

Blockchain is already here. What does that mean for records management and archives?

Sharmila Bhatia^a and Anthony D. Wright de Hernandez^{b*}

^aNational Archives and Records Administration, College Park, MD, USA; ^bVirginia Tech, Blacksburg, VA, USA

Sharmila Bhatia, ORCID: 0000-0002-9891-9327; LinkedIn: in/sharmila-bhatia-3874655; *Anthony D. Wright de Hernandez, antwri@vt.edu; ORCID: 0000-0002-2569-6556; LinkedIn: in/anthonydwright; Twitter: @RoganWright

Biographical Notes:

Sharmila Bhatia joined NARA in 2001 and is currently an electronic records format specialist in the Office of the Chief Records Officer with the National Archives and Records Administration (NARA). She researches digital formats and emerging technologies and the implications for records management and archives. Previously she was an archivist with the Electronic Records Division, also at NARA. She has experience in working with electronic records in a variety of formats and of multi-stages in the lifecycle: transfer, processing, accessioning, preservation, and description. She began her archival career as a reference archivist with the South Carolina Department of Archives and History. She has MA/MLIS degrees from the University of South Carolina and a BA in history from the University of Maryland.

Anthony D. Wright de Hernandez is the Community Collections Archivist at Virginia Tech in Blacksburg, VA where he works primarily with materials related to historically marginalized groups. He worked previously with corporate records management and database administration before moving into academic archives work. He has an interest in digital preservation and recordkeeping, particularly as they relate to preserving the history of historically marginalized communities. This interest has led him to explore the ways that these communities are represented in archival description and how emerging technologies, including blockchain, might benefit these communities in the future. He holds an MLIS from the University of Washington iSchool and a BA in Community Studies from the University of Massachusetts, Boston.

Note: This is a Preprint

This article has been accepted for publication in the Journal of Archival Organization, published by Taylor & Francis.

Abstract

This article provides a brief overview of blockchain or distributed ledger technology for archivists and records managers. While in the nascent stage, records are already being created or managed by blockchains. Therefore, it is necessary for archivists and records managers to become familiar with the basics of the technology, along with its benefits and shortcomings, in order to make informed decisions regarding the accessioning of blockchain systems and/or their use in managing record authenticity.

Keywords: blockchain, distributed ledger technology, records management

Introduction

Blockchain is not a new technology. The foundations were defined in 1991 to secure document timestamps or digital signatures in an electronic ledger. In 2008, a pseudonymous author, known as Satoshi Nakamoto, improved functionality for the application of digital currency. Because of Nakamoto's paper, blockchain has been closely associated with digital or cryptocurrency although there are many potential uses for this technology. Within the last three years there have been explorations and pilots in government and industry for supply chains, identity management, and recordkeeping.

Technology Overview

Blockchain, a type of distributed ledger technology, is a database that is shared, and replicated, among numerous parties. The database can only be updated by following pre-established rules and once updated, it is shared with all of the parties. These transactions are linked together in a chain that ensures that all parties have an accurate copy of the ledger. This replication and storage of transactional data on a blockchain network is known as a distributed ledger. The distributed ledger builds upon this concept and replicates the data across many nodes. Since blockchain is a peer-to-peer network that timestamps a transaction, it eliminates the need for third party services to validate the transactions. This form of recordkeeping contains consensus or trust-based transactions and is tamper resistant. Blockchains are not used for general data storage.

Archivists and records managers are familiar with ledgers, which are used for recordkeeping. Traditional, centralized recordkeeping requires a third party or repository to ensure trust. For example, banks perform this task. When depositing funds with a bank, customers trust that the money will be there until they decide to exchange the money for a purchase. Customers trust that the bank will maintain accurate records of the transactions such as amount, depositor, debtor, date and time. Blockchains are already disrupting third party financial institutions and organizations; however, banks are beginning to explore adopting the technology.¹

Blockchains contain three parts: block, chain, and network. The block contains the list of transactions and these transactions can track any type of activity. Common uses include tracking goods, single purchases, or assets. The rules, such as limiting the size or number of transactions, are established when the network is first created. When the block reaches the maximum size, it is chained or linked by a hash. A hash is an algorithm that generates a fixed length value from a variable string of data. Inserting the hash value of one block into the next block links the new block with the preceding block. Repeating a hash function on an unaltered block of data will always generate the same fixed-length value. If a block of data is altered, the resulting hash output will be different. A different output indicates that the original block has been altered and may no longer be trustworthy. Multiple hash values can be brought together, and hashed, creating a single hash or a Merkle root. Additional hashes are added to the root thus creating a Merkle tree.

¹ Darryn Pollock. "Forget Cryptocurrencies: How can financial institutions make use of blockchain?"

Forbes, March 7, 2019.

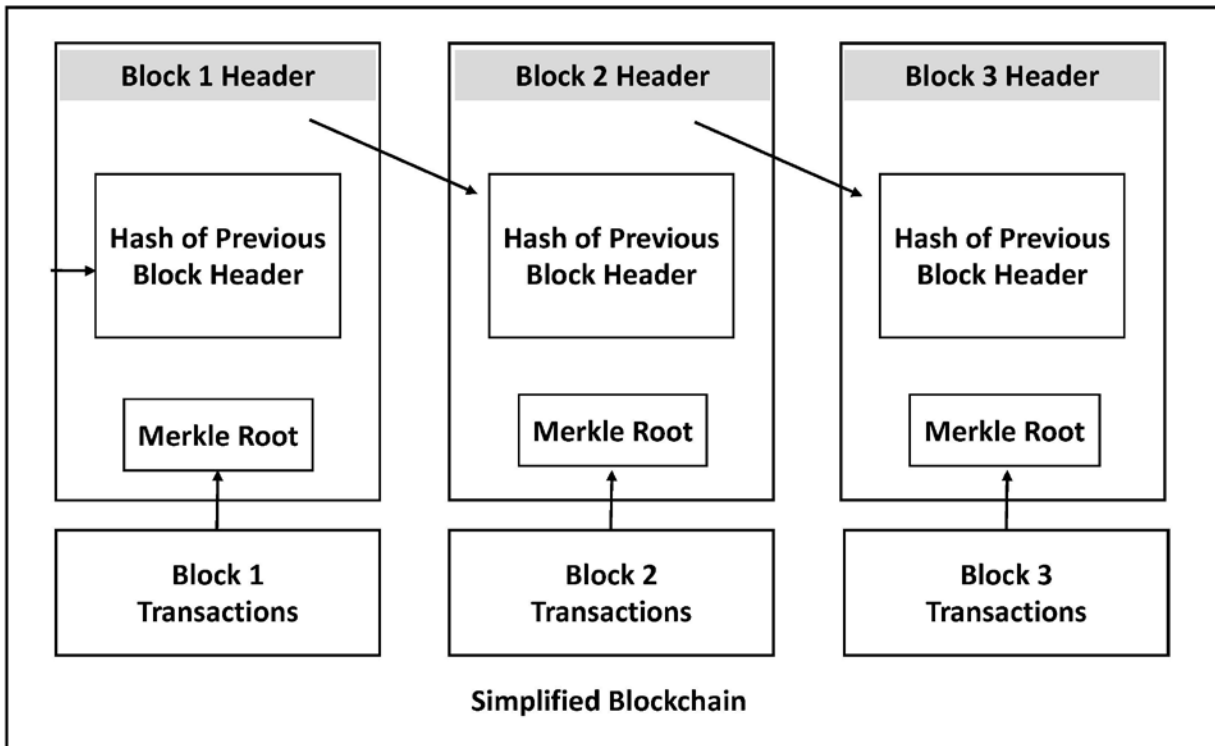


Figure 1: Simplified Blockchain

The blockchain network consists of nodes each containing a complete record of all the transactions. No centralized copy exists, and no node is trusted more than another. The data integrity is maintained by the blockchain being replicated on all of the nodes. There are three types of networks: public, permissioned, and private. A public network, which is usually a large network, allows anyone to participate. This is the most common network used for cryptocurrency exchanges. A permissioned network is limited to certain parties, but may also be large. The smallest network is a private network, which is tightly controlled, and typically established between trusted entities to share confidential information.

There are a multiple blockchain platforms with differing approaches. The most common is Bitcoin's blockchain cryptocurrency platform. Bitcoin is an open-source permissionless platform. Since Bitcoin assumes there is no trust among the parties, it requires numerous decentralized nodes, which ensures that the blockchain has not been corrupted by malicious actors. The decentralized nodes are maintained by active users called miners. Cryptocurrency platforms require miners who solve crypto-puzzles and which in turn is verified by other miners or nodes. This process is known as mining and the miners who solve and verify the solution are rewarded

with cryptocurrency. While required for cryptocurrency platforms, incentivized miners are not necessary for other blockchain platforms²

Currently, the two most common blockchain platforms that expand functionality beyond cryptocurrency are Ethereum and Hyperledger. Ethereum, launched in 2015, provides programming language to allow users to build applications with an integrated blockchain. Ethereum is an open-source programming language, and can be used to program executable smart contracts and decentralized applications using the blockchain.³ Hyperledger launched shortly after Ethereum, in 2016, seeks to develop an open-source, distributed ledger framework. Although there are numerous platforms, many corporations and organizations are using either Ethereum or Hyperledger. SIMBA chain, a cloud-based, blockchain-as-a-service (BaaS) which supports Ethereum, was developed as a messaging and transaction platform for the United States military. This service is expanding to other platforms in addition to Ethereum, and offers a build-your-own blockchain with scalable pricing and options.⁴ On April 30, 2019, Amazon Web Services announced Amazon Managed Blockchain, a managed service to create and manage scalable blockchain networks. At launch, the service supports Hyperledger Fabric, with Ethereum scheduled to be added later.⁵

Blockchain and Records Management

One of the main arguments in favor of using blockchain for records management is that blockchain will provide a means for establishing trust in the validity of a record. Public concern over trust in government, and concerns of marginalized communities over potential abuse of institutional power, make this aspect of blockchain very attractive to any agency or entity looking for a way to prove that their records are authentic and unaltered. However, as has been observed by Dr. Victoria Lemieux—the leading researcher on blockchain and archives—and the government of Vermont in their 2018 blockchain records law, the technology can only provide evidence that a record has not

² “Blockchain White Paper,” (National Archives and Records Administration, 2019); Dylan Yaga, Peter Mell, Nik Roby, and Karen Scarfone, “Blockchain Technology Overview,” (NISTIR 8202, National Institute of Standards and Technology, 2018).

³ Linda Xie. “A beginner’s guide to Ethereum,” *The Coinbase Blog*, February 23, 2017, <https://blog.coinbase.com/a-beginners-guide-to-ethereum-46dd486ceecf>. (accessed May 11, 2019).

⁴ “SIMBA Chain,” SIMBA Chain, Simba Chain Inc., accessed May 11, 2019, <https://simbachain.com/>.

⁵ Richard Kastelein. “Amazon Managed Blockchain Launched – AT&T, Nestlé, and Singapore Exchange Limited Onboard.” *Blockchain News*, 2019, <https://www.the-blockchain.com/2019/05/06/amazon-managed-blockchain-launched-att-nestle-and-singapore-exchange-limited-onboard/>.

been altered since it was added to or linked to the blockchain.⁶ It cannot speak to the validity of the contents of the record before being linked to the blockchain. Records management policies and institutional integrity are still necessary to provide full faith in stored records.

Other benefits being explored beyond cryptocurrencies are transactional automation and identity confirmation for official purposes. This has led governments to explore the use of blockchain for:

- vital records: the first blockchain-linked birth certificate was issued in Bengal, India in December 2018 and blockchain-linked marriage certificates were issued in Washoe County, Nevada beginning in April 2018
- health records: the Estonian eHealth system was established even before Bitcoin and Walmart won a patent in 2018 to house health records in a blockchain for use by first responders
- land deeds: the first official transfer of ownership was done by HM Land Registry's Digital Street project as a proof-of-concept in Spring 2019
- academic credentials: a group of universities, including Massachusetts Institute of Technology (MIT), Harvard, Delft University of Technology (TU Delft), and more are working to offer degree verification via blockchain

Each of these implementations offers potential reductions in the amount of paperwork processing to verify credentials, and adds a level of credibility based on trust in the immutability of blockchain systems. However, it also raises new concerns. Each has an expected life cycle of at least decades. Blockchain implementation today may address short-term concerns, but there needs to be careful consideration of the long-term viability of tying these records to this type of technology, including consideration of how to migrate records off the blockchain without them losing credibility if it becomes necessary to do so.⁷

⁶ Victoria L. Lemieux. "Blockchain Technology for Recordkeeping: Help or Hype?" (Volume 1: Report, The University of British Columbia, Vancouver, CAN, 2016); Blockchain enabling. *Vt. Stat. Ann. tit. 12. § 1913*. 2016. *Justia Legal Resources* (accessed May 28, 2019).

⁷ Udit Prasanna Mukherji and Suman Chakraborti. "A 1st in Bengal, baby gets blockchained birth certificate." *Times of India*, December 20, 2018, <https://timesofindia.indiatimes.com/india/a-1st-in-bengal-baby-gets-blockchained-birth-certificate/articleshow/67170551.cms>; "Some Nevada governments using blockchain for public records." *AP News*, January 6, 2019, <https://www.apnews.com/8eb100aa9ce9418b9ad34473e732c1e3>; Oscar Williams-Grut. "Estonia is using the technology behind bitcoin to secure 1 million health records." *Business Insider*, March 3, 2016, <https://www.businessinsider.com/guardtime-estonian-health-records-industrial-blockchain-bitcoin-2016-3>; Sy Mukherjee. "Walmart Just Scored a Patent for a Blockchain-Based Health Care Records System." *Fortune*, June 22, 2018,

The ability to migrate data off the blockchain is already something of growing importance. Consider the new data privacy laws in the European Union (EU) and California. The EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) both allow individuals to have direct control over some portion of their personal data, allowing them to have it removed from online sources. For data that lives in a blockchain (for example, data created by smart contracts such as a land transfer), this poses potential risks to the chain itself. The very immutability of the system means that removal of data would require some version of a hard fork in order to comply with a request for data removal. There are two types of forks that can happen on a blockchain, a soft fork and a hard fork. Both are changes in the rules that govern the operation of the blockchain. Soft forks are backward compatible while hard forks are not. These differences are visualized in Figure 2.

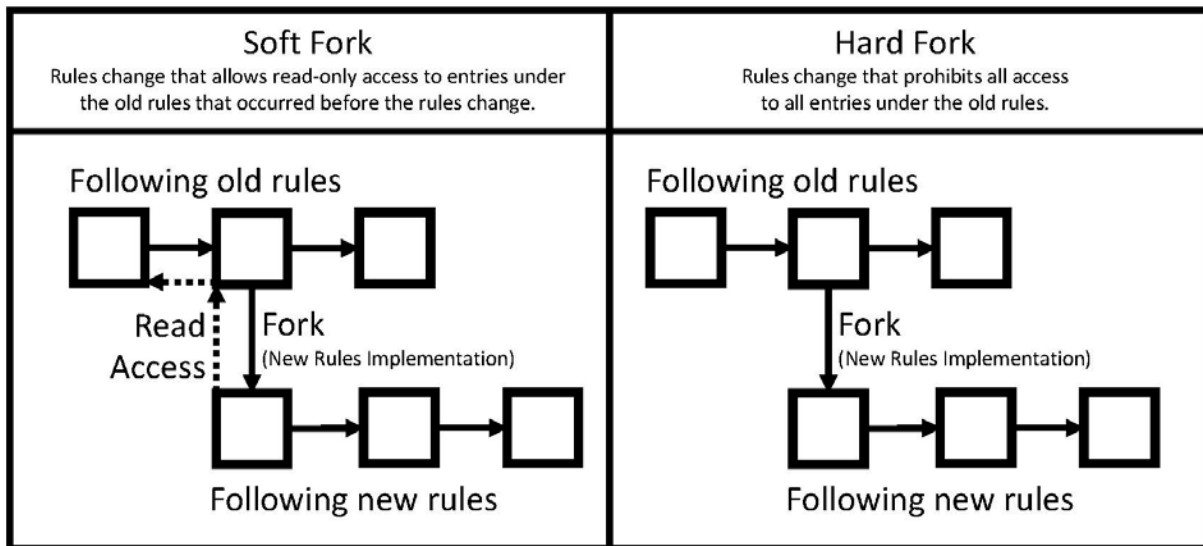


Figure 2: Blockchain forks

A hard fork is a drastic measure that can be used to invalidate previously valid data on a blockchain network by changing the rules of what is acceptable on that network and specifying what old data will be accessible by the network going forward. Ethereum has done this multiple times to deal with necessary changes to their system policies. However, while a hard fork can change the validity status of data blocks, it cannot actually remove them from the Internet, and

<https://fortune.com/2018/06/22/walmart-blockchain-patent-health-records/>; Dan Bindman.

“Land Registry successfully transfers house using blockchain.” *Legal Futures*, April 12, 2019,

<https://www.legalfutures.co.uk/latest-news/land-registry-successfully-transfers-house-using-blockchain/>; Leo Jakobson. “MIT leads coalition putting academic credentials on

blockchain.” *Modern Consensus*, April 24, 2019,

<https://modernconsensus.com/uncategorized/mit-universities-academic-credentials-blockchain/>.

would not satisfy requirements under GDPR or CCPA. To fully remove the data, it would need to be deleted from every node running that blockchain. This would essentially make blockchain unusable for storing any data subject to these privacy laws. Further, since legislation can change, it is important to consider how changes in the law could require the removal of data from a blockchain system, or even the complete shutdown of a blockchain itself. A possible alternative to storing this data directly in a blockchain is to store hash data in Merkle trees on the chain and use that to authenticate documents. However, simply storing hashes to authenticate files does not necessarily require a blockchain, and the particular use case should be reviewed to see if an easier solution exists.

In early 2019, Mark Stuart Day examined the “shutdown problem” for blockchain systems (i.e. what happens to the system when one eventually wants to stop its operation) and concluded that blockchain is not the best approach to provide authentication for permanent or long-term records because there are significant risks posed if the blockchain has too little participation by active users or nodes.⁸ In these cases, it becomes easy for dishonest nodes to provide false information—breaking the supposed immutability and trustworthiness of the system. Dishonest nodes can be avoided if one is using a permissioned or private blockchain, although that also eliminates the decentralized trust structure that is one of the main selling points of blockchain. Additionally, Day explored the challenges of maintaining trust in documents once a blockchain eventually ceases active operation. In order to continue providing documentary validation of a record, the system must remain minimally functional which might require the maintenance of a distributed set of nodes and active miners—both aspects of the system that would be expected to cease if the system were shut down.

These potential problems could be addressed through international standards organizations and consultation with experts in records management and archival theory. However, as Dr. Lemieux has observed, and as one can verify by scanning the available literature, there are few people versed in archival theory taking part in the development and deployment of blockchain systems.⁹ The primary research into blockchain is from those within the computer science and finance industries. For a technology that is fundamentally a form of recordkeeping, there is a significant lack of participation in its development from records managers and archivists. Compounding that lack is the fact that this technology is already being used for official recordkeeping. For records managers in government, industry, and academia, this is a technology that is not just coming, it is already here, being used to record information that they will be required to manage reliably, and affordably, for its entire lifecycle.

Blockchain in Archives

The implications of blockchain for archives and archivists are comparable to those for records management. Potential donors are already using this technology; researchers are developing smart contracts and developing new uses for it. In all likelihood, someday records creators and donors

⁸ Mark Stuart Day. “The Shutdown Problem: How Does a Blockchain System End?” *arXiv*, 2019, [arXiv:1902.07254](https://arxiv.org/abs/1902.07254).

⁹ Lemieux, “Recordkeeping,” 17-18.

will want to put the products of all this activity into an archives.

For archival practice, blockchain presents opportunities similar to those for records management. In the area of provenance tracking, blockchain offers a means of proving the ownership of material over time with each transfer recorded and timestamped. This type of tracking is already in use, tracking physical materials such as diamonds and other conflict minerals; De Beers and Ford Motor Company use the technology to demonstrate to consumers that their materials are ethically sourced, while Fogo de Chão uses it to provide information for consumers about the health of the beef they are eating.¹⁰ For future researchers visiting an archives, where blockchain data is available, this type of data could provide invaluable information about the provenance of materials, but only if we can ensure that it is still available and usable when that future arrives.

The British National Archives recently published a report about their proof-of-concept project ARCHANGEL. This project created a blockchain system, shared among multiple archives and memory institutions. The purpose of the ARCHANGEL blockchain network is to provide a method of proof that digital public records accessioned by these archives were neither censored, nor altered, once received. This project provides an example of how a permissioned blockchain could work to improve public trust in official archival materials.¹¹

Extrapolating slightly from this project, blockchain has some potential for facilitating collaboration between marginalized communities and archives housed at institutions with a history of ignoring and mistreating them. A current area of focus for many archives is addressing the gaps in our collections with regard to communities of color, LGBTQ+ communities, and other marginalized groups. A small permissioned blockchain shared between a community and an institution could provide the foundation for building a trusting relationship over time. The shared oversight of materials could allow the community to apply a “trust, but verify” approach to partnership with the institution allowing them to know if their records were altered or destroyed and who had done so. Or, as Norman Weekes observed in a piece for *Blacks in Bitcoin*, “... a

¹⁰ Stefan Nicola. “Using Blockchains to Help Fight Conflict Minerals.” *Bloomberg Businessweek*, April 24, 2019, <https://www.bloomberg.com/news/articles/2019-04-24/using-blockchain-to-help-fight-conflict-minerals>; Lana Bandoim, “Can Blockchain and Chip Technology Improve Beef Sourcing Transparency?” *Forbes*, April 30, 2019, <https://www.forbes.com/sites/lanabandoim/2019/04/30/can-blockchain-and-chip-technology-improve-beef-sourcing-transparency/#757460132846>.

¹¹ John Collomisse, Tu Bui, Alan Brown, John Sheridan, Alex Green, Mark Bell, Jamie Fawcett, Jez Higgins, and Oliver Thereaux. “ARCHANGEL: Trusted Archives of Digital Public Documents,” *arXiv*, 2018, [arXiv:1804.08342](https://arxiv.org/abs/1804.08342).

decentralized open sourced technology based on mathematics makes the possibility for manipulation, discrimination and exclusion extremely low.”¹²

Mark Stuart Day identifies another major implication for blockchain within archives, the shutdown problem. Once a blockchain has reached the end of its active use, people will look for ways to archive the contents of these systems or the entire systems themselves. Archivists need to understand how they operate to either extract the useful information from the system to be stored elsewhere or provide a scaled-down means of keeping a blockchain running for incoming queries against its data. Archiving a blockchain system will require specific thought, and a plan for implementation because the longer the system is left alone without a community of active users, the less reliable the data stored in it becomes, due to blockchain’s reliance on its distributed nature to create trust. Without enough active users, the potential for dishonest nodes to affect the validity or trust of the information stored in the blockchain increases. An archives accepting a blockchain system, or the data from a system, needs to understand this technology well enough to know what is important to harvest from it for storage in another form or how to approach preserving the data in some other way that still makes it useful. They need to be ready to execute archival processing promptly, and precisely, so that the integrity of the data can be preserved before dishonest nodes and spam bots destroy the trust that the system once maintained.

Conclusions and Implications for archival education

Currently, information schools (iSchools) are looking at how to educate students about blockchain technology, while exploring its uses in archival and records management fields. The University of British Columbia (UBC) began offering coursework on blockchain in 2017. UBC has a multidisciplinary research cluster, which includes Prof. Victoria Lemieux, a faculty member from the iSchool. Students at the iSchool are learning how to create blockchains using Hyperledger. That same year, San José State University (SJSU) received a grant from the Institute of Museum and Library Services to research blockchain applications for libraries. The Blockchain Project defined three micro-projects: inter-library loan fee management using cryptocurrency for micro payments; universal library card, and credentialing. As part of the grant, in early 2019 SJSU offered a massive open online course, or MOOC, regarding blockchain. The 6-week course covered the technology, as well as use cases specific to libraries, museums, and archives. SJSU has made the content of the course available to the public. Other universities are offering their students workshops or short introductory courses.

For both archivists and records managers, it is important to learn about blockchain because it will likely come up in future practice. Governmental entities are already mandating its use for official records. Businesses and academic institutions are beginning to follow suit. Learning about the technology, and how it is being used now, can help one prepare for the inevitable accession of

¹² Norman Weekes. “Why Black Folks Should Like Bitcoin and Love Blockchain,” *Blacks in Bitcoin*, December 30, 2017, <http://www.blacksinbitcoin.com/2017/12/guest-opinion-why-black-folk-should.html> (accessed May 15, 2019).

materials linked to or stored on the blockchain, and the potential accession of an entire blockchain system.

Beyond simply preparing for the future, however, it is important for archivists and records managers to be a part of the development of these systems, and discerning how people choose to use them. Blockchain, as a ledger system, is a means of recordkeeping, which is our area of expertise. We should make an effort to understand the technology, and help guide its growth, so that it uses good recordkeeping practices from the beginning.

Education about blockchain for our profession is compatible with and complementary to other digital records and archives instruction. Archivists and records managers will likely need additional training regarding the methods that blockchain uses to provide authentication of contents based on external queries, and about how a blockchain network maintains its distributed trust mechanism. However, much of our current training in digital archives (for example, validating data transfer using hashes) is transferable to the understanding of blockchain and can provide a solid foundation as we move forward with this technology.