Metadata Issues in Blockchain Technology

Ellen Phillips

INFO 285 Metadata

Literature Review

April 16, 2018

Introduction

Blockchain is a new technology that is poised to radically change many business processes. While the most famous use is the cryptocurrency Bitcoin, other industries with a need for secure, yet verifiable transactions, as well as those with complex supply chains, are starting to adopt blockchain. Libraries and publishers are have taken notice and new initiatives are announced every day. This paper will give a brief overview of blockchain technology and how they can be used to manage administrative metadata such as access, rights management, and preservation. It will look at some other uses such as improving peer review for scholarly articles and safeguarding free speech by providing an alternative way to register and maintain web addresses. Included in the discussion will be an examination and critique of the current state of blockchain technology for metadata management with a focus on how well blockchain can support the complex metadata systems that libraries rely on.

Definition & Context

Blockchain is a distributed record of transactions that can help guarantee trust when two parties do not know or trust one another. Casey and Vigna (2018) said it was a representation of what “cryptographer Ian Grigg described as “triple-entry bookkeeping”: one entry on the debit side, another for the credit, and a third into an immutable, undisputed, shared ledger.”

In traditional financial systems, a trust agent, typically a bank, will be the intermediary for secure, transactions. They may even assure anonymity and protect privacy rights like a Swiss bank account. Blockchain removes the need for an agent and

replaces it with a public register where trust is built into the system because it is distributed.

Libraries are used to being a trusted intermediary, a central source for a publicly shared information base. Garcia-Barriocanal et al. (2017) wrote that blockchain “has the potential to impact the economics of current archival institutions, their funding models, and eventually, their archival cycles and responsibilities (p.40). In order for blockchain to succeed in libraries, there will need to be an infrastructure and user interfaces that enable easy participation.

Potential Uses

There are many potential uses for blockchain including rights management, preservation, scholarly publishing, and even safeguarding free speech. Garcia et al. (2017) observed that “pricing, rights management, and interchange are not granted as a direct consequence of using a blockchain, but the blockchain enables the creation of new mechanisms for such applications” (p.41). Moreover, there seems to be a desire to cultivate “polycentric governance principles [for] the digital preservation community” (Whitt, 2017, 198). These two desires, the commercial exploitation of a new technology coupled with a societal need to create more transparent shared resources point toward a landscape that is receptive to this type of disruptive innovation.

Rights Management

Being able to establish copyright is important to all authors, but it is particularly important to anonymous and pseudonymous creators, who have little to no protection for their works. Bell (2016) said that blockchain can “give pseudonymous authors robust

identities over time, allowing them to develop reputations and increase credibility of their claims to hold good title to transferable copyrights" (462). The ability to track derivative works is also of great importance to certain types of content creators. These include composers, choreographers, musicians, and playwrights, just to name a few. Blockchain will allow them to more easily track works that are associated with their original creations. Rights management for open access publications could be much improved with blockchain as versions will stay together along with rights information circumventing the need for a lot of Internet searching and flat storage of rights metadata.

Provenance

Preservation metadata seems to be a very likely candidate for blockchain. One major advantage is that it has the potential to reduce the amount of work that it takes to check and verify metadata stored in the cloud. As Liang et al. (2017) stated the current tools are"not effective in cloud computing systems . . . due to several layers of interoperating software and hardware components spread across geographical and organizational boundaries" (469). They detail a blockchain system called ProvChain that performs audits of cloud-based entities and produces proof in the form of a block (476). This block is the product of a federated search across several cloud platforms, and so addresses a key piece of interoperability between business systems.

Scholarly Publishing

The entire lifecycle of a scholarly publication runs on metadata. This includes all of the types of metadata outlined above, and it also must accommodate peer review. A major criticism of scientific publishing is reproducibility, the ability to replicate the results

of an experiment (Economist 2016). Flawed methods and skewed results should be detected in a rigorous system of peer review, but are often missed. Systems that are able to produce a more open, transparent peer review through blockchained metadata can help solve this issue.

Commercial publishers are interested in blockchain for open peer review. Digital Science announced in early 2018 that they were looking for partners who wanted to join them along with Springer Nature and Katalysis Publishing to explore “practical solutions that leverage the distributed registry and smart contract elements of blockchain technologies” for scholarly communications. Katalysis is already using blockchain with their product Katalysis DecPub; a Digital Ownership and MicroPayment solution. This is a free Wordpress plugin that utilizes blockchain to track rights and control access (Katalysis, 2018).

Safeguarding Free Speech

If information sources can be decentralized, they can be a safeguard against state and/or corporate censorship. Namecoin.org (n.d.) describes themselves as being an “experimental open-source technology which improves decentralization, security, censorship resistance, privacy, and speed of certain components of the Internet infrastructure such as DNS and identities.” One of the primary uses of Namecoin is that you can register a .bit domain. These are outside of the ICANN registry and therefore not prone to either government or corporate intrusion.

Although the data size of the blocks in Namecoin is limited to 520 bytes, their provenance as the “first altcoin from Bitcoin with its own blockchain” (Kalodner et al. 2015, 2) could mean it is well-positioned to play a major role in shaping the infrastructure

of this technology. However, Kalodner et al. also stated that Namecoin is “a system in disrepair” (2), few domains have content that is deemed “non-trivial” (1, 2, 12) , and that the “market for domains is thin-to-nonexistent” (1). Compounding the problem is that Namecoin will not scale up well as the “supply of names that are memorable and meaningful to humans is scarce” (2). However, the real potential for metadata here is that Namecoins can support namespaces (Namecoin.org, FAQ, n.d; Kalodner et al., 2015).

Technical Underpinnings: Limits to the current state

The distributed nature of blockchain requires that users download the entire chain before adding new items (called blocks) to it. This ensures that each copy of the database is identical. According to Scott (2018), it is not only very costly but also “very hard to compute an alternate chain . . . which will catch-up with and overtake the one true chain.” This also impacts computation times. In Bitcoin for example, it can take up to ten minutes to verify the transaction and make the block (Yi-Huumo, 2016).

The technology relies on hashing where files are encrypted and compressed to a 256 character string. This not only controls access to the files, but it also keeps the chain smaller. This poses some limitations, however, since large amounts of data cannot be stored, which might point to the persistent need to trust a centralized intermediary. Gupta (2017) and García-Barriocanal (2017, p.42) both point to the use of the InterPlanetary File System’s (IPFS) use of blockchain on their Peer to Peer (P2P) platform in order to facilitate the transfer of large files. Gupta also named several other blockchains that can be used for larger documents such as Ethereum which was designed to handle contracts (para. 6), as well as Filecoin, Enigma, and Storj (para. 7).

Data storage is not the only issue, however. As Chow (2016) said, “One critical characteristic of a secure cryptographic hash function is that it is one-way. This means that from the output, it is virtually impossible, or mathematically and computationally improbable, to determine what the input is.” García-Barriocanal (2017) et al. described several large structural issues around the use of blockchain for bibliographic needs. This includes a lack of indexing although “using conventional indexing and retrieval engines as Apache Lucene is an option, . . . it requires a copy of the resources to be indexed, thus becoming a trusted party” (p.41). Any solution that involves storing things off-chain is not operating in a trust free environment.

Although García-Barriocanal et al. felt that some descriptive metadata could be captured along with provenance metadata, they did not feel it would work for showing relationships between entities (p. 40-1). An additional point of concern was the lack of semantic interoperability brought about because “decentralization entails the exposure of a heterogeneity of autonomous, incompatible media repositories and it is unlikely that there will ever exist a single agreed-upon metadata schema” (p.41).

There is also a matter of cost. If Bitcoin is any indication, transaction fees can spike unexpectedly. As Gulker (2017) observed average Bitcoin transaction fees increased by 1186% in the first half of 2017. While some of this increase is due to financial speculation, there are serious environmental ramifications that should be addressed as well, as much of the cost is seen in the form of energy consumption (Serrells, 2018, para. 7, Yli-Huumo, 2017).

A (very) Short Library Critique

Does blockchain offer any real benefit or is it just a novelty? According to MacManus (2018) “we’re in the Geocities era of blockchain.” Indeed, an analysis of content hosted on .bit domains revealed that most had content that the authors felt was a trivial use of the technology (Kalodner, 2015, 1, 2, 12). How does it differ from other forms of peer review; how will it fix a potentially broken business model as was asked several times in the comments section of a recent online article (Meadows, 2018). It comes down to a matter of trust. Institutions are trusted based on how they behave. Given the changes in scholarly publishing, nothing seems certain at this time.

Conclusion

There are many potential applications for blockchain technology, but the technology is still in an immature state and cannot support all of the metadata needs of a digital library or repository. The inability to store data in the blocks is a major hurdle since it leads to back to the problem of trust as data must be stored off of the blockchain. Blocks rely on hashed versions of files. Files are reduced to 256 character strings, and it is not possible to dereference them. However, many metadata systems rely on shared resources. The issue of trust could be found again in a different way. The fact that there is a developed system for bibliographic metadata that is centered around e-commerce suggests that industry will provide some of the infrastructure. It is wise to wonder, whether or not this will create new trust issues. Like all new technologies, ambiguity and uncertainty abound.

References

Casey, M. & Vigna P., (April 9, 2018) In blockchain we trust. MIT Technology Review. Retrieved from https://www.technologyreview.com/s/610781/in-blockchain-we-trust/

Chow, J. (January 2, 2016). Blockchain Underpinnings: Hashing, light bulbs, bits, and bytes. [Blog post]. Retrieved from https://medium.com/@ConsenSys/blockchain-underpinnings-hashing-7f4746cbd66b

Economist Editorial Staff, The. (February 6, 2016). The scientific method: Let's just try that again. The Economist. Retrieved from https://www.economist.com/news/science-and-technology/21690020-reproducibility-should-be-sciences-heart-it-isnt-may-soon

García-Barriocanal E., Sánchez-Alonso S., Sicilia MA. (2017). Deploying Metadata on Blockchain Technologies In: Garoufallou E., Virkus S., Siatri R., Koutsomiha D. (eds) Metadata and Semantic Research. MTSR 2017. 755 Communications in Computer and Information Science. Retrieved from https://doi.org/10.1007/978-3-319-70863-8_4

Giddings, J. March 7, 2018. Get Involved in Our Initiative to Explore Blockchain Technologies for Peer Review. News Blog: Shaking up science. Retrieved from https://www.digital-science.com/blog/news/blockchain-technologies-for-peer-review/

Gulker, M. (June 23, 2017) Who Pays to Keep Bitcoin Running? American Institute for Economic Research. Retrieved from https://www.aier.org/blog/who-pays-keep-bitcoin-running

Gupta, S. (June 2, 2017) Storing documents in blockchain. [Blog post.] Retrieved from https://medium.com/blockchain-musings/storing-documents-in-blockchain-1d4d361ca962

Kalodner, H., Carlsten, M., Ellenbogen, P., Bonneau, J., & Narayanan, A. (2015). An empirical study of Namecoin and lessons for decentralized namespace design. [Working paper]. Retrieved from https://prod.coss.io/documents/white-papers/namecoin.pdf

Katalysis. (2018). We are doing this for: Online publishing. Retrieved from https://www.katalysis.io/

Liang, X, Shetty, S., Deepak, T., Kamhoua, C., Swiat, K., Njilla, L. (May 14, 2017). ProvChain: A Blockchain-based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability. CCGrid '17 Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing
Pages 468-477

MacManus, R. (March 20, 2018). We're in the Geocities era of blockchain. Blocksplain [Blog]. Retrieved from https://blocksplain.com/2018/03/20/were-in-the-geocities-era-of-blockchain/

Meadows, A. (April 3, 2018). Everything* You Always Wanted To Know About Blockchain (But Were Afraid To Ask). Scholarly Kitchen [Blog]. Retrieved from https://scholarlykitchen.sspnet.org/2018/04/03/everything-always-wanted-know-blockchain-afraid-ask/

Serrells, M. (April 16, 2018). Inside the coal power plant opening its doors to bitcoin mining. C|NET. Retrieved from https://www.cnet.com/news/blockchain-coal-power-plant-mining-bitcoin-cryptocurrency/

Whitt, R.S. (2017). Through A Glass, Darkly" Technical, Policy, and Financial Actions to Avert the Coming Digital Dark Ages. Clara High Tech. L.J. 33(117) 2017. Retrieved from http://digitalcommons.law.scu.edu/chtlj/vol33/iss2/1

Yli-Huumo et al. Where Is Current Research on Blockchain Technology?—A Systematic Review. PLoS One 11(10) 2016. DOI: 10.1371/journal.pone.0163477