

ICS 35.030; 35.240.99

**Privacy by Blockchain Design: Ein standardisiertes Verfahren für die
Verarbeitung personenbezogener Daten mittels Blockchain-Technologie;
Text Englisch**

Privacy by Blockchain Design: A standardised model for processing personal data using
blockchain technology;
Text in English

Zur Erstellung einer DIN SPEC können verschiedene Verfahrensweisen herangezogen werden:
Das vorliegende Dokument wurde nach den Verfahrensregeln einer PAS erstellt.

Gesamtumfang 58 Seiten

Contents

	Page
Foreword	4
Introduction.....	6
1 Scope	7
2 Normative references	7
3 Terms and definitions.....	7
4 Symbols and abbreviations.....	12
5 Personal data	12
5.1 Personal data in general	12
5.2 Defining Personal data	12
5.3 Practical consideration about identifiability and identifiers.....	13
5.4 Identifying personal data in a blockchain context	15
5.5 Requirement of an Anonymity assessment.....	17
6 GDPR awareness	18
7 Principles of data protection and their risks from the perspective of Privacy by Design.....	21
7.1 General	21
7.2 Fundamental principles of data protection	22
7.3 Assessing the risk of processing personal data	24
7.3.1 General	24
7.3.2 Traditional risk assessment methodology.....	24
7.3.3 Risk assessment from the perspective of data protection law	26
7.4 Initial assessment of risks in a blockchain application.....	26
8 Mitigating the risk of processing and decreasing identifiability through technical measures.....	26
8.1 General	26
8.2 Technical measures	28
8.2.1 Categories of technical measures of data protection:	28
8.2.2 Techniques to improve data protection or mitigate risk of processing.....	29
8.3 Architectural blueprint for an IT system processing personal data utilizing a blockchain-based tamper-proof access log.....	31
8.3.1 General	31
8.3.2 DLT-based tamper-proof access log	33
8.3.3 Decentralized Personal Data Storage.....	34
8.3.4 Consent Management System.....	34
Annex A (normative) Recommendations for handling personal data in blockchain applications	35
Annex B (normative) GDPR awareness	36
B.1 General	36
B.2 Controllership and processors in a BC/DLT-system.....	36
B.3 Right to Erasure (art. 17 GDPR)	38
B.4 Justifications for immutability.....	38
B.5 Right to rectification.....	39
B.6 Data Portability (art. 20 GDPR)	40
B.7 Processing Agreements between Controllers and Processors	40
B.8 Household exemption	40

B.9	Identification requirements for controllers.....	41
B.10	Personal data vs Privacy enhancing technology (ISO/IEC 27018).....	41
B.11	Automated decision making (art. 22 GDPR).....	41
B.12	Staff training + obligation (art. 29 and art. 32(4) GDPR).....	42
B.13	Data protection impact assessment (art. 35 GDPR)	42
B.14	Documentation + record of processing activities (art. 5(2) GDPR).....	43
B.15	Right to information (art. 13, 14 GDPR).....	43
B.16	Data minimization (Art. 5 (1) lit. c GDPR)	44
B.17	Data Protection Officer (Art. 37 (1) GDPR)	45
B.18	Privacy by Design & Default	45
B.19	Notification of data breach to authorities and data subjects (art. 33/34 GDPR).....	45
B.20	Right of access by the data subject	46
B.21	Right to object (art. 21 GDPR)	47
B.22	Transfer to third countries	47
Annex C (normative) Questionnaire: Extent of the implementation of data protection principles in a DLT, in particular blockchain solution.....		49
Annex D (informative) Summary of applicable risk assessment methodologies		52
Annex E (informative) Additional Information on Technical Measures		53
Bibliography		54

Foreword

This DIN SPEC has been developed according to the PAS procedure. The development of a DIN SPEC according to the PAS procedure is carried out in DIN SPEC (PAS)-consortiums and does not require the participation of all stakeholders.

This document has been developed and adopted by the initiator(s) and authors named below:

- Outright Solutions Gbr
Christian Wirth
- Forschungsinstitut für öffentliche Verwaltung Speyer
Michael Kolain
- Privacy by Blockchain Design
Anja Grafenauer
- TechGDPR DPC GmbH
Alex Carroll, Silvan Jongerius
- Nubessom Consulting
Alexis Cosson
- Juranos - Erbguth & Kirchert GbR
Jörn Erbguth, Katrin Kirchert
- Kedula GmbH
Oliver Guderjahn, Frank Jander
- Lightcurve GmbH
Frederic Hannesen
- KPMG Law Rechtsanwaltsgesellschaft mbH
Dennis Hillemann
- QRCrypto S.A.
Stiepan Aurélien Kovac
- AKQA GmbH
Harald Krefting
- IBM Deutschland GmbH
Elke Kunde
- BlockAxs GmbH
Dominik Müller, Fabian Pohl
- Cleopa GmbH
Detlef Olschewski

- Centrifuge GmbH
Martin Quensel
- zkSystems GmbH
Diana Rees
- Cryptom Technologies UG
Craig Sailor
- msg Systems AG
Steffen Schwalm
- WINHELLER Rechtsanwaltsgesellschaft mbH
Olga Stepanova
- HERE Technologies
Lisa Trujillo
- DIN Verbraucherrat
Kristina Unverricht

At present, there are no standards covering this topic in the body of German Standards.

DIN SPEC (PAS)s are not part of the body of German Standards.

A draft of this DIN SPEC (PAS) has not been published.

Despite great efforts to ensure the accuracy, reliability and precision of technical and non-technical information, the DIN SPEC (PAS)-consortium cannot give any explicit or implicit assurance or warranty in respect of the accuracy of the document. Users of this document are hereby made aware that the consortium cannot be held liable for any damage or loss. The application of this DIN SPEC (PAS) does not release users from the responsibility for their own actions and is applied at their own risk.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. DIN [and/or DKE] shall not be held responsible for identifying any or all such patent rights.

Provision of this document free of charge as a PDF via the Beuth WebShop has been financed in advance.

Introduction

The adoption of distributed ledger technology, in particular blockchain technology has been proceeding slowly in established markets. This is partly due to a high degree of legal uncertainty as to its compliance with data protection regulations. One of the reasons being that legal frameworks, such as the General Data Protection Regulation apply to data processing in single server structures operated by a legally and technically tangible intermediary. The resulting difficulty is to retrofit decentralized and distributed IT systems onto a normative concept.

One of the challenges concerning the compliance with data protection regulation is determining whether remains particularly unclear under which circumstances the data processed in an IT system using DLT, in particular blockchain technology (hereinafter called BC/DLT-systems) is considered personal data. Another such challenge with BC/DLT-systems, comes from legal analyses that rely on popular scientific simplifications of DLT, in particular blockchain technology. This tends to disregard potential technical approaches to resolving legal provisions.

The legal uncertainty regarding DLT, in particular blockchain technology and data protection law leads to the widespread assessment that “personal data should not be processed on the blockchain”. But (how) is this even possible? First attempts to resolve the issue by storing personal data “off-chain”, “anonymizing” it or only feeding M2M data (machine to machine) into a BC/DLT-system in order to, seemingly, circumvent data protection law (which only regulates personal data) have proven to be suboptimal and under-complex.

The DIN SPEC’s immediate aim is to help analyze the degree to which a natural person, in BC/DLT-systems, is identifiable, providing developers with tools and methods that preserve privacy and reduce the risk of legal uncertainty. The DIN SPEC looks at the risk for the data subject and lists and maps technical measures that increase the effort required to recreate any personal reference to a data subject.

The DIN SPEC classifies different approaches of handling personal data and exemplifies them with architectural blueprints. The latter describes technical design patterns in order to mitigate the risk and therefore raise the data protection level in an IT system.

Overall, the work in this DIN SPEC is guided by the “Privacy by Design” principle (art. 25 GDPR) and specifies technical and organizational measures in line with the principles laid out in the GDPR. It aims to lay a framework accessible to lawyers and computer scientists alike. The DIN SPEC 4997 committee claims to achieve the following goals by providing:

- A description of the functional requirements for DLT systems, in particular blockchain, to achieve compliance with General Data Protection Regulation requirements;
- A guide to the handling of personal data using blockchain technology;
- Architectural blueprints to illustrate the uses of blockchains to improve privacy;
- Procedures for business processes for the iterative maintenance and quality assurance of data protection.

1 Scope

This DIN SPEC establishes general principles for and methods of handling personal data in BC/DLT-systems. It specifies technical and organizational measures for data protection while taking into account the principles of “privacy by design” as well as specifications that are inspired by legal frameworks, such as the GDPR.

The document defines relevant terms for both technical as well as legal experts and establishes a methodological framework that helps to identify types of data (encrypted and not encrypted) as well as mapping legal principles of the GDPR to technical measures available to improve data protection or mitigate the risk of processing personal data in BC/DLT-systems.

This DIN SPEC is aimed towards establishing a high level of privacy in BC/DLT-systems. This document is applicable to all BC/DLT-systems.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

DIN SPEC 16597:2018-02 *Terminology for blockchains; Text in English*

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>

BSI, The Standard Data Protection Model, 7 The generic measures for the implementation of the Protection Goals. Available from: https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methodology_V1.0.pdf, cited as [NM1]

3 Terms and definitions

For the purposes of this document, the terms and definitions given in DIN SPEC 16597:2018-02, the GDPR and the following apply.

DIN and DKE maintain terminological databases for use in standardization at the following addresses:

- DIN-TERMinologieportal: available at <https://www.din.de/go/din-term>
- DKE-IEV: available at <http://www.dke.de/DKE-IEV>

3.1

Anonymity assessment

Reidentification assessment

a procedure to evaluate the residual risks of identifying a natural person with data that is claimed to be anonymous

3.2

Approved security function

security function (e.g. cryptographic algorithm) that is referenced in Annex C of ISO/IEC 19790:2012

[SOURCE: ISO/IEC 19790:2012, 3.8, modified — the reference to ISO/IEC 19790:2012 added to the definition]

3.3

Architectural blueprint

methodology developed by Wirth and Kolain [17] how to conclude a collection of system specifications and design patterns derived from legal requirements

Note 1 to entry: The purpose of an architectural blueprint is to show developers how their solutions might comply with the principle of Privacy by Design (Art. 25 GDPR). Its main function is not to be implemented in a specific product, but rather to give an idea of a technical reflection of legal requirements about IT-systems.

3.4

Bitcoin

an example of a blockchain using Proof of Work

[SOURCE: ITU FG DLT Terms and definitions [88], 6.5]

3.5

Cryptographic algorithm

well-defined computational procedure that takes variable inputs, which may include cryptographic keys, and produces an output

[SOURCE: ISO/IEC 19790:2012, 3.20]

3.6

Cryptographic key

key

sequence of symbols that controls the operation of a cryptographic transformation

[SOURCE: ISO/IEC 19790:2012, 3.23, modified — example deleted]

3.7

Certificate

entity's data rendered unforgeable with the private or secret key of a certification authority

Note 1 to entry: Not to be confused with a modules validation certificate issued by a validation authority.

[SOURCE: ISO/IEC 19790:2012, 3.12]

3.8

Decentralized application

DApp

application that runs on a decentralized system

[SOURCE: ISO/DIS 22739:2019, 3.18]

3.9

Distributed ledger

ledger that is shared across a set of DLT nodes and synchronized between the DLT nodes using a consensus mechanism

Note 1 to entry: A distributed ledger is designed to be tamper resistant, append-only and immutable containing confirmed and validated transactions.

[SOURCE: ISO/DIS 22739:2019, 3.22]

3.10**Distributed ledger technology***DLT*

technology that enables the operation and use of distributed ledgers

[SOURCE: ISO/DIS 22739:2019, 3.23]

3.11**Entity**

person, group, device or process

[SOURCE: ISO/IEC 19790:2012, 3.37]

3.12**Entropy**

measure of the disorder, randomness or variability in a closed system

[SOURCE: ISO/IEC 19790:2012, 3.38, modified — note deleted]

3.13**Ephemeral key**

cryptographic key that is generated for each execution of a key-establishment process and that meets other requirements of the key type (e.g., unique to each message or session)

Note 1 to entry: In some cases, ephemeral keys are used more than once within a single session (e.g., for broadcast applications) where the sender generates only one ephemeral key pair per message, and the private key is combined separately with each recipient's public key.

3.14**Hashed object**

input value mapped to the hash value through a hash function

3.15**Hash value**

output of a cryptographic hash function

[SOURCE: ISO/IEC 19790:2012, 3.53]

3.16**Input data**

information that is entered into a cryptographic module may be used for the purposes of transformation or computation using an approved security function

[SOURCE: ISO/IEC 19790:2012, 3.57]

3.17**Interoperability**

ability of two or more systems or applications to exchange information and to mutually use the information that has been exchanged

[SOURCE: ISO/DIS 22739:2019, 3.42]

3.18

IT System

set of one or more computers, associated software, peripherals, terminals, human operations, physical processes, information transfer means, that form an autonomous whole, capable of performing information processing and/or information transfer

[SOURCE: ISO/IEC 14662:2010, 3.13]

3.19

Key management

administration and use of the generation, registration, certification, deregistration, distribution, installation, storage, archiving, revocation, derivation and destruction of keying material in accordance with a security policy

[SOURCE: ISO/IEC 19790:2012, 3.64]

3.20

Ledger

information store that keeps records of transactions that are intended to be final, definitive and immutable

[SOURCE: ISO/DIS 22739:2019, 3.44]

3.21

Off-chain

related to a blockchain system, but located, performed or run outside that blockchain system

[SOURCE: ISO/DIS 22739:2019, 3.52]

3.22

On-chain

located, performed or run inside a blockchain system

[SOURCE: ISO/DIS 22739:2019, 3.54]

3.23

Peer-to-peer

relating to, using, or being a network of equal peers that share information and resources with each other directly without relying on a central entity

[SOURCE: ISO/DIS 22739:2019, 3.56]

3.24

Privacy by Design

a concept of a design principle addressed in GDPR Art.25

Note 1 to entry: It stipulates that data protection is to be taken into account in the conception and development of software and hardware that processes personal data.

3.25

Organisational Measures

any tactical, managerial and operational actions that support the chosen controls that secure information and protect personal data

3.26**Pruning**

function that deletes leaves, containing the transaction, and their hashes from the merkle tree while leaving the upper levels of the merkle tree and therefore the block's root hash intact

Note 1 to entry: The method has been proposed by Satoshi Nakamoto primarily "to save disk space" [68].

3.27**Record**

information created, received and maintained as evidence and as an asset by an organization or person, in pursuit of legal obligations or in the transaction of business

Note 1 to entry: This term applies to information in any medium, form or format.

[SOURCE: ISO/DIS 22739:2019, 3.67]

3.28**Security function**

cryptographic algorithms together with modes of operation, such as block ciphers, stream ciphers, symmetric or asymmetric key algorithms, message authentication codes, hash functions, or other security functions, random bit generators, entity authentication and SSP generation and establishment all approved either by ISO/IEC or an approval authority

[SOURCE: ISO/IEC 19790:2012, 3.106, modified — note deleted]

3.29**Technical Measures**

functions, processes, systems, procedures, controls and measures that are taken in order to protect and secure personal data that an organization processes

3.30**Timestamp**

time variant parameter which denotes a point in time with respect to a common time reference

[SOURCE: ISO/DIS 22739:2019, 3.75]

3.31**Trust**

degree to which a user or other stakeholder has confidence that a product or system will behave as intended

[SOURCE: ISO/DIS 22739:2019, 3.80]

3.32**Validation**

function by which a ledger record or block is validated

[SOURCE: ISO/DIS 22739:2019, 3.82]

3.33**Validator**

DLT node or other device or process in a distributed ledger system that participates in validation

Note 1 to entry: In some distributed ledger systems the DLT node that has the role of validator can digitally sign a ledger record or block

[SOURCE: ISO/DIS 22739:2019, 3.83]

4 Symbols and abbreviations

For the purposes of this document, the following abbreviated terms apply.

AML	Anti-Money Laundering
DLT	Distributed Ledger Technology
DPIA	Data Protection Impact Assessment
DPA	Data Processing Agreement
ECJ	European Court of Justice
ESB	Enterprise Service Bus
GDPR	General Data Protection Regulation
ISP	Internet Service Provider
TOM	Technical and Organisational Measures
URL	Uniform Resource Locator (cf ISO 19790 abbreviations)
ZKP	Zero-Knowledge Proof

5 Personal data

5.1 Personal data in general

The (legal) classification of personal data is technology-agnostic. The GDPR gives no normative indication on applying the term to specific technologies or systems. In order to lay a foundation for BC/DLT-systems, it is necessary to — in a first step — clarify what personal data is and what the properties for such a classification looks like *in general terms*. In order to apply the general classifications onto BC/DLT-systems, in particular blockchain systems, 5.4. will then discuss blockchain specific aspects of personal data, especially concerning identifiers.

NOTE The information in this section gives a short explanation of the legal term of “personal data” as used in the GDPR. It then provides notions of data linkage which are not specific to blockchain applications.

5.2 Defining Personal data

Art. 4(1) and recital 30 GDPR define personal data as information relating to a natural person that can be identified with that person.

Personal data = information + identifier

However, both the information and the identifier do not need to be explicitly present in the data. Having only the identifier might be enough when information can be derived from the context or with the help of available external data sources.

EXAMPLE In a team, referring to the tallest or shortest person is as good as naming them explicitly.

A person is *identified* if there is an immediate inference, while she is *identifiable* if it is possible to link a data to it by intermediate steps. In any case, the controller does not have to actually have identified a natural person — it is enough if there is the potential to do so.

Recital 26 GDPR defines a threshold that needs to be overcome in order to identify whether a particular data shall be qualified as personal data. In a decision by the European Court of Justice [1], the court ruled that dynamic IP addresses link back to a person, because there is a realistic and legal way to combine them with the information that the internet service provider (ISP) stores. This is due to the fact that the ISP keeps records which client has been attributed which IP address during a specific time interval.

Concerning the question, *whose* knowledge needs to be considered *to which degree* in identifying the data with natural persons, the GDPR takes a stance in its recital 26: Account should be taken of all means that the controller or another person may reasonably likely use. However, a means is only reasonably likely to be used, when there is a possibility that the data and the means to identify the data can be brought together.

NOTE This definition is in line with the definition of PII in ISO/EC 27000 (“To determine whether a PII principal is identifiable, account should be taken of all the means which can reasonably be used by the privacy stakeholder holding the data, or by any other party, to establish the link between the set of PII and the natural person.”). However, this DIN SPEC takes the terminology of the GDPR as its foundation.

When assessing the question if data that one wants to process is considered “personal data” under the GDPR, one must take into account the means which they themselves or another person is likely to use. As a result, if the identification of the person concerned is impracticable, for example because it would require a disproportionate amount of time, expense and manpower, the risk of identification is de facto negligible. By adding technical safeguards, any data processed by the entity can namely be moved towards the threshold of impracticability.

However, the exact threshold of impracticability has not (yet) been specified by legislation or jurisdiction. Furthermore, the classification of data to be personal might change over time through technological advancements, like improvements of computing power (Moore’s Law) as well as quantum computing. Both developments will impact blockchains in the future with considerable probability. Therefore, in case of doubt, an entity processing data which can theoretically be identified to a natural person can avoid risk by considering their application to be within the scope of data protection regulation, such as GDPR.

Before the legal background, there is thus the possibility that a specific data set is to be considered personal at some point, but is processed in a way using technical and organizational measures that it afterwards cannot be traced back to a specific person anymore with practical means. As a result, there can be parts of a data ecosystem where a personal information is stored as personal data, and other parts where the (original) data is transformed and stored in a way that the identification of a person behind it is impracticable. This will be the case if the means to be used to trace a stored data back to a natural person are so costly or unavailable that practically nobody can trace it back to a natural person anymore.

5.3 Practical consideration about identifiability and identifiers

One might understand the identifiability through the following example. Consider a data set that does not contain direct identifiers but overlaps with information that is publicly available.

Take for example a supposedly “anonymized” history of accessed web pages that does not contain any direct reference to the individual that has accessed these web pages. When all personal information (parameters etc.) in the URLs are removed, it seems impossible to identify a person with a list of web pages. However, linking it to Twitter profiles where people have shared a number of links can be enough to identify their history of accessed web pages.

Therefore, any combination of innocent looking attributes might be used to identify a person. Even hashed or randomly generated identifiers can turn into personal data, if some other data that is stored with that data can be used to identify a person with the identifier.

To get a clearer understanding of the term “identifier”, it is helpful to consider its practical implications in scenarios of the processing of personal data.

Typical direct identifiers in a digital ecosystem can be

- name;
- photo;
- fingerprints;
- biometric data.

Typical identifiers that can be used to identify a person are

- passport number;
- telephone number;
- email address;
- IP address;
- public key of a blockchain account.

Other data that might be used — especially in combination — to identify a person seems to be limitless:

- surfed web pages;
- location data;
- affiliation with certain user groups;
- purchasing behaviour (specific purchases or purchasing preferences);
- health records.

When considering whether you are in possession of personal data, ask yourself:

- a) Is this information about natural persons or can any kind of information be derived from that data using external data sources or tools. The most trivial kind of information is sufficient to confirm this question. Only data that does not imply any information about a natural person at all, will not be considered personal data.

If the answer to this first question is “yes”, ask:

- b) Is it possible to identify the persons concerned by the data? Is this not only theoretically possible, but can be done by you or somebody else through reasonable practical means? Can it be done with current or with future technology?

If your answer is both times in the affirmative, we recommend that you assume the information in your possession to be personal data.

5.4 Identifying personal data in a blockchain context

When working with DLT, in particular blockchain solutions, particular challenges are brought up by

- the combination of transparency of data and metadata on the blockchain,
- the immutability aspects,
- the present and future potential for data analytics, and
- third party data to be linked which could identify blockchain users and other natural persons.

In order to evaluate the possibility to identify a natural person, a closer look needs to be taken at typical data that is put on blockchains:

Public addresses and transaction data

Public addresses and transaction data might be considered personal data when they are combined. Most often, a pure public address is a random sequence of numbers and letters. A transaction only reveals the transactional data which is by itself of technical nature. However, when combined, public addresses and transaction data often allow the identification of transactions with natural persons. This is especially true for systems that underly AML regulation like exchanges and wallets service providers because they are required to store identification records related to public addresses. Service providers also use internet traffic information to offer identification services for addresses. Using new addresses for every transaction already makes identification harder. Using Zero Knowledge Proofs (ZKPs) might render identification difficult enough to not reach the threshold of recital 26 GDPR.

Encrypted personal data

Encryption transforms the original data object into a ciphertext that conceals the information contained in the original data object. Only when an approved security function is used, it is guaranteed that the entity in possession of the private key is the only one that can recover the original data object providing the information in question.

Furthermore, technological advancements like quantum computers need to be taken into account [2]. This DIN SPEC recommends using approved security functions in order to adhere to the requirements of the GDPR to take foreseeable technological advancements into account. The former consideration applies to both asymmetric and symmetric encryption algorithms (including hashing functions based on symmetric crypto primitives), both being affected by quantum computers once large-enough such computers become available to attackers. For details, refer to [2]. In particular, the attention of the reader is drawn to the NIST competition on post-quantum asymmetric cryptography, likewise the ISO study period on the inclusion of eAES [84] in ISO/IEC standards, as well as X.1197amd1 [85] for emerging standards catering to the need of replacing existing cryptography.

Besides this only the availability of the private key leads to the classification of encrypted data as personal data under the following conditions:

- The private key is securely deleted. Then the data is not considered personal data anymore, because there is no practicable way to decrypt and take notice of the data;
- The private key is only known to the data subject. Here, it is not clear, whether the encrypted data is considered personal data. Even if not, the data subject might give away or lose the key and the data is considered personal data;
- The private key is known to other people. Then the data is considered personal data.

Storing encrypted personal data on immutable blockchains might also violate ICT-security standards [3]. This is so, because encryption on blockchains works like access control with an unchangeable password.

NOTE A possible solution to this is pruning, that can extend the capabilities to support a delete-operation on transactions without compromising the integrity of the Blockchain on a technical level. Since pruning is not a mandatory feature of every blockchain this DIN SPEC recommends to consider all the implications and responsibilities of a potential controller, when deciding to write encrypted personal data to ledger that will under no circumstances allow deletion of this data.

Anonymization, deletion and publication

Anonymization can be considered equal to deletion [4]. However, in case some newly available data makes it possible to reverse the (failed) anonymization, the data is neither considered to be anonymized nor deleted. When anonymous data is published, it is yet unclear, who is responsible, when the anonymous data can be identified through means that were neither available nor perceivable at the time of publication. The term anonymization is often only used when some personalized data sets are taken and by stripping some column or doing other changes in order to not call the remaining data personal data anymore. However, the process is different in the context of blockchains: Data is often not retroactively anonymized, but a system design tries to avoid storing personal data on a blockchain.

Hash values

Hash values can be used in BC/DLT-systems in many contexts, not specifically limited to constructing the root hash through a Merkle tree.

When storing hash values of personal data on blockchains, this needs to be considered personal data in certain situations:

- The data has **low entropy**. This makes it possible to derive the original data from the hash value. Bitcoin-miners are extremely fast and can reverse hashes, by trying all possible values of the original object. There should be at least more than 10^{30} equally possible values for the original object.
- **Part of the data is known**; the rest of the data has low entropy. Here again, reversing the hash value is possible.
- When the **hash is used as a pseudonym to connect it to other data stored alongside the hash on the blockchain**. In this case, the original object can be connected to this other data. The same is true when the additional information is given through the context of the hash value on the blockchain. This enables connecting the data in the hashed object with additional data.
- If the above factors (high entropy; no parts of data known to others; hash not used as a pseudonym; no additional context stored on blockchain that allows re-identification) are met and the **hashed object is securely deleted** (without remaining copies anywhere), the hash does not reveal any information about the data subject.
- A special case applies, when only those who have access to the hashed object itself can make sense of the information behind the hash. In this case, however, the information about a natural person that can be derived by re-connecting the hash value to the hashed object is already contained in the hashed object. In other words: for those who already have access to the hashed document, the hash does not reveal any additional information (again — only if the above factors are met, especially if the context of the stored data on a blockchain does not reveal any additional personal information). However, this line of argumentation has not yet been confirmed by a data protection authority or a court [5].

For those using hash values as a technical vehicle to store (a reference to) personal data on a blockchain, this DIN SPEC requires only to use data with high entropy. Special attention must also be paid to where and by

whom the hashed object (e.g. a scanned document) can be processed and in which context the hash is stored on a blockchain.

NOTE The hash-functions given in DIN SPEC 3104 can be used.

5.5 Requirement of an Anonymity assessment

Since the term “personal data” is not defined in a mathematical sense, the burden to identify a natural person with some information can be increased beyond the level of recital 26 and the threshold of the European Court of Justice. This can be done through a combination of **technical measures** and **organizational measures**.

This DIN SPEC does not give a detailed guideline on when a specific DLT, in particular blockchain application can be considered not using personal data, but rather describes possible technical measures to move the threshold of identifiability towards the threshold of impracticability.

- However, this DIN SPEC requires those who publicly claim to have combined technical and organizational measures in a way that there are no practical means to identify a natural person behind data stored on a blockchain to **carry out an (external) Anonymity-assessment** on their running system.
- This DIN SPEC also requires them to clearly point out which parts of the BC/DLT-systems are storing personal data and which are not. For those parts of the BC/DLT-systems that operate with personal data, the controllers have to pay close attention to fulfilling the legal requirements put forward by data protection regulation such as GDPR (see Annex B). They also have to consider the need for a design approach that follows the virtues of Privacy by Design (see the following sections).

Anonymity-Assessment specifications

In order to “measure the required effort, necessary to reidentify personal data that has been hashed to become non-personal data in the blockchain-context” we require the controller to obtain an assessment by an external independent party. The assessment must contain a clear description of:

- 1) The underlying measures that have been taken leading to the assumption that the data processed by a DLT, in particular blockchain application is considered to be non-personal in the blockchain-context (provided by the controller);
- 2) The risks and resulting reidentification-scenarios that have been taken into account to penetrate the measures to conceal the personal information (provided by the external party carrying out the Anonymity-assessment assisted by the controller);
- 3) The report and conclusion about the penetration test that have been conducted to stress-test the hypothesis of the data-controller (provided by the penetration-tester independently from the data controller).

6 GDPR awareness

A BC/DLT-system that operates with personal data falls into the scope of data protection regulation, such as the GDPR (art. 2(1) GDPR). It therefore has to comply with several legal requirements. Since there is currently no jurisdiction concerning blockchain technology, this DIN SPEC is confined to give an outline of challenges that might arise during the development of BC/DLT-systems. Thus, the following section merely serves as a check-list for DLT, in particular blockchain applications in order to create awareness for possible legal challenges in a development process that follows the ideal of Privacy by Design. While this section is deliberately kept short, Annex B gives a more detailed introduction to the normative requirements of data protection law that have to be considered while developing a state of the art DLT, in particular blockchain system.

1) Controllership in a BC/DLT-system

Legal science has not yet found a uniform way to attribute the role of “controller” (legal definition in art. 4(7) GDPR) to a specific entity within the framework of BC/DLT-systems. This constitutes a major source of legal uncertainty, since the data controller is the central figure when it comes to fulfilling legal requirements of data protection law.

Determining the participants on a permissioned blockchain is relatively easy. However, it still can be a challenge, to determine, whether only the central authority providing permissions or also others are to be considered (joint-)controllers. However, different participants could also be considered as joint controllers according to art. 26 GDPR. The situation is different for permissionless blockchain networks that include several participants, including the miners, validators, and other roles depending on the nature and particular implementation of the blockchain and its network. Here, the legal debate is still ongoing. [45]

2) Right to Erasure (art. 17 GDPR)

The right to erasure is technically feasible on a single computer from a technical point of view. Implementing this GDRP request in a decentralized network is relatively more difficult because deleting it on the computer of a single node does not result in deletion across the network. Although some technical measures have been proposed to delete specific transaction content from a distributed ledger (e.g. pruning), it strongly depends on the existence of an enforceable governance mechanism that allows deletion “without undue delay” (Art. 17 (1) GDPR) upon request of a data subject.

3) Justifications for immutability

The essence of DLT, in particular blockchain technology to store data in a temper proof ledger, constitute a barrier for the achievement of the GDRP goal of accuracy. It poses a technical challenge to implement the right to erasure in BC/DLT-systems (see 2.). However, Art. 17(3) GDPR regulates exemptions from the duty to erase data upon request of a data subject. If a DLT, in particular blockchain application can persistently comply with one of these exemptions (e.g. “processing is necessary (...) for the establishment, exercise or defense of legal claims”, Art. 17 (3)(e) GDPR), the immutability feature of a blockchain application can be justifiable.

4) Right to rectification (Art. 16 GDPR)

When data is incorrect, the data subject has the right to demand that the data is corrected (art. 16 GDPR). In BC/DLT-systems, it is not entirely clear whether an additional entry that corrects incorrect information but does not delete it entirely will be sufficient to comply with the legal requirement. Although information or transactions can be invalidated with newly appended blocks, anyone can view incorrect data. Even if this technical solution was sufficient, its implementation would still be a significant hurdle especially in large public permissioned DLT, in particular blockchain networks.