-called consensus mechanisms, e.g., Proof of Work (PoW) or Proof of Stake (PoS), to find agreement in a decentralized system and, thus, to secure the network from attacks [17, 18].

One of blockchains' core features is immutability, meaning that retrospectively tampering with data written to the ledger is hard or practically impossible. For example, to gain control of the network in PoW systems, large amounts of hashing power and, thus, energy and hardware expenditures are required [17, 18]. Furthermore, blockchains can have a tailor-made design subject to specified parameters that affect access, participation and governance. Bitcoin and Ethereum, for instance, comprise permissionless blockchains that are open to everyone, while blockchains used in consortia are typically permissioned, meaning that only authorized members can join the network and participate in consensus [15, 19]. Consequently, different roles and rights, e.g., for writing, reading, updating, validating, or deleting data, arise. In addition, a blockchain can be either public or private. In public blockchains, the data is exposed to everyone while in a private blockchain the records are only visible to the authorized entities participating on the blockchain [17].

Blockchain systems are not limited to store transaction data only. So-called smart contracts incorporate programming logic on a blockchain [20, 21], and if invoked, the corresponding methods are executed by all blockchain nodes [16, 21]. Therefore, smart contracts provide new opportunities for the automated processing of data in decentralized applications, e.g., in supply chain management or e-government [19, 22].

2.2. Merkle Trees and Zero-Knowledge Proofs

As pointed out, in any design, the list of transactions is visible to every network participant, thus allowing for the replicated verification of transactions [17]. This intrinsic visibility of transaction data immediately results in significant challenges related to the processing and storage of sensitive information [23]. Consequently, blockchain system are often supplemented by cryptographic techniques to hide information that is not necessary to be disclosed for transaction validation and business processes that aim to be managed. Probably the simplest way is to only write hashes of data to a blockchain to allow for latter proofs that data has not been manipulated, without disclosing sensitive

information on-chain. In this context, Merkle trees [24] can be used to improve privacy and efficiency. A Merkle tree is created by recursive pairwise hashing; only the root is then stored on-chain. One can then prove that data was included at the point of tree generation by providing the correct path and its adjacent hashes that lead to the corresponding Merkle root [24]. This approach allows to represent large datasets by a single on-chain hash and facilitates the efficient verification of parts of this dataset without revealing further, potentially sensitive information in a "Merkle proof" [25].

However, most operations, like conducting payments and invoking smart contracts, require operations beyond the checking the integrity of data. In this context, ZKPs have received increasing attention and usage in blockchain ecosystems, for example to add a privacy-preserving layer for digital identities [26] or distributed payment systems [27]. ZKPs are cryptographic protocols that convince a verifier that a (mathematical) statement about data is correct without revealing the information itself [12]. The prover only provides information required to assess the correctness of the statement via a proof but does not have to disclose any (potentially confidential) additional information [28, 29]. In the context of blockchains, ZKPs enable data parsimony as the proof can be verified succinctly on-chain while sensitive data is protected and only stored off-chain [30]. For example, Zcash's protocol relies on ZKPs to prove the legitimacy of a transaction while hiding information about the transaction's sender, receiver, and amount [31].

2.3. General Data Protection Regulation

In 2018, the EU passed the GDPR for the protection of natural persons concerning the processing of their personal data by organizations. The GDPR not only harmonizes data privacy laws across European member states but also applies to all European citizens, irrespective of where the data collection and processing takes place [32]. According to Art. 1 GDPR, the imposed law aims at protecting "fundamental rights and freedoms of natural persons and their right to the protection of personal data". In particular, the GDPR lists principles relating to the processing of personal data, defines data subjects' rights, and specifies the provisions of a data controller and processor. In general, the data subject that can grant permission to process personal data can also revoke its consent at any time (see Art. 7 GDPR). Notably, Art. 6, para. 1 b) to f) GDPR list certain cases in which the data processing does not require an explicit permission. In general, the processing of sensitive data generally is prohibited (see Art. 9 para. 1 GDPR) except for certain special cases (see Art. 9 para. 2 GDPR). According to Art. 16 GDPR, the data subject always has the right to have inaccurate personal data rectified. In addition, the data subject can exercise his or her right to demand the erasure of personal data against the data controller at any time (Art. 17 GDPR). Both the right to rectification and the right to erasure of personal data are generally considered the most challenging issues in promoting blockchain use cases [33]. In addition, Haque et al. [6] outline conflicts regarding responsibilities of controllers and processors (see Art. 24, 26, 28, GDPR), as well as the technology's territorial scope (see Art. 3 GDPR). The distribution of personal information to third parties in a blockchain may also conflict with GDPR's purpose limitation, data minimization, storage limitation, and confidentiality [34]. For the reconciliation of blockchain with the strict requirements of the GDPR, e.g., the right of rectification and erasure of personal data, we present current findings of the relevant literature in Section 4. We also want to point out that GDPR is not the only data protection regulation that challenges blockchain applications; for example, the California Consumer Privacy Act (CCPA) and the recent Chinese Personal Information Protection Law have similar objectives. However, the GDPR is generally acknowledged as particularly strict, which is the main reason why most research focuses on GDPR and assumes that a GDPR-compliant blockchain solutions will also address other privacy regulations' requirements.

3. Research Design

To answer our research question, we take a pluralistic approach to obtain rich and reliable results, thus expanding the literature on blockchain and data protection in a rigorous way [35]. In specific, we follow an ADR approach [36] to develop a framework that supports organizations in aligning their blockchain-based applications with GDPR and, thus, to facilitate adoption. Proposed by Sein et al. [36], ADR supports in the creation of innovative artifacts like methods or constructs that aim to solve organizational problems [37, 38]. In total, the research process involves four stages and seven principles applied by both practitioners (e.g., individuals with first-hand experience or end-users) and researchers. We present all four stages of the ADR process in more detail in Section 4.

Motivated by the problems identified in the ADR process, we additionally conducted a systematic literature review following the guidelines of Webster and Watson [39] and vom Brocke et al. [40]. By

systematically reviewing the relevant literature on blockchain and GDPR, we provide a solid foundation for promoting knowledge, thereby enhancing theory development and identifying areas that need to be explored in this domain [40]. We present the results of the systematic literature review in the problem formulation's stage of Section 4.

4. Development of the Framework

We now describe the development of our generic framework to guide practitioners in their initial design and development process to create a blockchain-based application that complies with the GDPR requirements by design in the context of the ADR approach.

Stage 1: Problem formulation The first stage aims to identify and conceptualize a problem based on insights from practice or research. Thus, both "practice-inspired research" and "theory-ingrained artifact research" support the mutual understanding of the research aim [36]. We first encountered the underlying problem in a practical environment in the energy industry. In various one-on-one discussions and workshops, we identified that both energy and law practitioners considered compliance with GDPR as a major challenge for adopting blockchain technology. Additionally, domain-specific regulations, such as the German Metering Point Operation Act [41], complicate the use of blockchain technology. When investigating privacy-preserving solutions, we found that many projects already lever basic hashing functionalities, but the use of more advanced cryptographic methods including Merkle trees and ZKPs as well as knowledge about their benefits and limitations so far is rare. In specific, we conducted interviews with leaders of twelve selected research projects that address similar use cases and explore the opportunities of blockchain technology in the energy sector [blinded for review]. We found that all interviewees were aware of privacy protection being a significant hurdle for implementing a blockchain-based solution that complies with regulation. Still, most described the legal situation as unclear. Therefore, their research focused on the technical feasibility and postponed regulatory considerations or intended to ensure GDPR-compliance through individual data usage contracts (although this approach is highly questionable because consent can still be withdrawn). Nevertheless, certain technical solutions were applied, including off-chain data management or permissioned blockchain environments; often without being aware of the related tradeoffs indicated in Section 1. Accordingly, we defined the problem as "lack of a systematic approach on how blockchain technology and data protection under GDPR can be reconciled".

In addition to the insights gained from practice, we substantiated the perceived need for systematic knowledge on the reconciliation of blockchain technology and GDPR by systematically reviewing existing literature. We identified that the compatibility of blockchain technology and GDPR represents a prevailing problem to the blockchain research community, too. Our findings also allowed us to structure existing solution approaches regarding the reconciliation of blockchain technology and GDPR. We derived our search string applied to articles' titles from the main keywords of our research question: ("blockchain" OR "distributed ledger") AND ("GDPR" OR "general data protection regulation"). Our initial search yielded 126 results including articles from the databases ACM Digital Library (2), AIS eLibrary (4), Google Scholar (94), IEEE Xplore (10), ScienceDirect (2), and Web of Science (14) until 12th of March 2021. Following vom Brocke et al. [40], we then defined inclusion and exclusion criteria. We included only peer-reviewed scientific articles published in English, which resulted in a data set of 66 articles. We also removed duplicates and excluded articles not relevant for our underlying research objective, i.e., those without an explicit focus on the compliance of blockchain applications with GDPR. Applying our exclusion criteria resulted in 17 papers in total for full-text reading.

Our results indicate that GDPR-compliance and blockchain technology represents a relevant issue for both computer science and IS scholars. From our data set of 17 papers, nine articles take a technical perspective and propose blockchain architectures adhering to the GDPR. For example, Farshid et al. [41] proposes a prototype for the financial industry that deletes data after a pre-defined time range. Also, Dauden-Esmel et al. [42] and Truong et al. [43] propose a GDPR-compliant personal data management platform. Moreover, Precht and Marx Gomez [44] develop a prototype that enforces joint controllership agreements before any data is processed. Generally, Rieger et al. [33] and Guggenmos et al. [45] conclude that personal data should not be stored on a blockchain. The authors propose to use an off-chain mapping architecture when the use case requires personal data storage, yet questions around the management of this mapping remains open. Regarding legal assessments of blockchain and GDPR, Poelman and Iqbal [7] identified a dispute between purist and fundamentalist approaches. While purists are convinced that blockchain and GDPR

cannot be reconciled as any adjustments break with original principles, fundamentalists believe in technical adaptions of the technology to fulfill regulatory aspects. However, the latter often includes redactable and mutable blockchains relying on backdoor solutions like chameleon hashes [46], which undermine trust and security [10] or are not applicable to permissionless blockchains. Many of these approaches are also not applicable to existing, popular blockchains like Ethereum but would require the development of entirely new ecosystems. They also do not discuss issues beyond the right to erasure described in Section 2.3, for instance, purpose limitation.

Thus, the results from our systematic literature review confirm the relevance of the identified problem. Most research exclusively focuses on redactability to address the right to reasure or considers specific domains or use cases, e.g. traffic management [47], e-government [33], or finance [42]. In total, previous literature on blockchain and GDPR reveals only three general design principles: Rieger et al. [33] recommend avoiding personal data storage on blockchain, using private and permissioned pseudonymization approaches to process personal data, and implementing off-chain identifier mappings for the coordination of cross-organizational workflows. We also found that authors taking a legal perspective often state that blockchains need to be tailored to conform to the GDPR, e.g., [3]. Sağlam et al. [5] find that most blockchain projects do not acknowledge legal challenges of blockchain-based solutions and the GDPR and highlight the urgency for research on data protection regulations and blockchain technology. Against this backdrop, also Zemler and Westner [11] call for frameworks for GDPR compliant processing of personal data using blockchain technology.

Both the survey of industry experts and the literature review yield that merely recording encrypted or hashed data on a blockchain does not necessarily prevent from violating data protection. For example, repeatedly referring to a hash on a blockchain can make it personally identifiable. Moreover, storing encrypted or hashed data on-chain makes it useless for the smart contracts observed as they cannot perform useful computations on obfuscated data (we have not encountered the use of cryptographic tools that would allow to do so, such as homomorphic encryption, in these publications; probably for their computational complexity). We therefore examined further techniques and in particular Merkle proofs and ZKPs. The latter are already being used in the context of cryptocurrencies and allow sensitive data to be stored and processed in a fully anonymized and, thus, data-protection-compliant manner. We conclude that the underlying problem is of high relevance both for practice and theory (Principle 1). Following Principle 2, we aim to solve the defined problem by creating knowledge that is transferable to similar contexts [36]. Thus, we aspire to create a generic framework for data protection compliance on blockchain-based applications that can be used across various domains. In specific, we aim to enable researchers and practitioners to evaluate the applicability of anonymization techniques according to their underlying objective.

#	Role	Expertise	Organization
1	Project lead	Energy & blockchain	Research ass.
2	Project member	Energy & blockchain	Research ass.
3	Researcher	Law & energy	Foundation
4	Researcher	Law & energy	Foundation
5	Researcher	Law & energy	Foundation
6	Project lead	Law & energy	Foundation
7	Attorney	Law	Law firm
8	Researcher	Blockchain	Research inst.

Table 1. Interview partners

Stage 2: Building, intervention, and evaluation In the second stage, we follow the generic scheme for the organization-dominant development of artifacts to generate design knowledge, since our primary source of innovation is organizationally driven [36]. This approach guided the iterative design and development of our artifact in multiple intervention and evaluation cycles to create knowledge that can be transferred to similar problems (Principles 3, 4, and 5). The alpha version of our framework was concurrently challenged by the assumptions, expectations, and knowledge of practitioners, including legal experts in the energy domain (see Table 1). In this light, we conducted unstructured interviews, open discussion rounds, and multiple workshops in an organizational environment for a project to build a blockchain platform in the energy industry. The interview partners 1–6 (see Table 1) were working with with the authors on a blockchain-based project in the energy sector. The various forms of consultation with domain experts were performed iteratively in order to continuously improve and evaluate the framework.

After this first iteration cycle, the beta version was challenged by experts from the blockchain and legal domain (see Table 1). To refine our artifact, we further conducted workshops with practitioners and dedicated experts in the fields of GDPR. These interventions and evaluations resulted in a framework generalizable to other domains. Prior to generalizability and thus transferability to similar problems in other industries,

the framework lacked certain key elements: First, it required further information on the stage that data can be considered sufficiently anonymized. Second, the possibility of data triangulation and correlation, i.e., identifying an individual based on metadata that third parties outside the blockchain may possess (e.g., location coordinates or energy consumption), needed to be integrated into the framework. Third, the applicability of the GDPR and other relevant legislations (e.g., Art. 50 Metering Point Operation Act in the energy industry) regarding the processing of personal data was initially reflected in a unified, undifferentiated model. Fourth, the restriction of Art. 16 and 17 of the GDPR, the right to rectification and erasure, was not explicitly addressed. Fifth, the framework lacked a stringent sectioning regarding the different steps involved. Sixth, the aspect of alternative options of processing personal data when anonymization is not possible needed to be clarified and explained in more detail.

Stage 3: Reflection and learning In parallel to these stages, learnings should be generalized "to a broader class of problems" (p. 44) in stage three [36]. As outlined in the previous stage, we first developed an industry-specific framework that incorporates requirements from both the energy sector and the GDPR. To generalize from the industry specific context and to address the first iterations' shortcomings, we refined our artifact by making substantial changes to the design, meta-design, and meta requirements [36, 48, 49]. The iterative discussion with practitioners, law experts, and researchers represented an integral part of continuous reflection and learning. In this way, we received valuable feedback to adapt our research approach and to critically reflect on the current state of the framework, including its shortcomings. Overall, we refined the artifact in this stage (Principle 6). Against this backdrop, we developed an artifact supporting end-users to assess the reconcilability of blockchain and the GDPR beyond the energy industry context and thus promote the implementation of blockchain applications holistically (see Figure 1).

Stage 4: Formalization of Learning On this basis, learnings and reflections can be formalized for a range of field problems in the form of design principles [36]. By developing an industry-independent framework, we have iteratively formulated design knowledge on the reconciliation of GDPR and blockchain technology. Originally stemming from the context of energy industries, we claim generalizability by incorporating knowledge and evaluations from legal and blockchain experts working in other sectors. We fulfill Principle 7 of the ADR not only by generalizing both the problem and the solution, but also by discussing the literature and deriving design principles. In this light, we present four design principles (see Section 6) to extend the design theory for blockchain and guide the process of developing blockchain applications to meet the requirements of the GDPR by design.

5. Presentation of the GDPR Framework

The framework presented in this section aims to provide guidance for practitioners and researchers in the development of GDPR-compliant blockchain solutions. The purpose is to provide this guidance based on initial questions about the underlying use case and its data processing requirements. On this basis, a design recommendation is given. In particular, the contribution of our framework resides in the distinction between *data integrity* and *computational integrity*, which has not yet been made explicit in the literature we reviewed. In the following, we describe the three steps of our artifact for reconciling blockchain with GDPR.

Verification of personal data In the first step, it is determined whether the GDPR is applicable. Furthermore, the user of the framework needs to assess whether the data that is to be stored on the blockchain is cryptographically obfuscated, i.e., the data is encrypted or hashed and hence is neither readable by the human eye nor can be interpreted by simple computer programs. If data is obfuscated, it is important to assess to what extent the data can be revocered using the technical means available today. In determining whether technical means are reasonably likely to be used to identify the natural person, all objective factors, such as the cost of identification, metadata or additional data from other contexts needed, the time required to do so, and the technology available at the time of processing should be considered (see also recital 26 of the GDPR). Obfuscated personal data is not considered personal data only if completely irreversible anonymization has been carried out (thick arrow line in Figure 1). Although the GDPR is not applicable in such cases, it is important to examine further applicable laws that may be relevant for the blockchain use case. In certain scenarios, a specific kind of data is processed and requires a thorough review of industry-specific laws before data can be recorded on-chain.

If the data at hand is not cryptographically obfuscated, the question of persons' direct identification arises. This could also apply if someone successfully

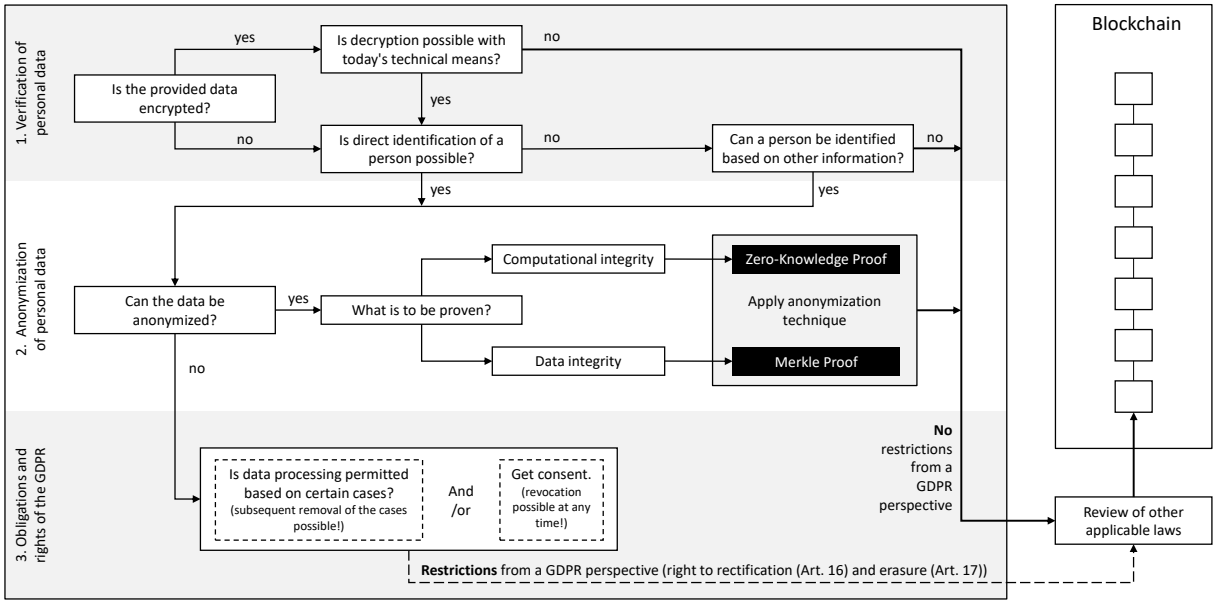

Figure 1. Generic framework for processing personal data in the scope of the GDPR

decrypts the data with negligible computing power, e.g., if a weak encryption technique, a non-cryptographic hashing algorithm, or too little entropy was used to initially obfuscate the data. Against this backdrop, the direct identification of a person is always assumed if the information available allows the natural person behind the data to be identified without further ado. In the energy sector, this would be the case with habit-related and, thus, correlatable information, such as unencrypted consumption data, comfort requirements or information on willingness to pay [50, 51]. Also potentially increasing computational power in the future or disruptive decryption possibilities with quantum computers should be considered by taking into account whether the information will still be personally identifiable in the future. Also data that has been de-personalized to a certain extent needs to be regarded as personal data within the meaning of the GDPR if it can be used to infer the identity of the person with the assistance of other information, provided that the effort required for identification is not unreasonably high. When pseudonymized or non-personal data is being combined to derive the identity of a person, this is described as "data triangulation" or the "bundling problem" [52]. This can be caused, e.g., by linking geographic data and the individual consumption of a consumer to identify a person.

Anonymization of personal data In the second step, personal data is being anonymized. In any of the scenarios in which it is possible to draw conclusions about the identity of a natural person, the question arises as to whether data can be anonymized in the first place. In certain scenarios, data processing requirements do not allow for the anonymization of personal data. In addition, anonymization is hindered when there is a lack of skills or resources to implement these techniques. However, in most cases data can be anonymized through obfuscation and involves the question of what exactly the obfuscated data is needed for. Both the specific requirements and objectives of the blockchain use case determine the anonymization technique that will be applied. If the use case requires the verification of correct calculations (computational integrity), such as balances, coordinates, or signatures, we recommend using ZKPs. It follows that only a (typically succinct) proof of the statement is provided on-chain, potentially including additional obfuscated input data, e.g., a Merkle root that commits to inputs or outputs of the computation. If, on the other hand, anonymized data on the blockchain should only ensure that the data has not been altered in any way (data integrity), Merkle proofs are appropriate. In this case, only the Merkle root is stored on-chain. Using ZKPs or Merkle proofs and ensuring that the data is sufficiently obfuscated so that it becomes technically impossible to reveal the identity of individuals can achieve a privacy-preserving record of data on the blockchain that does not fall within the scope of the GDPR. As mentioned above, additionally, reviews of industry-specific requirements regarding the

processing of personal data should be conducted.

Obligations and right of the GDPR In the third step, obligations, and rights in the context of the GDPR are examined. While it may be possible that in certain events personal data might not be anonymizable, the GDPR highlights several exemptions, which allow for the processing of this data anyhow. Part of these exemptions are a list of certain cases in which the processing of personal data is permitted, e.g., when it is necessary in the fulfillment of a contract, pre-contractual measures or when there is a legal obligation (also see Art. 6, para. 1 b) to f) of the GDPR). In addition, there are ways for data processing under national law, which is permitted in certain cases. Thus, beyond the applicability of the GDPR, the end-user should consider that other cases may need to be acknowledged, too.

6. Discussion and conclusion

By providing a framework for the GDPR-compliant processing of personal data in blockchains-based architectures, we are following the call by Zemler and Westner [11] and Sağlam et al. [5] to contribute to the theory in this research domain [33, 42]. Our framework can help practitioners assess the specific requirements related to their blockchain use case imposed by the GDPR and take measures to implement GDPR compliance by design. Following the ADR method also allows us to derive four key learnings, which serve as Design Principles (DPs) for reconciling data protection compliance on blockchain-based platforms that can be used across various domains. These DPs should be considered as guidelines to establish GDPR-compliant blockchain-based architectures by design.

DP1: Acknowledge GDPR compliance by design: Consider requirements of GDPR throughout the whole development cycle of the blockchain prototype.

To achieve GDPR compliance of blockchain-based solutions, data protection requirements should be considered from the very beginning of the development cycle. This ensures that compliance of GDPR is achieved by design. For example, if researchers and practitioners notice conflicts with the GDPR as data cannot be properly anonymized, cryptographic techniques like ZKPs can be integrated in architectural considerations from the very beginning.

DP2: Use state-of-the-art cryptography: Ensure to always use latest but established cryptographic techniques for hiding personal data.

Secondly, we recommend to always use the latest cryptographic techniques to prevent security issues owing to outdated or broken cryptography. On the other hand, cryptographic mechanisms need to be sufficiently established to prevent incidents as occurred recently in IOTA [51]. This DP also underlines that we do not claim the comprehensiveness of our framework for all future developments. As both regulation and cryptographic techniques and methods change at fast pace, the frameworks need to account for recent and expected future developments.

DP3: Differentiate aims of data processing: Different anonymization techniques allow to prove data or computational integrity.

Thirdly, our research shows that the underlying use case should determine the choice of appropriate anonymization techniques. Generally, we distinguish between proving the integrity of historic data and proving the integrity of computations performed on data. Based on this distinction, appropriate anonymization techniques may be chosen that prove integrity but can be reconciled with the GDPR. In specific, most blockchain applications struggle not only with Art. 16 and 17 GDPR but also with the initial distribution of potentially sensitive information. In many applications, these issues can be overcome from a legal perspective using the techniques mentioned above as only non-sensitive information is shared. Thus, we encourage researchers and practitioners to differentiate their data processing aims before developing a blockchain-based solution.

DP4: Review all relevant laws: Do not only evaluate the reconciliation of GDPR but also further industry-specific laws.

Lastly, our research highlighted that the GDPR is not the only relevant law in the context of legal compliance of blockchain applications. As GDPR represents a domain-independent regulation for data processing, industry-specific laws must be evaluated as well. For example, the German Metering Point Operation Act further regulates data processing in the energy sector. Only if domain-specific laws are taken into consideration, fully compliant solutions can be promoted. Information exposure can also be problematic beyond personal data, for example, antitrust regulation restricts the sharing of business secrets with competitors.

To conclude, our field research and literature review highlighted that the exposure of sensitive information in general and GDPR-compliance in particular represent a pressing challenge for productive blockchain applications. Against this backdrop, we contribute a generic framework that supports the assessment and creation of GDPR-compliant blockchain applications. To the best of our knowledge, our framework is the first that distinguishes between

data integrity and computational integrity. Moreover, we derive design principles for the development of compliant blockchain-based solutions. Our framework helps practitioners to assess the implications of the data at hand to be used in the blockchain use case from a GDPR perspective. Depending on the use case, different anonymization techniques can enable GDPR compliance by design. The design principles that we developed aim to guide practitioners to design and develop blockchain applications that are GDPR-compliant by design.

Although having pursued a rigorous research approach, we acknowledge limitations of our study. While Haque et al. [6] provide a comprehensive overview of compliance issues, we mainly focused on the right of rectification and erasure of personal data and only briefly reflected on additional requirements like purpose limitation. Thus, we motivate future researchers to investigate further compliance issues in a detailed manner. In addition, while the GDPR represents one of the most comprehensive privacy legislations [53], we expect that our results also apply to other regulations such as the CCPA. Also, the proposed framework has been validated by practitioners in the energy sector and legal experts successfully applied to create a blockchain prototype that reconciles GDPR by design. Most importantly, due to the fast-paced development of the research field, we do not claim comprehensiveness of the covered anonymization techniques. While the framework takes into account Merkle trees and ZKPs, we motivate future researchers to integrate further developments and standards for secure computation (e.g., multi-party computation and homomorphic encryption) and methods for statistic information disclosure (e.g., differential privacy) [52].

References

[1] S. Nakamoto, "A peer-to-peer electronic cash system," 2008.

[2] G. Fridgen, S. Radszuwill, N. Urbach, and L. Utz, "Cross-organizational workflow management using blockchain technology – Towards applicability, auditability, and automation," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 3507–3517, 2018.

[3] U. Tatar, Y. Gokce, and B. Nussbaum, "Law versus technology: Blockchain, GDPR, and tough tradeoffs," *Computer Law & Security Review*, vol. 38, 2020.

[4] European Union Blockchain Observatory & Forum, "Blockchain and the GDPR," 2018.

[5] R. B. Sağlam, Ç. B. Aslan, S. Li, L. Dickson, and G. Pogrebna, "A data-driven analysis of blockchain systems' public online communications on GDPR," in *International Conference on Decentralized Applications and Infrastructures*, pp. 22–31, IEEE, 2020.

[6] A. B. Haque, A. K. M. N. Islam, S. Hyrynsalmi, B. Naqvi, and K. Smolander, "GDPR Compliant Blockchains–A Systematic Literature Review," *IEEE Access*, vol. 9, pp. 50593–50606, 2021.

[7] M. Poelman and S. Iqbal, "Investigating the compliance of the GDPR: Processing personal data on a blockchain," in *5th International Conference on Cryptography, Security and Privacy*, pp. 38–44, IEEE, 2021.

[8] D. Deuber, B. Magri, and S. A. K. Thyagarajan, "Redactable blockchain in the permissionless setting," in *Symposium on Security and Privacy*, pp. 124–138, IEEE, 2019.

[9] M. Florian, S. Henningsen, S. Beaucamp, and B. Scheuermann, "Erasing data from blockchain nodes," in *European Symposium on Security and Privacy Workshops*, pp. 367–376, IEEE, 2019.

[10] M. Kuperberg, "Towards enabling deletion in append-only blockchains to support data growth management and GDPR compliance," in *International Conference on Blockchain*, pp. 393–400, IEEE, 2020.

[11] F. Zemler and M. Westner, "Blockchain and GDPR: Application scenarios and compliance requirements," in *Portland International Conference on Management of Engineering and Technology*, IEEE, 2019.

[12] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM Journal on Computing*, vol. 18, no. 1, pp. 186–208, 1989.

[13] D. Chaum, "Blind signatures for untraceable payments," in *Advances in Cryptology*, pp. 199–203, Springer, 1983.

[14] A. Back, "Hashcash – a denial of service counter-measure," 2002.

[15] F. X. Olleros and M. Zhegu, *Research handbook on digital transformations*. Edward Elgar Publishing, 2016.

[16] R. Beck, J. Stenum Czepluch, N. Lollike, and S. Malone, "Blockchain – the gateway to trust-free cryptographic transactions," in *24th European Conference on Information Systems*, 2016.

[17] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *International Congress on Big Data*, pp. 557–564, IEEE, 2017.

[18] J. Sedlmeir, H. U. Buhl, G. Fridgen, and R. Keller, "The energy consumption of blockchain technology: Beyond myth," *Business & Information Systems Engineering*, vol. 62, no. 6, pp. 599–608, 2020.

[19] K. Wüst and A. Gervais, "Do you need a blockchain?," in *Crypto Valley Conference on Blockchain Technology*, pp. 45–54, IEEE, 2018.

[20] N. Szabo, "Formalizing and securing relationships on public networks," *First Monday*, vol. 2, no. 9, 1997.

[21] V. Buterin, "A next-generation smart contract and decentralized application platform," 2013.

[22] J. Amend, J. Kaiser, L. Uhlig, N. Urbach, and F. Völter, "What do we really need? A systematic literature review of the requirements for blockchain-based e-government services," in *16th International Conference on Wirtschaftsinformatik*, 2021.

[23] M. Platt, R. J. Bandara, A.-E. Drăgnoiu, and S. Krishnamoorthy, "Information privacy in decentralized applications," in *Trust Models for Next-Generation Blockchain Ecosystems*, Springer, 2021.

[24] R. C. Merkle, "A digital signature based on a conventional encryption function," in *Conference on the Theory and Application of Cryptographic Techniques*, pp. 369–378, 1987.

[25] A. Djamali, P. Dossow, M. Hinterstocker, B. Schellinger, J. Sedlmeir, F. Völter, and L. Willburger, "Asset logging in the energy sector: A scalable blockchain-based data platform," in *The 10th DACH+ Conference on Energy Informatics*, Springer, 2021.

[26] Q. Stokkink and J. Pouwelse, "Deployment of a blockchain-based self-sovereign identity," in *International Conference on Internet of Things and Green Computing and Communications and Cyber, Physical and Social Computing and Smart Data*, pp. 1336–1342, IEEE, 2018.

[27] P. Chatzigiannis, F. Baldimtsi, and K. Chalkias, "SoK: Auditability and accountability in distributed payment systems," in *International Conference on Applied Cryptography and Network Security*, pp. 311–337, Springer, 2021.

[28] E. Ben-Sasson, A. Chiesa, D. Genkin, E. Tromer, and M. Virza, "SNARKs for C: Verifying program executions succinctly and in zero knowledge," in *Annual Cryptology Conference*, pp. 90–108, Springer, 2013.

[29] E. Ben-Sasson, I. Bentov, Y. Horesh, and M. Riabzev, "Scalable zero knowledge with no trusted setup," in *Annual International Cryptology Conference*, pp. 701–732, Springer, 2019.

[30] J. Partala, T. H. Nguyen, and S. Pirttikangas, "Non-interactive zero-knowledge for blockchain: A survey," *IEEE Access*, vol. 8, pp. 227945–227961, 2020.

[31] A. Biryukov and D. Feher, "Privacy and linkability of mining in Zcash," in *Conference on Communications and Network Security*, pp. 118–123, IEEE, 2019.

[32] European Union, "General data protection regulation (GDPR)," 2016.

[33] A. Rieger, F. Guggenmos, J. Lockl, G. Fridgen, and N. Urbach, "Building a blockchain application that complies with the EU general data protection regulation," *MIS Quarterly Executive*, vol. 18, no. 4, pp. 263–279, 2019.

[34] European Parliamentary Research Service, "Blockchain and the General Data Protection Regulation – Can distributed ledgers be squared with European data protection law?," 2019.

[35] J. Mingers, "Combining IS research methods: Towards a pluralist methodology," *Information systems research*, vol. 12, no. 3, pp. 240–259, 2001.

[36] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, "Action design research," *MIS Quarterly*, vol. 35, no. 1, pp. 37–56, 2011.

[37] S. Gregor and A. R. Hevner, "Positioning and presenting design science research for maximum impact," *MIS Quarterly*, vol. 37, pp. 337–355, 2013.

[38] S. T. March and G. F. Smith, "Design and natural science research on information technology," *Decision Support Systems*, vol. 15, no. 4, pp. 251–266, 1995.

[39] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002.

[40] J. vom Brocke, A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, "Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research," *Communications of the Association for Information Systems*, vol. 37, no. 1, 2015.

[41] S. Farshid, A. Reitz, and 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*, pp. 7087–7095, 2019.

[42] C. Daudén-Esmel, J. Castellà-Roca, A. Viejo, and J. Domingo-Ferrer, "Lightweight blockchain-based platform for GDPR-compliant personal data management," in *5th International Conference on Cryptography, Security and Privacy*, pp. 68–73, IEEE, 2021.

[43] N. B. Truong, K. Sun, G. M. Lee, and Y. Guo, "GDPR-compliant personal data management: A blockchain-based solution," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1746–1761, 2020.

[44] H. Precht and J. Marx Gómez, "Towards GDPR enforcing blockchain systems," in *16th International Conference on Wirtschaftsinformatik*, 2021.

[45] F. Guggenmos, J. Lockl, A. Rieger, A. Wenninger, and 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*, 2020.

[46] G. Ateniese, B. Magri, D. Venturi, and E. Andrade, "Redactable blockchain – or – rewriting history in Bitcoin and friends," in *IEEE European Symposium on Security and Privacy*, pp. 111–126, 2017.

[47] L. Campanile, M. Iacono, F. Marulli, and M. Mastroianni, "Designing a GDPR compliant blockchain-based IoV distributed information tracking system," *Information Processing & Management*, vol. 58, no. 3, 2021.

[48] S. D. Gregor and J. Iivari, "Designing for mutability in information systems artifacts," in *Information Systems Foundations: Theory, Representation and Reality*, pp. 3–24, ANU E Press, 2007.

[49] E. H. Glattfeld and L. Keller-Herder, "Die Datenschutz-Grundverordnung und ihre Umsetzung durch EVU," *ER EnergieRecht*, no. 4, 2018.

[50] M. Köhler and I. Müller-Boysen, "Blockchain und Smart Contracts – Energieversorgung ohne Energieversorger?."

[51] E. Heilman, N. Narula, G. Tanzer, J. Lovejoy, M. Colavita, M. Virza, and T. Dryja, "Cryptanalysis of Curl-P and other attacks on the IOTA cryptocurrency," *IACR Transactions on Symmetric Cryptology*, vol. 2020, pp. 367–391, 2020.

[52] G. Munilla Garrido, J. Sedlmeir, Ö. Uludağ, I. S. Alaoui, A. Luckow, and F. Matthes, "Revealing the landscape of privacy-enhancing technologies in the context of data markets for the IoT: A systematic literature review," 2021.

[53] L. Bari and D. P. O'Neill, "Rethinking patient data privacy in the era of digital health," 2019.