# Blockchain: Disrupting data protection?

150 Privacy Laws & Business International Report, December 2017

Henry Chang,
Adjunct Associate Professor
Department of Law, The University of Hong Kong

There is much hype around blockchain and maybe not so much discussion about the security and privacy aspects. This article introduces the issues on how blockchain has exhibited characteristics that are challenging to data protection, and what steps are being taken to resolve them.

Bitcoin: The Technology

Bitcoin was invented right after the global financial crisis as a form of digital money that is free of control by intermediaries such as banks or clearing houses, with a low transaction cost and is anonymous to its users (rather like cash in this aspect).

One would need to join as a node in the Bitcoin network to start using Bitcoin. When a transaction of Bitcoin between two nodes takes place, the transaction is digitally signed and time stamped by those involved and then verified by other nodes in the Bitcoin network. Once the transaction is verified, it is stored in the form of a "block" of storage in the transaction ledger. When the block gets to a certain size, it is closed and a new consecutive block is created to hold newer transaction records.

When a block is closed, a hash (which may be considered as a sophisticated checksum) is calculated and stored in the next block. This additional step ensures that even when one manages to tamper with transactions in previously stored blocks protected by digital signatures, any such change of the data will be immediately discovered by comparing against the hash previously stored in the next block.

If one manages also to change the hash previously stored in the next block, given the hash value is part of the block too, such change will then be detected by the hash previously generated for this block which is stored in the block that follows.

Finally, each closed block is copied to all other nodes in the Bitcoin network where every node compares its own version of the ledger with each other. If any discrepancy is found, the ledgers of the minority will have to follow the ledgers of the majority resulting in any such anomaly or rogue records to be overwritten.

*See Whitepaper on Distributed Ledger Technology. Hong Kong: The Hong Kong Monetary Authority; 2016. p.9-36. Available from: www.hkma.gov.hk/media/eng/doc/ key-functions/finanicalinfrastructure/Whitepaper_On_Distributed_ Ledger_Technology.pdf*

You may have heard frequent talk about Bitcoin, blockchain and distributed ledger technology (DLT), and how they may – like cloud computing a few years back – be the next disruptive technology that would transform how businesses operate. Interestingly enough, similar to cloud computing, while the technology creates many opportunities, its design has

certain features that are inherently unfriendly to data protection. Apart from explaining the technology (see inset), this article also picks out some data protection issues for discussion:

**Blockchain or Distributed Ledger Technology?** DLT is a generic description of a distributed database that often holds transaction records (i.e. a ledger). Blockchain is a specific kind of DLT that utilises blocks and chains to store records and to ensure their integrity. Bitcoin is a specific cryptocurrency application based on blockchain. In other words, Bitcoin is a sub-set of blockchain, which is a subset of DLT. However, due to the success of Bitcoin, its characteristics are often mistaken as also representative of blockchain and DLT.

**Anonymity? =** Anonymity is an option and not an inseparable feature in blockchain or DLT. Bitcoin is implemented as an "unpermissioned" blockchain where anyone can join as a public node without the need for identification and verification. However, Blockchain or DLT applications can also be designed and implemented as "permissioned", where proper governance is in place to validate users' identities before they are allowed to join as a node (See 'Bitcoin: The Technology')

**Security and transparency**? Security often includes three elements: confidentiality, integrity and availability. Unpermissioned blockchain, including Bitcoin, allows anyone – as a node or not – to access the ledger so there is no confidentiality protection. It even creates a disclosure problem if personal data is involved. In the cases of permissioned blockchain or DLT, whether the ledger is allowed to be read by other nodes may depend on the design. This is because data in the permissioned ledger may be encrypted or the ledger (or parts of the ledger) may not need to be copied to all other nodes, but only to a handful of trustworthy nodes for validation, backup or resilience. Finally and in terms of security, the attack on the "weakest link" always the easiest and in the past, hacking of Bitcoin and the like was possible at the "end-point" – be they the Bitcoin exchange, the eWallet or the person holding the password.

**Immutable?** Immutability is perhaps the best core feature of blockchain and DLT. Altering or removing data inside a block would simply render all the data inside the block no longer trustworthy. While from a record management point of view it is a great feature, it creates issues when the legal basis for the original collection of personal data is no longer valid and the data needs to be removed, or when a data subject exercises the right to erasure under the GDPR in the future.

**Undecryptable = erased?** Some technologists have come up with the idea that the personal data stored in blockchain or DLT can be encrypted and when the time comes for their erasure, all that needs to be done is to erase the decryption key.[1] This idea seems to be based on the notions that encrypted personal data is not personal data, and that encrypted data cannot be decrypted without the key. Naturally both assumptions are fraught with problems[2] and cannot be accepted when technology, including quantum computing that can crack some encryptions in a much-quicker way, is developing rapidly.

[1] Piscini E, Dalton D, Kehoe L. Blockchain & Cyber Security. Let's Discuss. Dublin: Deloitte; 2017. P.7. www2.deloitte.com/content/dam/Delo itte/ie/Documents/Technology/IE_C_ BlockchainandCyberPOV_0417.pdf.

[2] Opinion 05/2012 on Cloud Computing (WP196). Brussels: Article 29 Data Protection Working Party; 2012. ec.europa.eu/justice/dataprotection/article- 29/documentation/opinionrecommendation/files/2012/wp196_e n.pdf

**Redactable blockchain?** Usually when a block in a blockchain is altered, the hash will change. A vendor, however, has created a redactable or editable blockchain technology by weakening the hash mechanism so that the edited block keeps the same hash. This weakened hash mechanism (called chameleon hash) allows an algorithm to find ways to add back some information to the block to "compensate" for the editing/deletion so that the resulting hash stays the same. The vendor promises that a chameleon hash can only be re-calculated under strict governance control, and a marker will be put in the block to say that the block has been edited.3[3] While this solution is technologically innovative, one has to ponder where the value of blockchain lies when it becomes editable.

**Hybrid solution?** Blockchain and DLT were originally designed to record transactions which usually contain only small amounts of data. As such, for any other non-financial-transaction applications that involve storing large amount of data, including the growing interests of using them to process medical data, an external database may well be needed to meet these needs. In such a hybrid system, data can be stored in an external database, which may also be distributed to all or some of the nodes, while hashes of the data are stored in the blockchain or DLT. This arrangement allows for the best of both worlds. Actual data is stored in a traditional database where there are a number of granular controls, such as controls on record-level access based on the need-to-know principle, duplication of selective records to other designated nodes for trustworthy back-up, and deletion of individual records upon the expiry of the legal basis and/or upon requests from data subjects. At the same time, the immutable hashes that are retained in the blockchain or DLT ensure any alteration to the data stored in the external database is easily detected.

**Searchable?** Blockchain and DLT were designed as a ledger to store transaction records so it is easier to trace records (such as tracing back the previous owners of a particular piece of Bitcoin) but not for conducting query-type searches (such as to find out the number of transactions exceeding a certain amount). For general searching, an index of the data stored in blockchain or DLT will have to be kept in an external database so that it can be searched. With this being the case, a hybrid model could be all the more appropriate in many types of applications.

**Cross-border?** Blockchain and DLT are good for applications spanning across multiple organisations so inevitably nodes may reside in different jurisdictions. With cross-border restrictions and the uncertainty on where the next node may reside, it would be impracticable to obtain meaningful consents from all data subjects to store personal data in foreign jurisdictions. The hybrid model again provides the flexibility for personal data to be stored externally in a central, distributed or partiallydistributed database where each piece of personal data may be selectively copied to specific nodes that are located in jurisdictions with comparable data protection laws.

**Who is in charge?** In the cases of unpermissioned blockchain or DLT such as Bitcoin or Decentralised Autonomous Organisation where participants join as equal parties of the network or the organisation, they may potentially share the liability of such network or organisation as data controllers. This is somewhat uncharted territory that requires some in-depth studies. In the case of permissioned blockchain or DLT where there is a governance structure, depending on precise circumstances, both the entity that collects personal data or

[3] *Financial Times* 'Accenture to unveil blockchain editing technique' 2016. www.ft.com/content/f5cd6754-7e83- 11e6-8e50-8ec15fb462f4?mhq5j=e6

the governance body may be considered as the data controller. Even in cases where an entity other than the governance body is considered as the data controller, the blockchain or DLT may well be considered as a data processor sharing some responsibilities. To complicate the matter, if certain participants are involved as validation nodes that have to process personal data as a result, there is a possibility that they find themselves unknowingly caught under data protection laws. [4]

**Any other consideration?** This article has only discussed a few data protection concerns that are specific to the use of blockchain and DLT. There are other regulatory and legal issues associated with the use of blockchain and DLT, such as governance and control, liability, legal basis, applicable jurisdictions, settlement finality, smart contracts, etc. that would need to be handled but are beyond the scope of this article.

[4] *Financial Times* 'Accenture to unveil blockchain editing technique' 2016. www.ft.com/content/f5cd6754-7e83- 11e6-8e50-8ec15fb462f4?mhq5j=e6

4 Zetzsche D, Buckley R and Arner D. The Distributed Liability of Distributed Ledgers: Legal Risks of Blockchain. 2017. ssrn.com/abstract=3018214