NetIKX July 2017 seminar on blockchain and information management

Report and discussion by Conrad Taylor

Blockchain is a technology first developed as the technical basis for the cryptocurrency Bitcoin, but there has been recent speculation that it might be useful for various information management purposes too. There is quite a 'buzz' around the topic, yet it is too complex for many people to figure out, so it's not surprising that the 6 July 2017 NetIKX seminar, 'The implications of Blockchain for KM and IM', attracted the biggest turnout of the year so far.

The seminar took the form of three presentations, two from the consultancy **Metataxis** and one from **The National Archive**. The table group discussions which followed were simply open and unstructured discussions, with a brief period at the end for sharing ideas.

The subject was indeed complex and a lot to take in. In creating this report I have gone beyond what we were told on the day, done some extra research, and added my own observations. I hope this will make some things clearer, and qualify some of what our speakers said, especially where it comes to technical details.

Marc Stephenson gives a technical overview

The first speaker was Marc Stephenson, Technical Director at Metataxis, the information architecture and information management consultancy. In the limited time available, Marc attempted a technical briefing.

Marc's first point was that it's not easy to define blockchain. It is not just a technology, but also a concept and a framework for ways of working with records and information; and it has a number of implementations, which differ in significant ways from each other. Marc suggested that, paradoxically, blockchain can be described as 'powerful and simple', but also 'subtle, and difficult to understand'. Even with two technical degrees under his belt, Marc confessed it had taken him a while to get his head around it. I sympathise!

The first, the largest and the best-known implementation of blockchain so far is the infrastructure for the digital cryptocurrency '**Bitcoin**' – so much so that many people get the two confused (and others, in my experience, think that all of the features of Bitcoin are essential to blockchain – I shall be suggesting otherwise).

About NetIKX

The Network for Information and Knowledge Exchange is an independent community of interest with a focus on practical issues of managing knowledge and information in the workplace. It holds six meetings a year in London. For information and details of membership see the Web site **www.netikx.org**.

NetIKX also maintains a blog at **https://netikx.wordpress.com**, and there is a LinkedIn discussion group.

Wikipedia (at https://en.wikipedia.org/wiki/Blockchain) offers this definition:

> A blockchain […] is a distributed database that maintains a continuously growing list of ordered records called blocks. Each block contains a timestamp and a link to a previous block. By design, blockchains are inherently resistant to modification of the data — once recorded, the data in a block cannot be altered retroactively. Through the use of a peer-to-peer network and a distributed timestamping server, a blockchain database is managed autonomously… [A blockchain is] an open, distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way. The ledger itself can also be programmed to trigger transactions automatically.

Marc then dug further into this definition, but in a way which left some confused about what is specific to Bitcoin and what are the more generic aspects of blockchain. Here, I have tried to tease these apart.

Distributed database — Marc said that a blockchain is intended to be a massively distributed database, so there may be many complete copies of the blockchain data file on server computers in many organisations, in many countries. The intention is to avoid the situation in which users of the system have to trust a single authority.

I am sceptical as to whether blockchains necessarily require this characteristic of distribution over a peer-to-peer network, but I can see that it is valuable where there are serious issues of trust at stake. As we heard later from The National Archive, it is possible to create similar distributed ledger systems shared between a smaller number of parties which already trust each other.

Continuously growing chain of unalterable 'blocks' — The blockchain database file is a sequential chain divided into 'blocks' of data. Indeed, when blockchain was first described by 'Satoshi Nakamoto', the pseudonymous creator of the system in 2008, the phrase '*block chain*' was presented as two separate words. When the database is updated by a new transaction, no part of the existing data structure is overwritten. Instead, a new data block describing the change or changes (in the case of Bitcoin, a bundle of transactions) is appended to the end of the chain, with a link that points back to the penultimate (previous) block;

which points back to the previous one; and so on back to the 'genesis block'.

One consequence of this data structure is that a very active blockchain that's being modified all the time grows and grows, potentially to monstrous proportions. The blockchain database file that maintains Bitcoin has now grown to 122 gigabytes! Remember, this file doesn't live on one centralised server, but is duplicated many times across a peer-to-peer network. Therefore, a negative consequence of blockchain could be the enormous expense of computing hardware resources and energy involved in a blockchain system.

(As I shall later explain, there are some peculiar features of Bitcoin which drive its bloat and its massive use of computational resources; for blockchains in general, it ain't necessarily so.)

Timestamping — when a new block is created at the end of a chain, it receives a timestamp. The Bitcoin 'timestamp server' is not a single machine, but a distributed function.

Encryption — According to Marc, all the data in a blockchain is encrypted. More accurately, in a cryptocurrency system, crucial parts of the transaction data do get encrypted, so although the contents of the blocks are a matter of public record, it is impossible to work out who was transferring value to whom. (It is also possible to implement a blockchain without any encryption of the main data content.)

Managed autonomously — For Bitcoin, and other cryptocurrencies, the management of the database is done by distributed software, so there is no single entity, person, organisation or country in control.

Verifiable blocks — It's important to the blockchain concept that all the blocks in the chain can be verified by anyone. For Bitcoin, this record is accessible at the site bitcoin.info.

Automatically actionable — In some blockchain systems, blocks may contain more than data; at a minimum they can trigger transfers of value between participants, and there are some implementations – **Ethereum** being a notable example – which can be programmed to 'do' stuff when a certain condition has been met. Because this happens without user control, without mediation, all of the actors can trust the system.

Digging into detail

In this section, I am adding more detail from my own reading around the subject. I find it easiest to start with Bitcoin as the key example of a blockchain, then explore how other implementations vary from it.

'Satoshi Nakamoto' created blockchain in the first place to implement Bitcoin as a digital means to hold and exchange value – a currency. And exchange-value is a very simple thing to record, really, whereas using a blockchain to record more complex things such as legal contracts or medical records adds extra problems – I'll look at that later. Let's start by explaining Bitcoin.

Alice wants to pay Bob. Alice 'owns' five bitcoins – or to put it more accurately, the Bitcoin transaction record verifies that she has an entitlement to that amount of bitcoin value: the 'coins' do not have any physical existence. She might have purchased them online with her credit card, from a Bitcoin broker company such as **eToro**. Now, she wants to transfer some bitcoin value to Bob, who in this story is providing her with something for which he wants payment, and has emailed her an invoice to the value of 1.23 BTC. The invoice contains a 'Bitcoin address' – a single-use identifier token, usually a string of 34 alphanumeric characters, representing the destination of the payment.

To initiate this payment, she needs some software called a 'Bitcoin wallet'. Examples are **breadwallet** for the iPhone and iPad, or **Armory** for Mac, Linux and Windows computers. There are also online wallets. Users may think, 'the wallet is where I store my bitcoins'. More accurately, the wallet stores the digital credentials you need to access the bitcoin values registered in the blockchain ledger against your anonymised identity.

Launching her wallet, Alice enters the amount she wants to send, plus the Bitcoin address provided by Bob, and presses Send.

For security, Alice's wallet uses **public key cryptography** (see below) to append a scrambled digital signature to the resulting message. By keeping her private key secret, Alice is guaranteed that no-one can spoof Bitcoin into thinking that the message was sent to the system by anyone else other than her. The Bitcoin messaging system records neither Alice's nor Bob's identity in the data record, other than in deeply encrypted form: an aspect of Bitcoin which has been criticised for its ability to mask criminally-inspired transactions.

At this stage, Alice is initiating no more than a proposal, namely that the Bitcoin blockchain should be altered to show her wallet as that bit 'emptier', and Bob's a bit 'fuller'. Implementing computers on the network will check to see whether Alice's digital signature can be verified with her public key, that the address provided by Bob is valid, and

Public key cryptography

In public key cryptography, messages are secured using a pair of keys – the public key which may be disseminated widely, and a private key known only to the owner (in this case, the owner of a Bitcoin account). These methods can be used either for **encryption** of messages using someone's public key, which only they can read using their private key, or for **authentication** (e.g. a 'digital signature') in which someone creates an authentication code using their private key, and the receiver can check using their public key to ensure it really came from them.

The algorithms used in public key cryptography rely on very difficult mathematical problems. For example the RSA method uses two randomly chosen very large prime numbers as the public and private keys. It's computationally relatively easy to multiply the primes, but nearly impossible to 'factorise' them.

that Alice's account does in fact have enough bitcoin value to support the transaction.

If Alice's bitcoin transaction proposal is found to be valid and respectable, the transaction can be enacted, by modifying the blockchain database (updating the ledger, if you like). As Marc pointed out, this is done not by changing what is there already, but by adding a new block to the end of the chain. Multiple transactions get bundled together into one Bitcoin block, and the process is dynamically managed by the Bitcoin server network to permit the generation of just one new such block approximately every ten minutes – for peculiar reasons I shall later explain.

Making a block: the role of the 'hash'

The blocks are generated by special participating servers in the Bitcoin network, which are called '**miners**' because they get automatically rewarded for the work they do by having some new Bitcoin value allocated to them.

In the process of making a block to add to the Bitcoin blockchain, the first step is to gather up the pending transaction records, which are placed into the body of the new block. These transaction records themselves are not encrypted, though the identities of senders and receivers are. I have heard people say that the whole blockchain is irreversably encrypted, but if you think about it for a second, this has to be nonsense. If the records were rendered uninspectable, the blockchain would be useless as a record-keeping system!

However, the block as a whole, and beyond that the blockchain, has to be protected from accidental or malicious alteration. To do this, the transaction data is put through a process called 'cryptographic hashing'. Hashing is a well-established computing process which feeds an arbitrarily large amount of data (the 'input' or 'message') through a precisely defined algorithmic process, which reduces it down to a fixed-length string of digits (the 'hash'). The hashing algorithm used by Bitcoin is SHA-256, created by the US National Security Agency and put into the public domain.

By way of example, I used the facility at **http://passwordsgenerator.net/sha256-hash-generator/** to make an SHA-256 hash of everything in this article up to the end of the last paragraph (in a previous edits, I should add; I've made changes since). I got 9F0B 653D 4E6E 7323 4E03 B04C F246 4517 8A96 DFF1 7AA1 DA1B F146 6E1D 27B0 CA75 (you can ignore the spaces).

The hash string looks kind of random, but it isn't – it's 'deterministic'. Applying the same hashing algorithm to the same data input will always result in the same hash output. But, if the input data were to be modified by even a single character or byte, the resulting hash would come out markedly different.

Note that the hash function is, for all practical purposes, 'one-way'. That is, going from data to hash is easy, but processing the hash back into the data is impossible: in the case of the example I just provided, so much data has been discarded in the hashing process that no-one receiving just the hash can *ever* reconstitute the data. It is also theoretically possible, because of the data-winnowing process, that another set of data subjected to the same hashing algorithm could output the same hash, but this is an extremely unlikely occurrence. In the language of Bitcoin, the hashing process is described as '**collision-resistant**'.

The sole purpose of this hashing process is to build a kind of internal certificate, which gets written into a special part of the block called the 'header'. Here, cryptography is not being used to hide the transaction data, as it might in secret messaging, but to provide a guarantee that the data has not been tampered with.

Joining the hash of the transaction data in the header are some other data, including the current timestamp, and a hash of the header of the preceding block in the chain. These additions are what gives the blockchain its inherent history, for the preceding block also contained a hash of the header of the block before that, and so on down the line to the very first block ever made.

The role of the 'miner' in the Bitcoin system

Now, as far as I can tell, there is nothing in principle wrong with having the blockchain-building process run by one trusted computer, with the refreshed blockchain perhaps being broadcast out at intervals and stored redundantly on several servers as a protection against disaster.

But that's not the way that Bitcoin chose to do things. They wanted the block-writing process to be done in a radically decentralised way, by servers competing against each other on a peer-to-peer network; they also chose to force these competing servers to solve tough puzzles which are computationally very expensive to process. Why? Because intimately entangled in the way the Bitcoin ecology builds blocks, is the way that new bitcoins are minted; at present the 'reward' from the system to a miner-machine for successfully solving the puzzle and making the latest block in the chain is a fee of 12.5 fresh new bitcoins, worth thousands of dollars at current exchange rates. That's what motivates private companies to invest in mining hardware, and take part in the game.

This reward-for-work scheme is why the specialised computers that participate in the block-building competition are called 'miners'.

Let's assume that the miner has got as far through the process as verifying and bundling the transaction data, and has created the hash of the data for the header. At this point the Bitcoin system cooks up a mathematical puzzle based on the hash, which the 'miner' system making the block has to solve. These mathematical puzzles (and I cannot enlighten you more about their precise nature, it's beyond me!) can be solved only by trial and error methods. Across the network, the competing miner servers are grinding away, trying trillions of possible answers, hashing the answers and comparing them to the header hash and the puzzle instructions to see if they've got a match.

This consumes a lot of computing power and energy – in 2014, one bitcoin 'mining farm' operator, Megabigpower in

Washington State, USA, estimated that it was costing 240 kilowatt-hours of electricity per bitcoin earned, the equivalent of 16 gallons of petrol. It's doubtless gone up by now. The hashing power of the machines in the Bitcoin network has surpassed the combined might of the world's 500 fastest supercomputers! (See 'What is the Carbon Footprint of a Bitcoin?' by Danny Bradbury: https://www.coindesk.com/carbon-footprint-bitcoin/).

When a miner 'thinks' it has a correct solution, it broadcasts to the rest of the network and asks other servers to check the result (and thanks to the hash-function check, though solving the problem is hard, checking the result is easy). All the servers which 'approve' the solution – strangely, it's called a 'nonce' – will accept the proposed block, now timestamped and with a hash of the previous block's header included to form the chainlink, and they update their local record of the blockchain accordingly. The successful miner is rewarded with a transaction which earns it a Block Reward, and I think collects some user transaction fees as well.

Because Bitcoin is decentralised, there's always the possibility that servers will fall out of step, which can cause temporary forks and mismatches at the most recent end of the blockchain, across the network ('loose ends', you might call them). However, the way that each block links to the previous one, plus the timestamping, plus the rule that each node in the network must work with the longest extant version it can find, means that these discrepancies are self-repairing, and the data store is harmonised automatically even though there is no central enforcing agency.

The Bitcoin puzzle-allocation system dynamically adjusts the complexity of the puzzles so that they are being solved globally at a rate of about only six an hour. Thus although there is a kind of 'arms race' between competing miners, running on ever faster competing platforms, the puzzles just keep on getting tougher and tougher to crack, and this is what controls the slow increase in the Bitcoin 'money supply'. Added to this is a process by which the rate of reward for proof-of-work is being slowly decreased over time, which in theory should make bitcoins increasingly valuable, rewarding the people who own them.

As I shall shortly explain, this computationally very expensive 'proof-of-work' system is not a necessary feature of blockchain *per se*, and other blockchains use a less expensive 'proof-of-stake' system to allocate work.

Disentangling blockchain from Bitcoin

To sum up, in my opinion the essential characteristics of blockchain in general, rather than Bitcoin in particular, are as follows (and compare this with the Wikipedia extract quoted earlier):

- ◆ A blockchain is a data structure which acts as a consultable ledger for recording sequences of facts, statuses, actions or transactions which occur over time. So it is not a database in the sense that a library catalogue is; still less could it be the contents of that library; but the lending records of that library could well be in blockchain form, because they are transactions over time.

- ◆ New data, such as changes of status of persons or objects, are added by appending blocks of re-formed data; each block 'points' towards the previous one, and each block also gets a timestamp, so that together the blocks constitute a chain from oldest to newest.

- ◆ The valuable data in the blocks are not necessarily encrypted (contrary to what some people say), so that with the right software, the record is open to inspection.

- ◆ However, a fairly strong form of cryptographic hashing is applied to the data in each block, to generate a kind of internal digital certificate which acts as a guarantee that the data has not become corrupted or maliciously altered. The hash string thus generated is recorded in the head of the block; and the whole head of the block will be hashed and embedded in the head of the following block, meaning that any alteration to a block can be detected.

And I believe we can set aside the following features which are peculiarities of Bitcoin:

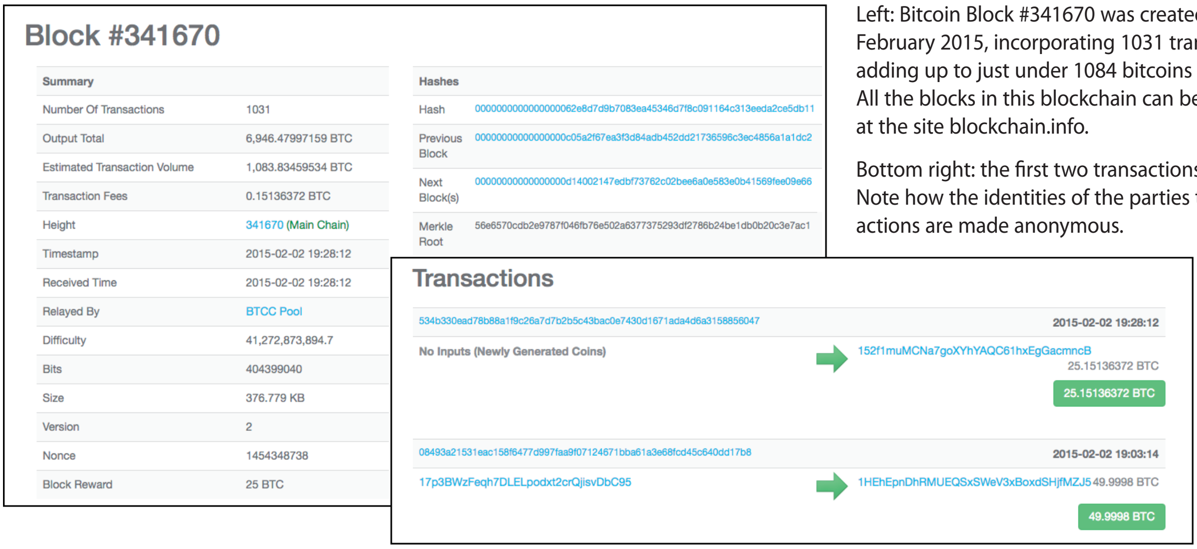
- ◆ The Bitcoin blockchain is a record of all the transactions which have ever taken place between all of the actors within the Bitcoin universe, which is why it is so giganormous (to coin a word). Blockchains which do not have to record value exchange transactions can be much smaller and non-global in scope – my personal medical record, for example, would need to journal only the experiences of one person.

- ◆ All the data tracked by the Bitcoin blockchain has to live inside the blockchain; but blockchain systems can also be hybridised by having them store secure and verified links to other data repositories. And that's a sensible design choice where the entire data bundle contains binary large objects (BLOBs) such as x-rays, scans of land title deeds, audio and video recordings, etc.

- ◆ The wasteful and computationally expensive 'proof of work' test faced by Bitcoin miners is, to my mind, totally unnecessary outside of that kind of cryptocurrency system, and is a burden on the planet.

Marc shows a block

In closing his presentation, Marc displayed a slide image of the beginning of the record of block number 341669 inside the Bitcoin blockchain, from back in February 2015 when the 'block reward' for solving a 'nonce' was 25 Bitcoins. You can follow this link to examine the whole block on bitcoin.info: https://blockchain.info/



Block #341670

Summary	
Number Of Transactions	1031
Output Total	6,946.47997159 BTC
Estimated Transaction Volume	1,083.83459534 BTC
Transaction Fees	0.15136372 BTC
Height	341670 (Main Chain)
Timestamp	2015-02-02 19:28:12
Received Time	2015-02-02 19:28:12
Relayed By	BTCC Pool
Difficulty	41,272,873,894.7
Bits	404399040
Size	376.779 KB
Version	2
Nonce	1454348738
Block Reward	25 BTC

Hashes	
Hash	0000000000000000062e8d7d9b7083ea45346d7f8c091164c313eeda2ce5db11
Previous Block	00000000000000000c05a2f67ea3f3d84adb452dd21736596c3ec4856a1a1dc2
Next Block(s)	00000000000000000d14002147edbf73762c02bee6a0e583e0b41569fee09e66
Merkle Root	56e6570cdb2e9787f046fb76e502a6377375293df2786b24be1db0b20c3e7ac1

Transactions

534b330ead78b88a1f9c26a7d7b2b5c43bac0e7430d1671ada4d6a3158856047 2015-02-02 19:28:12

No Inputs (Newly Generated Coins) → 152f1muMCNa7goXYhYAQC61hxEgGacmncB 25.15136372 BTC

25.15136372 BTC

08493a21531eac158f6477d997faa9f07124671bba61a3e68fcd45c640dd17b8 2015-02-02 19:03:14

17p3BWzFeqh7DLELpodxt2crQjisvDbC95 → 1HEhEpnDhRMUEQSxSWeV3xBoxdSHjfMZJ5 49.9998 BTC

49.9998 BTC

Left: Bitcoin Block #341670 was created on 2nd February 2015, incorporating 1031 transactions adding up to just under 1084 bitcoins of value. All the blocks in this blockchain can be inspected at the site blockchain.info.

Bottom right: the first two transactions in the block. Note how the identities of the parties to the transactions are made anonymous.

block/0000000000000000062e8d7d9b7083ea45346d7f8c-091164c313eeda2ce5db11 (ignore the hyphens). I've also put a couple of screen captures from that record at the top of this page.

That block carries records of 1,031 transactions, of a value of 1,084 BTC, and it is about 377 KB in size (and remember, these blocks add up!) The transaction record data can be clearly read, even thought it will not make much sense to human eyes because of the anonymisation provided by the encrypted user address of the sender, and the encrypted destination address for the receiver. Thus all we can see is that '17p3BWzFeqh7DLELpodxt2crQjisvDbC95' sent 50 BTC to '1HEhEpnDhRMUEQSxSWeV3xBoxdSHjfMZJ5'

Other cryptocurrencies, other blockchain methods

Bitcoin has had quite a few imitators; a July 17 article by Joon Ian Wong listed nine other cryptocurrencies – Ethereum, Etherium Classic, Ripple, Litecoin, Dash, NEW, IOTA, Monero and EOS. (Others not mentioned include Namecoin, Primecoin, Nxt, BlackCoin and Peercoin.) That article also points to how unstable the exchange values of cryptocurrencies can be: in a seven-day period in July, several lost over 30% of their dollar values, and $7 billion of their market value was wiped out!

From our point of view, what's interesting is a couple of variations in how alternative systems are organised. Several of these systems have ditched the 'proof-of-work' competition as a way of winning the right to make the next block, in favour of some variant of what's called 'proof-of-stake'.

As an example, consider **Nxt**, founded in late 2013 with a crowdsourced donation campaign. A fixed 'money' supply of a billion NXT coins was then distributed, in proportion initially to the contributions made; from this point, trading began.

Within the Nxt network, the right to 'forge' the next block in the transaction record chain is allocated partly on the basis of the amount of the currency a prospective 'forger' holds (that's the *Stake* element), but also on the basis of a randomising process. Thus the task is allocated to a single machine, rather than being competed for; and without the puzzle-solving element, the amount of compute power and energy required is slight – the forging progess can even run on a smartphone! As for the rewards for 'playing the game' and forging the block, the successful block-forger gains the transaction fees.

Marc specifically mentioned **Ethereum**, founded in 2014–15, the currency of which is called the 'ether'. In particular he referred to how Ethereum supports '**Smart Contracts**', which are exchange mechanisms performed by instructions in a scripting language being executed on the **Etherium Virtual Machine** – not literally a machine, but a distributed computing platform that runs across the network of participating servers.

Smart contracts have been explored by the bank UBS as a way of making automated payments to holders of 'smart bonds', and a project called The DAO tried to use the Ethereum platform to crowdfund venture capital. Scripts can execute conditionally – the Lighthouse project is a crowdfunding service that makes transfers from funders to projects only if the funding campaign target has been met.

Other uses of blockchain distributed ledgers

In October 2015, a feature article in *The Economist* pointed out that 'the technology behind bitcoin lets people who do not know or trust each other build a dependable ledger. This has implications far beyond the cryptocurrency.' One of the areas of application they highlighted was the secure registration of land rights and real estate transactions, and a pioneer in this has been **Lantmäteriet**, Sweden's Land Registry organisation.

Establishing a blockchain-based publicly inspectable record about the ownership (and transfer of ownership) of physical properties poses some different problems than those which simply transfer currency. The base records can include scans of signed contracts, digital photos, maps and similar objects. What Lantmäteriet aims to collect in the blockchain are what it dubs 'fingerprints' for these digital assets – SHA-256 hashes computed from the digital data. You cannot tell from a fingerprint what a person looks like, but it can still function as a form of identity verification. As a report on the project explains:

> 'A purchasing contract for a real estate transaction that is scanned and becomes digital is an example. The hash that is created from the document is unique. For example, if a bank receives a purchasing contract sent via email, the bank can see that the document is correct. The bank takes the document and run the algorithm SHA-256 on the file. The bank can then compare the hash with the hash that is on the list of verification records, assuming that it is available to the bank. The bank can then trust that the document really is the original purchasing contract. If someone sends an incorrect contract, the hash will not match. Despite the fact that email has a low level of security, the bank can feel confident about the authenticity of the document.'

('The Land Registry in the blockchain' — http://ica-it.org/pdf/Blockchain_Landregistry_Report.pdf)

In the UK, **Her Majesty's Land Registry** has started a project called 'Digital Street' to investigate using blockchain to allow property ownership changes to to close instantaneously. Greece, Georgia and Honduras have similar projects underway.

In Ghana, there is no reliable nationwide way of registering ownership of land and property, but a non-profit project called **Bitland** is drawing up plans for a blockchain-verified process for land surveys, agreements and documentation which – independent of government – will provide people with secure title (www.bitland.world). As they point out, inability to prove ownership of land is quite ommon across Africa, and means that farmers cannot raise bank capital for development by putting up land as security.

Neocapita is a company which is developing Stoneblock as a decentralised blockchain-based registration service for any government-managed information, such as citizen records. They are working in collaboration with the United Nations Development Program, World Vision, and two governments (Afghanistan and Papua New Guinea), initially around providing a transparent record of aid contributions, and land registry.

Noeleen Schenk on blockchain and information governance

After Marc Stephenson had given his technical overview of Blockchain, **Noeleen Schenk** (also of Metataxis) addressed the issue of what these developments may mean for people who work with information and records management, especially where there are issues around governance.

Obviously there is great interest in blockchain in financial markets, securities and the like, but opportunities are also being spotted around securing the integrity of the supply chain and proving provenance. **Walmart** is working with IBM on a project which would reliably track foodstuffs, from source to shelf. **The Bank of Canada** is looking towards using blockchain methods to verify customer identities onwards, on the basis that the bank has already gone through identity checks when you opened your account. Someone in the audience pointed out that there are also lots of applications for verified records of identity in the developing world, and Noeleen mentioned that **Microsoft and the UN** are looking at methods to assist the approximately 150 million people who lack proof of identity.

Google DeepMind Health is looking at using some blockchain-related methods around electronic health records, in a concept called 'Verifiable Data Audit' which would automatically record every interaction with patient data (changes, but also access). They argue that health data needn't be as radically decentralised as in Bitcoin's system – a federated structure would suffice – nor is proof-of-work an appropriate part of the blockmaking process in this context. The aim is to secure trust in the data record (though ironically, DeepMind's was recently deemed to have handled 1.6 million Royal Free Hospital patient records inappropriately).

Noeleen referred to the ISO standard on records management, **ISO 15489-1**, which gives as the characteristics of 'authoritative records' – meeting standards for authenticity, reliability, integrity and usability. What has blockchain to offer here?

Well, where a blockchain is managed on a decentralised processing network, one advantage can be distributed processing power, and avoidance of the 'single point of failure' problem. The use of cryptographic hashes ensures that the data has not been tampered with, and where encryption is used, it helps secure data against unauthorised access in the first place.

Challenges to be solved

Looking critically at blockchain with an information manager's eye, Noeleen noticed quite a few challenges, of which I highlight some:

- Private blockchains are beginning to make their appearance in various sectors (the Walmart provenance application is a case in point). This raises questions of what happens when different information management systems need to interoperate.

- In many information management applications, it is neither necessary nor desirable to have all of the information actually contained within the block (the Lantmäteriet system is a case in point). Bringing blockchain into the picture doesn't make the problem of inter-relating datasets go away.

- Blockchain technology will impact the processes by which information is handled, and people's roles and responsibilities with that process. Centres of control may give way to radical decentralisation.

- There will be legal and regulatory implications, especially where information management systems cross different jurisdictions.

- Noeleen has noticed that where people gather (with great enthusiasm) to discuss what blockchain can do, there seems to be very poor awareness amongst them of well-established record-keeping theory, principles, and normal standards of practice. The techies are not thinking about information management requirements enough.

These issues require information professionals to engage with the IT folks, and advocate the incorporation of information and record keeping principles into blockchain projects, and the application of information architectural rigour.

Intermediate discussion

Following Noeleen's presentation, there were some points raised by the audience. One question was how, where the blockchain points to data held externally, that external data can itself be verified, and how it can be secured against inappropriate access.

Someone made the point that is is possible to set up a 'crypotographic storage system' in which the data is itself encrypted on the data server, using well established public-private key encryption methods, and therefore accessible only to those who have access to the appropriate key. As for the record in the blockchain, what that stores could be the data location, plus the cryptographic hash of the data, so that any tampering with the external data would be easy to detect.

What blockchain technology doesn't protect against, is bad data quality to start with. I'm reminded of a recent case in which it emerged that a sloppy clinical coder had entered a code on a lady's record, indicating that she had died of Sudden Infant Death Syndrome (happily, she was very much alive). That transaction can never be erased from the blockchain – but it doesn't stop the record being corrected after.

Blockchain and the Archive: the TNA experience

Our third presentation was from **John Sheridan**, Digital Director at The National Archives (TNA), with the title 'Application of Distributed Ledger Technology'. He promised to explain what kinds of issues the Archive worries about, and where they think blockchains (or distributed ledgers more generally) might help. On the digital side of TNA, they are now looking at three use-cases, which he would describe.

John remarked that the State gathers information 'in order to make Society legible to it' – so that it might govern. Perhaps The Domesday Book was one of the world's first structured datasets, collected so that the Norman rulers might know who owned what across the nation, for taxation purposes.

The Archive's role, on the other hand, is to enable the citizen to see the State, and what the State has recorded, by perusing the record of government (subject to delays).

Sir Hilary Jenkinson, an important contributor to archive theory, established the principles on which The National Archive runs today. But it is a very paper-based way of looking at the job.

Much of the ethos of the TNA was set by Sir Hilary Jenkinson, of the Public Record Office (which merged with three other bodies to form TNA in 2003). He was a great contributor to archive theory, and in 1922 wrote *A Manual of Archive Administration* (text available in various formats from The Internet Archive, https://archive.org/details/manualofarchivea00jenkuoft). TNA still follows his attitude and ideas about how information is appraised and selected, how it is preserved, and what it means to make that information available.

An important part of TNA practice is the Archive Descriptive Inventory – a hierarchical organisation of descriptions for records, in which is captured something of the provenance of the information. 'It's sort of magnificent… it kind of works,' he said, comparing it to a steam locomotive. But it's not the best solution for the 21st century. It's therefore rather paradoxical that TNA has been running a functional digital archive with a mindset set that is 'paper all the way down' – a straight line of inheritance from Jenkinson, using computers to simulate a paper record.

Towards a second-generation digital archive

It's time, he said, to move to a second-generation approach to digital archive management; and research into disruptive new technologies is important in this.

For the physical archive, TNA practice has been more or less to keep everything that is passed to it. That stuff is already in a form that they can preserve (in a box), and that they can present (only eyes required, and maybe reading spectacles). But for the digital archive, they have to make decisions against a much more complex risk landscape; and with each generation of technological change, there is a change in the digital preservation risks. TNA is having to become far more active in making decisions about what evidences the future may want to have access to; and, which risks they will seek to mitigate, and which ones they won't.

They have decided that one of the most important things TNA must do, is to provide evidence for purposes of trust – not only in the collection they end up with, but also in the choices that they have made in respect of that collection. Blockchain offers part of that solution, because it can 'timestamp' a hash of the digital archive asset (even if they can't yet show it to the public), and thereby offer the public an assurance, when the archive data is finally released, that it hasn't been altered in the meantime.

Some other aims TNA has in respect of the digital archive is being more fluid about how an asset's context is described; dealing with uncertainties in provenance, such as about when a record was created; and permitting a more sophisticated, perhaps graduated form of public access, rather than just now-you-can't-see-it, now-you-can. (They can't simply dump everything on the Web – there are considerations of privacy, of the law of defamation, of intellectual property and more besides.)

The Archangel project

Archangel is a brand new project in which TNA is engaged together with the University of Surrey's Centre for the Digital Economy, and the Open Data Institute. It is one of seven projects which EPSRC is funding to look at different contexts of use for distributed ledger technology. Archangel is focused specifically on public digital archives, and they will try to work with a group of other memory institutions.

The Archangel project will not be using the blockchain methods which Marc had outlined. Apparently, they have their own distributed ledger technology (DLT), with 'permissioned' access.

The first use-case, which will occupy them for the first six months, will focus on a wide variety of types of research data held by universities: they want to see if they can produce sets of hashes for such data, such that at a later date when the findings of the research are published, and the data is potentially archived, any question of whether the data has been tampered with or manipulated can be dealt with by cryptographic assurance spread across a group of participating institutions. (The so-called 'Climategate' furore comes to mind.)

The second use-case is for a more complex kind of digital object. For example, TNA preserves the video record of proceedings of The Supreme Court. In raw form, one such digital video file could weigh in at over a terabyte! Digital video transcoding methods, including compression algorithms, are changing at a rapid pace, so that in a decade's time it's likely that the digital object provided to the public will have to have been converted to a different file format. How is it possible to create a crypographic hash for something so large? And is there some way of hashing not the bit sequence, but the informational content in the video?

It's also fascinating to speculate about how machines in future might be able to interpret the informational content in a video. At the moment, a machine can't interpret the meaning in someone's facial expressions – but maybe in the future?

For this, they'll be working with academics who specialise in digital signal processing. They are also starting to face similar questions with 'digital surrogates' – digital representations of an analogue object.

The third Archangel use case is about Deep Time. Most people experimenting with blockchain have a relatively short timescale over which a record needs to be kept in verifiable form, but the aspirations of a national archive must looks to hundreds, maybe thousands of years.

Another important aspect of the Archangel project is the collaboration which is being sought between memory institutions, which might reach out to each other in a concerted effort to underscore trust in each others' collections. On a world scale this is important because there are archives and collections at significant risk – in some places, for example, people will turn up with Kalashnikovs to destroy evidence of human rights abuses.

Discussions

NetIKX meetings typically feature a 'second half' which is made up of table-group discussions or exercises, followed by a summing-up plenary discussion. However, the speakers had not organised any focused discussion topics, and certainly the group I was in had a fairly rambling discussion trying to get to grips with the complexity and novelty of the subject. Likewise, there was not much 'meat' that emerged in the ten minutes or so of summing up.

One suggestion from Rob Begley, who is doing some research into blockchain, was that we might benefit from reading Dave Birch's thoughts on the topic – see his Web site at http://www.dgwbirch.com. However, it's to be borne in mind that Birch comes at the topic from a background in electronic payments and transactions.

Some closing thoughts

There is a lot of excitement – one might say hype – around blockchain. As Noeleen put it, in the various events on blockchain she had attended, a common attitude seems to be 'The answer is blockchain! Now, what was the problem?' As she also wisely observed, the major focus seems to be on technology and cryptocurrency, and the principles of information and records management scarcely get a look-in.

The value of blockchain methods seem to centre chiefly on questions of trust, using a cryptographic hashing and a decentralised ledger system to create a hard-to-subvert timestamped record of transactions between people. The transactional data could be about money (and there are those who suggest it is the way forward for extending banking services in the developing world); the application to land and property registration is also very promising.

Another possible application I'm interested in could be around 'time banking', a variant of alternative currency. For example in Japan, there is a scheme called '**Fureai Kippu**' (the 'caring relationship ticket') which was founded in 1995 by the Sawayaka Welfare Foundation as a trading scheme in which the basic unit of account is an hour of service to an elderly person who needs help. Sometimes seniors help each other and earn credits that way, sometimes younger people work for credits and transfer them to elderly relatives who live elsewhere, and some people accumulate the credits themselves against a time in later life when they will need help. It strikes me that time-banking might be an interesting and useful application of blockchain – though Fureai Kippu seems to get on fine without it.

When it comes to information-management applications which are non-transactional, and which involve large volumes of data, a blockchain system itself cannot cope: the record would soon become impossibly huge. External data stores will be needed, to which a blockchain record must 'point'. The hybrid direction being taken by Sweden's Lantmäteriet, and the Archangel project, seems more promising.

As for the event's title ' The implications of Blockchain for KM and IM' — my impression is that blockchain offers nothing to the craft of knowledge management, other than perhaps to curate information gathered in the process.

Conrad Taylor, NetIKX rapporteur
July 2017