Association for Information Systems

AIS Electronic Library (AISeL)

Wirtschaftsinformatik 2021 Proceedings

Track 12: Information Security, Privacy and Blockchain

# Blockchain and Data Protection: An Evaluation of the Challenges and Solutions mentioned by German Stakeholders

Frank Ebbers
*Fraunhofer Institute for Systems and Innovation Research ISI, Deutschland*

Murat Karaboga
*Fraunhofer Institute for Systems and Innovation Research ISI, Deutschland*

# Blockchain and Data Protection: An Evaluation of the Challenges and Solutions mentioned by German Stakeholders

Frank Ebbers, Murat Karaboga
Fraunhofer Institute for Systems and Innovation Research ISI,
Breslauer Str. 48, 76139 Karlsruhe, Germany
{firstname.lastname}@isi.fraunhofer.de

**Abstract.** This paper analyzes data protection challenges and possible solutions associated with the usage of the blockchain (BC) technology from the perspective of 94 German companies and organizations. This paper clusters 537 data protection-relevant statements into three subject areas: (1) relevance of data protection in BC, (2) articulated challenges and (3) proposed solutions. Each group is then collated with insights from computer science. The results show that a majority of the respondents do see data protection issues with using BC, which mainly relate to data erasure and identifying the data controller. However, the majority also consider these problems to be solvable utilizing already available technologies, e.g. off-chain storage, encryption, pseudonymization or usage of private BCs. Comparing these proposals with the findings in computer science literature shows that especially off-chain storage, encryption and redactable blockchains can be regarded as adequate solutions.

**Keywords:** Blockchain, Data Protection, Protection of Personal Data, Privacy, Content Analysis

## 1 Introduction

Distributed ledger technologies (DLT) use decentralized data storage on the computers of many users. One variant is the blockchain (BC) technology, which has become one of the top five strategic decisions for many companies worldwide [1]. Since BC has gradually been approaching commercialization [2], the discussion concerning an appropriate regulatory framework has also recently gained momentum worldwide. The data protection requirements of the General Data Protection Regulation (GDPR) are the most important regulatory challenge in the European Union (EU). Among the various data protection (DP) challenges, the following two are the most serious for BC applications: the difficulty of identifying the data controller, and the DP implications of the immutability of the data stored in a blockchain [3].

Companies expect disruptive changes and economic gains if BCs are operated in compliance with data protection [e.g. 4]. Therefore, an intensive debate is ongoing in academia, industry and regulatory authorities on the compatibility of BC applications with the data protection framework [5–7]. Although this debate is intensive, it lacks

16th International Conference on Wirtschaftsinformatik,
March 2021, Essen, Germany

empirical evidence of how stakeholders actually assess the DP challenges. Research in this area generally focuses on two strands. First, there are studies which only ask companies whether they perceive regulations such as DP as an obstacle. Second, there is intensive research, particularly from a legal and technical perspective, into solutions to DP challenges (see Section 2.2).

Several research gaps can be identified here. The analyses of DP challenges usually remain on an abstract level (in the sense of: is regulation/data protection considered an obstacle to the use of blockchains in your company?). Quantitative analyses are missing of the concrete data protection-related challenges that companies face. In addition, surveys among companies generally do not enquire whether they consider solutions to the challenges possible.

Thus, our study aims to find out whether and how companies and organizations are trying to overcome these challenges. Since there is a lack of quantitative analyses of the possible solutions being discussed, computer science research also lacks analyses of the usefulness of such proposed solutions. Against this background, we ask the following research questions:

- RQ1: What challenges and possible solutions do stakeholders see with regard to data protection and blockchain issues?
- RQ2: How are the proposed solutions to be evaluated from a technical point of view?

We answer these questions in two steps. To answer RQ1, we rely on a text-based qualitative content analysis of the statements of 130 actors who participated in the 2019 blockchain consultation of the German government. To answer RQ2, we assess the challenges articulated and the proposed solutions based on state of the art technical BC research in computer science literature.

Our study has implications for academia, practitioners and policy makers by providing qualitative insights into how companies and organizations evaluate the DP challenges and solutions concerning BC technology.

## 2 State of Research

### 2.1 Background of Blockchain Technology

DLT are a type of database with globally decentralized data storage across multiple computers, so-called nodes [8]. Each node has a partial or full copy of the ledger and can interchange data formally as a peer-to-peer network without a central authority [9]. Blockchain is a DLT, but consolidates new data into blocks, which are chained to the preceding one by means of cryptographic hash functions [10]. This results in an append-only structure, where prior blocks cannot be deleted or edited without changing all the subsequent blocks. If something tries to change a block, the corresponding hash value changes, resulting in a breakage of the chain [8]. As every node has a copy of the ledger, tampering with a single node cannot manipulate the blockchain. This should ensure full transparency and traceability. This consensus mechanism is, however, only possible for financial data, as the nodes check whether the total amount, e.g. of Bitcoins, is still valid after a transaction. This is to ensure that no one can transfer coins they do not own. In public BCs, nodes are unknown and there is no administrator. In contrast, permissioned/private or consortium BCs have one or a group of known nodes with special rights able to grant access to new users (such as nodes, miners or programmers) and thus control the BC.

The concept emerged during the global financial crisis in 2008, when an author with the pseudonym Satoshi Nakamoto introduced the crypto currency Bitcoin to provide non-manipulable financial transactions on the internet by avoiding intermediaries, e.g. financial institutions [9, 10]. The term “blockchain” only appeared 2013, but the underlying technology of storing chained hash values of documents already existed in 1991 [9].

### 2.2 Data Protection in Blockchain Technologies

The EU considers data protection a fundamental right (see Art. 8 of the Charter of Fundamental Rights of the EU) and strives for economic growth (see Recitals 2 and 7 GDPR). Consequently, the GDPR aims to guarantee a high and harmonized level of data protection for the personal data of EU citizens on the one hand, and to strengthen the digital single market by removing obstacles that impede the free movement of personal data on the other (see Recitals 9 and 10 GDPR.). The GDPR represents a comprehensive and complex set of rules with which these goals are to be achieved. A wide range of provisions must be complied with in order to ensure that personal data is processed in accordance with data protection regulations. These include the material and territorial scope, the definition of personal data, the rules on the lawfulness of processing, the rights of the data subject, and the obligations of the data controller. The GDPR has two requirements that stand out in particular for blockchain applications.

First, Art. 4 (7) GDPR is based on the assumption that one or more relatively clearly identifiable data controllers are responsible for the processing operation, and against whom the data subject can assert his or her DP rights. However, the technical mode of operation of public blockchain technology does not provide for clear responsibility. Instead, it explicitly relies on the decentralization of responsibility. Due to their influence on determining the means and purposes of BC data processing, nodes and users can be considered controllers (but not only these depending on the BC). Since both nodes and users are controllers, the provisions of Art. 26 GDPR on joint controllers must be fulfilled. In the case of traditional data processing, the joint responsibility of the data controllers has to be regulated in an agreement. However, as there is insufficient knowledge of all controllers in public BCs, a contractual sharing of joint responsibility is not possible [3]. The European Data Protection Board (EDPB) has finally clarified in its most recent Guidelines (see recital 167) that, in such cases, each individual controller must comply with the obligations of the GDPR [11]. In a public blockchain, data is passed on to an unmanageable group of people and the granting of data subject rights is extremely difficult. Therefore, nodes and users will generally not be able to meet their obligations [3, 7]. The Schrems II ruling has further complicated the issue. The ruling clarified that (joint) data controllers must also ensure compliance with the GDPR in the case of transfers of personal data to countries outside the EU that do not have an adequate level of data protection [12]. Since, in a public blockchain, every person from any part of the world can become a node or a user, fulfilling this requirement is also difficult.

Second, the GDPR grants the data subjects the right to rectify and erase their data. However, the distinctive feature of BC technology is the immutability of the stored data, in order to achieve maximum transparency and data integrity [3]. As this brief discussion has made clear, there are still challenges with regard to the privacy-compliant operation of a public BC, which make it necessary to adapt BC architecture to the legal requirements.

Nonetheless, many companies expect disruptive changes due to BC technology. Thus, awareness of the challenges posed by DP law has led to an intensive debate. A number of recent surveys have been carried out to identify the concerns of the business community. A PricewaterhouseCoopers survey [13] of 600 corporate executives from 15 countries found that 27% considered regulatory uncertainty the biggest barrier to the use of BCs. A Bitkom survey [14] of 1,004 companies revealed that 66% regarded DP requirements as a challenge. According to Deloitte's survey [1], 32% of 1,488 companies surveyed named regulatory issues as hindering BC adoption. This study is the only one that asked companies whether and to what extent they saw possibilities to overcome the DP challenges. Indeed, 83% of respondents indicated that they were very or somewhat confident that they would be able to meet the regulatory requirements [1]. Such surveys mainly focus on finding out whether companies see DP as a challenge, but they do not ask what exactly they regard as challenges or how they will try to overcome them.

Until a few years ago, IS literature considered DP to be guaranteed, e.g. due to the anonymity of the nodes [15]. Today, a growing number of papers, mainly in the fields of computer science [10, 16] and law [3], acknowledge DP challenges, and search for possible solutions. Some of these deal with the challenges in great depth. To the authors' best knowledge, there is no literature on the possible solutions discussed in the business world and no literature comparing stakeholder views.

The German government published its BC strategy in September 2019 [17]. It laid down the framework conditions for the further development of the technology and announced several dozen measures in five fields of action. The subsequent reactions ranged from clear support [18] to skeptical relief [19] and fundamental criticism [20]. In particular, the lack of uniform goals and a binding timetable was criticized.

With regard to DP issues in connection with BC applications, the German government stated that it saw no need to amend the GDPR. Instead, the uncertainties of developers and users with regard to data protection law were to be addressed using existing technical solutions (including hash values, pseudonymization, ZKP) and holding a "round table" on the topic of blockchain and data protection [17].

## 3 Methodology

We conducted a text-based, qualitative content analysis (CA) following [21], because "it is a research technique for making replicable and valid inferences from texts" [21].

**Unitizing:** As the basis for our analysis, we used the document provided by [22], containing all the officially published answers given in the consultation process for the blockchain strategy of the German government. In total, there were 6,261 answers from 130 respondents, which we transferred into a machine-readable format.

**Sampling:** From this corpus of data, we identified data protection-relevant answers and questions by applying a keyword search to the statements. Since the blockchain consultation was conducted in German, we decided to use the terms shown in Table 1 (case insensitive, incl. substrings). These are terms found in the German and English version of the GDPR as well as synonyms in the relevant data protection literature. In Table 1, we also provide the percentage of mentions of each keyword within questions (Q) and given answers (A). We ended up with 537 relevant questions and answers in total, which we refer to as statements from now on (8.6%).

Table 1: Keywords and percentage of mentions in all statements (Q=question, A=answer)

| **Keyword** | **% in Q** | **% in A** | | **Keyword** | **% in Q** | **% in A** |
|---|---|---|---|---|---|---|
| Datenschutz | 14.90% | 35.94% | | private Daten | 0.00% | 0.74% |
| Privatsphäre | 21.23% | 5.40% | | DS-GVO | 0.00% | 0.19% |
| personenbezogen | 29.24% | 18.06% | | Privatheit | 0.00% | 0.13% |
| DSGVO | 3.91% | 22.53% | | private data | 0.00% | 0.03% |
| privacy | 0.00% | 5.96% | | personal data | 0.00% | 0.00% |
| GDPR | 0.00% | 3.72% | | Data protection | 0.00% | 0.00% |
| persönliche Daten | 0.00% | 1.68% | | | | |

**Coding:** Two persons coded the statements. Both were familiar with the subject of data protection; the principal coder (C1) had additional technical expertise in BC technology. C1 processed all 537 statements and defined the final category set. The second coder (C2) processed a representative sample of 50 statements from more than ten percent of the respondents, as suggested by [23].

C1 applied iterative inductive-deductive coding. Based on our research question, we predefined three subject areas, i.e. “problem relevance”, “articulated challenges” and “proposed solutions”. In total, 57 codes were identified for each subject area. Additionally, C1 always noted the usage domain in which the question was situated.

After this initial coding, C2 relied on the codes by C1 and processed the coding of the representative sample to ensure reproducibility [21]. Our analysis showed a moderate strength of agreement, manifesting in a Cohen’s Kappa of 0.53 for intercoder reliability.

**Reducing:** We applied hybrid card sorting in order to reduce the number of codes and created clusters based on insights from literature. For example, [16] and [24] considered pruning, Merkle trees, and chameleon hash functions as erasure methods, as all aim at (physically) deleting data. This process resulted in a reduction from 57 to 31 categories

(Table 2). Finally, we consulted an expert panel of n=8 to check the codes were collectively exhaustive and mutually exclusive, as suggested by [21].

Table 2. Quantitative representation of codes and categories for each subject area

| Group of themes | No. of code categories | No. of codes |
|---|---|---|
| Problem description | 4 | 5 |
| Articulated challenges | 15 | 19 |
| Proposed solutions | 8 | 26 |
| General remarks | 4 | 7 |
| Total | 31 | 57 |

**Inferring:** In order to "bridge [...] the gap between descriptive accounts of texts and what they mean, refer to, entail, provoke, or cause" [21], we compared the respondents' statements with findings from the literature. This yields important insights, as most respondents are generalists rather than BC experts.

**Narrating:** We present the procedures and results, theoretical and practical contributions, and upcoming questions of the content analysis in the next chapters.

## 4 Results

In this section, we provide a descriptive overview of all DP-related statements of the respondents. First, we show how relevant DP issues are for the respondents. Second, we present the specific challenges stated and, finally, we give an overview of the proposed solutions to these challenges. In total, there were 130 companies and organizations, of which 94 (72%) provided DP-relevant answers. Their 537 statements form the basis for this analysis. The respondents are active in different industries, with a majority in the IT sector (38%), followed by research institutions (18%) and Fintech (13%). Although our focus is on companies, the analysis involves different types of stakeholders. Based on our analysis, we cannot identify structural differences in the answers between the different sectors, as there were not enough data available. The available data indicate that the automotive, financial and energy sectors are particularly skeptical about the compatibility of data protection and blockchain. In contrast, the healthcare and IT sectors are particularly optimistic.[1]

### 4.1 Problem Relevance

The first subject area represents how relevant companies and organizations assess DP in the field of blockchain. The relevance is shown in eight coding categories (Figure 1).

[1] Compare the *online appendix* for a table in which the assessment of how serious the problems are and whether there are possible solutions was divided by sector.

These results are independent of specific usage domains, but an analysis of the provided application domains shows similar results.

Out of the total of 94 respondents, a majority (81%) agree that DP is a challenge in BC applications, whereas only 21 respondents (22%) see no DP challenges, at least in some usage domains[2]. 19 respondents (20%) consider DP issues to be so serious that they expect a showstopper effect for companies. This means that they see no possible solution to DP issues of BC, either now or in the future. Very few respondents believe these challenges will discourage companies from adopting blockchain applications. A majority of 66 respondents are confident that solutions to DP issues are possible. About one third (29%) of respondents even argue that BC can enhance DP. For example, they mention "self-sovereign identity" (7%), which helps users to track and configure how their data must be processed [25].

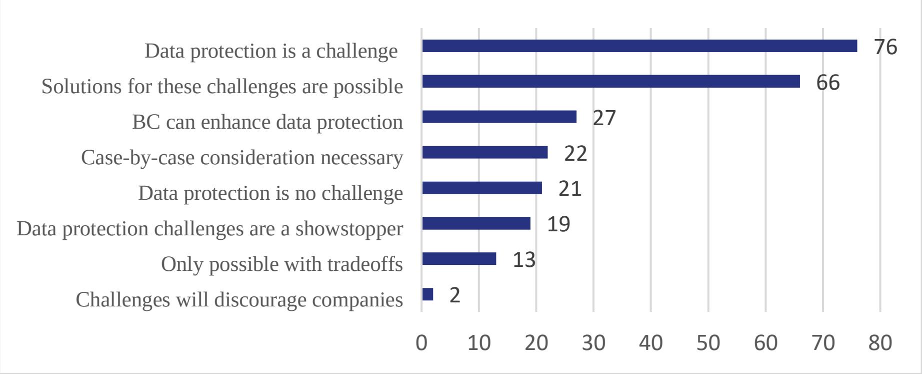

Figure 1. Relevance of data protection issues in blockchain applications in absolute numbers, n=94 (own analysis, multiple responses possible)

### 4.2 Articulated Challenges

Guaranteeing the data subject’s rights is seen by 53 respondents as the biggest challenge. In Figure 2, we refer to these rights with the heading “grouped”. Of these, 50 respondents (94%) consider deletion to be the main problem, 38 (72%) say rectification and six (11%) see the right of data portability as challenging. 12 respondents (23%) mention guaranteeing these rights in general as problematic. The second most mentioned challenge by 43 respondents (46%) is that all personal data in the BC is visible to everyone. The third most frequently mentioned problem was the (non-)identifiability of the data controller (31 times, 33%). 15 respondents (16%) criticized the current encryption technologies for personal data, fearing they could be cracked in the near future. 14 actors (15%) doubted the effectiveness of pseudonymization and regard de-pseudonymization as a problem. Seven (7%) criticized the security of storage, which is particularly important in the case of possible off-chain storage or storage of the keys required for encryption. Four respondents (4%)

[2] As the respondents see no challenges in some specific usage domains, the total number of mentions is 97 (103%), instead of max. 94.

expressed concerns about the integrity or quality of the input data and stated that it is difficult to verify the correctness of data relating to objects in the physical world. Finally, three respondents (3%) criticized the high computing power requirements of zero-knowledge proofs (ZKP) and the effectiveness of anonymizing personal data.

In addition, seven respondents (7%) drew attention to the legal problems arising from the transfer and data storage outside the EU. In the same context, seven respondents (7%) criticized the unclear legal situation, both with regard to divergent legal frameworks worldwide, which made the use of a global BC more difficult, and with regard to the - from their perspective - unclear legal situation in the EU.

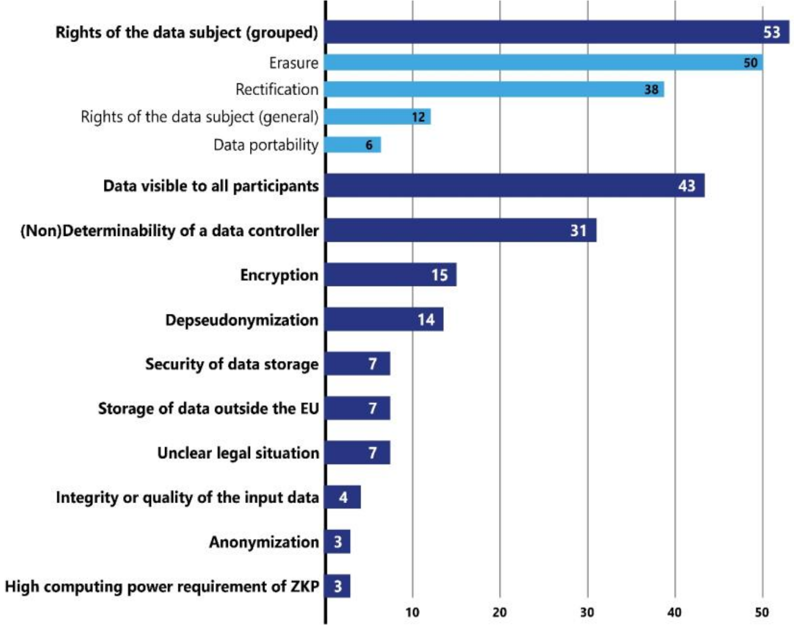

Figure 2. Overview of articulated challenges (multiple coding possible)

### 4.3 Proposed Solutions

A majority of the respondents believe that solutions to DP issues in blockchain technology are generally possible. In the following section, we briefly present the proposed solutions that were mentioned by the respondents.

Figure 3 shows that the majority of respondents (62 or 66%) believed that a solution to the DP challenges was already possible using existing technology. This included off-chain storage in the first place (43 or 46%), closely followed by encryption technologies (37 and 39%). 18 actors (19%) considered pseudonymization to be a useful approach. The use of zero-knowledge proofs (ZKP) represented a possible solution for 13 actors (14%). Twelve actors (13%) thought that there are possibilities for deleting data. The use of existing anonymization methods was mentioned by eleven actors (12%).

The second most frequently mentioned proposal (38 actors/40%) was to simply refrain from storing personal data in a BC. 31 stakeholders (33%) were in favor of restricting access by using a private BC. 29 actors (31%) mentioned legal adjustments. Of those, 27 (29%) mention the concretization and amendment of the GDPR, four (4%) the creation of legal bases that apply worldwide (4 or 14%), and two (2%) cooperation with other countries.

29 respondents (31%) expressed the opinion that further technical developments and standards were necessary to operate BCs in conformity with data protection laws. With the help of organizational security measures, such as user roles or data aggregation, a solution to DP problems could be found, according to 17 actors (18%). This included user roles and rights (9 or 10%), assigning different protection levels to different types of data (8 or 9%), and the exclusive storage of data aggregates (3 or 3%).

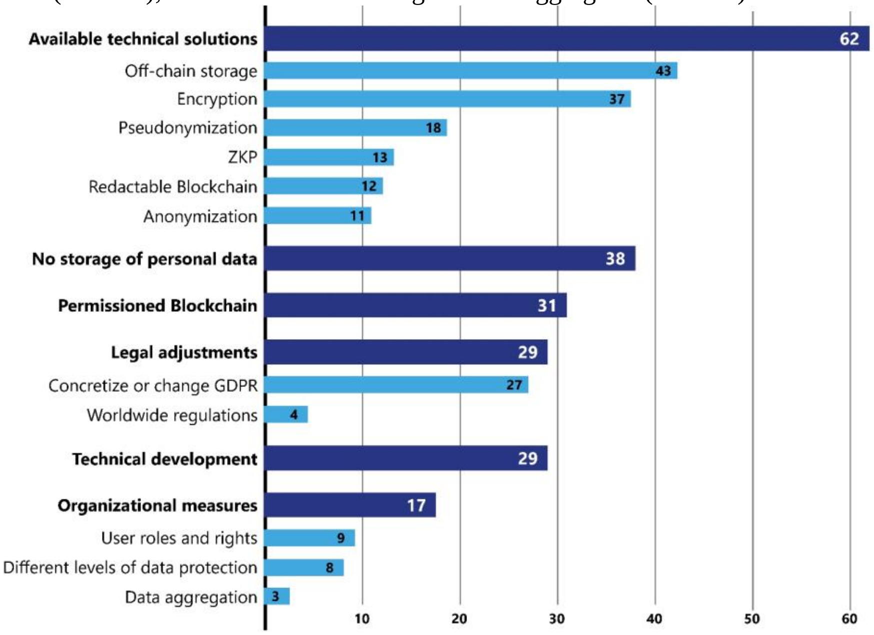

Figure 3. Overview of proposed solutions (multiple coding possible)

## 5 Discussion

In this section, we compare the articulated challenges and proposed solutions with the state of research. The results show that the majority of respondents (81%) consider data protection a challenge in BC. However, many of the respondents in the BC consultation consider the challenges to be manageable. Our results also show that many respondents promote BC technology, whereas others are rather critical or cannot make a generally valid statement. Further, our results suggest that many companies and organizations are not very familiar with "the BC issue" or have no desire to delve deeper into it. This can be seen, for example, in the fact that almost a third of the respondents urged to concretize or amend the GDPR and called on public authorities to offer guidelines. Whereas publishing guidelines is certainly a realistic option for action (as the recent

EDPB guidelines on accountability have shown), amendments of the GDPR are most unlikely at this point in time [26]. In the following subsections, we discuss the statements articulated by the respondents.

### 5.1 General Statements

Several respondents raise questions concerning information security goals, i.e. confidentiality, integrity and availability, which we want to discuss briefly. While immutability and distributed storage fulfill integrity and availability requirements; the blockchain "is not specifically designed to support or maintain data confidentiality" [27]. Although the company Achelos suggests using *encryption* to tackle these problems, encryption cannot prevent internal errors, such as misuse.

Additionally, a few respondents note that the inserted data is not verified. Thus, data quality control is an issue. The literature agrees on that point, as the BC technology "does not guarantee or improve data quality" [28]. Neither the respondents nor the literature suggest possible solutions.

Almost one third of the respondents articulate the need for technical developments and standards to operate BCs in a DP-compliant manner. According to the company DB Systel, there are currently only "trade-offs between security, data protection, efficiency, flexibility, platform complexity and user-friendliness for developers" (own translation). The literature acknowledges that the prerequisites for successful research have been created and it is likely that a large number of new applications will emerge in the near future [29].

Finally, for about one fifth of the respondents, technical solutions do not seem to be sufficient at this point of time. Instead, they suggest three types of *organizational measures*: (1) introduction of user roles and rights, (2) allocation of different protection levels to different types of data, e.g. medical data, and (3) exclusive storage of data aggregates. The literature also discusses such measures for BC applications, e.g. [30] find that "organizational measures need to be taken to fulfill the boundary conditions, before blockchain can be used successfully". Generally, private BCs can comply with these three types of measures, as a central authority can be defined. For (1), a predefined data controller could allow new users and assign rights and roles, which could be themselves stored in the BC using smart contracts [7]. For (2), it is technically possible to assign data categories that define conditions of use and specify a group of privileged recipients [3]. [31] describe a method for service providers to carry out data processing directly in the user's network without accessing raw data. This can ensure that users can only access certain data. At present, there does not seem to be a sufficient solution for the (3) measure. Researchers agree that performing privacy-preserving data aggregation is challenging due to advancements in data processing using big data and artificial intelligence [32].

### 5.2 Rights of the Data Subject: Erasure, Rectification and Data Portability

**Problem description**: By far the most frequently mentioned problem relates to guaranteeing the rights of data subjects to the erasure or rectification of their data. This is hardly possible, because of the immutable nature of BCs. Several researchers agree that these are the most pressing points when considering DP in BC applications [24]. Another problem mentioned by the respondents is data portability. This calls for data being stored in structured, commonly used and machine-readable formats. However, there is no standard exchange format in the blockchain [33].

**Suggested solutions and evaluation:** Most respondents suggest *off-chain storage*. In this case, only a reference hash value of the original file will be stored in the BC. As hash values are collision-free, changes to the original file can be detected and thus transparency achieved, while being able to continue the blockchain [34]. Contrary to the original BC idea, the operator must be trusted, since there are ways to change stored data afterwards. Such methods are often referred to as *redactable blockchains* [24]. One prominent example mentioned by the respondents are *chameleon hash functions*, which have collision-free algorithms that enable a group of controllers to delete data while leaving a "scar" [35]. However, deletion is only possible in permissioned BC architectures, because data controllers need to coincide and deletion is an exception [24]. Another drawback is the possible identification of a hash value in a small search space [36]. For example, a modern graphic card can calculate the double-sha256 hashes of all human names (7.6 bn) in under 4 seconds [37]. Thus, there is the need to add random data, a so-called "secret". To identify users as data holders, this secret must be transmitted. However, attacks could compromise the transition and intercept the secret.

To overcome this problem, some respondents suggest *Merkle trees,* which are used, e.g. in the Bitcoin blockchain, to reclaim disk space [9]. These combine hashes of different data fields, such as the hash of the name and the hash of the birthdate and create a new hash. This results in a tree-like structure. Only the uppermost hash (analogous to the tree trunk) is saved in the BC [38], so that no single person can identify the original hash. Especially in the Bitcoin BC, Merkle trees are used to verify data blocks to ensure that no miner transmits a manipulated financial transaction.

*Zero knowledge proofs (ZKP)* represent another solution, which dispenses with the release of the "secret". They do not use a person's real data, but rather a proof that the datum exists or is correct [39]. In such a way, one could easily check whether a person is older than 18, without revealing the actual birthdate. However, there are two disadvantages. First, the literature suggests that an attacker could try out all possible input values until the proof is verified [40]. Second, the respondents note that ZKP require considerable computing power. Whereas the transactions are stored on-chain, the computation and storage are performed off-chain.

Other respondents mentioned *tombstones* and *data revocation keys* as deletion methods. However, both these methods only mark data as invalid and do not physically delete it, as required by Art. 17 GDPR. Here, at first glance, *forking* seems a solution. It presents an irreversible separation from a BC. However, if data changes and deletions have to occur frequently, the chain needs to be split very often. This contradicts the basic BC idea, since short chains may counteract transparency [41]. The so-called *pruning* is a deletion method already used in the Bitcoin blockchain. Pruning removes old and thus no longer needed parts of the chain while maintaining the integrity of the whole chain using Merkle trees [9]. However, this can only delete transaction values that have been "consumed" (i.e. spent Bitcoin). No other data could be pruned [6].

Concerning the right of data portability, it is unlikely that a standard file format will evolve in the near future, as lock-in effects are economically advantageous to companies. The current diversity of the blockchain market underlines this problem [2]. The question also arises as to who should guarantee the rights of those affected (see the following subsection).

### 5.3 Identification of the Data Controller

**Problem description**: To protect the rights of a data subject, the GDPR provides for a (joint) data controller who, for example, acts as an addressee for data subjects to assert their rights, or who fulfills the transparency requirements of the Articles 13 and 14. One third of all respondents were of the opinion that the difficulty in identifying the data controller (see also section 2.2) makes it considerably harder to comply with the legal requirements.

**Suggested solutions and evaluation:** Respondents suggest using a permissioned *private or consortium blockchain,* as the participants are known there. Researchers, such as [3], agree with this, even though this contradicts the basic BC idea of transparency and distributed responsibility. Nevertheless, in light of recent developments in case law and EDPB recommendations, we concur with this assessment. In view of the vast number and geographical location of different nodes and users, we do not expect that it will be possible to fulfill the data protection requirements regarding the obligations of data controllers in a public BC. Alternatively, the problem of responsibility could be solved by limiting the personal nature of data to a manageable group of actors (see the following subsection on encryption) [3, 7].

### 5.4 Encryption

**Problem description**: In connection with the above-mentioned problems of identifying the data controller and guaranteeing the rights of those affected by data processing, an important strand of the debate is devoted to solutions using *encryption*.

**Suggested solutions and evaluation:** Many respondents point out that the DP challenges related to erasure could be solved by *encryption*. The respondents' assumptions sound simple: data would only be considered personal for those actors who have the access key. In this regard, eco - the Association of the Internet Industry demands that the verified destruction of a decryption key should be considered sufficient for anonymization.

The French Data Protection Authority CNIL agrees with this opinion. However, [3] points out the need for further regulatory advice on this issue, as under the current conditions, even nodes that do not actually have significant control over the encrypted data could still be considered responsible.

From a technical point of view, encryption is not identical to physical erasure - it only makes data inaccessible. In this respect, both respondents and researchers fear that current encryption methods could be cracked in the future and data made accessible [7]. However, (a)synchronous encryption algorithms such as AES or RSA are commonly used to encrypt bulk data [42] and cracking these is very unlikely in practice, as AES-

256 and RSA-2048 are considered secure for the next decades [43]. However, BC creates an immutable technology architecture, which relies on cryptographic procedures. In case of an error in a procedure, the entire chain would be affected forever. Thus, as long as no further regulatory guidance is provided, only erasure methods  as discussed in Section 5.2 could help to overcome the problems.

### 5.5 Storage outside the European Union

**Problem description**: The GDPR requires that if any personal data is transferred outside the EU, it must meet the requirements of Articles 44-49 GDPR. As a public BC is distributed among many (unidentifiable) users, personal data could be stored outside of the EU, which causes compliance difficulty.
**Suggested solutions and evaluation:** The respondents referred to the use of a *private or consortium blockchain*. Indeed, this would solve some challenges, as only EU citizens could be allowed to join. However, it would thwart the BC's goal of maximizing transparency. While relying on a public chain, *geo-blocking* could be a solution, but is not in accordance with European law [44]. Furthermore, VPN software can easily circumvent geo-blocking.[3] Apart from the possibilities mentioned above, the transfer of data outside the EU is currently an ongoing problem (not only) in the BC context.

### 5.6 Data Readability and Writability for BC Participants

**Problem description**: Another important challenge articulated by the respondents is that all blockchain users can read and write all data, even personal data, as there is no possibility of verification for non-transactional data. Here a dilemma arises: while the visibility of all data ought to support transparency, readability and writability also raises significant DP concerns, because any user can add personal data to the BC. This poses a challenge even if limiting responsibility to the owner of the private keys would enable DP-compliant operation of a BC. Indeed, other users could enter unencrypted personal data into BC at any time and thus invalidate its data protection compliance.
**Suggested solutions and evaluation:** Most respondents suggest using *access-restricted (private) blockchains* to let only registered users participate. Although this could relieve several DP challenges, it runs counter to the BC intention of ensuring full transparency. Other respondents suggest a rather pragmatic approach: simply *no storage of personal data*. However, technical or organizational measures cannot fully achieve this. Firstly, data that are not personal today, could become so in future [45]. Secondly, content filters could be easily circumvented by experienced users, and excluding these users is also very difficult [46]. Respondents also discuss *pseudonymization* as an effective method to veil personal references. However, it has become relatively easy to re-assign data. For example, [47] de-pseudonymized up to 60% of the IP addresses used to execute Bitcoin transactions. The literature does not provide effective solutions to render de-pseudonymization impossible, as even TOR

[3] *Please note that this assessment was made before July 16, 2020, when the European Court of Justice ruled that the EU-US Privacy Shield is invalid.*

network users can be identified by their Bitcoin transactions [48]. Another possible solution is a Bitcoin mixer, which combines several transactions into a large bitcoin pool and then distributes the coins to the receivers [49]. However, the service provider still knows the user's bitcoin address and could de-anonymize data. Furthermore, the mixing service could be a honeypot set up by governments to identify users. For the same reason, *anonymization* cannot be guaranteed for all BC applications. Even Monero, which claims to be an anonymous cryptocurrency, is prone to de-anonymization errors [50].

## 6 Conclusions and Outlook

The data protection-related challenges of BC technology are taking center stage for many companies in different industry sectors. Despite a broad debate, little was known about the concrete challenges facing companies and how they intend to overcome them.

Our results augment the existing literature in several regards. First, our study contributes to undermining the view that DP regulations are an insurmountable hurdle to the use of BCs. The number of actors perceiving a challenge (81%) and believing it can be overcome (70%) is very similar. Second, our analysis provides insights into which challenges the stakeholders regard as particularly important. Immutability is the biggest challenge for most respondents. Whereas many answers relate to public BCs and guaranteeing data subjects' rights, some respondents even promote BC as improving DP, e.g. via self-sovereign identities. Third, our results also show that the majority consider the problems solvable with already available technologies, in particular off-chain storage, encryption, pseudonymization and ZKPs. However, a comparison with state-of-the-art scientific literature reveals that only off-chain storage and encryption are advisable. Finally, a considerable number of actors also demanded the use of a private BC, promotion of further technical developments, and concretization or modification of the GDPR. These results show that stakeholders are aware that there is no silver bullet to overcome DP-related challenges. Instead, solutions depend strongly on the specific implementation and use case. Ultimately, our results show that most challenges arise in the field of public BCs. Thus, academia should focus on solutions here (e.g. chameleon hash functions). Furthermore, computer science research could benefit from empirical insights into how BC stakeholders perceive the challenges and solutions, and how these coincide with research. Practitioners can benefit from the evaluation of solutions to installing a BC architecture that best addresses DP demands. Additionally, our results offer important insights for policy makers, as they can see what specific challenges companies face and which research to support.

Although we rely on a large sample of 94 stakeholders, it is not representative for all industry sectors, since the sample contains only actors who took part in the consultation. The quantity and quality of coders is another common criticism of content analysis. However, our reliability measure shows moderate strength.

Future research could make a quantitative analysis of how often the proposed solutions are mentioned by stakeholders across all industry sectors. This would also pave the way for sector-specific analyses to find out whether, for example, certain sectors see greater challenges, or whether economic actors and researchers hold different views.

**Funding and Acknowledgement**
This research has been funded by the German Federal Ministry of Education and Research (BMBF) within the Project "Forum Privatheit und selbstbestimmtes Leben in der Digitalen Welt". We would like to thank our colleagues from the "Forum Privatheit" as well as Dirk Achenbach and Jochen Rill from FZI Research Center for Information Technology, for detailed and helpful feedback during the development of this paper.

## References

1. Deloitte: Deloitte’s 2020 Global Blockchain Survey (2020)
2. Grover, P., Kar, A.K., Janssen, M.: Diffusion of blockchain technology. JEIM 32, 735–757 (2019)
3. Finck, M.: Blockchain and the General Data Protection Regulation. Can distributed ledgers be squared with European data protection law? European Parliament, Brussels (2019)
4. Holotiuk, F., Pisani, F., Moormann, J.: The Impact of Blockchain Technology on Business Models in the Payments Industry. Proc. Wirtschaftsinformatik (2017)
5. European Commission (2018): European countries join Blockchain Partnership, https://ec.europa.eu/digital-single-market/en/news/european-countries-join-blockchain-partnership
6. Farshid, S., Reitz, A., Roßbach, P.: Design of a Forgetting Blockchain: A Possible Way to Accomplish GDPR Compatibility. In: HICSS-52 2019
7. Fridgen, G., Guggenberger, N., Hoeren, T., Prinz, W., Urbach, N.: Chancen und Herausforderungen von DLT (Blockchain) in Mobilität und Logistik. Berlin (2019)
8. Nofer, M., Gomber, P., Hinz, O., Schiereck, D.: Blockchain. Business & Information Systems Engineering 59, 183–187 (2017)
9. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008)
10. Hughes, L., Dwivedi, Y.K., Misra, S.K., Rana, N.P., Raghavan, V., Akella, V.: Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda. IJIM 49, 114–129 (2019)
11. EDPB: Guidelines 07/2020 on the concepts of controller and processor in the GDPR (2020)
12. CJEU: Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems: Case C 311/18 (2020)
13. PwC: Global Blockchain Survey 2018 (2018)
14. Gentemann, L.: Blockchain in Deutschland – Einsatz, Potenziale, Herausforderungen. Studienbericht 2019. Berlin (2019)
15. Yli-Huumo, J., Ko, D., Choi, S., Park, S., Smolander, K.: Where Is Current Research on Blockchain Technology?-A Systematic Review. PloS one 11 (2016)

16. Florian, M., Henningsen, S., Beaucamp, S., Scheuermann, B.: Erasing Data from Blockchain Nodes. In: 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 367–376. IEEE (2019)
17. Press and Information Office of the Federal Government: Blockchain Strategy of the Federal Government (2019)
18. Sausen, T. (2019): BVDW lobt Blockchain-Strategie der Bundesregierung, https://www.bvdw.org/der-bvdw/news/detail/artikel/bvdw-lobt-blockchain-strategie-der-bundesregierung/
19. Brandenburg, M. (2019): Die Blockchain-Strategie der Bundesregierung – ein überfälliges Positionspapier, https://www.btc-echo.de/die-blockchain-strategie-der-bundesregierung-ein-ueberfaelliges-positionspapier/
20. Streim, A. and Hansen, P. (2019): Bitkom: Blockchain-Strategie gibt Aufbruchsignal, https://www.bitkom.org/Presse/Presseinformation/Bitkom-Blockchain-Strategie-gibt-Aufbruchsignal
21. Krippendorff, K.: Content analysis. An introduction to its methodology. Sage Publ, Thousand Oaks, Calif. (2004)
22. BMWi, BMF (2019): Online-Konsultation zur Erarbeitung der Blockchain-Strategie der Bundesregierung, https://www.bmwi.de/Redaktion/DE/Downloads/-Stellungnahmen/Stellungnahmen-Blockchain/stellungnahmen.pdf
23. Lombard, M., Snyder-Duch, J., Bracken, C.C.: Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. Human Comm Res 28, 587–604 (2002)
24. Ateniese, G., Magri, B., Venturi, D., Andrade, E.: Redactable Blockchain – or – Rewriting History in Bitcoin and Friends. In: IEEE European Symposium on Security and Privacy (EuroS&P), pp. 111–126. IEEE, Piscataway, NJ (2017)
25. Schwerin, S.: Blockchain and Privacy Protection in the Case of the European General Data Protection Regulation (GDPR): A Delphi Study. The JBBA 1, 1–77 (2018)
26. EC: Data Protection as a Pillar of Citizens' Empowerment and the EU's Approach to the Digital Transition. Communication from the Commission to the European Parliament and the Council (2020)
27. Warkentin, M., Orgeron, C.: Using the security triad to assess blockchain technology in public sector applications. IJIM 52 (2020)
28. Piscini, E., Dalton, D., Kehoe, L.: Blockchain & Cyber Security (2017)
29. Belchior, R., Vasconcelos, A., Guerreiro, S., Correia, M.: A Survey on Blockchain Interoperability: Past, Present, and Future Trends (2020)
30. Behnke, K., Janssen, M.F.W.H.A.: Boundary conditions for traceability in food supply chains using blockchain technology. IJIM 52 (2020)
31. Zyskind, G., Nathan, O., Pentland, A.S.: Decentralizing Privacy: Using Blockchain to Protect Personal Data. In: 2015 IEEE Security and Privacy Workshops (SPW), pp. 180–184. IEEE, Piscataway, NJ (2015)
32. Memon, I.: An Analysis of Privacy Preserving Data Aggregation Protocols for WSNs. In: Park, J.J., Zomaya, A., Yeo, S., Sahni, S. (eds.) Proceedings of the 9th IFIP international conference, NPC 2012, 7513, pp. 119–128. Springer, Berlin (2012)

33. Jaikaran, C.: Blockchain: Background and Policy Issues (2018)
34. BSI: Blockchain sicher gestalten. Konzepte, Anforderungen, Bewertungen (2019)
35. Lumb, R., Treat, D., Jelf, O.: Editing the uneditable Blockchain. Why distributed ledger technology must adapt to an imperfect world (2016)
36. Kohn, W., Tamm, U.: Mathematik für Wirtschaftsinformatiker: Grundlagen und Anwendungen. Springer Berlin Heidelberg (2019)
37. Bitcoin Wiki (2020): Non-specialized hardware comparison, https://en.bitcoin.it/wiki/Non-specialized_hardware_comparison
38. Fill, H.-G., Haerer, F.: Knowledge Blockchains: Applying Blockchain Technologies to Enterprise Modeling. In: HICSS-51 2018 (2018)
39. Fatz, F., Hake, P., Fettke, P.: Confidentiality-preserving Validation of Tax Documents on the Blockchain. In: Gronau, N., Heine, M., Poustcchi, K., Krasnova, H. (eds.) WI2020 Zentrale Tracks, pp. 1262–1277. GITO (2020)
40. Yung, M.: Zero-Knowledge Proofs of Computational Power. In: Quisquater, J.-J., Vandewalle, J. (eds.) Advances in cryptology - EUROCRYPT '89, 434, pp. 196–207. Springer, Berlin (1990)
41. Avital, M., Beck, R., King, J.L., Rossi, M., Teigland, R.: Jumping on the Blockchain Bandwagon: Lessons of the Past and Outlook to the Future. In: Proceedings of the 37th ICIS (2016)
42. Thambiraja, E., Ramesh, G., Umarani, R.: A survey on various most common encryption techniques. International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) 2 (2012)
43. Bernstein, D.J., Lange, T.: Post-quantum cryptography. Nature 549, 188–194 (2017)
44. European Parliament (2018): Distributed ledger technologies and blockchains: building trust with disintermediation, http://www.europarl.europa.eu/doceo/document/TA-8-2018-0373_EN.html
45. Tönnissen, S., Teuteberg, F.: DSGVO und die Blockchain. Datenschutz und Datensicherheit 44, 322–327 (2020)
46. Matzutt, R., Henze, M., Ziegeldorf, J.H., Hiller, J., Wehrle, K.: Thwarting Unwanted Blockchain Content Insertion. In: Chandra, A. (ed.) 2018 IEEE IC2E, pp. 364–370. IEEE, Piscataway, NJ (2018)
47. Biryukov, A., Tikhomirov, S.: Deanonymization and Linkability of Cryptocurrency Transactions Based on Network Analysis. In: Proceedings of the 4th IEEE EuroS&P, pp. 172–184. IEEE (2019)
48. Jawaheri, H.A., Sabah, M.A., Boshmaf, Y., Erbad, A.: Deanonymizing Tor hidden service users through Bitcoin transactions analysis. Computers & Security 89 (2020)
49. Ciaian, P., Rajcaniova, M., Kancs, d.'A.: The digital agenda of virtual currencies: Can BitCoin become a global currency? Inf Syst E-Bus Manage 14, 883–919 (2016)
50. Möser, M., Soska, K., Heilman, E., Lee, K., Heffan, H., Srivastava, S., Hogan, K., Hennessey, J., Miller, A., Narayanan, A., et al.: An Empirical Analysis of Traceability in the Monero Blockchain. Proc. of PoPETs (2018)