Why Blockchain is not inherently at odds with GDPR[1]

Lokke Moerel and Marijn Storm

"*In an almost direct clash of intentions, the GDPR has effectively banned the use of blockchain technology in Europe because of its immutable nature.*" – **Forbes**

***Summary**:*
The current perception is that blockchain is not compatible with GDPR. We disagree that this is the case and explain why none of the issues identified by legal scholars and stakeholders are likely to pose issues for blockchain applications. Our conclusion is that GDPR is well able to regulate this new technology.

Current perception
The current conception amongst industry stakeholders is that blockchain (**BC**) is not compatible with GDPR, resulting in a call for urgent revision right after GDPR came into force. The concerns are fed by public statements of Jan-Philipp Albrecht (the MEP responsible for coordinating the Parliament's input for GDPR) that BC "*probably cannot be used for the processing of personal data*" and the CNIL cautioning in draft guidelines that public BC may not be the most appropriate technology for the processing of personal data and that priority should be given to other processing solutions that can achieve the same purpose. As the same results currently can always be achieved with another solution, this is a difficult standard to meet. The above conception is fed by recent in-depth publications on the data protection aspects of BC, which indeed paint a grim picture: the characteristics of public BC would be "*on a collision course*" and "*profoundly incompatible at a conceptual level*" with GDPR.

What are the issues?
GDPR requires identification of a central 'controller' who is responsible for compliance with GDPR, while a public BC decentralizes the storage and processing of personal data, as a result of which there is no such central point of control. For lack of a better alternative, the authors conclude that all 'nodes' involved in operating a BC qualify as a controller under GDPR, raising enforcement and jurisdictional issues that make it impossible for individuals to enforce their rights. The transparency and immutability of a public BC would further not sit well with principles of data confidentiality, data minimization, data accuracy, and the rights of individuals to correction and deletion of their data.

A different perspective
We disagree with this analysis. The main reason is that the authors focus on the shortcomings of the initial public (Bitcoin) BC when many new types of *permissioned private* and *consortium* BC already have been developed that significantly diverge from the original, permissionless public BC. In fact, these types of permissioned BC have been developed exactly in response to the shortcomings of public BC. The authors further consider the data processing implications of BC as if this technology itself constitutes a data processing activity for which a controller has to be identified. Controllership is, however, decided based on a specific use or deployment of a certain technology. BC, like the internet, is a general purpose technology (**GPT**) that is subsequently deployed by actors for a certain purpose in a specific context. We will explain below why none of

---

[1] ***Note:***
This blog is a summary version of a full publication of Lokke Moerel published in European Review of Private Law 6-2019 [825 – 852] and in The Cambridge Handbook of Smart Contracts, Blockchain Technology and Digital Platforms (September 2019).

these identified issues are currently hampering application of the GDPR to the internet and are equally unlikely to pose issues for BC applications.

The conclusion is that GDPR is well able to regulate this new technology. This does not, however, mean that BC will thus be suitable for all use and deployment cases.

Intermediaries will not become obsolete

We consider it highly unlikely that BC will make intermediaries obsolete; rather, it will replace the current intermediaries. The BC revolution is well described by the World Economic Forum (2017 report), indicating that where the last decades brought us the *internet of information*, we are now witnessing the rise of the *internet of value*, whereby we can send money and soon any form of digitized value – from stocks and bonds to intellectual property – directly and safely between us.

As BC is about *value* (rather than just 'information'), and therefore whether someone has ownership of money, stocks, houses, or not (as evidenced by the BC), the participants will insist that their stakes be safeguarded before the BC will be trusted. The prediction, therefore, is that, whenever BC applications are built for evidence and transfer of value, there will always be a set of *governance rules* reflecting the terms agreed by the participants of the ecosystem to regulate their relationship. The first examples indeed show new entities being set up mostly as a *consortium* (often including or funded by incumbents, such as banks), which are in charge of the governance of the BC platform, as well as separate entities operating a BC *application* on top of the BC platform for specific ecosystems. These BC are *permissioned*, in the sense that they implement membership rules that determine which parties have read or read/write authorization. To avoid jurisdictional and enforcement disputes, these rules will provide who the *responsible entity* is, as well as a choice of law and forum. The jurisdiction and enforcement issues raised by the authors, therefore, are likely not a realistic reflection of how these issues will be encountered in practice. The controller issue is solved as, in any event, this central entity deciding on purposes and means of the BC platform will qualify as the controller under GDPR. The entities operating the BC application on top of the BC platform will also qualify as controllers in their own right (potentially jointly with the controller of the BC platform).

Deja Vu

We here recall that early predictions in respect of the internet foresaw similar enforcement and jurisdictional issues. Every encounter of consumers in cyberspace would raise the possibility that diverse laws would apply and multiple courts would have jurisdiction, and a myriad of court cases was predicted. Contrary to these early expectations, there have been only isolated court cases dealing with online cross-border consumer disputes. One of the explanations is that stakeholders quickly found practical work-arounds in the form of contractual self-regulatory systems. Examples are the use of credit cards for online payments, which bring their own dispute resolution system, and the emergence of large intermediaries like eBay, which was at first just regulated by the ratings and review consumers could post, but later introduced full-fledged dispute resolution. Also, here the old intermediaries (retailers) were replaced by new intermediaries, again generating the required trust to do business. In fact, it is fair to say that there is very little happening on the internet that is not governed by some form of contract. The use of websites is regulated by their website , online purchases are governed by purchase terms, access to the internet is governed by the terms and conditions of ISPs, App stores have their own Terms & Conditions ("T&Cs"), search functionality is governed by the T&Cs of the provider of the search engine, etc. As happened with the internet, it is a justified expectation that the stakeholders involved in BC will implement their own contractual self-regulatory mechanisms to ensure adequate dispute resolution.

GDPR applies to the use of a technology, not the technology itself
The authors try to determine controllership in respect of BC technology at large, which would indeed raise the identified issues. Controllership is, however, decided based on a specific use or deployment of a certain technology. BC, like the internet, is a *GPT* that is subsequently deployed by actors for a certain purpose in a specific context. None of the issues raised by the authors have hampered the development of the internet, for the simple reason that controllership is not decided based on the technical level of operation of the relevant technology, but is based on who deploys this technology for a certain purpose. For example, a website owner uses the internet to offer its website. It is the website owner who qualifies as the controller in respect of the processing of any personal data via the website and not the operator of the technical infrastructure.

GDPR does not impose requirements on designers of technology
GDPR includes an obligation for the controller to set up data processing functions on the basis of *privacy-by-design* (Article 25 GDPR). GDPR does not impose this requirement on providers of software and infrastructure that are used to process personal data. As a consequence, individual controllers need to expressly instruct each of their technology suppliers to provide software and infrastructure that incorporate privacy-by-design in order to meet their controller obligations.

Although this indirect manner of regulating seems inefficient, the reality is that for technology developers, it is often difficult to foresee all possible deployments of their technology. As a consequence, it is difficult to implement all requirements into their product from the outset. It is often in the feedback loop of the users, customers, or society at large when the technology is deployed in practice that the design issues become apparent and are addressed. Too-strict upfront design requirements (in the form of standards) may even hamper innovation, and it may even lead to "widespread adoption of inferior technology" (as explained in this report of the World Economic Forum). In the words of Behlendorf (CEO of the Linux Foundation):

> "*The space is still so young that the desire for standards, while well-placed, runs the risk of hardening projects that have just come out of the lab," and "we need to avoid making serious architectural decisions that first become legacy and then become a hindrance.*"

GDPR is, just as its predecessor, technology agnostic (see Recital 15) in the sense that it provides for general data protection principles and requirements but does not prescribe any technology or technical manner for how these principles and requirements should be implemented. As BC is an emerging technology still in its infancy, GDPR works exactly as it is intended, challenging developers to think of creative ways for how to develop the technology in such a manner that the impact on the privacy of individuals can be mitigated and basic principles of GDPR can be complied with. That this may take several development cycles to be achieved is fully understood. The conclusion of the authors that GDPR is thus unable to embrace this new technology is missing the point that GDPR is intended to provide guidance on how to develop new technology in the first place. In Part 2 of this sequel, we will discuss how the transparency and immutability issues raised by BC can be addressed by implementing innovative privacy-by-design measures.

Why Blockchain is not inherently at odds with GDPR[2]

What are the real data protection issues?

The fact that BC, both public and private, is inherently transparent and immutable may clash with data minimization principles and may make it impossible to respond to rights of individuals to have their data corrected or deleted. BC is further, by definition, unable to forget; as a result, the right to be forgotten will be impossible to enforce. The transparency and immutability issues can, to a large extent, be addressed by implementing innovative privacy-by-design measures (see below for examples). Noteworthy is that these innovations are not necessarily triggered by privacy considerations, but mostly out of efficiency considerations.

In its most basic form, BC can be used to store plain text information on the ledger, which information can be accessed by those who have read rights. Storing all information on BC takes up a large amount of space on BC and takes a lot of energy both to run and cool the machines. Block space can be saved by separating (segregating) the signature ('witness') information from the transaction data (the 'payload'), so the network can increase the transactions processed. These measures may also, to a certain extent, mitigate transparency issues.

The immutability of BC further does not sit well with, for example, **smart contracts** in more complex transactions (as contracts often have to be amended for unforeseen circumstances); with **technological malfunction**, including in case of interference by hackers; and, more generally, with **human error** (known to lose their BC private key). Solving these issues will require solving the immutability of BC, which may also solve the issue of being able to respond to requests of individuals for deletion and the right to be forgotten.

Immutability is not always an issue

As a side note, we mention that the immutability of BC is not always an issue. For certain applications (in particular, in case of public registries), immutability is actually a requirement. Illustrative here is the judgment of the European Court of Justice (ECJ) in the *Manni* case. The plaintiff (Mr. Manni) requested deletion of his personal information from the Italian public company register, where information on his prior bankruptcy was recorded. He argued that this record in the company register was widely reused by data brokers, and as a result, his reputation was prejudiced, having a detrimental effect on his new business. The ECJ balanced the public interest in the legal certainty in trade and transparency of business information in the company register with the fundamental right to data protection and concluded that, in this case, the interference with the right to data protection was not disproportionate, taking into account the limited amount of personal information held in the company register.

In line with the above ruling, registering limited personal data in a BC for public registers like land ownership, trademark ownership, and company registers may, therefore, be justified. The above case entails that a balancing of interests should be made for each BC application. For other use cases, the balancing test may conclude that BC will not be suitable, as the impact on data protection will be disproportionate. An example of the latter would be if BC would be

---

[2] ***Note:***
This blog is a summary version of a full publication of Lokke Moerel published in European Review of Private Law 6-2019 [825 – 852] and in The Cambridge Handbook of Smart Contracts, Blockchain Technology and Digital Platforms (September 2019).

applied to provide air passengers with expedited access through the airport, while also recording all money spent in shops and restaurants at airports, subsequent transport, and accommodations on the BC for purposes of a loyalty program. Using BC for the commercial loyalty program would likely be disproportionate.

Privacy-by-Design Options

Limit ledger storage. The original Bitcoin BC stores the full ledger on every node, making it impossible to make changes to prior blocks and thus providing an indisputable ledger for all prior transactions. However, this also means that the personal data included on the ledger is shared with a large number of nodes (Bitcoin has approximately 9,500 nodes). Storing so many instances of personal data is at odds with the data minimization principle of GDPR, which requires access to personal data to be limited to the fewest possible recipients.

A privacy-by-design solution is to no longer store the entire ledger on all nodes. In most Bitcoin instances, the validity of a new block is verified by a consensus mechanism. This means that the creator of the block provides a unique hash of the information. The nodes make the same mathematical equations and, if the outcome of this hash is the same, the block is verified. This requires the nodes to have access to the information included on the block. However, the nodes would still be able to fulfill their verification function if they deleted the information after verification. This would increase the confidentiality of the personal data included on the block and, at the same time, has economic advantages. If each node has to store a full copy of the ledger, a large amount of storage capacity is required, which, in turn, requires a large investment in data storage and uses a lot of energy. Therefore, storing the ledger in one (or a few) instances, rather than on every node, has both privacy and economic advantages.

***Pruning*.** Most BC applications store all transactions since the start of the chain, dating back to the 'genesis block', which means that all transactions on that BC are stored infinitely (and, as set out above, are sometimes stored on all nodes). Storing data infinitely is, by definition, at odds with GDPR's data minimization requirement but also brings ever-increasing storage requirements. For example, during a stress test, the size of the BC of an Ethereum client increased to 40 gigabytes in the first three months of the test.

A privacy-by-design solution to this storage issue is pruning, which enables the node to verify a new block without processing historical transactions by having the node download as many block headers as it can and determine which header is on the end of the longest chain. Starting from this header on the longest chain, the node goes back 100 blocks to verify that the chain matches up. Because this verification process removes the need for retaining the entire chain history for verification purposes, this allows for the removal of unused blocks, which drastically lowers the required storage and implements data minimization into the BC. To ensure that no data is lost, the unused blocks can be stored in one or more archive nodes, which store all data just in case the rest of the network needs them in the future, but the 'active' nodes no longer have to process these archived blocks.

Privacy-friendly consensus. A privacy-by-design solution for the infinite storage issue is the concept of non-interactive zero-knowledge proof, which makes it possible to verify the correctness of a computation, e.g., a hash, without having to execute the computation or even learning what was executed. For example, the proposed currency Zerocoin works as follows. When a coin is purchased, a serial number is attributed to the coin, which can only be revealed using a random number. Using these two numbers, a user can generate a zero-knowledge proof for the fact that the user knows both the serial number and the random number. This

zero-knowledge proof can then be verified by the network without having access to the coin's serial number or the random number.

The potential use of zero-knowledge proof is not limited to the transfer of coins using BC but can be used to verify any computation without having access to the underlying information. This enables nodes to reach consensus on a new block without accessing the information on that block, and thus without sharing the personal data included on that block with the nodes.

Editable BC. A more radical approach that solves a number of BC data protection issues is the editable BC, for which Accenture has been awarded a patent. The editable BC uses the 'chameleon' hash function, which allows for changing the underlying information without changing the outcome of the hash function. This allows for changes to the underlying information for which the hash is already included on the BC, which makes it possible to correct (human) error or intentional (fraudulent) inaccuracies on the BC. This would allow for the execution of individuals' rights under GDPR, e.g., to correction and to be forgotten.

Solving the immutability of BC comes at a price. To a large extent, the trust in a BC application relies on the network's consensus on the content of a block and the immutability of the content thereafter. When removing this immutability, other measures should be implemented to retain (or gain) sufficient trust in the BC application for individuals and organizations to use it as a record of their transactions. The trust in a BC application could be retained if, for example, only a single trusted entity can make these changes, similar to the fact that only governments can make certain changes to governmental public registries. A different solution could be to implement a very strict change management procedure, which could include a consensus mechanism that verifies the legitimacy of a change. In any event, changes will have to be strictly logged to ensure that changes can always be reviewed and explained in the future.

BC 'self-sovereign' identity management. The well-known use cases of BC are mostly focused on administering transactions, but BC can also be deployed for *privacy enhancing* purposes, for example, by facilitating 'self-sovereign' identity management.

In the offline world, an individual's identity is mostly established by verifying an individual's driver's license or passport. The strength of this system follows from a trusted central governmental authority that provides these proofs of identity. However, because the online world does not follow the national boundaries of the offline world, it is difficult to appoint such a trusted centralized authority for an online proof of identity. By now, there are many initiatives to provide individuals with a digital identity. An example of how BC can be deployed for online identify management is the initiative of Microsoft and Accenture providing a BC-based solution designed to allow individuals with direct control over who has access to their personal data. Rather than all service providers each collecting and storing the personal data required for providing services to an individual, the personal data are stored off-chain, and the system only calls on these data when the individual grants access, whereby access can be limited both in scope and in time. For example, when an individual needs to prove his or her identity when renting a car, the access to the identifying information can be limited to what is necessary to provide this proof and for a short period of time only.

Decentralized identity management has a number of benefits. From a privacy point of view, it enables individuals to take back control over their digital identity, coined the 'self-sovereign identity'. Currently, many individuals are, for example, not aware of the use of their digital identity and personal data, e.g., for advertising purposes. By using a decentralized identity system, individuals would be able to decide who to give access to which information, for which

period of time. A single decentralized identity system also has economic benefits. Right now, a large number of companies are storing similar information about the same individuals. A decentralized identity management system makes this duplicated storage obsolete and ensures that companies have access to up-to-date information on an individual, insofar as the individual wants the company to have such access.