ALA Neal-Schuman
CHICAGO | 2020

ALA CENTER FOR THE FUTURE OF LIBRARIES

BLOCKCHAIN

EDITED BY SANDRA HIRSH AND SUSAN ALMAN

LIBRARY FUTURES 3

libraries to use blockchain technology in credentialing in the mastery of information literacy units. Libraries could also educate community members about how blockchain technology can help them achieve self-sovereign identity.

DISTRIBUTED ACCESS TO LIBRARY METADATA

Timothy A. Thompson, Discovery Metadata Librarian, Yale University Library

Authority control, controlled vocabularies, and the intellectual and physical control of collections—the idea of control is part of a professional ethos that tends to privilege centralization and uniformity in the organization of information. Centralization on the level of organization and representation maps to centralization on the level of maintenance and exchange: libraries rely on gatekeepers and centralized services to manage knowledge organization systems such as the LOC Subject Headings and related authority files. Ultimately, questions of control are questions of trust: centralized systems and workflows allow libraries to define boundaries within which community norms and standards can be enforced. At the same time, they are exclusionary and limit participation to a set of authorized contributors. Blockchain technology has the potential to reconfigure the relations of information exchange among libraries and to shift the locus of trust from centralized services to distributed systems built on the premise of peer-to-peer interactions.

As an illustration, consider the example of the NACO-México project, which began in 2003 as a cooperative effort among academic libraries in Mexico to contribute records to the Name Authority Cooperative

Program (NACO), maintained by the Program for Cooperative Cataloging and the LOC.[60] Currently, an institution that wishes to contribute to the Name Authority File must do so through one of two authorized gateways, OCLC or SkyRiver, both of which charge a membership fee.[61] In 2011, participants in NACO-México were compelled to move from OCLC to SkyRiver because they could no longer afford the cost of OCLC membership. Subsequently, the project dealt with discrepancies in its contribution statistics as recorded by the LOC, which has reported contribution totals that were lower than the project's internal numbers. If NACO itself were to run on a blockchain network rather than as a centralized service, membership fees might be eliminated or replaced by marginal transaction fees, making participation affordable for a broader range of international contributors. Discrepancies in contribution statistics could be eliminated through reference to the blockchain as a permanent, time-stamped record of transactions. Individual libraries might sell or purchase data on demand from peer contributors, thereby decreasing local workloads and creating new opportunities for international collaboration.

For several years, academic libraries have been discussing and undertaking efforts to migrate from legacy metadata formats to linked open data. This transition has been a difficult one, however. Although work developing semantic ontologies such as BIBFRAME has allowed the library community to examine and reconsider some of its fundamental data models, the implementation of linked open data in academic libraries has been impeded by the absence of an underlying computational

60. Julia Margarita Martínez Saldaña, "Informe sobre las actividades del proyecto NACO-México," PowerPoint presentation, Sites/Cites, Texts, and Voices in Critical Librarianship: Decolonizing Libraries and Archives, Seminar on the Acquisition of Latin American Library Materials (SALALM) 63, Mexico City, Mexico, July 2, 2018.

61. https://www.oclc.org; https://theskyriver.com.

architecture that can support new models for data-sharing and production.[62] A major selling point of linked open data is its support for internationalization and integration with the wider web of data. However, without an open, distributed market for the exchange of data, centralized bottlenecks will continue to undermine attempts at systemic change.

Many information professionals have been skeptical of blockchain applications for the cultural heritage domain, echoing the standard critique of blockchains in general: that blockchain applications have been peddled as a panacea when in reality these applications are little more than inefficient databases that offer limited functionality in exchange for troubling amounts of energy (in Proof-of-Work systems) or wealth (in Proof-of-Stake systems).[63] Even leading Bitcoin advocates such as Jimmy Song have argued that the constraints imposed by blockchain technology make it appropriate for only a very limited set of applications (in his view, currency and the exchange of value).[64] For individuals and organizations that are investigating blockchains as a technical solution, it is important from the outset to establish a framework for evaluating their applicability and appropriateness.[65] On its own terms, a blockchain is simply a means to an end, one possible approach to achieving consensus among nodes in a distributed network—and doing so even in the presence of arbitrary system failure or malicious behavior (the so-called Byzantine Generals' Problem in computer science). The question for libraries is whether participation in decentralized networks is desirable

62. https://www.loc.gov/bibframehttps.
63. https://web.archive.org/web/20180812122624/https://go-to-hellman.blogspot.com/2018/06/the-vast-potential-for-blockchain-in.html.
64. https://web.archive.org/web/20181031220327/; https://medium.com/@jimmysong/why-blockchain-is-hard-60416ea4c5c.
65. Brian A. Scriber, "A Framework for Determining Blockchain Applicability," *IEEE Software* 35 (July/August 2018): 70–77.

for a given use case or can help further their core mission in ways that may not have been possible before.

For example, the problem of entity resolution, also known as record linkage or data-matching, is one that has a direct impact on the work of information professionals in academic and research libraries. In library units responsible for catalog management, many workflows center on the procedure known as copy cataloging, which aims to expedite the processing of new acquisitions. Copy cataloging involves searching a shared database for records created by another cataloging agency, but which describe identical publications that have been acquired by one's local institution. In the current environment, the global exchange of library catalog data is controlled in large part by OCLC. Although OCLC provides data aggregation and storage services that allow libraries to share their data, this vendor-driven paradigm entails the acceptance of a business model that, in effect, charges libraries for serving their own data back to them, albeit enhanced with different forms of added value.

Libraries have a tradition of experience with data-matching and automation, but now stand to benefit from the increasingly mainstream availability of algorithms and routines developed within the context of data science and machine learning. Sophisticated algorithms for string comparison and probabilistic record linkage have long been available, but are not widely used by libraries, with notable exceptions such as the Virtual International Authority File,[66] which is itself a project of OCLC. As machine learning tools and methods have become more accessible, large-scale, real-time access to library metadata has not necessarily followed suit. The catalog of a large academic library may contain several million records. By comparison, as of August 2018, the OCLC catalog database, WorldCat, contained 427,501,671 bibliographic records in

66. https://viaf.org/.

491 languages.[67] As long as central hubs or service providers maintain control over the aggregated metadata of research libraries, the large-scale computational analysis and utilization of this data will remain out of reach for most.

Of course, when discussing decentralization, there are a range of new technologies that should be considered. Blockchain may or may not be the most appropriate one for a particular use case—or it may need to be used in conjunction with other technologies in order to enable decentralized exchange. Several efforts are underway to develop systems for decentralized file storage using distributed hash tables, one of the most prominent being the IPFS. In a way similar to the software versioning protocol Git, IPFS uses hash values to capture the state of a file at a particular point in time and then serves it on a peer-to-peer network. IPFS hashes might be referenced as links in blockchain transactions in order to decouple the storage layer from the accounting layer.[68]

Technologies and protocols for distributed systems, such as blockchains and distributed hash tables, could allow research libraries to form robust peer-to-peer networks that would enable data-sharing on both macro and micro levels. Public blockchains such as Ethereum and Bitcoin are severely limited in the amount of data that can feasibly be stored on the chain, but alternative blockchain platforms that address this limitation have recently been developed. For example, the blockchain-based database service BigchainDB offers a robust data-storage solution while ensuring Byzantine fault tolerance and providing blockchain features such as immutability and an asset-based transactional

67. https://web.archive.org/web/20181029213224/https://www.oclc.org/en/worldcat/inside-worldcat.html.
68. https://web.archive.org/web/20181010171040/https://medium.com/@mycoralhealth/learn-to-securely-share-files-on-the-blockchain-with-ipfs-219ee47df54c.

model.[69] By running a "consortium blockchain" network using a system like BigchainDB, academic libraries could be empowered to move away from centralized models and begin managing their data collectively.[70] Instead of paying a centralized metadata hub to distribute their catalog records, libraries could use blockchain technology to share their metadata—whether in batch or as discrete bits—on a peer-to-peer exchange. Many blockchain systems support the creation of so-called smart assets, a term used prominently by the New Economy Movement project.[71] Smart assets can be modeled as nonfungible tokens that represent an object on a blockchain and allow it to be exchanged. Metadata professionals are familiar with the concept of the record or descriptive unit as a surrogate for a real object. As smart assets, metadata objects themselves could be tokenized and represented at the appropriate level of granularity, whether as linked data statements (triples) or as record-like objects. By providing an international peer-to-peer marketplace, blockchain networks could facilitate the free flow of library metadata, potentially creating new revenue streams for individual libraries and replacing some of the costly subscription services that currently predominate.

DATA COLLECTION AND ASSESSMENT

Annie Norman, State Librarian of Delaware and Director, Division of Libraries

Blockchain may provide the salvation for libraries' data challenges. Effectively quantifying and showcasing the value of libraries, especially

69. https://www.bigchaindb.com.
70. https://web.archive.org/web/20181027023321/https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/.
71. https://web.archive.org/web/20190420165029/https://nem.io/technology/.