How Blockchains May Comply With GDPR Mandates: Part 1

By **Kennedy Luvai** (May 2, 2018)

Kennedy Luvai

The May 25, 2018, enforcement date of the European Union's General Data Protection Regulation is fast approaching. All companies that process the personal data of EU citizens must comply with GDPR requirements or risk significant financial penalties — fines of up to four percent of annual global turnover or €20 million (whichever is greater). Ultimately, the GDPR shifts balance of power in favor of individuals and against organizations that collect or use such data for commercial purposes.

Blockchain: A Brief Primer

In essence, a blockchain is a distributed ledger of shared digital records saved in concatenated blocks and spread across multiple nodes.[1] The blocks comprise cumulative records which are interconnected so that each subsequent block contains a cryptographic hash or signature of the previous block. New blocks are created using a consensus protocol that solves the so-called Byzantine Generals' Problem — a thought experiment illustrating the difficulty of ensuring nodes in a distributed computer network are in agreement. The consensus protocol serves as a built-in trust-enhancing mechanism that secures the integrity of the data to be included in the new block. This linking process secures the blockchain against manipulation or hacking.

While blockchain technology initially garnered public attention in the context of cryptocurrencies — notably boitcoin — the nature and scope of blockchain applications now extends well beyond virtual currencies to include applications in areas such as digital identity, supply chain management, smart contracting, efficient settlement of market transactions and decentralized messaging, to name a few. The built-in trust mechanism in a functional blockchain reduces or eliminates the need for trusted third parties or intermediaries to validate transactions and, therefore, has the potential for increasing efficiencies and reducing costs associated with transactions.

Personal Data Under the GDPR

The GDPR was enacted against a backdrop of a siloed, centralized data storage framework. Blockchains, by design, go beyond decentralizing data storage — they distribute data across multiple nodes. The result is a functionally redundant network that lacks a single (centralized) point of failure or control.

The GDPR strengthens privacy rights of EU citizens by increasing control over personal data. The concept of personally identifying information is key to the GDPR, which embraces a fairly broad definition of personal data — "any information relating to an identified or identifiable natural person ('data subject')."[2] Data that qualifies as personal data can only be processed subject to the GDPR's conferral of data subjects with specific substantive rights to their personal data.[3] The GDPR does not apply to data that "does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."[4]

The GDPR also introduces the new concept of pseudonymization: "[T]he processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information."[5] To pseudonymize data, the "additional information" must be "kept separately and ... subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable person."[6] While the GDPR recognizes that pseudonymization "can reduce risks to data subjects concerned and help controllers and processors meet their data-protection obligations,"[7] pseudonymization, alone does not exempt data from the GDPR's regulatory reach.

Alternately, anonymized data is not subject to the GDPR. Recital 26 of the GDPR defines anonymized data as "data rendered anonymous in such a way that the data subject is not or no longer identifiable." This definition emphasizes that anonymized data must be stripped of any identifiable information, making it impossible to derive insights on a discreet individual, even by the party that is responsible for the anonymization. When done properly, anonymization places the processing and storage of personal data outside the scope of the GDPR. The Article 29 Working Party has made it clear, however, that true data anonymization is an extremely high bar, and data controllers often fall short of actually anonymizing data.

The Implications of the Erasure Mandate on Blockchain Solutions

Under Recital 26 of the GDPR, "[p]ersonal data which have undergone pseudonymisation, which could be attributed to a natural person by use of additional information should be considered to be information on an identifiable natural person." Accordingly, it is safe to state that personal data that is hashed or encrypted to a blockchain may still qualify as personal data subject to the GDPR as such techniques fall short of irretrievably anonymizing the personal data.[8]

Hashing refers to the concept of taking an arbitrary amount of input data, applying some algorithm to it to generate a fixed-size output data called the hash.[9] It is unfeasible (though technically possible) to determine the original input from its hash value. The only means currently available to find an original input is the "brute force method," which involves picking a random input, hashing it, comparing the output with the target hash value and repeating the process until a match is found. While possible in theory, it is unfeasible in cases where the size of the data creates an inordinate number of possible combinations. Nevertheless, because of the possible "linkability" of the unreadable hash to the original input, hashing is deemed to be pseudonymization of data thus subjecting the unreadable hash to the GDPR.

Encryption refers to the concept of taking input data and translating it into an unreadable form so that only those with decryption keys can decipher the text back to readable plain text. Encryption is a two-way function. With the right decryption key, the unreadable encrypted data may be deciphered and reverted to its original readable state. Because the decryption key deciphers unreadable encrypted text, the decryption key counts as "additional information" linking personal data to the data subject. Encryption pseudonymizes data but does not anonymize it under the GDPR. Thus, the GDPR regulates encrypted personal data saved to a blockchain.

The treatment of hashed or encrypted data on a blockchain as pseudonymous data, subject to the GDPR, has far-reaching implications, including the right to erasure, the right to data portability and the right to rectification. The first part of this two-part article is limited to a discussion of the erasure mandate.

Article 17 of the GDPR provides that data controllers must erase personal data “without undue delay” if the data is no longer needed, the data subject objects to the processing or the processing was unlawful. While many properties of blockchains actually promote goals of GDPR (e.g., availability, transparency and integrity of stored data), the right to erasure presents a notable challenge to blockchain solutions.

Selected Approaches That May Address the Erasure Mandate

Several basic approaches aimed at implementing blockchain solutions compliant with the right to erasure have been explored. Indisputably, the safest way to avoid GDPR compliance issues with data stored to a blockchain requires ensuring that no personal data is stored to the blockchain or properly anonymizing any personal data stored to the blockchain. If these options are unavailable or unfeasible, the following are potentially available, general approaches:[10] (1) editing an otherwise “immutable” blockchain to remove personal data; (2) physical deletion of personal data stored in an off-chain database but with retention of unreadable hashed references to that data on a blockchain, and (3) logical deletion of encrypted personal data on a blockchain through the permanent destruction of decryption keys, rendering encrypted data undecipherable.

Editing an “Immutable” Blockchain

Accenture PLC is advancing an option that may address erasure issues in permissioned blockchains, which involves modifying an existing blockchain to allow designated authorities to edit, rewrite or remove previous blocks of information without breaking the chain. This solution could be used to edit an otherwise immutable blockchain to remove personal data at the request of a data subject. The option enables blockchain editing by using a new variation of the “chameleon” hash function. The chameleon hash key is used to unlock the link between the block to be edited and its successor, allowing substitution of the block to be edited with a new block without breaking the hash chain.

This solution does have potential drawbacks. First, it requires the introduction of a trusted third party to manage the editing process, which may undercut the benefits of distributed computing upon which blockchains are built. Second, it is unclear how this approach will scale to accommodate large numbers of requests that may impact many, if not all, blocks on the chain. Third, with the loss or destruction of the chameleon hash key, the blockchain again becomes immutable.

Logical Deletion of Unreadable Hashed Data

In some instances, personal data may be stored in a traditional database with a cryptographic hash reference to that external database being stored on a blockchain. Only an unreadable hashed output of the personal data would appear on the blockchain. This technique allows the blockchain to maintain its integrity or immutability while simultaneously storing information which cannot be easily linked back to the deleted personal data.

Some argue that removing the personal data from the traditional database renders the unreadable hash reference to that data on the blockchain functionally defunct. In this instance, the blockchain has not been altered or edited. However, as discussed above, hashing of input data results in pseudonymized data to which hashed output data may be “linked” through a computationally intensive (and likely economically unfeasible) brute force operation.

Logical Deletion of Unreadable Encrypted Blockchain Data

Logical deletion in the context of blockchain may be accomplished by permanently destroying decryption keys to make the associated encrypted data unreadable without expending an inordinate amount of time, effort or money to decipher the data. Because the permanent destruction of such keys would make the content practically unreadable, it can be contended that such a permanent key destruction should result in the logical erasure of data for purposes of the GDPR's right to erasure provision. The encrypted personal data would remain on-chain but would not be accessible at all given the private key destruction.

Pseudonymization of Blockchain Data May, in Certain Cases, Satisfy the Erasure Mandate

The question of whether the hashing or encrypting of transaction data on a blockchain would satisfy the erasure mandate under the GDPR depends on the extent to which the pseudonymized data can be reidentified. This issue is of critical importance, particularly given the fact that "erasure" is not defined under the GDPR, leaving open interpretations of "erasure" that extend beyond the "traditional" physical deletion of data.

In assessing whether pseudonymized data can be reidentified, the GDPR focuses on whether the reidentification is "reasonably likely." Recital 26 of the GDPR provides that "to determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly." The recital goes on to state that "to ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."

Accordingly, the GDPR provides a measure of flexibility with regard to what counts as a rendering of pseudonymized data such that it cannot be reidentified, making it functionally anonymized or erased. The GDPR does not foreclose the use of appropriate hashing or encryption techniques that would require an inordinate amount of time and costs to decipher the hashed or encrypted data. Obviously, whether such techniques entail such expenditures of time and money will depend on the technological innovations available at the time.

With the GDPR yet to come into force, it remains to be seen how the various data protection authorities in the EU member states will assess whether pseudonymization of data in particular instances satisfies the erasure mandate. Prompt guidance by relevant authorities will be required to address uncertainties inherent in the application of the erasure mandate to blockchain solutions.

Kennedy K. Luvai is a shareholder at Parsons Behle & Latimer PLC in Salt Lake City, and a former software developer.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] A node is computer connected to a blockchain network and which is capable of performing the task of validating and relaying transactions.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council (hereinafter "GDPR") Art 4.

[3] These rights include the right to access of data by the data subject, the right to rectification, the right to restrict processing, the right to data portability, among others.

[4] GDPR Recital 26.

[5] GDPR Art. 4.

[6] GDPR Art 4.

[7] GDPR Art. 28. Also, a 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 while a processor is defined as the natural or legal person, public authority or other body which processes personal data on behalf of the controller. GDPR Art. 4.

[8] This discussion is limited to transactional data to the extent that it contains personal data, and does not relate to public keys on a blockchain which comprise a string of numbers and letters that allows for the pseudonymous identification of a natural and legal person.

[9] Many blockchain applications use the hashing algorithm SHA-256 (Secure Hashing Algorithm 256) that was designed by the United States National Security Agency(NSA). Any changes to the input, no matter how small, will result in a completely different output hash that is easily detectable.

[10] This list is being provided for discussion purposes and is by no means intended to be exhaustive. Other potential avenues that may make blockchain solutions compliant with the erasure mandate under the GDPR may exist or are in the process of being developed.