Blockchain and Recordkeeping: Editorial

Victoria L. Lemieux

School of Information, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada; v.lemieux@ubc.ca

Abstract: Distributed ledger technologies (DLT), including blockchains, combine the use of cryptography and distributed networks to achieve a novel form of records creation and keeping designed for tamper-resistance and immutability. Over the past several years, these capabilities have made DLTs, including blockchains, increasingly popular as a general-purpose technology used for recordkeeping in a variety of sectors and industry domains, yet many open challenges and issues, both theoretical and applied, remain. This editorial introduces the Special Issue of Computers focusing on exploring the frontiers of blockchain/distributed ledger technology and recordkeeping.

Keywords: blockchain; distributed ledger technology; records; recordkeeping; records management; computational archival science

1. Introduction

Records provide evidence of business processes, activities and transactions and are information assets [1]. To serve this purpose, records must be authoritative; that is, they must be authentic, reliable, complete, unaltered and useable [1]. A records system is an “Information system which captures, manages and provides access to records over time” [2], with a well-designed records system being one that will enable the creation, use and preservation of authoritative records. All organizations will have at least one, and generally more than one, records system. Records systems can consist of technical elements such as software as well as non-technical elements including policy, procedures and stakeholders [2]. Distributed ledger technologies (DLT), including blockchains—defined by the International Standards Organization with the input of over 300 international experts from 50 countries as a ‘distributed ledger with confirmed blocks organized in an append-only sequential chain using cryptographic links’ [3], with a distributed ledger being a ‘ledger that is shared across a set of (distributed ledger technology (DLT)) nodes and synchronized between the DLT nodes using a consensus mechanism’ [3]—combine the use of cryptography and distributed networks to achieve a novel form of records system designed for tamper-resistance and immutability. Over the past several years, these capabilities have made DLT increasingly popular as a general-purpose technology used for recordkeeping in a variety of sectors and industry domains. Indeed, developers were already experimenting with this possibility in the early days of DLT, as evidenced by a number of what former lead maintainer of Bitcoin’s codebase is reported to have referred to as ‘bizarre hacks’ to embed content into the Bitcoin blockchain [4]. Observers of these early developments took note, however, recognizing the potential of the underlying ledger to meet the need for ‘a trustworthy record, something vital for transactions of every sort’ [5].

Though many DLT systems purportedly seek to offer trustworthy records and recordkeeping, there is a noted absence of records management and archival knowledge in much of the literature discussing the design and implementation of these systems. This Special Issue seeks to address this gap and contributes to ongoing work in the area of ‘computational archival science’ [6], bringing the records management and archival perspective to research on the design and application of DLT in recordkeeping. The papers in this Special Issue illustrate a number of recordkeeping use cases, covering the application of DLT in digital preservation, government recordkeeping, public procurement, supply chain...
management and financial payments. Each paper discusses the need for authoritative (trustworthy) records in a specific context, the value that DLT offers, and how relevant recordkeeping theory, principles and standards apply and open research challenges. In each case, the papers not only discuss technical elements, but also how technical elements interact with social and informational elements of solution design and implementation to realize the key advantages of using DLT, that is, trust across an entire DLT solution operating as a socio-informational-technical system [7].

2. Important Considerations in DLT Recordkeeping Use Cases

Each paper in this Special Issue highlights a number of key challenges to be considered when designing and implementing DLT solutions for recordkeeping and offers examples of how these challenges might be addressed. While not all of these considerations are unique to recordkeeping use cases, nor do they represent an exhaustive list of DLT recordkeeping system requirements, they do emerge in the papers in this Special Issue and elsewhere as recurring themes across recordkeeping use cases and thus can provide designers and developers of DLT records systems with useful guidance.

2.1. Use of Records Management and Archival Theories, Models and Standards in DLT System Design

As already noted, many DLT solutions are designed and implemented without reference to existing records management or archival theories, models and standards, which often, as a result, produces sub-optimal outcomes and unintended real-world consequences that lead to the diminution of the value of records as evidence or assets. A number of papers in this Special Issue highlight the importance and value of drawing upon existing records management and archival knowledge in the design of DLT solutions for recordkeeping. Stančić and Bralić [8], for example, provide a data model of a record and model its relationship to a digital signature and an archival bond, as well as discussing the concept of the archival bond, which expresses the fact that records derive their meaning not just from their semantic content but also from their relationship to other records and the activities that give rise to them and in which they participate. Weingärtner et al. [9] introduce the Generally Accepted Recordkeeping Principles and map this well-known records management model to features of blockchains.

2.2. Records Authenticity

The need to establish authenticity has received much recent attention in the literature on non-fungible tokens (NFTs) and blockchains (see, e.g., [10]). Authenticity (i.e., in simple terms, that the record is what it purports to be) is a critical feature if distributed ledgers are to be relied upon as evidence and assets. Weingärtner et al. [9] grapple with this issue in the context of public procurement documentation. In their solution design, a smart contract checks the authenticity of the records presented by the winning bidder to execute the next transaction through an external call to a public agency database (e.g., a call to the revenue agency database to verify the authenticity of a tax regularity certificate). This checking process must be registered in the public procurement documentation on the blockchain, and it is not possible to award a contract between the government and the winning business without it. Aside from assuring the authenticity of bid documentation, this process also improves the reliability of the records, since the smart contract encodes and executes a regulated procedure to attest to the veracity of the certificates presented by the winning bidder.

Weingärtner et al.’s [9] approach to establishing the authenticity of bid documentation highlights another common challenge in the design of DLT solutions for recordkeeping: the well-known ‘oracle problem’ [11]. Services that update a distributed ledger using data from outside of a DLT system—called ‘oracles’—must be guaranteed to be trustworthy. Questions about the trustworthiness of these services give rise to the oracle problem. Weingärtner et al. [9] caution that, due to possible threats to system security, oracle services must be designed and implemented carefully. To retain the decentralized nature of a
blockchain and avoid introducing single points of failure into the process, they advise using multiple, independent oracle services providing the same information, which allows a majority decision in case of contradictory information.

A related issue when establishing the authenticity of records is to know that the entity purporting to be the creator of the record is who they say they are. Thus, establishing records creators’ identity and authorization or competence to participate in a process is key to authenticity. To overcome this tension, Weingärtner et al. [9] suggest using a Self-Sovereign Identity (SSI) [12]—a decentralized identity that is managed by the entity that the identity distinguishes, or its representative, and that allows that entity to make cryptographically verifiable claims, typically using Verifiable Credentials, about its identity. Weingärtner et al. [9] use the SSI identity layer in their public procurement DLT recordkeeping solution to ensure that suppliers participating in public procurements are compliant with legal prerequisites such as the correct payment of taxes or compliance with working conditions. These prerequisites can be tender-specific, such as quality seals or permission to manufacture medical devices or reused for multiple bids. Weingärtner et al.’s [9] design allows each vendor to create its own identity, which is then certified by an official body with proof that is stored on the blockchain. Additional certifications can be linked to this identity by authorized issuers, and all certifications are stored on the blockchain. During the verification phase of the procurement, the required certifications are checked without the involvement of the individual issuers. Certifications also can have expiration dates or can be revoked.

2.3. Privacy

DLTs’ assemblage of cryptography, distributed networking, transparent ledger and consensus to incentivize honesty in network participants (or trust without a central intermediary) is novel. It does, however, create a challenge relating to privacy and tracking that must be addressed in many use cases involving recordkeeping. In their paper on the application of DLT to track the distribution of donor funds, Rehman et al. [13] acknowledge the tension between transparency and privacy, noting that transparency might not be desirable in the context of transmission of donor funds but also that too much privacy can undermine the integrity of a DLT. In the context of their solution, they propose several approaches to achieving the necessary balance between transparency and privacy.

To keep details of transactions private, they suggest placing only an encrypted version of a transaction on ledger, with the transaction being encrypted using the private key of one of the participating members of the DLT. To check the validity of the transaction, the parties involved in the transaction verify that the details of the transaction are consistent (e.g., they conform to predefined input controls). If there is ever a need to look into the details of the transaction, the NGO or donor that encrypted the transaction is publicly known and may be contacted. In the Wang and Yang paper [14] discussing a government recordkeeping use case, the authors observe that an administrative agency may require confidentiality and security, and hence, a private blockchain is more suitable even if such a DLT system is slightly less stable and secure. To mitigate this risk, they suggest that several closely cooperating institutions or authorized groups should participate in operation of the private network. Henninger and Mashatan [15] advocate for the use of SSI in the context of supply chain recordkeeping, citing it as a way to give users of the system greater control over the data they wish to disclose. They note that the use of SSIs in DLT-based supply chains is a fundamentally different approach to privacy than Privacy by Design [16], which still depends on data stewards and the ethical processing of personal data. With SSIs, the ownership, control and decision-making regarding the disclosure of records shifts to the person that the data are about.

2.4. Records Storage

DLT records system architectures can use a number of different models for records storage. For example, all records can be stored on-ledger, or alternatively, only records
metadata can be stored on-ledger, while the records themselves are stored off-ledger. Further, records stored off-ledger can be stored in a centralized local or cloud data store, or in a decentralized data store (e.g., Inter Planetary File System (IPFS) [17]). In their paper, Stančić and Bralić discuss the various options for records storage, considering three different digital archive models: ARCHAIN [18], Cilegon E-Archive system [19] and Lekana [20]. In addition to these archival systems, they considered several general data storage systems including BigChainDB [21], ChainSQL [22], EthernityDB [23] and Mystiko [24] and show how their blockchain solution can be integrated with the use of distributed noSQL database systems (e.g., Cassandra [25] and MongoDB [26]).

2.5. Records Deletion and Modification

In traditional records systems, such as databases, records can be deleted or modified. While this capability makes correcting errors or updating records easier and provides a means to comply with legislative or other requirements to remove records, it also makes it easier to alter records fraudulently or inadvertently. On the other hand, the relative immutability of DLTs, while providing better protection for records against tampering or manipulation, means that it can be challenging to delete, cancel, correct and update records when necessary for legitimate purposes. In the context of donor payments to educational NGOs, for example, Rehman et al. [13] cite the need for cancellation of transactions, such as when a student no longer requires donor funds. To provide the capacity to cancel transactions whilst still attending to the need for security and immutability of the ledger, the authors describe a process for recording cancellation messages on chain and an ordered approach to broadcasting the cancellation to other nodes. They also propose a novel method of computing timestamps for both regular transaction and cancellation records. The paper by Stančić and Bralić [8] also addresses this issue as it manifests in the conflicting requirements of blockchain immutability and the archival bond, which, they note, is changeable until records become inactive. They propose a supporting data system separate from the blockchain to enable users to alter certain metadata information (e.g., add archival bond information which might be created after the record has entered the system), which also allows the blockchain to be indexed and easily searchable. In order to achieve these goals, the system cannot rely on an immutable data structure, so a dual storage system is proposed, consisting of an immutable blockchain core, which guarantees data integrity, and a partially mutable supporting system.

2.6. Descriptive Metadata

Metadata—structured or semi-structured information, which enables the creation, management and use of records through time and within and across domains [27]—is an essential component of any records system. Metadata for records captures such information as the business context of the records, dependencies and relationships among records and records systems, the records relationship to legal and social contexts, and relationships to agents who create and manage records [1]. Some records metadata is automatically derived or attributed at the time of records creation, and other metadata will be manually or automatically added to records over time as part of records management and digital preservation processes. The need to update metadata requires that DLT records systems have flexibility to add new records metadata. At the same time, records metadata must be protected from unauthorized deletion, alteration and manipulation, and logical relationships or linkages between a record’s content and its associated metadata must be created and maintained using automated or manual processes [1]. This is a topic specifically addressed in the paper by Stančić and Bralić [8], in which they suggest that their immutable blockchain core coupled with a partially mutable supporting system addresses the need for both updating and data integrity protection. Metadata for records also must be described and documented in authoritative metadata schemas. The paper by Rehman et al. [13] discusses the challenge of creating, updating and ensuring the authenticity of such schemas. In their solution design for donor funds tracking, they describe a novel process for defining
an authoritative donor/NGO list. Updating of the list is only allowed when it is signed by not just two, but a significant number of NGOs as well as donor agencies. These signatures are included in the description of the table fields and indicate that a certain number of NGOs and donor agencies have put their reputation on the line in case the update is found to be faulty.

2.7. Discoverability and Accessibility

In their paper, Stančić and Bralić [8] discuss the inefficiency of discoverability and retrieval of records in DLT systems, especially when block or transaction IDs or dates are unknown. Metadata for records can support improved discoverability and accessibility of records by providing information that may be needed to retrieve and present them, such as identifiers, format or storage information [1]. In their paper, Stančić and Bralić [8] compare four relevant records metadata standards—DACS [28], ISAD(G) [29], PREMIS [30] and one general digital object description standard, Dublin Core [31]—and illustrate their use to aid records discoverability and retrieval in the context of their DLT solution design.

2.8. Balancing Decentralization and Centralization

As noted above, in many cases, in order to protect the privacy and confidentiality of transaction records, solution designers re-introduce a degree of centralization and authorization into the architecture and operation of the DLT. For example, they might use private, permissioned DLT solutions, that is, solutions that are accessible to only a limited number of users and for which authorization is needed to participate in sending/receiving transactions or in the operation of the network. This arrangement, argue Henninger and Mashatan [15], is typical of existing supply chain management DLT solutions, which tend to have a certain degree of centralization in which solution providers maintain control over the blockchain. In a similar vein, Rehman et al. [13] suggest a private, permissioned solution for their donor funds tracking solution, but seek to mitigate the risks to security and integrity of the ledger that more centralized solution architectures can introduce using a novel consensus-based approach to ensuring only trusted NGO or donors can be added to the network. Specifically, the solution envisions a distributed list of valid donor agencies maintained by each node. When a node receives a request to add a new NGO or donor to its list, if the majority of nodes have the requesting donor agency in their lists, the NGO or donor agency is allowed to join. Rehman et al. [13] note that the list is maintained outside of the ledger, so human intervention is required. The paper further contributes a novel approach to encouraging ongoing node participation for networks that rely upon persistent node participation to maintain the security properties of the network. In the Government of Korea use case, Wang and Yang [14] describe a policy-based approach, rather than a technical solution, to protecting the security of the system against the manipulation or collusion of a limited number of network participants, as can be the case with the more centralized private, permissioned DLTs. For example, they argue that administrative agencies should be encouraged to participate as nodes in the blockchain network through enforcement and incentive strategies (e.g., budgetary measures). In contrast, Weingärtner et al. [9] suggest the use of a public, permissionless DLT, given that transparency is the overriding consideration in the context of protecting public procurement processes from fraud and corruption. In order to prevent spam transactions, which can be a problem in permissionless systems, they propose the introduction of a deposit, which must be paid to prevent spam procurement offers, noting that the deposit is reimbursed after the final supplier has been determined.

2.9. Interoperability

DLT interoperability—whether at the level of data semantics or technical protocols—is a key challenge faced by most use cases. This is a topic addressed at some length in the review paper presented by Henninger and Mashatan [15] given that effective global supply chain management typically requires integration of many interconnected systems and
processes that create, use and exchange records that were not designed to interact with one another. They note that, as goods travel across the supply chain between multiple actors who each use varying formats for their records, the associated records become fragmented, are transformed, migrate across hardware and software configurations and even migrate between manual and automatic data processing functions. This, note Henninger and Mashatan, slows down the supply chain process, is inefficient, can introduce errors and can impact the reliability of records. As such, they argue that the key to unlocking the full potential of supply chain management using DLTs is interoperability across participating records systems and networks. They observe that most of the technologies addressing interoperability propose validators to bridge between different blockchain networks (see, e.g., [32]).

3. Open Research Challenges and Conclusions

Despite the many benefits that DLT, including blockchains, can bring to recordkeeping, and the existence of a number of approaches to address the challenges, there remain many open challenges and issues that must be addressed if the value of DLT records systems is to be fully realized. These open challenges are both of a theoretical and applied nature.

In relation to more theoretical challenges, there is a need for solutions that combine computing and engineering knowledge with records and archival knowledge. This begins with greater understanding of the important differences between terminology that conveys very different meaning in the context of computing and software engineering than in archival science. The term record is one very important example of this. In computing, a record might be understood as somewhat synonymous to a row in a classical database, while in records and archival conceptualizations it is more often thought of as a representation (e.g., a document) affording evidence of transactions and, ideally, includes metadata about the social and business context of records creation, management, use and preservation. The development of a standard international vocabulary relating to DLT and blockchain technology, such as exemplified in the work of the ISO Technical Committee (TC) 307 [3] in addition to work of the Joint Working Group of ISO TC 46 and TC 307 on DLT and records management, will help to provide those engaged in DLT system design and implementation with shared conceptualizations and a common working language for use in the application of DLTs for recordkeeping.

In terms of the need for more applied research and development of a technical nature, there also remain many open challenges. Human usability challenges are not really discussed in the papers in this Special Issue, but usability issues remain an important barrier to widespread adoption of cryptography-based decentralized trust systems [33] and have also been identified as preventing widespread adoption of classic centralized records systems [34]. Much more work is needed on the human-centred design and usable security aspect of DLT-based records systems. Despite that AMIs have responsibility for long-term preservation of records, both the paper by Wang and Yang [14] and the paper by Weingärtner et al. [9] acknowledge that it is still challenging to assure the preservation of records for long periods in such a novel technology as DLT. Another open challenge is the ‘oracle problem’. Though DLTs can support records immutability and reliability and address the double-spending problem, they do not automatically reflect real-world events and cannot ensure the authoritativeness of records manually entered into DLT systems. While the oracle problem might be solved when all information is recorded using DLT, as Henninger and Mashatan [15] suggest, this eventuality is a long way off, so novel solutions to the oracle problem are still required. Similarly, the problem of DLT system interoperability is not solved, but realization of DLT recordkeeping at scale will require it. While many research and development efforts are focused on solving these technical challenges, not all of them do so with an eye to the requirements for recordkeeping. There is, therefore, a need to evaluate emerging approaches through the lens of records management and archival knowledge about the requirements for creation and preservation of authoritative records.
At this point, research which combines computing and engineering research and development with archival research and development is still quite shallow, as evidenced by the number of papers accepted for this Special Issue. The hope is that the excellent contributions to this Special Issue will stimulate future work and many more contributions to this journal that combine these two areas of knowledge for the design of innovative and effective DLT solutions for recordkeeping.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The author would like to express appreciation of all contributors to this Special Issue and thank the editorial staff for support and assistance with this Special Issue.

**Conflicts of Interest:** The author declares no conflict of interest.

**References**


8. Stančić, H.; Bralić, V. Digital Archives Relying on Blockchain: Overcoming the Limitations of Data Immutability. Computers 2021, 10, 91. [CrossRef]


11. Caldarelli, G. Understanding the Blockchain Oracle Problem: A Call for Action. Information 2020, 11, 509. [CrossRef]


13. Rehman, E.; Khan, M.A.; Soomro, T.R.; Taleb, N.; Afifi, M.A.; Ghazal, T.M. Using Blockchain to Ensure Trust between Donor Agencies and NGOs in Under-Developed Countries. Computers 2021, 10, 98. [CrossRef]

14. Wang, H.; Yang, D. Research and Development of Blockchain Recordkeeping at the National Archives of Korea. Computers 2021, 10, 90. [CrossRef]

15. Henninger, A.; Mashatan, A. Distributed Interoperable Records: The Key to Better Supply Chain Management. Computers 2021, 10, 89. [CrossRef]


