Blockchains and Data Protection in the European Union

*Michèle Finck**

This article examines data protection on blockchains and other forms of distributed ledger technology. Whereas the General Data Protection Regulation was fashioned for centralised methods of data collection, storage and processing, blockchains decentralise each of these processes. We engage with the resulting tensions in the below analysis.

I. Introduction

This article examines data protection on blockchains and other forms of distributed ledger technology (DLT).[1] The EU General Data Protection Regulation's (GDPR) imminent entry into force coincides with pronounced hype surrounding blockchain as a new paradigm of data storage and management.[2] A blockchain is in essence an append-only decentralised database that is maintained by a consensus algorithm and stored on multiple nodes (computers). While the technology is still immature and applications remain rare it is widely viewed as a disruptive force, capable of decentralising business models, forms of human interaction and markets.[3] From a data protection perspective, the rise of the blockchain may be no less transformative. Whereas the GDPR was fashioned for a world where data is centrally collected, stored, and processed, blockchains decentralise these processes. With a paradigm shift of such radical contours, we must enquire about the applicability of a legal framework constructed for a sphere of centralisation to one of decentralisation.

We will observe that at least at first sight blockchains (especially those that are public and unpermissioned) and the GDPR are profoundly incompatible at a conceptual level as the data protection mechanisms developed for centralised data silos cannot be easily reconciled with a decentralised method of data storage and protection. Even where data is encrypted or hashed it qualifies as personal data under EU law. The cryptographically modified data stored on a distributed ledger, in addition to public keys, are hence subject to the GDPR. Herefrom results a risk that data protection legislation renders the operation of blockchains unlawful, hence asphyxiating the development of an innovative technology with much promise for the Digital Single Market. To distill how this consequence should be accounted for we must reflect on the status of innovation in EU law. The tension between the GDPR and these novel decentralized databases indeed reveals a clash between two normative objectives of supranational law: fundamental rights protection on the one hand, and the promotion of innovation on the other. The article will highlight, however, that legal interpretation techniques and technological solutions can facilitate an at least partial reconciliation of these apparently conflicting rationales. Blockchains are a technology that might in the future achieve some of the objectives inherent to the GDPR through technological means, although through mechanisms distinct from those envisaged by the legal framework itself.

In their current state DLT will in most, if not all, instances be incompatible with the GDPR as the specific requirements of the EU data protection framework cannot be easily applied to distributed ledgers. In the future, however, they could be compatible on a meta-level, as, if properly designed, blockchains can pursue the GDPR's underlying goal of giving a data

DOI: 10.21552/edpl/2018/1/6

[*] Dr Michèle Finck, Senior Research Fellow at the Max Planck Institute for Innovation and Competition, Munich; Lecturer in EU Law, Keble College, University of Oxford. The author thanks Christian Sillaber for very helpful comments and discussions. For correspondence: <Michele.Finck@ip.mpg.de>.

[1] The article will refer to blockchains and other forms of distributed ledger technology interchangeably. This should not, however, obscure the significant technical distinctions between them and also between various models of blockchains themselves.

[2] Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data [2016] OJ L119/1.

[3] It is often claimed that blockchain 'disintermediate' the economy. This remains to be seen as, for the time being, more intermediaries have created by the technology than replaced.

subject more control over her data. The analysis will conclude by underlining that in order to achieve the latter, we must be willing to adapt law to technological change and be accepting of greater techno-legal interoperability. This does not mean that data protection should be weakened but rather that it is worth exploring whether the GDPR's objectives can be achieved through means different from those originally envisaged. This does not, however, mean that blind trust should be placed in the technology. Blockchains by no means automatically support data sovereignty but rather must be purposefully designed to do so as blockchains can also constitute a danger to data protection. Regulators must, in insisting on the core of data protection regulations whilst also showing flexibility regarding the specific mechanisms employed, nudge blockchain developers to design their products in compliance with this important public policy objective.

This argument unfolds in five steps. We shall first briefly introduce distributed ledger technology before evaluating the application of the GDPR to blockchains to establish that public and unpermissioned blockchains, built to achieve decentralisation, cannot be straightforwardly reconciled with a legal framework targeting centralised data silos. The implications of that finding are then evaluated. We conclude by arguing that a compromise is needed where the legal certainty of data protection in the Union is reconciled with the desired promotion of innovation and thus also alternative, and maybe more effective, means of data protection.

II. Data on Blockchains

The present section lays out the background of our analysis in providing a cursory overview of blockchains and other forms of distributed ledger technology. It must be clear from the outset that there is huge variance in distributed ledgers and their internal governance structures. In its essence, a distributed ledger can be described as a shared and synchronized digital database that is maintained by a consensus algorithm and stored on multiple nodes (computers). Blockchains are both a new technology for data storage as well as a novel variant of programmable platform and network that enables new applications such as smart contracts.[4] The term 'blockchain' is often used to denote any kind of distributed ledger, including those that do not store data in blocks. Technically, however, blockchains only designate the variants of DLT that record data in packages ('blocks') that are hashed ('chained') to another. For the sake of simplicity, and to reflect the as of yet unsettled terminology in this domain, we shall refer to both notions interchangeably.

Rather than being a completely novel technology, a blockchain is better understood as a combination of previously existing mechanisms such as distributed ledgers, asymmetric encryption and merkle trees, that were linked together to enable Bitcoin in 2009.[5] In the years following the emergence of this cryptoasset, more and more observers stressed DLTs' capacity to widely serve as a replicated record of data and digital assets that can be operated between parties that do not know or trust each other without the need for a trusted third party. This has led developers to build on the Bitcoin blockchain[6], create new blockchains[7] and other forms of distributed ledger technology to fashion a wide range of use cases. Even though the technology is still in its early stages of development, applications facilitated by DLT range from different forms of digital assets over mobile banking[8], tracking goods in international trade[9], arranging payments for the Internet of Things[10] and land registries[11], to name just a few.

To understand blockchains' implications from a privacy perspective, we must delve a bit deeper into their technical details. On a 'blockchain', data is usually grouped into blocks that, upon reaching a certain size, are chained to the existing ledger through

[4] For an overview of smart contracts, see Kevin Werbach and Nicolas Cornell, 'Contracts Ex Machina' (2017) (forthcoming Duke Law Journal 2018); Markus Kaulartz and Jörn Heckmann, 'Smart Contracts – Anwendungen der Blockchain Technologie' (2016) 9 Computer und Recht 618.

[5] Satoshi Nakamoto, 'Bitcoin: A Peer-to-Peer Electronic Cash System' (2009) <https://bitcoin.org/bitcoin.pdf> accessed 5 March 2018 (hereafter Nakamoto, 'Bitcoin').

[6] These blockchains are not necessary cryptocurrency related but can take a wide range of forms.

[7] Such as the Ethereum blockchain.

[8] An example would be BitPesa, which has revolutionised mobile payments in sub-Saharan Africa.

[9] Everledger tracks diamonds while Walmart is using blockchains to track its goods.

[10] Iota provides a DLT solution specifically for this domain.

[11] Gertrude Chavez-Freyfuss, 'Sweden tests blockchain technology for land registry' *Reuters* (16 June 2016) <https://uk.reuters.com/article/us-sweden-blockchain/sweden-tests-blockchain-technology-for-land-registry-idUKKCN0Z22KV> accessed 5 March 2018.

a hashing process. Through this process, data is chronologically ordered in a manner that makes it difficult to tamper with information without altering subsequent blocks.[12] Tamper-evidence is indeed one of blockchains' most heralded features and some would consider it its core value proposition. It is in this context often stated that blockchains are 'immutable'. This terminology is misleading as even though it is very difficult and to amend blockchains, it is not impossible and has in the past been done such as on the occasion of the DAO hack.[13]

DLTs rely on a two-step verification process with asymmetric encryption. Every user has a public key (a string of letters and numbers representing the user), best thought of as an account number that is shared with others to enable transactions. In addition, each user holds a private key (also a string of letters and numbers), which is best thought of as a password that must never be shared with others. Both keys have a mathematical relationship by virtue of which the private key can decrypt data that is encrypted through the public key. Public keys thus hide the identity of the individual unless they are linked to additional identifiers. The nodes are the computers on which the ledger is stored. Some DLTs operate a distinction between 'full' and 'lightweight' nodes whereby only full nodes store an integral copy of the ledger from the genesis block whereas lightweight nodes only store those parts of the ledger of relevance to them. In public and permissionless blockchains, anyone can entertain a node by downloading and running the relevant software. Some (but not all!) nodes also function as 'miners', which aggregate transactions into candidate blocks and hash a new block to the chain on the basis of a predetermined consensus protocol (such as proof-of-work or proof-of-stake).

It must be plain from the outset that on a decentralised ledger data can be stored in a variety of different forms. First, it is possible to store data, such as a document or digital art, on the ledger in plain text. This is however problematic for a number of reasons. On a permissionless blockchain, anyone can arbitrarily read such data, which is of course highly undesirable from a privacy perspective. Blocks have moreover limited storage capacity and storage is often expensive so that this would not be an economical solution. Rather than storing data in plain text, it is usually encrypted or hashed before it is added to a blockchain. Most DLTs contain two types of data: (i) the header which includes the timestamp, the identity of the data's source such as an address and the previous block hash, whereas the block content (or payload) contains the data to be stored (on the Bitcoin blockchain this would be the relevant transactions as well as the coinbase transaction[14]). Whereas the header is usually not encrypted, the payload normally is.

Where data is encrypted, in principle, only a user in possession of the private key can decrypt the documents. On blockchains, asymmetric cryptography is used as a means to generate digital signatures. Encryption is a two-way function, meaning that with the right cryptographic key, previously encrypted data can be 'unlocked' and reverted to its original state. This security technique renders data unintelligible to individuals without authorised access.[15] While data is in practice often encrypted, this is a completely optional process that developers must chose. The block header is usually[16] not encrypted given that for nodes to process a cryptoasset transaction, they for instance need to verify whether the relevant wallet holds the required funds.[17] Data can also be hashed to a distributed ledger. The hashing process can register large amounts of data with a small digital fingerprint. Under the common SHA 256 hashing algorithm, any amount of data will be reduced to a 32-byte hash value.[18] A cryptographic hash is a one-way function that cannot be reverse engineered, meaning that there is no key that can unlock data that has been hashed.[19] Hashes allow for the verification of whether a certain

[12] Whereas data stored on a blockchain is often described as 'immutable', this is not quite the case as such information can be modified in exceptional circumstances through human intervention, which however requires the collusion between a majority of the network's nodes.

[13] See further Conte de Leon et al, 'Blockchain: Properties and Misconceptions' (2017) 11 Asia Pacific Journal of Innovation and Entrepreneurship 268.

[14] This refers to the transaction realizing the mining reward.

[15] Lessig euphemistically declared it 'the most important technological breakthrough in the last one thousand years'. See Lawrence Lessig, *Code and Other Laws of Cyberspace* (Basic Books 1999) 35 (although cryptography has been used before).

[16] This is not always the case. Zcash for instance encrypts the sender and recipient as well as amount of data within single-signature transactions.

[17] This can all be a bit abstract. The following website provides live coverage of the Bitcoin blockchain and illustrates this further: <https://tradeblock.com/bitcoin> accessed 5 March 2018.

[18] SHA-256 is a hashing algorithm created by the NSA, which is considered particularly secure. It always generates a 32-byte hash value, notwithstanding the size of the original data.

[19] This, as many things, may change with quantum computing.

document was stored in a database at a given time, as re-hashing the off-chain version of that document will produce the exact same hash.[20] We have already noted that there is a large diversity of DLTs and related applications. Importantly, we also cannot predict which blockchains or blockchain-like databases will see broad adoption in the future. There is indeed at least a possibility that the technologies eventually deployed to enable use cases that are now experimented with will have considerably different properties from first and second generation blockchains.

The original Bitcoin blockchain is a public and unpermissioned (or 'permissionless') blockchain, which means that it is open-source and open-access so that anyone can create a Bitcoin address and download or design software to run nodes. Unpermissioned blockchains are the farthest away from standard conceptions of traditional databases, and unsurprisingly raise the highest conceptual challenges under data protection law. Blockchains can however also be private and permissioned, which means that they can run on a private network such as intranet or a VPN (as opposed to the Internet) and an administrator needs to grant permission to individuals wanting to maintain a node. The key distinction between permissioned and unpermissioned blockchains is indeed that while one needs access permission to join the former, this is not necessary in respect of the latter. In addition to public and private blockchains, hybrids have emerged· Given that unpermissioned blockchains offer most novelty and complications from a data protection perspective our focus rests mainly on them. We now turn to an analysis of blockchains' implications from a data protection perspective.

III. Blockchains: Promises and Perils for Data Protection

Blockchain developers are currently struggling to determine whether they can legally store and process personal data on their ledgers. This answer will largely depend on whether such activity falls within the scope of the EU's data protection regime. Before turning to a detailed analysis of the GDPR, we first engage with the implications of DLT for data protection to set the scene.

For our purposes, the most relevant aspect of DLT is its degree of differentiation to conventional forms of data storage. Blockchains offer a record-keeping function that dispenses from the need for third-party intermediation[21] and by analogy can decentralise the collection, storage and processing of data. This stands in sharp contrast with the current data economy, characterised by economic centralisation in the form of 'platform power'.[22] Large intermediaries such as Google, Amazon, Apple and Facebook control how we search, shop and connect. They autonomously collect, store, process and monetise our data trails.[23] This, in turn, enables them to expand their position of power in building on the data mountains they sit on, for instance to train new algorithms. Such market power has caused concern from a competition policy perspective as it burdens market entry. The issues engendered by these circumstances are two-fold, relating on the one hand to economic operators' market position, and, on the other, the protection of privacy.

Regarding the latter, blockchains offer the promise of the decentralised handling of data and data sovereignty, a concept that focuses on giving individuals control over their personal data and allowing them to share such information only with trusted parties.[24] The GDPR shares the data sovereignty objective as it aims to give natural persons 'control over their own personal data'.[25] The right to data portability in Article 20 GDPR enshrines this objective in allowing a data subject to receive data from a controller in order to give it to another controller. The right to data portability is an emergent concept in EU law, the contours of which remain largely undefined. There is no doubt, however, that it seeks to give data subjects more control over personal data. The Article 29 Working Party for instance considers that the 'primary aim of data portability is enhancing individuals' control over their personal data and

[20] This has enabled solutions that offer a timestamping service. See by way of example: <https://www.bernstein.io/> accessed 5 March 2018.

[21] Unless we count miners as intermediaries. It is worth noting that even if we do, these would be a different class of intermediaries as they are perfectly interchangeable.

[22] Orla Lynskey, 'Regulating "Platform Power"' (2017) LSE Legal Studies Working Paper 1/2017.

[23] See also recital 6 GDPR.

[24] For an overview, see 'Identity as a Bottleneck for Blockchain' (*BlockchainHub*, 17 October 2017) https://blockchain-hub.net/blog/blog/decentralized-identity-blockchain/ accessed 5 March 2018.

[25] recital 7 GDPR.

making sure they play an active part in the data ecosystem'.[26]

It is important to note that the precise meaning of data portability and sovereignty, in the GDPR and elsewhere, remains unsettled. This is an important point as there are as of yet no solutions that would provide data subjects with *total* control over their data, just those that provide *more* control compared to the current status quo. Many predict that DLTs can be fashioned so that only the user has access to the public and private key, deciding freely as to when she reveals her data with external parties.[27] Unlike ID cards or conventional medical records, blockchains promise selective data sharing through adequate applications, ensuring privacy and reducing the risk of identity theft.[28] Blockchains *could* thus facilitate new forms of identity management by enabling individuals 'to control access to their identity information and to create, manage and use a self-sovereign identity'.[29] Whether this will be the case, however, remains to be seen. It is for instance true that selective sharing is possible, yet what about the fact that once data is revealed, those with access will generally be able to copy and extract data and store it perpetually? Yet, as the technology develops many proposals for the decentralised personal data management system circulate that would empower users to own and control their data. [30] These projects must be evaluated with a critical eye yet should not be dismissed from the outset as technology could indeed come to realize the objectives set out in the GDPR.

While the promise of DLT for data sovereignty should not be downplayed, it is also important to remain realistic and vigilant at a time where blockchain hype and hybris sometimes cloud rational judgment. Blockchains are authenticity solutions that do not, in themselves, provide any privacy guarantees so that for data sovereignty objectives to be achieved, they must be combined with additional mechanisms. Indeed, despite the technology's promises for data sovereignty, there are also perils for if the necessary safeguards are not implemented; blockchains can reveal any and all data stored on them. As a new technology, blockchains are malleable and can develop in a number of directions. It is here where law, technology and innovation must meet and where dialogue between innovators and regulators must occur to ensure that innovation can occur, yet in a fashion that is desirable for the public good. Much will thus depend on blockchains' design, which must reflect technological requirements as well as public policy considerations. Section 5 returns to the examination of regulators' incentivising role to make sure that rights are adequately protected in the face of technological transformation. In order to determine how DLTs relate to one such consideration, namely data protection we now turn to examine blockchains from the perspective of the GDPR.

IV. The EU's General Data Protection Regulation

To pursue the dual objectives of data protection and the free movement of personal data in the internal market, the European Union has opted for an ambitious data protection framework, the General Data Protection Regulation that becomes binding on 25 May 2018, replacing Directive 95/46/EC.[31] Technological developments such as the rise of platform intermediaries have triggered new challenges for data protection as the scale of data sharing and collection have steadily increased. In this context, a stronger and more coherent legal regime was deemed necessary.[32] This novel legal framework will apply to the personal data of natural persons that is wholly or partly automated or stored in a filing system.[33] Given that May 2018 is just around the corner, blockchain developers and entrepreneurs are currently anxiously trying to determine whether the GDPR applies to their activities, for if this is the case their leeway for experi-

[26] Article 29 Working Party, 'Guidelines on the Right to Data Portability' (2017) 16/EN WP 242, 4, fn 1.

[27] Michael Mainelli, 'Blockchain could help us reclaim control over our personal data' (5 October 2017) Harvard Business Review <https://hbr.org/2017/10/smart-ledgers-can-help-us-reclaim-control-of-our-personal-data> accessed 5 March 2018.

[28] Instead of having to show your ID at a supermarket to buy alcohol or reveal all medical data to a doctor to indicate prescription medicine currently used, these pieces of information could be revealed in isolation. For an example, see <https://shocard.com/>.

[29] Clare Sullivan and Eric Burger, 'E-Residency and Blockchain' (2017) 33 Computer Law & Security Review 460, 475.

[30] Guy Zyskind et al, 'Decentralizing Privacy: Using Blockchain to Protect Personal Data' (IEEE Security and Privacy Workshops, 2015) 180 <http://ieeexplore.ieee.org/document/7163223/> accessed 5 March 2018 (hereafter Zyskind et al, 'Decentralizing Privacy').

[31] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31.

[32] recitals 6 and 7 GDPR.

[33] arts 1 and 2 GDPR.

mentation and innovation risks being considerably constrained. Bearing in mind the important distinctions between various forms of DLT and the corresponding need for a case-by-case analysis, we attempt to provide a general overview of the application of the GDPR framework to DLTs, starting with the question of whether data related to a natural person stored on a decentralised ledger qualifies as personal data as a matter of EU law.

1. The GDPR's Material Scope: Does Data Stored on a Blockchain Qualify as Personal Data?

This section enquires whether public keys and other data fall within the scope of the GDPR. The Regulation only applies to 'personal data', defined as 'any information relating to an identified or identifiable natural person'; the 'data subject'.[34] An 'identifiable person' is defined as a natural person that

> can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, psychological, genetic, mental, economic, cultural or social identity of that natural person[35]

Where data is rendered completely anonymous, it no longer amounts to personal data and thus falls outside the scope of the legal framework. Where data is rendered pseudonymous, however, it continues to qualify as personal data as the indirect identification of a natural personal by an identifier remains possible. Two sets of data stored on blockchains can potentially be defined as personal data for the purposes of the GDPR; transactional data stored in the blocks as well as public keys.[36]

a. Personal Data Stored on a DLT

Depending on the respective DLT's use case, data stored blocks may be data related to an identified or identifiable natural person such as data revealing individual behaviour in Internet of Things use cases; digital identities; or financial and medical data. To distinguish this data, which often contains personal information, from other data such as personal keys we will refer to it as 'transactional data'. Many current use cases revolve around transactions, which usually contain specific information related to a person. We have already observed that this data can be stored on a blockchain in three alternative fashions: in plain text, in encrypted form, or by hashing it to the chain. This section evaluates whether these processes can sufficiently anonymise personal data to allow it to evade the GDPR's scope of application. It is worth noting that the distinction between personal and non-personal data is likely to vanish over time as sophisticated machine learning techniques may enable the identification of individual characteristics and behaviour through non-personal data.[37]

The threshold for anonymisation under the Regulation is high and only results 'from processing personal data in order to *irreversibly prevent identification*'.[38] Personal data stored on a blockchain in plain text clearly remains personal data for the purposes of the GDPR so that this option does not merit any further analysis. Where data is encrypted it can still be accessed with the correct keys, meaning that it is not irreversibly anonymised. Encrypted data can for example be connected to the data subject where transactions are effected for off-chain goods or where cryptoassets are converted into fiat currency. Encryption is considered a pseudonymisation technique under the EU data protection regime given that the data subject can still be indirectly identified so that it can, on its own, not be considered as an anonymisation technique.[39] The conclusion that transactional data that has been encrypted remains personal data for the purposes of the GDPR is accordingly unavoidable.

Transactional data that has been subject to a hashing process also qualifies as personal data under the GDPR. Whereas a one-way hash function that cannot be reverse-engineered can offer stronger privacy guarantees than encryption it will not allow data to evade the qualification as personal data for GDPR purposes. The Article 29 Working Party has been un-

[34] art 4(1) GDPR.

[35] ibid.

[36] It is important to remember that there is a huge variance in blockchains and that the link between the encrypted data hashed to the chain and an individual will accordingly vary.

[37] Similarly data that is now anonymous may become personal data due to technological developments.

[38] Article 29 Working Party, 'Opinion 04/2014 on Anonymisation Techniques' (2014) 0829/14/EN, 20 (emphasis added) (hereafter Article 29 Working Party, 'Anonymisation Techniques').

[39] ibid.

equivocal that hashing constitutes a technique of pseudonymisation, not anonymisation as it is still possible to link the dataset with the data subject.[40] We thus conclude that transactional data that is encrypted or has undergone a hashing process will still be considered personal data for the purposes of the GDPR.

The conclusion that transactional data stored on a blockchain is subject to GDPR requirements may however be avoided in future times. First, it is imaginable that over time, some cryptographic processes such as SHA-256 or its SHA-3 successor will be declared capable of anonymising data by courts or the European Data Protection Supervisor.[41] Second, a number of technical solutions are currently being developed that may prevent transactional data from being directly stored on the blockchain. Buterin considers cryptographically secure obfuscation[42] as the 'holy grail' of privacy on blockchains but concedes that the tool is not sufficiently developed to be used.[43] While this solution remains unavailable, others can more readily be deployed. First, personal data could be stored off-chain and merely linked to the blockchain through a hash pointer. In such a scenario, personal data is recorded in a referenced encrypted and modifiable database and not on the blockchain.[44] A number of data-management and sovereignty solutions are currently being developed that for instance combine blockchain and off-chain storage to 'construct a personal data management platform focused on privacy'.[45] Developers working on such solutions must, however, be careful to ensure that metadata is also treated appropriately as it can reveal personal information even where personal data is not directly stored on-chain.[46] Off-chain storage solutions may further require the reintroduction of a trusted third party, which could then defeat the very motivation for relying on DLT as opposed to other forms of data storage. There are, however, attempts to design GDPR compliant chains that hold data in a private store where the blockchain merely holds proof that the data is valid.[47] It is further worth noting that where off-chain data is also distributed, enforcing the GDPR in relation to that data also become more burdensome.[48]

Eberhardt and Tai designed a series of off-chain storage solutions that do not require the reintroduction of a trusted third party. These include challenge response patters; off-chain signature patterns; delegated computing patterns; low contract footprint patterns; and content addressable storage patterns.[49] The latter is particularly relevant for our purposes. Here, data is stored off-chain in a content-addressable storage system rather than on the blockchain. For example, a smart contract would merely contain the hash to said data rather than the data itself.[50] This pattern allows the 'trustless outsourcing of data to an off-chain storage system since a modification in the data would immediately change its address and with that invalidate its references'.[51] The benefits of this approach are not limited to data protection but also drastically limit an application's storage costs. Developers designing such a solution must however be careful that off-chain data doesn't become unavailable as this threatens the availability of the on-chain part of the application and they must also avert data-leaks as leaked data can be immediately confirmed to be authentic by recalculating its address.[52]

While only time will reveal whether the Court of Justice of the European Union (CJEU) and the European Data Protection Supervisor agree it seems safe to assume, for the time being, that solutions storing all personal data off-chain are the most important step developers must take to ensure GDPR compliance. Next, we evaluate whether a user's public key constitutes personal data under EU law.

[40] ibid.

[41] If this is to be done such standards would require continued updating to account for evolutions in cryptography.

[42] Perfect cryptographically secure obfuscation is however mathematically impossible.

[43] Vitalik Buterin, 'Privacy on the Blockchain (*Ethereum Blog*, 15 January 2016) <https://blog.ethereum.org/2016/01/15/privacy-on-the-blockchain/> accessed 5 March 2018 (hereafter Buterin, 'Privacy on the Blockchain').

[44] We turn to this topic further below.

[45] Zyskind et al, 'Decentralizing Privacy' (n 30) 180.

[46] James Smith et al, 'Applying blockchain technology in global data infrastructure' (2016) Technical Report ODI-TR-2016-001, Open Data Institute.

[47] Such as the collaboration between LuxTrust and Cambridge Blockchain: Business Wire, 'LuxTrust and Cambridge Blockchain Announce Privacy-Protecting Identity Platform' (*Sys-Con Media*, 15 May 2017) <http://news.sys-con.com/node/4080523> accessed 5 March 2018.

[48] A number of current projects such as Swarm, Storj and Filecoin are experimenting with such options.

[49] Jacob Eberhardt and Stefan Tai, 'On or Off the Blockchain? Insights on Off-Chaining Computation and Data' 3 <http://www.ise.tu-berlin.de/fileadmin/fg308/publications/2017/2017-eberhardt-tai-offchaining-patterns.pdf> accessed 5 March 2018.

[50] ibid 10.

[51] ibid 11.

[52] ibid.

b. Public Keys

Public keys are a string of letters and numbers[53] that allows for the pseudonymous identification of a natural or legal person for transactional or communication purposes. The father of the first blockchain, Satoshi Nakamoto, himself considered that consensus mechanisms require information that limits the way in which access to the actual data can be limited. [54] Privacy, he argued, is maintained not by encrypting data but rather by 'breaking the flow of information in another place: by keeping public keys anonymous'.[55] From a GDPR perspective, the pertinent question is whether public keys are really anonymous data. Article 4(5) GDPR defines pseudonymisation as

> the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable person.[56]

A public key is data that 'can no longer be attributed to a specific data subject' unless it is matched with 'additional information' such as a name or an address. Where these two sets of information are combined, identification is plausible, explaining why public keys cannot qualify as anonymous data. We have already seen that for data to qualify as being anonymous identification must be irreversibly prevented.[57] Practice reveals that this cannot be said to be the case in relation to public keys. DLTs' short history testifies that despite asymmetric encryption identification remains possible. Connecting public keys with additional information permitting identification has been facilitated through users' voluntary release of such information, such as where they disclose their public key to receive funds; through illicit means, or where additional information is gathered in accordance with regulatory requirements, such as where cryptoasset exchanges perform KYC and AML duties.[58] On the Bitcoin blockchain, encrypted data has been proven capable of revealing a user and transaction nexus that allows for transactions to be traced back to the users.[59] Law enforcement agencies have moreover long developed forensic chain analysis techniques to identify suspected criminals on the basis of their public keys, and a range of professional service providers performing related services have emerged.[60] Academic research has moreover shown that public keys can be traced back to IP addresses, aiding identification.[61] What is more, where a user transmits a transaction to the network, they usually connect directly to the network and reveal their IP address. The GDPR leaves no doubt that personal data that has 'undergone pseudonymisation, which could be attributed to a natural person by the use of additional information' qualifies as personal data.[62] To determine whether a person can be identified on the basis of pseudonymous data account has to be taken of 'all the means reasonably likely to be used'.[63] Considering that public keys are in fact being used to identify individuals, they should be presumed to be a means 'reasonably likely to be used'.[64]

The CJEU's adjudicative practice reinforces our conclusion that public keys qualify as personal data. In *Patrick Breyer v Germany* it classified dynamic IP addresses as personal data.[65] The Court ruled that IP addresses assigned to a computing device when connected to a network may constitute personal data even if a third party (such as an internet service provider) holds the data relevant to identify an indi-

[53] Keys are technically always numbers, derived from large primes, that are however encoded alphanumerically to save space.

[54] Some might object to designating Nakamoto as male. Given that the person(s) behind the pseudonym have chosen a Japanese masculine given name for themselves, I respect that choice.

[55] Nakamoto, 'Bitcoin' (n 5).

[56] art 4(5) GDPR.

[57] Emphasis added.

[58] Kelly Philipps Erb, 'IRS Tries Again To Make Coinbase Turn Over Customer Account Data' *Forbes* (20 March 2017) <https://www.forbes.com/sites/kellyphillipserb/2017/03/20/irs-tries-again-to-make-coinbase-turn-over-customer-account-data/#1841d9e5175e> accessed 5 March 2018.

[59] Fergal Reid and Martin Harrigan, 'An Analysis of Anonymity in the Bitcoin System' (2012) <https://arxiv.org/abs/1107.4524> accessed 5 March 2018.

[60] Such as the appropriately named Chainalysis: <https://www.chainalysis.com/>.

[61] Biryukov et al, 'Denanonymisation of Clients in Bitcoin P2P Network' (2014) <https://arxiv.org/abs/1405.7418> accessed 5 March 2018.

[62] recital 26 GDPR.

[63] ibid.

[64] ibid (requiring that relevant factors are 'all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments').

[65] Case C-582/14 *Patrick Breyer* [2016] EU:C:2016:779.

vidual. This situation is in many ways analogous to the information exchanges or other service providers that are legally obliged to collect data under KYC and AML requirements.

We conclude that public keys are pseudononymous data caught by the EU data protection regime. Unlike transactional data, public keys cannot however be moved off-chain as they are quintessential components of the technology and form part of a transaction's 'metadata' required for its validation. GDPR-compliant solutions are accordingly more difficult to identify.

Some have suggested the use of a stealth address, which uses a one-time transaction that relies on hashed one-time keys. The cryptocurrency Monero for example hides the recipient of the transaction by generating a new dedicated address and a 'secret key'.[66] The use of one-time accounts for transactions foresees that every transaction must completely empty one or more accounts and create one or more new accounts. [67] This so-called 'merge avoidance'[68] can be deployed on the Bitcoin blockchain but some consider that even where this is done that system 'has proven to be highly porous and heuristic, with nothing even close to approaching high guarantees' of privacy protection.[69] The Bitcoin White Paper itself recommends that 'a new key pair should be used for each transaction to keep them from being linked to a common owner', while conceding that this is merely a security rather than anonymisation technique as

> [s]ome linking is still unavoidable with multi-input transactions, which necessarily reveal that their inputs were owned by the same owner. The risk is that if the owner of a key is revealed, linking could reveal all other transactions that belonged to the same owner.[70]

Cryptographic research has moreover developed 'zero-knowledge proofs' that provide a binary true/false answer without providing access to the underlying data.[71] The Zcash cryptocurrency relies on the process to ensure that even though transactions are published on a public blockchain its details (including the amount as well as its source and destination) remain hidden.[72] The ledger merely reveals whether a transaction has occurred, not which public key was used or what value (if any) was transferred. [73] Other options that are currently being deployed involve state channels for two-party smart contracts that only share information with outside parties in the event of a dispute.[74] Ring signatures on the other hand hide transactions within other transactions by tying a single transaction to multiple private keys even though only one of the initiated the transaction.[75] The signature proves that 'the signer has a private key corresponding to one of a specific set of public keys, without revealing which one'.[76] Whether any of the above solutions can be considered to anonymise public keys remains to be seen.

Another possible solution consists in adding 'noise' to the data.[77] Here, several transactions are grouped together so that from the outside it is impossible to discern the identity of the respective senders and recipients of a transaction. Algorithms similar to this model have already been defined for the Bitcoin[78] and Ethereum blockchains[79]. What is promising about this privacy technique is that the Article 29 Working Party has already recognized that, provided that the necessary safeguards are complied with, the addition of noise may be an acceptable anonymisation technique.[80] For this to be the case, it should be combined with additional privacy tech-

[66] The cryptocurrency Monero uses stealth addresses to ensure privacy. See further Monero, 'Stealth Address' https://getmonero.org/resources/moneropedia/stealthaddress.html accessed 5 March 2018.

[67] Buterin, 'Privacy on the Blockchain' (n 43).

[68] Mike Hearn, 'Merge Avoidance? (*Medium*, 11 December 2013) <https://medium.com/@octskyward/merge-avoidance-7f95a386692f> accessed 5 March 2018.

[69] See further Buterin, 'Privacy on the Blockchain' (n 43).

[70] Nakamoto, 'Bitcoin' (n 5).

[71] Zcash, 'What are zk-SNARKs?' <https://z.cash/technology/zksnarks.html> accessed 5 March 2018.

[72] This solution is currently being relied on by Zcash. See ibid

[73] ibid.

[74] Buterin, 'Privacy on the Blockchain' (n 43).

[75] See further, Monero, 'Ring Signature' <https://getmonero.org/resources/moneropedia/ringsignatures.html> accessed 5 March 2018.

[76] ibid.

[77] This has been explored by the MIT ENIGMA project and uses modified distributed hashables to store secret-shared data in combination with an external block chain for identity and access control.

[78] See further, Pablo Martin and Amir Taaki, 'Anonymous Bitcoin Transactions' <https://sx.dyne.org/anontx/> accessed 5 March 2018.

[79] Vlad Gluhovsky and Gavin Wood, 'The Witness Algorithm: Privacy Protection in a Fully Transparent System' (*GitHubGist*, 2015) <https://gist.github.com/gavofyork/dee1f3b727f691b381dc> accessed 5 March 2018.

[80] Article 29 Working Party, 'Anonymisation Techniques' (n 38) 12-13 (discussing the technique in general, not specifically with respect to blockchains).

niques 'such as the removal of obvious attributes and quasi-identifiers'.[81]

It is, at this stage, difficult to predict whether any of these techniques will be considered capable of anonymising public keys for GDPR purposes. It is true that for data to be considered as anonymous under the GDPR, it must not be perfectly impossible to link it to a natural person, as there is always a residual risk of identification.[82] The identified options require further observation and study to determine whether they can be considered suitable anonymisation techniques. We conclude that public keys as well as the transactional data stored on blockchains will often qualify as personal data. Where blockchain use cases are caught by the GDPR, its various substantive rights come to apply. The subsequent section investigates how these rights can be deployed on DLTs.

V. Applying the GDPR to Blockchains

We have already observed that transactional data and public keys generally constitute personal data for the purposes of the EU data protection framework. To pinpoint the precise legal consequences flowing from this state of affairs we must start by determining to whom the GDPR's obligations are addressed. We first evaluate who qualifies as the data controller on a decentralised ledger given that this entity must enforce its substantive rights and then consider the territorial scope of the corresponding obligations.

1. The Data Controller(s)

The GDPR defines a data controller as any natural or legal person that 'determines the purposes and means of the processing of personal data'.[83] The use of the singular indicates that in centralised data silos there is often only one entity that qualifies as a data controller. It is to them that the GDPR is addressed. When it comes to private blockchains, it might still be possible to identify a central intermediary that can qualify as *the* data controller such as the systems operator that will be the addressee of the data subject's claims.[84] For other DLTs, there is no central point of control as the network is operated by all nodes in a decentralised fashion. Permissionless blockchains are distributed and decentralised peer-to-peer networks that everyone can participate in to interact with unknown or untrusted counterparties. In such a setting, either no node qualifies as the data controller in the absence of independent determination of the means and purposes of processing, or, more likely, *every* node qualifies as a data controller. Nodes are indeed not subject to external instructions, autonomously decide whether to join the chain, and pursue their own objectives. As a consequence, it appears that the Regulation's legal obligations would rest on each node, meaning that data subjects can invoke claims vis-à-vis each node independently.

Nodes do not, in principle, qualify as 'joint controllers' under Article 26(1) GDPR as they do not 'jointly determine the purposes and means of processing'. This requires a clear and transparent allocation of responsibilities.[85] Nodes are free to determine whether to join the unpermissioned ledger and in what function (i.e. as a full or lightweight node). Nodes do not commonly determine applicable rules in the sense of Article 26 GDPR; the system is rather shaped by the nodes' individual behaviour. While a blockchain is fuelled by the interplay of various nodes they don't determine the modalities of data processing of other nodes.

Determining that each node is a data controller raises considerable complications. The exact number, location and identity of nodes on a chain cannot be established without difficulty. Depending on the perspective adopted, nodes are either passive agents subject to the directions of software designed by developers or active participants in blockchain governance.[86] What is more, nodes (i) only see the encrypted or hashed version of the data; and (ii) are unable to make any changes thereto. Nodes are thus decentralised entities that cannot respond to the tasks the GDPR requires of centralised agents.

The enforcement of obligations resting on nodes is thus burdened by significant difficulty. For the Bitcoin blockchain, there are currently approximately 11,000 nodes around the planet, of which about 1800

[81] ibid 12.

[82] Article 29 Working Party, 'Anonymisation Techniques' (n 38) 7.

[83] art 4(7) GDPR.

[84] This can be a single firm, or a joint venture in the case of consortia.

[85] recital 79 GDPR.

[86] On the legal implications of blockchain governance, also from the GDPR perspective, see Michèle Finck, *Blockchain Regulation and Governance in Europe* (Cambridge University Press 2018).

are in Germany and 800 in France.[87] The Ethereum blockchain currently counts around 19,000 nodes.[88] If one were to address each of these nodes, some of which may not be found[89] in a single jurisdiction this would create two sets of problems. First, a large amount of nodes would need to be contacted and compelled to comply, as opposed to a single controller in a data silo scenario. Second, this may lead to forcing all nodes to stop running the blockchain software where GDPR rights cannot be achieved through alternative means. This would result in a situation where an entire blockchain would be taken down in one jurisdiction for non-compliance with a single data subject's rights, which may be considered disproportionate. It is moreover unclear how fines will be calculated where a data controller on an unpermissioned blockchain has failed to comply with data protection requirements given that Article 83 GDPR calculates them on the basis of annual worldwide turnover.[90] Besides the determination problem, further questions arise as to how ordinary nodes could ever pay the hefty fines associated with the GDPR.

It is also worth remembering that through blockchains, data subjects can gain control over their own data through the private key, which triggers the question of whether the data subject herself can be considered a controller. Indeed, where an individual hashes personal information concerning herself to the blockchain, she might be both the data subject and data controller. The 'means of processing' are determined by the software run by miners and nodes as well as the hardware they use. The purposes of a data subject's reliance on a blockchain will vary and we may thus also consider the data subject to, at least in some instances, be able to qualify as a data controller is adding personal data to a blockchain. On private blockchains, nodes are moreover more likely to be qualified as data processors rather than controllers.[91] The role of data processors on blockchains cannot be addressed in detail either due to concerns of space but it is also worth nohing that blockchain data is further being used by intermediaries that process and analyse such data, which could also be considered to be data processors.[92] Ultimately, a given distributed ledger's governance arrangements need to be considered to determine why the respective controllers and processes of data are.

Next we turn to examine the Regulation's territorial scope to specify which nodes will be controllers under EU law.

2. The GDPR's Territorial Scope

Unpermissioned blockchains usually run on nodes located in various jurisdictions across the globe, leaving creators with no control over the geographic spread of the network. This makes DLTs inherently transnational in nature, triggering a range of jurisdictional issues. The GDPR applies 'to the processing of personal data in the context of the activities of an establishment of a controller or processor in the European Union, regardless of whether the processing takes place in the Union or not'.[93] This establishment clause is designed to avoid that firms escape their obligations by simply outsourcing data processing out of the Union. Pursuant to its Article 3(2), the GDPR also applies where the controller or processor are not established in the Union but where processing activities relate to either the offering of goods or services (paid or unpaid) to a data subject based in the EU[94] or where they monitor behaviour that takes place in the Union.[95] Where a controller not established in the EU processes personal data in a place where Member State law applies by virtue of public international law, the GDPR also applies.[96] The GDPR's broad territorial scope accordingly likely entails that its obligations bind many blockchain-based applications with only an indirect link to the EU.

A further jurisdictional question relates to the application of European data protection requirements to the transfer of data to third countries.[97] On permissionless ledgers we can presume that there is al-

[87] See further, Bitnodes, 'Global Bitcoin Nodes Distribution' <https://bitnodes.earn.com/> accessed 5 March 2018.

[88] Ethernodes, 'Network number 1' <https://www.ethernodes.org/network/1> accessed 5 March 2018

[89] ibid. Through a 'getaddr' message, nodes are asked for information about known active peers.

[90] Fines for breaches of data protection requirements can be as high as €20 million or 4% of global turnover, whichever is higher.

[91] art 4(8) GDPR defines a processor as 'a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller'.

[92] See, by way of example, blockchain.info.

[93] art 3(1) GDPR.

[94] art 3(2)(a) GDPR.

[95] art 3(2)(b) GDPR.

[96] art 3(3) GDPR.

[97] arts 44-50 GDPR.

ways an element of cross-border data processing. The GDPR provides that whenever there is a 'transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country or to an international organization' shall only occur subject to a number of conditions.[98] The data stored in blocks is hashed to the chain by a miner that can be based anywhere. The ledger is subsequently updated on each node to reflect the addition of the new block. The conditions allowing such cross-border processing include the possibility for the Commission to declare a 'third country, a territory or one or more specified sectors within a third country' or an international organization to ensure an adequate level of protection[99], where the controller or processor themselves provide appropriate safeguards and where 'enforceable data subject rights and effective legal remedies for data subjects are available'.[100] Competent supervisory authorities may moreover approve binding corporate rules governing data protection.[101] In theory, the chain's protocol could be designed to account for these concerns, yet, as seen below, the substantive requirements of data protection cannot easily be reconciled with DLT. A more realistic solution is enshrined in Article 49(1)(a) GDPR that foresees the possibility of a data subject providing explicit consent for such a transfer, subject to being informed about possible risks. This could be easily implemented on a private blockchain where access is controlled and can be subjected to terms and conditions but it is not obvious how such consent could be acquired in respect of a permissionless chain.

In attempting to determine the GDPR's personal, material and jurisdictional scope, we have observed that the EU's data protection regime, fashioned for the centralised collection, storage and processing of data, cannot be easily transposed to decentralised digital ledgers. An analysis of the application of the Regulation's substantive rights to distributed ledgers further validates this conclusion.

3. Enforcing Substantive Data Protection Rights on Blockchains

The GDPR creates a number of rights for data subjects in respect of their personal data. After having established that data stored on a distributed ledger as well as public keys in fact constitute personal data, this section evaluates whether data subjects can invoke their rights vis-à-vis data controllers that operate in a decentralised data environment. Numerous frictions can be identified regarding data subjects' rights and the ability of nodes to respond to them. While from a legal perspective a data subject can invoke her rights vis-à-vis every single node, it is far from obvious how, from a technical perspective, nodes could implement related requests to correct, erase or restrict data. Yet, as blockchain technology and literacy develop, technical solutions may provide relief. We limit our analysis to substantive rights arising under the GDPR for reasons of space. This does not mean that the Regulation's procedural obligations are any less problematic when applied to DLT. How a data subject can consent to the processing of her personal data on a blockchain indeed remains an, as of yet, unresolved question.[102] In examining the application of various GDPR substantive rights to DLTs we must always distinguish the two categories of personal data: transactional data as well as public keys.

a. Data Minimisation

The spirit of data minimisation is profoundly at odds with data storage on a DLT. The GDPR mandates that personal data be 'collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes'.[103] Data once added to a blockchain will perpetually remain part of the chain, given that it is an append-only database that continuously expands.[104] Distributed ledgers are by definition ever-growing creatures, which augment and accumulate further data with each additional block. What is more, integral copies of the chain are stored on each full node, quite the opposite of the data minimisation spirit. Once data has been added to the chain, it can in principle no longer be amended or deleted, which makes it diffi-

[98] art 44 GDPR.

[99] art 45(1) GDPR.

[100] art 46(1) GDPR.

[101] art 47 GDPR.

[102] art 4(11) GDPR.

[103] art 5(1)(b) GDPR.

[104] Blockchains can however perish if nodes stop running them, which creates a whole range of different legal questions.

cult not to say impossible to implement the minimisation principle and storage limitation requirements. It is worth recalling that the conflict between data minimisation requirements and novel forms of data processing are by no means novel and limited to the DLT context. Rather, they have also been stressed in respect of big data.[105]

A second look however reveals that technical solutions to these difficulties might be on the horizon. Transactional data that is stored off-chain can be modified and minimised in line with these legal requirements without touching the distributed ledger itself. The situation is however more difficult in relation to the pseudonymous public keys that cannot be retroactively removed from the ledger. A similar state of affairs exists in relation to the GDPR's right to amendment.

b. The Right to Amendment

The GDPR requires that personal data be accurate and up to date.[106] Where this is not the case, 'every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay'.[107] Data subjects' right under Article 16 GDPR includes the right to obtain rectification from the controller without undue delay. This would mean that the data subject could address any or all nodes with a request rectify personal data subject to the provided conditions. Two practical impasses arise in this context. First, a data subject cannot possibly identify any or all of a blockchain's full nodes.[108] Second, even if the data subject succeeds in addressing a claim under Article 16 GDPR, nodes are simply unable to change any of the encrypted data stored in a block. Blockchains are branded as 'immutable' ledgers precisely because information stored on them can no longer be changed except in very exceptional circumstances.[109]

While it seems that, in principle, the right to modification cannot be implemented on blockchains, the provision explicitly provides that the principle of amendment must be applied with regard to the specific technology at stake. The 'purposes of the processing' must be accounted for and data can be rectified 'by means of providing a supplementary statement'.[110] This leaves us to wonder whether the addition of new data to the chain of blocks, which rectifies data previously added (without however deleting the original entry) could be considered to comply with the requirements of Article 16 GDPR. This solution could be easily applied to an append-only ledger, yet does not lead to the modification of the problematic data itself. A more suitable solution would be to store transactional data off-chain, so that it can be modified in line with data protection requirements without the need to touch the blockchain itself. Off-chain storage can again facilitate GDPR compliance in relation to transactional data but not public keys.

Article 19 GDPR moreover requires that the controller communicate any rectification or erasure of personal data to 'each recipient to whom the personal data have been disclosed'. This, can however be presumed to not apply to nodes as the same provision clarifies that controllers are dispensed from said obligation where 'this provides impossible or involves disproportionate effort'. The application of the GDPR's right to access to a DLT is burdened by similar complications.

c. The Right to Access

In accordance with Article 15 GDPR, a data subject has the right to obtain confirmation from the controller whether or not her personal data is being processed.[111] Where this is the case, she can request additional information including but not limited to the purposes of such processing, the categories of personal data concerned, the recipients to which the data will be disclosed, the duration of storage and the existence of automated decision-making, including profiling.[112] Under Article 15(2) GDPR, data subjects are moreover entitled to be informed about safeguards that apply where data is transferred to third countries – a pertinent question in respect of blockchains given that a node validating a block in the EU will thereafter share that information with all

[105] Tal Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 Seton Hall Law Review 995.

[106] art 5(1)(d) GDPR.

[107] ibid.

[108] Reasons include that nodes may be online part time, may have closed ports, or frequently change IP addresses.

[109] The Ethereum code was for instance changed to reverse an objectionable transaction in 2016.

[110] art 16 GDPR.

[111] art 15(1) GDPR.

[112] ibid.

nodes of the blockchain, irrespective of their geographical location. Similarly to what we have already seen, Article 15 GDPR raises important questions in relation to its application to DLT given that controllers do not know which data is stored on the blockchain as they often only handle the encrypted or hashed version thereof. Even if a data subject were successful in contacting a node, the latter would be incapable of verifying whether a data subject's personal data is being processed. The data subject could of course join an unpermissioned network and obtain a copy of all data, including her own but it is questionable whether this would be regarded as a satisfactory solution in the eyes of the GDPR. As a corollary of the right to access Article 15(3) GDPR moreover entitles data subjects to obtain a copy of their personal data undergoing processing from controllers, which would be equally impossible where its has been cryptographically pseudomyised.[113] Again, storing personal data off-chain is to be preferred for transactional data but remains unfeasible for public keys. We now consider the GDPR's most famous provision: the right to be forgotten.

d. The Right to be Forgotten

Article 17 GDPR mandates that the data subject shall have the right to obtain from the controller 'the erasure of personal data concerning him or her without undue delay'.[114] Controllers are obliged to delete personal data subject to a number of conditions, such as (i) that personal data is no longer necessary for the purposes it was collected or otherwise processed; (ii) that the data subject withdraws consent on which the processing is based or where there is no other ground for processing; (iii) that the data subject objects to the processing and that there are no overriding legitimate grounds for processing; that (iv) data has been unlawfully processed; (v) that personal data has to be erased for compliance with national or supranational law to which the controller is subject; or that (vi) personal data has been collected in relation to the offer of an information society service to a child under 16 years of age.[115]

Immutability is one of blockchains' most heralded (although exaggerated) features. They are, by definition, unable to forget as they were specifically designed to be censorship-resistant.[116] A straightforward application of the right to be forgotten to DLTs can be excluded. We again distinguish between transactional data and public keys. With regard to transactional data, a number of possible solutions can be envisaged. Where personal data is recorded in a referenced encrypted and modifiable database as opposed to the blockchain itself, it can be deleted in line with data protection requirements without the need to touch the blockchain.

With regard to public keys, compliance is again more burdensome. First, it must be recalled that the right to be forgotten is not an absolute right. Article 17(2) GDPR rather provides that when faced with a request for erasure, the data controller shall take '*account of available technology* and the cost of implementation'[117] and then take 'reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of those personal data'.[118] Here, the question arises as to whether the reference to 'available technology' could lead to an interpretation of the GDPR that dispenses from outright erasure in light of blockchains' technical limitations in favour of an alternative solution. Some have moreover suggested that formalised procedures of transmitting a key to the data subject or deleting the private key in a supervised setting could amount to erasure for the purposes of the GDPR.[119] Unlike outright erasure, the encrypted data would still exist on-chain but could only be accessed by the data subject (through her exclusive control of the private key) or simply no longer be accessed at all. Pruning can be used to delete obsolete transactions in older blocks that are no longer necessary for the continuation of the chain but the idea remains controversial.[120] A fur-

[113] It is in this context worth recalling that encryption cannot be reverse-engineered.

[114] art 17(1) GDPR.

[115] ibid. Additional limitations to the right to be forgotten that are not of specific interest in the context of blockchains, such as public policy reasons, can be found under art 17(3) GDPR.

[116] Nakamoto, 'Bitcoin' (n 5).

[117] Emphasis added.

[118] art 17(2) GDPR.

[119] For an overview of other techniques that can be used to employ privacy on blockchains, see Primavera De Filippi, 'The Interplay Between Decentralization and Privacy: The Case of Blockchain Technologies' (2016) 9 Journal of Peer Production 1 (hereafter De Filippi, 'The Interplay between Decentralization and Privacy').

[120] Emanuel Palm, 'Implications and Impact of Blockchain Transaction Pruning' (Master's Thesis, Luleå University of Technology 2017) <http://www.diva-portal.org/smash/get/diva2:1130492/FULLTEXT01.pdf> accessed 5 March 2018

ther option would be the use of chameleon-hashes to re-write the content of blocks on a DLT by authorised authorities under specific constraints, and with full transparency and accountability.[121] There are however a number of problems with this approach. First, if the lock key is destroyed or lost the chain reverts to being immutable. This solution would moreover reintroduce the need for a trusted third party such as special bodies or arbitrators, which some will find unacceptable given that it arguably defeats the very benefit of DLTs. Secondly, chameleon hashes can't eliminate old copes of the blockchain that will still contain the redacted information and miners also have discretion as to whether to accept the changes or not. [122]

It should be stressed that hard forks, which can be used to mutate blockchains in very exceptional cases, are not viable GDPR compliance-tools. Hard forks only make sense for the most recently mined block as all subsequent blocks are rendered invalid so that all the transactions stored in these blocks would have to be reprocessed, which would be too costly regardless of the consensus protocol that is used and take a very long time (equal to the time that has passed since the block was mined, assuming equal mining power).

Whether any of these solutions can satisfy the requirements of Article 17 GDPR remains to be seen. We note that the precise meaning of 'erasure' is not defined in the GDPR, opening the door to other interpretations than absolute deletion.[123] It is however worth noting that certain national 'implementing' laws have already directed themselves towards a softer version of the right to be forgotten.[124] The German framework accepts that data is not deleted where the specific mode of storage makes this impossible.[125] In such circumstances, an alternative solution of not deleting but merely limiting the processing of data is tolerated. How this will apply to DLT remains to be seen given that as long as a public key is on the blockchain it will always be 'processed' in the sense that it forms part of the chain of blocks to which new blocks are hashed. This is nonetheless interesting as it shows that the GDPR can be interpreted to combine its objectives with the respective technological characteristics of the instrument at issue. This further seems to, at least as a matter of principle, open the door for interpretations of the right to be forgotten that account for the ledger's immutability and the need for alternative solutions. Other Member States have not, however, foreseen that option, which risks fragmenting applicable rules, which is precisely what the GDPR sought to eliminate.[126] Next, we look towards the GDPR's principles of data protection by design and data protection by default.

e. Data Protection by Design and Data Protection by Default

Data protection by design and data protection by default are two overarching guiding principles of the GDPR. Whilst they are not individual rights as such we nonetheless briefly examine these principles as they confirm the tension between blockchains' promises and perils for data protection. Under Article 25(1) GDPR

> the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimization, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and to protect the rights of data subjects.[127]

The above obligations are addressed to controllers which must 'implement' such mechanisms defined by software developers.[128] Systems architects must from the beginning account for the GDPR's objectives, which should include

> minimizing the processing of personal data, pseudonymizing data as soon as possible, transparen-

[121] Giuseppe Ateniese et al, 'Redactable Blockchain – or – Rewriting History in Bitcoin and Friends' (2017) 2 <http://ieeexplore.ieee.org/document/7961975/> accessed 5 March 2018.

[122] ibid 3.

[123] Such as removal from the search index.

[124] While the GDPR is a Regulation and does thus not require implementation under art 288 TFEU this is nonetheless possible through the existence of flexibility clauses.

[125] art 35 of the Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680.

[126] See, by way of example, art 16 of the Luxembourg implementing legislation.

[127] art 25(1) GDPR.

[128] Whereas in a centralised setting the controller could determine and implement the principles.

> cy with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features.[129]

While data minimisation will always be challenging on DLTs, Article 25(1) GDPR underlines that encryption can be a desirable feature, which may be a reason for regulators and courts to look favourably at the technology. This is an important point, which underlines that technology can be used to achieve legal objectives. The minimising of transactional data can be achieved by moving it, as far as possible, off-chain. The remaining question is whether the pseudonymisation of public keys can be fashioned so as to be compliant with the GDPR. The Regulation considers that the pseudonymisation of personal data 'can reduce the risks to the data subject concerned and help controllers and processors to meet their data-protection obligations'.[130] Data protection by design and default can be achieved in

> minimizing the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor data processing, enabling the controller to create and improve security features.[131]

Article 32 GDPR obliges data controllers to adopt appropriate technical and organisational measures to ensure a level of security that is appropriate to the risk. Article 25(2) GDPR however also requires the controller to implement 'appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed'.[132] This obligation applies to the amount of personal data that is collected, the extent of its processing as well as the period of storage and accessibility'.[133] Given that each full node holds a complete copy of each blockchain and that a new block is added to the complete preceding chain, this provision cannot be complied with in respect of public keys. The only way to ensure compliance in this respect would be to recognize specific key-handling techniques such as particularly strong encryption formulas or zero-knowledge proof as GDPR compliant.

The preceding analysis has revealed an undisputable lack of legal certainty when it comes to the application of the EU's data protection framework to blockchains and other forms of distributed ledger technology.[134] Ultimately, the application of the GDPR to a specific blockchain or blockchain use-case will come to be determined by the specific governance arrangements in practice. Only a close examination of governance arrangements on a case-by-case basis will allow for a determination of the respective data controller. For the time being, the safest advice for blockchain developers is that transactional data should never be stored on a blockchain. Regarding public keys, the necessary risk-management solutions must be adopted and detailed Data Protection Impact Assessments must be carried out.[135] It is obvious that the GDPR was designed for centralised models of data collection, storage and processing that cannot readily be transposed to decentralised and distributed databases. Only time will reveal how regulators and judges will approach the tension between the GDPR and DLT. In order to make sense of this tension we must consider it from a meta-perspective and evaluate the two conflicting normative objectives of EU law at play; fundamental rights protection on the one hand and the promotion of innovation on the other.

VI. Reconciling the Protection of Fundamental Rights and the Promotion of Innovation

Blockchains, in particular those of a public and permissionless character, and the EU's data protection framework stand in tension. Whereas the GDPR was fashioned for an age of centralised data silos, blockchains promise a future of decentralised data management. This highlights that, even before the new supranational data protection framework enters into force, it is already partly outdated in respect of its application to distributed ledgers for it simply can-

[129] recital 78 GDPR.

[130] recital 28 GDPR.

[131] recital 78 GDPR. On the desirability of pseudonymisation, see also arts 6(4)(e), 31(1)(a) and 89(1) GDPR.

[132] art 25(2) GDPR.

[133] ibid.

[134] Additional questions arise regarding the compatibility of the GDPR and blockchains, such as the application of art 22(1) GDPR to smart contracts.

[135] art 35 GDPR.

not account for the technology's characterising features. Similar concerns have emerged in relation to big data.[136] While law has always lagged behind technological change, this divide becomes more acute as the pace of innovation speeds up in the digital age. Specifically in respect of the GDPR we have observed that pivotal features thereof such as the rights to amendment and erasure cannot be easily applied to new technologies for data storage and processing. We have however also seen that blockchains, if adequately designed, and the GDPR can share a common objective: giving a data subject more control over her data. This is of course only the case where blockchains are specifically fashioned to achieve that objective. De Filippi has warned that if this is not the case, these decentralised structures 'might turn out to be much more vulnerable to governmental or corporate surveillance than their centralized counterparts'.[137] The challenge thus lies in bringing law and technology together to ensure that law does not unnecessarily hinder technological progress but also that technologically develops in a normatively desirable fashion. In this specific context, the challenge consists in applying the EU data protection framework in a manner that doesn't asphyxiate blockchains' innovative potential, yet at the same time ensures that data protection is guaranteed.

DLTs that store personal data are caught by the GDPR, which causes concern for many operators. The foolproof solution would be to simply refrain from storing such data on chains, which might be feasible for data itself but not the keys and signatures without which these ledgers cannot function. Considering that both fundamental rights protection and the promotion of innovation are supranational objectives, a purposive interpretation of the GDPR should be adopted whenever possible. Blockchains indeed bear the promise of realizing the GDPR's objectives through technological means and such techno-legal interoperability should not be stifled at inception. While we are used to seeing technology and privacy as antagonists they do not have to be - technology can help achieve the GDPR objectives. A purposive approach would further reflect the need for legislation to be technology- and business-model neutral as a textual interpretation risks disadvantaging blockchains over other technologies.[138] The European Commission has stressed that the GDPR is a technologically neutral legislation that will enable 'innovation to continue to thrive'.[139] Indeed, even the fiercest data protection proponents have argued that although the GDPR will change 'nothing less than the world as we know it', it also underlines that 'it is possible to achieve common action through a democratic process on the basis of high standards for citizens' and consumers' rights as well as a competitive and innovative single market'.[140]
Blockchains can provide an alternative means of achieving the Regulation's objective of allowing data subjects to control their own personal data and bear much promise for the Single Digital Market project, which still remains to be successfully completed.

The protection of natural persons in relation to the processing of personal data constitutes a fundamental right under Article 8(1) of the Charter of Fundamental Rights and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU). Its importance can thus not be overestimated. Innovation is, however, also a normative objective of the EU and its legal order. As per Article 173 TFEU, the EU and the Member States must work towards the EU's competitiveness, which includes the fostering of innovation and technological development. The 'Innovation Union', part of the Europe 2020 initiative, was designed to make the EU an 'innovation-friendly environment that makes it easier for great ideas to be turned into products and services that will bring our economy growth and jobs'.[141] In competition law, agreements caught by the prohibition of illicit collusion in Article 101(1) TFEU may further be allowed to stand where, as per Article 101(3) TFEU they contribute to 'promoting technical or economic progress'. In his 2017 State of the Union speech, Commission President Juncker announced that the EU's new Industrial Policy Strategy is designed to make European industries 'the number one in innova-

[136] Tal Zarsky, 'Incompatible: The GDPR in an Age of Big Data' (2017) 47 Seton Hall Law Review 995.

[137] De Filippi, 'The Interplay between Decentralization and Privacy' (n 120) 1.

[138] See more generally, Mireille Hildebrandt and Laura Tielemans, 'Data Protection by Design and Technology Neutral Law' (2013) 19 Computer & Security Review 509.

[139] European Commission, 'Questions and Answers – Data Protection Reform' (Press Release, 21 December 2015).

[140] Jan Philipp Albrecht, 'How the GDPR will change the World' (2016) 2(3) EDPL 287, 289.

[141] See further, Commission, 'Innovation Union' <https://ec.europa.eu/research/innovation-union/index_en.cfm> accessed 5 March 2018.

tion'.[142] While 'innovation' certainly is a term easy to use yet hard to define[143], there can be no doubt that the EU currently considers it as a normatively desirable objective, just as it is hard to deny that DLTs are innovative technologies and despite numerous technological hiccups blockchain promise to emerge as 'an important technological and economic phenomenon'.[144]

This is not to say that the promotion of innovation should outweigh fundamental rights protection. Rather than seeing these two objectives as antagonists, future blockchain development might reveal them be allies. If fashioned appropriately, DLT does not undermine the data protection objective, but rather changes the means of its realisation. The European Data Protection Supervisor recognizes that even though 'advanced technologies increase the risk to privacy and data protection, they may also integrate technological solutions for better transparency and control for the persons whose data is processed'.[145] As a blockchain industry develops in the EU, regulators must not shy away from using the variegated incentivising mechanisms available to them to ensure that the technology evolves in a normatively desirable manner. The relationship between law and innovation is multifaceted and stringent data protection requirements in the EU can work as an incentive to refine privacy-protecting blockchain solutions and develop a corresponding industry in the EU. Provided that innovators are given the necessary flexibility, the GDPR could spur innovation to evolve in a direction compliant with these important public policy objectives. For this to materialise, discussion and mutual learning between the industry and policy-makers cannot be avoided.[146]

It is in this context useful to remember that data protection operates in a wider context. The GDPR furthers two objectives: that of data protection but also that of the free movement of data.[147] Data protection is to be 'designed to serve mankind'.[148] If we accept that innovation has also served mankind[149], the conclusion that innovation is a consideration to be accounted for in interpreting the GDPR is reinforced. Data protection is not an absolute right but must rather 'be considered in relation to its function in society'.[150] The GDPR's pivotal principles of data protection by design and default even require technological innovation in mandating that new products and services account for data protection considerations.[151] It is in this context encouraging that in its 2017 Annual Report, the European Data Protection Supervisor indicated that

> it is essential that data protection experts begin to examine the concepts behind blockchain technology and how it is implemented in order to better understand how data protection principles can be applied to it. An integral part of this process should be the development of a privacy-friendly blockchain technology, based on the principles of privacy by design.[152]

New technology doe not just change how we apply existing regulations to new facts but may also profoundly unsettle the foundations upon which existing regulation rests. In the eyes of the GDPR, the onus of personal data stewardship rests on singular data controllers and processors that handle singular data silos. The technological innovation that brought us blockchains may however turn individuals into data sovereigns that can themselves, copy, change, share, move their data. It is now, in the still relatively early stages of blockchain technology, that appropriate data protection safeguards must be implemented and strongly encouraged by regulators. While some degree of transparency on a DLT is unavoidable to allow the network to reach decentralised consensus, transparency is only unavoidable at the ledger's most basic layer that applies the consensus algorithm. Just

[142] This speech is available online: <http://europa.eu/rapid/press-release_SPEECH-17-3165_en.htm> accessed 5 March 2018.

[143] John Kao has defined innovation as 'the ability of individuals, companies and entire nations to continuously create their desired future'. See John Kao, *Innovation Nation* (Free Press 2007).

[144] Juho Lindman et al, 'Executive Summary' in Roman Beck et al, 'Opportunities and Risks of Blockchain Technologies' (2017) 7 Dagstuhl Reports 99, 102.

[145] See further, European Data Protection Supervisor, 'Technology Monitoring' <https://edps.europa.eu/data-protection/our-work/technology-monitoring_en> accessed 5 March 2018.

[146] On this, see further Michèle Finck, 'Blockchain Regulation' (forthcoming, German Law Journal 2018).

[147] recital 13 GDPR ('The proper functioning of the internal market requires that the free movement of personal data within the Union is not restricted or prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data').

[148] recital 4 GDPR.

[149] The invention of sanitation should be an uncontroversial case in point.

[150] recital 4 GDPR.

[151] recital 78 GDPR.

[152] This Annual Report is available online: <https://edps.europa.eu/sites/edp/files/publication/17-04-27_annual_report_2016_en_1.pdf> accessed 5 March 2018.

as with the TCP/IP layer for the Internet, additional layers of encryption and obfuscation can be build on top to conceal personal data.[153] Only time will reveal whether blockchains' potential for data sovereignty is confirmed and whether the interpretation of the EU's data protection framework allows such models to develop. In this context, those called upon to interpret and apply the GDPR should of course not blindly trust DLTs to be by definition furthering of data sovereignty. It is rather also regulators' role to make sure that these considerations are incorporated into the software from the beginning.

[153] De Filippi, 'The Interplay between Decentralization and Privacy' (n 119).