# BLOCKCHAIN AND THE GDPR

a thematic report prepared by

THE EUROPEAN UNION BLOCKCHAIN OBSERVATORY AND FORUM

EUBlockchain
Observatory and Forum

An initiative of the

European
Commission

# About this report

The European Union Blockchain Observatory & Forum has set as one of its objectives the analysis of and reporting on a wide range of important blockchain themes, driven by the priorities of the European Commission, and based on input from its Working Groups and other stakeholders. As part of this, it will publish a series of thematic reports on selected blockchain-related themes. The objective of these thematic reports is to provide a concise, easily readable overview and exploration of each theme suitable for the general public. The input of a number of different stakeholders and sources are considered for each report. For this paper these include:

- Members of the Observatory and Forum's Working Groups.
- On Blockchains and the General Data Protection Regulation by Luis-Daniel Ibáñez, Kieron O'Hara, and Elena Simperl, an academic research paper on the theme prepared by the University of Southampton, one of the Observatory's academic partners.
- Input from participants at the EU Observatory and Forum GDPR Workshop on June 8 in Brussels, with special thanks to Olivier Micol, Alexis Berolatti, Jörn Erbguth, Michèle Finck, and Elizabeth Renieris.
- Input from the Secretariat of the EU Blockchain Observatory & Forum (which includes members of the DG CONNECT of the European Commission, and members of ConsenSys).
- Blockchains and Data Protection in the European Union, by Michèle Finck.
- Blockchain et RGPD : quelles solutions pour un usage responsable en présence de données personnelles? by CNIL.
- Input from national initiatives such as the Dutch Blockchain Coalition.

## CREDITS

This report has been produced by ConsenSys AG on behalf of the European Union Blockchain Observatory and Forum.

Written by: Tom Lyons, Ludovic Courcelas, Ken Timsit
Workshop moderator: Vitus Ammann
Report design: Benjamin Calméjane
Images and icons: Unsplash, Flaticon, The Noun Project

First edition published on 16 October 2018.

## DISCLAIMER

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

# Contents

4 Executive summary

7 Introduction

10 Evolution from above: Introduction to the GDPR

Personal data, the heart of the GDPR 10

GDPR roles 11

Principles, rights and obligations 12

14 Revolution from below: Blockchain and the tools of decentralisation

The decentralized database model 14

Public blockchains and permissioned blockchains 14

Is there a GDPR-compliant blockchain? 16

17 Tensions between the GDPR and blockchain

Accountability and roles: who is the controller? 17

How should personal data be anonymised? 19

Blockchains and the GDPR's rights and obligations 24

28 Opposites attract: Resolving the tensions between blockchain and the GDPR

32 Appendix

Blockchain terminology 32

Infographic 35

# Executive summary

The General Data Protection Regulation (GDPR), which entered into force in the European Union in 2016 and into application in 2018, is the latest development in the European Union's ongoing efforts to protect the personal data of its citizens.

Designed to reach a balance between data protection and the free movement of personal data, the GDPR was written during the rise to prominence of what is considered to be one of the most disruptive new information technologies on the horizon today: blockchain.[1]

At its core a database technology that enables radical decentralisation of data storage and processing, blockchain implies an environment and operating paradigms that would seem to make it difficult to interpret some of the GDPR's rules. In this new environment, where information does not flow linearly from users to providers and back, the necessary compliance with the GDPR may provide technological challenges.

The issue of compliance of blockchain with the GDPR is, however, an important one.

Government agencies and regulators in Europe have embraced this new technology for its potential for innovation in Europe, and have stated many times that while their goal is the protection of individual rights, they are by no means looking to end blockchain.

So while there are certainly tensions between the GDPR and blockchain, we argue that there are paths for reconciliation too.

As this paper will explain, **GDPR compliance is not about the technology, it is about how the technology is used.** Just like there is no GDPR-compliant Internet, or GDPR-compliant artificial intelligence algorithm, **there is no such thing as a GDPR-compliant blockchain technology. There are only GDPR-compliant use cases and applications.**

Among other things, in this report we observe that many of the GDPR's requirements are easier and simpler to interpret and implement in private, permissioned blockchain networks than in public, permissionless networks. Yet public networks are here to stay, and represent a vital space of innovation that has the potential to create jobs and thriving companies

[1] More information about blockchain in Appendix.

## EXECUTIVE SUMMARY

in the same way that the world wide web did over the last twenty years.

The tensions between the GDPR and blockchain revolve mainly around three issues:

- **The identification and obligations of data controllers and processors.** While there are many situations where data controllers and data processors can be identified and comply with their obligations, there are also cases where it is difficult, and perhaps impossible, to identify a data controller, particularly when blockchain transactions are written by the data subjects themselves.
- **The anonymisation of personal data.** There are intense debates, and currently no consensus, on what it takes to anonymise personal data to the point where the resulting output can potentially be stored in a blockchain network. To take one example, the hashing of data cannot be considered to be an anonymisation technique in many situations, and yet there are cases where the use of hashing to generate unique digital signatures of data that is stored off-chain, is potentially conceivable on a blockchain.
- **The exercise of some data subject rights.** We note that if personal data is recorded in a blockchain network, it may be difficult to rectify or remove it. Defining what can be considered erasure in the context of blockchains is under discussion.

**To be clear, these issues have not been conclusively settled by the data protection authorities, the European Data Protection Board (EDPB) or in court.** In our view, it is important that regulators take the time to deeply understand each use case of blockchain technology, as well as the impact that various interpretations of the GDPR can have on the European ecosystem.

**Meanwhile, we propose four rule-of-thumb principles that entrepreneurs and innovators can consider:**

1. Start with the big picture: how is user value created, how is data used and do you really need blockchain?

2. Avoid storing personal data on a blockchain. Make full use of data obfuscation, encryption and aggregation techniques in order to anonymise data.

3. Collect personal data off-chain or, if the blockchain can't be avoided, on private, permissioned blockchain networks. Consider personal data carefully when connecting private blockchains with public ones.

4. Continue to innovate, and be as clear and transparent as possible with users.

## EXECUTIVE SUMMARY

Blockchain technology is new and complex to understand. Additionally, it is still immature and it should not be surprising to citizens and regulators that not every ‘t’ has been crossed.

**There are many promising research and development efforts** under way to make it easier for blockchain application developers to comply with the GDPR. Even more excitingly, we are seeing many projects exploring how blockchain could be used to support the GDPR.

By finding ways to ensure the robust protection of personal data in decentralised systems, Europe could very well replace the tensions and hurdles we have outlined in this report with a much more virtuous circle of secure information.

# Introduction

Protecting personal data has long been an important policy goal in the European Union, so much so that it is enshrined in the EU Charter of Fundamental Rights.[1]

In an increasingly digital world, in which large amounts of information are generated, collected and processed, data has gained significance beyond what the framers of that charter could have imagined.[2] In many ways data has become the life-blood of our society as well as an increasingly valuable asset. Just how valuable can be seen, among other things, by the success of business models based on users trading their personal data, whether consciously or not, for free services.

Such models financed much of the development of the modern world wide web, and are responsible for many services now deemed essential to our lives. But there is a downside. Some companies misuse individuals' personal data,[3] selling it on to third parties without their knowledge and for purposes over which they have no say. Some companies, governments and other institutions have struggled to safeguard personal data, as the almost daily revelations of hacks and data breaches attest. This exposes individuals to both nuisance risks, like unwanted advertising, and existential ones, like identity theft.

The European Union, keen to protect its citizens, has long sought to regulate the use of personal data. Yet the task is not as straightforward as it may seem. To function, democracies must balance the fundamental right of safeguarding personal data with other fundamental rights. Whether in support of free speech, the public's right to know, legitimate security concerns, or fostering economic growth, there are many reasons why the right to data protection needs to be balanced with other equally weighty concerns.

The General Data Protection Regulation (GDPR),[4] which entered into

[1] EU Charter of Fundamental Rights, Article 8 - Protection of personal data.
[2] Data, Data Everywhere, The Economist, February 2010.
[3] Companies are making money from our personal data, but at what cost?, Jathan Sadowski, The Guardian, 31 August 2016.
[4] 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).

## INTRODUCTION

force in the European Union in 2016,[5] lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data. It is also meant to take the significant technological developments of the past 25 years into account, in particular the rise of networked information spaces[6] like the world wide web. It was, however, conceived and written before the rise to prominence of what is considered to be one of the most disruptive new information technologies on the horizon today – blockchain.

Originally invented to power Bitcoin, the world's first cryptocurrency, blockchain has spawned a revolution in how we think about money, transactions, commerce, and data in a digital world. It can be argued that this technology was in part inspired by a desire to both protect individual freedom and data (e.g., information on financial transactions) and facilitate secure information and value exchange. The issue, as we will see, is that the GDPR was fashioned with the implicit assumption that data in our digital world is controlled by identifiable actors. Blockchain technology seeks to achieve radical decentralisation of data, a very different approach.

In this new environment where information does not flow linearly from users to providers and back, the necessary compliance with the GDPR may provide technological challenges. There are those who claim that the two cannot coexist,[7] and even cases of blockchain projects shutting down for fear of GDPR's fines.

The issue of compliance of blockchain with the GDPR is an important one. By specifying how personal data is to be protected, the GDPR will play a fundamental role in shaping digital markets in the Union. Considering its strong support of this nascent technology, the European Union clearly believes that blockchain technology has an equally important role in these markets too, offering new paradigms for the ways we transact and interact with each other.

Government agencies and regulators in Europe have embraced this new technology for its potential for innovation in Europe, and have stated many times that while their goal is the protection of individual rights, they are by no means looking to end blockchain.

[5] The GDPR came into application on 25 May 2018.

[6] On Blockchain and the General Data Protection Regulation.

[7] Software Engineers Discovering How GDPR Limits the Use of Blockchain, Eweek, 11 June 2018; Will GDPR Compliance Kill Blockchain?, Stanly Johnson, 4 July 2018.

## INTRODUCTION

With this goal of supporting innovation as a backdrop, we have written this paper with two main constituencies in mind. First, there are the entrepreneurs and developers who will use blockchain to devise and develop new businesses, platforms, products and services, and who will want to do so in a GDPR-compliant way. Second, there are lawyers, lawmakers and regulators who will ponder – and hopefully settle – many of the questions that stand in their way.

We have tried — as best we can — not to write a technical paper, but to examine these issues in as accessible and understandable a way as possible. With the GDPR's ambition to protect one of our most fundamental rights as individuals, and the blockchain's ambitions to reshape many fundamental social, economic and political structures, this is an issue that ultimately touches on all Europeans. It is our sincere hope that this paper can be of some service to them.

# Evolution from above: Introduction to the GDPR

As comprehensive and far-reaching as it is, the GDPR is an evolution, not a revolution. It builds on concepts and principles in place in the EU since the European Data Protection Directive (DPD) of 1995, and can trace its roots back through 1991's Right to Privacy Convention, the early national data protection laws of the 1970s and all the way to the European Convention on Human Rights of 1953.[1]

The GDPR is conceived as an update to 1995's DPD, one of the aims of which is to strengthen the protections that legislation provides individuals — referred to in the law as 'data subjects' — over their personal data.[2] It also significantly increases the potential fines for non-compliance, giving the regulation more teeth.

While it is a long and detailed document, to understand the GDPR it helps to think of it in terms of what it applies to, who it affects and what protections it aims to provide.

## PERSONAL DATA, THE HEART OF THE GDPR

The material and territorial scope[1] of the GDPR is broad: it applies to all personal data of data subjects in the European Union, with personal data defined as 'any information relating to an identified or identifiable natural person'.[2] That does leave a lot of data out: for example, information pertaining to companies or machines which do not contain personal data does not fall under the GDPR's protections. But when it comes to natural persons, the net is cast very wide.

That's because this notion of a potentially "identifiable" natural person means that the GDPR applies not just to data that seems obviously related to us — our name or phone number, say — but also any data that, perhaps when combined with other data, can potentially at some point and by some means be used to identify us.

To take a simple example, if an online retailer collects our name, home address, email address and credit card number to carry out a purchase, then that is clearly personal data.[3] If that same retailer has a physical store at which we buy something with cash but use a numbered discount coupon that we had

[1] Convention for the Protection of Human Rights and Fundamental Freedoms.

[2] Compared to the DPD, the GDPR does not add many new concepts. Among other things, however, it more clearly defines the respective obligations of data controllers and processors, strengthens language around the right to be forgotten, and creates a new right to portability of some types of data.

[1] GDPR, op. cit., Art 2(1) and 3(2) .

[2] GDPR, op. cit., Article 4 (1).

[3] Here too things are not as straightforward as they might seem. For instance, our name alone is not necessarily personal data, since other people may have the same name. A record in a database that contained our name, home address, birth date and eye and hair colour would be, because that combination is likely unique to us.

downloaded from the store's website, the record of that transaction would be personal data too, because the store could use the coupon number to connect the transaction to our email address and hence to us.[4] As we will see, this notion of data that can be used to indirectly identify us has important implications for many blockchain-based platforms.

## GDPR ROLES

To understand how the GDPR aims to protect personal data, it can also be helpful to think of the law in terms of the interplay of three main actors.

- **Data subject.** There is the data subject referred to above: the person to whom the data relates.
- **Data controller.** The data controller is defined as the 'natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data'[1]. This is the key role in the regulation. It is the data controller to whom data subjects turn to exercise their rights, and who is ultimately accountable for compliance and liable if the rules are breached.
- **Data processor.** The data processor is a body which 'processes personal data on behalf of the controller'.[2] Depending on how the data controller designs its system, there can be many data processors, or none at all. And while data processors have their own sets of obligations under the GDPR, they are ultimately working for the data controller.

The data controller, as architect and point of accountability for the data processing, is the key role. **As far as the GDPR is concerned, it must be possible to identify a data controller. As we will see, this is not always easy in the blockchain world.**

[4] This example has been modified from GDPR EU.org, which is an excellent source for understanding the GDPR. The GDPR itself, in Article 4(1), provides a non-exhaustive list, including location data or an online identifier.

[1] GDPR, op cit, Article 4(7).
[2] GDPR, op cit., Article 4 (8).

EVOLUTION FROM ABOVE: INTRODUCTION TO THE GDPR

# PRINCIPLES, RIGHTS AND OBLIGATIONS

With this constellation in mind, the GDPR then lays out the means by which personal data is to be protected. These are founded on a set of six core data processing principles:[1]

- **Lawfulness, fairness and transparency.** The GDPR states that personal data should be 'processed lawfully, fairly and in a transparent manner', meaning that the data controller must have lawful grounds to collect the information as defined in the GDPR, must be transparent about how it intends to use the data, and must ensure that it does nothing unlawful with it.
- **Purpose limitation.** Personal data shall be 'collected for specified, explicit and legitimate purposes', and only used for those purposes that have been stated. This makes it unlawful for a controller to indiscriminately repurpose data it has collected for one purpose for another, unrelated purpose. There are, however, some exceptions, like archiving purposes in the public interest or for scientific or historical research.
- **Data minimisation.** The personal data collected should be 'adequate, relevant and limited to what is necessary' for the stated purposes. This makes it unlawful to ask for data that isn't necessary, for example if an online pizza delivery company were to ask for someone's marital status in order to complete an order.
- **Accuracy.** Personal data held by a controller must be 'accurate and, where necessary, kept up to date'. Under this principle, data subjects have the right, among other things, to ask that inaccurate data held by a controller be amended, and controllers must take every reasonable step to either fix or erase such data.
- **Storage limitation.** Personal data must be 'kept in a form which permits identification of data subjects for no longer than is necessary' for the stated purposes. This makes it unlawful to keep data indefinitely and means that controllers must be capable of deleting the data when it is no longer necessary. Here too there are some exceptions for archiving in the public interest and for scientific use.
- **Integrity and confidentiality.** Personal data should be 'processed in a manner that ensures appropriate security', meaning that controllers have a responsibility to safeguard the personal data they collect, whether against hacks or against accidental loss, destruction or damage.

In addition to these principles, the GDPR says that processing can be lawful if a data subject consents to it, as long as such consent is 'freely given, specific, informed and unambiguous'.[2] Once given, the data subject also has the 'right to withdraw his or her consent at any time'.[3] However, the GDPR specifies a number of grounds on which data controllers have the right to keep and process personal data even without consent. These include where processing is necessary for the performance of a contract, to protect the vital interests of the data subject or someone else, for the performance of a task carried out in the public interest and if necessary for the legitimate

[1] GDPR, op. cit., Article 5 (1).

[2] GDPR, op cit., Article 4 (11).

[3] GDPR, op. cit., Article 7(3).

## EVOLUTION FROM ABOVE: INTRODUCTION TO THE GDPR

interest pursued by the controller.[4]

The principles mentioned above are used to derive a detailed set of rights for data subjects and obligations for data controllers. The most important of these in our context are:

- **Right of rectification.** Under the GDPR data subjects have a right to have incorrect data updated in a timely manner.[5]
- **Right to erasure** ('right to be forgotten'). Data subjects also have the right to have personal data that is no longer needed for the purpose of lawful processing be deleted.[6]
- **Right of access.** This means that data subjects have a right to enquire of a data controller if their personal data is being processed and, if it is, to receive certain details about how this is being done and where.[7]
- **Rights related to automated processing.** The GDPR also recognises certain rights of the data subject pertaining to automated processing of data. These are primarily concerned with profiling and/or subjecting individuals to legal or other significant effects solely as the result of a decision taken by a machine.[8]

For their part, data controllers have a number of obligations. The paramount one, of course, is the obligation to process personal data lawfully or face the consequences.

Other obligations are more specific. Two of these of note in our context are 'data protection by design and default'[9] and 'security of personal data'.[10] The first means that controllers should consider how to comply with the GDPR both when designing their processes and implementing their systems, while the latter means that they should do everything they reasonably can to ensure that data is secure.

There are also obligations in terms of where data processing can take place, also known as 'transfers of personal data to third countries'. The GDPR specifies that personal data can generally only be transferred to third countries if they are deemed 'adequate' — that is, if they are deemed to provide data protection that is essentially equivalent to that in the EU — or if the data controller can otherwise introduce appropriate safeguards that the data will be processed in a manner consistent with that law. In any event, transfers to third countries may only be carried out in full compliance with the GDPR.

Finally, under certain circumstances data controllers may be obliged to integrate data protection explicitly into their governance processes, for example when they are obligated to designate a data protection officer (DPO) or carry out a data protection impact assessment (DPIA).

There are other obligations as well. Controllers will want to get them right. The GDPR clearly states that the data controller is 'responsible for, and be able to demonstrate compliance with',[11] the above principles. This 'accountability' pervades the whole of the legislation and is important, as the GDPR also lays out stiff fines for its breach that can reach as high as 20 million euro or 4% of a company's worldwide annual turnover.[12]

[4] Except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject.
[5] GDPR, op. cit., Article 16.
[6] GDRP, op. cit., Article 17.
[7] GDPR, op. cit., Article 15.
[8] GDPR, op. cit., Article 22.
[9] GDPR, op. cit., Article 25.
[10] GDPR, op. cit., Article 32.
[11] GDPR, op. cit., Article 5 (2).
[12] GDPR, op. cit., Article 83 (5).

# Revolution from below: Blockchain and the tools of decentralisation

## THE DECENTRALISED DATABASE MODEL

At its core, blockchain is a decentralised database technology. It allows large numbers of actors, including strangers or even adversaries, to store synchronised copies of the same data. The data is typically organised in the form of an append-only ledger, meaning data can only be added, not taken out. The specifics, however, vary by technology, as there are several types of blockchains.

The technology has the potential to bring about immense economic benefits, making it possible for actors, financial or otherwise, to transact with each other almost in real-time without requiring several layers of intermediaries.

On the flip side, the technology brings about a new paradigm of data storage and governance, leaving many questions open as to how the GDPR applies to ecosystems where there is no single, centralised, third-party data storage platform.

## PUBLIC BLOCKCHAINS AND PERMISSIONED BLOCKCHAINS

Broadly speaking, a blockchain network consists of a group of server nodes that store synchronised copies of the same data. There are usually two types of nodes:

- **Validating nodes.** Validating nodes are allowed to add data to the ledger, according to an agreed-upon algorithm called a consensus mechanism.
- **Participating nodes.** Participating nodes store synchronised copies of the data. Depending on the specific technology, not all nodes may necessarily store all data. If a user is connected to a participating note, they can add new data to the ledger, but this data needs to be sent to the participating node first, and then submitted to a validating node.

As mentioned above, there are many different types of blockchains. The original blockchain, which was invented to power Bitcoin, is what is known as a public, permissionless blockchain. **In a public, permissionless network,** anyone is allowed to become a participating node or a validating node.

Doing so simply requires one to install the client (software which is almost always open source) and download a full copy of the blockchain, and so become a full node that can take part in the process of storing and/or adding data.[1] There is no network owner, no sign-up procedure, no registration, and no restrictions on who can do this. The software is developed and maintained by changing groups of volunteers, and it exists 'in the wild' as a tool that people can choose to use or not.

In a public, permissionless network, all nodes can see all data, as well as the addresses of the sender and the receiver. That said, anyone can obviously decide to encrypt the data before submitting it to the ledger, in the same way that they would encrypt an email or a credit card payment over the internet. They can also use a third-party redirection service to obfuscate the sender's or the receiver's address.

Since the advent of the original blockchain, other variants have arisen.

**Some networks are public and permissioned.** This means that anyone can be a participating node and see all data, but only pre-approved actors can become validating nodes and add data to the ledger.[2]

**Some networks, particularly the ones used by financial institutions, are private and permissioned.**[3] This means that validating nodes and participating nodes must be pre-approved by a governance of actors, generally in the form of a consortium of companies or government agencies. Also, in some cases, there are rules in place that define who is able to see what data.

While the analogy is far from perfect, we can think of private, permissioned blockchains as akin to private intranets, only accessible to those within the organisation. Public, permissioned blockchains are more akin to (sometimes very large) extranets. They could also be compared to America Online (AOL) in the early days of the world wide web, a service open to all but built, maintained and managed by a single entity. Public, permissionless blockchains are more akin to the open Internet, and represent a base platform of trust available for use by all for any purpose.

[1] See the Appendix for a more detailed description of how blockchains work.
[2] See the Alastria project in Spain or the Sovrin network.
[3] See, for example, WeTrade, Enerchain or TradeLens.

REVOLUTION FROM BELOW: BLOCKCHAIN AND THE TOOLS OF DECENTRALISATION

# IS THERE A GDPR-COMPLIANT BLOCKCHAIN?

As this paper will explain, **GDPR compliance is not about the technology, it is about how the technology is used.** Just like there is no GDPR-compliant Internet, or GDPR-compliant artificial intelligence algorithm, **there is no such thing as a GDPR-compliant blockchain technology. There are only GDPR-compliant use cases and applications.**

Understanding the interplay between blockchain and the GDPR should therefore take place on a case-by-case basis, by analysing where the personal data appears, how it is processed and who is responsible for that processing.

**Private, permissioned blockchain networks operated by consortiums of companies or government agencies, will find it easier to apply the letter of the GDPR** than public, permissionless networks. Such consortiums are in a position to define the roles of their participants and the information flows, and they can impose strict data processing rules by making sure that all network participants commit to a set of terms and conditions. However, they have challenges too: just because consortium members are tied by contractual terms and conditions does not for instance mean that they all have a legitimate reason to see the data of each subject.

When a blockchain application involves personal data and can be deployed on a private, permissioned blockchain, it certainly makes sense to stick to a private, permissioned network, in line with the 'privacy-by-design' requirements of the GDPR.

Having said that, **public, permissioned blockchain networks already exist, and are likely to stay with us for the foreseeable future.** They currently represent an estimated 80% of all blockchain application developers and transactions and – to borrow from our intranet/Internet analogy – have the potential to become the glue that keeps the world's private blockchain networks interoperable.

**Public, permissionless blockchains represent the greatest challenges in terms of GDPR-compliance, because of their extremely distributed nature.** For this reason, a large portion (but not all) of the analysis that follows focuses on examining the interplay between the GDPR and public, permissioned blockchains of the type introduced by Bitcoin.

# Tensions between the GDPR and blockchain

There is no contradiction in principle between the goals of the GDPR and those of blockchain technology. Most GDPR requirements can be applied to most blockchain applications.

For example, many blockchain-based applications are operated by an identified entity, or consortium of entities, and they post data on a blockchain ledger on behalf of their users. In this case, the identity that operates the application is a data controller, and it must comply with its obligations under the GDPR.

However, the **GDPR does not offer clear answers to all the questions asked by entrepreneurs and technologists** when it comes to the development of innovative applications on blockchain networks. In this section, we discuss these open questions in more detail.

## ACCOUNTABILITY AND ROLES: WHO IS THE CONTROLLER?

Accountability is a central issue in the GDPR, particularly when it comes to the responsibilities of the data controller.

In the traditional client-provider model, it is relatively easy to identify the controller. There is almost always an entity that is offering some product or service, or an agency fulfilling some function, that determines the purpose and means for processing, sets up the systems to do it, and collects and processes the data for the data subject. If several entities are jointly offering a product or service, they can be identified as joint controllers.

**Private, permissioned blockchain networks lend themselves well to the above approach.** As recommended by the French CNIL,[1] blockchain consortiums should identify the controller or joint controllers as soon as possible in the project.

**In public, permissionless blockchains**, where the idea is to replace the traditional client-provider model with one based on collective processing of data via a shared protocol, the question of how to identify a data controller is less straightforward, and the object of debate.

**To be clear, this debate has not been conclusively settled by the data protection**

[1] Blockchain et RGPD : quelles solutions pour un usage responsable en présence de données personnelles?, CNIL, September 2018.

**authorities, the EDPB or in court.** What follows is a summary of common views within the blockchain community in terms of what could be a desirable outcome of this discussion.

In many instances, it would be desirable that the **protocol developers** who create and maintain open-source blockchain technology, as is the case for example with Bitcoin, **should not be considered data controllers.** They volunteer to work on an open source project and in many cases are not directly compensated for their efforts and are in essence simply creating a useful tool, not prescribing how this tool should be used. Holding developers accountable under these circumstances would be like holding DARPA or Tim Berners-Lee accountable for everything that happens on the world wide web, or MySQL creators accountable for every use of that database technology.

In many instances, it would be desirable that the actors who run the blockchain protocol on their computers in order to act as **validating nodes or participating nodes in public, permissionless networks should not be considered data controllers either.** Here too there is much debate. On the one hand it can be argued that nodes do not determine the purpose and means of processing. They are running the protocol in the hope of winning a reward, or in order to contribute to the stability of the network, and/or as a way to access the data that is relevant to them without relying on third-party intermediaries. Others argue the opposite: that through the act of actively downloading and running the software, nodes are indeed determining the purpose and means of the processing. They often point out that when a new version of a protocol is released, nodes are free to run it or not, and through this act have an influence on how the platform evolves.

Either way, the question of accountability remains tricky. On top of this, even if it were possible to approach the owners of individual nodes with a request to edit or delete data, due to the immutable structure of the data it is very unlikely that they would be able to comply unless they shut themselves down completely.

What about the **network users** who sign and submit transactions to the blockchain network via a node? **If they submit personal data to the blockchain ledger as part of a business activity, they are most likely to be considered data controllers.** This would include entities that operate software as well as products or services that post personal data onto a blockchain (which is not recommended). However, **if they submit their own personal data for their own personal use**, for example to buy or sell crypto-assets, **they are likely to fall under the household exemption of the GDPR and may not be considered data controllers.**

What about the **publishers of smart contracts?** Smart contracts are pieces of software that can be deployed to a blockchain network and that, once deployed, may be executed independently from their publisher(s). Most importantly, this software is only executed when called by a network user, so **there is a debate as to whether this software should be seen as being operated by its publisher, by the network user calling it or by both. This debate will probably have to be resolved on a case-by-case basis.**

TENSIONS BETWEEN THE GDPR AND BLOCKCHAIN

# HOW SHOULD PERSONAL DATA BE ANONYMISED?

## 1. Personal data, pseudoanonymous and anonymous data

The GDPR applies to processing of personal data unless it has been anonymised; thus, the GDPR does not apply to anonymised data. The bar for what qualifies as anonymised is, however, set very high. Not only must the anonymisation technique be good enough to make it impossible to identify a natural person through any and all of the means 'reasonably likely to be used', the process must also be irreversible. It should not be possible to reconstitute the original data from the anonymised form.[1]

Any techniques that do not meet this standard are considered 'pseudonymous', not anonymous. Pseudonymised data remains subject to GDPR obligations.

Given the immutability of data in most blockchain networks, there is consensus within the community that the storage of personal data in a clear (i.e. unencrypted) way on a shared ledger is a bad idea. This recommendation applies to both public, permissionless and to private, permissioned networks.

Application developers have many data obfuscation, encryption and aggregation techniques at their disposal that can be used to turn personal data into digital signatures that are cryptographically linked to the original data without actually revealing that data.

**There are intense debates within the community as to what specific techniques may be used to turn personal data into anonymous data.** These debates have not been fully settled by law nor by regulators. They are important because, in many business applications, software developers would like to be able to use the blockchain to store digital signatures of data that exists outside of the blockchain in order to create immutable proof that this data has been generated or validated at a specific point in time by a specific actor. Supply chain and workflow tracking applications are an example of this.

In practical terms, when considering the use of obfuscation, encryption and aggregation techniques to process personal data, one must evaluate two risks in detail:

- **Reversal risk**, or the risk that, despite the cryptographic technique used, it is possible to reverse the process and reconstitute the original data, for example by using brute force decryption.
- **Linkability risk**, or the risk that it is possible to link encrypted data to an individual by examining patterns of usage or context, or by comparison to other pieces of information.

## 2. Obfuscation of personal addresses

### a. Public keys or addresses on a blockchain are generally personal data

Many blockchains use 'public/private key' cryptography as a means to provide or derive addresses of the senders and receivers of

[1] Opinion 05/2014 on Anonymisation Techniques, Article 29 Working Party.

transactions. A public key or the address derived from it is akin to a number on a post office box. Someone can send information to this number, but only the possessor of the private key can open the box and get the information out.

Since the public key is a long string of quasi-random characters there is, on the surface, no way to tell anything about the owner of the public key from the private key.

Because on some public blockchains the addresses of the senders and receivers of transactions can be seen by everyone, under the GDPR such addresses would often be considered pseudonymous, especially in cases where there is a clear linkability risk.

If for example someone uses the same address for multiple transactions, then patterns begin to emerge. These patterns can, perhaps combined with other types of information, be used to indirectly identify individuals, and such techniques are indeed already used.

### *b. Address obfuscation techniques*

The most common address obfuscation technique is called a **third-party indirection service.** It consists in asking a third party to aggregate many blockchain transactions and post them to the ledger using their own public key. This is, for example, what sometimes happens when somebody asks an online trading platform to purchase crypto-assets on their behalf. The person's own single transaction is usually not revealed on the public blockchain.

**Ring signatures** are another technique whereby multiple parties sign a given transaction in such a way that an outsider can be sure that one of the parties is the legitimate signer, but not which one.

Address obfuscation techniques can be implemented in many ways, and each needs to be examined in detail on a case-by-case basis in light of the GDPR. There are also some blockchain technologies that do not reveal public keys or addresses on the shared ledger.

## *3. Encryption of personal data*

### *a. A simplified taxonomy of encryption techniques*

Cryptography is a highly technical subject, but for the sake of this discussion — and at the risk of over-simplifying — we describe two main techniques that are relevant to the GDPR.

- **Reversible encryption.** Reversible encryption involves scrambling a piece of data in such a way that its contents cannot be understood. Only the person in possession of the encryption key can decrypt it. There are various types of reversible encryption, such as symmetric encryption (the same key is used for encryption and decryption) and asymmetric encryption (different keys are used, also reffered above as public/private key encryption).
- **Hashing (non-reversible encryption).** Blockchains make heavy use of hashes. A cryptographic hash is a mathematical technique that allows you to generate a unique, fixed length string of characters from any set of digital data. There is no limit to how large a file you can hash. Whether it's a short note or the complete contents of the Internet, when you run the hashing function you will always get a unique text phrase of a certain fixed length, say 64

characters (it depends on the hashing function used). More importantly, if you change even one byte of the underlying data, the hash itself will be dramatically different, making it extremely clear that this underlying data was modified. Hashes are often referred to as digital fingerprints: no two are alike. Blockchains use hashes, among other things, to secure the current state of the chain (using the fingerprint like a seal to lock the chain every time a new valid block is entered) and to provide a means to uniquely reference data that is kept off the chain.

There are other, more advanced cryptographic techniques that are increasingly talked about in the blockchain ecosystem, and are described below.

### b. Reversibly encrypted personal data is personal data

It may seem surprising at first, but **even if strong encryption is employed on personal data, the result is almost surely pseudonymous, not anonymous.** This is for the simple reason that, as long as the key exists somewhere, the data can be decrypted, leading to a reversal risk.

On top of this, the technology and the science of cryptography constantly evolves. We have seen many reversible encryption techniques that once were secure eventually be cracked. We can expect that the techniques used today may be cracked in the future.

This means that reversibly encrypted personal data remains in the scope of the GDPR.

### c. Hashed personal data is a grey area

Hashing is at the heart of many of the most important properties of blockchains, providing much of the “magic” of decentralisation. This question of **whether hashed personal data should be considered personal data is hotly debated at present**, and unfortunately much of this debate relies on rather complex details.

Also, it should be kept in mind that not all hashing algorithms are equal and that the most advanced algorithms[2] should always be preferred.

As stated above, these issues have not been conclusively settled by the data protection authorities, the EDPB or in court. **At this stage, a desirable outcome of the debate regarding the status of hashed personal data could be: it depends.** The gist of it could potentially come down to the question of identifying potential reversibility or linkability risks.

When it comes to the reversibility risk, a brute force attack could succeed in reversing the hash if the original data is of a known, and relatively small size, as pointed out by the Article 29 Working Party:

> *‘If the range of input values the hash function are known they can be replayed through the hash function in order to derive the correct value for a particular record. For instance, if a dataset was pseudonymised by hashing the national identification number, then this can be derived simply by hashing all possible input values and comparing the result with those values in the dataset.’* [3]

[2] At the time of writing this report, SHA-3 is considered one of the most advanced hashing algorithms in the world.

[3] Op. cit.

## TENSIONS BETWEEN THE GDPR AND BLOCKCHAIN

It is possible to mitigate this risk using techniques such as ‘salting’ or ‘peppering’ hashes, which involve adding extra information to the data to make it large enough that a brute force attack would be extremely unlikely to reverse the data in, say, the next fifty years. (The difference between a salt and a pepper is that the salt is stored off-chain alongside the hash by the actor who generated the hash, whereas a pepper is stored secretly or even not stored at all).

**When it comes to the linkability risk**, there are situations where pattern analysis makes it possible to uncover information regarding a particular individual. For example, let’s imagine that you are using an application that performs certain buy or sell transactions on your behalf, and posts a hash of your address to a blockchain ledger to notarise each transaction.

In this case, the recorded hash is the same every time that a given user orders a transaction, which makes it possible for me to analyse the times and frequency of the transactions of each user. I can uncover your entire transaction behaviour if I happen to learn about one particular transaction that you have completed at a specific date and time. This example is similar to the one examined by the French Conseil d’Etat when the company JCDecaux was storing hash identifiers of mobile phones along with their location coordinates.

On the other hand, let’s imagine that the application posts the hash of a complex dataset each time that you make a transaction. The original dataset could include the details of the trade (investor name, asset, price, date, etc.) as well as random characters to make it larger. In this case, the hash would be unique for every single transaction, and it would be practically infeasible for a third party to derive personal data from the analysis of these unique hashes.

**To recap, it is desirable that the reversal risk and the linkability risk should be assessed on a case-by-case basis.** The table below illustrates such a potential assessment in the situations that we have mentioned, assuming that a state-of-the-art hashing algorithm is used. One should note, however, that it is entirely possible that the data protection authorities, the EDPB or the courts will adopt a much more cautious evaluation of these risks.

| | Situation a)<br>Hash is used to replace a unique attribute in a dataset | Situation b)<br>Hash is used as a one-time value to notarise the state of a dataset |
|---|---|---|
| Reversal risk (reverse engineering) | **Medium.** Brute force can be considered viable if the size of the input is known or within a small range (e.g. SSN, password, name)<br>Can potentially be mitigated using a salt or pepper. | **Low.** Reverse engineering is non-trivial as the size of the input can range from a few bytes to hundreds of terabytes and be coupled with multiple layers of hashing. |
| Linkability risk (via data analysis) | **High.** It is possible to conduct pattern analysis and trace data back to the individual, potentially with the help of other sources of information. | **Low.** Each hash is unique. There is no obvious way to cross-analyse the data. |

### *d. Many advanced cryptographic techniques are promising for the mid-term*

Many advanced cryptographic techniques are being developed in the context of blockchain that could eventually allow application developers to implement even more robust data anonymisation approaches.

**Zero-knowledge proofs (ZKP)**[1] are advanced cryptographic techniques that allow someone to produce proof of a statement without disclosing the data underlying that statement. For example, someone can produce proof that they are over 18 years old without disclosing their actual age. ZKP applications hold great promise when it comes to privacy-by-design and self-sovereign ownership of personal data.

However, there are few, if any, large-scale implementations of these techniques, and many subtleties in terms of how to apply them. For instance, the fact that someone is over 18 years old is still personal data.

**Homomorphic encryption** techniques are advanced cryptographic methods that allow someone to request distributed computations to be performed by private servers. While the underlying data of these computations is never revealed or shared on the blockchain, it is theoretically possible to obtain a cryptographic proof that the aggregated result of these computations is correct. These techniques would be implemented outside of the blockchain network ('off-chain') but it could potentially be useful to use the blockchain to store these proofs of computation for every stakeholder to see.

In **secure multi-party computation**, a group of actors jointly carry out the computation needed for a transaction in such a way that each party only has part of the underlying data, and no party can deduce from their particular part what the full data set was. Future improvements to this technology could very well lead to truly anonymised underlying data, if methods to ensure data cannot be reinstated are proven effective.

Here again, there are few, if any, large-scale implementations of these techniques, and many subtleties in terms of how to apply them.

Suffice it to say at this stage that these very promising areas of research and development are likely to play an integral role in how blockchain-based applications can be made compliant with the GDPR during the coming years. **The many possible uses of these technologies will need to be evaluated on a case-by-case basis.**

## *4. Aggregation of personal data*

Data aggregation techniques can be used in conjunction with above-mentioned obfuscation and encryption techniques. For example, large amounts of data from many data subjects can be aggregated into a single digital signature that is added to the blockchain ledger. That digital signature can then serve as proof-of-existence of every single underlying piece of data.

We will not discuss each technique in detail here, except to mention that many of them rely on data structures called Merkle trees that involve hashing functions

[1] Huixin Wu and Feng Wang, A Survey of Noninteractive Zero Knowledge Proof System and Its Applications, The Scientific World Journal, vol. 2014, Article ID 560484, 7 pages, 2014.

and make the hashing process even more robust and potentially anonymised. **Under certain conditions, it should be possible to anonymise personal data using these data aggregation techniques.**

**The possibilities offered by data aggregation techniques to anonymise personal data are likely to be instrumental for the development of the blockchain ecosystem.** Many blockchain experts believe that private, permissioned blockchain networks are best suited to log individual transactions. However, these scattered networks are unlikely to deliver transformative economic value if they are unable to inter-operate with each other. One hypothesis is that interoperability can be achieved by creating bridges between these private networks and public blockchains. Such bridges involve communication between private and public blockchains, leveraging data aggregation techniques in order to post anonymised data to public blockchains.

Again, each application needs to be examined on a case-by-case basis in light of the GDPR.

# BLOCKCHAINS AND THE GDPR'S RIGHTS AND OBLIGATIONS

Now that we understand the most important issues related to applying the GDPR in a blockchain world, we can look at some of the **other tensions related to the GDPR's data protection principles and the rights and obligations it specifies.**

At this point, it should be noted that the GDPR does not exist in a regulatory vacuum. Quite the contrary, it is part of a universe of other regulations including financial (and anti-money laundering) regulations. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality.

## *1. Lawfulness of processing*

In a decentralised network, it is not always so straightforward to determine on what legal basis data is being processed. According to the GDPR,[1] personal data can be processed only if one of the six legal bases mentioned earlier applies.

Consider the issue of consent. Who does a user give consent to in a public, permissionless situation, when it is unclear who the controller is?

It could of course be argued that, by choosing to use a decentralised network like Bitcoin, the user is *de facto* providing consent. The GDPR, however, stipulates that consent be specific and unambiguous, which seems to imply an

[1] GDPR, op. cit., Article 6.

active granting of permission, not a passive one. Similarly, one could argue that by initiating a transaction a user is entering into a contractual obligation with the platform and that this could form the basis for processing. But here too we are dealing with a passive act. And without explicit terms or a named counterparty, it would be an odd contract and one difficult to enforce.

This means that we are dealing with a grey area where, in some cases, it will not be possible to identify a controller.

This does not mean that all applications built on public, permissionless blockchains fall into that grey area. In many cases, it will be possible to identify an entity that operates the product or service and acts as an intermediary between individual users and the blockchain.

The issue of lawfulness is more straightforward (but still complex) in the context of a private, permissioned network, as it is possible to require that each network participant should agree to certain terms and conditions before being granted access to the network.

## 2. Data minimisation and right to erasure and rectification

Many of the rights and obligations specified in the GDPR seem to clash with the way blockchains store data. As we have seen, blockchains are generally designed so that data, once written to the chain, can't be changed. This immutability is a key property of the technology.

Under these circumstances, how can a data subject exercise his or her right to erasure or rectification? Even if a data controller could be found, on the Bitcoin network, for example, it is impossible to go back and delete or update the record of a transaction without destroying the chain. The whole point of such a blockchain is to ensure that transactions, including the parties to them, are never forgotten in order to enable decentralised trust.

These issues are not resolved just by moving to a private, permissioned blockchain network, unless that network is designed in a way that each and every piece of data is readable by only the parties that absolutely need to, and can be rectified or erased at the request of the data subject.

However, it should be noted that the GDPR doesn't specify what constitutes erasure. In this context, the French CNIL acknowledges[2] that some encryption techniques, coupled with key destruction, can potentially be considered erasure even if it's not erasure in the strictest sense.

## 3. Right of access

The GDPR understands a 'right of access',[3] meaning that the data subject has the right to find out from the controller if his or her data is being processed, and if so, for what purpose, who the data is being shared with, and so on.

Here too we have the problem of who to turn to get this information if there is no identified controller. And even if the data subject could identify and communicate with a specific node, the node wouldn't necessarily be able to answer these questions.

[2] Blockchain et RGPD : quelles solutions pour un usage responsable en présence de données personnelles?, CNIL, September 2018.
[3] GDPR, op. cit., Article 15.

## 4. Automated processing

As part of their right to access, data subjects can enquire from the data controller whether or not their data is being used for automated decision making. This brings up a special issue with regards to newer blockchain technologies.

The GDPR cares about automated decision making[4] because, among other things, it wants to protect people against indiscriminate profiling, or being subject to some legal or other consequence solely on the basis of a decision made by a machine.

For this reason the regulation stipulates that data subjects have the right to be informed that such processing is taking place and that they have a right to request human intervention or challenge a decision.

There are those who believe that this provision could have an effect on how people use smart contracts. Smart contracts have been heralded for their potential to introduce radical automation in many use cases. The question arises, however, of how to square them with the provisions of the GDPR. If smart contract developers have to introduce measures to allow for human intervention, the trust that transaction participants have in smart contracts could be dramatically curtailed.

We should acknowledge, however, that the issue of automated processing may not be the most pressing one at the moment, as there are not many blockchain use cases where smart contracts are used for individual profiling, credit or insurance underwriting decisions, or the like.

## 5. Territoriality

There are also obligations in terms of where data processing can take place, also known as 'transfers of personal data to third countries'. The GDPR specifies that personal data can generally only be transferred to third countries if they are deemed 'adequate'[5] — that is, if they are deemed to provide data protection that is essentially equivalent to that in the EU — or if the data controller can otherwise introduce appropriate safeguards that the data will be processed in a manner consistent with that law. In any event, transfers to third countries may only be carried out in full compliance with the GDPR.

This can be problematic if personal data is stored on a permissionless blockchain, and is also an issue for permissioned blockchain networks if their scope is global, as is often the case.

## 6. Data protection by design and by default

Finally, there are issues involved with how blockchains are designed and governed. For instance, the GDPR stipulates that data protection should be 'baked in' to platforms, and not added on top. This is the principle of data protection by design and by default.[6]

As blockchain technology is still immature and often developed by open source communities, the way that personal data protection is built in may leave some room for improvement. The good news is that the technology is at a stage where the foundations are still being built,

[4] GDPR, op. cit., Article 22.

[5] GDPR, op. cit., Article 45.

[6] GDPR, op. cit., Article 25.

and some of these foundations will be able to incorporate the spirit and the letter of the GDPR over time.

# Opposites attract: Resolving the tensions between blockchain and the GDPR

Based on the above, readers might have the impression that blockchain and the GDPR are incompatible, especially in situations where data is stored and processed but no controller can be clearly identified.

This is far from the truth. While there are serious tensions, we do not believe that the GDPR means the end of blockchain innovation, or even the end of public blockchain networks in the European Union.

**To be clear, these tensions cannot be resolved by this thematic report.** Only the EDPB, the courts, and other regulators and government agencies are in a position to do this.

Going forward, we expect regulatory agencies to gradually bring forth proposals that will clarify the issues outlined in this report, such as:

- **The identification and obligations of data controllers and processors,** acknowledging that there are situations where it is difficult, and perhaps impossible, to identify data controllers, for example when individual users are posting transactions or calling decentralised smart contracts on a public, permissionless blockchain for their own individual purpose.
- **The anonymisation of personal data,** and the validity of various techniques that allow users to record 'proofs of data' on the blockchain without actually revealing the data.
- **Other issues** such as lawfulness, data minimisation, right to erasure and rectification, right of access, automated processing, territoriality and data protection by design and by default.

Meanwhile, in this section, we propose four rule-of-thumb principles that entrepreneurs and innovators can consider when designing blockchain-based applications.

Let's be clear again that it's not all about the technology, it's about how the technology is used. As we mentioned in the introduction, there is no such thing as a GDPR-compliant blockchain, just as there are no such things as GDPR-compliant Internet or artificial intelligence. There are only GDPR-compliant use cases and applications.

## *Principle 1. Start with the big picture: how is user value created, how is data used, and do you really need blockchain?*

The interplay between the GDPR and blockchain is complex and it is easy for entrepreneurs to get lost in the minute details.

Instead, entrepreneurs must start with the key question of how data will be used to create user value: **what kind of data do they**

**need, who must be able to query it, for what purpose, on what legal basis, and for how long?** Only then, and with the GDPR principle of data protection by design and by default as laid out in Article 25 in mind, should they look to architect their solution.

In doing so, another key element to consider is that blockchain technology is not the solution to every problem. **Entrepreneurs should not assume that using blockchain automatically makes an application more secure or cheaper, or that it automatically equates to data protection or privacy.**

**One example is business-to-business applications.** Many public and private blockchains are used to eliminate the need for intermediaries in business-to-business transactions. Each entity should be able to manage the personal data of its users separately from the blockchain ('off-chain') and use blockchain technology to transact with other businesses on an aggregated basis in a faster, cheaper way that does not involve posting the details of individual user transactions to the blockchain.

## *Principle 2. Avoid storing personal data on a blockchain. Make full use of data obfuscation, encryption and aggregation techniques in order to anonymise data.*

This report describes a number of tensions and complexities related to how the GDPR applies to blockchain networks and applications. While some of the tensions are specific to public, permissionless networks, many of them have an impact on private and permissioned networks too (such as data minimisation and the right of erasure and rectification).

**Practically, this means that if a business is likely to be identified as a data controller by a regulator or the courts, this business should avoid storing any personal data on any blockchain. This recommendation applies even if the data is encrypted using reversible encryption techniques.**

In this report, we have described a number of data obfuscation, irreversible encryption and aggregation techniques that can potentially be used to anonymise personal data. These techniques are hotly debated and there is no official guidance on how they can be used in blockchain networks.

Even if we cannot recommend specifics, **it could be argued that blockchain networks should be used to store immutable proofs that certain data exists, rather than to store the data itself.**

For example, let's imagine an innovative platform that uses a public blockchain to help job seekers to provide proof of their academic background and school reports to potential employers.

As your school report contains personal data, it cannot be stored on the blockchain, even in reversibly encrypted form. Instead the platform could use hashing and aggregation techniques to generate a single-use digital signature of your school report, and store that digital signature in the blockchain, along with a timestamp and the cryptographic signature of the institution that generated the report. Subsequently, as a job seeker you will show the school report to the employer, completely outside of the blockchain. The employer, in turn, will be able confirm that the report is genuine by locating the signature and its timestamp in the blockchain.

In this example, we note that the right of rectification can be implemented. For example, let's imagine that your school report contains an erroneous grade. You can destroy the report stored outside of the blockchain, and ask the institution to generate a new report which will have its own distinct digital signature on the blockchain. The previous digital signature will simply be 'left hanging', with no off-chain data to point to.

**It would be beneficial to the blockchain industry that a hash in this context is not systematically interpreted by the EDPB as personal data, based on the arguments explained in the previous sections.**

## *Principle 3. Collect personal data off-chain or, if the blockchain can't be avoided, on private, permissioned blockchain networks. Consider personal data carefully when connecting private blockchains with public ones.*

To the extent that some personal data must be stored or processed on a blockchain, for example in the regulated financial services sector, **it is absolutely essential that the data should be stored and processed in a blockchain that is as tightly controlled as possible.**

Personal data could be restricted to a permissioned consortium blockchain with a small number of nodes, where it is possible to require consortium members to agree on contractual terms that define precisely their roles and duties and the privacy policy towards end users, as well as the process for amending the data if needed. Consortium members may form a separate legal entity that will act as the data controller, or they may elect to act as joint controllers. It is, in turn, possible for the data controller to present clear terms and conditions to end users.

Over time, we expect to see many of these applications adopt **multi-layered blockchain designs.** For instance, a two-layer design typically involves two interoperable blockchains:

- **A private, permissioned consortium blockchain network, operated by just a few dozen nodes,** where the actual real-time processing takes place. This network could, for example, be running an exchange marketplace for crypto-assets. Such a network is fast but not very decentralised, and it is not interoperable with other networks worldwide.
- **A base, public blockchain, operated by thousands of nodes,** is on the other hand very decentralised and very difficult for anyone to disrupt or take control of, but not private and not very fast. The base blockchain can be used to store crypto-assets over long periods of time without exchanging them, or it can be used as a bridge to move these crypto-assets from one private trading network to another.

In such a design, the base blockchain makes it possible for the private network to interoperate with other networks worldwide, but consortium members must be extremely careful that no personal data is compromised when data is exchanged back and forth between the two layers.

## *Principle 4. Continue to innovate, and be as clear and transparent as possible with users.*

At the time of writing this report, there are many open questions as to the precise interpretation of the GDPR to applications built on blockchain technology. This should not deter developers and entrepreneurs from innovating, especially if they are convinced that they are doing the right thing for their users.

Many considerations will have to be looked at on a case-by-case basis. Innovators should apply common sense and work in collaboration with regulators and the community to get feedback on their solutions.

**In situations where application developers or consortiums act as intermediaries between individual users and blockchain networks, they will most likely be considered data controllers,** and must ensure that they can carry out their obligations. Their responsibilities would include informing data subjects of what is happening with their data, conducting data protection impact assessments, and ensuring they have the means to carry out requests from data subjects to exercise their rights, for example the right to amendment or erasure. This would be handled in different ways depending on what data is on-chain or off-chain, but fundamentally would involve terms of service, privacy policies and consent forms, as is the case for other web and mobile applications.

However, in situations where no data controller can be clearly identified, regulatory intervention or clarification could be needed to allow peer-to-peer ecosystems to flourish rather than fear that any ecosystem participant could potentially be found liable as a joint controller.

We note that many new technology developments could make it easier for innovators and entrepreneurs to comply with the GDPR in the long run. For example:

- **Developers are working on pruning techniques that allow data to be removed from blockchains when it is no longer needed or wanted.** This work is generally done with the idea of improving performance by reducing the size of the chain, but pruning techniques could also theoretically be employed to meet the GDPR's right to erasure requirements.
- Other cryptographers are working on reversible encryption techniques that they claim to be **quantum-resistant,** i.e. that cannot be broken by quantum computers.[1]
- **Some projects are exploring the use of 'chameleon' hashes.[2]** In simple terms, such a hash contains a 'trapdoor' that allows the hashed data to be broken. If a block associated with the hash needs to be changed, this trapdoor can be used to open that block, change the data, and regenerate the block. Although this functionality can't be added retroactively to an existing blockchain, and while there is still a problem of unamended versions of the blockchain remaining available, this technique could be useful in specific use cases.

Hundreds of developers around the world are currently working on these techniques and others. We can be hopeful that accepted standards and best practices will emerge within the next three to five years.

[1] See Post Quantum Cryptography Standardization, FALCON and CRYSTALS.
[2] Chameleon-Hashes with Ephemeral Trapdoors.

# Appendix – Blockchain Terminology

### What is a blockchain?

Blockchain is one of the major technological breakthroughs of the past decade. A technology that allows large groups of people and organisations to reach agreement on and permanently record information without a central authority, it has been recognised as an important tool for building a fair, inclusive, secure and democratic digital economy. This has significant implications for how we think about many of our economic, social and political institutions.

### How does it work?

At its core, blockchain is a shared, peer-to-peer database. While there are currently several different kinds of blockchains in existence, they share certain functional characteristics. They generally include a means for nodes on the network to communicate directly with each other. They have a mechanism for nodes on the network to propose the addition of information to the database, usually in the form of some transaction, and a consensus mechanism by which the network can validate what is the agreed-upon version of the database.

Blockchain gets its name from the fact that data is stored in groups known as blocks, and that each validated block is cryptographically sealed to the previous block, forming an ever-growing chain of data. Instead of being stored in a central location, all the nodes in the network share an identical copy of the blockchain, continuously updating it as new valid blocks are added.

### What is it used for?

Blockchain is a technology that can be used to decentralise and automate processes in a large number of contexts. The attributes of blockchain allow for large numbers of individuals or entities, whether collaborators or competitors, to come to consensus on information and immutably store it. For this reason, blockchain has been described as a 'trust machine'.

## APPENDIX – BLOCKCHAIN TERMINOLOGY

The potential use cases for blockchain are vast. People are looking at blockchain technology to disrupt most industries, including from automotive, banking, education, energy and e-Government to healthcare, insurance, law, music, art, real estate and travel. While blockchain is definitely not the solution for every problem, smart contract automation and disintermediation enable reduced costs, lower risks of errors and fraud, and drastically improved speed and experience in many processes.

### Glossary

The vocabulary used in the context of blockchains is quite specific and can be hard to understand. Here are the essential concepts you should know in order to navigate this breakthrough technology:

- **Node:** A node is a computer running specific software which allows that computer to process and communicate pieces of information to other nodes. In blockchains, each node stores a copy of the ledger and information is relayed from peer node to peer node until transmitted to all nodes in the network.
- **Signature:** Signing a message or a transaction consists in encrypting data using a pair of asymmetric keys. Asymmetric cryptography allows someone to interchangeably use one key for encrypting and the other key for decrypting. Data is encrypted using the private key and can be decrypted by third-party actors using the public key to verify the message was sent by the holder of the private key.
- **Transaction:** Transaction are the most granular piece of information that can be shared among a blockchain network. They are generated by users and include information such as the value of the transfer, address of the receiver and data payload. Before sending a transaction to the network a user signs its contents by using a cryptographic private key. By controlling the validity of signatures, nodes can figure out who is the sender of a transaction and ensure that transaction content has not been manipulated while being transmitted over the network.
- **Hash:** A hash is the result of a function that transforms data into a unique, fixed length digest that cannot be reversed to produce the input. It can be viewed as the digital version of a fingerprint, for any type of data.
- **Block:** A block is the data structure used in blockchains to group transactions. In addition to transactions, blocks include other elements such as the hash of the previous block and a timestamp.
- **Smart contract:** Smart contracts are pieces of code stored on the blockchain that will self-execute once deployed, thus leveraging

## APPENDIX – BLOCKCHAIN TERMINOLOGY

the trust and security of the blockchain network. They allow users to automate business logic and therefore enhance or completely redesign business processes and services.

- **Token:** Tokens are a type of digital asset that can be tracked or transferred on a blockchain. Tokens are often used as a digital representation of assets like commodities, stocks and even physical products. Tokens are also used to incentivise actors in maintaining and securing blockchain networks.
- **Consensus algorithm:** Consensus algorithms ensure convergence towards a single, immutable version of the ledger. They allow actors on the network to agree on the content recorded on the blockchain, taking into consideration the fact that some actors can be faulty or malicious. This can be achieved by various means depending on the specific needs. The most famous consensus algorithms include Proof-of-Work, Proof-of-Stake and Proof-of-Authority.
- **Validator nodes:** Validator nodes are specific nodes in a network that are responsible for constituting blocks and broadcasting these blocks with the network. To create a valid new block they have to follow the exact rules specified by the consensus algorithm.

**Learn more about blockchain by watching a recording of our Ask me Anything session.**

# BLOCKCHAIN & THE GDPR

**The General Data Protection Regulation (GDPR) is a regulation in EU law that aims to give control to citizens and residents over their personal data.**

## HISTORY OF THE GDPR

### 1995>>>>>>>>>>>>>>>>

Data Protection Directive (DPR) is passed, it regulates the processing of personal data within the EU.

### 2012>>>>>>>>>>

European Commission introduces plan to develop GDPR.

### <<<<<MAR 2014

European Parliament adopts GDPR. Votes in favour: 621 against: 10 abstentions: 22

### <DEC 2015<<<<

The European Parliament, the Council and the Commission reach an agreement on the GDPR.

### >>>>MAY 2015>>>>>>>>>>

The regulation (EU) 2016/679 enters into force.

### MAY 2018>>>

Companies must be in full compliance.

## CORE PRINCIPLES

**LAWFULNESS, FAIRNESS AND TRANSPARENCY.**
Personal data should be «processed lawfully, fairly and in a transparent manne

**PURPOSE LIMITATION.**
Personal data shall be «collected for specified, explicit and legitimate purposes», and only used for those purposes that have been stated

**DATA MINIMISATION.**
The personal data collected should be «adequate, relevant and limited to what is necessary» for the stated purposes

**ACCURACY.**
Personal data held by a controller must be «accurate and, where necessary, kept up to date»

**STORAGE LIMITATION.**
Personal data must be «kept in a form which permits identification of data subjects for no longer than is necessary» for the stated purposes

**INTEGRITY AND CONFIDENTIALITY.**
Personal data should be «processed in a manner that ensures appropriate security»

## GLOSSARY

### PERSONAL DATA

Any information related to a natural person or 'Data Subject', that can be used to directly or indirectly identify the person.

### THE CONTROLLER

The entity that determines the purposes, conditions and means of the processing of personal data.

### THE PROCESSOR

The entity that processes data on behalf of the Data Controller. A data controller can be processor at the same time.

### PSEUDONYMISATION

The processing of personal data such that it can no longer be attributed to a single data subject without the use of additional data, so long as said additional data stays separate to ensure non-attribution.

### ANONYMISATION

The process of removing personal identifiers, both direct and indirect, that may lead to an individual being identified.

### CONSENT

Refers to any freely given, specific and informed indication of the wishes of a data subject, by which he/she agrees to personal data relating to him/her being processed

*Read the full report on eublockchainforum.eu*

EUBlockchain
Observatory and Forum

# BLOCKCHAIN & THE GDPR

## TENSIONS BETWEEN THE GDPR AND BLOCKCHAIN

The tensions between the GDPR and Blockchain revolve mainly around three issues. These issues have not been conclusively settled by the data protection authorities, the European Data Protection Board (EDPB) or in court.

### IDENTIFICATION AND OBLIGATIONS OF DATA CONTROLLERS AND PROCESSORS.

While there are many situations where data controllers and data processors can be identified and comply with their obligations, there are also cases where it is difficult to identify a data controller.

### THE ANONYMISATION OF PERSONAL DATA.

What does it take to anonymise personal data to the point where the resulting output can potentially be stored in a blockchain network?

### EXERCISE OF SOME DATA SUBJECT RIGHTS.

Blockchain implies an environment and operating paradigms that may make it difficult to exercise some data subject rights such as the right to erasure or rights related to automated processing.

## RECOMMENDATIONS

Four rule-of-thumb principles that entrepreneurs and innovators can consider when designing blockchain-based applications

1 *Start with the big picture: how user value is created, how is data used, and what is blockchain really used for*

2 *Avoid storing personal data on a blockchain. Make full use of data obfuscation, encryption and aggregation techniques in order to anonymise data*

3 *Collect personal data off-chain or, if the blockchain can't be avoided, on private permissioned blockchain networks. Consider personal data carefully when connecting private blockchains with public ones*

4 *Continue to innovate, and be as clear and transparent as possible with users*

*Read the full report on eublockchainforum.eu*

EUBlockchain Observatory and Forum