A Typology of Blockchain Recordkeeping Solutions and Some Reflections on their Implications for the Future of Archival Preservation

Victoria L. Lemieux
School of Library, Archival and Information Studies
The University of British Columbia
Vancouver, Canada
vlemieux@mail.ubc.ca

Abstract—**This paper presents a synthesis of original research documenting several cases of the application of blockchain technology to land transaction, medical, and financial record keeping. Using a thematic synthesis of the cases, the paper describes a typology of blockchain solutions for managing current records representing three distinct design patterns. It then considers the different types of solutions in relation to implications for recordkeeping and long-term preservation of authentic records.**

Keywords—blockchain; distributed ledger; recordkeeping; digital preservation

I. INTRODUCTION

Blockchain technology - a novel from of distributed ledger that cryptographically secures records of transactions – is transforming the creation and keeping of records [1].

Throughout time, similar changes in the technical form of records and information creation, storage and use have led to concomitant changes in the practice of preserving records, either because these changes required different approaches to archival preservation or because they offered new technical capabilities that archivists have been able to apply to the preservation of archival records.

This paper presents a synthesis of original research conducted under the auspices of the "Records in the 'Chain'" Project [2] by the author and her research collaborators between October 2015 to October 2017 documenting several cases of the application of blockchain technology to land transaction, medical, and financial record keeping. Using a thematic synthesis of the case studies, the paper describes a typology of blockchain solutions for managing current records representing three distinct design patterns. It then considers the different types of solutions in relation to recordkeeping and long-term preservation of authentic records, discussing new archival projects that illustrate how blockchain technology might alter future archival work.

The paper is organized into three sections as follows:

(1) A brief overview of blockchain technology

(2) A typology of blockchain recordkeeping solutions and some recordkeeping implications

(3) A discussion of archival implications

II. AN OVERVIEW OF BLOCKCHAIN TECHNOLOGY

Blockchain is a type of distributed ledger technology in which validated sets of transaction records are grouped into blocks, which are then chained together cryptographically (i.e., using hashes, 256-bit random numbers generated from input information), computationally validated and broadcast throughout a peer-to-peer mesh network [1]. This architecture and mode of operation is said to offer immutability and transparency of transaction records, which solves an important problem: the problem of trust.

Trust is necessary in any form of interaction. Through cryptographically securing records and distributing copies that can be compared, it is possible to protect and validate the integrity of records as one of the key elements necessary to be able to trust them. Trusted records are an important foundation for other types of trust, such as the trust between citizen and state, business counterparts, or communicating system components. This ability to provide a foundation for trust is what sets blockchain technology apart from other information processing technologies and makes it a unique innovation as well as an especially important one from an archival perspective.

To illustrate the process of blockchain transaction recording in blockchains that record financial transactions, for example, a transaction record may reflect that a certain amount of value, or coin, has been deducted from the value attributed to one wallet and added to the value attributed to another wallet [3]. Wallets are actually just addresses on the network. These addresses are denoted by the hash of a public key - a hash that functions somewhat like a postal code indicating the destination of a particular transfer of value. For each public key there is a matched private key, which unlocks the wallet [1]. The person who holds the private key controls the wallet and the assets in it. When individuals want to transfer value to someone else's wallet, they must use their private key to digitally sign the transaction in order to give it effect. Since there is nothing inherent in the operation of the blockchain that links a person's real world identity to their public or private key pairs, blockchains are said to operate pseudonymously [1].

When a new block reaches its maximum size, it is chained together with all the previous blocks. This is accomplished cryptographically by combining a hash of information in a current block with the hash of the previous block to create a new block hash. (See Fig. 1) and computationally by the operation of a consensus mechanism) [1].

The consensus mechanism is an algorithm designed to ensure that updates to the blockchain are agreed and

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

communicated across the entire network in a transparent manner, that the order in which records of transactions entered the blockchain is undisputed and that any changes to what has been written to the blockchain will be detectable. Once written to the blockchain, transaction records are meant to be unchangeable.

Each of these consensus mechanisms incentivizes the nodes on the network to behave slightly differently depending on its reward or incentive mechanism [1]. An example of such incentives is the financial reward given to "miners" (i.e., nodes that validate transactions) on the Bitcoin blockchain. This reward incentivizes the miners to validate and secure transactions on the Bitcoin network [1]. Once validated by means of the consensus mechanism, each node on the peer-to-peer mesh network receives an update to its copy of the ledger [1].

The linking together of blocks results in a long continuous chain of hashes, hence the name blockchain and makes tampering without detection difficult [1]. In some blockchains, when the chain becomes quite long (Bitcoin, 2017), it is shortened into a Merkle Tree, where the root of the tree is the hash that is formed from the hashes of all previous transactions. This allows storage space to be saved without breaking the integrity of the chain [6]. When the continuous chain splits into two (e.g., when there is a change in the blockchain protocol or software) the blockchain is said to have forked [5].

In addition to being distributed systems with many dispersed components, some blockchains (i.e., public blockchains) operate as decentralized systems; that is, nodes do not operate under the control of a centralized server, but in an independent albeit coordinated manner. These blockchains may also be characterized by decentralized governance; that is, they may not operate under the formal authority of a single person or organization (even though groups of individuals or organizations may wield informal control over their operation). Examples of these types of blockchains include Bitcoin and Ethereum. Other blockchains and distributed ledgers operate under the control of a single authority (e.g., Ripple, Guardtime) or the authority of a consortium (e.g., R3).

Blockchains (and distributed ledgers) may be public or private, permissioned or permissionless. Public blockchains are those that any participants may use and access. Public blockchains are often permissionless; that is, participants do not require special authorization or authentication to access, read, write and be participants in transactions and in the consensus process [5]. Permissioned blockchains, on the other hand, are ones in which nodes must have a member identity and participants must have authority and authentication to access [5]. These are often private blockchains, meant for the use of only members of a shared ledger or a single ledger that multiple participants may access and use. Permissioned blockchains have membership services that manage identities of the members of the blockchain system.

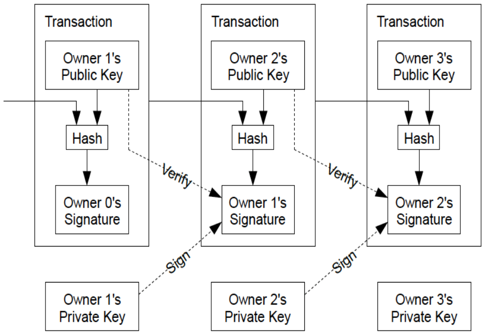

Figure 1. Blockchain structure (Source: [4])

As use of blockchain technology has expanded, new applications and services have emerged. Blockchain applications, for instance, run over blockchain networks and permit participants to easily interact with these networks. Smart contracts (also termed chaincode, programmable asset or programmable contract) are blockchain applications that express business logic associated with a transaction and execute on a blockchain platform [5]. Smart contract code determines what transactions are recorded into the blockchain and the information they will contain. Through the use of smart contracts, their proponents claim that many kinds of contractual clauses may be made partially or fully self-executing or self-enforcing, or both. In this sense, smart contracts represent a digital means to administer and enforce property rights [7]. This does not mean, however, that the outcome of a smart contract is legally binding.

Asset registries link digital currencies to other assets or records on top of a distributed ledger [8]. An asset may be a piece of land, an artwork, food or anything of value. When these assets are linked to digital currencies, they are said to have been 'tokenized'. When the digital currency to which an asset is connected is transferred, as in per the process described above, the transaction results in a changes in the state of that unit of value or asset (e.g., a payment or a transfer of ownership) as well. [3].

III. A Typology of Blockchain Recordkeeping Solutions

Many current and proposed applications of blockchain technology aim to address recordkeeping challenges, such as more efficient and secure processing of land title transfers [9], greater patient control over sensitive health information [10], and more efficient recording of financial payments and settlements [11]. All of the blockchain systems offer a new form of records generation use, storage and/or control.

From a synthesis of original research data collected between October 2015 to October 2017 pertaining to three pilot blockchain-based land transaction recordkeeping solutions (Brazil, Sweden, and Honduras), one blockchain e-health record keeping solution (Estonia), and one proposed cryptocurrency solution (Sweden), it is possible to infer an emergent typology of blockchain solutions for recordkeeping. The model posits a transition from types of systems that are closest to current practices (least innovative) on the left to most innovative, or those that are least like current recordkeeping practices on the left (see Fig 2). Each of these types of solutions exhibits its own particular design pattern, and each presents unique challenges for record keepers. Though for the purpose of this discussion, each type of system is portrayed as being unique and atomistic, in the "wild" these systems may bear a mix of the characteristics of different types of systems (e.g., a system that mirrors digital records, but also uses an underlying token to represent an asset). This is a reminder that the "map is not the territory" and that, while models can be useful, they abstract away from the complexity of reality, sometimes helpfully and sometimes less helpfully.

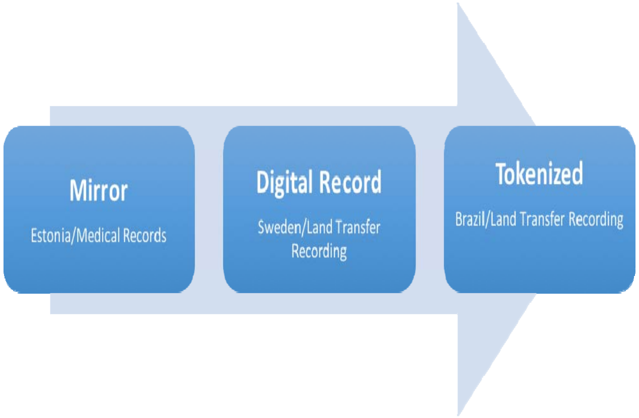
Mirror
Estonia/Medical Records
Digital Record
Sweden/Land Transfer Recording
Tokenized
Brazil/Land Transfer Recording

Figure 2: A Typology of Blockchain Recordkeeping Solutions

A. *Mirror Type*

In the first type of system, the "mirror" system, the blockchain serves as a repository of "digital fingerprints", or hashes, of the records in an originating system. The original records, which may be born paper or digital, but now exist in digital form, are hashed. This produces a sort of digital fingerprint of the record. These hashes are anchored into the blockchain, with the blockchain being used as a means of validating the integrity of the records.

This design pattern may be implemented using either a public blockchain (e.g., the Brazilian land registration pilot project, which uses the Bitcoin blockchain) or a private, permissioned blockchain (e.g., the Estonian e-health solution, which uses the Guardtime system). An important point about this type of blockchain recordkeeping system is that it does not result in the creation of records natively on chain but rather runs in parallel to, and cryptographically mirrors, existing digital systems. In the case of the Brazilian land transaction recording pilot, the blockchain solution mirrors the land registry system, while in the case of the Estonian e-health blockchain solution, the e-health database is mirrored. The primary purpose of this type of blockchain system is to protect the integrity of the records, since the hashing of the original records and ability to compare that hash with a hash recorded on the blockchain provides a means to determine whether, at some point, the records have been tampered with (i.e., if the two hashes do not match, they records have been altered in some way).

As mentioned, an example of a mirror system is the Government of Estonia's e-health solution. The application of a blockchain solution in the context of the e-Health Information System aims to ensure the integrity of Estonian e-health records. Mandated by the "Act of the e-Health Information System," integrity must be ensured at the highest level of the Estonian Information Security Guideline (ISKE) for such records. The source of the information used to populate the blockchain is the government's e-Health Oracle Database. The most important data elements managed in the e-Health register are (see Fig. 3):

- documents (*Dokumendid*): documents about a healthcare occurrence (e.g., a visit to a doctor, surgery, analyse in a lab)
- patient data (Patsiendid)
- Relations register (Viidaregister): relations between patients and documents
- Relations audit (Viidaregistri audit) and patient audit (Patsientide audit): technical audit data on the integrity of the according relations and patient tables

The "documents" are kept in XML format. In addition, most documents are also digitally signed and kept in a hash table in the system. New entries to documents and the audit tables are hashed and added to a hash chain / hash tree (Räsiahel) which is internal to the Oracle database. All actions (view, change, add queries) are logged and saved in plain text SQL format and exported from the system as a standalone log. Every business day around 40,000 documents are created in the system with 1,000,000 actions (change, add, log, audit). The size of the database is currently around 2 TB with an annual growth of 300-400 GB [12].

In addition, the AuditLog (all executed SQL queries/actions that are exported from the system) are hashed and added to the internal system hash tree. One hash tree contains about 1000 hashes. The top of the hash tree is sent to the Guardtime Keyless Signature (KSI) network around every minute, ensuring that third party verification is possible a minute after the creation of data. The e-Health system is connected to the GuardTime KSI network using the Catena middleware (formerly called Fusion). The main benefit of Fusion is that it enables asynchronous communication, as such preventing the KSI solution from situations of network failure [12].

Despite the fact that records are not actually created or stored on chain in mirror solutions, only their digital fingerprints in the form of hashes, a number of solution providers have claimed that records are, in fact, "archived" on chain (for a further discussion of this see [3]). This is misleading, and leads to a key oversight in the design solutions being that the digital originals (or digitized copies of paper originals) must be digitally preserved alongside the blockhained hashes of the records. The requirements for preserving the originals (or digitized copies) are not fundamentally changed with this type of blockchain system (e.g., compliance with ISO 14721 [13] to conform to the OAIS

Reference Model and establish an Open Archival Information System (or OAIS) or a trusted digital repository). Digital preservation of the hashes of the records is more complex, with the theory being that "Lots of Copies Keeps Stuff Safe" (the LOCKS principle [14]); that is, if one node on a distributed blockchain system failed, a copy of the hashes in the ledger would be available on another node. In practice, the whole system is dependent on their being a full copy of the ledger on all nodes, which is not always the case, or at least enough full copies that one full node would survive

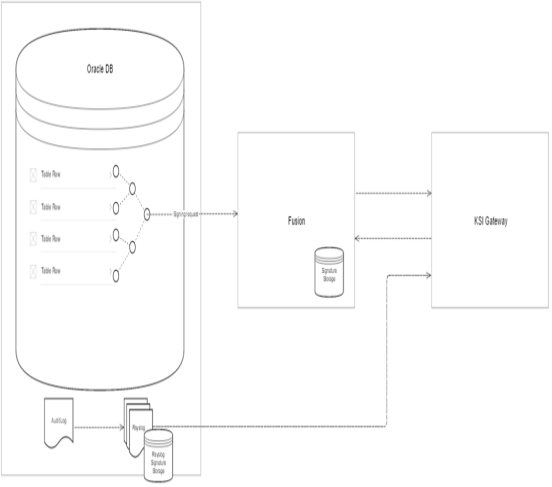

Figure 3. Generic view of GuardTime KSI implementation[12]..

If only one full node survived, however, it would be impossible to determine whether that node had been tampered with, since the integrity of the node is dependent upon matching its copy of the ledger with other surviving copies of the ledger. The more matching the better. For example, on the Bitcoin network it is typical to wait for at least for confirmation of a transactions by at least six nodes before accepting a transaction as valid [15]. Moreover, continued operation of the network is premised upon the continued attractiveness of the incentives that entice nodes to validate transactions. In the Bitcoin network, for instance, each of the validating nodes, the miners, earns a financial reward paid in Bitcoin. If the price of Bitcoin were to drop precipitously, would these miners still wish to validate transactions? Likely not. Each blockchain may have its own unique consensus algorithm (e.g., there is "Proof of Stake", "Proof of Work" and "Practical Byzantine Fault Tolerance" [16]), which establishes its own unique incentives for validating transactions. If those incentives change, the blockchain system may not be sustainable.

B. Digital Record Type

In the second type of system, records are no longer just mirrored or "fingerprinted" on chain, they are actively created on chain in the form of "smart contracts". This marks a more fundamental departure from the traditional form of digital records creation and storage in centralized databases or cloud-based platforms. Typically, these smart contracts encode procedures that execute among a multi-stakeholder network as part of a work process flow. In the case of the Swedish land transfer pilot, for example, the stakeholders in the network include the buyers and sellers of property, the banks that offer mortgages, and the Swedish land registration authority, among others. In these "digital records" systems, execution of the smart code results in an update to the distributed database of records, or ledger, implementing state change once the smart contract has completed.

The pilot Swedish land transfer registration system offers an example of this type of blockchain recordkeeping system. This system uses ChromaWay's two products: Esplix, a smart contract enabled workflow middleware which implements processes and workflows to be described using code, and Postchain, which provides a permissioned technical ecosystem combining enterprise databases with private, permissioned blockchains [17].

To buy or sell a property, a user accesses the ChromaWay solution through a mobile or web-based end user interface, such as that shown in Fig. 4. Behind-the-scenes, the end-user interface interacts with the Esplix application contract engine, where the smart contracts are defined and executed for each property transfer process. Using the contract editor, it is possible to see the actual code of a smart contract. The Esplix user interface also shows a block explorer for the blockchain, revealing the transactions being added to the blockchain as each element of a smart contract defining the property transfer process is executed, from the initial offer to sell a property to the final recording of the fully executed deed of sale transferring ownership of the property [17].

Each new transaction represented in the business logic encoded in the smart contract generates a message from Explix and creates a new block within the Postchain ecosystem. Postchain is described as a "hybrid" between a database and blockchain. According to ChromaWay, it is suitable for storing and retrieving data in the same way as a traditional database but incorporates the distribution and redundancy of a blockchain. It is aimed at providing secure replication of data between a "consortium" of databases, with every node operating as part of the consortium and possessing a copy of the database. Postchain's architecture differs from traditional databases in that the blockchain sits inside the database. There are tables for raw transaction data, and for blockchain data (headers and hashes). Postchain can be thought of as a "manager" for the database, which acts as a transaction handling layer for the database to ensure that the transactions are properly ordered, deterministic, etc.; whereas, in other blockchains, databases (e.g., noSQL or MongoDB) are linked through APIs to a blockchain processing layer [17].

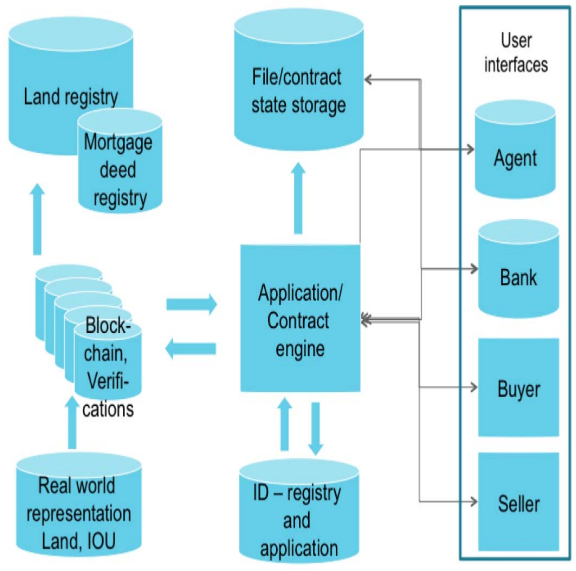
Land registry
Mortgage deed registry
Block-chain, Verifi-cations
Real world representation Land, IOU
File/contract state storage
Application/ Contract engine
ID – registry and application
User interfaces
Agent
Bank
Buyer
Seller

Figure 4. Overview of ChromaWay Postchain architecture for registration of Swedish property transactions [18].

Postchain is said to be able to function with a variety of different consensus algorithms, but is currently working with a modified version of Practical Byzantine Fault Tolerance (PBFT) [19]. In standard PBFT, a client node sends a message to the network (e.g., a primary node or may be more than one node, depending on the system design). The primary node, or the nodes in receipt of the message, broadcast the message out to the other nodes. Once enough identical responses from nodes occur, then the transaction is validated and added to the blockchain.

This type of blockchain solution takes the record professional into uncharted territory. The status of smart contracts as legally admissible evidence of business transactions is, as yet, uncertain. Their admissibility has not been tested by the courts, and there is currently no case law or other guidance that could reasonably be relied upon. Further, what is the actual record? It may be considered to be the instructions or procedures drafted in narrative or diagrammatic form to be implemented in a smart contract. It may be the raw code written in the smart contract scripting language (e.g., Solidity, Serpent, etc.). It may be the compiled code produced after a compiler converts the instructions into a machine-code or lower-level form so that they can be read and executed by a computer. Thus, the question arises as to what is the actual record – a narrative or diagrammatic representation of the will of the creator, the original script, or the compiled code? Another issue is that most of the smart contract languages are not "functional", so different initial conditions may yield different, non-deterministic final states, potentially undermining their evidential quality.

C. *Tokenized Type*

This brings the discussion to the third, and most innovative, type of blockchain recordkeeping solution, the "tokenized solution". With this type of system, not only are records captured on chain, but assets are represented and captured on chain via linking them to an underlying cryptocurrency. As already mentioned, these assets can represent anything of value - land, fine wine, food, diamonds, artworks, etc.

Readers may ask whether these assets are records. For answers we may turn to the well-known archival theorist Sir Hilary Jenkinson. Jenkinson writes in his *Manual of Archive Administration* that there is some question about the treatment of what under English law at the time that he wrote in 1937 were called "exhibits", being non-textual objects that form part of what is formed in the process of an official transaction and set aside for future reference [20]. Says Jenkinson, " . . . there is a case where an old pair of military epaulettes; and among enclosures to letters, forming in each case an integral part of the document, the writer can recall portraits, human hair, whip-cord (part of cat-o'-nine-tails), a penny piece inscribed with disloyal sentiments, and a packet of strange powder destined to cure cancer." [20, 6-7]. Thus, clearly in Jenkinson's view, these exhibits formed part of the archive, or collective body of records.

Unlike in the era in which Jenkison wrote, in the digital era of the blockchain, what once had a material form has become determaterialized. Paper currency becomes cryptocurrency and land, fine wine, artwork, diamonds, food and other material objects, though still physically in existence, can be "tokenized", or dematerialized in the form of virtual tokens that represent their physical form. In this way, in a tokenized blockchain recordkeeping system, literally every "thing" potentially becomes a record.

An example of this type of blockchain recordkeeping solution is found in Ubitquity's Brazilian pilot land titles registration recordkeeping solution. Ubitquity's solution operates using a software-as-a-service (SaaS) business model, for the recording of land transactions on behalf of companies and government agencies. Fees are charged for adding and updating documents on its blockchain platform. An overview of the Ubitquity platform is depicted in Fig. 5.

The solution comprises a web front end that captures information taken from the real estate register, as well as a web server and backup storage. These components communicate with the Colu Application Programming Interface (API), translating what is entered using the front end web user interface into a format that permits assets (i.e., land) and transactions involving those assets (i.e., land transfers) to be recorded on a blockchain. At present the solution uses the Colu "Colored Coins" protocol to tokenize the land (i.e., represent it as a coin on the blockchain to enable its transfer) and for the recording of transactions on the Bitcoin blockchain. Colored Coins is a group of protocols and methods for representing and managing real world assets, such as real estate, as a data layer on top of a blockchain. Each asset – in this case – a piece of land is represented by a Bitcoin that can, once tokenized in this way, be transferred between owners. In this case, Bitcoin is also being used as the blockchain transaction recording layer, but it is possible to use other blockchains. Storage of information on chain in this manner allows for association of that transaction output (more commonly referred to as a "utxo") with a piece of property – a process known as "coloring", hence the use of the label Colored Coins as the name of the protocol [21].

Since the data storage space on the Bitcoin blockchain is limited and may be insufficient for the amount of data a user wishes to associate with a particular transaction, Colu's 'coloring scheme' also allows for association of unlimited amounts of metadata through the use of publicly available torrent files. In this way, data or metadata relating to the asset can be stored and associated with a transaction using BitTorrent [22]. This is a peer-to-peer protocol in which peers coordinate to distribute requested files, much as Bitcoin nodes coordinate to record transactions on a distributed ledger. And, as with Bitcoin, peers can be located anywhere in the world. The continued existence of the data online depends upon at least one, preferably many, peers holding the downloaded data and continuing to participate in the public BitTorrent network. In theory, Colu handles the uploading of metadata content to BitTorrent, which is called 'seeding'.

Now, this form of recordkeeping, though technologically novel in the digital era, is not so new after all as it turns out. M.T. Clanchy tells us that it existed in the medieval era, during the transition from oral to written forms of memoralization:

> In the transition from memory to written record, such symbolic objects play a crucial role. Indeed, early writings are seen as being similar in kind to symbolic objects and the first archivists were keepers of precious things rather than documents as such. The Crowland chronicler writing of the Norman conquest remarks that "at first many grants were conferred by the bare word (*nude verbo*) without a writing or charter, but only with a sword or helmut or horn or cup. A well known medieval example of such an object is the "ancient and rusty sword" which the Earl Warenne allegedly exhibited as evidence of title before Edward I's judges saying: "Look at this, my lords, this is my warrant!" A less familiar but better example, because it still exists, is the broken knife of Stephen de Bulmer kept in the archives of Durham cathedral. To its horn handle is attached a parchment label recording the details of the gift (made in the middle of the twelfth century) which the knife symbolizes. Likewise on the handle is inscribed "*signum de capella de Io*wic (the sigh for the chapel of Lowick)". As the parchment label records details of the gift more clearly than the knife does, the interesting question arises of why the knife was kept. The best explanation is that to people at the time the knife was as important a record as the label, if not more so [23, 117].

Indeed, knives, horns, cups, finger rings, and other objects were customarily used in the conveyance of land during this period [23].

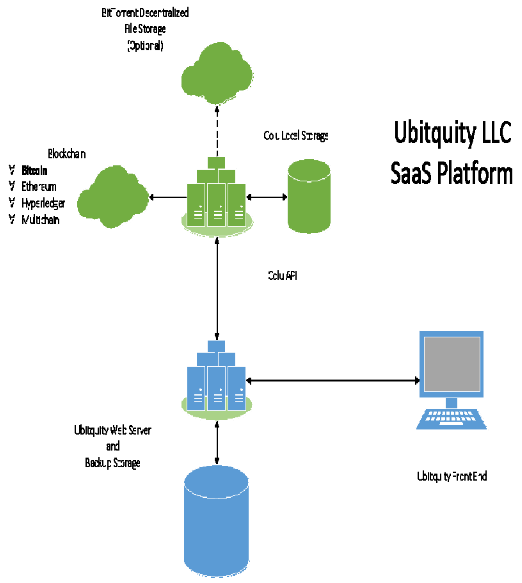

Figure 5. Ubitqity Platform Architecture (Source: Ubitquity)

Although this system of blockchain recordkeeping, which not only records the transfer of title to an asset, but transfers a tokenized version of the asset itself, is novel in that allows for rapid, disintermediated peer-to-peer asset transfers, there are dangers. For one thing, it is possible to accidentally 'spend' the coin, or token, that represents the asset, thereby accidentally extinguishing both the representation of the asset along with title to it. In addition, with increased ease of transfer and disintermediation, any loss of control over a private key, or keys, needed to give effect to an asset transfer, could mean loss of control of the asset. Careful key management, which is never easy undertaking in an cryptographic system, is thus a requirement to safe operation of such systems [24].

IV. ARCHIVAL IMPLICATIONS

We begin this section by reflecting again on how changes in the nature of record creation, use and storage impact upon the methods of archival preservation. With the rise of blockchain recordkeeping, there will, no doubt, be an opportunity, if not a necessity, to alter current archival preservation practices. Indeed, it is possible to observe the early beginnings of these changes.

In a manner similar to the first type of blockchain recordkeeping system, which simply augments the existing recordkeeping system by adding an integrity-checking layer, the blockchain does not fundamentally alter archival practices so much as being conceived as a new tool that archivists can apply to enhancing existing archival practices.

As an example, the InterPARES TrustChain project [25] seeks to use the blockchain to preserve the "chain of preservation" in digitally signed records. In traditional digitally signed records, a trusted third party certificate authority issues a public key which establishes the authenticity of the digital signatory. These certificates are issued for a specific period of time. In many cases, the certificates expire after the records are transferred to the archives, essentially breaking the chain of preservation needed to establish the authenticity of the records from creation to a given future point of time. Blockchain recordkeeping, however, does not require the existence of certificate issuing authorities, since public-private key pairs are self-generated within the system. Thus, incoming archival records can be re-signed while their original certificates are still valid, and cryptographically secured in a blockchain in time ordered sequential manner as a basis for establishing their continued authenticity. In a similar vein, the State Committee for Archives of the Republic of Tatarstan is reported to have decided to experiment with the use of blockchain technology for the acceptance of documentation for archiving. The report indicates that the State Archive, when transferring documents to the archive, will create electronic imprints (i.e., hashes) of documents which will be registered in the blockchain network. After that, a secure transfer of data to the data center is initiated, and at the end of the data transfer, their integrity will be checked [26]. The UK National Archives has also begun similar explorations, under project "Archangel" though it is at a very early stage [27].

This approach assumes the continuance of traditional centralized digital recordkeeping, with use of the blockchain to augment current digital preservation practices. The second type of blockchain system – wherein records are natively created and stored on chain – foreshadows a need for fundamental changes in digital preservation paradigms. Such records must be preserved in their original state, or at least must be preserved in such a way as to make it possible to emulate, or in some other way experience, their original state. But how? The challenge of preserving distributed, decentralized blockchain recordkeeping systems is an open one, with current digital preservation reference models being largely reliant on centralized trusted digital repositories and centralized notions of trusted third parties, i.e., a single preservation authority. Blockchain recordkeeping challenges this centralized approach, and re-opens the debate about non-custodial versus custodial models of digital preservation. Indeed, Peter Van Garderen [28] has posited the notion of "decentralized autonomous collections" (DACs) which he defines as "a set of digital information objects stored for ongoing re-use with the means and incentives for independent parties to participate in the contribution, presentation, and curation of the information objects outside the control of an exclusive custodian" [28]. Van Garderen's proposal sees DACs as an antidote to a number of the problems associated with traditional, centralized institutional repositories: shortage of resources, political interference, and colonial attitudes. For Van Garderen, blockchain technology has the potential to displace traditional institutional archives as curators of digital content.

In a less extreme vision of a distributed and networked future for archival preservation, archives could work collaboratively as nodes on a distributed purpose built blockchain consortium aimed at preserving "born blockchain" records (see Fig. 6). The archival preservation blockchain network could be connected to an active blockchain recordkeeping network via a multihomed (to ensure security) "interchain interoperability service". Archives nodes could be incentivized to preserve blockchain transaction records by means of a novel archival consensus mechanism that would see them receive transaction fees, in a manner similar to the Bitcoin network, for preserving blockchain records.

Clearly, many details of a blockchain-based, or even simply a distributed and networked, archival preservation model would need to be worked out. Much research relating not only the design but the varied effects of such an approach in relation to archival preservation and the operational efficiency of active blockchain recordkeeping solutions would be needed in order to move from speculative ideas to blockchain preservation reality. Moreover, how to integrate various interconnected components of blockchain-based record-keeping solutions, including centralized or distributed off-chain digital records storage, would also require specification. The challenges should not be underestimated, but it is possible to chart a course to addressing them if the archival community works collaboratively with the blockchain developer community.

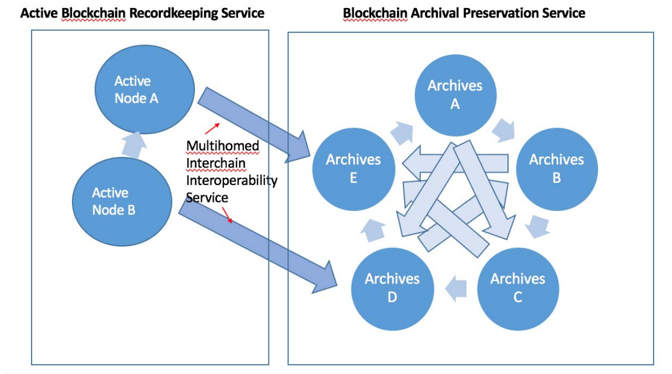
Active Blockchain Recordkeeping Service
Blockchain Archival Preservation Service
Active Node A
Active Node B
Multihomed Interchain Interoperability Service
Archives A
Archives B
Archives C
Archives D
Archives E

Figure 6. A Possible High-Level Future Distributed and Networked Blockchain-based Preservation Model

This brings the discussion to the third type of blockchain recordkeeping solution, the tokenized system. In addition to the changes suggested by the second type, this type of system could truly transform the fundamental nature of archives. Since the late 18th century, archives have been store houses of documents primarily used for historical research. With the tokenization of assets, we could again see archives become stores of treasure, or treasuries, as they were in medieval times. In these times, the kind of things found in the archives were not only textual documents, but also rings and other objects of value, as for example the emerald ring sent by Pope Adrian the IV to King Henry II as a symbol of his investiture as overlord of Ireland [23]. "The first medieval archives were therefore the special places, the *secretarium*, or *archiva*, where valuables of all sorts were kept . . . such archives did not just contain writings but all sorts of memory retaining objects. On looking into such an archive . . . the viewer would have seen . . . bones of saints encased in gold, charters and seals wrapped in Asiatic silks, finger rings, knives symbolizing conveyances, and so on." [23, 122]. These kinds of objects, which in modern times have arguably been taken out of context and treated a museum objects, may in the blockchain era take on their original role (albeit in virtualized form) with the result that the ancient role of archives as stores of tokens of inherent as well as symbolic value could re-emerge.

Such a future could certainly bring new challenges for archivists, but also new opportunities, as in the possibility that originators of blockchain transaction records and related tokenized assets may see an economic value, in addition to a socio-cultural value, in archival preservation. This may, in turn, provide a basis to design a novel archival consensus mechanism as mentioned above, placing the work of archival preservation on a more secure financial footing.

V. CONCLUSION

This paper has considered how blockchain technology - a novel form of distributed ledger that cryptographically secures records of transactions – is transforming the creation and keeping of authentic records and long-term preservation of archives. Using a thematic synthesis of a series of case studies of blockchain recordkeeping solutions, representing original research conducted as part of the "Records in the 'Chain' Project" [2], it has presented a typology of blockchain solutions for managing current records representing three distinct design patterns and discussed some aspects of their impact on recordkeeping. It also has documented several new archival projects that aim to leverage this innovative technology in support of digital preservation, and the potential that exists to see fundamental shifts in current digital preservation paradigms and, indeed, the very role of archives in society as a result of this technology.

ACKNOWLEDGMENTS

The author would like to express thanks to those who have contributed to the development of the case studies in the "Records in the Chain" Project that have provided a basis for this paper: Kuldar Aas from the Estonian National Archives; Artur Novek from the Estonian e-health service; Ivo Lohmus from Guardtime; Claudia Lacombe Rocha from the National Archives, Brazil; Daniel Flores from the Federal University of Santa Maria, Brazil; Rafael Mezzari from the Cartario de Imoveis de Registro, Pelotas, Brazil; Nathan Wosnack of Ubitquity; Mats Snäll of the Lantmäteriet in Sweden; Jörgen Modin and Augustus Botsford of ChromaWay; and Goran and Hans Almgren and Mats Stengard from Enigio.

REFERENCES

[1] Arvind Narayanan, Joseph Bonneau, Edward Felton, Andrew Miller, and Steven Goldfede, *Bitcoin and Cryptocurrency* Technologie.s Princeton University Press, 2016.

[2] Https://blockchainubc.ca/portfolio/records-in-the-chain/.

[3] Lemieux, V.L. Blockchain Technology for Recordkeeping: Help or Hype? Unpublished report, 2016, http://www.idees-ideas.ca/sites/default/files/sites/default/uploads/general/2016/2016-sshrc-ksg-lemieux.pdf.

[4] Nakamoto, S. *Bitcoin: A Peer-to-Peer Electronic Cash System*, 2008.

[5] InterPARES Trust Terminology Project: Key Blockchain Terms and Definitions, 2017, http://arstweb.clayton.edu/interlex/blockchain/.

[6] Bitcoin.org. Bitcoin Developer Guide, 2017, https://bitcoin.org/en/developer-guide#proof-of-work.

[7] Garrod, J.Z. The real world of the decentralized autonomous society. *triple C: Communication, Capitalism & Critique*, 14(1): 2016, pp. 62–77.

[8] Blockchain Council.org. How Blockchain Can Be Used in Asset Registry & Tracking?, 2017, https://www.blockchain-council.org/use-cases/blockchain-within-asset-registry-how-it-works/.

[9] Victoria L. Lemieux, 'Trusting records: is Blockchain technology the answer? *Records Management Journal*' 26(2):2016, pp. 110-139.

[10] A.A. Shrier, A. Chang, N. Diakun-thibault, L. Forni, F. Landa, F., J. Mayo, and R. van Riezen, R., Blockchain and Health IT: Algorithms, Privacy, and Data, 2016. Office of the National Coordinator for Health Information Technology US Department of Health and Human Services.

[11] Gareth W. Peters, Panayi Efstathios. Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money." *Banking Beyond Banks and Money*. Springer International Publishing, 2016., pp. 239-278.

[12] Records in the 'Chain' Project, Estonian E-Health System (RCPEU-01) – Case Study 1, draft version 0.2, 24 May, 2017.

[13] ISO/IEC *ISO 14721: 2012– Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model*, ISO, Geneva, 2012.

[14] LOCKSS, Lots of Copies Keeps Stuff Safe, n.d., https://www.lockss.org.

[15] Https://bitcoin.stackexchange.com/questions/8360/how-many-confirmations-do-i-need-to-ensure-a-transaction-is-successful.

[16] Briscoe, G. Blockchain: distributed consensus protocols, unpublished. Peer-to Peer Financial Systems Workshop, London, UK, 201-21 July, 2017.

[17] Lemieux, V.L. Evaluating the Use of Blockchain in Land Transactions: An Archival Science Perspective. *European Property Law Journal* (2017 forthcoming).

[18] Mats Snäll, Blockchain and Land Register – a new "trust machine"? World Bank Land and Proverty Conference, 22 March, 2017/

[19] Miguel Castro and Barbara Liskov, Practical Byzantine Fault Tolerance,' *Proceedings of the Third Symposium on Operating Systems Design and Implementation, New Orleans, USA,* February 1999, http://www.pmg.lcs.mit.edu/papers/osdi99.pdf.

[20] Hilary Jenkinson, A Manual of Archive Administration, 2 ed., London: Percy Lund, Humphries & Co. Ltd., 1937.

[21] Colored Coins, *Bitcoin* Wiki, 2015, https://en.bitcoin.it/wiki/Colored_Coins#Colu.27s_ColoredCoins.org_Block_Explorer accessed 31 July, 2017

[22] Bram Cohen, The BitTorrent Protocol Specification, *BitTorrent.org,* February 4, 2017, http://www.bittorrent.org/beps/bep_0003.html accessed 31 July, 2017.

[23] Michael T. Clanchy, "Tenacious Letters": Archives and Memory in the Middle Ages. *Archivaria* 11 (1980), pp. 115-125.

[24] S. Eskandari, D. Barrera, E. Stobert, J. Clark, 'A First Look at the Usability of Bitcoin Key Management' (*USEC 2015*, San Diego, CA) http://www.internetsociety.org/sites/default/files/05_3_3.pdf accessed 21 November, 2015.

[25] InterPARES, The TRUSTER Preservation Model (EU31), 2017, https://interparestrust.org/trust/about_research/studies

[26] Anatol Antonovici, Tatarstan (Russia) Plans to Transfer its State Archive to Blockchain, *Cryptovest* October 10, 2017, https://cryptovest.com/news/tatarstan-russia-plans-to-transfer-its-state-archive-to-blockchain/.

[27] Engineering and Physical Sciences Research Council, ARCHANGEL - Trusted Archives of Digital Public Records, 2017, http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/P03151X/.

[28] Peter Van Garderen, P., Decentralized autonomous collections, 2016, https://medium.com/on-archivy/decentralized-autonomous-collections-ff256267cbd6.