Sangeeta Desai. New Recordkeeping on the Block: An Assessment of 2 Blockchain-Based Recordkeeping Systems. A Master's Paper for the M.S. in I.S degree. December 2018. 90 pages. Advisor: Helen R. Tibbo

This study explores the application of blockchain technology to recordkeeping practices. To that end, two blockchain-based recordkeeping platforms—ARCHANGEL and RecordsKeeper—were evaluated according to the three criteria outlined in T. D. Smith's evaluation framework for blockchain-based recordkeeping platforms—dependabilty, security, and trust.

The results of these two evaluations demonstrate blockchain technology's inability to provide viable long-term solutions for sustainable records management as yet. This study also suggests supplementing Smith's framework with more blockchain-specific questions to ensure a more comprehensive evaluation of the use of blockchain in such platforms. Finally, this study recommends adding a fourth criteria, sustainability, to the framework.

Headings:

Blockchains

Data integrity

Digital preservation

Electronic records

Records management

NEW RECORDKEEPING ON THE BLOCK:
AN ASSESSMENT OF 2 BLOCKCHAIN-BASED RECORDKEEPING SYSTEMS

by
Sangeeta Desai

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

December 2018

Approved by

_____

Dr. Helen R. Tibbo

Table of Contents

1 Introduction

With its introduction into the financial and technology sectors in 2007 as the distributed ledger technology underlying the digital currency bitcoin,[1] blockchain technology has quickly spread beyond the realm of cryptocurrency and finance to other fields. Perhaps no other technology since the internet has managed to capture the interest, resources, and imagination of such a wide variety of public and private institutions as blockchain has in just the last few years (Mougayer, 2016). Described by The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) as "an open-source technology that supports trusted, immutable records of transactions stored in publicly accessible, decentralized, distributed, automated ledgers," blockchain technology still lacks a standard definition (Pearce-Moses, Duranti, Michetti, Andaur, Banard, Barlaoura & Pan, 2017). Regardless of its definition, what ultimately makes blockchain technology so compelling is its potential to eliminate the need for trusted third parties and ability to generate secure, immutable records.

As a result, a steady stream of investors, developers, and dreamers continue to explore blockchain-based solutions for issues great and small. Describing the high hopes some blockchain advocates hold for the future of the technology, Klint Finley writes:

> Its biggest boosters believe blockchains can not only replace central banks but usher in new era of online services outside the control of internet

[1] Bitcoin with a capital "B" refers to Bitcoin software, technologies and/or protocols, while bitcoin with the lower case "b" refers to the digital coin. (Bitcoin Project, n.d.)

> giants like Facebook and Google. These new-age apps would be impossible to censor, advocates say, and would be more answerable to users. (2018)

Accordingly, banks, private businesses, start-ups, tech firms and even some governments betting on such potential, have begun to develop blockchain-based applications to solve problems ranging from managing agricultural supply chains, transferring ownership of deeds and titles, preventing voter fraud, providing secure digital identities for refugees and homeless persons, managing electronic health records, sharing and archiving genomic research data and more (Galen et al., 2018).

Skeptics of blockchain, on the other hand, argue that blockchain technology has been overhyped. While some staunch critics feel that blockchain is just a passing fad with no future (Stinchcombe, 2018; Volpicelli, 2018; Walker, 2018), others feel that considerably more research and conceptual development is needed for it to live up to its purported potential (Galen et al., 2018; Lemieux, 2016a; Pisa and Juden, 2017; Zīle and Strazdiņa, 2018). As many critics have noted, while technological development and investment into blockchain has grown exponentially within the last few years, critical research evaluating it has not (Lemieux, 2016a; Zīle and Strazdiņa, 2018). Notably, in their recent article "Blockchain Use Cases and Their Feasibilty" (2018), Kaspars Zīle and Renāte Strazdiņa point out various obstacles to the successful implementation of blockchain, including ongoing technical issues with blockchain technology, differing understandings of the uses and goals of blockchain technology and an overall lack of available research about blockchain (Zīle and Strazdiņa, 2018). Having surveyed much of the available research on block chain, Zīle and Strazdiņa (2018) observe that much of this research is either focused primarily on Bitcoin blockchain or security and privacy

improvements. There is thus, a general tendency to ignore other blockchain technologies, potential applications for blockchain, and technological issues.

Zīle and Strazdiņa (2018) also noted that aside from limited scientific articles and a few books on blockchain, much of the published materials consist of articles and white papers published by developers and enthusiasts on blogs and discussion boards. While those articles and white papers provide some interesting technical details and descriptions of various blockchain technologies, overall most of the available material on blockchain does not enable larger theoretical framing of blockchain technologies, comparisons of various applications or even a standard definition of blockchain (Zīle and Strazdiņa, 2018).

In a similar vein, though focused more specifically on blockchain solutions for archival preservation and recordkeeping, Victoria Lemieux (2016a) notes in "Blockchain Technology for RecordKeeping: Help or Hype?" that without the critical analysis and evaluation of blockchain technology that more comparative and theoretical research might produce, it becomes harder to assess the relevance and effectiveness of using blockchain technology for any one particular application. Put simply, without frameworks for evaluation, how does one separate fact from fiction—or rather, true potential from hype—when it comes to using blockchain technology for a particular application?

This paper seeks to explore that question by using an evaluation framework to assess two blockchain-based recordkeeping solutions: ARCHANGEL and RecordsKeeper. More specifically, by using the evaluation framework developed by T.D. Smith (2017) in his recent article "The Blockchain Litmus Test," this paper will

examine how well these two blockchain-based solutions execute recordkeeping functions. Smith's framework seeks to determine the overall utility of blockchain-based systems by rating the project's performance in three primary categories: dependability, security and trust. While Smith applies his framework to the Bitcoin blockchain and four other blockchain-based initiatives, this paper seeks to expand his work by appraising two specific blockchain-based recordkeeping projects that have not yet been evaluated by Smith's framework. In addition to analyzing the assessments of ARCHANGEL and Recordkeeper resulting from application of Smith's framework, insights regarding the application of Smith's framework and suggestions for improving such assessments will then be discussed.

2 Literature Review

2.1 What is Blockchain?

At present, there is no consensus regarding a standard definition for blockchain. Similarly, concepts to attach to any potential definition are just as varied. Defining blockchain is complicated by the fact that the term is used in different ways. "The blockchain," "a blockchain," "blockchain," and "blockchain technology" when used, may refer to the same or different things depending on the context. Thus, the term blockchain has been (and can still be) used to refer to: 1) a data structure 2) an algorithm 3) the specific distributed ledger system underlying Bitcoin 4) derivative distributed ledger systems based off of the original Bitcoin blockchain ledger, or 5) the blockchain concept itself—a distributed peer-to-peer system—with no specific implementation in mind (Burniske & Tartare, 2017; Drescher, 2017). The blockchain-based recordkeeping solutions assessed in this paper fall under the fourth category listed above.

To be able to understand the derivative blockchain systems and the assessments included in this study requires some fundamental background regarding the Bitcoin blockchain, types of blockchains, and the Ethereum blockchain. This section will provide an overview of the Bitcoin blockchain, which includes its origins, how it functions, and potential vulnerabilities. The section will expand further to discuss the differences between public and private blockchains. Another prominent blockchain platform

Ethereum, will be introduced, particularly to examine the use of "smart contracts" in blockchain.

The Bitcoin blockchain was first detailed in 2008 when a person (or group of people) under the pseudonym of Satoshi Nakamoto[2] published a white paper titled "Bitcoin: A Peer-to-Peer Electronic Cash System." (2008). Published on the cryptography listserve metzdown.com, the white paper detailed Nakamoto's concept of bitcoin, as well as the Bitcoin blockchain technology underlying the cryptocurrency (Redshaw, 2017). Therein, Nakamoto described the new peer-to-peer distributed ledger he had developed as "an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party" (Nakamoto, 2008, p. 1). In addition to replacing mediating trust networks to facilitate these cryptocurrency transactions, Nakamoto's blockchain also produced a permanent, immutable record of all these transactions. While many aspects of Nakamoto's blockchain technology were not exactly new, it was Nakamoto's innovative use of immutable records that generated the breakthrough, providing trustless, immutable capabilities.

The key to understanding Nakamoto's innovation begins with understanding the main aspects that allow the Bitcoin blockchain to carry out its primary function—to provide a tamper-proof, reliable record of all bitcoin transactions. As Phil Champagne (2014) succinctly clarifies in his chapter "How and Why Bitcoin Works," these aspects are:

---

[2] The real person or persons behind the name Satoshi Nakamoto have not yet been revealed. Many theories abound as to the real identity of Nakamoto, but as this is outside the scope of this paper, Nakamoto will be referred to throughout this paper using the singular male pronouns as indicated in online profiles of Nakamoto.

> - A public ledger (called Bitcoin's *block chain*). Consider this as essentially a giant book that is publicly available and contains the bookkeeping records of all transactions ever made in the Bitcoin system, with new pages constantly being added.
>
> - A cryptographic algorithm called asymmetric encryption used for authorization of the transactions.
>
> - A distributed network of computer *nodes* (also commonly known as *miners*) that verify and validate Bitcoin transactions and update the public ledger (p.10).

Each of these three components—a public ledger, a cryptographic algorithm, and a distributed network of computer nodes—performs particular tasks that make the entire blockchain work. The Bitcoin blockchain's public ledger is essentially a data structure. More specifically, as Champagne explains it is a distributed register shared by all members of the Blockchain network that is constantly appended, keeping a stored record of every bitcoin transaction (though not the actual bitcoins themselves).

Transactions are not added individually, but in encrypted units, or "blocks," that are linked or "chained" together, hence the term 'blockchain.' Each block contains a list of the most recent transactions and a hash pointer. The hash pointer forms the link to the next block in a long a continuous chain of hashes, thereby making it impossible to delete a block or insert a new one into the middle of the chain (Lemieux 2017a). Sometimes when chains grow too long, the chain can be compressed into a structure known as a hash tree or a Merkle Tree (Lemieux 2017a). A hash derived from all the previous hashes in the chain forms the root of the Merkle Tree, thereby maintaining the integrity of the chain while also saving storage space (Lemieux 2017a).

As Champagne (2014) further explains, using asymmetric encryption, transactions—i.e. blocks—are verified and further validated by Bitcoin miners. Also referred to as public-key cryptography, asymmetric encryption is used to determine who

is authorized to spend the bitcoins. The encryption algorithm generates a pair of different digital signatures or hashes referred to as the public and private key (Champagne, 2014). While the public key can be easily calculated through the algorithm by the private key, it is impossible for the public key to determine the private key (Champagne, 2014). Thus, the public key serves as a visible Bitcoin address for the user, while the private key connected to that public key remains an encrypted hash only visible to the owner through his/her Bitcoin wallet, a password-protected account (Champagne, 2014). Any transactions—transfers of bitcoin to and from the account, current balance, etc.—related to the Bitcoin address constitute a public key, open to public viewing (Champagne, 2014). The owner's identity, however, remains private by means of the private key. Moreover, only the owner of the private key can access, spend, or transfer the bitcoins associated with the key.

Finally, as Champagne (2014) points out it is the network of nodes or miners that essentially keep the Bitcoin blockchain operating. Blocks cannot be added into the chain without reaching network consensus. In other words, all active miners in the network must verify the block's authenticity. Any time Bitcoin owners sign off on a transaction, the transaction is put into a pool of unconfirmed transactions (Champagne, 2014). Miners then generate a block of transactions by selecting transactions from these pools (Champagne, 2014). Approximately every ten minutes a new block is added that lists all the bitcoin transactions to have transpired in the last ten minutes (Champagne, 2014). In a process explained further below, miners race to be the one to verify the block by solving a cryptographic puzzle and thereby add the block to the blockchain. Every other active miner in the network must then validate the block, thus ensuring all miners have the most

current blockchain (Champagne, 2014). As the network continues to validate the transaction it becomes irreversible (Champagne, 2014).

By waiting for confirmation of a payment's receipt, miners also guard themselves from double spending fraud (Champagne, 2014). A double spending fraud occurs when someone successfully spends their money more than once, sending money and then reversing the transaction to return these funds back to their account (Champagne, 2014). One of Nakamoto's major innovations was to use the proof-of-work protocol to ensure transactions become irreversible, preventing double-spending. The proof-of-work protocol, a computationally intensive protocol invented in 1992 by Cynthia Dwork and Moni Naor to deter spam and denial-of-service attacks to email accounts, works well with the decentralized nature of the blockchain network (Camp, 2018). In the most basic sense, a proof-of-work protocol issues a somewhat difficult mathematical challenge—usually in the form of a cryptographic puzzle—to the service requester (Camp, 2018). The requester must solve the puzzle to obtain service (Camp, 2108). The asymmetric protocol must be just difficult enough to slow the requester's computer down, but simple enough for the service provider to check (Camp, 2018).

In the Bitcoin blockchain, the proof-of-work is a cryptographic puzzle in the form of a SHA256 hash. Each block in the chain has its own hash value or checksum, as well as other transactional data: a summary of the proposed transactions included in the block, a set of identifiers associated with the previous block, and a random variable called a nonce (Champagne, 2014). The nonce is a random variable assigned to a new block (Champagne, 2014). Miners compete with other active nodes to solve the puzzle—that is, miners compete to find blocks with specific nonce values that cause the block's hash

value to be unusually small (Smith, 2017). Once a miner has found the right nonce value through what is essentially trial and error, the block can be verified. True to a proof-of-work protocol, the verification—authenticating the block's hash with a SHA256 checksum—is easy (Smith, 2017). Finding the new nonce—hence the term mining—is computationally intensive as it requires hundreds of checksum calculations (Smith, 2017). The first miner to solve the proof-of-work algorithm earns the bitcoins associated with the block. The rest of the miners validate that block by confirming it adhered to all the rules. Once the block is validated, miners drop whatever block they were working on and the entire cycle begins again (Champagne, 2014). It takes about ten minutes to validate a block on the Bitcoin blockchain (Champagne, 2014).

As one might surmise, proof-of-work is inefficient and not only consumes a great deal of computing resources, but also electricity both to power the computers as well as the cooling systems required to keep those systems from overheating. An analysis of Bitcoin blockchain performed in May 2017 determined that at that time, it cost an estimated $50,000 an hour to maintain this hardware (Aste, 2016). According to the Digiconomist's "Bitcoin Energy Consumption Index" (n.d.-a), Bitcoin consumes as much energy as the entire country of Switzerland. The electricity consumed per transaction was estimated at 638 KWh, the same amount of energy that could power 21.57 U.S. households for one day (Digiconomist, n.d.-a). While these figures relate specifically to Bitcoin blockchain, an older blockchain with an extensive mining community and long blockchain, these results are still useful for thinking comparatively about other blockchain projects—particularly ambitious projects aspiring to manage millions of records on their respective blockchains.

Storage is also another concern for any blockchain, as storing these large quantities of data could be impractical and lead to scalability issues. Blockchains are not to be confused with cloud storage, wherein data is stored off-site usually on servers controlled by a third-party. As every node must replicate the entire blockchain, the storage required for each node scales out linearly with the number of nodes (Smith, 2017). As more miners join the network, enabling more transactions to take place, the blocks begin to reach their maximum data limit, creating a chain with full blocks—a phenomenon known as blockchain bloat (Buntinx, 2017). Blockchain bloat will slow down the transaction rate of the entire blockchain. To avoid bloat, block sizes must be increased in order to broadcast more transactions (Buntinx, 2017). Bigger block sizes lead to an increase in transaction fees per block and require nodes to use up more storage space (Buntinx, 2017). To stave off blockchain bloat many blockchains attempt to impose block size limits. If an application with many participants enabled users to save large quantities of data—such as files, images, documents, etc.—on the blockchain, the amount of storage required would soon eclipse each participating node's available storage.

Another problematic issue with blockchain is its claim regarding immutabilty. While it is true that the proof-of-work consensus model and append-only structure of blockchain makes it extremely hard to modify, it is not completely invulnerable to tampering (LearnCryptography.com, n.d.; Smith, 2017). A double-spend attack, also known as a 51% attack on a blockchain, would have the power to interfere with verification of blocks. As its name suggests, a 51% attack can occur when one entity controls more than half of the network's computing power (LearnCryptography.com, n.d.). Because the blockchain relies upon a consensus mechanism to approve

transactions, if one person controlled more than half the nodes they could form and control an artificial consensus (LearnCryptography.com, n.d.). They could then deny other transactions, while approving the sale of their coins multiple times in a double-spend attack (LearnCryptography.com, n.d.).

Double-spend attacks are neither easy nor inexpensive to execute. For those reasons, they are considered to be not only extremely improbable, but easy to fend off, and in the unlikely event of their occurrence, ultimately incapable of causing too much damage. In order to gain a majority consensus of miners, an attacker would need to acquire a great deal of computing equipment that would have the capability to power more than half the nodes in the blockchain network (S., 2018). Aside from acquiring, maintaining, and housing that computing equipment, the attacker would need to pay for the electricity required to run those nodes, as well as the transaction fees incurred for each node per transaction (S., 2018). Moreover, in any attack only the most recent blocks could be altered, as blocks closer to the beginning of the chain are more secure (LearnCryptography.com, n.d.). This means an attacker could likely gain coins only from the most recent transactions, and they would also be unable to mint new coins (LearnCryptography.com, n.d.). For these reasons, many blockchain enthusiasts believe the actual double-spending damage these hypothetical attackers could perpetrate would be minor, and real damage would instead be in the form of a potential loss of trust for the blockchain (LearnCryptography.com, n.d.). As the blockchain might lose its potential legitimacy, the value of any associated coins might drop and other miners might abandon the chain entirely (LearnCryptography.com, n.d.).

Within the past year for example, five cryptocurrencies—Monacoin, bitcoin gold, zencash, verge and litecoin cash—have reported double-spend attacks (Hertig, 2018). These attacks have all occurred on smaller blockchains; blockchains having smaller networks that are easier to overcome (Hertig, 2018). The proof-of-work algorithm provides greater security when there are more active miners hashing and competing against one another. A potential attacker has much more competition, making the double-spend attack far more difficult to execute (Hertig, 2018). Smaller blockchains usually provide less competition for a potential attacker, as their mining networks are not as large (Hertig, 2018). Moreover, attackers may then rent computing power, making it that much easier to amass the computing power needed to execute a highly lucrative double-spend attack (Hertig, 2018). Within the past year, attackers netted nearly $20 million in profits during those five double-spend attacks alone (Canellis, 2018).

The five cryptocurrency blockchains that endured the attacks were, like the Bitcoin blockchain, public, permissionless blockchains. These blockchains are considered public, because anyone downloading the respective blockchain software may access the blockchain, join the mining network and start mining (Lemieux, 2017a). Moreover, these blockchains are considered permissionless, as all participants may access, read, write, and verify transactions without requiring special authorization or authentication (Lemieux, 2017a). Public, permissionless blockchains also often run as decentralized systems—operating without any overseeing central authority figure or institution, coordinated instead by consensus protocols (Lemieux, 2017a).

Blockchains may also be private and permissioned. Private blockchains, unlike public ones, are accessible only by invitation, made exclusively for member use

(Lemieux, 2017a). These permissioned blockchains require miners to possess member identities. Participants must have authorization and be authenticated in order to access the ledger. Permissioned blockchains usually must manage their member identities through membership services (Lemieux, 2107a). One of the key issues with private, permissioned blockchains is that they have a more centralized authority controlling access, authentication, permissions, what transactions can be written to the blockchain, and the method of consensus. While private blockchains are always permissioned, permissioned blockchains do not necessarily have to be private. As the ARCHANGEL project discussed below demonstrates, there are applications that would necessitate a permissioned blockchain that provides some limited public access.

Ethereum is another important example of a public, permissionless blockchain. While both Bitcoin and Ethereum are among the most globally-recognized public decentralized blockchains, Ethereum differs from Bitcoin in many ways. For one thing, Ethereum is managed by a group—co-founder Vitalik Buterin and the Swiss non-profit group, the Ethereum Foundation (Ethereum Foundation, n.d.-a). Ethereum is also focused on doing far more with its blockchain platform than trading its cryptocurrency, ether. Instead, Ethereum is concentrating on its role as a blockchain application platform (Ethereum Foundation, n.d.-a). Currently, Ethereum serves as the underlying blockchain for numerous decentralized applications (DApps) and various blockchain-based projects, including ARCHANGEL (Ethereum Foundation, n.d.-a; Collomosse, et al., 2018).

What makes Ethereum such a popular blockchain platform is the infrastructure it provides to new blockchain-based applications, which enables developers to write their own smart contracts. Ethereum, built in a Turing-complete language, was the first

blockchain built to specifically enable users to generate complex smart contracts (Ethereum Foundation, n.d.-c). Smart contracts are applications that directly embed the terms of a transaction into lines of code (Ethereum Foundation, n.d.-c). More specifically, the smart-contract is a type of code or algorithm that allows for doing more on a ledger than simply exchanging coins: the smart contract automates obligations and payments between agreed parties (Lemieux, 2017a). The contract to be executed—a payment on a particular date, given certain outcomes—is then appended to a timestamp in the blockchain sequence (Lemieux, 2017a). As a result, these smart contracts make it possible for the pre-arranged delivery of payments upon the outcome of an external event. Running on blockchain networks, smart contracts can be made fully or partially self-executing and/or self-enforcing (Lemieux, 2017a). Based on fulfillment of the terms of the contract, smart contracts govern which transactions are written onto the block, as well as the information it will contain (Lemieux, 2017a). It is important to note that despite its nomenclature, a smart contract is not legally binding on its own. Laws must exist in the non-virtual, outside world that validate the legality of a smart contract (Lemieux, 2017a).

While smart contracts expand the possibilities of what blockchains can be used for, they may also serve as a source of vulnerability for a blockchain platform. Because the smart contract is essentially code added to the blockchain, a talented hacker could potentially hack the smart contract code and attack the blockchain. One such example, the DAO attack, shows how flaws in a smart contract code, and not the blockchain itself, could lead to security issues (Lemieux, 2017b). In May 2016, the DAO, or Decentralized Autonomous Organization, was introduced by members of the Ethereum blockchain as a

decentralized, public venture capital platform (Falkon, 2017). The DAO was actually written onto the blockchain in the form of a smart contract code (Falkon, 2017). Consequently, anyone seeking funding for a project could then present it to the community via the DAO platform. Anyone with DAO tokens could vote on any project and invest their coins. If the project eventually became profitable, those who invested in the project would receive rewards. On June 17, 2016 a hacker was able to find a fault within the smart contract code that allowed him/her to steal 3.6 million ether, or roughly $70 million dollars, in an instant (Falkon, 2017). Fortunately, as this loot was held in a 28-day holding account, Ethereum developers were able to create a hard fork and send the hacked funds back to the original owners (Falkon, 2017). By then however, user trust in the DAO was destroyed and it was subsequently delisted from cryptocurrency exchanges (Falkon, 2017).

2.2 An Overview of TD Smith's evaluation framework

T. D. Smith's evaluation framework, or blockchain litmus test, seeks to provide potential stakeholders the criteria necessary to assess the overall utility of a blockchain-based project (2017). In other words, Smith's evaluation framework helps stakeholders to assess whether or not a proposed blockchain-based solution can successfully produce the immutable records it has been designed to produce. Smith (2017) notes that as the use of blockchain expands beyond the use of cryptocurrency into applications such as recordkeeping and big data, certain fundamental concerns such as maintainability and scalability of storage persist. While these issues are problematic for all data management applications, they are better understood on more established centralized databases. Moreover, while blockchain affords certain advantages regarding redundancy and

availability to an application, Smith (2017) notes blockchain also presents some major limitations regarding an application's means for change and growth.

To better evaluate these issues with regards to blockchain, Smith therefore developed his evaluation framework based on criteria derived from the fields of dependable and secure computing and archival science. Essentially, Smith's evaluation framework combines aspects of the framework found in Avizienis, Laprie, Randell, and Landwehr's (2004) "Basic Concepts and Taxonomy of Dependable and Secure Computing" with Victoria Lemieux's more recent Archival Theoretic Evaluation Framework (2017b). Avizienis et al. define the attributes required for evaluating computer systems based on dependability and security (see Figure 1), while Lemieux's framework develops a taxonomy that stems from the archival concept of trust (see Figure 2). As a result, Smith's framework for the evaluation of blockchain-based recordkeeping systems places utility at the top of the taxonomy and places three primary criteria below it: dependability, security and trust (See Figure 3).

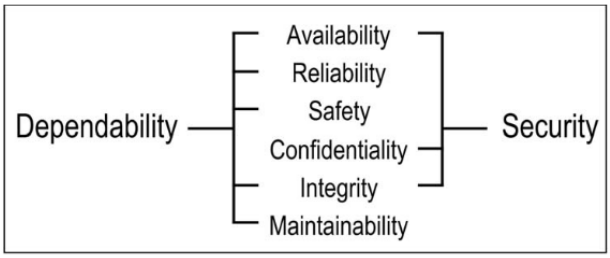

Figure 1. Attributes for dependable and secure computing. Reprinted from 'Basic concepts and taxonomy of dependable and secure computing,' by A. Avizienis, J.C. Laprie, Brian Randell and Carl Landwehr, 2004, *IEEE transactions on dependable and secure computing 1*, no. 1, p. 14. Copyright 2004 IEEE. Reprinted with permission.

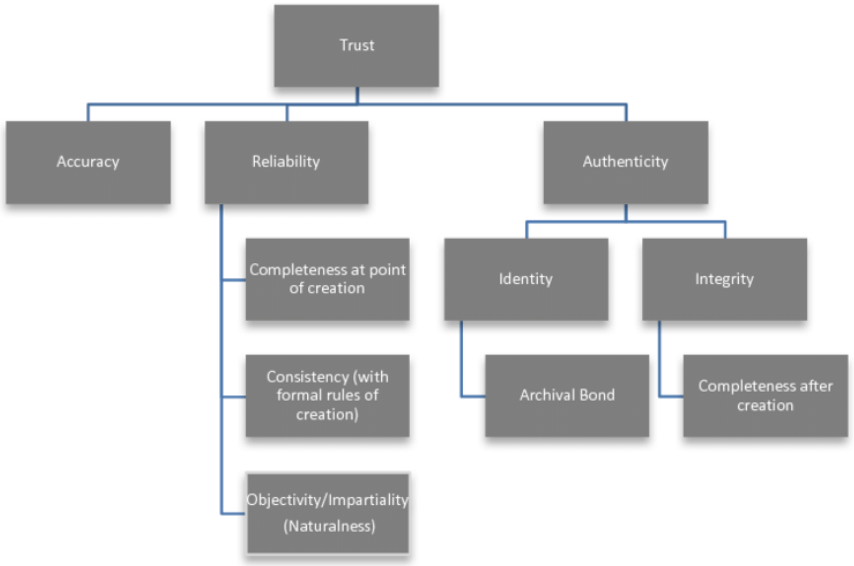

Figure 2: Lemieux's taxonomy of key archival concepts and their relationship to trust. Reprinted from "Blockchain and Distributed Ledgers as Trusted RecordKeeping Systems: An Archival Theoretic Evaluation Framework," by V. Lemieux, 2017, *IEEE Future Technologies Conference,* p. 3. Copyright 2017 IEEE. Reprinted with permission.

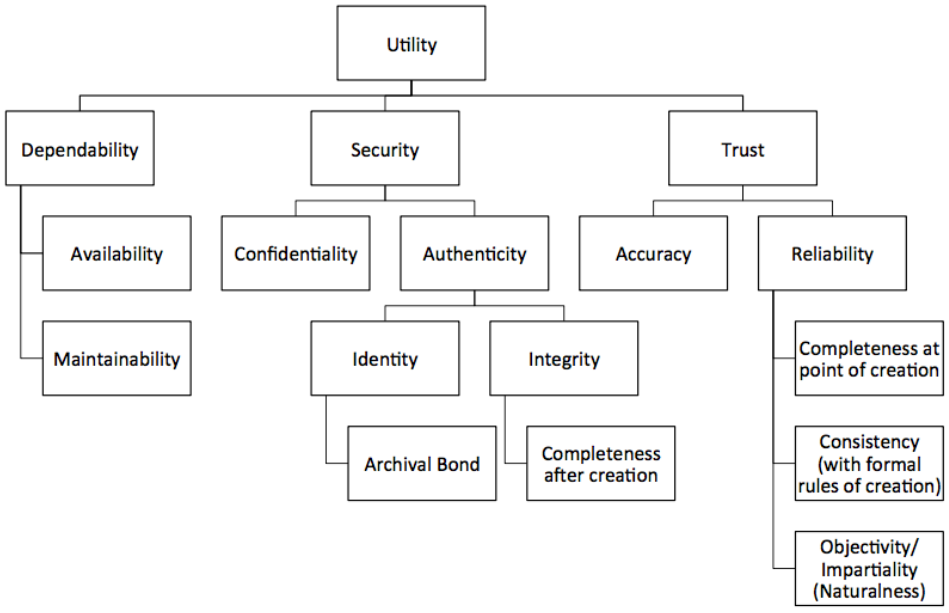

Figure 3: T.D. Smith's Taxonomy of utility concepts. Reprinted from "The blockchain litmus test," by T.D. Smith, 2017, *2017 IEEE International Conference on Big Data*, p. 2302. Copyright 2017 IEEE. Reprinted with permission.

Smith's placement of utility at the top of the taxonomy stems from his assertion that blockchain is in essence simply a data structure (2017). Noting that provided "enough investment an application using any data structure can be made secure"(2017, p. 2301), Smith suggests that the framework is best used to consider whether blockchain technology actually addresses the needs of the given application. In other words, the more terms from the taxonomy the blockchain can address, the better and more cost-efficient the blockchain application. For each term the blockchain application does not address, something else—another platform, program, system, etc.—addressing it will need to be provided, in turn adding more cost.

That being said, Smith does not proceed to parse out exact definitions of his usage for every term, rather explaining them more or less throughout his evaluation of the Bitcoin blockchain. This paper endeavors to explain each term through a combination of Smith's explanations and referrals back to the original source papers by Avizienis et al. (2014) and Lemieux (2107b). Given that Lemieux's work is especially relevant to the topic of recordkeeping and as it composes the greater bulk of Smith's taxonomy, the review of Lemieux's work is more extensive, meriting a section of its own.

Avizienis et al. (2014) consider the dependability and security of a system from the perspective of system faults. Recognizing that a system can and usually does fail, Avizienis et al. define a system's dependability in terms of the system's "ability to avoid service failures that are more frequent and more severe than acceptable"(2014, p. 13). In other words, how much system failure is too much? To that end, they argue that dependability can be characterized by the following five attributes: availability, reliability, safety, integrity and maintainability. Of those listed traits, Smith retains

availability—the "readiness for correct service" (Avizienis et al., 2014, p. 13)—and maintainability—the "ability to undergo modifications and repairs" (Avizienis et al., 2014, p. 13)—as elements of dependability. Smith (2017) eliminates safety as an attribute—"the absence of catastrophic consequences on the user(s) and the environment" (Avizienis et al., 2014, p. 13)—arguing that the blockchain systems he analyzed had no physical interaction with users and thus, pose no liability of direct physical damage.

Integrity—the "absence of improper system alterations" (2014, p. 13)—is also a required trait of security according to Avizienis et al. Integrity, along with availability and confidentiality—"the absence of unauthorized disclosure of information" (Avizienis et al., 2014, p. 13)—are conditions that must be met to ensure security. More specifically, within the context of security, availability means "the availability of authorized actions only" (Avizienis et al., 2014, p. 13) and integrity means the absence of unauthorized system alterations. Smith does include both confidentiality and integrity under the security branch of his taxonomy, but for reasons that will become more evident in the section below, he defers to Lemieux's more nuanced understanding of integrity.

2.3 Victoria Lemieux's Archival Theoretic Framework

Among the forefront of academics researching blockchain, Victoria Lemeiux was one of the first to produce an in-depth academic study of blockchain applications for recordkeeping. In her unpublished study, "Blockchain Technology for Record Keeping: Help or Hype," Lemieux (2016a) determined that blockchain solutions for records management were overhyped. While she notes that there was a lot of focus on and subsequently great potential for blockchain to provide increased transparency, more privacy protections, improved efficacy, current solutions were not meeting those desired

goals (2016a). Lemieux argued that part of the problem had to do with the lack of critical studies and evaluation metrics of blockchain solutions (2016a). Without such tools how could one properly assess and implement blockchain solutions? In particular, Lemieux demonstrates this by pointing out the irony in the fact that no developers of blockchain-based recordkeeping had actually consulted with archivists, records managers or academics well-versed in the archival sciences (2016a). As a result, their solutions for the long-term archival preservation of trustworthy records were designed with no awareness of archival theory or even best practices and standards for records management (2016a). Thus, at the time of her study, Lemieux concludes blockchain based solutions for records management were mainly hype and that in order to truly understand and actualize the potential of blockchain-based recordkeeping systems, more archival-science based research had to be done.

Lemieux has subsequently published a series of articles providing more detailed research into the use of blockchain solutions for recordkeeping. Her studies include a case study of the government use of blockchain technology for the recording and transferring of land titles (2017c), a risk analysis of blockchain-based land-registry system using records management and digital preservation standards (2016b), a co-authored study with Manu Sporny considering how to establish the archival bond in blockchain-based recordkeeping systems (Lemieux and Sporny, 2017), and a typology of blockchain records-keeping solutions currently available (2017a). From this original research, Lemieux has designed the evaluation metric she originally cited the need for in her earlier article “Blockchain Technology for Record Keeping: Help or Hype” (2016a).

In her paper, “Blockchain and Distributed Ledgers as Trusted Recordkeeping Systems: An Archival Theoretic Evaluation Framework” (2017b), Lemieux applies archival science to generate a preliminary framework to assess the capability of blockchain-based recordkeeping systems to produce trustworthy immutable records. More specifically, Lemieux (2017b) lays out an archival theoretic framework based upon the theory and principles underlying trustworthy recordkeeping, which she then uses to analyze a generic blockchain recordkeeping reference architecture and operating model.

Lemieux’s framework is built upon archival science’s three criteria for trustworthiness: accuracy, reliability and authenticity (2017b). Pulling from archival sources—mainly the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) terminology database (Pearce-Moses et al., 2017) and the Society of American Archivist’ glossary—Lemieux defines each criteria and its related attributes, further diagramming a taxonomy of these archival concepts and attributes in relation to trust (see figure 2 above). While these definitions provide some clarity regarding what constitutes trustworthiness (at least in the archival sense), Lemieux notes that assessments of a record’s trustworthiness are often probabilistic, as humans must evaluate these criteria using often incomplete and sometimes uncertain information regarding the record and its origins (2017b, p.2).

Of the three criteria, accuracy is perhaps the simplest to define. Fairly similar to common understandings of the term, accuracy relates to how precise, correct, truthful, and pertinent the contents of records are, or in other words, how well the records reflects reality. Lemieux defines accuracy concerned with “the truth-value of the contents (facts) of the record” (2017b, p. 3).

Reliability, as the chart demonstrates, is a bit more complex as it is composed of three attributes: completeness at the point of creation, consistency with formal rules of creation, and "naturalness" (Lemieux, 2017b). Completeness at the point of creation speaks to the transactional aspect of a record, namely its ability to effect consequences. More specifically, Lemieux posits that for a record to have completeness, it must have all the necessary components required by the record creator and a legal-administrative system to enact whatever transaction the record is about (2017b, p. 3). Lemieux cites as an example a sale contract for land—without the requisite signatures and date, the contract is not complete. In archival science, therefore, completeness is considered an innate characteristic of a record related to its formal aspects (2017b, p.3).

Consistency refers to the document's consistency—or similarity—with other authentic documents created of similar provenance according to Lemieux (2017b). More specifically, the record should share the same physical and formal elements—ink from that time period, computer font that does not postdate the document, contemporaneous language and style—of other authentic records from that time period (Lemieux, 2017b).

Naturalness refers to the nature of the record's creation, particularly how deliberately the record was created (Lemieux, 2017b). Records, unlike books or other types of publications, are not created to disseminate knowledge as end products in and of themselves. Records are created more as byproducts of daily business or life processes. As a result, Lemieux further explains that records "possess qualities of unselfconsciousness that underpin their reliability as records" (2017b, p. 3).

The final criteria, authenticity, is the quality of the record that verifies that it is in fact what it claims to be and that it has not been corrupted, altered and/or falsified in any way (Lemieux, 2017b, p. 4). To be considered authentic then, the record must demonstrate that it was created by the entity represented as its creator. To that end, a signature—physical or digital—both identifies the creator while also validating the record as the product of the creator (Lemieux, 2017b). It is important to note, as Lemieux does, that authenticity of the record has no bearing on the truth-value of the content, as "it merely establishes that the purported creator of the record is genuine and that the creator possesses the authority to make the record" (2017b, p. 4). Thus, it is very possible for a news paper article to be authentic—it can be proven that it was written by a journalist and approved and published by the journalist's newspaper—yet have false content—the journalist did not verify their facts, was given false information, or lied.

Further elaborating on authenticity, Lemieux (2017b) discusses identity and integrity—the two required preconditions for establishing authenticity. Identity refers to "the whole of the characteristics of a document or record that uniquely distinguish it from any other document or record" (Pearce-Moses, 2017). In other words, identity is the sum of all aspects of the document or record that establish the record to be uniquely and authentically itself, differentiated from other similar copies or forgeries.

As Lemieux explains further, identity in the archival context hinges upon proving and maintaining the archival bond (2017b, p.4). The archival bond refers to the relationship a record has to the event or activity it serves to document, to other records documenting the same activity, and to the individual who saved it as a record (Lemieux, 2017b, p. 4). Through the archival bond a record is connected to its "specific context of

creation and use" (Lemieux, 2017b, p.4) as well as to other records from the same archival aggregation—i.e. records that emerged from the same context. It is impossible to tell whether a record is genuine or forged without examining the archival bond, which in turn is usually ascertained through studying the record's provenance.

Aside from confirming the identity, the integrity of the document must be verified in order to establish authenticity of a record over time (Lemieux, 2017b). For a record to maintain its integrity means it has not been altered, corrupted, or tampered with over time. To ensure the integrity of a document, a system is devised to monitor and record the chain-of-custody of a record. In terms of digital archiving, this means a whole series of measures related to operation and infrastructure of recordkeeping information systems. As Lemieux (2017b) writes, some processes that ensure integrity include "access controls, user authentication and verification, audit trails, as well as documentation that demonstrates the normal functioning, regular maintenance, and frequency of upgrades or records systems" (p. 4).

Lemieux (2017b) further notes that an important aspect of integrity is the archival concept of completeness after creation. This relates to both the record's physical integrity and interpretability over time (Lemieux, 2017b). Within the context of digital preservation, this means that preserving the bit structure of data is insufficient if interpretability and accessibility are lost. Because bit level preservation often incurs semantic loss, the digital curator must also ensure that enough contextual information is preserved with the record in order to allow current and future users to understand and access it (Lemieux, 2017b). As Lemieux (2017b) explains, while one might be able to preserve the bit stream of a record and even the software needed to render the bitstream,

if the contextual information necessary to understand the record is missing, then the record is no longer a valid record. That is, a record that is not interpretable and/or remains inaccessible can no longer produce its "real world effect," whether that was to serve as proof of land transfer or as a certificate of ownership or identity, etc. (Lemieux, 2017b, p. 4-5). Therefore, digital preservation of records, as Lemieux (2017b) succinctly concludes, "involves preservation of the integrity of the identity of records, through preservation of the archival bond, in addition to preservation of the integrity of the general semantic context, content and form of data" (p. 5).

3 Methods

This study applies the criteria found in T. D. Smith's blockchain evaluation framework to two blockchain-based recordkeeping projects: ARCHANGEL and RecordsKeeper. Using publicly available information, each project was assessed and rated according how well these recordkeeping systems address the evaluation framework's three primary criteria—dependability, security and trust. Ratings range from Low, Medium and High indicating that the criteria was not addressed at all, partially addressed or fully addressed, respectively (See Table 1 in Discussion section).

Though a variety of blockchain-based platforms devoted to recordkeeping have emerged, the platforms chosen for this study—ARCHANGEL and RecordsKeeper—exemplify two primary recordkeeping applications. The ARCHANGEL project represents one of the first attempts by an archival institution to solve a recordkeeping issue related to public records management by archives and memory institutions. In contrast, the RecordsKeeper platform is a private company's attempt to address business and administrative electronic records management needs for organizations and individuals.

More specifically, the ARCHANGEL project focuses on preserving the provenance and integrity of public digital documents curated by archives and memory institutions (Collomosse et al., 2018). While archives and memory institutions would primarily add records to the ARCHANGEL blockchain, other public users would be

allowed to use the platform to authenticate the integrity of public documents whose hash record have already been entered on the platform. The ARCHANGEL project is a collaboration between a trusted government archive, an academic institution, and an independent non-profit organization—The National Archives, London, UK, the University of Surrey, and the Open Data Institute, respectively (Collomosse et al, 2018).

RecordsKeeper, on the other hand, is a private company managed by its two founders—blockchain developers Toshendra Sharma and Rohendra Singh—and other blockchain developers and experienced marketers (RecordsKeeper, n.d.-b). RecordsKeeper is a blockchain-based recordkeeping solution for businesses and individuals to use for their various electronic records management needs (RecordsKeeper, n.d.-c). Aside from the comparative value of these two different systems, at the time of this study's writing, no previous evaluations of this type had been conducted on these two blockchain-based recordkeeping solutions.

This study used publicly available information regarding each platform to make the evaluations. The primary source of information for ARCHANGEL consisted of a four-page summary of the ARCHANGEL project written by Collomosse et al. (2018) and blog posts posted by project partners (Keller 2018a, 2018b). The RecordsKeeper white paper (RecX Technologies Limited, n.d.) and website (RecordsKeeper, n.d.-a, n.d.-b, n.d.-c,n n.d.-d, n.d.-e) served as primary sources regarding the RecordsKeeper platform. For more technical information regarding the MultiChain stream technology underlying the RecordsKeeper platform, this study referred to blog posts on the MultiChain website (Greenspan, 2018, 2016; MultiChain, n.d-a, n-d-b.).

Ratings are based on the how well each platform addressed Smith's three evaluation criteria: dependability, security, and trust. Each of the three criteria is further composed of two major traits—availability and maintainability (dependability); confidentiality and authenticity (security); and accuracy and reliability (trust). In order to determine each platform's performance with regard to each listed trait as defined in Smith's framework, this study examined each platform's proposed use cases, intended participants, consensus protocol, blockchain type, technical features, and other available information. If a platform was able to address both traits of one criterion efficiently, then it earned a 'High' ranking for that criterion. If a platform was only able to address one trait or could partially address both traits, then it earned a 'Medium' ranking for that criterion. If the platform was unable to sufficiently address both traits, then it earned a 'Low' ranking for that criterion.

4 Research Results

4.1 ARCHANGEL Assessment

Initiated in 2017 by The National Archives (UK) in partnership with the University of Surrey and the Open Data Institute, ARCHANGEL is a blockchain-based decentralized platform focused on securing the long-term integrity of digital documents preserved in public archives and memory institutions (Collomosse et al., 2018). Aiming to guarantee trustworthiness through distributed ledger technology (DLT)—as opposed to institutional reputation as archival institutions currently do—ARCHANGEL "cryptographically guarantees the provenance, immutability and so the integrity of archived documents" (Collomosse, et al., 2018, p.1). ARCHANGEL seeks to record the digital signature, or hash, of the digital documents onto the blockchain, along with the relevant accompanying metadata to aid in the identification and verification of said documents. Thus, while the digital public documents themselves would not be preserved in ARCHANGEL—instead remaining under the stewardship of the archive or memory institution—the hash recorded on the blockchain provides a public record of the document's provenance and allows one to verify the integrity of the document over long periods of time (Collomosse et al., 2018).

Collomosse et al. (2018) propose to use a permissioned blockchain, which enables them to designate which users are able to perform various activities on the network. As such, Collomosse et al. (2018) explain only approved archives and memory institutions

would be authorized to append the blocks to the blockchain, but the public would still be able to access and read the blockchain, as well as verify transactions. More specifically, the archivists (or approved agents) responsible for depositing documents into the archive or memory institution would be authorized to append the ARCHANGEL blockchain with the document's content hash upon the moment of the document's deposition into the archive (Collomosse et al., 2018). Likewise, these permissioned agents would be able to update the blockchain if any authorized changes were made to the document—i.e. redactions, etc.—thereby creating a transparent audit trail (Collomosse et al., 2018). By still enabling the blockchain to be publicly readable, moreover, the ledger remains transparent, allowing anyone to openly authenticate digital objects released from an archive at any time (Collomosse et al, 2018).

Using the ARCHANGEL platform, a hash of the digital document is extracted upon deposition of the document (Collomosse et al., 2018). A file format identification tool examines the binary information within the file to determine the file format (Collomosse et al., 2018). A content hashing algorithm is then used to extract format-dependent hash or content hash from the document. For now, ARCHANGEL uses the classic binary hashing algorithm, the SHA-256 (Collomosse et al., 2018). Once the content hash has been generated through this process, the hash, a document global unique identifier (GUID), and a unique identifier representing the content hashing process, along with any supporting metadata provided by the archivist, will be appended in a new block at the end of the blockchain (Collomosse et al., 2018).

Though not developed yet, Collomosse et al. (2018) plan to eventually use algorithms (codes or models) customized to file formats in order to generate the content

hashes for the ARCHANGEL platform (Keller, 2018b). Once this is implemented, users will need to be sure that they use the same customized algorithm to generate the new content hash (Collomosse et al, 2018). In this case, the content hashes and algorithm hashes must match to accurately verify integrity. While Collomosse et al. (2018) do not provide a specific framework or model regarding the content of the supporting metadata, they do suggest including the archivist's notes, deposition date, versioning information, and for customized content hashing, the algorithmic hash of the code or model used to extract the content hash (p. 2).

Having opted to implement ARCHANGEL on the Ethereum platform, ARCHANGEL must use the same proof-of-work protocol as Ethereum to append and validate transactions to the ledger (Collomosse et al., 2018). In accordance with their permissioned DLT model, Collomosse et al. (2018) suggest two means of consensus checking for ARCHANGEL. In the first method, the ledger is maintained through proof-of-work performed among a private set of nodes sustained collectively by multiple archives and memory institutions, ideally from different disciplines and nations (Collomosse et al., 2018). In the second method, the ledger is maintained through proof-of-work that is performed across a public, globally-maintained blockchain, such as Ethereum, by miners of the public blockchain (Collomosse et al., 2018). Collomosse et al. (2018) explain that a smart contract granting authorization to write is used in this case to append data, provided that all conditions of the smart contract are met. Once the conditions are met, a secret key grants entry to the smart contract at its end-point (Collomosse et al., 2018).

At any point in the future, as Collomosse et al. (2018) explain, if anyone wishes to authenticate the provenance and integrity of a document curated or released by an archive or memory institution, they must rehash the content hash from the same copy of the document and then compare it with the original hash securely stored within the ARCHANGEL blockchain. In order to find the appropriate data block with the original content hash created upon deposition, the user may search through the publicly available contents using the GUID, the content hash, and the metadata (Collomosse et al., 2018). Once the appropriate data block has been identified, the hashes can be compared. According to Collomosse et al. (2018), if the hashes match, then the document's integrity has been verified. If the hashes do not match, the integrity has been compromised. In situations where the changes were required for legitimate reasons (i.e. redacting sensitive information, preservation purposes, etc.), the archive will then append the ledger documenting those changes by recording the new content hash of the altered version along with information and metadata detailing who made the changes, when the changes were made and possibly why (Collomosse et al., 2018).

Dependability

Availability of the ARCHANGEL blockchain depends on its ability to maintain a robust network of nodes to continually validate updates to the blockchain. Given that they have chosen to use the expensive proof-of-work protocol to establish consensus, the costs of maintaining a node on the ARCHANGEL blockchain seem rather cost prohibitive for archival institutions and/or independent scholars. Though Ethereum does not consume as much energy as Bitcoin, it still uses as much energy per year as the entire country of

Angola (Digiconomist, n.d.-b). One transaction—recall multiple transactions fit into one block—requires an estimated 40 KWh or enough to power 1.35 U.S. households per day (Digiconomist, n.d.-b). At an average price of $0.10 per KWh, one transaction could cost up to $4.00 to maintain.

The National Archive, which holds records from the last 1000 years, 5% of which have been digitized, could produce millions of transactions alone, just by seeking to make hash records of their digital documents (The National Archives, n.d.-b). A quick search of three of the online collections out of 66—the Royal Navy ratings' service records 1853 (700,000 records), the Royal Marines service records (~110,000 record), and the British Army Medical index cards 1914-1920 (5 million records)—yields 5,810,000 records (The National Archives, n.d.-a, n.d.-c, n.d.-d, n.d.-e). This does not include records from the UK Government Web Archive, which includes a Video Archive (at least 10,000 videos), Twitter Archive (at least 100,000 posts), and approximately 500,000 archived websites (U.K. Government Web Archive, n.d.-a, n.d.-b, n.d.-c). This simple estimate of records alone generates 6,420,000 hash records to be transacted and stored on ARCHANGEL's blockchain. At a cost of $4.00 per transaction paid strictly for energy consumption—not including transaction fees, cost of computing, or other associated expenses—this can add up fast.

The costs would continue to rise for other archives and memory institutions—particularly national archives joining ARCHANGEL—which would likely desire the addition of thousands of their own records to the blockchain. As these transactions become more blocks added to the chain, a robust network of miners will be needed to handle all the transactions. In the first proposed method, the archives and memory

institutions involved in project ARCHANGEL would provide private nodes that perform the proof-of-work verifying transactions appending blocks to the blockchain. In order to maintain a node on the network, archives and memory institutions that join will need to be well-resourced in terms of adequate computing hardware, funding to pay for electricity, adequate cooling systems for their mining equipment, and technically proficient staff. It is unclear, moreover, whether one node is sufficient for an archive or memory institution to join the blockchain, or if a minimum number of nodes are required to join. While some archives and memory institutions might be interested in ARCHANGEL for its ability to increase public trust in their records and practices, generally speaking, memory institutions are often underfunded and understaffed.[3] In his blog post critiquing the ARCHANGEL project, David S. H. Rosenthal also notes Collomosse et al.'s disconnect with the pragmatic realities of most archival institutions, writing that:

> These institutions are under severe budget pressure and competition for skilled staff. They are being forced to outsource their IT operations to "the cloud," and are unlikely to take on new or maintain existing in-house tasks. (2018)

Given how few memory institutions do not face the above-mentioned constraints, it is very difficult to imagine how ARCHANGEL will be able to recruit enough institutions to maintain sufficient active nodes to sustain the ARCHANGEL blockchain.

The second method whereby public miners perform the necessary proof-of-work to maintain the ARCHANGEL ledger requires more incentive to ensure a robust network.

[3] This was the dominant motif amongst participants at a stakeholder workshop held for the ARCHANGEL platform. The participants—13 experts from a variety of AMIs dealing with public documents—were instructed about the ARCHANGEL project and given time to interact with platform prototype for an hour in a lab-based setting at the University of Surrey. (Collomosse et al, 2018, p. 3).

At this time, there does not seem to be any incentive, such as earning cryptocurrency tokens or assets of any kind, for miners to mine. Collomosse et al. (2018) have mentioned that ARCHANGEL is still exploring potential business models to encourage sustainability, including the idea of having users seeking to verify public documents contribute mining effort as payment for the service (Collomosse et al., 2018; Keller, 2018b). This would require expanding smart contracts to enable search and verification of the content of the block in addition to the current write function (Collomosse et al., 2018; Keller, 2018b). It does not seem likely that there is enough market demand to pay for the authentication of public documents, at least, not enough to sustain the entire blockchain. As Rosenthal (2018) is quick to point out, moreover, scholars using such documents are not very likely to pay for such services.

Given its expense and the fact that permissioned blockchains do not require proof-of-work to maintain consensus, it is surprising ARCHANGEL has opted to use proof-of-work. In fact, other less expensive consensus models can be used, such as proof-of-stake or Byzantine Fault Tolerance protocols (Rosenthal, 2018). A proof-of-stake protocol is an algorithm for validating transactions that, unlike proof-of-work, determines the creator of the next block through a formula based on random selection and their stake, or how much coin they own in the blockchain (Blockgeeks, 2018). Nodes are not called miners, but forgers, as they no longer compete to acquire the next block (Blockgeeks, 2018). Forgers also do not get block rewards, but instead win the transaction fees associated with the block (Blockgeeks, 2018).

Implemented by IBM's permissioned blockchain Hyperledger, a Byzantine Fault Tolerance protocol is designed to resist faults or attacks so long as two-thirds of the

nodes are honest nodes (Hyperledger, 2018).. This generally works well in a permissioned ledger as participants have legal, business, or some other goal-oriented incentive to remain honest actors (Hyperledger, 2018). The Byzantine Fault Tolerance protocol is also easier and more efficient to implement in a network with a lower number of nodes (Hyperledger, 2018). In his blog, Rosenthal (2018) argues that ARCHANGEL's first model for consensus based upon a permissioned ledger of only memory institutions would be much better implemented on Hyperledger where the Byzantine Fault Tolerance protocol would provide a much more efficient and inexpensive consensus protocol.

Availability also hinges upon the participating memory institutions and archives responsible for off-site storage of the actual documents. ARCHANGEL relies upon the archives and memory institutions to preserve the actual records correlating with the content hashes being preserved on its blockchain. Preserving a content hash means very little if the document the hash is meant to refer to is not available. Should the repository fail to preserve these documents for whatever reason—i.e. error, malware attacks, damage to the physical hardware, etc.—or if the repository chooses to quit the project, then these documents would no longer be available to the system. This would constitute a system failure. Thus, system availability also depends upon the sustained preservation of these documents in the off-chain archives and institutional repositories.

In terms of maintenance, the platform interface will likely require upgrades as they continue to develop the platform and its functionality. A recent blog post by the Open Data Institute's J. R. Keller (2018b) announced plans to develop the ARCHANGEL user interface in order to make it easier for both the archival institutions and public citizens to use for their varying purposes. Some future plans also include

creating more specialized hashing for particular file formats—i.e. PDF files, image files, video files, etc. (Keller, 2018b).

It is unclear if the new hashing algorithms will be applied retrospectively to documents that have already been entered onto the blockchain initially with the binary SHA-256 algorithm. If so, a central question becomes how those changes will be reflected in the ARCHANGEL platform, given that blockchains are an append-only structure. In other words, how does ARCHANGEL reflect any changes to records entered into the system? Does it produce an audit trail to reflect changes made to the public record, either in terms of applying a new content-specific hash via the ARCHANGEL platform, or in the more common instance of an archivist making necessary updates (i.e. redactions, format changes for system migration, etc.) to the record held in the repository? Will users be able to use the GUID to connect all transactions regarding a particular record on the blockchain, or will a new GUID be issued for every single transaction involving the same document? Brief mention was made about adopting W3C PROV standards for document versioning, but it is again not clear exactly how that would be integrated into the system. While no clear answers to these questions have been provided in Collomosse et al.'s paper (2018), it is fair to assume that someone will be in charge of maintaining aspects of the platform such as the interface, permissions or memberships, should membership fees be charged. It is not clear who would do this maintenance work and how they would get paid.

Given the cost-prohibitive, inefficient, and basically unnecessary use of proof-of-work consensus protocols it seems highly unlikely ARCHANGEL will attract enough archives or memory institutions to sustain the network. Similarly, without any incentive

for public miners to mine for blocks, it seems unlikely that they will attract other miners to support their network. Reliance on off-chain storage of documents in repositories can negatively impact availability as well. With regards to maintainability, very little information regarding how changes to document hash records has been mentioned. Without much of a financial plan it is also difficult to understand who will continue to maintain the entire ARCHANGEL platform in the near future or over time. As such, ARCHANGEL earns a low rating for dependability as neither availability nor maintainability has been sufficiently addressed.

Security

Confidentiality of user identities is maintained through the public-private key protocol of blockchain. This works well when the mining network is composed of various miners on a public blockchain. In the case of the private, permissioned blockchain constituted of archives and memory institutions however, it does not. Generally, in permissioned blockchains participant identity is not private, as all participants would need to be identified just to attain membership and permissions. Moreover, given that archives and memory institutions would be the only entities that can append records to the blockchain—and would be doing so in high volume and frequency—it seems that their user identity would be easy to uncover. Indeed, it would be desirable to know who adds the digital signature for a document, as this demonstrates the provenance of the hash record. The entity adding the record should be the institution that actually holds the document. Otherwise, a malicious actor could falsify a hash record. Moreover, being

publicly identifiable increases the transparency of the institution's practices, further enhancing the trust in record.

While transparency of user identity is likely not an issue, given that this is a permissioned blockchain, it is imperative that the authorizations—whatever protocols and mechanisms are given to the archives and memory institutions to authorize them to add content data to the chain—remain confidential. Currently, it is unclear what mechanisms and protocols would be used to authorize archives and memory institutions to commit blocks to the chain in the permissioned model. If using smart contracts, the code is public so care must be taken to ensure that the smart contract reveals nothing that bad actors could exploit.

The transaction information recorded in each block is meant to be publicly viewable and thus, no private or sensitive information should be included in these blocks. This would follow similar protocols archivists follow in the creation of finding aids for their records. The actual public records themselves are held off-chain within the archives, so access to them would be handled through the archive. For public records that cannot be publicly released until after a certain time period (i.e. classified documents), ARCHANGEL could be used to provide a transparent record of the provenance while keeping said documents confidential. Because the document's hash could be checked upon its eventual release to the public, its integrity could be verified.

Identity of the content hash records stored on the blockchain is essentially predicated on the included document GUID, metadata, and transaction information. The document GUID represents the document from which the content hash was derived, thereby linking the document with the content hash in the system. The metadata might

also provide a way of linking the content hash to its document. The block transaction information can provide information regarding the memory institution that submitted the hash record to the blockchain. However, as accuracy of records cannot be guaranteed—the wrong GUID or wrong metadata might have been entered—neither identity in this case. No apparent means of preserving the archival bond are available.

Aside from being cost-prohibitive, the proof-of-work consensus protocol adopted by the ARCHANGEL project, despite all the hype, is not tamper-proof. As mentioned above, 51% or double-spend attacks are becoming more prevalent, particularly for newly forming blockchains like ARCHANGEL with smaller networks. Of the two proposed methods for building consensus, the permissioned network of authorized archives and memory institutions would probably be more secure than the network of miners on a public, globally maintained blockchain. In the first method, only permissioned archives and memory would be allowed to join and append records. The membership process would likely weed out bad actors. Moreover, one would assume that the institutions joining would have some stake in ensuring system integrity is maintained. Tampering with the blockchain would cast doubt on the records of their own institutions and records. Collomosse et al. (2018) also argue that incorporating diverse archives (government, university, etc.) from various nations operating independently under their own suitable governance structures would provide a further check against collusion, as coordinating such an attack against diverse actors would require immense effort (p.3).

The latter mode of using smart contracts and public blockchain with a large network of independent miners can be exploited in two ways. The first is by hacking or exploiting a loophole in the smart contract code as was exemplified by the DAO contract

exploits mentioned earlier. The second remains, of course, through 51% attacks. Collomosse et al. (2018) mistakenly suggest that proof-of-work consensus on such a well-established public blockchain like Ethereum is tamper proof because the network of nodes is so large and diverse that collusion is difficult to arrange. Though the recent successful spate of 51% attacks have been on blockchains with networks much smaller in size than Ethereum, the fact of the matter is that Ethereum has suffered double-spend attacks in the past (Smith, 2017). Fortunately, Ethereum's infrastructure was strong enough that the attack was countered in time with hard forks (Smith, 2017). Even so, as Rosenthal (2018) notes, most of the large public blockchains have mining pools—collectives of nodes—that do not operate independently. Ethereum has multiple mining pools, with the largest 3 pools—ethpool/ethermine (24.7%), f2pool (21.6%), and dwarfpool (13.3%)–controlling about 60% of the hashrate (Tuwiner, 2018). Thus, proof-of-work, particularly in this instance, is not a guarantee of tamper-proof blockchain records at all.

While much importance is given to securing the immutability of transactions in a block, the integrity of the entire blockchain relies upon its replication in multiple nodes (Lemieux, 2017a, Rosenthal, 2018). Each node is technically supposed to be carrying a full copy of the digital ledger as it continues to validate transactions. As Lemieux (2017a) notes, the tricky thing with a blockchain of hashes is that multiple copies of the entire ledger MUST exist in at least 2 nodes, though hopefully more. As Lemieux explains, "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" (Lemieux, 2017a). In

other words, the entire notion of the decentralized ledger works, assuming that if one of the nodes fails and thus, does not retain a full copy of the ledger, the many other nodes in the system will have full copies. Because there is no guarantee that every node will contain a full copy of the digital ledger, the more nodes in the system, the greater the probability that multiple full copies of the ledger survive (Lemieux, 2017a).

Another potential risk to the physical integrity of records is the storage of the actual public records off-chain in their respective repositories. As Smith (2017) notes, "Attacking the blockchain directly is computationally difficult but any application that uses 'off-chain' resources runs the risk of lost data or value to malware or storage failure" (p. 5). Archives and memory institutions have the advantage of being singularly focused upon the preservation of their records. This means they have at least some level of archival infrastructure—trained archivists and/or records managers, practices, equipment, etc.—in place to preserve their records according to the best possible standards. Nevertheless, these institutions are not invulnerable to accidental loss or malicious damage of records.

In the end, the hashes stored on the blockchain hold very little meaning if the original documents they were to be compared to are not also preserved. Any lack of preservation would negatively impact completeness after creation. In terms of the interpretability of records, the GUID and metadata provide the information and context necessary to connect the content hash to its related public document. Providing the content hashing algorithm would further provide information necessary to accurately verify the content hash, once ARCHANGEL is modified to utilize more customized

hashing algorithms for documents. However, if the document the content hash refers to is not available, then the hash record loses any real world effect.

Confidentiality is not particularly relevant in the permissioned blockchain model, and available in terms of public-private key identity encryption for the public blockchain model ARCHANGEL proposed. Document confidentiality is managed by the archival and memory institutions, so again not under the purview of ARCHANGEL. Authenticity is difficult to guarantee in ARCHANGEL, as there is no means to provide the archival bond. Moreover, ARCHANGEL's proof-of-work consensus model does not guarantee the physical integrity of its records. Finally, off-chain storage of the documents in separate repositories increases risk to the physical integrity of those documents, which in turn does not allow for completeness after creation. ARCHANGEL therefore earns a low rating for security.

Trust

Accuracy and reliability are always difficult to manage in blockchain systems. Even though ARCHANGEL proposes a permissioned model, wherein only records professionals would be entering the hash data, there is still a chance that a record might have inaccurate data. In particular, the archivist might make a mistake in the metadata fields of the hash record, entering a typographical error, inaccurate information or incomplete information. Similarly, it is unclear from the ARCHANGEL documentation how the document GUID will be determined and if its inclusion in the record is automated or not. Nevertheless, a failure in either producing truly unique document GUID or in ensuring that the accurate document GUID is included in the hash record

would also produce inaccurate records. These inaccuracies would certainly negatively affect a future user's ability to search and/or identify hash records for particular public documents.

The use of smart contracts can also produce inaccuracies with records. As Lemieux (2017b) points out there is always the possibility of incongruity between what the smart contract code was intended to do and what is actually executed. The DAO exploit was in fact a matter of a loophole that was exploited by hackers. It is possible that if not tested and proofed enough a smart contract might yield unintended results through loopholes in the code.

Reliability is out of the scope of ARCHANGEL since the content hash records alone cannot prove the authenticity of the records they were derived from. Rather, these hash records will need to be compared to another content hash of the same record at a future date to verify the authenticity and accuracy of the document.

As both accuracy and reliability cannot be guaranteed by ARCHANGEL, the platform earns a low ranking for trust.

4.2 RecordsKeeper Assessment

With the stated vision to "create a global open ecosystem for data sharing and verification," RecordsKeeper promises to provide users the ability to store and verify records of any data object through their platform (RecX Technologies Limited, n.d., p. 5). Unlike ARCHANGEL, RecordsKeeper, founded in November 2016 by Toshendra Sharma and Rohendra Singh, is a private company seeking to serve the recordkeeping needs of organizations and individuals (RecordKeeper, n.d.-b; RecX Technologies

Limited, n.d.).[4] RecordsKeeper seems mostly to be oriented towards businesses, as is evident in their suggested use cases for the platform: insurance record, health record manifests in judicial proceedings, enterprise know your customer (KYC) needs, employee verification, corporate compliances, land ownership records, government regulations, trustless file sharing, verifying academic certifications, and supply-chain management (RecordsKeeper, n.d.-e; RecX Technologies Limited, n.d., pp. 16-21).

Built upon MultiChain technology, the RecordsKeeper platform is meant to operate as an open-source, immutable public global database enabling users to store both the hashes of the records as well as the records themselves (RecordsKeeper, n.d.-c). The RecordsKeeper platform, moreover, promises out-of-the-box functionality, thereby allowing users to seamlessly integrate the platform into their recordkeeping workflows (RecordsKeeper, n.d.-c).

There are essentially four main aspects to the RecordKeeper platform: network nodes, JSON-RPC APIs, MultiChain streams, and the RecordsKeeper blockchain. In order to use RecordsKeeper each user—be it an individual or organization—must first set up a RecordsKeeper private node (RecordsKeeper, n.d.-c). Once the RecordsKeeper node is set up on the user's local or cloud infrastructure, the user can also then set up a RecordsKeeper XRK Light Wallet (RecordsKeeper, n.d.-c).. Through the Light Wallet, users obtain XRK tokens, RecordKeeper's own cryptocurrency, which must be used to pay for blockchain transaction fees (RecordsKeeper, n.d.-c; RecX Technologies Limited, n.d., p. 7). Users can also generate, store, and receive records from the RecordsKeeper blockchain using the Light Wallet (RecX Technologies Limited, n.d., p. 14).

[4] RecordsKeeper is according to the company's website and documentation a registered limited company by the name of "RecX Technologies Limited" in Gibraltar, based in Singapore with offices in India (RecordKeeper, n.d.-b; RecX Technologies Limited, n.d.)

After the RecordsKeeper node is set up on the user's computing infrastructure, the user can synchronize existing applications with the node and publish records on the RecordsKeeper Blockchain (RecordsKeeper, n.d.-c). The JSON-RPC APIs provided by the RecordsKeeper platform enable the existing applications on the user's node to issue notifications and calls to the rest of the RecordsKeeper platform, ensuring integration of the user's system with the RecordsKeeper (Morely, n.d.). Moreover, RecordsKeeper provides a wide array of APIs and open-source libraries to allow users to implement this seamless integration in various ways, including programming, websites, backend services, servers and mobile and desktop applications (RecX Technologies Limited, n.d., p.7).

To upload and publish records on the blockchain, the user must first acquire XRK tokens either through mining or purchase via fiat money (RecX Technologies Limited, n.d., p.7). The user then uploads their record (or data) in key-value pair format into the RecordsKeeper blockchain, paying the required 0.1XRK per KB upload fee. Only one key-value pair can be added at a time, though the same key may be used several times with different values—i.e. records/data (RecX Technologies Limited, n.d., p.7). The record key is used for retrieval and verification of the record later (RecX Technologies Limited, n.d., p.7). Once a miner confirms the transaction, the new block is added to the chain. The miner earns the XRK transaction fees as a block reward for mining (RecX Technologies Limited, n.d., p.7). After the record is published on the RecordsKeeper blockchain, it may be viewed and verified by authorized parties with whom the user has shared the record key or transaction ID (RecX Technologies Limited, n.d., p.7).

RecordsKeeper strongly discourages its users from uploading an actual file or record in its native format. Rather, they suggest using formats ranging from JSON, XML, Hex, Objects, or simple text (RecordsKeeper, n.d-a). Noting that files often contain a lot of unnecessary raw data that will increase the size of the transaction and its cost, RecordsKeeper suggests all files be rendered in these simpler formats (RecordsKeeper, n.d-a). For files larger than 5GB, RecordsKeeper recommends only uploading the record's hash (RecordsKeeper, n.d-a). Thus, not all records archived in the RecordsKeeper platform can be saved in their original format, unless their original format is one of the ones listed above. Depending on the use case, this may or may not be problematic. For example, if part of the records requires an image, the file will have to be converted to binary hexadecimal format first. In this case, the original record is not being stored by the RecordsKeeper system at all, but in an off-chain, off-system storage system.

Within the RecordsKeeper system, the transaction information, document hashes and other metadata regarding the uploaded record are saved on the blockchain, while the record is usually not. Instead, the record/data is moved to off-chain storage in the encrypted storage layer (RecX Technologies Limited, n.d., p.7). Built upon MultiChain technology, RecordsKeeper uses an abstraction layer—MultiChain streams—to provide secure, queryable, offchain, key-value data storage that sits on top of the RecordsKeeper blockchain (RecX Technologies Limited, n.d., p.7). Thus, while the blockchain is primarily used to support timestamping, notarization, and immutability, the Multichain streams enable the RecordsKeeper platform to perform as a database for storing records as well (MultiChain, n.d.-a).

Each stream functions as a separate append-only ordered list of items. The RecordsKeeper blockchain may have an unlimited number of streams. The data published on each stream is further archived by every node in the RecordKeeper network. By electing to subscribe to a stream, the node can then index the stream's contents to allow efficient data retrieval (MultiChain, n.d.-a).

Streams are generated individually on the blockchain through a distinct transaction that can only be signed by addresses that have been granted create permissions (MultiChain, n.d.-a). Stream creators automatically gain administrative, activation and write permissions for the streams they author. Moreover, no more than one stream can be created per transaction (MultiChain, n.d.-a).

Each stream item is also represented by a blockchain transaction that publishes the item to the stream (MultiChain, n.d.-a). Only one item can be added to a particular stream at a time (MultiChain, n.d.-a). Depending on whether or not the stream is created as an open stream or a closed one, publishing items to a stream may or may not require write permissions (MultiChain, n.d.-a). Open streams allow any address that has permission to send blockchain transactions to publish items in the stream (MultiChain, n.d.-a). A closed stream, on the other hand, requires that the stream item publisher must have write permission in order to validate the item and the transaction (MultiChain, n.d.-a). A closed stream will require one or more administrators who are able to manage and change write permissions as needed over time (MultiChain, n.d.-a).

Each item in the stream has four basic features: data, publisher(s), a record key, and block transaction information (MultiChain, n.d.-a). The data is the record—or whatever data—the user wishes to save in the RecordsKeeper database (MultiChain, n.d.-

a). This data may range in size from small bits of text to multiple megabytes of raw binary data. Each item must also be signed by one or more publishers (MultiChain, n.d.-a). The item also includes a record key (a number between 0 and 256 bytes in length) to aid in future retrieval of the item (MultiChain, n.d.-a). Lastly, transaction and block information—i.e. transaction identification number (transaction ID), timestamp, and blockhash—is taken from the header of the block in which the item is confirmed (MultiChain, n.d.-a).

MultiChain streams can be referenced through their transaction ID, streamref, or if available, a stream name (MultiChain, n.d.-a). The streamref encodes a combination of transaction information, specifically the block number, byte offset of the stream creation transaction and the first two bytes of the transaction ID (MultiChain, n.d.-a). Stream names are optional and must be selected at the time the stream is created. Special care must be taken that the stream name is unique on the blockchain—that it is not the same as any other stream or asset already on the blockchain (MultiChain, n.d.-a). Stream names are stored as UTF-8 encoded strings, case insensitive and go up to 32 bytes (MultiChain, n.d.-a).

Dependability

RecordsKeeper depends upon miners to validate its transactions. As it is still a new platform developing its network of miners, RecordsKeeper currently provides a permission-based consensus scheme with plans to shift to a proof-of-work consensus protocol once its network of miners is large and diverse enough to theoretically stave off a 51% attack (RecX Technologies Limited., n.d., p. 9, 12). In a permission-based

consensus scheme, only nodes that have been given mining permissions may participate in the consensus algorithm, while only the select nodes that have administrative (admin) permissions can add or delete all other permissions, including mining. In order to maintain the sense of a decentralized ledger that is not controlled by any one entity, the blockchain is set up so that no single administrator may change such important permissions as mining. Instead, a certain percentage of the administrators pre-defined in the blockchain parameters must agree to a permissions modification. Ideally, there would be multiple administrators from different organizations—i.e. multiple nodes—to make this work properly (MultiChain, n.d.-b).

RecordsKeeper will manage consensus through their permission-based consensus scheme until block number 4,204,800 in the blockchain (RecX Technologies Limited., n.d., p. 10, 12). After block 4,204,800 is verified, mining permissions will be granted to every node in the network and RecordsKeeper will move to adopt an open proof-of-work consensus model (RecX Technologies Limited., n.d., p. 10, 12).

RecordsKeeper relies upon their native XRK utility tokens to sustain and maintain their miner network regardless of the consensus model. Their primary token economy would revolve around users paying upload transaction fees using XRK tokens, which in turn are delivered to the miners who mine the block (RecX Technologies Limited., n.d., p. 15). RecordsKeeper further plans to reward users—both miners and initial adopters—XRK tokens periodically throughout the year. Businesses that are early adopters will also be granted advance permissions and other rewards (RecX Technologies Limited., n.d., p. 15).

While cryptocurrency rewards for mining and for early adoption may work to draw users into joining the new platform, it is unclear whether or not it will draw in and retain enough miners. One of the issues with XRK tokens is that they are utility tokens, as opposed to the more traditional cryptocurrency like bitcoin. Utility tokens are used for paying for the services provided by the company and thus, largely only have value within that platform or company (Camacho, n.d.). In other words, the XRK token can only buy services within the RecordsKeeper platform. Utility tokens can, however, increase in value through speculation in crypto exchanges—markets where people can buy and sell utility tokens (Camacho, n.d.). Namely, if an investor feels that a token like XRK will be valuable to many users (based on their confidence in the platform or project it supports), they can try to buy up many XRK tokens at a lower valuation and then try to sell them once their value increases (as their demand increases). While it is not possible to predict the future valuation of XRK tokens with absolute certainty, it would seem from the recent cancellation of the XRK token sale that the XRK is struggling to appeal to the cryptocurrency market (Sharma, 2018). In his cancellation announcement, RecordsKeeper co-founder Sharma (2018) cites poor marketing on their part, a recent overall drop of 80-90% in the cryptocurrency markets, and the recent spate of cryptocurrency scams as issues preventing the XRK from attracting miners.

Aside from this type of speculation, the only other major source of demand for the XRK token would be from users of the actual RecordsKeeper platform. Because the XRK tokens can be used to pay for record keeping services on the platform, the XRK still has some value for those platform users. Mining for tokens would, therefore, be a worthwhile incentive for these users. Thus, availability would be contingent primarily upon users that

actually use the RecordsKeeper platform for recordkeeping functions. As it is a relatively new platform, having just opened up record keeping functions in April 2018, it is difficult to discern whether the low numbers—a total of 279 records and 93 active miners—are typical of a new start up or forecast an unsuccessful road ahead (RecordsKeeper, n.d.e; RecX Technologies Limited., n.d., p. 24). While RecordsKeeper continues to operate using a permission-based consensus model, the smaller network of nodes might be enough to sustain the platform for a while. However, once a shift to a proof-of-work protocol is instituted, it seems likely Recordskeeper will require a larger network of nodes to sustain the blockchain and fend off 51% attacks.

As more businesses and nodes are added to their network, scalability might also become an issue in terms of availability and maintenance for RecordsKeeper. Given that each user will have multiple records and related transactions to store on the platform, storage might become an issue. Assuming that all users comply and store all records off-chain via MultiChain streams, the fact remains that along with the blockchain, every stream item belonging to the blockchain is stored by every node in the network as well. Using a binary format and LevelDB index, each node stores all the off-chain data in a special directory of the blockchain directory (Greenspan, 2018). A node will further generate a separate subdirectory of the items of each stream it is subscribed to and that it has created (Greenspan, 2018). Within the subdirectory the relevant stream data is duplicated once more. While retrieval time is greatly reduced by use of MultiChain streams, and the blockchain itself remains smaller by storing the records off-chain, it is unclear whether this distributed storage capacity of network nodes will successfully store the blockchain and all of the off-chain records archived by users of the network.

To more fully support the open source community, RecordsKeeper will also offer developers at large a fixed amount of XRK tokens for any code developed that improves the security, functionality, or design features of the Recordskeeper platform. Developers will be able to fork the codebase to make changes and upon developing, updating and testing a successful code improvement, also make a merge request (RecX Technologies Limited, n.d., p.15). If the request is approved, the developer will be rewarded with tokens. This small measure may improve the overall maintainability of the blockchain (RecX Technologies Limited, n.d., p.15).

RecordsKeeper relies upon the recordkeeping services of their platform and their XRK token to attract miners to their network. Unfortunately, as a utility token, XRK tokens are only useful to those using the token to pay for services within the RecordsKeeper platform. For miners not interested in using the RecordsKeeper platform for recordkeeping, XRK tokens may not hold much incentive to mine. There is no indication that the tokens will gain in value, or that even if they do, they will stabilize in value. Should RecordsKeeper shift to a proof-of-work model, they will need to build an even larger network of nodes. With regards to maintainability, it is unclear how RecordsKeeper would manage storage, as every node in the system saves all stream items in addition to the blockchain. This would be particularly problematic should RecordsKeeper expand to include many businesses and corporate clients storing thousands and thousands of records to the system each. RecordsKeeper provides no contingency plan in case their blockchain and stream data becomes too large for a node to store. As such, RecordsKeeper earns a low rating for dependability as neither availability nor maintainability has been sufficiently addressed by the system.

Security

In terms of confidentiality, data that is published on a MultiChain stream is not private unless it is first encrypted. Given that a stream's data is saved on every node in the network, effective read permissions for streams are not possible as the node could read the data through its disk drive. This might give records managers some pause, particularly those dealing with sensitive information, such as health records, social security numbers, banking information, etc. While RecordsKeeper provides two options for encryption of files below, the fact remains that RecordsKeeper does not automatically encrypt files for the user. Rather, users must do this encryption themselves. Ensuring encryption occurs with every upload would likely require an automatic aspect in their records management workflow. However, as mistakes may occur and Recordskeeper is a digital, append-only ledger, should a user accidentally forget to encrypt their files and still upload them to the RecordsKeeper blockchain, those files will become accessible even to those who have not subscribed to the stream.

Therefore, any sensitive or private data must be encrypted. RecordsKeeper enables two forms of encryption. In the first, users can encrypt their data through their own application layer prior to publication. Decryption keys are then shared with other authorized users. It is important to note that RecordsKeeper does not automatically encrypt records for users. Thus, the user must make sure that their application layer and workflow enables and ensures encryption of data.

The second method, which RecordsKeeper advocates for certain cases, is to use symmetric cryptography and a combination of three streams. The first stream distributes

the public-key for participating, permissioned users in the public cryptography scheme (Greenspan, 2016). The second stream publishes the segments of data, all of which are encrypted with a unique key via symmetric cryptography (Geenspan, 2016). Finally, the third stream enables data access to participating permissioned users (Greenspan, 2016). For each segment of data that a participant is permitted to see, a stream entry is generated containing the data's secret key, which in turn is encrypted with the user's public key (Greenspan, 2016). This method ensures that only authorized users may see the archived data on the blockchain (Greenspan, 2016).

For records managers dealing with sensitive information—i.e. health records, social security numbers, banking information—the fact that all record information stored in the stream would be accessible to every node through the node's disk drive might be cause for alarm.

In terms of the physical integrity of the transaction data, record hashes, and metadata stored on the blockchain, neither the permission-based consensus scheme nor the proof-of-work algorithm RecordsKeeper hopes to eventually adopt provide guaranteed immutability. In the permission-based consensus scheme, RecordsKeeper also employs mining diversity parameters to ensure no one can launch a 51% attack. Mining diversity parameters limit miners from overpowering the network or manipulating transaction data by prohibiting miners from mining a continuous series of blocks (RecX Technologies Limited, n.d., p. 10). RecordsKeeper employs a mining diversity factor of 0.2, which means that in a community of 100 miners, each time a miner mines a new block he/she would have to wait for 20 blocks to be mined before mining another block

(RecX Technologies Limited, n.d., p. 10). Thus, permissioned miners would be rotated within the algorithm.

The mining diversity parameters, a safeguard implemented to protect against the possibility of someone launching a 51% attack, might actually make it easier for miners to tamper with blocks as well (MultiChain, 2018a). Given that a node can possibly predict the next block it will mine, an attacker may use that prediction to tamper with transactions in two ways (MultiChain, 2018a). First, the attacker could use their node to willfully censor a transaction, thereby delaying its confirmation until the next block generated by an honest miner (MultiChain, 2018a). The second way can occur when two conflicting transactions are waiting to be confirmed (MultiChain, 2018a). Normally, the first of the two transactions to be confirmed by multiple nodes is validated, while the other is invalidated and orphaned (MultiChain, 2018a). However, in this case, the attacker can use the node to decide, based on preference, which of the two transactions ought to be confirmed (MultiChain, 2018a).

Another potential issue with the permission-based consensus scheme is its reliance on designated administrators to determine mining permissions. None of the documentation for RecordsKeeper clarifies exactly who the administrators are and how they attain administrative status. It is unclear from the RecordsKeeper documentation if admin status is relegated strictly to RecordsKeeper administrators or if users adopting RecordsKeeper for records management obtain admin status as well. If the latter is the case, would each client receive one node with admin status? Or would larger companies that claim multiple nodes in the network procure multiple nodes with administrative status? It is difficult to ascertain how the admin permissions would be attained or

distributed, and thus, difficult to understand how much influence a particular party might have to grant or deny mining permissions. It is also unclear as to whether there are any guidelines the admin must follow regarding granting or denying these permissions. Without more transparency regarding these admin factors, it is difficult to determine whether or not the system would be safe from collusion between a/an administrator(s) and the several nodes to which they grant mining permissions.

Because stream items are encoded into block transactions, they receive the same protections as any other blockchain transaction. This means that records that are published in streams are as immutable as the blockchain. Furthermore, each record published in a stream will exist in every node along the network. Every stream item on the blockchain will thus be saved by every network node, regardless of whether or not that node is subscribed to the stream. This process will ensure multiple copies of the record exist, ensuring preservation regardless of whether or not the original publisher node leaves the system or is corrupted.

With regards to completeness after creation, the renderability and interpretability of records may be compromised in RecordsKeeper, as records that are stored in the RecordsKeeper platform must adhere to particular file formats. If a company has all of their files in JSON, XML, or any of the other recommended file formats, then their files and data may be preserved in the original file format within the RecordsKeeper stream. However, if the original files are images, PDF files, or some other format not recommended for direct preservation into the RecordsKeeper stream, the files must be reformatted before being uploaded and published to the stream. This is problematic for a user seeking retrieval of the original electronic file.

For example, an image file could not be stored in RecordsKeeper in its original format, but would need to be rendered into a binary hexadecimal format (Multichain, 2018b). While the record key or transaction id would pull up the necessary information for a user to retrieve and verify that record of the image from the stream, the record would need to further indicate and/or provide whatever tools were necessary to render the binary hexadecimal into a viewable image. However, if an actual copy of the original image file is required, then the user must be able to access the off-chain, off-system repository where the original record is archived.

Currently, there is no indication in RecordsKeeper's documentation as to how original records might be accessed by other individuals. Moreover, there is no clear sense of how an audit trail would be generated between records saved in the preferred file formats on the RecordsKeeper streams and the original records in their original file formats that are preserved in their institutional repositories. Thus, for users whose files are originally formatted in simple text, Objects, JSON, XML, or Hex, RecordsKeeper does enable them to store, share, and verify their records with others through the platform. For others users whose records are otherwise formatted, RecordsKeeper is not exactly allowing them to store, share, and verify the original electronic record, but an altered version of it. Depending on the use case, this may entirely defeat the purpose of uploading and sharing records through the RecordsKeeper platform.

Though RecordsKeeper does not automatically encrypt records as they are uploaded, it does provide users the option to do so, either prior to upload or through the symmetric encryption offered through the streams. Oddly enough though, because stream data is stored on every node, each node can access the data through their disk drive,

rendering read permissions ineffective and thus, absolutely requiring encryption of sensitive data. Thus, while means for confidentiality are provided, the urgency for confidentiality is also inherently implicated within the stream storage system. With regards to the physical integrity of the records, permission-based consensus model provides an efficient and viable solution to preventing a 51% attack for a newly started blockchain like RecordsKeeper. The mining diversity factor, which enables the round-robin mining of blocks, however, is not completely tamper-proof. Likewise, neither is the proof-of-work protocol RecordsKeeper will eventually adopt.

The major threat to integrity with RecordsKeeper, however, is the inability to store records in their original file formats, but instead in reformatted versions. Without some semblance of an audit trail between the newly formatted records in the stream and the off-chain, off-system originals, it is unclear how records saved on the stream can stand in for the originals in various use cases where other users need to access and/or verify the original document. Likewise, RecordsKeeper provides no access to the original documents, should another user request it. None of those issues exist, on the other hand, for records that are originally managed in the preferred file formats.

RecordsKeeper's use of key-value pairs and MultiChain streams provides users the potential to link certain records by their procedural context, thereby instantiating the archival bond. Users upload their records into RecordsKeeper using a key-value format, where the key is the record key (Greenspan, 2016). This key-value format is not to be confused with the public or private keys used for transactions (Greenspan, 2016). The record key may be named whatever the user decides (Greenspan, 2016). The value is the encrypted record or data (Greenspan, 2016). Because the same record key may have

multiple values, a user could base the record key on the procedural context—i.e. “XYZ event files”—and upload multiple records that belong to that aggregate under that same record key onto the same stream (Greenspan, 2016). Whenever someone seeks to retrieve those files, they would use the record key to find those files (Greenspan, 2016). So long as the records were published on the same stream, each record would be listed along with its unique transaction information—i.e. transaction id, time stamp, record hash, and any other metadata entered in the block (Greenspan, 2016). Moreover, it is possible to encode more procedural context in the name of the stream.

For example, if two users, say, an event planner Jolene Doe and a local business, Standard Business, wanted to keep a records of all events she planned for the business, they could open a closed stream with exclusive write permissions for both parties. They could name it Standard Business Events planned by J. Doe. Then for each event she plans, the record key could be the event name or whatever event identification number they choose, such as HolidayPartyDecember 2018. All records regarding the event could be entered by each party under that key value. Other event records could be added to the same stream under different record keys—e.g. Sal’sRetirementParty, SB10YearAnniversity, etc. Given the issues mentioned above, this would work best in cases where records are already in the accepted standard formats mentioned or where standard formatted versions of records are acceptable substitutes.

In terms of confidentiality, RecordsKeeper provides options to ensure encryption, even though the onus to do so is placed upon the user. With regards to authenticity, RecordsKeeper provides a way to instantiate the archival bond in certain use cases and

moderate protections for integrity. Thus, RecordsKeeper earns a medium rating for security.

Trust

As with most blockchain-based platforms that require users to upload records created off-chain, complete accuracy is difficult to ensure. It is the responsibility of the user to ensure that s/he has uploaded the correct record with its correct record ID and accompanying metadata into the RecordsKeeper system. While the added functionality of streams enables users to group particular records together to preserve the archival bond, it also produces another potential opportunity for error. In particular, users publishing records must ensure that they publish the records onto the correct streams (in situations where they might be subscribed to multiple streams) and that they also select the accurate record key. If a record is to be aggregated with other records under a particular record key, or conversely isolated from other record groups, then it is important that the correct record key is used. Otherwise, it could be problematic for later retrieval by both the user and/or any other agencies seeking the record on the particular stream and/or under a particular record key. To avoid such issues, users will need to maintain some documentation regarding their record key naming systems.

The key-value system also provides an opportunity to try to correct certain errors. For example, if an error has been made on Record A, which is listed under the record key "Group 1," then the corrected record could be added to the same Group 1 record key, with notation of the correction. The correction would be visible to users who pull up the entire record key list. However, if a user simply searches out the record by its particular

transaction ID, then the corrected record will not appear, as only the specific record corresponding to the transaction ID is pulled up, and not the entire group of records.

RecordsKeeper offers little in terms of reliability. Completeness at the point of creation is beyond the scope of RecordsKeeper as many records are created off-chain. For example, in a supply-chain records use case, RecordsKeeper does not have a way to verify assets have changed hands in the real world. Someone must enter a record of the transactions that occurred off chain. Consistency with the formal rules of creation is also problematic in the RecordsKeeper context, as records originally in more dense formats—i.e. images, audio recordings, PDF files, etc.—are reformatted and saved into the system that way. Thus, the records made within RecordsKeeper are not consistent.

Given its inability to ensure accuracy or reliability of records, RecordsKeeper earns a low rating for trust.

5 Discussion

Based on the low results of the ARCHANGEL and RecordsKeeper assessments (See Table 1), one can conclude that blockchain-based platforms for recordkeeping at this point in time might just be more hype than help. Neither of the two blockchain-based recordkeeping platforms earned high ratings. ARCHANGEL earned a "Low" rating for all three criteria, while RecordsKeeper was able to manage a "Medium" rating for security. To better explain these results, the discussion section will provide a comparative analysis of each criterion in the study and contextualize these results in terms of T. D. Smith's evaluations of blockchain-based platforms. Finally, based on these analyses, suggestions for improving Smith's evaluation framework are provided.

Project	**Dependability**	**Security**	**Trust**
ARCHANGEL	Low	Low	Low
RecordsKeeper	Low	Medium	Low

Table 1: Assessment results

The first criterion, dependability, had two major aspects the platforms had to address: availability and maintainability. As the evaluations above demonstrated, determining availability for a blockchain platform largely hinges upon how the platform manages to get nodes to continue to validate and store transactions. Without the nodes validating the transactions, the blockchain ceases to continue. Moreover, the nodes provide redundancy as they continue to store the latest version of the chain, so long as the

blockchain is running. In order to determine availability, therefore, the stakeholder needs to know the following information related to the blockchain: proposed consensus protocol, blockchain type (i.e. public/private, permissioned/permissionless), intended participants, and proposed incentive for miners. Knowing the consensus mechanism enables one to understand just how extensive a network of nodes one might need, as well as costs for upholding such a network. Knowing the type of blockchain one is proposing and what type of participants are sought for the blockchain then makes it easier to understand whether or not the proposed incentive is actually viable.

ARCHANGEL, for example, proposed to use a proof-of-work consensus protocol. The proof-of-work protocol requires a large network of nodes to validate the transactions on the blockchain and is very expensive to maintain. Smaller networks with lower numbers of nodes run the risk of a double-spend or 51% attack, as mentioned earlier. ARCHANGEL's permissioned ledger model proposed either using only archives and memory institutions as miners or using a public blockchain to enable anyone to mine the blockchain. This proves to be problematic on multiple levels. In the first instance, while the transparency regarding the integrity of their public documents might serve as sufficient incentive for archives and memory institutions to join the platform, the costs of proof-of-work protocols are quite high for these generally resource-strapped institutions. In the latter model, there is very little incentive for non-archival participants to join the network. Even for individuals who would be interested in checking the integrity of particular documents, the cost of mining would likely be more expensive than verifying the integrity of a public document. Availability could likely have been improved had ARCHANGEL chosen a different consensus mechanism, such as proof-of-stake or

Byzantine Fault Tolerance, both of which actually work more efficiently and cost-effectively with permissioned and private blockchains.

RecordsKeeper, though opting for a different permissioned-consensus model until it builds its network of nodes, seeks to use proof-of-work, as well. Unlike ARCHANGEL, RecordsKeeper does provide an incentive to all its miners—its own XRK utility tokens. Unfortunately, the problem with this is that the utility tokens appeal primarily to the users of the RecordsKeeper platform. As the tokens cannot be used for transactions outside of the platform, it is hard to predict if miners not directly using the platform for records storage would really want to collect the utility tokens just based on potential speculative value on the cyptocurrency exchange. Thus, in terms of availability, both platforms scored low.

Evaluation of maintainability requires consideration of the changes and/or repairs the entire system will undergo over its lifetime. Both ARCHANGEL and RecordsKeeper did not sufficiently address how proposed changes or repairs would be addressed. One key issue regards documents being stored off-chain in repositories. While the maintenance of participating repositories is out of the scope of ARCHANGEL, their continued participation in ARCHANGEL is not. The hash records on the blockchain lose their value without the documents to compare them with. If a member institution is destroyed along with its records, or a participating institution decides to quit the ARCHANGEL project, those documents are no longer available. Not only is this an availability issue, but as records cannot be deleted from the chain, it becomes a maintenance issue. More pointedly, in the hypothetical situation where a document collection from a participating archive is transferred to another participating institution,

then how would the institutional transfer be denoted within the system? Would entirely new records be created denoting new ownership? Would those be linked to the previous records? ARCHANGEL does not as yet have any solutions to these issues.

ARCHANGEL also proposes future improvements such as new content hashing algorithms, as well as a possible membership subscription model. If institutions wish to update the previously hashed documents with the new content hashes, then, once again, some solution regarding how such changes will be noted within the append-only blockchain will need to be addressed. Similarly, should ARCHANGEL choose to adopt a membership subscription model, some level of administrative management will be required. Currently, none of their materials address these issues.

RecordsKeeper had an even greater issue regarding scalability of its storage. Unlike ARCHANGEL, or many other blockchain platforms for that matter, RecordsKeeper's use of MultiChain streams means that all record data that participants upload to streams, regardless of whether or not the data is on or off the chain, will end up being stored in every single node in the network. As participants—and, more importantly, the amount of files they upload—increase, the storage capacity of nodes will continue to decrease until eventually the node is unable to continue storing the data. RecordsKeeper has presented no solution for this problem at this time.

Security according to Smith's rubric hinges upon two aspects: authenticity and confidentiality. As with attempts to examine dependability, knowing the blockchain type, intended participants, and consensus model is required in order to properly evaluate security. Confidentiality for ARCHANGEL and RecordsKeeper could be evaluated in terms of maintaining the confidentiality of the users and of the records. With regards to

determining the confidentiality of users, one must know the blockchain type and intended participants. In the case of ARCHANGEL, two different users have been proposed—any individual who wishes to mine the blockchain and/or verify documents, and archives and memory institutions. The identity of individual miners can remain private through the standard public and private key system.

Because ARCHANGEL's proposed blockchain is a permissioned ledger, it was unclear whether or not participating archives and memory institutions could keep their identities private, should they even want to. A permissioned blockchain requires users to gain permissions for various actions—i.e. to upload hash records. As a result, the administrators of the blockchain must know their identities. It is unclear from ARCHANGEL's documentation whether there will be a central administrative body, and, if so, how that body is determined. The documentation is also unclear as to the technical specifications of how such permissions will be managed. Moreover, as one of the main points of the hash records is to provide more transparency to the preservation practices of these institutions, it would seem counterintuitive for the identities of participating institutions to remain private. As a matter of provenance, furthermore, the identity of the institution submitting a hash record on the blockchain should be transparent.

ARCHANGEL does not provide much by way of records encryption, either, though encryption of hash records seems somewhat contrary to the enterprise. In terms of the documents the hash records refer to, confidentiality of sensitive information is handled by the participating institution that is preserving the document.

Confidentiality in RecordsKeeper is a little less complex. The platform is modeled as a public blockchain wherein all users can remain pseudonymous by the public-private

key cryptographic system. Confidentiality of records is, for the most part, completely up to the user in the RecordsKeeper platform. RecordsKeeper does not provide any automated service to encrypt one's records upon upload to the system. Users must ensure that they encrypt any sensitive documents and information, as whatever is saved directly onto the blockchain is visible to all users of the platform. Moreover, with the underlying MultiChain streaming architecture that saves all steam items on every node, all records uploaded to the platform are technically accessible by any node through its hard disk. This seems problematic for users dealing with records containing sensitive information—should they forget to encrypt their records, or if somehow their encryption key is stolen, then a hacker with a node in the system could access all their records through his/her own node's hard disks.

Of all the aspects to assess, authenticity is perhaps the most complex, as it requires one to examine the integrity and identity of records within the system. Integrity is further split into physical integrity and interpretability. In order to assess the physical integrity of records in a blockchain-based system, it is necessary to know the consensus protocol being used to validate the blockchain transactions. In this regard, both ARCHANGEL and RecordsKeeper propose to use the proof-of-work algorithm, which is still considered by some to be near immutable, to protect the physical integrity of on-chain records. But as the recent spate of 51% attacks has shown, such attacks are becoming more common and not as difficult to execute. In a bid to deflect such attacks, which are easier to execute on networks with lower numbers of nodes, RecordsKeeper does propose to operate on a permissions-based consensus model with a mining diversity algorithm until the network grows. As seen in the evaluation above, however, there are

vulnerabilities with this method as well. Thus, neither platform can guarantee the physical integrity of their records.

Interpretability is also somewhat problematic for both platforms. ARCHANGEL's hash records seem fairly straightforward to understand, assuming that the information is properly labeled and formatted consistently throughout the system. The GUID and metadata provide the information and context necessary to connect the content hash to its related public document. One issue, however, is that the content hash records should always include the content hashing algorithm, so there is no confusion for future users as to what algorithm was used. As Collomosse et al. (2018) discuss the possibility of adding more content-specific hashes to the system beyond the current SHA-256 default, it will become imperative that all records—even those that are default SHA-256—specify the hashing algorithm used. Otherwise, future users will not be able to properly interpret the content hash and correctly verify documents.

Interpretability, specifically in terms of renderabilty, might be diminished by RecordsKeeper's preferred file formats for uploading records. Files such as PDF files, image files, and audio files must be reformatted to simpler binary formats to be uploaded into the RecordsKeeper platform. Users seeking an image file would find a file stored in binary hexadecimal format that they would then need to be able to render into an image on their systems. While this might suit the user's purposes, in most situations where a user needs the original document in its original format, this does not work. RecordsKeeper does not provide any information of how original records in original formats can be shared among users.

The other aspect of authenticity is identity. According to Lemieux (2017b), it is difficult to ensure identity without being able to establish the archival bond. Lemieux also notes that archival bond is often beyond the purview of most blockchain platforms (2017b). This certainly is the case with ARCHANGEL. Though hash records provide fields for metadata, metadata alone cannot establish the archival bond between the hash record and its document, or to other hash records. Moreover, the archival bond among off-chain documents cannot be instantiated in the blockchain.

Despite its inability to guarantee tamper-proof record storage, RecordsKeeper's use of key-value pairs in conjunction with MultiChain streams did enable the platform to instantiate the archival bond in certain use cases. Moreover, as the same MultiChain stream technology enables storage of records (in certain formats) on every node, the multiple copies of the records grant better assurance of the long-term security of these records. Because RecordsKeeper was able to demonstrate some assurances regarding confidentiality of users, identity and redundancy of records, it received a ranking of medium for security.

Both platforms also have difficulty addressing the last of the three criteria, trust. This was to be expected, as Lemieux had noted that accuracy and reliability typically fall beyond the scope of blockchain-based recordkeeping solutions (2017b). ARCHANGEL and RecordsKeeper are no exception. As both platforms demonstrate, accuracy is difficult to guarantee when off-chain records are involved. In the case of ARCHANGEL, it is possible that incorrect or incomplete information might mistakenly be entered into the metadata fields of the hash record. Similarly, another potential source of error would be an incorrect document GUID or duplicate document GUID's for two different hash

records. Not only could these inaccuracies cause confusion and false results when comparing hash records in the future, they could also result in the inability to access the correct hash record and/or correct document altogether. Lastly, ARCHANGEL's use of smart contracts could lead to errors, as sometimes the code does not accurately execute the intended actions.

Similarly, RecordsKeeper has many opportunities to generate inaccurate records, as records are produced off-chain and then must be uploaded to the system. In general, there is always the chance for someone to upload the wrong record, thus ensuring the wrong content for the record. As RecordsKeeper requires record entries to be uploaded onto MultiChain streams in key-value format, there is further chance for mislabeling records, as the user could enter the wrong key or mistype/misspell the key name, upload inaccurate or incomplete records, and/or upload the record onto the wrong stream(s). This could make later retrieval for both the user and any other agencies seeking the record on the particular stream or under a particular record key. Fortunately, the key-value system does potentially afford the opportunity to post corrections, by linking a corrected record into the same record key. However, such corrections would only be visible if a user searched records by record key name, thereby pulling up all the records named with the same record key. If, on the other hand, a user looks up a specific record by the particular transaction ID, then only the specific record corresponding to the transaction ID is pulled up, and not the entire group of records with the corrected record.

In order for the platforms to prove their records are reliable, the records had to demonstrate all three of the following traits: completeness at the point of creation, consistency with the formal rules of creation, and naturalness. If just one of those traits is

absent, then the record is not reliable. Both platforms had difficulty demonstrating completeness at the point of creation. In the case of ARCHANGEL, the content hash records need to be able to demonstrate that the content hash is the accurate hash of the document to which it refers. Nothing inherent in the record proves that it represents the document referred to in the record by the document GUID. The hash must match another hash created from the same document with the same algorithm to prove that it is in fact the hash record of that document. Thus, ARCHANGEL's hash records lack completeness at the point of creation and are not reliable.

RecordsKeeper cannot ensure reliability, either. For the most part, records are generated off-chain and then uploaded onto the RecordsKeeper platform in their preferred file formats. Reformatting particular documents—such as property titles, certifications, and photo identifications—into JSON, XML, or any of the other preferred formats runs the risk of removing the legitimating aspects of the record, i.e. the signatures, official seals, photo image, etc. Not only does this mean the record is no longer complete at the point of creation, as it only represents the record but holds no real-world effect per se, but records are also no longer consistent with the formal rules of creation. Moreover, though RecordsKeeper does propose certain use cases that might involve the creation of records online, no details are provided on how such online records would be created, monitored, or verified to ensure that the real-world actions the records signify have actually occurred.

The above rankings for ARCHANGEL and RecordsKeeper might at first blush not seem consistent with the higher range of scores T. D. Smith (2017) awarded to the Bitcoin blockchain and four other blockchain-based recordkeeping platforms: MedRec,

Storj, Blockchain for the Internet of Things, and Bitcoin for Decentralized Trusted Timestamping. However, of the blockchain-based solutions Smith (2017) evaluated, only one, MedRec, comes close to providing large-scale recordkeeping services for electronic records management and archival purposes. Storj is simply a decentralized storage management system for individual data—not necessarily even records. No records sharing or even verifying is necessary. Bitcoin for Timestamping simply considers the timestamp function on the Bitcoin blockchain. MedRec, on the other hand, is a blockchain-based platform seeking to solve electronic medical records management issues using blockchain. MedRec also deals with electronic records that are produced off-chain and stored off-chain. MedRec scored "Low" in every single category (Smith, 2017, p. 3305). MedRec, ARCHANGEL, and RecordsKeeper, as well-intentioned as they may be, demonstrate that at least from an archival perspective, blockchain is not well-suited for long-term records management and preservation.

These rankings are consistent, moreover, with the broader archival context. As much of Victoria Lemieux's work has demonstrated, blockchain-based records management platforms are generally not designed in consultation with archivists, leading to blockchain-based solutions that completely ignore key aspects of the long-term digital preservation of records (2016a). In particular, based on the application of her own archival theoretic framework to a generic blockchain-based structure, Lemieux posits that accuracy, reliability, the archival bond, and persistence through time are all aspects of recordkeeping that blockchain-based recordkeeping solutions cannot fully address.

While the low scores earned by the platforms above are consistent with previous research, there are some limiting factors that may have impacted these rankings. First and

foremost, the knowledge used to inform these evaluations was limited to information that was publicly available at the time of the study. In the case of ARCHANGEL, this information consisted of a four-page summary of the ARCHANGEL project written by Collomosse et al. and blog posts posted by project partners. As ARCHANGEL is still in its early developmental stages, some of the criteria that are not currently addressed might be resolved in future versions.

The RecordsKeeper assessment relied upon the RecordsKeeper project white paper, the official website of the company, and the MultiChain website. The information provided by RecordsKeeper was particularly challenging, as in certain cases it was too general and vague, or, in other cases, a bit inconsistent. In particular, the figures and information provided by RecordsKeeper regarding proposed use cases only detailed problems and promised solutions, but gave no information regarding how such solutions would be reached through the platform (RecordsKeeper, n.d.-d). No technical details or step-by-step instructions matching the various aspects of the figures are provided. Similarly technical information regarding the MultiChain stream technology underlying the RecordsKeeper platform could not be found in RecordsKeeper materials. Fortunately such information is available through blogposts on the MultiChain website.

Inconsistencies throughout their documentation—born out of what appears to be lack of editing—made certain aspects even more confusing. In the description of proof-of-work protocol of their white paper, for example, they also use the acronym for proof-of-stake (PoS). Such inconsistencies make it difficult to be certain of the interpretation of their documents.

Another potential limitation of this study is the fact that both of the recordkeeping solutions evaluated use the proof-of-work consensus model. It would be interesting to see how blockchain-based recordkeeping systems based on other consensus models, such as proof-of-stake and Byzantine Fault Tolerance, would rate using Smith's evaluation framework. Likewise, including another blockchain-based recordkeeping use case—such as a data-sharing platform—might lend some more comparative insights. Such comparisons could be a direction for future research.

Another potential area for future research would be to use Smith's evaluation framework on non-blockchain-based recordkeeping platforms. Smith's evaluation framework is general enough to be applied to any recordkeeping platform, regardless of its use of blockchain technology. The three criteria—dependability, security, and trust—can serve to illuminate the utility of any recordkeeping system. In fact, if stakeholders were to use Smith's evaluation framework to determine whether or not to adopt a blockchain-based recordkeeping platform, it might be useful to apply the same framework to evaluate their current recordkeeping system, as well any other recordkeeping systems they might wish to use, blockchain or not. Such a comparative evaluation would enable a stakeholder to make a fully informed decision regarding what recordkeeping system would best fit their needs.

While Smith's evaluation framework was helpful in determining how well ARCHANGEL and RecordsKeeper addressed the three criteria, the framework did not provide anything unique to the evaluation of blockchain-based platforms. Though Smith's evaluations of blockchain-based recordkeeping systems provide examples of evaluations for stakeholders to attempt to model their own evaluations after, no explicit

instructions have been provided. To that end, based upon the evaluation of ARCHANGEL and RecordsKeeper, this study proposes a set of blockchain-specific questions a stakeholder could use to better evaluate blockchain-based recordkeeping platforms. First and foremost, one must be able to answer, what recordkeeping functions does the platform promise to perform? Second, what is the blockchain being used for, precisely—is it to record transactions? To store records? To verify records? Third, what is the consensus mechanism being used to verify transactions? As seen above, the consensus mechanism determines important aspects regarding dependability and security. Fourth, what type of blockchain is this—public, private, permissioned, and/or permissionless? And last, though not least: who benefits, and how, from this blockchain? For instance, are participants only users of the services provided by the platform, or are there also users who earn tokens for mining? With regards to Smith's evaluation framework, answers to these basic questions about the blockchain-based recordkeeping platform under review will provide much of the requisite information needed to generate an informed appraisal. These questions would also make such evaluations simpler for those not familiar with blockchain technology.

Perhaps however, the most important question one should ask about a blockchain-based system is whether or not it is appropriate or even necessary to use for one's recordkeeping needs. Given that such systems are expensive and volatile, one must give serious consideration as to whether or not blockchain technology is the most appropriate solution for one's needs. Various flowcharts, such as the one produced for the World Economic Forum by Mulligan, Warren and Rangaswami, might be useful places for

stakeholders to begin their initial evaluations of blockchain-based recordkeeping solutions (2018).

In addition to ascertaining whether or not blockchain technology is necessary or appropriate for one's recordkeeping needs, one should consider conducting an explicit evaluation of the sustainability of using such blockchain-based programs. To that end, this study suggests adding sustainability as the fourth criteria to Smith's evaluation framework. Smith's framework never directly addresses issues related to sustainability and the long-term persistence of records. Instead, one might indirectly touch upon issues related to sustainability when examining availability. For example, in determining the feasibility of ARCHANGEL's plan to attract other archives as miners to the blockchain network, the expense of maintaining the proof-of-work consensus protocol became a major flaw.

The addition of sustainability to Smith's evaluation framework would force stakeholders to directly address issues—such as costs—that affect the long-term viability of such recordkeeping systems. While Smith's evaluation framework makes no explicit mention of long-term viability, Lemieux does reference the importance of considering the "persistence through time" (Lemieux, 2017b, p. 10) of records created in blockchain-based recordkeeping systems. Though not an explicitly linked to her taxonomy of trust, persistence through time does appear on the periphery of the taxonomy diagram (Lemieux, 2017b, p. 7). Lemieux concludes that persistence through time generally is beyond the scope of blockchain-based platforms, as blockchain is somewhat volatile, and as a relatively new technology, no one can predict its long-term viability (2017b). As one

of the co-founders and developers of Stanford's LOCKSS program, David Rosenthal further argues:

> Sustainability is job #1 for archives. There's no point in setting up an archiving system and filling it with content only to have it fail after a decade or so. Sustainability has to be designed into both the technology and the organization into which it is embedded from the start if the contents are to survive the wide range of threats to which archived data is subject. Layering it on afterwards isn't going to be effective. (2018)

As the above statement demonstrates, sustainability can never be a peripheral consideration, or an afterthought, for any archive. In fact, as Rosenthal asserts, sustainability is the primary consideration. As such, any recordkeeping system that cannot guarantee sustainable results from the beginning, cannot guarantee long-term persistence of records over time. As Rosenthal states, sustainability cannot be added to the system later. Measures to preserve records and operate sustainably must be in place from the start.

Moreover, while Rosenthal argues that sustainability is of central importance regarding long-term preservation for archives and memory institutions, his statements can also be applied to any records management system. Most records management systems will contain records collections that need indefinite storage; but even a collection that has a 10-year retention schedule needs to be managed in a recordkeeping platform that will be around at minimum for the next decade. Thus, any stakeholder considering storing their records with a recordkeeping platform will need to know how long they can expect their records to be safely preserved within that system.

To assess the long-term sustainability of a program, two factors must be considered: financial cost and environmental impact. As David Rosenthal so aptly states,

"digital preservation is primarily an economic problem" (2018). Digital preservation requires a long-term commitment of resources and money. Knowing what a recordkeeping system will cost to set up and run annually is a key factor in establishing the system's viability as an option for the long-term storage of records. Stakeholders need to know not only what finances and resources they are committing in the near future to the new platform, but estimated costs for the next 10 years, 20 years, and so on. Without such practical information, stakeholders will not be able to determine if they have enough funds (or could procure enough funds) to ensure long-term preservation. Thus, stakeholders need to ascertain the actual costs of computing equipment required, electricity costs to power their program, storage costs, and any other resources that would need to be allocated to the recordkeeping platform. While it is hard to imagine that most stakeholders would not examine the costs associated with any new platform under consideration, by making it an explicit criterion in the evaluation framework, potential users would need to scrutinize such costs more directly. In so doing, potential users would likely become aware of blockchain's high electricity consumption, an important hidden cost that has great ramifications for long-term sustainability.

Environmental impact is another important issue to consider in terms of blockchain-based programs. Though often ignored or downplayed by blockchain enthusiasts and cryptocurrency investors, blockchain technology consumes a lot of energy and, as a result, emits a huge amount of CO2 emissions into the atmosphere. Current estimates by Digiconomist (n.d.-a) regarding Bitcoin blockchain's carbon emissions are a staggering 241.91 kg of CO2 per transaction and 22,293 kilotons of CO2 annually. Not only is the electricity consumption by Bitcoin massive—constituting an

estimated 0.20% of the entire world's electricity consumption—but most of it is powered by coal, an especially potent source of CO2 emissions (Digiconomist, n.d.-a). As climate change continues to get worse, consumers should be aware of their energy consumption both in terms of amount and type (clean or dirty). In a sense, assessing environmental impact ensures that there will be future users of the records an institution works so hard to preserve.

Finally, it is important to state that while this paper does not provide favorable evidence for using blockchain-based platforms for records management, blockchain technology is still developing and may yet provide certain solutions otherwise not possible. With the amount of resources invested in blockchain, and awareness of its shortcomings, it is possible that some new solutions could be developed to make it more secure and viable. Developers continuing to work on blockchain technology might come up with innovations that can counter nefarious 51% attacks. For other investors, the shortcomings might not outstrip the potential as the decentralized nature of blockchain is perceived by some as the only means for solving complex social issues, such as identity management for homeless persons and migrant refugees (Galen et al, 2018).

6 Conclusion

This paper examined two blockchain-based recordkeeping platforms—ARCHANGEL and RecordsKeeper—using T. D. Smith's evaluation framework, which applies the secure-computing principles of Avizienis et al. and Victoria Lemieux's archival theoretic framework. Application of Smith's framework to both platforms yielded low rankings for each program, raising severe doubts about the capability of blockchain-based programs to ensure secure, accurate, reliable, long-term recordkeeping.

Based upon the insights gained through using Smith's evaluation framework, this study also offers two suggestions for improving Smith's evaluation framework. The first suggestion is the addition of a set of questions to make Smith's evaluation more blockchain-specific. Such blockchain-specific questions would enable stakeholders to quickly identify and assesss the strengths and weakness of the proposed blockchain technology. The second suggestion is to add a fourth criterion to Smith's framework, sustainability. Based upon two factors, financial cost and environmental impact, an appraisal of a system's sustainability would enable a more practical and complete assessment of the long-term viability of that recordkeeping platform. Given that a primary focus of recordkeeping is the long-term persistence of records, sustainability is a crucial factor for evaluating recordkeeping platforms, particularly ones that are as unpredictable and expensive as blockchain-based platforms.

Bibliography

Avizienis, A., Laprie, J. C., Randell, B., & Landwehr, C. (2004). Basic concepts and taxonomy of dependable and secure computing. *IEEE transactions on dependable and secure computing*, *1*(1), 11-33.

Aste, T. (2016, June 27). The fair cost of bitcoin proof of work. Retrieved from: http://dx.doi.org/10.2139/ssrn.2801048

Bitcoin Project. (n.d.). Some bitcoin words you might hear. Retrieved from: https://bitcoin.org/en/vocabulary#bitcoin

Blockgeeks. (2018). Proof of work vs proof of stake: Basic mining guide. Retrieved from: https://blockgeeks.com/guides/proof-of-work-vs-proof-of-stake/

Buntinx, J. P. (2017, June 15). What is blockchain bloat? Accessed on November 20, 2017 at: https://nulltx.com/what-is-blockchain-bloat/

Burniske, C., & Tatar, J. (2017). *Cryptoassets: The innovative investor's guide to bitcoin and beyond*. McGraw Hill Professional.

Camacho, J. (n.d.). Utility tokens: A general understanding. [blog post]. Retrieved from: https://medium.com/coinmonks/utility-tokens-a-general-understanding-f6a5f9699cc0

Camp, L. J. (2018, November 5). *Challenges to bitcoin*. Keynote presented at the Symposium on Blockchain and Trusted Repositories, Chapel Hill.

Canellis, D. (2018, October 23). Report: Cryptocurrency hackers earned $20M with 51-percent attacks in 2018. Retrieved from: https://thenextweb.com/hardfork/2018/10/23/cryptocurrency-51-percent-attacks/.

Champagne, P. (2014). The book of Satoshi: The collected writings of bitcoin creator Satoshi Nakamoto. *E53*.

Collomosse, J., Bui, T., Brown, A., Sheridan, J., Green, A., Bell, M., …Thereaux, O. (2018). ARCHANGEL: Trusted Archives of Digital Public Documents. *DocEng '18: ACM Symposium on Document Engineering 2018, August 28–31, 2018, Halifax, NS, Canada.* https://doi.org/10.1145/3209280.3229120

Drescher, D. (2017). *Blockchain basics*. Apress.

Digiconomist. (n.d.-a). "Bitcoin Energy consumption index" Retrieved November 20, 2018 from: https://digiconomist.net/bitcoin-energy-consumption

Digiconomist. (n.d.-b). "Ethereum Energy Consumption Index (beta)." Retrieved Dec. 7, 2018 from: https://digiconomist.net/ethereum-energy-consumption

Ethereum Foundation. (n.d.-a). About the Ethereum Foundation. Retrieved from: https://www.ethereum.org/foundation

Ethereum Foundation. (n.d.-b). Main page. Retrieved from: https://www.ethereum.org/

Ethereum Foundation. (n.d.-c). A next-generation smart contract and decentralized application platform. [White Paper] Retrieved from: https://github.com/ethereum/wiki/wiki/White-Paper#ethereum

Falkon, Samuel. (2017, December 24). The Story of the DAO—Its history and consequences. Retrieved from: https://medium.com/swlh/the-story-of-the-dao-its-history-and-consequences-71e6a8a551ee

Finley, K. (2018, February 1). The Wired Guide to the Blockchain. *Wired Magazine*. Retrieved from: https://www.wired.com/story/guide-blockchain/.

Galen, D., Brand, N., Boucherle, L., Davis, R., Do, N., El-Baz, B.,…& Lee, J. (2018, April 11). Blockchain for social impact: Moving beyond the hype. *Center for Social Innovation, RippleWorks*. Retrieved from: https://www. gsb. stanford.edu/sites/gsb/files/publication-pdf/study-blockchain-impact-moving beyond-hype_0. pdf.

Greenspan, G. (2018, June 13). Scaling blockchains with off-chain data: When a hash is worth a million words. [blog post] Retrieved from: https://www.multichain.com/blog/2018/06/scaling-blockchains-off-chain-data/

Greenspan, G. (2016, September 15). Introducing MultiChain streams: For shared immutable key-value and time series databases. [blog post] Retrieved from: https://www.multichain.com/blog/2016/09/introducing-multichain-streams/

Hertig, A. (2018, June 8). Blockchain's once feared 51% attack is now becoming regular. *Coindesk*, https://www.coindesk.com/blockchains-feared-51-attack-now-becoming-regular Accessed December 12, 2018.

Hyperledger. (2018). Blockchain and the enterprise. But what about security? Webinar Q & A. Retrieved from: https://www.hyperledger.org/blog/2018/07/17/blockchain security-webinar-q-a

Keller, J.R. (2018a). Blockchain's potential role in the future of archiving. [blog post] Retrieved from: https://theodi.org/article/blockchains-potential-role-in-the-future-of-archiving/

Keller, J.R. (2018b). Challenges in using blockchains to build trust in digital archiving. [blog post] Retrieved from: https://theodi.org/article/challenges-in-using-blockchain-to-build-trust-in-digital-archiving/

LearnCryptography.com. (2018). 51% attack. Retrieved from: https://learncryptography.com/cryptocurrency/51-attack

Lemieux, V. L. (2017a). A typology of blockchain recordkeeping solutions and some reflections on their implications for the future of archival preservation. In *Big Data (Big Data), 2017 IEEE International Conference on* (pp. 2271-2278). IEEE.

Lemieux, V. (2017b). Blockchain and Distributed Ledgers as Trusted Recordkeeping Systems: An Archival Theoretic Evaluation Framework. In *IEEE Future Technologies Conference*.

Lemieux, V. L. (2017c). Evaluating the Use of Blockchain in Land Transactions: An Archival Science Perspective. *European Property Law Journal*, *6*(3), 392-440.

Lemieux, V. L. (2016a). Blockchain technology for recordkeeping: Help or hype. *Unpublished report*.

Lemieux, V. L. (2016b). Trusting records: is Blockchain technology the answer?. *Records Management Journal*, *26*(2), 110-139.

Lemieux, V. L., & Sporny, M. (2017, April). Preserving the archival bond in distributed ledgers: A data model and syntax. *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 1437-1443).

Morely, M. (2018). JSON-RPC 2.0 Specification. Retrieved from: https://www.jsonrpc.org/specification

Mougayer, W. (2016). The business blockchain: Promise, practice and application of the next internet technology. Hoboken: John Wiley & Sons.

Mulligan, C., Scott, J. Z., Warren, S., & Rangaswami, J. (2018). Blockchain beyond the hype; a practical framework for business leaders. *Geneva. URL: http://www3. weforum. org/docs/48423_Whether_Blockchain_WP. pdf (visited on 18th May 2018)*.

MultiChain. (n.d.-a). MultiChain data streams: For shared immutable key-value and time series databases. [blog post] Retrieved from: https://www.multichain.com/developers/data-streams

MultiChain. (n.d.-b). Permissions consensus: Creating a consensual governance model. [blog post]. Retrieved from: https://www.multichain.com/developers/permissions-consensus/

MultiChain. (2018a). MultiChain developer q & a: Does mining diversity make it easier to tamper with the blockchain? Retrieved from: https://www.multichain.com/qa/8764/does-mining-diversity-make-easier-tamper-with-the-blockchain

MultiChain. (2018b). MultiChain developer q & a: How do i store images in multichain streams? Retrieved from: https://www.multichain.com/qa/11306/how-do-i-store-images-in-multichain-streams?show=11306#q11306

Nakamoto, S. (2008). Bitcoin. *A peer-to-peer electronic cash system*.

Pearce-Moses, R., Duranti, L., Michetti, G., Andaur, S. B. H., Banard, A., Barlaoura, G., ... & Pan, W. (2017). InterPARES Trust Terminology Database.

RecordsKeeper. (n.d.-a). Frequently asked questions. Retrieved from: https://docs.recordskeeper.co/en/latest/faq.html

RecordsKeeper. (n.d.-b). Our company. Retrieved from: https://www.recordskeeper.co/our-company/

RecordsKeeper. (n.d.-c). Overview. Retrieved from: https://www.recordskeeper.co/overview/

RecordsKeeper. (n.d.-d). RecordsKeeper Stats. Retrieved from: https://stats.recordskeeper.co/

RecordsKeeper. (n.d.-e). Use cases. https://docs.recordskeeper.co/en/latest/usecases.html

RecX Technologies Limited. (n.d.). Record Keeping & Data Security Solution [White Paper].

Redshaw, T. (2017). Bitcoin beyond ambivalence: Popular rationalization and Feenberg's technical politics. *Thesis Eleven*, *138*(1), 46-64.

Rosenthal, D. S. H. (2018, September 13). Blockchain solves preservation! [blog post] Retrieved from: https://blog.dshr.org/2018/09/blockchain-solves-preservation.html

S., J. (2018, May 5). Blockchain: how a 51% attack works (double spend attack). Retrieved from: https://medium.com/coinmonks/what-is-a-51-attack-or-double-spend-attack-aa108db63474

Sharma, T. (2018, November 16). RecordsKeeper token sale cancellation announcement. Retrieved from: https://www.recordskeeper.co/blog/recordskeeper-token-sale-cancellation-announcement/

Smith, T. D. (2017). The blockchain litmus test. *Big Data (Big Data), 2017 IEEE International Conference on* (pp. 2299-2308). IEEE.

Stinchcombe, K. (2018). "Ten years in, nobody has come up with a use for blockchain." *Hackernoon,* https://hackernoon.com/ten-years-in-nobody-has-come-up-with-a-use-case-for-blockchain-ee98c180100, Accessed December 12, 2018.

The National Archives. (n.d.-a). British army medical index cards 1914-1920. Retrieved November 5, 2018 from: http://www.nationalarchives.gov.uk/help-with-your-research/research-guides/british-army-medal-index-cards-1914-1920/

The National Archives. (n.d.-b). Home page. Retrieved from: http://www.nationalarchives.gov.uk/

The National Archives. (n.d.-c). Research guides. Retrieved November 5, 2018 from: http://www.nationalarchives.gov.uk/help-with-your-research/research-guides/

The National Archives. (n.d.-d). Royal navy ratings' service records 1853-1928. Retrieved November 5, 2018 from: http://www.nationalarchives.gov.uk/help-with-your-research/research-guides/royal-navy-ratings-service-records-1853-1928/

The National Archives. (n.d.-e). Royal marines' service records 1842-1925. Retrieved November 5, 2018 from: http://www.nationalarchives.gov.uk/help-with-your-research/research-guides/royal-marines-service-records-1842-1925/

U.K. Government Web Archive. (n.d.-a). Home page. Retrieved November 5, 2018 from: http://www.nationalarchives.gov.uk/webarchive/

U.K. Government Web Archive. (n.d.-b). Twitter archive. Retrieved November 5, 2018 from: https://webarchive.nationalarchives.gov.uk/twitter/

U.K. Government Web Archive. (n.d.-c). Video archive. Retrieved November 5, 2018 from: https://webarchive.nationalarchives.gov.uk/video/

Volpicelli, G. (2018). Does blockchain offer hype or hope? Retrieved from: https://www.theguardian.com/technology/2018/mar/10/blockchain-music-imogen-heap-provenance-finance-voting-amir-taaki

Walker, D. (2018). "It's not much use: blockchain won't be big." *CEO Magazine*. https://www.theceomagazine.com/opinion/its-not-much-use-blockchain-wont-be-big/ Accessed December 14, 2018.

Zīle, K., & Strazdiņa, R. (2018). Blockchain use cases and their feasibility. *Applied Computer Systems*, *23*(1), 12-20.