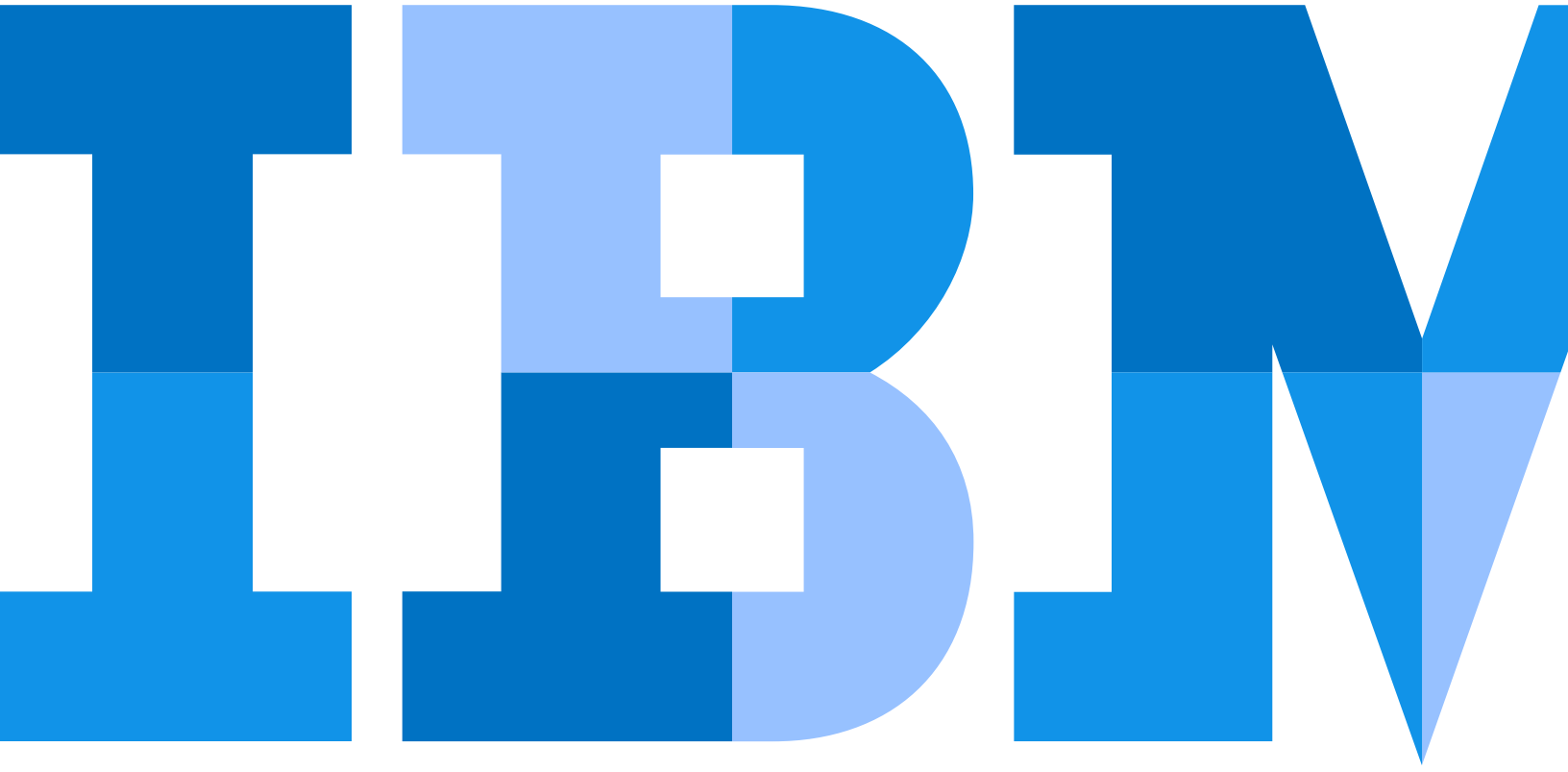
IBM Security
White paper

GDPR considerations for blockchain solution architects

Know your options before processing EU personal data on a blockchain



IBM

Legal notice

Clients are responsible for ensuring their own compliance with various laws and regulations, including the European Union (EU) General Data Protection Regulation (GDPR). Clients are solely responsible for obtaining the advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulations that may affect the clients' businesses and any actions the clients may need to take to comply with such laws and regulations.

The products, services, and other capabilities described herein are not suitable for all client situations and may have restricted availability. IBM does not provide legal, accounting or auditing advice or represent or warrant that its services or products will ensure that clients are in compliance with any law or regulation.

Introduction

The European Union (EU) General Data Protection Regulation (GDPR), which went into effect on 25 May 2018, is the most significant overhaul of EU data protection rules in 20 years. It imposes stringent requirements on companies that collect, store or process data about EU data subjects, regardless of whether their systems are in the EU. In addition, it extends several new rights to individuals in the EU (for example, the rights to data portability and to have data erased), and modifies some existing ones (for example, the right to access data). The penalties for non-compliance include fines of up to four percent of the global revenue of the firm or €20 million, whichever is greater; and regulators may have other powers, including the ability to issue "stop processing" orders.

The privacy rules established by the GDPR affect many business processes and technology solutions. Because of the GDPR's broad definition of "processing," compliance challenges arise in various contexts, including the collection, retention, modification, access and deletion of personal data. Projects with GDPR implications should have a clear idea of the types and amounts of data involved; the location of data storage; articulated policies for the retention, modification, access and deletion of that data; and the means to selectively delete data in an efficient, effective and demonstrable manner.

As a result, the GDPR may pose a challenge for architects of blockchain solutions that process EU personal data. This document outlines some of the blockchain properties that represent challenges from a GDPR perspective, describes various approaches for readying blockchain solutions for GDPR and outlines the need for a robust governance process if data pseudonymisation is used. In certain cases, blockchain can also help with achieving GDPR compliance, such as consent management or self-sovereign identities,[1] but that discussion is beyond the scope of this paper.

Selected GDPR terms:

- **"Processing."** Any operation performed on or using personal data, including collection, organisation, storage, adaptation, retrieval, use and transmission, among others.
- **Personal data.** Any information relating to a data subject.
- **Data subject.** An identified or identifiable individual in the EU.
- **Controller.** A natural or legal person, public authority, agency or other body that, alone or jointly with others, *determines the purposes and means of processing* personal data.
- **Processor.** Any natural or legal person, public authority, agency or other body that *processes personal data on behalf of a controller*.
- **Special data.** A separate category of personal data that has additional legal protections and requirements. Special data rules apply to several categories of data, including racial or ethnic origin, genetic data, biometrics and data concerning health, among others.

Blockchain and the challenge of GDPR

This paper focuses on three core features of blockchain that cause special challenges with the GDPR: distribution, immutability and permanence.

Distribution

Blockchain records are stored and replicated among a set of participants, regardless of their geographic location. In the simplest case, records are replicated among all participants in the blockchain. In other cases, the blockchain may allow storing records among a subset of participants. See the technical note on Hyperledger Fabric.

Distributing the ledger across multiple participants increases the difficulty of fraudulently altering past transactions, since the ledger would need to be modified in the same way on all nodes. In addition, the distributed network eliminates the need for a centralised, trusted intermediary to hold a master ledger on behalf of all participants.

However, the distributed nature of blockchain records may present challenges from a GDPR perspective. The GDPR prescribes strict guidelines for data controllers regarding the sharing of personal data, requiring that notice be provided to data subjects before sharing, allowing data subjects to object to processing and restricting access to those controllers and processors with a legal basis, for example, contractual or based on consent. By its nature, participants of a blockchain network may be in different regulatory regions. As a result, blockchain architects should give thought to how all participants on the chain account for data subjects' rights and restrictions, and whether those participants have appropriate controls in place prior to adding entities to the various communication channels of a blockchain network.

Immutability

Data on a blockchain is described as "immutable" because each block on the chain contains a cryptographic hash of the block contents, and each block is distributed among participants, meaning that it is infeasible to modify or delete data once it has been placed on a chain. Through cryptography, users of the shared ledger are authenticated and transactions are verified and stored in a such a way that they can be audited by all members of the network for the duration of the network. Network participants can't tamper with transaction records; and the transparency, security and durability of the network's data fosters trust among its members.

While the immutability of blockchain data provides several benefits in terms of transparency and security, it also poses some challenges from a GDPR perspective. The GDPR gives data subjects the right to request modification or erasure of their personal data and requires data controllers and processors to comply with such requests when necessary. If personal data is stored on the blockchain, and it is not possible to modify data on the blockchain, data subjects may not be able to exercise those rights.

Permanence

Records are added to a blockchain by append only and are stored permanently by some or all participants in the blockchain. The permanent storage of blocks is beneficial for several reasons, including auditability. However, this feature also appears to conflict with certain provisions of the GDPR, including the requirements of data minimisation and data retention. The GDPR requires that only necessary personal data be gathered, and that it be deleted once it's no longer necessary. In addition, if the lawful basis underlying the processing of the data is modified or rescinded (for example, individuals who originally consented to the use of their data later remove that consent), the data controller should update its records and cease that processing activity.

Summarising the challenge

Data on a blockchain is distributed among a set of nodes, theoretically immutable and stored permanently. Without these properties, a blockchain loses some or all its value. However, because data on a chain is immutable and permanent (that is, it can't be removed once it has been added), GDPR compliance can be challenging. To proceed with a blockchain project that includes the use of EU personal data, a solution architect needs to determine whether and how these opposing points of view can be reconciled.

Possible approaches to GDPR with blockchain

Despite the challenges articulated in the previous section, it may be possible to design GDPR-compliant solutions. This section discusses potential approaches to building blockchains that help fulfil GDPR requirements, as well as some associated pros and cons. A detailed technical description of these approaches is beyond the scope of this paper.

Avoid processing personal data

A simple approach to avoiding GDPR considerations is to avoid processing personal data on the blockchain. If no personal data processing occurs, the GDPR wouldn't apply. However, avoiding personal data altogether would limit the use of blockchain technology to projects dealing exclusively with non-personal data, which may not be practicable in all—or even most—cases.

Anonymise all personal data

Assuming the project in question involves the use of personal data, architects may need to find a way to store such data on the blockchain while also meeting GDPR requirements. One option may be to employ anonymisation before storing personal data on the blockchain, although this method also presents several challenges.

The authors of the GDPR made an explicit carve-out for anonymised data, stating that "principles of data protection should ... not apply to ... personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."[2] The problem is that full anonymisation is difficult to achieve because the GDPR text implies a high bar to successful anonymisation. Unlike previous interpretations of "anonymisation," which required only that identification of an individual was "not likely to take place," the GDPR requires one to consider the factors relevant to the likelihood of re-identification, including the costs, time and available technology.[2]

For various reasons, a zero likelihood of re-identification may be impossible to achieve. The blockchain controller would need to anonymise obvious personal data, for example, names, email addresses and other identifiers, as well as less-obvious information, such as Internet Protocol (IP)and Media Access Control (MAC) addresses. Even then, it's possible that an individual's identity could be discerned by analysing transaction information available to the network, or by devising a method of de-encrypting anonymised information to reveal the underlying information. In addition, techniques to support the re-identification of individuals from anonymised data sets are rapidly and continually improving.

Maintaining personal data on the blockchain presents risk of exposure and GDPR non-compliance. Anonymisation may be possible, but accomplishing it is challenging; and its durability is uncertain. If the anonymisation process were incomplete, it would indefinitely expose the underlying personal data to the blockchain network.

Encrypt all personal data

Another possible approach would be to simply encrypt personal data before storing it on the blockchain and manage access and erasure through control of encryption keys. However, in this scenario, personal data is still stored on the blockchain and can't be altered without creating a new block in the chain, thereby creating potential risk. Moreover, the encryption could be broken in the future, or the encryption keys disclosed, which would expose the underlying personal data.

This risk may be perceived to be too great by some customers and their supervisory authorities, considering that all data is replicated and permanent. Therefore, this paper suggests using other methods of managing personal data on a blockchain.

Pseudonymise all personal data

A more useful approach to meeting GDPR requirements and business needs in the context of blockchain is to make use of pseudonymisation. The text of the GDPR specifically addresses pseudonymisation, noting that it should separate visible identifiers from the underlying personal data "in such a way that the data can no longer be attributed to a specific data subject without the use of additional information."[3]

In this approach, personal identifiers are maintained off the blockchain, with some mechanism to link back to the chain. The controlled disclosure of this link provides the fine-grained access and logical delete mechanisms, for example, deletion of the links from the blockchain to personal identifiers off the chain, needed to meet GDPR requirements. This approach is in contrast to anonymisation, which removes personal identifiers altogether.

It's important to note that pseudonymised data is still personal data as defined by the GDPR, and so it's not "carved out" from the regulation in the way anonymised data would be. However, if the off-blockchain personal identifiers were subsequently deleted, the on-chain data would then be anonymised and thus out of scope for the GDPR.

Technical methods for pseudonymisation are beyond the scope of this paper, but architects would need to ensure that data on the chain is sufficiently pseudonymised to avoid a reasonable risk of re-identification.[4] As mentioned earlier, the re-identification risk also exists in anonymisation. Architects would need to determine whether any indirect identifiers, if combined, could be reasonably likely to lead to identification. Similarly, controls over the pseudonymised identifiers from the chain to the underlying personal data would have to be very robust, such that there would be no reasonable risk of accidental disclosure. The selected method would have to be clearly demonstrable and auditable.

Business governance for pseudonymisation

If architects choose to adopt the pseudonymisation approach, the overall business solution would also depend upon a rigorous business governance framework. There are several strands to this framework:

Governance of pseudonymisation links for access control

Access to personal data is managed by allocating access to personal identifiers off chain from pseudonymised links on chain. Managing this allocation of access is a critical ingredient to maintaining adequate privacy and meeting GDPR requirements.

Governance of pseudonymisation links for erasure

"Logical erasure" of personal data is the deletion of the links between the blockchain and personal identifiers off the chain, thereby irreversibly anonymising the data. Managing this process is critical to meeting GDPR requirements. Architects need to be sure that the links present no reasonable risk of re-identification of data subjects.

Governance of content—personal versus non-personal

Determining which data is "personal" is a critical part of determining which records, or parts of records, can be stored on a blockchain. Business governance mechanisms would need to be in place to control what data is gathered or entered through several channels, including comments fields and document uploads. Technical solutions to scan for personal data may be appropriate as part of that overall governance for a blockchain solution.

Moreover, architects would need to determine that a blockchain doesn't accidentally hold unprotected personal data. This concern especially arises in scenarios in which users don't directly interact with a blockchain, but rather indirectly interact with it through application programming interfaces (APIs). For example, users may accidentally upload scanned copies of original paper documents in PDF or image formats, which then are stored on the blockchain. These documents, in turn, may contain personal data. It may not be possible to selectively redact personal information from such documents as they are loaded to the blockchain, raising concerns from a GDPR point of view. In this scenario, documentation should be provided to inform users how the APIs will store data on the blockchain.

Conclusions

Blockchain represents an important new kind of network infrastructure with several valuable benefits, including a ledger that's distributed, immutable and permanent. However, these benefits also pose challenges to the principles of the GDPR.

If at all possible, it's preferable to avoid storing personal data on a blockchain. Where such storage does occur, anonymisation or encryption of personal data may not be a sufficient method to manage privacy over an indefinite period. Instead, pseudonymisation, coupled with a robust business governance framework, is recommended to sustainably meet GDPR requirements. If in doubt, it's recommended to engage legal counsel before storing personal data on a blockchain.

The GDPR challenges discussed in this paper aren't the only ones. The GDPR undoubtedly will also affect other aspects of blockchain solutions and regardless, will require solution architects to account for additional considerations around the use and protection of data and individuals' rights. Questions that require consideration, but, which we didn't discuss in this paper, include:

- Who is responsible for informing individuals about data processing involving their data?
- How is consent gathered from individuals, and how is it stored?
- What's the lawful basis for processing data on a blockchain network?
- In a blockchain network, which entities are considered controllers and which are processors?
- How should you prevent participants in the blockchain from including personal data in their entries?
- Do blockchain networks with nodes outside the EU require any special transfer rules or procedures?

Answering these and other questions is likely to be crucial in implementing a new blockchain project that involves EU personal data.

Technical footnote on Hyperledger Fabric

Hyperledger Fabric[5] is an open source software for creating "permissioned" blockchain networks. The blockchain network is called permissioned because admission of a new entity (node) in the network, as well as subsequent operations in the network, conform to an agreed-upon policy by existing members. When a new entity is added in the network, it's assigned digital identities, which it can then use to interact in the network. Presently, these identities are managed through X.509 certificates, which can contain multiple attributes. Depending on how the entity was added, these attributes can contain personal data. See Figure 1. To verify a transaction, the X.509 certificates of an entity issuing a transaction in Hyperledger Fabric are permanently stored in the ledger along with the transaction data. Obviously, all attributes within a certificate are also stored in the ledger. With version 1.2, Hyperledger Fabric allows the use of a new credential based on zero-knowledge proofs.[6] These credentials are derived from the X.509 certificate and allow selectively disclosing attributes or prove only predicates about their attributes without their actual value. These credentials are then stored as part of a transaction instead of an X.509 certificate. Architects need to be aware of these two approaches and the trade-off involved in using these choices.

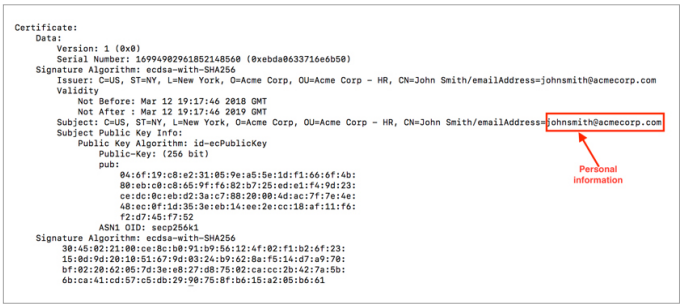


Figure 1: Example of an X.509 certificate

Hyperledger Fabric provides a way for organisations to interact through so-called "channels". Only nodes that participate in a channel can read or write information associated with this channel. This information isn't replicated to peers who don't participate in the channel. Further, through channel private data, only hashes of transaction data are exposed to the orderers.[7] Moreover, the transaction information can be encrypted; only users in possession of a key can decrypt such information.

For more information

To learn about ways IBM can help you address your GDPR obligations, please visit the following websites:

For more information on GDPR, see **ibm.com**/gdpr.

For more information on blockchain, see **ibm.com**/blockchain.

To learn how IBM® Security can help support your GDPR journey, see **ibm.biz**/PrepareForGDPR.

To assess the progress of your GDPR journey, see **ibm.biz**/GDPR-Ready.

Acknowledgements

The authors will like to thank Bertrand Portier, Maurizio Luinetti, Simon McDougall, Christel Lefort, Robert Hartley and other IBM colleagues for their contributions and input.

About the authors

Salman A. Baset (@salman_baset) is Chief Technology Officer (CTO) Security for IBM Blockchain Solutions. In this role, he oversees the security, compliance and identity management of blockchain solutions.

Graham Olsen is a business architect with IBM Global Business Services® (GBS) in the United Kingdom.

Joe Oehmke (@JoeOehmke) is a Principal at Promontory Financial Group, LLC, an IBM Company. His advisory work includes strategic and regulatory considerations related to blockchain and the GDPR.

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
August 2018

1 Blockchain and GDPR, https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=61014461USEN

2 GDPR Recital 26

3 GDPR Art. 4(5)

4 See GDPR Recital 26 (“Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means **reasonably likely** to be used…” (emphasis added).

5 Hyperledger Fabric, http://hyperledger-fabric.readthedocs.io

6 Hyperledger Fabric, MSP Implementation with Identity Mixer, https://hyperledger-fabric.readthedocs.io/en/release-1.2/idemix.html

7 https://hyperledger-fabric.readthedocs.io/en/release-1.2/private-data/private-data.html

Please Recycle

84018584-GBEN-00