Catagora: Shared Library Cataloguing on the Ethereum Blockchain

by

Kris Joseph

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts and Master of Library and Information Studies

Digital Humanities and School of Library and Information Studies

University of Alberta

Abstract

The 2008 debut of Bitcoin marked the first large-scale implementation of blockchain technology, and its decentralized approach to monetary systems has since been abstracted to more generalized purposes like distributed computing. Platforms like Ethereum, which function as a global, decentralized computing and data storage system, promise to bring the cost of decentralized knowledge production in line with the efficiencies afforded by the centralized, integrated computing systems that currently dominate the knowledge economy.

Blockchain technologies have been investigated for a wide range of information management purposes, but their exploration within the realm of library and information studies has largely been nascent. Though many applications within the field have been envisioned, few have been explored in depth. Among the many functions performed in the field of librarianship, the work of cataloguers—which has always been performed in a decentralized manner—represents an intriguing use case. A review of current shared-cataloguing practices reveals that catalogues have become largely-centralized, divorced from public participation, dominated by an ethos of efficiency at the cost of quality, and essentially unaltered since the shift from physical to electronic catalogue storage more than 40 years ago.

The evolution of blockchain technologies, paired with an intentional approach to shared catalogues that is open for use, transparency, and public participation, is explored in a conceptual framework and design based on the Ethereum platform. A theoretical design scheme grounded in the affordances of Ethereum, shaped by the principles of open source software development, and guided by the best practices of existing social information production systems results in a proposal for *Catagora*: an open source, open-for-use, transparent and participatory shared-cataloguing platform that reverses the trend towards architectural and political centralization and promises novel catalogue features such as complete revision history and distributed collaboration on the content and quality of catalogue entries.

Blockchain technology, alone, cannot disrupt shared cataloguing practices; such a shift involves the voluntary and eager participation of cataloguers and members of the public in order to sustain and grow the system. The *Catagora* design concept presented in this thesis incorporates accessibility, collaboration and reputational systems that are intended to foster open participation, but these alone cannot guarantee a thriving, shared-cataloguing alternative to existing systems. Further exploration, in the form of a live implementation, is warranted; and lessons from existing large-scale library technology projects suggest that a centrally-coordinated implementation, targeting key cataloguing partners and driven by a passionate project champion, may provide a more complete picture of the blockchain's potential to support open, shared cataloguing for the benefit of information seekers.

Preface

Data presented in chapter 4 of this thesis consists, in part, of information from the Bibliographic Dataset provided by the Harvard Library under its Bibliographic Dataset Use Terms. The Bibliographic Dataset includes data made available by, among others, OCLC Online Computer Library Center, Inc. and the Library of Congress.

The "Decentralization and Shared Library Catalogues" section in chapter 4 and the "Revision History for Catalogue Entries" section in chapter 5 have been published as part of Kris Joseph, "Wikipedia Knows the Value of what the Library Catalog Forgets," *Cataloging & Classification Quarterly*, volume 57(2-3), pp. 166-183, wherein I was the sole author.

Dedication

For Aaron Swartz, James Dolan, and everyone who furthers the causes of open access, open scholarship, the informational commons, and the public good.

Acknowledgments

I owe a debt of gratitude to the faculty and support staff of both the Digital Humanities department and the School of Library and Information Studies. Grace Jamieson and Nicola DiNicola's tireless work and advocacy for students is invaluable. Maureen Engel possesses the rare ability to stretch a student's brain in just the right way; it is a beautiful gift. Geoffrey Rockwell expresses endless enthusiasm for the work of students and is a font of collegiality. Dinesh Rathi is an ecstatic student advisor whose holistic approach to guidance is a reflection of his endless intellectual rigour. Harvey Quamen is an exquisite and quiet alchemist who forges rich scholarship in the shared caverns of technology and the arts. And Michael McNally's profound impact on me sits betwixt my deep passions about "openness" and my vast dedication to public scholarship. "Profound," "deep," and "vast" are all synonyms for "abysmal."

I would like to thank my husband, Garrick Burron, for the emotional and physical support that has sustained me through three years of hard work, despite my confused wandering through the dark valleys of work/life balance and my occasional storms of irritability and depression. You are my rock.

This work is graciously supported by the Social Sciences and Humanities Research Council.

Contents

1. Introduction 1

2. Background: Ranganathan's Five Laws and Modern Shared Cataloguing 7

One: Books are for use 9

Two and Three: Every reader his book, and every book its reader 15

Four: Save the time of the reader 19

Five: A library is a growing organism 22

A possible future 26

3. Literature Review 29

4. Design Framework for a New Shared Catalogue 37

Decentralization 39

Ethereum foundations 50

Open source software and social information production 68

The framework versus the modern shared catalogue 81

5. The Catagora Software Design 83

Introduction 83

Constraints imposed by Ethereum 84

System Architecture 90

System Requirements 95

Key design characteristics 108

A new engine for the catalogue 123

6. Discussion and Next Steps 125

Returning to Ranganathan 125

Envisioning a Catagora implementation: lessons from other initiatives 129

Next Steps 136

7. Conclusion 142

References 145

List of Figures

Figure 1: Catalogue record creators and modifiers (filtered view) for a subset of Harvard University's open catalogue data 48

Figure 2: Catalogue record creators and modifiers (filtered view) for a subset of University of Michigan's open catalogue data 49

Figure 3: Key Catagora data structures 92

Figure 4: Catagora architecture, with arrows indicating possible component communication paths 98

Figure 5: Workflow steps for a Catagora catalogue search 100

Figure 6: Workflow steps for creating or editing catalogue entry data 102

Figure 7: Workflow steps for editing user profile information 104

Figure 8: Workflow steps for updating a user's contribution score 106

1. Introduction

Library catalogues connect books and readers. This early nineteenth century assertion, made by Charles Ammi Cutter and later reinforced by S. R. Ranganathan,[1] has served as the canonical purpose for the catalogue's existence. Near-universal agreement on the *why* of cataloguing, however, has not constrained the universe of *how.* Scholars have debated the advent and evolution of cataloguing rules,[2] the form and function of online catalogues,[3] and the balance of cost and benefit in resource descriptions.[4] Two maxims seem to have emerged from these debates: first, that the main difference between contemporary catalogues and their nineteenth century equivalents is a "larger

---

[1] Charles A. Cutter, *Rules for a Printed Dictionary Catalog*, 1st ed. (Washington: Government Print Office, 1876), 10; S. R. Ranganathan, *Theory of Library Catalogue* (London: Madras Library Association, 1938); S. R. Ranganathan, *The Five Laws of Library Science* (Madras: The Madras Library Association, 1931), https://babel.hathitrust.org/cgi/pt?id=mdp.39015073883822.

[2] Michael Gorman, "Implementing Changes in Cataloging Rules," *Library Journal* 112, no. 3 (1987): 110–12; Gorman; Andrew D. Osborn, "The Crisis in Cataloging," *The Library Quarterly: Information, Community, Policy* 11, no. 4 (1941): 393–411; Mary K. Bolin, "Make a Quick Decision in (Almost) All Cases: Our Perennial Crisis in Cataloging," *Journal of Academic Librarianship* 16, no. 6 (1991): 357–61.

[3] Marcia J. Bates, "Designing Online Catalog Subject Access To Meet User Needs," in *55th IFLA Council and General Conference* (Paris, France: IFLA Division of Bibliographic Control/Section on Classification and Indexing, 1989), 40–24 to 40–26; Karen Markey, "The Online Library Catalog: Paradise Lost and Paradise Regained?," *D-Lib Magazine* 13, no. 1/2 (January 2007), https://doi.org/10.1045/january2007-markey; H Kalilur Rahman and J Dominic, "An Analytical Study of Online Public Access Catalogues in Comparison with Features of Amazon and Google: A Checklist Approach," *Asian Journal of Information Science & Technology* 2, no. 1 (2012): 17–23; Karen Calhoun et al., "Online Catalogs: What Users and Librarians Want" (Dublin, OH: OCLC, 2009), https://web.archive.org/web/20110101163633/http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf.

[4] Shawne D. Miksa, "You Need My Metadata: Demonstrating the Value of Library Cataloging," *Journal of Library Metadata* 8, no. 1 (April 9, 2008): 23–36, https://doi.org/10.1300/J517v08n01_03; Sheila Ayers, "The Outsourcing of Cataloging: The Effect on Libraries," *Current Studies in Librarianship* 27, no. 1/2 (2003): 17–28; Clare B. Dunkle, "Outsourcing the Catalog Department: A Meditation Inspired by the Business and Library Literature," *The Journal of Academic Librarianship* 22, no. 1 (January 1996): 33–44, https://doi.org/10.1016/S0099-1333(96)90032-4; Frederick G. Kilgour, "The Economic Goal of Library Automation," *College & Research Libraries* 30, no. 4 (July 1, 1969): 307–11, https://doi.org/10.5860/crl_30_04_307.

bibliographical superstructure;"[5] and second, that a *shared* approach to cataloguing is more efficient than an individual one.[6]

If a catalogue is supposed to connect a book and a reader, there ought to be merit in assessing the catalogue's effect on the figurative distance between these two points. Work in this area often examines information search processes,[7] enhancements to metadata that aid in information retrieval, or post-Boolean approaches to search that mimic the success of Google.[8] Though these investigations aim to connect information seekers and their quarry more efficiently, they are based on refinements to existing practices. By contrast, Buckland's *Redesigning Library Services: A Manifesto* prefixes its own reconstruction of the library catalogue with this caution: "To the extent that the card catalogue was a product of the limitations of what is no longer the preferred technology, the development of even the most sophisticated electronic version… could represent misguided creativity, reminiscent

---

[5] Michael Keeble Buckland, *Redesigning Library Services: A Manifesto* (Chicago: American Library Association, 1992), 29.

[6] Rebecca Mugridge, ed., *Cooperative Cataloging: Shared Effort for the Benefit of All* (London, UK: Routledge, 2012); Frederick G. Kilgour, "Computer-Based Systems, A New Dimension to Library Cooperation," *College & Research Libraries* 34, no. 2 (March 1, 1973): 137–43, https://doi.org/10.5860/crl_34_02_137; David Banush, "Cooperative Cataloging at the Intersection of Tradition and Transformation: Possible Futures for the Program for Cooperative Cataloging," *Cataloging & Classification Quarterly* 48, no. 2–3 (February 10, 2010): 247–57, https://doi.org/10.1080/01639370903535742; Christopher Cronin et al., "Strength in Numbers: Building a Consortial Cooperative Cataloging Partnership.," *Library Resources & Technical Services* 61, no. 2 (April 2017): 102–16; Joan E. Schuitema, "The Future of Cooperative Cataloging: Curve, Fork, or Impasse?," *Cataloging & Classification Quarterly* 48, no. 2–3 (February 10, 2010): 258–70, https://doi.org/10.1080/01639370903536088.

[7] Carol Collier Kuhlthau, *Seeking Meaning: A Process Approach to Library and Information Services*, 2nd ed (Westport, Conn: Libraries Unlimited, 2004).

[8] Jan Brophy and David Bawden, "Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources," *Aslib Proceedings* 57, no. 6 (December 2005): 498–512, https://doi.org/10.1108/00012530510634235; Markey, "The Online Library Catalog"; Sharon Q. Yang and Melissa A. Hofmann, "The Next Generation Library Catalog: A Comparative Study of the OPACs of Koha, Evergreen, and Voyager," *Information Technology and Libraries* 29, no. 3 (September 1, 2010): 141–50, https://doi.org/10.6017/ital.v29i3.3139.

of the continued refinement of sailing ships after steam had become the preferred source of power."[9] Instead, Buckland advocates starting from first principles.

This thesis heeds Buckland's advice: through a return to our longest-standing foundations for library catalogues, it theorizes a novel shared catalogue design. Starting with a re-interrogation of Cutter's *Rules for a Printed Dictionary Catalog* and Ranganathan's *Five Laws of Library Science*, it assesses the contemporary shared library catalogue's ability to connect books to readers. From there, it leaps past the historical shift from card-based catalogues to computer-based analogues and the subsequent transformation to responsive web-based catalogues to imagine what might be embodied in a new library catalogue that is built-from-scratch, shared and distributed.

A novel catalogue design suggests the use of a novel technology, and the embryonic incursion of blockchain technology into the Library and Information Studies (LIS) field presents an appealing opportunity. The heart of the blockchain is the concept of a *ledger*—an official record of data and the transactions that create and alter it—whose purpose aligns closely with librarianship's practices of data organization and management. Unlike traditional ledgers, however, blockchains are organized in a distributed manner, with no central control required to verify that information is current and accurate. Technologies of this kind are referred to as *decentralized ledgers*.

Outside the world of librarianship, interest in decentralized technologies and blockchains has stretched into the realm of inflated expectations. A technology that was

---

[9] Buckland, *Redesigning Library Services*, 32.

initially created to disrupt centralized monetary systems[10] is now viewed as a panacea that can absolve humanity of the need for trust and state-based governance.[11] Above the tumult of gushing blockchain discourse and the proliferation of blockchain-based technologies, however, we find claims that blockchain applications rarely survive past the ideation phase[12] and, in fact, may not uniquely solve any problem at all.[13]

Beyond the current hype, though, might decentralized ledgers find a beneficial home in the library? In an environment where information commodification has created a tug-of-war between private interests and the public good, systems that promise equitable, non-intermediated, and distributed access to information should hold appeal for librarians. Despite the application of blockchain technology to a wide range of information-related domains—from identity management[14] and content distribution[15] to data privacy[16] and

---

[10] Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," October 31, 2008, https://bitcoin.org/bitcoin.pdf.

[11] Marcella Atzori, "Blockchain Technology and Decentralized Governance: Is the State Still Necessary?," *Journal of Governance and Regulation* 6, no. 1 (2017), https://doi.org/10.22495/jgr_v6_i1_p5; John Thornton, "Blockchain on the Books: If Trust and Transactions Are a Major Part of an Organisation's Work, It Could Benefit from Secure Shared Records Created Using Blockchain Technology.," *Public Finance*, no. 9 (September 2017): 43–43.

[12] Olga Labazova, Tobias Dehling, and Ali Sunyaev, "From Hype to Reality: A Taxonomy of Blockchain Applications," in *Proceedings of the 52nd Hawaii International Conference on System Sciences* (HICSS 2019, Honolulu, HI, 2019), 4555, https://hdl.handle.net/10125/59893.

[13] Kai Stinchcombe, "Ten Years in, Nobody Has Come up with a Use for Blockchain," Hacker Noon, December 22, 2017, https://hackernoon.com/ten-years-in-nobody-has-come-up-with-a-use-case-for-blockchain-ee98c180100; Kai Stinchcombe, "Blockchain Is Not Only Crappy Technology but a Bad Vision for the Future," *Medium* (blog), April 5, 2018, https://medium.com/@kaistinchcombe/decentralized-and-trustless-crypto-paradise-is-actually-a-medieval-hellhole-c1ca122efdec.

[14] See the Sovrin project for identity management, at https://sovrin.org/

[15] LBRY is a decentralized, YouTube-like video platform at https://lbry.io/

[16] Tim Berners-Lee is at the forefront of Solid, a platform that empowers users to control their own data, providing permission-based access to services that use it. See https://solid.inrupt.com/

data transparency[17]—it has yet to find fertile soil in the practice of librarianship itself.

In the same way that the euphoria and excitement over blockchain technology overreaches its realistic usefulness, the technology is not universally-suited to every problem in the field of librarianship. Where do decentralized, trustless, unalterable ledgers of data add value to our services, and can they be used to more closely link users and information? Driven by three central research questions, this thesis posits that blockchain technology is best used for library services where information production and dissemination are already inherently distributed. On the assumption that one of librarianship's core functions—cataloguing and bibliographic description—fits this model, these research questions are:

1. How might contemporary cataloguing challenges be addressed with a design for a decentralized, collaborative library catalogue based on blockchain technology?
2. Can a truly decentralized, shared cataloguing system be created or are some areas of centralization required for such a service to be sustainable?
3. What policy and management conditions would support the creation and sustainable use of a decentralized, shared library catalogue based on blockchain technology?

[17] Factom emphasizes blockchains as a record management platform. Originally public, the company has pivoted to providing enterprise solutions. See https://www.factom.com/

Blockchain technology is complex and still evolving. For example, there are varying models for transaction verification and consensus,[18] purpose-built blockchains, and general-purpose distributed-computing platforms.[19] This project aligns the affordances of Ethereum's blockchain platform with the first principles of library cataloguing to create a distributed, shared cataloguing model that more closely links information seekers to information resources, and addresses many of the shortcomings of current cataloguing practices. Key features of the conceptual design outlined in this thesis include: catalogue records that are open for access and use; a facility for public participation in shared cataloguing work; social incentives intended to foster participation; and the ability to store and retrieve full revision history and provenance information for catalogue entries.

The "background" chapter provides context for this work by examining some of the current challenges of shared cataloguing as framed by Ranganathan's *Five Laws*. Following this stage-setting, the review of relevant literature illuminates a gap in exploratory research that this project intends to address. An introduction to the Ethereum framework and key findings from successful open source software projects in Chapter 4 serves as a methodological approach to a theoretical design, and the features of the blockchain-based shared cataloguing concept are discussed in Chapter 5. Subsequent chapters present a discussion of key findings, challenges, and next steps for a cataloguing system of this type.

---

[18] Zibin Zheng et al., "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," in *2017 IEEE International Congress on Big Data* (BigData Congress, Honolulu, HI: IEEE, 2017), 557–64, https://doi.org/10.1109/BigDataCongress.2017.85.

[19] Gavin Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," April 7, 2018, https://github.com/ethereum/yellowpaper; NEO Team, "NEO Smart Economy," 2019, https://neo.org/.

2. Background:

Ranganathan's Five Laws and Modern Shared Cataloguing

In 1931 S. R. Ranganathan published his *Five Laws of Library Science*, in which he laid out foundational principles for the practice of librarianship:

1. Books are for use.
2. Every reader his book.
3. Every book its reader.
4. Save the time of the reader.
5. A library is a growing organism.

At that time, library collections consisted primarily of books and periodicals, and the work of libraries was undergoing a sea change: towards accessibility, and towards a focus on the needs of people who seek information. Ranganathan used his five laws as a measuring stick, comparing the praxis of librarianship to what adherence to the laws might enable.

Though the scope of *The Five Laws of Library Science* was vast—addressing everything from shelf labeling to mobile "book vans" for rural areas—the catalogue was at the centre of nearly every chapter. In fact, when Ranganathan published his *Theory of Library Catalogue* seven years later, he insisted that the singular function of the library catalogue was to fulfil the laws of library science.[20] On the surface, at least, this connection is easy to draw: books may be for use, but without a catalogue's ability to render resources discoverable, they are effectively useless.

---

[20] Ranganathan, *Theory of Library Catalogue*, 20–21.

Nearly 90 years after Ranganathan's laws were articulated, both the library and its greater societal context have evolved drastically. Modern networking technologies, the range of available digital and analogue resources, and the ubiquity of information from non-library sources might render today's institution unrecognizable to Ranganathan, yet scholars have regularly reflected on his laws to assess the state of library services and forms.[21]

One recent paper measures modern cataloguing practices against Ranganathan's fifth law, but the examination isolates the growth of cataloguing as a practice and does not interrogate that growth's connections to the larger context of the library.[22] By contrast, the groundwork for this thesis is laid through a re-examination of Ranganathan's laws that acknowledges the integral position of the catalogue in the function of contemporary librarianship. Through this broader lens, challenges to the current cataloguing model are outlined. Ultimately, this work proposes that we address these challenges by reinventing the heart of the catalogue using an approach based on blockchain technology and informed by successful practices in the open source software and social information production communities. Ranganathan's five laws serve as our lodestar.

---

[21] Several examples of these analyses, as well as Michael Gorman's 1995 reframing of Ranganathan's laws for the digital age, can be found in Patrick L. Carr, "Reimagining the Library as a Technology: An Analysis of Ranganathan's Five Laws of Library Science within the Social Construction of Technology Framework," *The Library Quarterly* 84, no. 2 (April 2014): 152–64, https://doi.org/10.1086/675355.

[22] Carlo Bianchini and Mauro Guerrini, "A Turning Point for Catalogs: Ranganathan's Possible Point of View," *Cataloging & Classification Quarterly* 53, no. 3–4 (May 19, 2015): 341–51, https://doi.org/10.1080/01639374.2014.968273.

One: Books are for use

One of the most compelling terms that appears in 1931's *The Five Laws of Library Science* is "Open Access." In contrast to the contemporary definition, however, Ranganathan used the term to refer to emerging approaches for the physical arrangement and function of libraries at the beginning of the 20th century.[23] "Opening" a library meant committing to shelving units that did not require ladders (which present physical risk to readers in search of books); allowing visitors to wander the stacks on their own and locate material themselves; and to making the catalogue useful to anyone gifted with curiosity and a thirst for knowledge. For Ranganathan, Open Access means that users have the "opportunity to see and examine the book collection with as much freedom as in one's own private library."[24] The modern interpretation of Open Access celebrates the same spirit from a digital perspective: for example, research paid for with public funds should be available to the public, and creators of works are encouraged to make them available under liberal licence terms. Though the Open Access movement has touched many areas of the library, it has seldom been applied to the catalogue itself. But why might such an approach have value?

Physical books have historically been used as sites of control. Even though Ranganathan states that his first law, "books are for use," appears self-evident, the practice of librarianship has not always reflected the law's intent.[25] In the late 19th century, when

[23] William Warner Bishop, "A Decade of Library Progress in America" 66 (December 1904): 135–36.

[24] Ranganathan, *The Five Laws of Library Science*, 300.

[25] Ranganathan, 1.

the dominant mechanism for knowledge dissemination was the printed book, libraries historically restricted access to information by limiting physical access to material.[26] Reproduction of books was still expensive, and fears of physical damage or loss led to a desire to preserve rather than share bound volumes. Ranganathan relates the story of Harvard librarian John Langdon Sibley, who was asked why he seemed so pleased after finishing an inventory of the library's collection. "All the books are in excepting two," Sibley responded. "Agassiz has those and I am going after them."[27]

The underlying conditions outlined by Ranganathan have shifted completely since his time at the University of Madras: the concept of the "book" may now be abstracted to any number of physical and digital forms of information reproduction, and the cost of digital information duplication is so low as to be seen as nonexistent. Nonetheless, a modern application of the first law of library science reveals that it is still not self-evident. The tendency to view information as a unique commodity,[28] worthy of protection, has shifted beyond the walls of the library and become embodied in a corporate mindset that would rather control access to information than make it openly available. It can be seen in the application of "digital locks" to information goods and the slow, regressive creep of copyright and intellectual property laws.[29] This attitude towards controlled access to

---

[26] Jeffrey Pomerantz and Robin Peek, "Fifty Shades of Open," *First Monday* 21, no. 5 (April 12, 2016): para. 5, https://doi.org/10.5210/fm.v21i5.6360; Ranganathan, *The Five Laws of Library Science*, xxvii, 2–3.

[27] Theodore Wesley Koch, *On University Libraries* (Evanston, IL: Northwestern University, 1924), 27, https://archive.org/embed/onuniversitylibr00kochuoft; Ranganathan, *The Five Laws of Library Science*, 6.

[28] Samuel E. Trosow, "The Commodification of Information and the Public Good.," *Progressive Librarian*, no. 43 (2014): 17–29.

[29] Cory Doctorow, Amanda Palmer, and Neil Gaiman, *Information Doesn't Want to Be Free: Laws for the Internet Age* (San Francisco, CA: McSweeney's, 2014); Michael Geist, "Rethinking IP in the TPP: Canadian Government Plays Key Role in Suspending Unbalanced Patent and Copyright Rules," *Michael Geist* (blog),

information arises from the emergence of knowledge- and service-based economies, and the commensurate desire to treat information like a non-renewable resource: finite and lucrative, requiring central ownership and strict control.[30]

The Sibley example illustrates a preference for preservation and control of information over accessibility, but it serves to make another point: Sibley could not have conducted his inventory without a catalogue on which to base his work. The accuracy of the catalogue certainly affects the ability to identify resources, but its very existence highlights the catalogue as another site of control. Even in Ranganathan's day, when access to books was mediated by staff working behind a counter, the catalogue was always accessible;[31] it is also reasonable to assume that Sibley, whose goal was to keep Harvard's books safely locked away, would not have restricted access to their index. It is surprising, then, that the historical approach stands in contrast to the modern condition, where openly-accessible copies of catalogue data are only provided by a handful of progressive institutions,[32] and some owners of bibliographic services have attempted to restrict access to the catalogue itself.

November 11, 2017, http://www.michaelgeist.ca/2017/11/rethinking-ip-in-the-tpp/; Mariano Zukerfeld, "The Tale of the Snake and the Elephant: Intellectual Property Expansion under Informational Capitalism," *The Information Society* 33, no. 5 (October 20, 2017): 243–60, https://doi.org/10.1080/01972243.2017.1354107.

[30] Douglas Rushkoff, *Program or Be Programmed: Ten Commands for a Digital Age*, ed. Leland Purvis (Berkeley, CA: Counterpoint, 2011); Trosow, "The Commodification of Information and the Public Good."

[31] Ranganathan, *The Five Laws of Library Science*, 337–39.

[32] "Open" catalogues provide downloadable copies of catalogue record sets, but do not permit direct access to catalogue entries for update, alteration, or revision. A collection of open catalogue data, including records used in later sections of this thesis, can be found at the internet Archive. See Internet Archive, "Open Library Data," Open Library, May 3, 2018, https://archive.org/details/ol_data.

The most notable service provider in this area was created in 1971, when fifty-four libraries partnered to create the Ohio College Library Center's (OCLC) shared library catalogue,[33] now known as WorldCat. The catalogue is a primary source for information about monographs and serial publications and is currently used by more than 16,000 library systems in over 120 countries.[34] OCLC was formed as a not-for-profit corporation and still holds that designation, but in contrast to its original intent to ease "access to and use of…knowledge and information,"[35] it has created digital fences around its service that do the opposite.

Two examples involving OCLC are salient here. The 1990s saw the widespread implementation of the TCP/IP protocol for inter-computer networking—one of the fundamental technologies underpinning the modern-day Internet. In 1999 OCLC created a new vision for itself as a "leading global library cooperative" and a "globally networked information resource of text, graphics, sound, and motion."[36] The communications technology that had been used to interconnect library databases (called Z39.50) was rapidly adapted for use over the Internet. Within three years, OCLC non-members were using the freely-accessible Z39.50 protocol to harvest catalogue information from member libraries. In 2003, OCLC Vice-President Gary Houk spoke out against this practice by

---

[33] Charles P. Bourne and Trudi Bellardo Hahn, *A History of Online Information Services, 1963-1976* (Cambridge, MA: MIT Press, 2003), 344. Note that OCLC changed the definition of its acronym to "Online Computer Library Center" in 1981.

[34] OCLC, "OCLC Technology," OCLC, February 8, 2019, https://www.oclc.org/en/technology.html.

[35] Allen Kent, Harold Lancour, and William Z. Nasri, eds., *Encyclopedia of Library and Information Science*, vol. 45 (New York, NY: M. Dekker, 1968), 289.

[36] Jay Jordan, "OCLC 1998-2008: Weaving Libraries into the Web.," *Journal of Library Administration* 49, no. 7 (October 2009): 727–62.

coining the term "record nabbing." Arguing that record nabbing was theft that reduced "the economic viability of the cooperative cataloguing model,"[37] Houk asserted that access to records would be tightened to encourage participation in the cooperative. Record nabbing became a longstanding problem for OCLC and a source of great debate. In a scathing essay by Jeffrey Beall in 2008, the author pointed out that libraries were being encouraged to turn off Z39.50 access to their local catalogues at the same time they were being encouraged to use OCLC's fee-based interlibrary loan service: "In summary, when OCLC makes money on the deal, it's called 'resource sharing'; when OCLC doesn't make money on the deal, it's called 'record nabbing'."[38]

Perhaps a more egregious example of the catalogue as a site of control occurred even earlier, in the early 1980s, when OCLC announced a plan to copyright WorldCat.[39] Despite an internal committee's recommendation against the idea, the move was intended to reinforce the expectation that all of its member libraries should contribute to WorldCat equally, that freedom of access did not also mean freedom from cost, and that other commercial and non-profit organizations should be prevented from making copies of records from the shared catalogue without contributing to the cost of the catalogue's creation and maintenance.[40] After significant debate, the dispute was resolved with the

---

[37] Gary R. Houk, "OCLC Speaks out on Record Nabbing," *Library Collections, Acquisitions, & Technical Services* 27, no. 3 (2003): 278.

[38] Jeffrey Beall, "OCLC: A Review," in *Radical Cataloging: Essays at the Front*, ed. K. R. Roberto (Jefferson, N.C: McFarland & Co, 2008), 90.

[39] David F. Bishop, "OCLC Copyright: A Threat to Sharing," *Journal of Academic Librarianship* 11, no. 4 (September 1985): 202; Rowland C. W. Brown, "OCLC, Copyright, and Access to Information: Some Thoughts," *Journal of Academic Librarianship* 11, no. 4 (September 1985): 197.

[40] Brown, "OCLC, Copyright, and Access to Information: Some Thoughts," 197.

1987 creation of the *Policy for Use and Transfer of WorldCat Records*.[41] This policy specified that member libraries could use or share records of their own holdings, but only if they also shared them with OCLC. It also ordered that member libraries must not share records with any non-member or commercial entity without a written agreement.

Intellectual property regimes have driven the commodification of information and have created a modern-day perversion of Ranganathan's first law, transmuting "books are for use" into "books are for *permitted* use." In a broad response to these forces, information professionals have created Open Access models for information sharing and publishing,[42] the Creative Commons model for copyright management,[43] and the open source software movement.[44] At the same time, corporate interests have infiltrated every area of information management—including libraries, as the OCLC example illustrates—to create complex publishing models,[45] feudalist information ecosystems,[46] and expanded copyright

---

[41] OCLC, "Policy for Use and Transfer of WorldCat Records [Archived]," November 19, 2008, https://web.archive.org/web/20081203043912/http://www.oclc.org:80/worldcat/catalog/policy/recordusepolicy.pdf.

[42] Open Access Max-Planck-Gesellschaft, "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities," October 22, 2003, https://openaccess.mpg.de/Berlin-Declaration.

[43] Creative Commons, "Creative Commons: When We Share, Everyone Wins," Creative Commons, 2018, https://creativecommons.org/.

[44] Yochai Benkler, "Coase's Penguin, or, Linux and 'The Nature of the Firm,'" *The Yale Law Journal* 112, no. 3 (December 2002): 369, https://doi.org/10.2307/1562247; Eric S. Raymond, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, Rev. ed (Cambridge, MA: O'Reilly, 2001).

[45] Julian H. Fisher, "Scholarly Publishing Re-Invented: Real Costs and Real Freedoms," *The Journal of Electronic Publishing* 11, no. 2 (May 30, 2008), https://doi.org/10.3998/3336451.0011.204; Amy Forrester, "Barriers to Open Access Publishing: Views from the Library Literature," *Publications* 3, no. 4 (September 3, 2015): 190–210, https://doi.org/10.3390/publications3030190.

[46] Rijurekha Sen et al., "Inside the Walled Garden: Deconstructing Facebook's Free Basics Program," *ACM SIGCOMM Computer Communication Review* 47, no. 5 (October 25, 2017): 12–24, https://doi.org/10.1145/3155055.3155058.

and intellectual property rights.[47]

While it is true that library catalogues have embraced aspects of openness—especially connected to the Open Access movement, where tools like the Directory of Open Access Resources (DOAR)[48] and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[49] may be found—the catalogue itself has not been fully-exploited as an open resource. A catalogue that is "open" by design—transparent, authorized for liberal use, and accessible for public participation[50]—might not only make information resources more accessible but would serve to improve the links between information seekers and information resources. This thesis presents an architecture for such a catalogue.

Two and Three: Every reader his book, and every book its reader

Ranganathan writes at length about how to distinguish his second and third laws but says the two are complementary; accordingly, they have been treated together in this context. The second law is related to the power that books have to provide education and empowerment to people, transcending class, gender, and geography.[51] The third law asserts that there is a reader for every book, and that the role of the librarian (and, by

---

[47] Geist, "Rethinking IP in the TPP."

[48] Jisc, "OpenDoar: Directory of Open Access Repositories," accessed December 14, 2018, http://v2.sherpa.ac.uk/opendoar/.

[49] Open Archives Initiative, "Open Archives Initiative Protocol for Metadata Harvesting," accessed December 14, 2018, https://www.openarchives.org/pmh/.

[50] Pomerantz and Peek, "Fifty Shades of Open."

[51] Ranganathan, *The Five Laws of Library Science*, 76.

extension, the catalogue) is to ensure that every book and its reader can be connected.[52] Accessibility is central to both laws, but cataloguing takes the spotlight in the third law with its emphasis on robust, thorough, and reader-friendly resource descriptions.[53]

It is in the application of the second and third laws that we begin to see tension. Take, for example, the first conference of the American Library Association (ALA), where Melvil Dewey listed himself as member number one and declared that the organization's motto should be "the best reading for the greatest number at the least cost."[54] This utilitarian view conflicts with the goal of linking readers and books because it suggests that the time and effort invested by cataloguers in resource descriptions must be economized. Ranganathan understood this tension, pointing out that all five of his laws stood in opposition to the Law of Parsimony, which is also known as Occam's Razor. Though he believed strongly in working to address the needs of readers, he acknowledged that a primary mediator of librarianship would be "economy—in materials, space, time and energy."[55]

The evolution of cataloguing practices bears out the truth of Ranganathan's observation. In the early days of catalogues, libraries worked painstakingly to describe each of the items on their shelves, creating unique directories and finding aids for items in their collections. The shift from book-style indices to the use of catalogue cards was

---

[52] Ranganathan, 299–300.

[53] Ranganathan, 307–12.

[54] Ranganathan, 156.

[55] Ranganathan, *Theory of Library Catalogue*, 55.

monumental: searches were expedited with mechanical card-sorting systems and changes to collections were easy to manage with the addition or removal of cards. Perhaps most importantly, however, widespread use of a standardized 3-inch by 5-inch index card enabled information sharing between libraries: by copying cards, or by creating duplicate sets of cards, libraries could collaborate and reduce their individual cataloguing workload.

An entire industry of pre-printed catalogue cards rapidly appeared (and endured until the last set of cards was printed in 2015),[56] maximizing cataloguing efficiency for libraries. Ready-made bibliographic entries were distributed *en masse*, eliminating the need for local cataloguing in many cases.[57] This practice transitioned into the digital age when Fred Kilgour and Ralph H. Parker wrote a 1965 proposal to create a computerized network for Ohio colleges: a collection of academic libraries, all operating under the guidance of a single, shared catalogue system. In their proposal, Kilgour and Parker outlined several benefits including reduced time for cataloguing materials, the elimination of physical card catalogues, and increased service consistency among libraries due to commonly-shared catalogue information.[58] This proposal led to the creation of OCLC in 1971, and the organization remains the dominant provider of shared cataloguing services worldwide.[59]

---

[56] OCLC, "OCLC Prints Last Library Catalog Cards," OCLC.org, October 1, 2015, https://www.oclc.org/en/news/releases/2015/201529dublin.html.

[57] Library of Congress, ed., *The Card Catalog: Books, Cards, and Literary Treasures* (San Francisco, CA: Chronicle Books, 2017).

[58] Frederick G. Kilgour, "Report to the Committee of Librarians of the Ohio College Association," in *Collected Papers of Frederick G. Kilgour, OCLC Years*, ed. Lois L. Yoakam (Dublin, OH: OCLC Online Computer Library Center, 1984), 1.

[59] OCLC, "OCLC Technology."

Standardization of catalogue entries benefits users in one significant way: readers who search a catalogue at one library can be reasonably-assured that a similar search at a different library will provide similar results. But the drive to centralize and economize cataloguing also presents challenges for library users. Centrally-created catalogue records cannot account for regional or local contexts, where knowledge and information needs may be more specialized.[60] In fact, early experience with OCLC's WorldCat showed that there was high variation in the entry of catalogue records, and this was addressed through stronger adherence to cataloguing rulesets like the Anglo-American Cataloguing Rules (AACR2, for the second edition). These guidelines allowed for the use of "minimal" and "maximum" catalogue records, accounting for situations where local expertise did not allow for full resource descriptions.[61] Recent reviews of cataloguing quality issues have shown that cataloguers' practices are more often based on convenience than on users' needs,[62] that shifting catalogue rules create compliance problems, and that local libraries often "tweak" records to force a consistent appearance for users.[63]

The Law of Parsimony was observed in the creation of OCLC and the evolution of the shared catalogue model, arguably leaving Ranganathan's second and third laws unfulfilled.

---

[60] M. Cristina Pattuelli, "Modeling a Domain Ontology for Cultural Heritage Resources: A User-Centered Approach," *Journal of the American Society for Information Science and Technology* 62, no. 2 (February 2011): 314–42, https://doi.org/10.1002/asi.21453.

[61] Glenn Patton, "OCLC's Long Association with Less-Than-Full Cataloging," *Technical Services Quarterly* 9, no. 2 (February 12, 1992): 21–30, https://doi.org/10.1300/J124v09n02_04.

[62] Barbara Schultz-Jones et al., "Historical and Current Implications of Cataloguing Quality for Next-Generation Catalogues," *Library Trends* 61, no. 1 (2012): 78, https://doi.org/10.1353/lib.2012.0028.

[63] Donna Ellen Frederick, "Library Data: What Is It and What Changes Do Libraries Need to Make? (The Data Deluge Column)," *Library Hi Tech News* 34, no. 8 (October 2, 2017): paras. 7–10, https://doi.org/10.1108/LHTN-06-2017-0044.

An alternate approach might bring Occam's Razor into harmony with Ranganathan's third law: a cost-effective shared catalogue that also encourages robust descriptions of information resources.

Four: Save the time of the reader

In 1876 (the same year that the American Library Association was formed, with Melvil Dewey as its first member),[64] Charles Ammi Cutter developed a set of objectives for library catalogues that have governed the institution ever since. They opened with the words "to enable a person to find a book…",[65] emphasizing the value of the catalogue as a user-centred finding aid and placing special emphasis on the value of subject-based classification.

Half a century later, Ranganathan explores his fourth law— "save the time of the reader"—by examining every step of a user's interaction with the library, highlighting the value of information accessibility through the removal of physical barriers and human intermediaries. When his gaze falls on the catalogue, Ranganathan notes (in alignment with Cutter) that most readers prefer to look up information based on subject. To save a reader's time, then, cataloguers must understand the "aboutness" of a book and describe it with as many terms as possible. He notes that most information resources are "composite," requiring primary, secondary, and perhaps even deeper subject classification;[66] in fact,

---

[64] Ranganathan, *The Five Laws of Library Science*, 155.

[65] Cutter, *Rules for a Printed Dictionary Catalog*, 10.

[66] Ranganathan, *The Five Laws of Library Science*, 351–59.

Ranganathan developed a five-faceted classification system that is still widely-used in India today.[67]

Even with the use of deep subject classification, the act of describing resources is fraught with challenges. Some outside the field of LIS may think of resource descriptions as objective, but they rarely are. Ideally, library users would be able to discover resources rapidly by using personally-familiar search terms, but the likelihood that two people will use the same subject terms to describe a book (without knowledge of controlled vocabulary) is somewhere between 10 and 20 percent.[68] Even among experienced cataloguers, complete consistency in resource descriptions is rare.[69] How can these disparate views coalesce, and how can we "save the time of the reader" if we cannot accurately predict the language that readers use when posing their questions? Significant labour would have to be invested in the thorough description of every resource, placing Ranganathan's fourth law once more into conflict with the Law of Parsimony.

The struggle between economy and resource description—the same tension noted in my reflection on Ranganathan's second and third laws—is apparent in at least two areas related to the fourth law. The first is that cataloguers act as intermediaries between

---

[67] S. R. Ranganathan, *Colon Classification*, 6th ed (New Delhi, India: Ess Ess Publications, 2007).

[68] Dorothy Gregor and Carol Mandel, "Cataloging Must Change!," *Library Journal* 116, no. 6 (4/1/1991 1991): 46; Thomas Mann, "'Cataloging Must Change!' And Indexer Consistency Studies: Misreading the Evidence at Our Peril," *Cataloging & Classification Quarterly* 23, no. 3–4 (March 28, 1997): 3–45, https://doi.org/10.1300/J104v23n03_02; Bates, "Designing Online Catalog Subject Access To Meet User Needs."

[69] The issue of inter-indexer consistency is well-studied in the LIS field. Work from prominent scholars includes Lois Mai Chan, "Inter-Indexer Consistency in Subject Cataloging," *Information Technology & Libraries* 8, no. 4 (December 1989): 349–58; Lawrence E Leonard, "Inter-Indexer Consistency Studies, 1954-1975: A Review of the Literature and Summary of Study Results," *Occasional Papers (University of Illinois at Urbana-Champagne. Graduate School of Library Science)*, no. 131 (December 1977): 1–54.

readers and books—a relationship that is further abstracted because third-party cataloguing service providers are an intermediary between cataloguers and the catalogue. In other words, the drive to make cataloguing more efficient has created an external inefficiency for library users. For example, a library patron who identifies an issue with a catalogue record might speak to a front-line staff member, who then passes the information to a cataloguer, who then works with OCLC to resolve the issue. There is a less direct link between users and the catalogue than there was in Ranganathan's time. Moreover, the ability to access and edit raw catalogue data is restricted: OCLC provides online access to catalogue data through a web-based Applications Programming Interface (API), but only for paid subscribers;[70] and only subscribers with sufficient authorization may make updates to those entries. The robustness of resource descriptions would benefit from the direct input of those that use them—members of the public—but that affordance has been sacrificed for the sake of lowered costs.

Involving the public in discourse about resource descriptions would be invaluable but advocating for their participation leads indirectly to the second conflict between robust resource descriptions and economy: the more people that are involved in the discussion about how a resource is described, the more difficult it is to coordinate and manage the discussion. This is already an issue for the work of cataloguers, because contributions to shared catalogues are made in a distributed manner.

---

[70] OCLC, "WorldCat Search API (Web Service)," WorldCat, 2019, https://www.worldcat.org/affiliate/tools?atype=wcapi. It should be noted that WorldCat.org provides free search access to the global catalogue for anyone, but the interface does not permit users to report errors, make improvement suggestions, or easily download records in machine-readable formats..

F.K. Donnelly notes this challenge in her work on controversies over content and authorship in catalogue entries. She observes that there are wide and debatable descriptions being used for resources, but that no framework exists that might allow librarians to debate and resolve these challenges.[71] In an approach to shared cataloguing that is designed to include public participation, the path between users and resource descriptions can be shortened, and can include an efficient framework for librarians to discuss and resolve resource description challenges. Such a framework could be employed to improve the quality of resource descriptions stored in the catalogue, which may, in turn, save time for users engaged in search activity.

Five: A library is a growing organism

Ranganathan's fifth law deals with the growth of libraries on two fronts: scale and evolution. Insisting that organisms that do not grow will "petrify and perish,"[72] he notes that the number of readers and books are increasing, which necessarily leads to a demand for increased library resources. Ranganathan could not have predicted the technological revolution and its impact on society, or the ongoing pressures of economization (the Law of Parsimony, yet again) brought on by globalization and capitalism. He surmised that the constant growth of the library organism would lead to drastically-increased demands on

---

[71] F. K. Donnelly, "Catalogue Wars and Classification Controversies.," *Canadian Library Journal* 43 (August 1986): 247.

[72] Ranganathan, *The Five Laws of Library Science*, 382.

services and staff,[73] but he did not anticipate the form that the response to these demands would take.

Budgetary and operational pressures have often forced libraries to outsource or, in some cases, privatize functions that used to be performed in-house.[74] Over the past five decades, organizations like OCLC have become more dominant in the provision of services. This has gradually led to a situation where a small number of external organizations have become nearly-monolithic, horizontally-integrated providers of library services: centralized entities, outside the direct control of the libraries they serve, conducting much of the critical work of libraries.

Centralization is not inherently dangerous, and debates about the relationship between computing centralization and administrative centralization are as old as computing itself.[75] When centralized, critical processes can be controlled from end to end, optimizing them for speed, cost, integrity, and quality. In an economic system that values efficiency, these qualities are admirable.[76] On the other hand, centralization can lead to

---

[73] Ranganathan, 385–412.

[74] Ayers, "The Outsourcing of Cataloging: The Effect on Libraries."

[75] Brian P. Bloomfield and Rod Coombs, "Information Technology, Control and Power: The Centralization and Decentralization Debate Revisited," *Journal of Management Studies* 29, no. 4 (July 1992): 459–459, https://doi.org/10.1111/j.1467-6486.1992.tb00674.x; Joey F. George and John L. King, "Examining the Computing and Centralization Debate," *Communications of the ACM* 34, no. 7 (July 1, 1991): 62–72, https://doi.org/10.1145/105783.105796; Thomas Marschak, "On the Comparison of Centralized and Decentralized Economies," *The American Economic Review* 59, no. 2 (1969): 525–32.

[76] Robert D. Atkinson and Michael Lind, "Is Big Business Really That Bad?," *The Atlantic*, April 2018, https://www.theatlantic.com/magazine/archive/2018/04/learning-to-love-big-business/554096/; R. H. Coase, *The Firm, the Market, and the Law* (Chicago, IL: University of Chicago Press, 1988).

market consolidation that stifles innovation, increases the damage caused by the failure of systems, and creates barriers to new competition.[77]

"Centralization" and "monopoly" are words seldom used in connection with libraries, but both terms first appeared in connection with OCLC in 1979. To deter libraries from choosing alternative collaborations over OCLC's services, Fred Kilgour (then the President of the organization) brought academic library directors to Ohio at OCLC's expense. He outlined his plans and asserted the "long-standing policy that only libraries which input all their cataloguing would be permitted to use the OCLC system and access its data base."[78] Through a long succession of mergers and acquisitions,[79] OCLC has continued to attract criticism for its protectionist approach. In 2010, a commercial cataloguing company called SkyRiver and its parent organization, Innovative Interfaces, Inc., launched an anti-competitive lawsuit against OCLC, claiming that the organization held a monopoly over bibliographic services through its membership requirement, and that they were not a true non-profit because its members did not have direct control over OCLC's management or policies. Sadly, the suit never got its day in court: it was dropped in 2013, after a number of staffing and structural changes at Innovative Interfaces, with its new CEO Kim Masanna noting that "we decided to view a relationship with OCLC as a potential collaboration partner."[80]

---

[77] Barry C. Lynn, *Cornered: The New Monopoly Capitalism and the Economics of Destruction* (Hoboken, NJ: John Wiley & Sons, 2010).

[78] Richard De Gennaro, "From Monopoly to Competition: The Changing Library Network Scene," *Library Journal* 104 (January 1, 1979): 1216.

[79] OCLC, "Mergers and Acquisitions," 2017, https://www.oclc.org/en/about/finance/mergers.html.

[80] Gary Price, "III Drops OCLC Suit, Will Absorb SkyRiver," LJ infoDOCKET, March 4, 2013, http://www.infodocket.com/2013/03/04/innovative-interfaces-integrates-all-skyriver-services-and-

As the scale of librarianship has grown in accordance with Ranganathan's fifth law, then, we find the forces of consolidation at work to contradict the first law. When the ability to discover resources is compromised, a user's ability to use resources is diminished.

When discussing the growth of the library in terms of its evolution, Ranganathan turns his attention the catalogue once more in his observations about the medium and format of catalogues and their entries. He notes the shift from "paste-down" book-style catalogues, to loose-leaf binders, to card catalogue systems.[81] To this evolutionary trail we must add the computing revolution, which saw catalogue records move from cards to punch-paper tape[82] to magnetic tape and then to purely digital form.

Though the storage and transmission media for catalogue entries have both shifted, their evolution exhibits elements of path dependence: new innovations to the catalogue have been built iteratively on prior technology, constraining each new form of the catalogue to one that resembles those of its ancestors. Metadata standards have evolved but each catalogue record is still a list of field-value pairs;[83] storage and communications technologies have evolved, making it easier to grow and edit catalogue entries, but the

---

withdraws-antitrust-lawsuit-against-oclc/; Association of Research Libraries, "Summary of Antitrust Lawsuit: SkyRiver & Innovative Interfaces v. OCLC," January 1, 2010, http://www.arl.org/news/arl-news/2356-summary-of-antitrust-lawsuit-skyriver-innovative-interfaces-v-oclc.

[81] Ranganathan, *The Five Laws of Library Science*, 395–97.

[82] Michele Seikel and Thomas Steele, "How MARC Has Changed: The History of the Format and Its Forthcoming Relationship to RDA," *Technical Services Quarterly* 28, no. 3 (May 19, 2011): 324, https://doi.org/10.1080/07317131.2011.574519.

[83] The birth of the BIBFRAME standard is a novel exception here, but it is also illustrative of path dependence. It shifts the structure of catalogue records from field-value pairs ("doubles") to subject-predicate-object triples. See Library of Congress, "BIBFRAME - Bibliographic Framework Initiative," accessed April 17, 2019, https://www.loc.gov/bibframe/.

nature of copy cataloguing is essentially unchanged from its earliest roots. Modern approaches to cataloguing can be viewed as iterations of the work begun by Cutter and Ranganathan, suggesting that positive-feedback mechanisms have "locked in" early implementation decisions. This makes alternative approaches difficult to envision or implement.[84] The digital form of the catalogue record—based on early MARC standards—is essentially unchanged since it was created in 1965, and the paradigm used to store, edit and retrieve catalogue entries has been fundamentally unquestioned since WorldCat made its computerized debut in 1971.

Over time, the increasing scale and gradual evolution of shared cataloguing has created a deeply-entrenched system of centralized catalogue control and limited technological flexibility. A fresh look at the design of a shared catalogue system, built with new technologies and shaped by successful practices in the open source software world, might disrupt this pattern and illuminate a new path for evolution and growth.

A possible future

Returning to some of the earliest guiding principles for library science presents a tantalizing opportunity to revisit the function and goals of the library catalogue, looking to new and disruptive technologies that may give us a fresh perspective on how those goals can be achieved. One technology that promises to reverse trends towards centralization and information commodification is the decentralized ledger, or blockchain. If carefully

---

[84] W. Brian Arthur, "Positive Feedbacks in the Economy," *Scientific American* 262, no. 2 (February 1990): 92–99, https://doi.org/10.1038/scientificamerican0290-92; Buckland, *Redesigning Library Services*, 32.

designed, exercising the lessons that have been gleaned from open source software and social information production systems like Wikipedia and Linux, such a technology may enable a closer alignment of the shared catalogue to the values imbued by Ranganathan's five laws.

Though Bitcoin was created as a form of digital currency, using a bank-like *ledger* of blocks (linked in a chain and leading to the term *blockchain*) as its data storage mechanism, the underlying technology can be used to create decentralized systems capable of any form of computation. Since blockchains can independently-verify data or execute code across a distributed array of storage and computing nodes, they can effectively play the traditional economic role of the *firm*. A firm, following Ronald Coase's definition, is created when an organization internalizes the work of transforming an input into an output in order to reduce the cost of using the market (and its associated transaction costs) for the same purpose.[85] Unlike a traditional firm, however, whose mechanisms and operation are managed by a central entity, blockchains are able to process transactions and manage their costs using software-based consensus mechanisms that are shared by all of the participants in the system. This absolves the need for a central coordinating mechanism to manage transactions. In fact, the blockchain is also the *market*, since the work of price-setting, and of making matches between buyers and sellers, is performed by the blockchain itself.

One of the most active and prominent examples of general-purpose computing platforms using blockchain technology is Ethereum,[86] which separates distributed

---

[85] Coase, *The Firm, the Market, and the Law*; Benkler, "Coase's Penguin, or, Linux and 'The Nature of the Firm.'"

[86] Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger."

networking, data storage, and consensus mechanisms into abstract, independently-evolvable layers with defined roles. Programmers working with the platform use a fully-featured language called Solidity to create *contracts* that the network can execute without the requirement of human or corporate mediation.[87] The characteristics of distributed computing platforms like Ethereum enable features that Fred Kilgour and Ralph Parker could not have envisioned in their 1965 shared cataloguing proposal. Those features, as well as their alignment with the work of cataloguers, are explored as part of this project.

This thesis is an attempt to return to the first principles of cataloguing. With a view towards the values and principles of librarianship, it presents a conceptual design for a library catalogue engine that leapfrogs the existing paradigm and leverages the affordances of open source software design, collaborative information production, non-market incentive systems, and decentralized ledger technology.

[87] Chris Dannen, *Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners.* ([Place of publication not identified]: Apress, 2017); Ethereum, “Solidity — Solidity 0.4.21 Documentation,” 2017, https://solidity.readthedocs.io/en/v0.4.21/#.

3. Literature Review

The previous chapter examines contemporary cataloguing practices through the lens of Ranganathan's five laws to enumerate challenges and provide a broad context for this work. This section provides a more focused overview of the intersection between library catalogues and blockchain technology to illustrate a gap in current research that will be explored in the chapters that follow.

Two key subject areas provide the basis for framing this space: well-developed literature related to the design and efficacy of modern library catalogues, and less substantial literature related to nascent uses of blockchain technology in the field of Library and Information Studies. Within the first domain, there are other visions for the reinvention of the catalogue that can be broadly divided into three categories: models with benefits for users, models with benefits for librarianship, and models in favor of the abolition of traditional library catalogues. The second domain, reflecting the novelty of the blockchain paradigm itself, is a shallower exploration of how decentralized ledgers might impact the practice of librarianship in general. These domains are addressed in turn, beginning with literature related to contemporary cataloguing practices.

As was noted in Chapter 2, the dominant mechanisms for library catalogue creation and maintenance have largely been unaltered since the debut of the first electronic catalogue in 1971.[88] Catalogue data is stored in relational databases, with current implementations reliant on distributed, cloud-based file systems spread between multiple

---

[88] See pages 25-26.

data centres.[89] The largest of these—the WorldCat union catalogue—is centrally-maintained and administered by OCLC;[90] the corporation is based in Dublin, Ohio, but has data centres in Australia, Canada, the United States and the Netherlands.[91] Locally-stored copies of catalogues, such as those used by individual libraries and library systems, are also kept in relational databases, typically as part of larger Integrated Library Systems (ILS) platforms designed to manage collections and library operations.[92] Large libraries maintain cataloguing departments but it is not uncommon for smaller libraries to rely on "shelf ready" materials,[93] or to outsource their cataloguing work to OCLC or commercial organizations.[94] Institutional subscribers to OCLC services have the ability to update or contribute catalogue records to the global collection.[95]

The dominance and functionality of online services like Google Search, Google Scholar and Amazon have led some to claim that library catalogues are no longer

---

[89] Apache Software Foundation, "Apache Hadoop," 2018, https://hadoop.apache.org/; OCLC, "OCLC Technology."

[90] OCLC, "What Is WorldCat?," WorldCat.org, 2019, https://www.worldcat.org/whatis/default.jsp.

[91] OCLC, "OCLC Technology."

[92] Cynthia Lopata, "Integrated Library Systems. ERIC Digest" (ERIC Clearinghouse on Information Resources, April 1995), https://eric.ed.gov/?id=ED381179.

[93] The term "shelf ready" refers to the acquisition of library materials that arrive complete with premade cataloguing information. See ProQuest, "Shelf Ready Services," ProQuest, accessed May 24, 2019, https://www.proquest.com/products-services/Shelf-Ready-Services.html.

[94] Vicki Toy Smith, "Outsourcing Cataloging: An Evaluation," *Journal of Educational Media & Library Sciences* 34, no. 4 (1997): 380–95; Marie Kascus, Dawn Hale, and Association for Library Collections & Technical Services, eds., *Outsourcing Cataloging, Authority Work, and Physical Processing: A Checklist of Considerations* (Chicago: American Library Association, 1995); Claire Doran and Cheryl Martin, "Measuring Success in Outsourced Cataloging: A Data-Driven Investigation," *Cataloging & Classification Quarterly* 55, no. 5 (July 4, 2017): 307–17, https://doi.org/10.1080/01639374.2017.1317309; Ayers, "The Outsourcing of Cataloging: The Effect on Libraries."

[95] OCLC, "What Is WorldCat?"

necessary. Sites like Google are now the dominant starting point for searches,[96] and the depth and breadth of their results are unmatched by contemporary library catalogues.[97] Tennant raises this point as early as 2003, calling for Online Public Access Catalogues (OPACs) to compete with other online services by querying *across* databases to provide better information for users.[98] In 2014 Kortekaas and Kramer posited that libraries should prioritize content delivery over discovery and asserted that the quality of Google's results is unsurpassable by OPACs.[99]

Other work has sought to address catalogue quality by improving its user-facing features, often as part of a larger ILS. Early findings in this area state that users equate search quality with the interface's similarity to the function and output of popular websites.[100] Features and usability of "next-generation" catalogues are well-studied, with immediate access to full content, state-of-the-art web interfaces (including search term prediction and flexible handling of spelling and grammar), and user-contributed content regularly appearing as in-demand features.[101] A popularly-deployed system that

---

[96] Rahman and Dominic, "An Analytical Study of Online Public Access Catalogues in Comparison with Features of Amazon and Google: A Checklist Approach," 8; Ian Hargraves, "Controversies of Information Discovery.," *Knowledge, Technology & Policy* 20, no. 2 (2007): 84,88; Markey, "The Online Library Catalog," paras. 12–18.

[97] Brophy and Bawden, "Is Google Enough?"

[98] Roy Tennant, "Library Catalogs: The Wrong Solution," *Library Journal* 128, no. 3 (February 15, 2003): 28.

[99] Simone Kortekaas and Bianca Kramer, "Thinking the Unthinkable - Doing Away with the Library Catalogue.," *Insights: The UKSG Journal* 27, no. 3 (November 2014): 244–48.

[100] Calhoun et al., "Online Catalogs: What Users and Librarians Want," 14.

[101] Yang and Hofmann, "The Next Generation Library Catalog"; DeeAnn Allison, "Information Portals: The Next Generation Catalog," *Journal of Web Librarianship* 4, no. 4 (November 30, 2010): 375–89, https://doi.org/10.1080/19322909.2010.507972; Brophy and Bawden, "Is Google Enough?"

emphasizes these functions, integrating them with social-media-like connectivity between users, is BiblioCommons, which is currently in use by more than 125 public library systems in North America, Australia and New Zealand.[102]

Still other attempts have been made to redesign all or some part of the catalogue's back end, but in comparison to work on catalogue interfaces, scholarship in this area is limited. Projects of this kind tend to investigate solutions to individual catalogue challenges. Examples include: adding support for Resource Description Framework (RDF) queries on existing MARC records;[103] implementing new catalogue records record formats such as Resource Description and Access (RDA) or BIBFRAME,[104] which is based on the hierarchical Functional Requirements for Bibliographic Records (FRBR) model;[105] the creation of a catalogue based on automated harvesting of open journal data; and the formation of partnerships to create new, specialized catalogues for specific knowledge domains.[106]

---

[102] Norman Oder, "BiblioCommons:. A Breakthrough?," *Library Journal* 133, no. 13 (August 15, 2008): 14–14; BiblioCommons, "A Better Online Catalog Is Just the Beginning.," BiblioCommons, 2017, https://www.bibliocommons.com/about/.

[103] Martin Malmsten, "Making a Library Catalogue Part of the Semantic Web," in *Proceedings of the International Conference on Dublin Core and Metadata Applications* (DC-2008, Berlin, Germany: Universitätsverlag Göttingen, 2008), 146–50, http://www.oapen.org/download?type=document&docid=610315#page=162.

[104] Donna E. Frederick, "Metadata Specialists in Transition: From MARC Cataloging to Linked Data and BIBFRAME (Data Deluge Column)," *Library Hi Tech News* 33, no. 4 (June 6, 2016): 1–5, https://doi.org/10.1108/LHTN-03-2016-0015; Library of Congress, "BIBFRAME - Bibliographic Framework Initiative."

[105] David Mimno, Gregory Crane, and Alison Jones, "Hierarchical Catalog Records: Implementing a FRBR Catalog," *D-Lib Magazine* 11, no. 10 (October 2005), https://doi.org/10.1045/october2005-crane; Library of Congress, "BIBFRAME - Bibliographic Framework Initiative."

[106] Cronin et al., "Strength in Numbers: Building a Consortial Cooperative Cataloging Partnership."

As an introduction to the second domain of literature, which is related to the balance of decentralization and centralization in a shared cataloguing system, it is worth reviewing debates about centralization in the field of librarianship. The earliest discussion of the topic discovered in the literature is from 1939, and though it focuses on centralization in terms of a library's physical configuration, it echoes Ranganathan's warning about the Law of Parsimony[107] by framing the discourse as a balance of efficiency and adequacy in library services.[108] The cost of service provision is central to all discussions of library centralization since that time, whether it is framed positively or negatively: economic recessions in the late 1990s and 2000s suggested that austerity would drive consolidation and central control of services,[109] and the emergence of next-generation catalogues promised cost-efficient, centralized access to previously-siloed information sources.[110] Implicit in these debates is the assumption that decentralized systems are more expensive than their centralized counterparts; notably, however, a four-scholar interchange in 1983[111] presaged Yochai Benkler's more general finding in 2006[112]

---

[107] Ranganathan, *Theory of Library Catalogue*, 55.

[108] Robert A Miller, "Centralization versus Decentralization," *ALA Bulletin* 33, no. 2 (February 1939): 75.

[109] Jean Costello, "An Inflection Point for American Public Libraries," *In the Library with the Lead Pipe*, 2009, http://www.inthelibrarywiththeleadpipe.org/2009/an-inflection-point-for-american-public-libraries/; Lauren Seiler and Thomas Surprenant, "When We Get the Libraries We Want, Will We Want the Libraries We Get?," *Wilson Library Bulletin* 65, no. 10 (June 1991): 29–31, 152,157.

[110] Joshua Barton and Lucas Mak, "Old Hopes, New Possibilities: Next-Generation Catalogues and the Centralization of Access," *Library Trends* 61, no. 1 (2012): 83–106, https://doi.org/10.1353/lib.2012.0030.

[111] Hugh C. Atkinson, "A Brief for the Other Side," *Journal of Academic Librarianship* 9, no. 4 (1983): 200–201; Edward G. Holley, "Reaction to 'A Brief...,'" *Journal of Academic Librarianship* 9, no. 4 (1983): 201–2.

[112] Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, CT: Yale University Press, 2006), 2–8.

that new advances in technology would eliminate the cost differential for centralized and decentralized approaches.

As a means of effecting decentralization, blockchain-based technology has had no notable impact on librarianship and existing work is still preliminary. The largest effort to date was launched in 2017 when San Jose State University (SJSU) received an Institute of Museum and Library Services (IMLS) grant to explore library uses for blockchains;[113] funds from this grant catalyzed two symposia in 2018 to discuss possibilities. Librarianship's responses to blockchain technology reflect varying interpretations of the role and function of decentralized ledgers, likely due to the constantly-evolving nature of the technology itself. Many information professionals hold a limited view of blockchains as a secure data storage mechanism[114] and align possible library deployments with that model; though the catalogue has been mentioned,[115] commonly-cited applications are for copyright registration,[116] management of community-based collections (e.g. tools libraries),[117] record keeping and records management.[118] The notion of a blockchain as a general-purpose

---

[113] San Jose State University, "Blockchains for the Information Profession," Blockchains for Libraries, accessed December 21, 2018, https://ischoolblogs.sjsu.edu/blockchains/.

[114] Debbie Ginsberg, "Blockchain: What It Is, How It's Being Used, and What It Means for the Future of Law Libraries.," *AALL Spectrum* 22, no. 1 (2017): 37; Matt Enis, "SJSU-Led Team Explores Blockchain," *Library Journal* 143, no. 1 (January 2018): 16; Nancy K Herther, "Blockchain Technology in the Library," *Online Searcher*, October 2018, 37.

[115] Enis, "SJSU-Led Team Explores Blockchain," 17.

[116] David Ensign, "Copyright Corner: Blockchain and Copyright," *Kentucky Libraries* 82, no. 3 (2018): 4–5.

[117] "Use for Blockchain in Libraries," *Blockchains for Libraries* (blog), September 13, 2017, https://ischoolblogs.sjsu.edu/blockchains/blockchains-applied/applications/.

[118] Cassie Findlay, "Participatory Cultures, Trust Technologies and Decentralisation: Innovation Opportunities for Recordkeeping," *Archives and Manuscripts* 45, no. 3 (September 2, 2017): 176–90, https://doi.org/10.1080/01576895.2017.1366864; Victoria Louise Lemieux, "Trusting Records: Is

computer, using platforms like Ethereum, is less-reflected in LIS literature[119] and little work has been done in this area. Proposals have included a research data rights management system,[120] access and rights management for eBooks,[121] and royalty tracking and distribution for authors and musicians.[122]

Current investigations of the use of blockchain technology have been perceived as part of the "hype cycle," where excitement over a technology overpowers rational investigations into its ideal application.[123] Few in-depth models for library-focused blockchain applications have been suggested,[124] and no work in the LIS field thus far examines the alignment of decentralized ledger technologies with the organization of libraries or their services.

Blockchain Technology the Answer?," *Records Management Journal* 26, no. 2 (July 18, 2016): 110–39, https://doi.org/10.1108/RMJ-12-2015-0042.

119 Brendan Howley, "The Razor's Edge: Blockchain, Ledger Legerdemain, and the Public Library," *Information Today*, November 2016, 15; Ginsberg, "Blockchain: What It Is, How It's Being Used, and What It Means for the Future of Law Libraries.," 38–39.

120 Adrian-Tudor Pãnescu and Vasile Manta, "Smart Contracts for Research Data Rights Management over the Ethereum Blockchain Network," *Science & Technology Libraries* 37, no. 3 (July 3, 2018): 235–45, https://doi.org/10.1080/0194262X.2018.1474838.

121 Bill Rosenblatt, "Can Blockchain Disrupt the E-Book Market? Two Startups Will Find Out," Forbes, accessed August 30, 2018, https://www.forbes.com/sites/billrosenblatt/2018/08/18/can-blockchains-disrupt-the-e-book-market-two-startups-will-find-out/.

122 Howley, "The Razor's Edge: Blockchain, Ledger Legerdemain, and the Public Library," 15.

123 Matthew B. Hoy, "An Introduction to the Blockchain and Its Implications for Libraries and Medicine," *Medical Reference Services Quarterly* 36, no. 3 (July 3, 2017): 278, https://doi.org/10.1080/02763869.2017.1332261; Ginsberg, "Blockchain: What It Is, How It's Being Used, and What It Means for the Future of Law Libraries.," 36; Stinchcombe, "Ten Years in, Nobody Has Come up with a Use for Blockchain."

124 The most complete model found in the literature to date is Pãnescu and Manta, "Smart Contracts for Research Data Rights Management over the Ethereum Blockchain Network."

In summary, literature related to blockchain technology's implications for libraries is emerging and of interest to the field, but gaps exist for in-depth explorations of how blockchains may be used to address specific practices. Work undertaken by SJSU has suggested cataloguing as a potential use for the blockchain,[125] but the engine of the modern catalogue has not been problematized in recent literature, and no fulsome examination of its possibilities and challenges in this area has been undertaken. One method for this exploration is the creation of a theoretically-derived design framework for a blockchain-based shared catalogue and a high-level design for the catalogue itself. These artefacts, intended to address the opportunities outlined in Chapter 2, are used to address the research questions outlined in the introduction. To begin, the next chapter scaffolds the design by outlining theories, technologies and practices that might serve as the foundation for a shared cataloguing system based on blockchain technology.

---

[125] SJSU iSchool, "Blockchains for the Information Profession," Blockchains for Libraries, 2017, https://ischoolblogs.sjsu.edu/blockchains/.

4. Design Framework for a New Shared Catalogue

Following an outline of challenges facing contemporary shared cataloguing systems and the illustration of a gap in recent LIS-related research, this chapter presents detailed aspects of a design framework that can be used to drive a theoretical exploration of Chapter 1's research questions. The design framework consists of three components: a discussion of the approach to decentralization and its alignment with the practice of shared library cataloguing; an overview of the Ethereum framework, including the nature of distributed ledger (blockchain) technology and the unique affordances of Ethereum; and an overview of studies of open source software and social information production systems, which has been used by Yochai Benkler to abstract a set of design levers for fostering sustainable, open social information production.

This approach to the design framework, which combines technological and socio-behavioural mechanisms, is intentional. In order to address the research questions that have been posed, a shared cataloguing design must not only address the technical challenges of building a catalogue with decentralized blockchain technology, but also the systems intended to guide (but not pre-determine) the catalogue's sustainable implementation, its growth, and its embrace of public participation. This approach is informed by Darrin Barney's description of the social constructivist view of technology, of which he is also critical.[126] Briefly, the social constructivist view states that the outcomes of technological implementations are "underdetermined" by technology itself, since they are

---

[126] Darin David Barney, *The Network Society*, Key Concepts (Cambridge, UK: Polity, 2010), 39–43.

ultimately produced by the interactions between technologies and their social environments. Barney's criticism of this approach is that it may focus on local, contextual cases that resist later abstraction.

Barney's criticism is well-heeded, and he prefers a composite approach linking constructivism with both instrumentalism (technological neutrality) and substantivism (technology as an embodiment of specific values, generally associated with the idea of technological determinism),[127] but the arguments articulated in the preceding chapters are already implicitly rooted in the context of librarianship's core values[128] and, in particular, their interaction with Ranganathan's *Five Laws of Library Science*. Moreover, the scope of this thesis is limited to the *conceptual design* of a shared cataloguing system, leaving the success of an actual implementation subject to the effect of social incentives and other factors that cannot be predicted without additional study. The framework explored in this chapter is not divorced from North American librarianship's ethical context, and cannot be studied in hindsight, so a social constructivist approach seems appropriate. Implicitly-structured by the core values of librarianship, the framework elements outlined in this chapter are subsequently used to guide the design of the new shared-cataloguing engine presented in Chapter 5.

The modern catalogue's connections to Ranganathan's five laws have been addressed in the "Background" chapter, and the concept of decentralization has been introduced, but a characterization of decentralization has not yet been presented. Since the

---

[127] Barney, 35–43.

[128] American Library Association, "Core Values of Librarianship," Text, Advocacy, Legislation & Issues, July 26, 2006, http://www.ala.org/advocacy/intfreedom/corevalues.

focus of blockchain technology is the elimination of centralized control and points of failure, and since it serves as the core of the new shared-cataloguing engine, a more detailed model of decentralization and its application to blockchain technology must be addressed here first. The presentation of this model is complemented by a brief exploration of the distribution of shared cataloguing work, illustrating its inherently-decentralized character. This is followed by an in-depth description of the Ethereum framework, which has been chosen as this thesis' blockchain technology. Finally, this chapter outlines Yochai Benkler's design levers for the creation of sustainable social information production systems and draws connections between those levers and current practices in librarianship.

Decentralization

Ronald Coase's model of the *firm* is based on the idea that the transaction costs are minimized when processing steps are internalized within an organization.[129] Instead of relying on a market to negotiate and perform each step in a process—possibly involving the physical transport of goods, time spent negotiating prices and administering agreements, etc.—a firm handles these steps internally. Minimizing costs within a firm has traditionally meant removing factors related to time, distance, and processing, all of which can be addressed with centralization and automation. Modern communications technologies and their emphasis on networking ability have served to minimize these costs drastically within firms (referred to by Manuel Castells as *timeless time* and the *space of*

[129] Coase, *The Firm, the Market, and the Law*, 5.

flows),[130] but they have also economized them *among* organizations, within markets, and even among different markets. The power of networks to reduce costs is so significant that it has been seen by scholars such as Yochai Benkler as an enabler of new, decentralized forms of knowledge production that can compete with the efficiencies afforded to traditional firms.[131] To be clear: network effects have reduced the cost of both centralized *and* decentralized activity, but Benkler argues that their effect on decentralized activity has renewed interest in novel forms of organization and information production that do not require traditional forms of centralization.

The proposition made by networked blockchain technology is that it can match the efficiency of centralized systems with its novel implementation of the principles of decentralization. To set the stage for the design framework presented in this chapter, then, a more specific definition of decentralization is required. The concept of *decentralization* has multiple facets, but in the rhetoric associated with blockchain-based systems "decentralization" is often listed as a general, implicit benefit of the technology. Sadly, this renders its practical interpretation vague.

One approach to the study of centralization is to partition it along administrative and technical lines. For example, work has shown that the organization of systems technology and their associated administration tend to align: when administrative functions are centralized, the underlying technical infrastructure is likewise centralized.[132]

---

[130] Barney, *The Network Society*, 28–31.

[131] Benkler, *The Wealth of Networks*, 2–8.

[132] George and King, "Examining the Computing and Centralization Debate," 64.

Others have defined decentralization through the lenses of organizational structures[133] or information ecosystems.[134] Synthesizing these perspectives, Vitalik Buterin employs a technology-centered approach to his model of decentralization. He proposes a definition that divides decentralization's focal points along three technologically-linked axes:[135]

1. **Architectural decentralization**, which is related to the physical computing architecture of a system. Decentralization here could be measured by the number of systems that can break down before the system fails. At one extreme, a fully centralized system would be managed by a single computing platform that, if compromised, renders the system inoperable. The more distributed a computing architecture is, the more tolerant it is of points of failure.
2. **Political decentralization**, which addresses the number of individuals or organizations that control the computing devices within the system. A key example of centralization in this area is the realm of Content Delivery Networks (CDNs) like Amazon's CloudFront: though the infrastructure is fault-tolerant due to the widespread placement of computing devices in different areas of the network (architectural decentralization), they are all controlled by a single corporate entity that can arbitrarily create and enforce

---

[133] Ori Brafman and Rod A Beckstrom, *The Starfish and the Spider: The Unstoppable Power of Leaderless Organizations* (New York, NY: Portfolio, 2014), chaps. 1–2.

[134] Johanna Möller and M. Bjørn Von Rimscha, "(De)Centralization of the Global Informational Ecosystem," *Media and Communication* 5, no. 3 (September 22, 2017): 37, https://doi.org/10.17645/mac.v5i3.1067.

[135] Vitalik Buterin, "The Meaning of Decentralization," *Vitalik Buterin* (blog), February 6, 2017, https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274.

access policies (such as disallowing applications like Telegram from hiding their identity to bypass national network censorship).[136]

3. **Logical decentralization**, which consists of the data structures and interfaces that the system presents for interaction by people or other systems. The best definition for this concept is implicit in the test for its existence. Following the model of organizational decentralization named as the *starfish* by Brafman and Beckstrom:[137] if the system is cut in half, can each piece continue to operate as an independent unit? One example of a technology that functions in this way is the BitTorrent system, which provides a peer-to-peer mechanism for locating and downloading files over the Internet. As long as one copy of the file can be constructed, regardless of the source of its pieces and whether or not those pieces are, themselves, parts of complete files, the system can deliver an intact file download.[138]

Blockchain technologies have been suggested as a new "third way:" an efficient, decentralized alternative to purely market-driven systems and centrally- (or state-) planned ones.[139] By decentralizing governance through software-based protocols and heuristics, we can bypass the tendency towards centralization that the modern drive

---

[136] Russell Brandom, "Amazon Web Services Starts Blocking Domain-Fronting, Following Google's Lead," The Verge, April 30, 2018, https://www.theverge.com/2018/4/30/17304782/amazon-domain-fronting-google-discontinued.

[137] Brafman and Beckstrom, *The Starfish and the Spider*, 31–38.

[138] BitTorrent.org, "What Is BitTorrent?," 2015, http://www.bittorrent.org/introduction.html.

[139] Eric A Posner and E. Glen Weyl, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society* (Princeton, NJ: Princeton University Press, 2018), chap. 1.

towards efficiency demands of most organizations.[140] Though Buterin cites BitTorrent as an example of a system that has achieved decentralization on all three axes, he contradicts many blockchain evangelists by claiming that blockchain solutions *cannot* achieve decentralization at the same level. The flaw, he states, is in the blockchain's inability to achieve logical decentralization.

Buterin's assertion warrants further explanation. Fundamentally, blockchains cannot attain logical decentralization because they rely on statistical certainty supplied by a commonly-agreed-upon trail of data, and these data are stored in an ordered format—chronologically—in the blockchain itself. Applying the test for logical decentralization: if you cut a blockchain in half by separating more recent transactional data (the "top half") from older data (the "bottom half"), and then try to operate each half in parallel as an independent system, problems quickly manifest. The information on more recent transactions (the top half) is less reliable because it is not statistically-supported by a previous history of verified transactions (from the bottom half). In addition, the parallel system based on data from the bottom half of the original chain will process transactions differently than the top half. This is because data from newer transactions, which are locked into the structure of the top half, are not present in the bottom half. In summary, each of the "halves" of the chain represent a different *state* and, as a result, each half will process future transactions differently.

---

[140] MIT Technology Review, "In Blockchain We Trust," *MIT Technology Review* (blog), April 9, 2018, paras. 8–12, https://medium.com/mit-technology-review/in-blockchain-we-trust-1cafe1c914b2.

Buterin further points out that architectural decentralization in software systems is difficult to achieve because of the introduction of common modes of failure, often at the hands of political systems. For example, if all nodes in a network are running the same piece of software and that software has a fatal flaw, the whole system is vulnerable. An example of this has already occurred on the Ethereum platform: an entirely self-governed financial system called the Decentralized Autonomous Organization (DAO) failed in 2016 when hackers found and exploited flaws in the implementation of its contract execution code.[141]

As a result of the challenges outlined in this section, the best-case scenario for decentralization in a blockchain-based solution is one that enjoys architectural and political decentralization. The political component of blockchain-based architectures is dependent on human-centered factors of software design and organizational control, which is why the design of software elements and social incentives both play a pivotal role.

Though his model largely applies to software systems, Buterin does attempt to apply it to non-software entities as well. Here we see how his model connects with existing discourse in librarianship. Libraries are, for the most part, politically and logically decentralized,[142] but less architecturally decentralized due to their increasing reliance on

---

[141] Samuel Falkon, "The Story of the DAO — Its History and Consequences," *Medium* (blog), December 24, 2017, https://medium.com/swlh/the-story-of-the-dao-its-history-and-consequences-71e6a8a551ee.

[142] Political decentralization arises from the local and regional arrangement of libraries as organizations, mirroring the scope at which they are funded; though most library systems are members of associations, they are (for the most part) administratively independent of one another. Logical decentralization is less attributed to the availability of services than the general consensus on information storage and management standards, as well as the arrangement of systems for information organization and retrieval.

centralized computing architectures for cataloguing and other services. The work of cataloguers is increasingly being outsourced to external organizations, and these organizations have their own technical and political infrastructures. For example, the dominant cataloguing services provider, OCLC, is centralized along all three of Buterin's axes. This tendency follows the pattern outlined in the beginning of this section: the technical centralization of a service is linked to administrative/political centralization. It also raises the question of whether or not the ground-level, politically-decentralized structure of libraries can be sustainably-mirrored in an architecturally-decentralized solution for their services.

Decentralization and shared library catalogues

During the exploration of Ranganathan's five laws in Chapter 2, I asserted that the work of library cataloguers is inherently decentralized. Evidence presented here in support of this assertion was discovered in an exploration of how current cataloguing data—largely based on the MARC21 data format—might be used to record historical information in catalogue records.

A library that creates an original catalogue record may, optionally, stamp its identity into MARC field 040; evidence of subsequent revisions to the record can be marked with the addition of "modifying cataloguer" values in the same field. The use of this field is optional, but its implementation in open catalogue records obtained from Harvard

University[143] and the University of Michigan[144] exposes the distributed nature of shared cataloguing work. Roughly three million catalogue records were extracted from open catalogue data provided by these institutions, and these data were used to create two graphs of the relationships between original cataloguing and modifying institutions.

Figure 1 (p. 48) and Figure 2 (p. 49) represent filtered graphs of data from the 040 field,[145] outlining the connections between institutions that create and modify catalogue records. Node sizes are governed by the number of records created or modified by an institution, and links between nodes connect institutions that have created a record to the ones that have modified that record. Node names are as provided in MARC record data, with most terms conforming to library codes assigned by either the Library of Congress or OCLC. The dominance of the Library of Congress (node "dlc") is clear in Figure 1, as is the prevalence of OCLC-related entities ("oclcq" *et. al.*) in both graphs. Loopbacks (nodes linking back to themselves) are evident for some institutions, meaning that these libraries are editing their own records. Overall, however, the *centrality* of both graphs is of note: the variations in cataloguing participation fall within a narrow range, supporting the assertion that shared cataloguing work is inherently decentralized.

This finding, though not based on rigorous analysis, suggests that there is a positive alignment between the architectural decentralization of blockchain technologies and the

---

143 Harvard Library, "Harvard Library Open Metadata," Harvard Library, 2017, https://emeritus.library.harvard.edu/open-metadata.

144 University of Michigan, "University of Michigan MARC Catalog Records," Internet Archive, 2010, https://archive.org/details/UniversityOfMichiganMarcCatalogRecords.

145 Only the most active catalogue record editors are included in the diagrams to assist with visual clarity. Complete graphs contain thousands of nodes each, many of which represent institutions that contribute catalogue data on an infrequent basis.

political decentralization of cataloguing work. Unlike blockchain-based applications, such as those designed for financial systems, that aim to create political decentralization through architectural means, a blockchain-based catalogue aims to return the current, politically-centralized activity of shared cataloguing to its politically-decentralized roots.

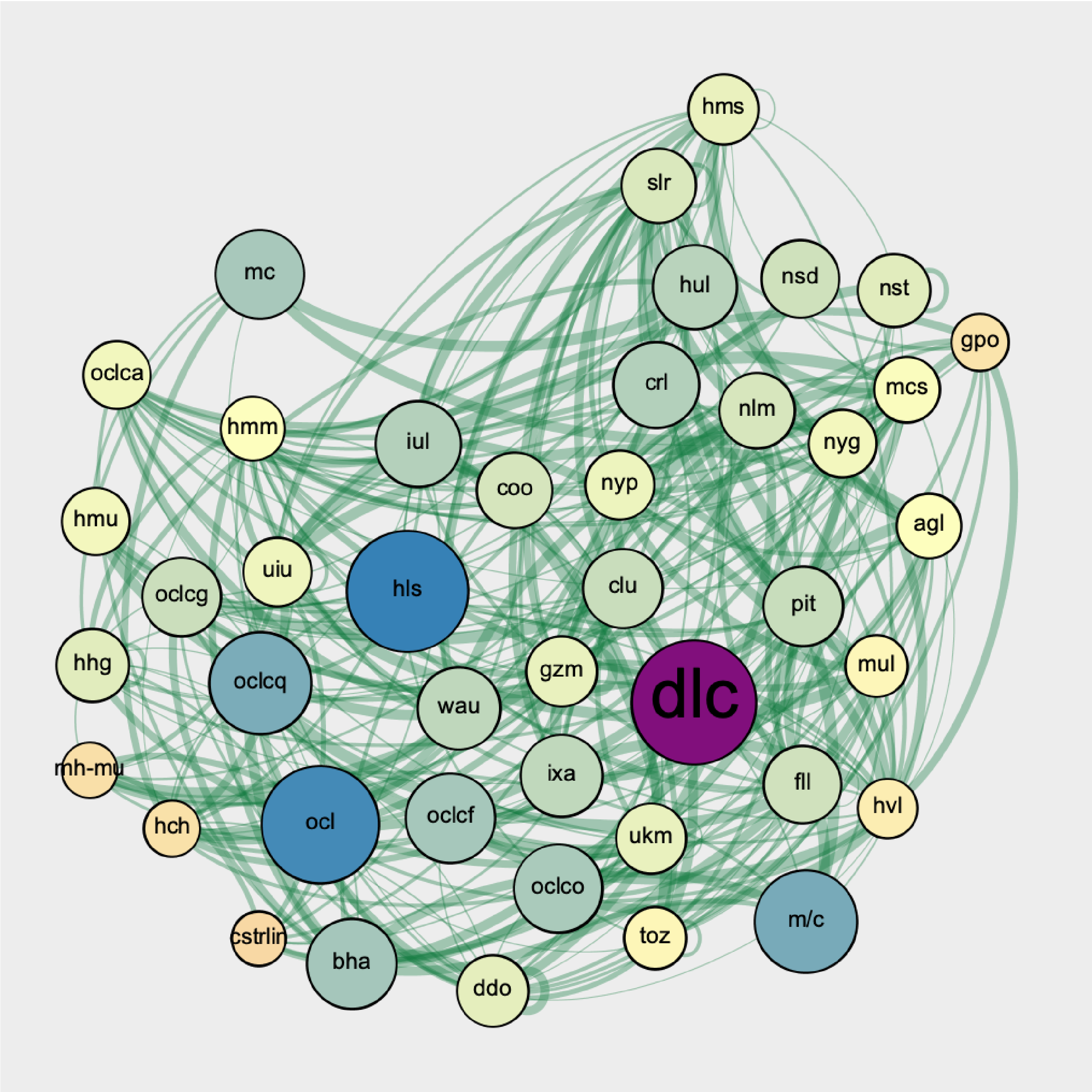

Figure 1: Catalogue record creators and modifiers (filtered view) for a subset of Harvard University's open catalogue data

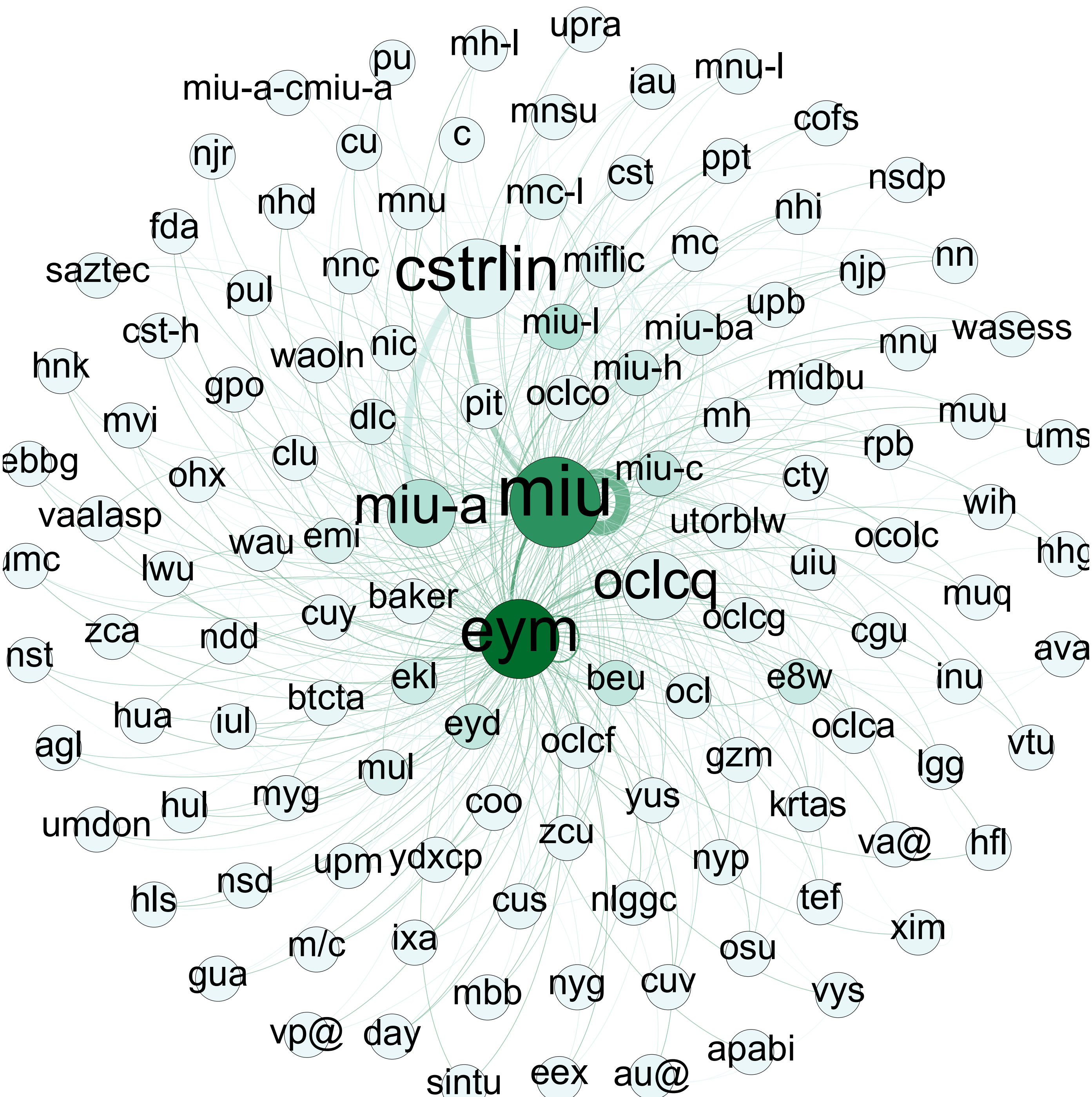

Figure 2: Catalogue record creators and modifiers (filtered view) for a subset of University of Michigan's open catalogue data

Ethereum foundations

With an acknowledgment of the decentralized nature of shared cataloguing work and the presentation of a functional model for assessing decentralization, our attention turns to the implementation of decentralized blockchain technology under the Ethereum model. The 2008[146] creation of the open source Ethereum project follows the 2001 debut of Bitcoin, after several years of other iterative advances on Satoshi Nakamoto's disruptive blockchain model.[147] Rather than a system intended to serve as a payment ledger, Ethereum was created to function as a general-purpose, blockchain-based computing platform. Vitalik Buterin, one of the co-creators of Ethereum, uses a simple metaphor to explain the difference.[148] He imagines Bitcoin as a basic electronic calculator, created to perform specific mathematical functions. Subsequent advancements of Bitcoin technology (for example, MasterCoin and PeerCoin) added features to the basic model but illustrated a path-dependent approach: the underlying calculator was the same, but with new buttons added to provide support for additional mathematical functions. Ethereum, by contrast, represents the shift from a calculator to a smartphone: a general-purpose computing device that can run a calculator (or an advanced scientific calculator) as an application, but can also run any software application envisioned and built by designers to use the smartphone's interface.

---

[146] Though it was conceived in 2008, the first iteration of software that created the public Ethereum network was not released until July, 2015.

[147] Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System."

[148] Ethereum Foundation, *Devcon2: Ethereum in 25 Minutes*, Ethereum Developer Conference, 2016 September 19-21 (Shanghai, China, 2016), https://www.youtube.com/watch?v=66SaEDzlmP4.

Retreating to a more technical definition, Ethereum is described as a *cryptographically-secure transaction state machine*: in other words, a system that provides a single, objective truth about the information it stores and how that information was derived, backed by accuracy and trust provided through encryption algorithms and digital signatures.[149] Underlying that premise is the fact that Ethereum—like its most well-known predecessor, Bitcoin—is intended to be distributed across a network of individual actors who do not need to be trusted in order to build trust into the system. Unlike Bitcoin, however, Ethereum is designed to function as a general-purpose computing platform instead of a global monetary system.

In a classical economic sense, the decentralized Ethereum blockchain may be viewed as an embodiment of a *firm* as defined by Coase: an organization that transforms input into outputs.[150] Anyone wishing to use the Ethereum system for computing tasks is welcome to submit the task to the network, where it is processed by other participants in the system. Unlike a traditional Coasian firm, however, there is no central body that manages the work of processing transactions. Just as anyone can submit a task to the system for processing, anyone can work to process tasks. The operation of the network is governed by software, which uses two novel techniques to eliminate the need for a central authority. These techniques are addressed, in turn, in the next two sections.

---

[149] Preethi Kasireddy, "How Does Ethereum Work, Anyway?," *Preethi Kasireddy* (blog), September 27, 2017, https://medium.com/@preethikasireddy/how-does-ethereum-work-anyway-22d1df506369.

[150] Coase, *The Firm, the Market, and the Law*, 5; Nick Tomaino, "The Slow Death of the Firm," The Control, October 21, 2017, https://thecontrol.co/the-slow-death-of-the-firm-1bd6cc81286b.

Cryptographic signatures

All elements of the Ethereum system are made verifiable and tamper-proof by using cryptographic digital signatures. The platform leverages open standards established by the National Institute of Standards and Technology (NIST). NIST's Shared Hashing Algorithm (SHA) version 3 specification[151] was released in 2015, and it outlines a publicly-visible and verifiable mechanism for creating secure, private encryption keys and digital signatures. Ethereum is currently implemented using 256-bit hashing algorithms.

The ability for network participants to generate secure hashes[152] and digital signatures is critical to the implementation of a reliable blockchain. Their implementation means that:

1. The creator of a piece of data can be independently-verified, even if the identity of the creator is unknown. The private key used to *sign* data is unique and not shared but plays a fundamental role in the computation of the digital signature. With a simple computation on the receiving end, the validity and uniqueness of the key can be easily proven. By way of analogy, think of a solved Sudoku puzzle: the exact steps used to find a valid puzzle solution, or how long it took, may be impossible to know, but it is easy to check a completed Sudoku puzzle to see if the proposed solution is a correct one.

---

[151] Morris J. Dworkin, "SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions" (National Institute of Standards and Technology, July 2015), https://doi.org/10.6028/NIST.FIPS.202.

[152] *Hash* is simply the name given to the value output by any function that can be used to verify data integrity. See Techopedia, "What Is a Cryptographic Hash Function?," Techopedia.com, 2019, https://www.techopedia.com/definition/27410/cryptographic-hash-function.

2. The integrity of any digitally-signed data can be verified by ensuring that the cryptographic hash of the data has not been altered. Changing even a single bit of information in the source data results in a completely different hash value that cannot be predicted with current computing technology. An altered hash value exposes the fact that data has been changed since its creation.

Most fundamentally: cryptographic hashes and digital signing techniques allow trust in data and its creators to be established by anyone, even if the creator of the data, or the data itself, is unknown. This lays the groundwork for a system that can foster trust even if its participants and constituent data are anonymous and/or geographically-distributed.

Blockchains

Since cryptographic technologies allow the source and integrity of data to be transparently verified by anyone, we are able to demonstrate at any time that the "state" of the system is accurate and has not been altered. This is not sufficient for the creation of a decentralized system, however. To remove the need for a central authority that maintains and manages this data, and to ensure that data is not lost, we need a reliable way to distribute the same data to all of the participants in the network (along the same model as librarianship's Lots of Copies Keep Stuff Safe (LOCKSS) paradigm).[153] If everyone has a copy of the full state of the system and everyone can validate the data, then no trusted, central authority is needed. The challenge lies in ensuring that the system can accurately-synchronize the data across the network as new information is added.

---

[153] Stanford University, "LOCKSS," accessed January 17, 2019, https://www.lockss.org/.

If the data were static and unchanging, this would be trivial: all that would be needed is to verify that a hash of anyone's stored data set matches everyone else's. The goal, however, is to create a dynamic system that can process new information as it is introduced. Since it is not possible to predict the hash of a data set based on an introduced change, we need a mechanism whereby all network participants can accept and process new data, with full agreement on the ordering and content of that data. The ability to independently arrive at a shared consensus on the content of the data, its source, and its placement within the data chain forms the second critical component of the blockchain paradigm.

Blockchain-based systems arrive at a consensus about the "truth" of an evolving data set in the following way:

1. Any new transactions that would alter the state of the system (by adding or changing data) are visible in a *pool* of unprocessed transactions on the network.
2. Network participants who are willing to process these transactions compete for the right to combine a set of transactions into a bundle called a *block*. The dominant mechanism for this requires a winning participant to show that it has solved a cryptographic problem posed by the network in an algorithmic manner. This proof acts as evidence that the participant did some computing work in order to arrive at a correct response (verified in the same fashion as the Sudoku puzzle mentioned earlier), leading to the term *proof of work*.[154]

---

[154] This is only one of many mechanisms under discussion for submitting proof of the right to build a block. Recognizing that "proof of work" systems are computing-intensive (and therefore costly in terms of

3. All possible winning participants create a block by processing a set of transactions from the pool and putting those results into a formalized data structure that is, itself, digitally-signed. This process is referred to as *mining*. Blocks sizes are limited by the network, and each transaction includes a *fee* that is claimed by the participant who successfully builds a block. Participants are therefore incentivized to create the largest possible block to maximize their reward of fees; this is intended to ensure that all available transactions in the pool are eventually processed.
4. Once completed, the block is cryptographically-signed and submitted to the network. Other network participants pick up these blocks and add them to the end of the existing set of blocks, lengthening the *chain* referred to in the term "blockchain".
5. Since it is possible (even likely) that multiple blocks will be submitted to the network concurrently, or that not all network participants receive new blocks at the same time, a *consensus* mechanism is required to ensure that all participants can agree on *which* block is the one that everyone should accept. Ethereum does this using the Greediest Heaviest Observable SubTree

---

energy consumption), other models are being investigated and used by other systems. These include proof of stake, delegated proof of stake, and even proof of space-time. Though Ethereum currently uses a Proof of Work model of the type used by Bitcoin, it hopes to transition to a distributed Proof of Stake system in the near future. The issues associate with these proof systems are complex and are beyond the scope of this thesis. For an overview of protocols and their challenges, see Vitalik Buterin and Virgil Griffith, "Casper the Friendly Finality Gadget," *ArXiv.Org*, October 25, 2017, http://arxiv.org/abs/1710.09437; Sunny King and Scott Nadal, "PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake," August 19, 2012, Internet Archive, https://archive.org/details/PPCoinPaper; Will Little, "A Primer on Proof-of-Stake and Why It Matters for the Future of Blockchains," Hacker Noon, May 14, 2018, https://hackernoon.com/a-primer-on-proof-of-stake-and-why-it-matters-for-the-future-of-blockchains-48764373d4b1.

(GHOST) protocol,[155] where participants choose the block that adds the most amount of overall work to the chain.[156] Since all participants apply the same algorithmic rules for selecting the authoritative block, the entire network will eventually arrive at the same decision about the state of the system's data.

6. The previous steps repeat in an endless cycle, with the creation of new blocks confined to targeted, time-based intervals. On the Ethereum network, a new block is issued every 15 seconds on average.[157]

The combination of cryptographic signatures and a distributed consensus mechanism allow participants of the Ethereum network, collectively, to act as a firm; however, the network also functions as a market (following Coase's definition) since it facilitates exchange.[158] An alternate view, espoused by some who see blockchain

---

[155] Kasireddy, "How Does Ethereum Work, Anyway?"; Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 2, 13–15.

[156] Since the network is distributed, it is possible for multiple participants to make different decisions about which block to include in the chain based on the information they have (some blocks may be delayed on the network; may not have propagated to all nodes, etc.). Situations where this occurs are called "forks." However, the GHOST protocol is designed in such a way that forked or inaccurate blocks - even if they have created small sub-chains - are eventually rejected in favor of the chain that represents the most amount of computing work. Forks are common in blockchain networks but resolve over time; as a result, the finality of data in a blockchain improves statistically as new blocks are embedded deeper in the chain. By corollary, the newest block in the chain should not be trusted until several additional blocks have subsequently been added.

[157] This protocol was introduced by Ethereum Improvement Proposal (EIP) 2. See Vitalik Buterin, "EIP-2: Homestead Hard-Fork Changes," Ethereum Improvement Proposals, November 15, 2015, http://eips.ethereum.org/EIPS/eip-2.

[158] Coase, *The Firm, the Market, and the Law*, 7.

technologies as the basis of a new field of *cryptoeconomics*,[159] is that these technologies eliminate the need for the traditional firm altogether.[160]

Having set a foundation with the two core concepts of blockchain technology, we may now explore some of the unique properties of Ethereum's blockchain implementation: *nodes*, *contracts*, and *gas*. Each of these topics is addressed in the following sections.

Nodes

As with other blockchain implementations, each connection point on the Ethereum network is referred to as a *node*. Instead of requiring all connection points on the network to maintain a complete copy of the system's full dataset, however, Ethereum creates flexibility by allowing three models for operating nodes and for connecting to them.

A *full node* processes transactions and validates cryptographic hashes for every block using a locally-stored copy of network data. To do this, a full node must store and maintain a full copy of the Ethereum blockchain, verifying all historical transactions before participating in the work of verifying all new transactions and blocks as they arrive. Full nodes also have the ability to function as *miners*, actively competing for the right to construct a block and reap its associated rewards.

One alternative to storing and verifying the full blockchain is the storage and verification of a lighter data structure consisting only of the cryptographic hashes of

---

[159] BlockChannel, "A Crash Course in Mechanism Design for Cryptoeconomic Applications," *Medium* (blog), October 17, 2017, https://medium.com/blockchannel/a-crash-course-in-mechanism-design-for-cryptoeconomic-applications-a9f06ab6a976; Josh Stark, "Making Sense of 'Cryptoeconomics,'" *L4 Media* (blog), November 16, 2017, https://medium.com/l4-media/making-sense-of-cryptoeconomics-5edea77e4e8d.

[160] Tomaino, "The Slow Death of the Firm."

previous blocks. Because the data is composed of cryptographic hashes, it is still possible to verify the integrity of the blockchain even though the full set of transaction data is not available. To do this, Ethereum makes heavy use of a binary data structure called a Merkle Tree,[161] where the parent of each pair of tree leaves is a hash of the leaves' combined values. A node of this kind is able to verify the hashes of data in the blockchain by validating the combined hashes in the tree hierarchy; since these hashes can be verified, it is assumed that the underlying data is also accurate. These *light nodes* store a much smaller subset of Ethereum blockchain data, and the integrity of the blockchain is deemed safe with the proviso that the system's data should still be anchored by a large number of distributed, independently-operated full nodes.[162]

As development on the Ethereum project has progressed in the application space, a third client model has emerged to support users whose network connections may be intermittent or who have limited access to data storage. A *remote client* works with blockchain data by communicating with a full node over Remote Procedure Call (RPC) connections. Though this model creates flexibility for access to the Ethereum network—in particular, enabling distributed application access for mobile and Internet of Things (IOT) devices that have minimal data storage and computation ability—it has also created a market for node providers like Infura,[163] who operates full nodes that can be used to build

---

[161] Zheng et al., "An Overview of Blockchain Technology"; Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 18–19.

[162] Thibaut Sardan, "What Is a Light Client and Why You Should Care?," Blockchain Infrastructure for the Decentralised Web, July 26, 2018, https://www.parity.io/what-is-a-light-client/.

[163] Infura, "Your Access to the Ethereum Network," Infura - Scalable Blockchain Infrastructure, 2019, https://infura.io.

and operate blockchain applications. The popularity of Infura, in particular, has recently exploded, causing some to observe that it forms a nexus of architectural and political centralization (and, therefore, a single point of failure) for the Ethereum network.[164]

Contracts

The Bitcoin platform largely serves as a ledger of account balances, as intended by its function as a monetary system. Ethereum aims to function as a general-purpose computer, though, so mechanisms must be created to allow computation to occur on the network. These software functions on the Ethereum platform are referred to as *contracts,* since they represent agreements on how input data should be processed to arrive at a result.[165]

The architecture of contracts is similar to that of the object-oriented software paradigm: contracts are written to accept a specific set of inputs and generate a predictable type of output. The distinction under the blockchain paradigm is that every contract on the Ethereum network is distributed to every network node and can be run on any of those nodes. Computing instructions in contracts are executed as part of transaction processing on *every* full node so that the result does not have to separately be distributed to the network.

---

[164] Sardan, "What Is a Light Client and Why You Should Care?," para. 7; Ian Kennedy, "Infura Manages Most of the Ethereum (ETH) Network's Nodes, but This Could Affect Decentralization," Crypto Daily Gazette, July 30, 2018, https://cryptodailygazette.com/2018/07/30/infura-manages-most-of-the-ethereum-eth-networks-nodes-but-this-could-affect-the-decentralization/.

[165] The astute reader may note that the decentralized processing of data with contracts must be deterministic—i.e. every node that executes the same piece of code must arrive at the same result. For this reason, the Ethereum network is particularly ill-suited, by itself, to processing that involves the generation of random numbers. Contracts that require randomness as part of their processing are encouraged to rely on external data sources, known as *oracles*.

Contracts on the Ethereum network are written in a JavaScript-like programming language called Solidity that is *compiled* into bytecode for execution on the Ethereum Virtual Machine (EVM) that each full or light node operates. Using Solidity, authors can create contracts to perform a nearly-unlimited set of computations: working with stored data by creating it, reading it, or modifying it; performing calculations and responding with results; or managing transactions between other agents on the network. As in the object-oriented programming paradigm, contracts can inherit functionality or interfaces from other contracts, and they can interact with existing contracts on the network.

All contracts and user accounts on the Ethereum network have an *address*. Code is executed when a transaction is sent to a contract's address from another account (which may be a human or non-human user, or another contract): the transaction contains any data inputs the contract may require, as well as a fee to cover the *cost* of processing.

The reference to cost in the previous paragraph is of critical importance, because the distributed nature of the blockchain creates unique challenges for computational work. For example, a contract may (intentionally or otherwise) implement an infinite loop that, when propagated across the network, could exhaust all of the network's computing resources. In a less drastic vein, a contract could be designed to perform CPU-intensive data processing or write large amounts of data to the blockchain. Since every node in the Ethereum network must process every transaction, execute its associated code, and store its associated data, the resource usage for computations and storage can grow exponentially over time. For a private or enterprise-level implementation of Ethereum, where all network participants are known, this may not be a concern, but the global public blockchain must address it in order to ensure that malicious or ignorant participants cannot damage, cripple or destroy the system.

Gas

Ethereum addresses this problem on the public blockchain by assigning a cost to computation and storage. The mechanism used is referred to as *gas*, and the connection to the concept of fuel is intentional. Every contract on the blockchain is assigned a gas cost, based on the number of computational steps performed and the amount of data written to the blockchain (created or modified).[166] Good software design for the Ethereum platform minimizes this cost by limiting the amount of permanent data storage needed; in addition, code optimization is more of a concern for Ethereum than for other programming languages (like Python or JavaScript) since fewer computational steps make better use of network resources.

The gas cost for executing a contract is paid by the account that submits the transaction. Gas is included in the transaction call and is spent by the contract that receives the transaction. If the gas provided by the calling account is exhausted before computation is complete, processing halts and the transaction fails; leftover gas at the end of a transaction is refunded to the calling account.[167]

The presence and use of gas on the Ethereum network creates a market. A mining node that creates a block is rewarded with the sum of gas fees across all transactions in the

[166] Fees for each EVM operation are defined in Appendix G of the Ethereum specification. Data storage costs, in particular, are intentionally priced at a premium. See Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 24–25.

[167] When a transaction fails, the cost of gas may or may not be refunded, depending on the function that is called and how that function is designed. Non-refundable transactions may deter some forms of abuse.

block. As part of their operation, mining nodes advertise a minimum gas price (representing the price of their willingness or ability to process transactions) whose weighted average varies over time and is visible on the Ethereum network.[168] When a transaction request is made, creators have the option of specifying their own gas price; this ensures that the transaction will be processed in a timely fashion (or with high priority), since mining nodes are free to build a block from any of the available pool of transactions. Typically, mining nodes will select transactions that maximize their rewards. The variability of gas prices on the Ethereum network is intended to balance the resources available on the network with the demand for transaction processing: if computing resources are low, gas prices are likely to increase.[169]

Implications for contract design

The features of Ethereum, as well as its decentralized nature and its openness to participation by anyone with sufficient network resources, create some implications for software design that do not commonly apply to other programming paradigms. These considerations include the following:

- **Security**: since every contract deployed on the public blockchain is visible, designers must consider all possible ways contracts may be used or abused:

[168] Etherscan, "Ethereum Gas Price Tracker," Etherscan: The Ethereum Block Explorer, 2019, https://etherscan.io/gasTracker.

[169] A real-world example of this phenomenon was network congestion caused by an online collectable game called CryptoKitties. Demand for the game was so high that it significantly affected the performance of the Ethereum blockchain. In response, the average gas price for the Ethereum network reached an all-time high on January 6, 2018. See ConsenSys, "The Inside Story of the CryptoKitties Congestion Crisis," ConsenSys Media, February 20, 2018, https://media.consensys.net/the-inside-story-of-the-cryptokitties-congestion-crisis-499b35d119cc.

thus, they should hide access to internal data wherever possible. Designs should carefully consider which functions are publicly available, the data parameters used as input for those functions, and how those functions may be abused by attackers. In the case of a shared library catalogue system, access to record storage by a malicious agent could result in data vandalism or destruction.

- **Ownership**: Related to the preceding consideration, data managed by a contract is open to manipulation by anyone willing to pay for a transaction unless the contract is written to manage, enforce, or restrict access to that data. For example, a data record can track the address of the account that was used to create it, and any subsequent attempt to read or modify that data can be checked against ownership or access rules that the contract defines. Considerations of data ownership are therefore of paramount importance for security and system functionality: for example, can public participation in the evolution of shared library catalogue records be hindered by a lack of access to those records, or by a process that places intermediaries between participants and the catalogue record changes they wish to make?
- **Immutability**: Like any data written to the blockchain, a contract is unalterable once it has been compiled and deployed on the Ethereum platform: this is because all contracts' compiled code is written to the blockchain, just like Ethereum transaction and account data. The only fix for a bug in a contract is the creation and deployment of a replacement contract at a new Ethereum address; the same goes for applications that want to add

new features or altered functionality. All functions must be carefully-planned, and allowances should be made in contracts so that any hard-coded values or constants can be altered later (keeping in mind the concepts of security and ownership that have already been mentioned).[170]

Ethereum platform risks and challenges

No system is a panacea, and Ethereum is not unique in this regard. An Ethereum-based implementation of a shared library catalogue bears some challenges and risks, and those are outlined here.

The public Ethereum blockchain has become more popular over time, and issues of scalability have taken focus in project development. The current chain can process between seven and 15 transactions per second globally, and this limit can be reached by a single, popular decentralized application running on the chain.[171] The currently-proposed solution to this issue is called *sharding*;[172] it involves splitting the global network into as many as 1024 sub-networks that can process transactions separately while maintaining the integrity and verifiability of a single, master blockchain. The design of the feature has been combined with an effort to move Ethereum from a proof-of-work-based consensus system to a proof-of-stake-based system where nodes that process transactions must *stake*

---

[170] The Ethereum development community has created design patterns for upgradable contracts, allowing for bug fixes and feature evolutions to occur. Most models involve using a permanent contract on the front end that acts as a proxy. For some discussion of this topic, see Trail of Bits Blog, "Contract Upgrade Anti-Patterns," *Trail of Bits Blog* (blog), September 5, 2018, https://blog.trailofbits.com/2018/09/05/contract-upgrade-anti-patterns/.

[171] See footnote 169.

[172] James Ray, "Sharding Introduction R&D Compendium," Ethereum Wiki, June 13, 2018, https://github.com/ethereum/wiki/wiki/Sharding-introduction-R&D-compendium.

Ethereum tokens against their honest participation in mining activity. Since it is not yet implemented, this thesis does not take the sharding feature into account. While the implementation should not affect the function of decentralized applications, it does have potential impact for libraries that wish to participate in the system: in order to function as miners under the proof of stake model, which would enable revenue generation as part of their participation, those libraries would need to make an initial monetary investment in Ethereum tokens in order to place a stake.[173]

Any Ethereum project may be implemented as part of the general, public blockchain (where it is anchored by publicly-available nodes operating anywhere in the world) or using a private blockchain that might be operated entirely by libraries and other stakeholders. Benefits of a private blockchain include a higher level of security due to limited participation from unknown agents; a low cost of transaction processing, likely due to enterprise-grade implementation of high-power full nodes; and the ability to more directly control and guide the evolution of the system. On the downside, the sustainability and success of the network can only be guaranteed by widespread implementation of Ethereum nodes to operate the private chain, and the use of a private chain creates barriers to the inclusion of contributions from members of the public or other interested organizations. One risk associated with creating an open source shared cataloguing system is that it could be implemented by anyone, on any private or public blockchain. This could fracture the system and lead to variations in implementations that diverge over time.

[173] Current designs require a stake of 32 tokens (called Ether or ETH). Based on January 2019 values, this investment would add $6,400 to the cost of implementing an Ethereum node. The cost could be recovered if and when a stake is removed and subsequently sold at Ether's market price.

Decentralization implies giving up control of implementation, and there is no way to predict the eventual growth and ownership of the Ethereum network. This is especially critical for the process of mining, since control of block creation is determined by the ownership of the nodes that perform mining functions. A large corporation could effectively take charge of the network by deploying a large number of nodes or by partnering with other node owners. This risk is lower on the public blockchain, but for a private blockchain implementation it could leave the network prone to a takeover by any organization (including existing cataloguing service providers) that chooses to deploy a majority number of nodes. The nature of this risk adds more weight to the idea of relying on a public, rather than private, implementation of a shared catalogue system.

The previous risks lead naturally to a more abstract issue associated with the creation of open source, decentralized software systems. One of the apparent paradoxes of the decentralized software movement is that the desire to create decentralized systems requires relinquishing control over those systems' creation and evolution. Though Brafman and Beckstrom advocate for project champions and evangelists to help efforts thrive,[174] there is no way to guarantee the success or evolution path of any decentralized effort. While it is entirely possible that work on this system may follow directions un-envisioned by this thesis, this risk may be minimized by thoughtfully combining centralized project coordination with decentralized content creation, and by emphasizing the value of a shared catalogue as a public good.[175] eBay and Amazon use hybrid models (supplementing

---

[174] Brafman and Beckstrom, *The Starfish and the Spider*, chap. 5.

[175] This approach is recommended as part of a blockchain-based catalogue implementation project; see the "Recommendations" section in Chapter 6.

traditional corporate control with decentralized contributions from users in the form of ratings and reviews),[176] and the Ethereum project itself, while open to contributions from anyone, is guided by a nonprofit foundation that coordinates efforts, focuses design work, and is open to parallel, private implementations of Ethereum technology based on open standards.[177]

Finally, though the Ethereum network's underlying software is almost four years old, it is still in a phase of early and rapid evolution. Many of the challenges outlined in this section are being actively researched and addressed by Ethereum developers, and this means the design framework outlined in this section is prone to being affected by new Ethereum features that address scalability, security, functionality and network consensus mechanisms. Changes to Ethereum will be made while this thesis is being written, but catalogue design choices will be based on the state of the Ethereum framework as of November 2018.

Now that a functional definition of decentralization and the core features of the Ethereum framework have been outlined, we turn our attention to lessons gleaned from the study of open source movements and popular models of social information production (such as Wikipedia).

---

[176] Brafman and Beckstrom, *The Starfish and the Spider*, 162–67.

[177] Ethereum.org, "About the Ethereum Foundation," Ethereum.org, 2018, https://www.ethereum.org/foundation.

Open source software and social information production

The implementation of a blockchain-based shared cataloguing system touches on two aspects of contemporary knowledge production on networks: open source software methodologies and social information production systems. The construction and evolution of the cataloguing system's technical components is based on open source methodologies, and the operation of the system is governed by its ability to attract, retain and grow the community of participants who build and modify catalogue data. This section outlines work aimed at enumerating factors that may contribute to the viability of projects that use these models.

Eric Raymond's *Cathedral and the Bazaar*, in which the author self-identifies as an "accidental revolutionary,"[178] presents a foundational description of the principles of open source software development. Some of its tenets are reflected in contemporary Silicon Valley culture: notably, to release early and often, and to rely on software users for information (for example: discovery of bugs, issues with usability, and suggestions for additional functionality). However, Raymond's original essay was created in 1996 during the explosive early stages Internet growth; though he had the hindsight of 13 years of GNU General Public License (GPL) licensing,[179] he also presaged observations that would be made by later scholars. For example, he noted that the development of Internet-based

[178] Raymond, *The Cathedral and the Bazaar*, 2001.

[179] Li-Cheng Tai, "The History of the GNU General Public License," July 4, 2001, https://www.free-soft.org/gpl_history/.

communications tools allowed for distributed development[180] and that intangible rewards such as ego satisfaction compelled developers to participate in open source projects.[181]

It is important to note here that the decision to commit to open source methodologies does not, of itself, guarantee success in terms of adoption or software quality. Objective data on the success of open source projects is difficult to come by due to the subjective definition of *success*,[182] but two analyses in 2002 and 2009 agreed that only about 17% of open source projects survive long enough to reach a stable production release.[183]

Despite this, many open source software development and related projects (such as social information production projects that rely on similar, distributed contribution models) have been studied in order to abstract a set of parameters that contribute to sustainable adoption and growth. Analysis of successful projects like Linux, Apache, Mozilla, and Wikipedia have led to interesting conclusions about what factors contribute to the health and sustainability of freely-available software tools, and coalescence around a

---

[180] Raymond, *The Cathedral and the Bazaar*, 2001, 23.

[181] Raymond, 24.

[182] In one study, forty-five publications were reviewed to arrive at a five-factor taxonomy for open source project success that includes user interest and product-linked measures of project activity, effectiveness, efficiency and quality. See Amir Hossein Ghapanchi, Aybuke Aurum, and Graham Low, "A Taxonomy for Measuring the Success of Open Source Software Projects," *First Monday* 16, no. 8 (July 28, 2011), https://doi.org/10.5210/fm.v16i8.3558.

[183] Sandeep Krishnamurthy, "Cave or Community? An Empirical Examination of 100 Mature Open Source Projects," *First Monday* 7, no. 6 (2002): para. 4, https://firstmonday.org/ojs/index.php/fm/article/view/1477/1392; Rich Gordon, "Six Things to Know about Successful Open-Source Software," Northwestern University Knight Lab, July 24, 2013, para. 10, https://knightlab.northwestern.edu/2013/07/24/six-lessons-on-success-and-failure-for-open-source-software/; Charles M Schweik and Robert C English, *Internet Success: A Study of Open-Source Software Commons* (Cambridge, MA: MIT Press, 2012), chap. 8.

set of design levers derived by Yochai Benkler that serve as part of the framework for developing features for a blockchain-based shared cataloguing system.

Benkler's model is applicable to this project because it derives its components from the open source software development paradigm,[184] the emergence of distributed computational systems like SETI@Home[185] and peer-to-peer file sharing networks,[186] and the explosion of social information production systems like the NASA Clickworkers project and Wikipedia.[187] Moreover, aspects of the model align strongly with the work of other scholars. The model's components, and those alignments, are outlined in the next seven sections.

Communication

While it is unsurprising that communication between collaborators can contribute to success on a project, its necessity is even more pronounced in networked environments where collaborators cannot see each other face to face. Conway's Law, coined in 1967 by programmer Melvin Conway, states that any organization created to design a system is "constrained to produce designs which are copies of the communication structures" of the

---

184 Benkler, "Coase's Penguin, or, Linux and 'The Nature of the Firm.'"

185 Though SETI@Home, created in 1999, is one of the more well-known early "open" distributed computing projects, it was not the first. The original "volunteer computing" project was The Great Internet Mersenne Prime Search (GIMPS), started in 1996. See BOINC, "Volunteer Computing," 2014, https://boinc.berkeley.edu/trac/wiki/VolunteerComputing.

186 Benkler, *The Wealth of Networks*, 81–85.

187 Benkler, "Coase's Penguin, or, Linux and 'The Nature of the Firm,'" 10; Benkler, *The Wealth of Networks*, 70–72; Yochai Benkler, *The Penguin and the Leviathan: The Triumph of Cooperation over Self-Interest*, 1st ed (New York, NY: Crown Business, 2011), 12.

group.[188] Its implications seem intuitive: if multiple developers are working on a software module and do not have the ability to communicate with one another about the design, the end result will be multiple, parallel instances of similar software.

As a response to Conway's law, and in addition to noting that "given enough eyeballs, all bugs are shallow,"[189] Raymond points out that collaborative software development requires a scalable communications mechanism.[190] Benkler extends this principle to social information production practices in general, asserting that the tools used to mediate communication and collaboration must include the means to facilitate participation and defend the goals of the common effort.[191] This is especially important in an online context, since studies have shown that physical contact among collaborators is typically key to the development of trust and social norms.[192]

In *The Penguin and the Leviathan*, Benkler refers to Wikipedia as a specific example of a scalable communications system in an online context: the collaborative mechanisms of "talk" pages and edit summaries for each article eases the process of arriving at consensus over how knowledge should be represented on Wikipedia. Analysis tools like *Contropedia* have illustrated the value of this feature by illuminating specific cases where inter-

---

[188] Melvin Conway, "How Do Committees Invent?," *Datamation* 14, no. 4 (1968): 28–31.

[189] Eric S. Raymond, "The Cathedral and the Bazaar," May 5, 2000, 6, https://web.archive.org/web/19990224193551/http://www.tuxedo.org:80/~esr/writings/cathedral-bazaar/cathedral-bazaar.ps.

[190] Raymond, 11.

[191] Benkler, "Coase's Penguin, or, Linux and 'The Nature of the Firm,'" 17.

[192] Brafman and Beckstrom, *The Starfish and the Spider*, 90; Benkler, *The Penguin and the Leviathan*, 103.

collaborator communication has been used to discuss and resolve issues with contentious Wikipedia articles.[193]

Framing, fit and authenticity

Benkler bases his *framing* design lever on behavioural economics work pioneered by Amos Tversky and Daniel Kahneman, whose research has showed that people's choices are influenced by how questions and situations are framed.[194] Tell a cataloguing librarian that they must complete entries to meet a daily quota, and they may resent the work; tell them that the completion of high-quality catalogue entries empowers people to satisfy their curiosity by locating information using the catalogue, and they may be more intrinsically motivated to complete the task.

The work of behavioural economists has been employed in many fields—notably, in service of the media of persuasion (marketing and advertising). At one end of the spectrum, the tools developed by this research can be used to *nudge* people into morally- or ethically-accepted behaviour, like filing taxes on time.[195] On the other end, they can be used to manipulate people into performing actions they might not otherwise do. This aspect of

---

[193] "Contropedia," accessed November 23, 2018, http://contropedia.net/; Erik Borra et al., "Societal Controversies in Wikipedia Articles," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea: ACM, 2015), 193–96, https://doi.org/10.1145/2702123.2702436; Fabian Flock et al., "Towards Better Visual Tools for Exploring Wikipedia Article Development --- The Use Case of Gamergate Controversy," in *AAAI Workshop - Technical Report* (Papers from the 2015 ICWSM Workshop, Oxford, UK: AI Access Foundation, 2015), 48–55, https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/viewFile/10656/10561.

[194] Benkler, *The Wealth of Networks*, 44–54.

[195] Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Rev. and expanded ed (New York, NY: Penguin Books, 2009).

framing connects with Benkler's related concept of *authenticity*: if the frame you choose for your system is perceived as being false or manipulative, it will fail.

Together, these two aspects contribute to the third: *fit*. This is described as the suitability of the work being done to the people who are doing the work. Though skill and ability are significant factors in this area, it also relates to the time available to perform the work and the granularity of the work (for example, the creation of full catalogue records versus making small changes to existing ones). The ability for participants to self-identify as "the best person for the job" serves as an intrinsic motivator; so, too, does the belief that participants are able to participate in the work autonomously.[196]

There is a special synergy for this lever that can be seen at play within the field of librarianship: the profession is ethically-grounded in notions of information accessibility, making the work of cataloguing essential to the practice. Education in librarianship includes an orientation to the ethics and culture of the field, which builds in an affinity for the framing device that a shared cataloguing system might use.

This lever is somewhat intrinsic to the design of a shared cataloguing system, since it partially explains the "why" of its existence. Economist Glen Weyl is skeptical about the current blockchain paradigm[197] but believes that a successful decentralized system can

---

[196] Autonomy over work tasks is discussed in-depth as an intrinsic motivator in Daniel H. Pink, *Drive: The Surprising Truth about What Motivates Us* (New York, NY: Riverhead, 2011). The high reliability of content created by infrequent, anonymous contributors, termed Good Samaritans, is also noted in Denise Anthony, Sean W. Smith, and Timothy Williamson, "Reputation and Reliability in Collective Goods: The Case of the Online Encyclopedia Wikipedia," *Rationality and Society* 21, no. 3 (August 2009): 283–306, https://doi.org/10.1177/1043463109336804.

[197] Weyl's two main criticisms are as follows: first, that blockchains are primarily used to create single, immutable information objects that mimic existing models of information commodification; and second, that the blockchain paradigm's obsession with anonymity and pseudo-anonymity make equitable governance and decision-making nearly impossible. For more detail, see Brian Fabian Crain, "Glen Weyl:

succeed without external regulation if it seen by participants as a one that serves the public good.[198] Authentic framing of the role and purpose of a decentralized, shared cataloguing system can contribute to this motivation.

Empathy and Solidarity

Benkler emphasizes the role of *empathy* and *solidarity* in the sustainability of distributed information production processes.[199] The two terms are contrasted as follows:

- **empathy**, as defined by Martin Hoffman[200] and Nancy Eisenberg,[201] is the set of responses (emotional and cognitive) that identify the emotional state of another person and replicate it internally for the participant.
- **solidarity** is the attachment to a group, or the sense of "us," that encourages us to work for the benefit of the group instead of in our own self-interest.

Solidarity, like fit, is somewhat in-built to the culture of librarianship, which is partly established by the accreditation and education systems used to forge librarians and other professionals who perform cataloguing work. These practitioners are generally aware of and are well-versed in the core values of librarianship, which include access, diversity,

---

Radical Markets – Uprooting Capitalism and Democracy for a Just Society," accessed September 5, 2018, https://soundcloud.com/epicenterbitcoin/eb-251.

[198] Crain.

[199] Benkler, *The Wealth of Networks*, 82–94.

[200] Martin L. Hoffman, *Empathy and Moral Development: Implications for Caring and Justice* (Cambridge, MA: Cambridge University Press, 2007), 29–30.

[201] Nancy Eisenberg and Janet Strayer, eds., *Empathy and Its Development*, Cambridge Studies in Social and Emotional Development (Cambridge, NY: Cambridge University Press, 1987), chap. 1.

service, social responsibility, and the public good.[202] The American Library Association (ALA) states that these values "define, inform and guide all professional practice."[203]

A challenge here is that there is no reason to believe that a distributed catalogue system will promote these values better than existing systems. Here, the lever of framing comes in to play: a new model for a shared catalogue that includes more direct input from librarians and patrons demonstrates a shift in activity towards the ALA's core values, removing intermediaries in current systems. These intermediaries include centrally-controlled authorities like OCLC who operate outside the absolute authority of the local institutional environment, whose employees may not be accredited library practitioners,[204] and who have the final say on changes made to global catalogue records.

Moral systems

The concept of moral systems, as defined by Benkler, consists of three components: *fairness*, *morality*, and *social norms*. Underpinning all three pieces is the belief that people, when given the opportunity, generally prefer to cooperate towards a positive end. This belief is backed by research: studies have shown that the majority of people are motivated (at least in part) by their subjective sense of fairness and ethics. Truly self-interested individuals—who are generally unaffected by moral systems—make up about 30% of the population,[205] and this segment is addressed in the next section.

---

[202] American Library Association, "Core Values of Librarianship."

[203] American Library Association, para. 1.

[204] Beall, "OCLC: A Review."

[205] Benkler, *The Penguin and the Leviathan*, 13–14; Samuel Bowles, *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*, The Castle Lectures in Ethics, Politics, and Economics (New Haven, CT; London, UK: Yale University Press, 2016), 41–44.

The subjective nature of fairness makes it a difficult concept to manipulate with incentives, but Benkler suggests at least one area where fairness can be built into a system. Contextual expectations play a significant role in the perception of fairness: for example, while people may agree that those with high incomes should be taxed at a high rate, the same people would likely disagree that lottery winnings should be similarly taxed.[206] Fairness is also perceived as merit-based, especially in corporate environments; here again, however, Benkler points out that merit-centered models may work or not work, depending on how and where they are implemented. The only predictive mechanism for implementing fairness is connected to the fairness of *processes*: where they are transparent and deemed to be fair, the outcome of the processes is also deemed fair, even if outcomes are inconsistent. In other words, says Benkler, fairness is strongly connected with *intention*.[207]

The role of social norms in collaborative and competitive systems has been studied in depth and has been shown to play a much stronger role than predicted in the *homo economicus* model of rational human actors.[208] Summarizing the motivation to adhere to social norms under the term *social preferences*, Samuel Bowles includes factors like aversion to inequity, intrinsic pleasure derived from helping others, ethical commitments, and reciprocity into a multi-variate system that causes people to adhere to social norms

---

[206] Benkler, *The Penguin and the Leviathan*, 126.

[207] Benkler, 133–40.

[208] Foundational studies are summarized in Bowles, *The Moral Economy*, 33–35; Benkler, *The Penguin and the Leviathan*, 145–48.

even at personal cost.[209] The power of social norms over decision making suggests that we should create signals for what counts as appropriate behavior, but economic modeling by Bowles illustrates an important caveat. The creation of a signal reveals something about the intention of the person who created it, which is connected with the perception of fairness: if the signal or incentive is interpreted negatively, it can backfire.[210] It is for this reason that Benkler intentionally chooses the word "signal"[211] over "incentive:" incentives tend to be transparent and political, and are seen by participants as a means of exercising control. The introduction of incentives in situations where collaborators already enjoy the work they are doing may render the desire for a controlled outcome visible and, as a result of removing participants' autonomy, create a negative response.[212]

The power of social norms, and the inability to predictably define them, appears to create a paradox for system design. However, scholars including Elinor Ostrom have illustrated that collaborators are more willing to adhere to norms when they have had input into their creation.[213] Accordingly, Benkler suggests that the path to accepted social norms lies in giving participants input or control over their construction.[214] This sentiment

---

[209] Bowles, *The Moral Economy*, 45.

[210] Bowles, 86–89.

[211] Benkler, *The Penguin and the Leviathan*, 145.

[212] Bowles, *The Moral Economy*, 96–99; Richard M. Ryan and Edward L. Deci, "Self-Regulation and the Problem of Human Autonomy: Does Psychology Need Choice, Self-Determination, and Will?," *Journal of Personality* 74, no. 6 (December 2006): 1580, https://doi.org/10.1111/j.1467-6494.2006.00420.x.

[213] Elinor Ostrom, James Walker, and Roy Gardner, "Covenants with and without a Sword: Self-Governance Is Possible.," *American Political Science Review* 86, no. 02 (June 1992): 412–14, https://doi.org/10.2307/1964229.

[214] Benkler, *The Penguin and the Leviathan*, 158.

is reflected by Brafman and Beckstrom, who specify shared ideology as one of the five key success factors for decentralized systems.[215]

Reward and punishment

This factor falls squarely under the authority of incentives, which has already been discussed as a mechanism that has the ability to backfire. The primary goal of this design lever is to address the 30% of the population whose self-interest renders them most responsive to extrinsic incentives. The challenge with implementing reward and punishment systems for social collaboration tools is to devise a system that motivates self-interested actors without negatively-affecting intrinsically-motivated ones.

The literature outlines some conditions under which basic rewards and punishments may work. "Carrot and stick" incentives work well for small, easy-to-complete tasks where time and energy investments are minimal.[216] As suggested in previous sections, the framing used for an incentive plays a role in its acceptance: terms like *bribe*, *bonus*, *prize*, *fine*, and *punishment* each have social connotations that may influence their palatability.[217] Here we see a connection to the earlier-discussed factor of authenticity: when the framing is genuine and is aligned with the vision for a project, it may succeed more universally.

Finally, it should be noted that reward and punishment systems need not be universally-applied. Approximately half of the programmers who work on open source

---

[215] Brafman and Beckstrom, *The Starfish and the Spider*, 95.

[216] Benkler, *The Penguin and the Leviathan*, 171–72.

[217] Bowles, *The Moral Economy*, 96–97.

projects claim that they are being paid for their work. At first this appears inequitable and unfair, yet it is a non-damaging condition for successful open source projects like Red Hat Linux and Apache. The combination of monetary and non-monetary rewards in the same system can be seen as fair simply because there is no universal definition of fairness.[218]

Reputation, transparency and reciprocity

One of the most powerful motivators for participation in open source software and social information systems is *reputation*. This was pointed out by Raymond, regarding the widespread participation in Linux development: "the 'utility function' Linux hackers are maximizing is not classically economic, but is the intangible of their own ego satisfaction and reputation among other hackers."[219] The reputation of Linux developers was established through participation in online forums and visible contributions to code, and reputational systems have been shown to drive participation in communities like Wikipedia,[220] but modern social production systems have also embraced karma-based mechanisms where points are awarded on a peer-to-peer basis. Benkler refers to Slashdot in his earliest example,[221] but many modern communications platforms have embraced the idea. Wikipedia and Reddit both establish credibility by displaying counts of each user's posting and editing history and social networks like Instagram and Twitter establish

---

[218] Benkler, *The Penguin and the Leviathan*, 181–82.

[219] Raymond, "The Cathedral and the Bazaar," May 5, 2000, 16.

[220] Anthony, Smith, and Williamson, "Reputation and Reliability in Collective Goods," 295–98.

[221] Benkler, "Coase's Penguin, or, Linux and 'The Nature of the Firm,'" 13–16.

authority through follower counts; in addition, Reddit allows posts and comments to be upvoted and downvoted by all other registered users.

These examples lead naturally to the second component of this factor: *transparency*. The act of rendering activity visible makes it easier for others to validate the authenticity of contributions and serves as a form of communication between project participants. Raymond speculates that the visibility of processes may be a driver for open source quality.[222]

The third component, *reciprocity*, is enabled by transparency. One benefit of contributing freely to a collaborative project is that it earns the contributor a right to profit from the output of the project at a later date. Benkler explores this through an examination of the Couch Surfing movement, whose success is credited with the idea that, eventually, everyone needs a place to crash when traveling.[223]

Build for diversity

The term *diversity* in this context refers to diversity of contributions. It has always struck me as odd that OCLC's expectation for contributions to library catalogues would be fully-reciprocal: policies put forth by the organization in the 1980s and 1990s were based on the idea that a library would contribute a record for every record used.[224] For most information production systems, the contributions of individual entities are governed by a power law, where a small number of actors contribute the largest portion of content, with a

---

[222] Raymond, "The Cathedral and the Bazaar," May 5, 2000, 23.

[223] Benkler, *The Penguin and the Leviathan*, 114.

[224] Houk, "OCLC Speaks out on Record Nabbing"; Brown, "OCLC, Copyright, and Access to Information: Some Thoughts."

largest number of participants contributing in small amounts. Benkler points out that the nature of peer production allows participants to self-identify for jobs for which their skills can be readily applied;[225] moreover, each participant responds to a subjective set of drives and motivational signals. Accordingly, he advocates that social information production systems must be constructed to allow for a diverse range of participation.[226]

The framework versus the modern shared catalogue

This section has knit together elements of behavioural economics, lessons from the open source movement, a three-faceted model for decentralization, and technical details of the Ethereum implementation to create a context for the reinvention of the engine of the shared library catalogue.

A reinvigoration of Ranganathan's first law, "books are for use,"[227] suggests that a new vision for the library catalogue should centre on the idea of openness, and the framework presented here embraces that concept. Political decentralization, as defined in this chapter, necessitates the relinquishment of control: though it has not yet explicitly been stated, this implies that any decision to participate in the sustainability and the growth of the cataloguing system becomes voluntary. Accordingly, the choice of technologies for system design must be paired with a set of incentives to motivate participation, and the framework presented here includes both. For example, the decision

---

[225] Benkler, *The Penguin and the Leviathan*, 5.

[226] Benkler, 239.

[227] See page 7 in chapter 2.

to use open source software implies the adoption of its associated paradigms, which are also reflected in Benkler's *transparency* design lever (transparency is also one of the definitions of openness as articulated by Pomerantz and Peek[228]) as a foundational value.

The principles of architectural, political and logical decentralization were used as an introduction to the design framework presented in this chapter, and their maximization serves as a guiding principle for a system design that addresses this thesis' research questions. Decentralization may enable the removal of intermediaries among the catalogue, cataloguers, and users, addressing the modern catalogue's incompatibilities with Ranganathan's fourth law and implicitly reducing the problem of centralization as described in the discussion of his fifth law. The introduction of a scalable communications system, as suggested by Benkler, enables discourse about the content and quality of cataloguing records to occur, tipping the hat to Linux founder Linus Torvald's assertion that "given enough eyeballs, all bugs are shallow"[229] and placing a focus on the values asserted in Ranganathan's second and third laws.

In summary, then, it appears that a decentralized approach to shared cataloguing, grounded in the principles of openness, has the potential to address many of the challenges of contemporary shared cataloguing practices. The ability of the design framework to achieve this potential is dependent on the design of such a system, and that design is the focus of the next chapter.

[228] Pomerantz and Peek, "Fifty Shades of Open," paras. 42–45.

[229] Raymond, *The Cathedral and the Bazaar*, 2001, 6.

5. The Catagora Software Design

Introduction

Having outlined a framework for the creation of a decentralized, blockchain-based shared cataloguing system with a focus on openness, attention is now turned to the creation of a conceptual design built on the Ethereum platform. The solution outlined in this chapter is referred to as "Catagora," which is a portmanteau of *catalogue* and *agora*, the Greek word for a place of congregation. The name functions as shorthand for the phrase "decentralized, blockchain-based shared cataloguing system" but also suggests that the library catalogue can serve as an open space where discussions about resource descriptions occur accessibly for all catalogue users, whether they are library professionals or members of the public.

This chapter is divided into three sections. The first section addresses some of the unique characteristics of the Ethereum blockchain platform to arrive at decisions that impact the architecture and implementation of the system. The second section outlines the high-level architecture of the system and traces the primary use cases for interactions between contributors and the catalogue. The final section discusses some unique features of Catagora that arise from combining the blockchain paradigm with lessons from the open source software movement and other social information production projects.

Constraints imposed by Ethereum

File Storage

The cost of file storage on the Ethereum blockchain is prohibitively high, by design, to create barriers to using the platform as a data warehouse. Based on Ethereum's yellow paper,[230] the price of one kilobyte of storage is 0.032 ETH, making one megabyte of data worth 32.768 ETH.[231] At Ethereum prices as of January 2019 ($165 CAD per ETH), one megabyte of storage would cost $5406.72. To put this in the context of catalogue data, a recent copy of the University of Toronto's catalogue consists of 5.2 gigabytes of MARC records, representing nearly 14 million unique records and more than $921 million in storage costs.[232]

Economic considerations clearly suggest there is little value in storing complete catalogue data on the Ethereum blockchain. To implement Catagora, then, it is necessary to link the blockchain's control and access mechanisms to a reliable, external, decentralized data storage solution. The Interplanetary File System (IPFS) is establishing itself as a peer-to-peer, distributed, open source option that can address needs of this kind.[233] The system builds on BitTorrent's file-sharing protocol design to implement an infrastructure-style platform for efficient, low-latency, distributed data storage. The major benefit of the design

[230] Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 24–25.

[231] 1Kosmos, "Costs Of Storing Data On The Blockchain," 1Kosmos BlockID, May 1, 2018, https://onekosmos.com/blog/cost-of-storing-data-on-the-blockchain/.

[232] University of Toronto, "University of Toronto Catalog," Internet Archive, February 11, 2019, https://archive.org/details/marc_university_of_toronto.

[233] Juan Benet, "IPFS - Content Addressed, Versioned, P2P File System," April 1, 2015, https://github.com/ipfs/papers/raw/master/ipfs-cap2pfs/ipfs-p2p-file-system.pdf.

is that it does not require all network participants to maintain local copies of all data. Each file in the IPFS storage space is addressable by its hash value and is accessible to anyone using an IPFS gateway: this can be through locally-installed IPFS software, or remotely-accessible servers using Hypertext Transfer Protocol (HTTP) connections.

To integrate IPFS with Ethereum and save costs of data storage, Solidity contracts are designed to store hashes of data files (which serve as pointers to files available on IPFS) rather than the files themselves. These hash values occupy 32 bytes each and are far more economical to store on the Ethereum blockchain than complete data files: for example, one megabyte of data contains pointers to as many as 15, 625 records.[234] In the case of the University of Toronto's full catalogue, this would reduce the required storage size from 5.2 gigabytes to 896 megabytes.[235]

Every edit to an IPFS file results in the creation of a new hash (and a new file), and the location of the most current version of a catalogue record can be recorded by updating a previously-stored hash value to a new one. Storing an updatable list of each catalogue record's previous hash values alongside the current hash creates the basis of an immutable, authoritative record of each catalogue entry's creation and history. This approach generates an implicit trail of data that traces the evolution of any catalogue record over time.

---

[234] Though 31, 250 32-byte records can be stored in one megabyte of data, it will be shown later in this chapter that the Catagora design requires two records per catalogue entry in order to track community discourse related to each entry.

[235] Though this still represents a total storage cost of $4.8 million dollars, the University of Toronto example is a stand-in for the value of a single, global, shared catalogue whose costs would be distributed across all participants in a distributed cataloguing system. As will be discussed in the "Key Design Characteristics" section, the cost of storing a single, new catalogue entry is approximately 51 cents.

In addition to providing file storage mechanisms, IPFS provides access to its data using local gateways and the HTTP protocol. This facilitates web-based access to all files extant on the IPFS system and also creates the opportunity for IPFS to host the HTML, JavaScript and associated files required to generate Catagora's user interface. Taking advantage of this feature means that the catalogue's web-based front end need not be served from a single, centralized location. This contributes to the overall architectural decentralization of the system.

IPFS, like Ethereum, is an evolving technology with planned features that have not yet been implemented. The core technology is stable, however, and is suitable for the creation of a Catagora prototype. Regardless, the decision to shift complete catalogue data from the Ethereum platform to a separate file system has other implications for the catalogue's software design. The following two points are significant:

1. Since Ethereum only stores hashes of files under this paradigm, it is unaware of the contents or contexts of its catalogue data files. As a result, the role of enforcing file format and metadata standards must be done "off-chain," likely as part of the code that manages Catagora's web-based front end. Sadly, this creates a barrier for downstream developers or users who wish to develop customized or modified client interfaces to the catalogue: unless this function is further-abstracted as part of the software design, the logic for managing compliance with metadata standards needs to be implemented as part of the client access component.
2. Since IPFS itself uses only hash values to reference files, it is difficult to search IPFS data. This poses a significant challenge for a library catalogue

that must be addressed in one of two ways: either create a separate index file of relevant metadata fields that can be used to facilitate searches, or use another layer of software, such as OrbitDB,[236] that stores IPFS files through an API-based access layer to provide query functionality. The former solution obviates the purpose of a catalogue as a means of information retrieval, so the latter solution is preferable.

Though a complete implementation should take these challenges into rigorous consideration, a Catagora prototype may rely on a simplified, JavaScript Object Notation (JSON) format for metadata storage and need not implement the OrbitDB layer for database functionality. Accordingly, a prototype would allow users to browse a small catalogue but would not provide a robust search interface.

User Interface

The cryptographic methods embedded in the Ethereum framework require that all user-agent interactions with the blockchain system be digitally-signed for transaction processing to occur. This must only take place when information on the chain is being created or edited (such as when someone creates a new catalogue entry, makes a revision to an existing one, or updates their user profile information: in other words, when the state of the blockchain data is actually affected). The client interaction model is as follows:

1. When data on the blockchain will be affected by a user action, the client interface builds a transaction that bundles together all of the associated

---

[236] haadcode, *Peer-to-Peer Databases for the Decentralized Web.*, JavaScript (2015; repr., OrbitDB, 2019), https://github.com/orbitdb/orbit-db.

Solidity contract function's required inputs. The transaction takes the form of a data structure incorporating input data, the Ethereum addresses of the sender and recipient, and the available amount of gas for the transaction.[237]

2. Before transmitting the application to the Ethereum network, the user is asked to sign the data package using their private encryption key. Once the transaction is signed, it is sent to the network.
3. The client interface waits for the Ethereum network to issue event notifications about state changes, indicating that the transaction has been processed. These events are based on definitions created in the Solidity contract.[238] Event notifications signify that changes to blockchain data have been made, triggering the user interface to update its information or take some other action.

Since private encryption keys must not be shared, the creation of digital signatures should ideally happen on the device the cataloguer is using. For web-based interfaces, the *de facto* programming language for this purpose is JavaScript, since it can be executed entirely within the cataloguer's web browser, on their local device.

JavaScript Libraries

For the sake of usability, a number of Ethereum *wallet* applications have been created to assist with the creation and management of transactions on the Ethereum

---

[237] Transactions also include other information such as nonce values (which are used to increase the cryptographic strength of the encrypted transaction). Full details on the anatomy of Ethereum transactions are provided in Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 4–5.

[238] "Contracts: Events," Solidity 0.5.3 documentation, February 11, 2019, https://solidity.readthedocs.io/en/v0.5.3/contracts.html#events.

network. A popular, open source tool that integrates with popular web browsers is **MetaMask**.[239] This utility lets people create Ethereum accounts (and associated addresses) with relative ease, can automatically detect web pages that are linked to the Ethereum network, and (once unlocked by users with a private key, used as a password) is able to detect and prompt users to sign Ethereum transactions with the press of a button and without leaving the web browser.

The most mature frameworks for interacting with Ethereum contracts have also been built in JavaScript, making the language ideal to use for the implementation of Catagora's web-based front end. The Ethereum ecosystem is currently mature enough that freely-available and open source JavaScript libraries exist for all of the catalogue's basic functions. These include the following modules:

- **Web3**[240] interacts with the MetaMask browser extension and has the ability to interface with the Ethereum network using Remote Procedure Call (RPC) connections. The library also simplifies the process of building, preparing, and signing Ethereum transactions.
- **Js-ipfs**[241] lets a JavaScript application interact with the Interplanetary File System (IPFS), including reading and writing data. The file hashes created as

[239] MetaMask, "MetaMask," accessed February 4, 2019, https://metamask.io/.

[240] *Ethereum JavaScript API*, JavaScript (2014; repr., ethereum, 2019), https://github.com/ethereum/web3.js.

[241] *IPFS Implementation in JavaScript*, JavaScript (2014; repr., IPFS, 2019), https://github.com/ipfs/js-ipfs.

a result of writing a file to IPFS can then be stored on the Ethereum network using Web3 transactions.

- **Orbit-db**[242] interfaces with the OrbitDB system, which is built on top of IPFS. This library provides a JavaScript interface to OrbitDB functions, enabling IPFS data search and retrieval. The OrbitDB platform provides an abstraction layer for IPFS and can serve as the conduit for reading from or writing data to the distributed file system.

In order to create a viable Catagora prototype, JavaScript must be used to bridge the client interface, the IPFS file system, and the Ethereum blockchain. Though the *orbit-db* library would be critical to a complete implementation, due to its ability to provide key-based search functionality, the *js-ipfs* library would suffice for the scope of a simple prototype implementation.

System Architecture

Data Structures

The back end of Catagora is governed by two basic high-level data structures: one for catalogue entries and one for user/cataloguer accounts. This section focuses on data stored on the Ethereum blockchain; as noted in the previous section, full catalogue record data and associated talk pages will be stored in IPFS and are not discussed here. Metadata and catalogue record formats are outside the scope of this thesis; while a complete

[242] *Peer-to-Peer Databases for the Decentralized Web.*, JavaScript (2015; repr., OrbitDB, 2019), https://github.com/orbitdb/orbit-db.

implementation might employ a modern standard like BIBFRAME,[243] a prototype could store records in a simple JSON format using key-value pairs, with talk pages stored as markdown files. The Ethereum data structures are outlined in Figure 3 (p. 92).

Catalogue entries

A catalogue entry has three components: the catalogue record itself, information about revisions that have been made to the catalogue record, and any inter-cataloguer discussion related to that record. To store this data, the "card" in the catalogue points to the most current IPFS hashes for the resource description entry and its associated talk page. In addition, the record stores a pointer to a list of edits (in the form of "edit records") that have been made to each of these two components. The structure of the edit record is outlined in the next section.

Edit record

A record of a specific edit consists of the address of the user/cataloguer, the date and time at which the edit was included in an Ethereum block, the content of the "edit summary" field and a link (consisting of an IPFS hash value) to the version of the record associated with the edit. This is a standardized format that can be used to track revisions of catalogue entries or their associated talk pages.

[243] Library of Congress, "BIBFRAME - Bibliographic Framework Initiative."

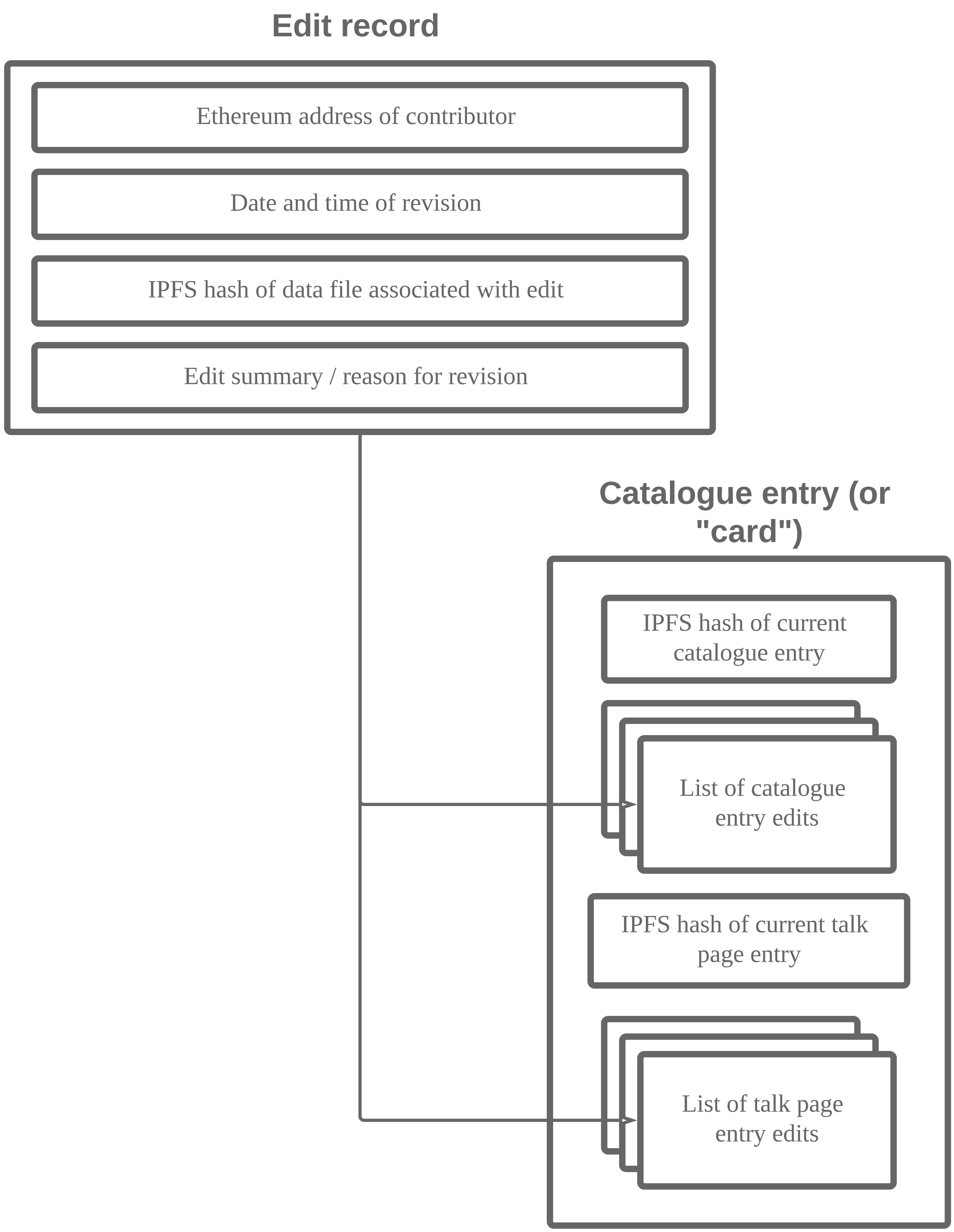

Figure 3: Key Catagora data structures

These records can be linked together in the form of an array. As revisions to a catalogue entry occur, new edit records are added to the existing *stack* of records for a content element, creating an ordinal record of each entry's revisions. This provides the ability to trace the evolution of a catalogue "card," including contributions to the catalogue record and its associated discussion. As a result, a "card" retrieved from the catalogue can also display references to its full history.

Cataloguer profiles

Each Ethereum account that participates in the editing, creation or evaluation of Catagora records has the option of creating a user profile to complement other data that is automatically tracked by the system. Elements of the user profile include a name and other basic profile information (such as contact information, institutional affiliations, and optional social media links).[244] In addition to the user-definable fields, several counters are used to provide the foundation for participation incentives and a reputation scoring system:

- A ***participation*** value is used to track interactions by a cataloguer that support the system. Points are given for up- or down-voting other users' contributions, for editing an existing catalogue entry, or for participating in a discussion related to a catalogue entry.

[244] The usefulness of institutional affiliation information is dependent upon its validity. Though the verification of user identities and affiliations is briefly addressed in "Vandalism detection and content monitoring" on page 139, its in-depth specification is beyond the scope of this thesis.

- a ***cards*** value is used to track the number of new catalogue records created by each cataloguer
- a ***user reputation score*** value is based on votes from other Catagora users, who increment or decrement a score by up- or down-voting contributions made by this cataloguer. User reputation is a social incentive, designed to encourage and reward contributions to the catalogue.
- A ***content reputation score*** value is algorithmically-determined, based on the durability of catalogue contributions made by a user, and is intended to serve as the basis for content moderation and vandalism detection systems.[245] The model for this score is drawn from work by Adler and de Alfaro, who showed a correlation between the durability of Wikipedia content edits (i.e. content that is not quickly-altered by other users and remains intact after a number of subsequent revisions) and the quality of those contributions.[246]

[245] As will be discussed in Chapter 6, formal content moderation mechanisms are beyond the scope of this thesis. The introduction of a field for content reputation score is an acknowledgment that there is a need for abuse management mechanisms in a full Catagora implementation. A method for determining a content score is suggested (see pages 139-140, which describe the approach outlined by Adler and de Alfero), but this design intentionally avoids *prescribing* how such a score should be used to safeguard catalogue data.

[246] B. Thomas Adler and Luca de Alfaro, “A Content-Driven Reputation System for the Wikipedia,” in *Proceedings of the 16th International Conference on World Wide Web - WWW ’07* (the 16th international conference, Banff, Alberta, Canada: ACM Press, 2007), 261, https://doi.org/10.1145/1242572.1242608.

System Requirements

Though it is possible for the web-based front end of Catagora to interact with existing IPFS and Ethereum nodes over HTTP or RPC connections, allowing for a "light" implementation that accesses existing infrastructure, it is suggested that a library-based Catagora deployment be built on hardware capable of running all of the system's main components. Though the system is designed to be publicly accessible to anyone with web access, its sustainable implementation will rely on the willingness of libraries and other institutions to install and configure *nodes* that support all aspects of Catagora's architecture.

It is reasonable to assume that all of Catagora's pieces can be set up on a single server or virtual machine, provided data storage and network bandwidth requirements can be met. A complete installation should include:

1. A full Ethereum node, whose basic hardware requirements are constrained by network bandwidth and data storage requirements.[247] CPU power for a full node need not be extreme, and memory requirements are not excessive, but the growing size of the Ethereum blockchain and the frequency of new block creation suggest that file input and output processing create the most significant performance bottleneck. Accordingly, an array of larger Solid-State Drives (SSD) should be employed for good system performance. As of

[247] Albert Palau, "Analyzing the Hardware Requirements to Be an Ethereum Full Validated Node," *Medium* (blog), September 24, 2018, https://medium.com/coinmonks/analyzing-the-hardware-requirements-to-be-an-ethereum-full-validated-node-dc064f167902.

January 2019, the Ethereum blockchain consists of about 125GB of data,[248] and upcoming developments such as pruning and sharding[249] will reduce the storage requirements for full nodes. Recent discussions of the cost of running full Ethereum nodes estimate the price of a local installation at about $30 USD per month. Hosted, cloud-based alternatives are also possible at costs of up to $100 USD per month.[250]

2. A full IPFS node, which is simply an installation of the IPFS software with the IPFS *daemon* enabled. As noted earlier in this chapter, an IPFS node does not store a full copy of all IPFS data. By default, nodes only store locally-created files and an internally-managed set of files that are often-requested by near-neighbours on the network.[251] The node then queries the *swarm* (the term used for the full IPFS node network) for any other data it needs. Data accessed over the network is subsequently stored locally but is automatically removed after a period of non-use. Records created on the local system are likely to be stored locally, as well as any records that are frequently-accessed. Catalogues like that of the University of Toronto consist of less than six

---

[248] Etherscan, "Ethereum ChainData Size Growth - Fast Sync," Etherscan, accessed February 4, 2019, https://etherscan.io/chart2/chaindatasizefast.

[249] Péter Szilágyi, "Geth 1.8 – Iceberg," Ethereum Blog, February 14, 2018, https://blog.ethereum.org/2018/02/14/geth-1-8-iceberg%c2%b9/; Hsiao-Wei Wang, "Ethereum Sharding: Overview and Finality," *Hsiao-Wei Wang* (blog), December 27, 2017, https://medium.com/@icebearhww/ethereum-sharding-and-finality-65248951f649; Ray, "Wiki."

[250] Ryan Todd, "Ethereum Essentials: Node Nuances – The Block," *CnbCrypto* (blog), January 23, 2019, https://cnbcrypto.com/2019/01/ethereum-essentials-node-nuances-the-block/. Though hosted solutions appear more expensive than locally-maintained servers, they include the overhead cost of system monitoring and maintenance.

[251] Benet, "IPFS - Content Addressed, Versioned, P2P File System," sec. 2.1.2.

gigabytes of data;[252] accordingly, the additional space required for IPFS file hosting is minor when compared to space required for the full Ethereum blockchain.

3. (optional) A copy of the files required for the web-based front end for the application. It should be noted here that these files, in practice, could also be stored in IPFS, obviating this requirement entirely.

A Catagora prototype should implement all of these components, and larger-scale development of the platform should presume that participating libraries, also, will dedicate computing and network capacity to the support of the platform with the implementation of full nodes. Though the Catagora system is intended to be publicly-accessible, it will need to be supported by a suite of participating libraries or related institutions, each hosting all of these infrastructure components.

[252] University of Toronto, "University of Toronto Catalog."

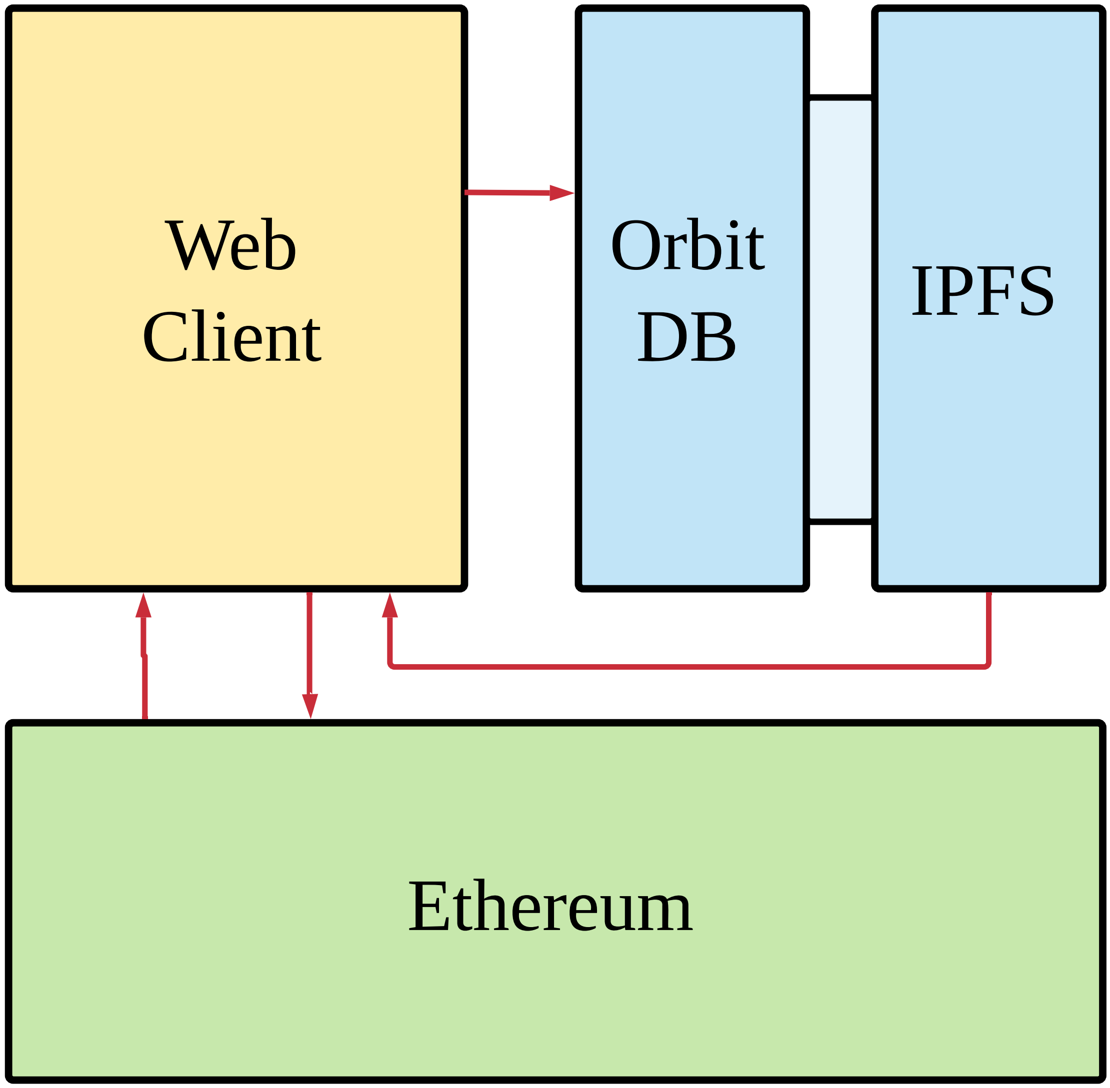

Figure 4: Catagora architecture, with arrows indicating possible component communication paths

System Architecture Components

The Catagora system architecture consists of four major components, as suggested in the previous section:

1. The Ethereum blockchain
2. An OrbitDB database
3. The IPFS file system
4. A web-based client front end, written with HTML, Cascading Style Sheets (CSS) and JavaScript

An abstract model of the architecture is presented in Figure 4 (p. 98). The user interface serves as the mediator between the IPFS and Ethereum components using functions defined in the Web3, OrbitDB, and JS-IPFS Javascript libraries.

It is important to note that this architecture is presented figuratively. Since each of the components of the Catagora system is accessible over the Internet, a user need only have direct access to the web-based interface in order to interact with the full system, wherever it is located and however it is distributed. The same is true for the IPFS and Ethereum components, which already exist as distributed systems that can be accessed over the Internet.

The remainder of this section outlines a primary set of use cases for Catagora, illustrating how the architecture's components interact with and support one another for primary operations. Accompanying diagrams outline the role that each of the system's components play in the execution of these cases.

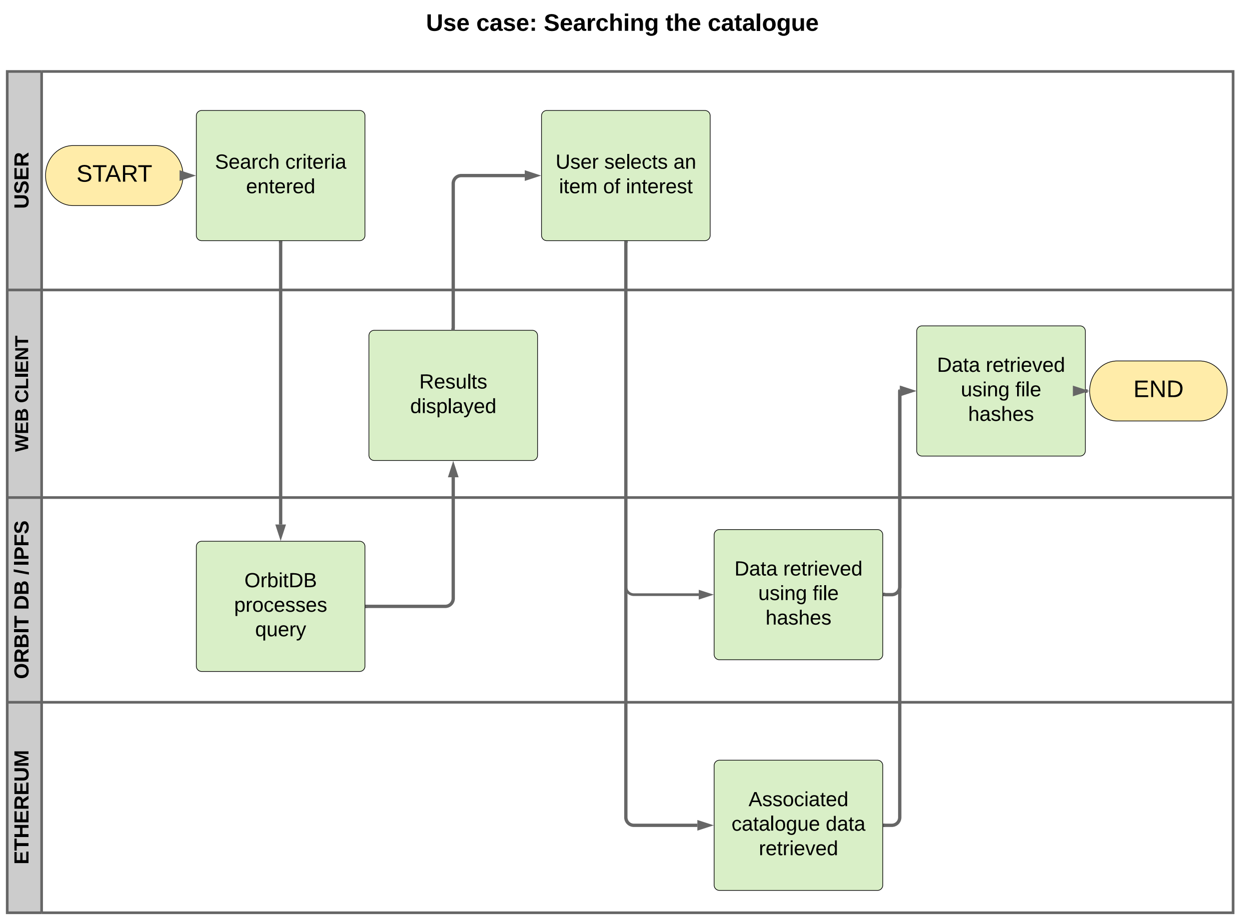

Figure 5: Workflow steps for a Catagora catalogue search

Use case: Searching the catalogue to display an entry

This scenario is depicted in Figure 5 (p. 100) and represents the most common usage scenario for the catalogue: executing a simple search query and returning a set of results. Steps are as follows:

1. A user enters their desired search criteria in an online form.
2. The query is sent to the OrbitDB database, which returns a set of results that include associated file hashes.

3. From the set of results returned by the database, the user selects a catalogue entry they wish to inspect.
4. The client interface pulls the associated data file from IPFS using a new query. At the same time, the interface queries the Ethereum-based catalogue for additional information related to the entry under examination.
5. The Ethereum blockchain returns a list of revisions to the catalogue entry, as well as its associated talk page, including all additional IPFS file hashes related to those items. This is combined with full catalogue entry data retrieved from IPFS.
6. Information in the catalogue record is displayed to the user.

Evolutions of this scenario can extend and augment functionality. For example, the database may support fuzzy or near-term searching, and the user agent may be a human actor or an automated system, such as a discovery agent, that wishes to retrieve and digest data to incorporate it into existing systems.

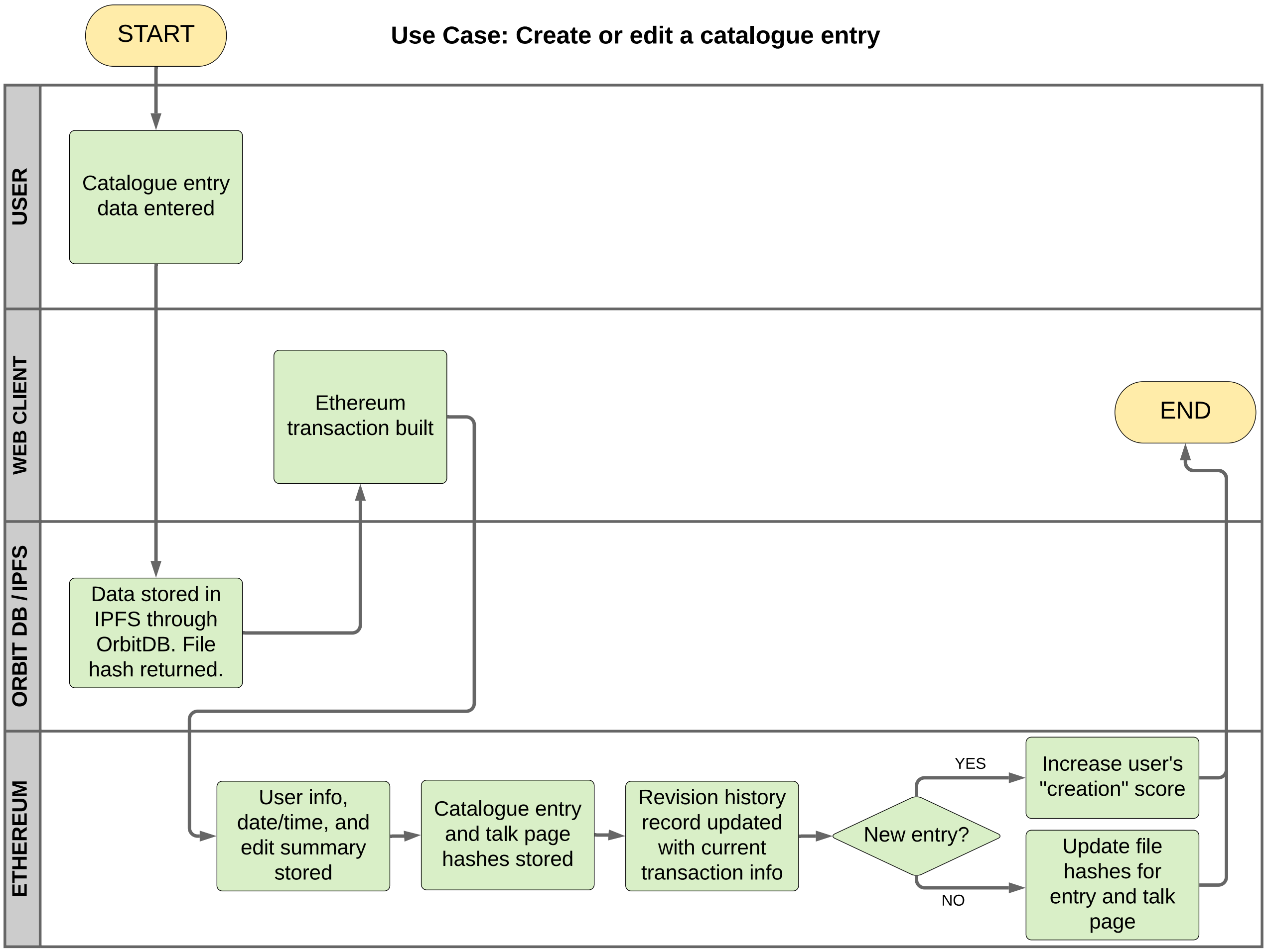

Figure 6: Workflow steps for creating or editing catalogue entry data

Use case: Creation or edit of a catalogue entry or talk page

Creation and editing of catalogue entries are handled in a similar manner and are treated together. The scenario is depicted in Figure 6 (p. 102), and follows these steps:

1. A user agent enters data in the web-based client and submits it for processing.
2. Updated data (which may be a catalogue entry, talk page information, or both) is stored in OrbitDB, which generates and returns an associated file hash. In the case of a newly-created catalogue entry, a new talk page (and associated hash) is also created for the record.

3. The resulting file hashes are bundled into an Ethereum transaction that also includes the address of the user who is creating or editing the record data.
4. The transaction is submitted to the Ethereum blockchain by the user interface.
5. If creating a new record:
    a. The blockchain stores the hash of the record and the hash of its associated talk page in its permanent storage, along with the address of the record's creator, edit summary information, a time and date.
    b. This activity also generates a revision history entry, which is also stored on the blockchain as part of the catalogue record.
    c. The entry creator's "cards" score is incremented to reflect that they have contributed a new entry to the catalogue.
6. If editing an existing record:
    a. The blockchain updates the existing record to point to the newly-created hashes of the catalogue record, talk page, or both.
    b. Information related to the change that was made (the editor's Ethereum address, edit summary, the current date & time and copies of the hashes related to the updated files) are added to the list of revisions for affected catalogue record.
    c. The editor's "contributions" score is incremented to reflect their participation in the system.

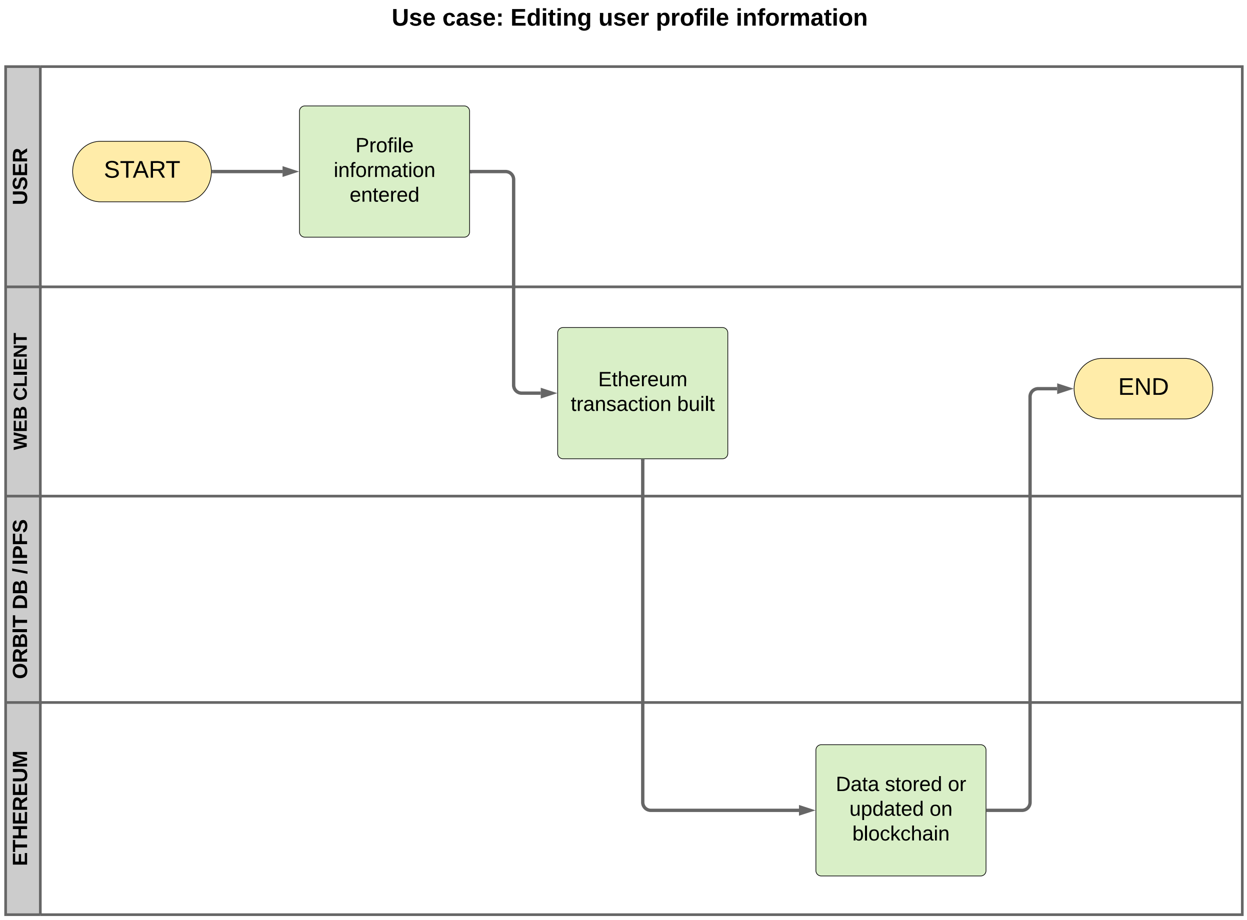

Figure 7: Workflow steps for editing user profile information

Use case: Updating a user profile

All contributors to Catagora have the option to maintain and update a user profile associated with their Ethereum address. The system assumes that all user profiles exist and do not need to be created. If a contributor tries to access a profile that has not yet had any data added, it is simply shown to be empty; the first update of user profile information results in data being stored on the blockchain.

For this design, it is assumed that user profile information is limited (for example, name and institutional affiliation) and can therefore be stored directly on the Ethereum blockchain at minimal cost. Should the requirements for user profiles expand to include long-form text, data could be stored in IPFS instead, using a similar procedure to the one outlined in Figure 6 (p. 102). The steps required for the basic scenario, which are depicted in Figure 7 (p. 104), are as follows:

1. A user agent enters profile information in the web-based client and submits it for processing.
2. The user interface constructs an Ethereum transaction that includes the updated profile information and the address of the user.
3. The transaction is submitted to the Ethereum blockchain by the user interface.
4. Data is stored on the Ethereum blockchain in a structure that maps directly to the user's Ethereum address.

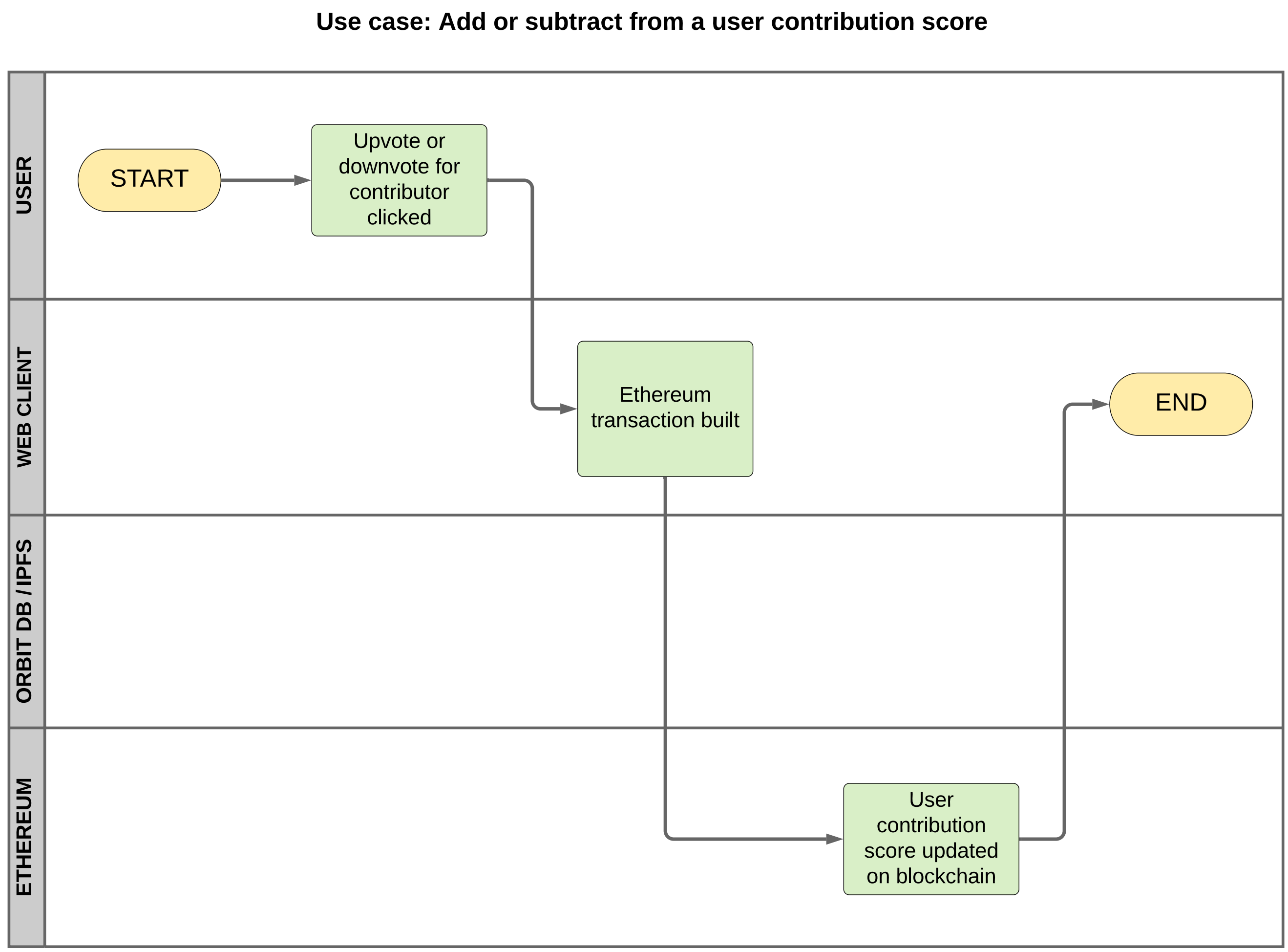

Figure 8: Workflow steps for updating a user's contribution score

Use case: Adding/subtracting from a user's contribution score

Catagora's success as a social information production system is supported with the implementation of a scoring system that can be used to rate the contributions and input of catalogue record contributors. One aspect of this system has already been mentioned in the record creation use case, where the creation of a new catalogue entry increases a contributor's "cards" score. Other details of this reputation system are described in detail later in this chapter, as part of the "public participation" subsection of Catagora's key design characteristics, but the process of incrementing or decrementing a user's contribution score is described by Figure 8 (p. 106) and the following steps:

1. An up- or down-vote for another user's contribution is triggered by a user interaction in Catagora's web interface.
2. An Ethereum transaction is prepared which includes the address of the contributor whose work is being voted on, as well as an indicate to increment or decrement that contributor's score. Since the act of up- or down-voting another user's contribution is a form of participation in the network, the transaction also includes an increment of the voting user's "participation" score.
3. The transaction is submitted to the Ethereum blockchain, which updates the contributor's user score (according to the upvote or downvote values) and the participation score of the voter.

As will be discussed in the "public participation" section, the implementation of user scoring systems is intentionally minimal and is expected to evolve over time. The cost of Ethereum transactions should minimize abuse, but additional criteria, such as a maximum frequency or volume of votes, could deter abuse of the system by automated agents.

Key design characteristics

Decentralization

The level of decentralization achieved by the Catagora design can be assessed with an evaluation of its alignment with the three-faceted model of decentralization presented in Chapter 4.[253]

Catagora's architectural decentralization is supported by the nature of Ethereum's blockchain technology and a preference for its implementation on public Ethereum network, where large numbers of geographically-distributed and independently-operated nodes are available to support the system. The architecture also benefits from the use of IPFS technology, which permits the storage of catalogue records to be likewise distributed. Even the client interface, which would traditionally be accessed on a centrally-controlled server or network, can be decentralized by placing its code on the IPFS infrastructure. Logical decentralization is not possible with current blockchain technology, as discussed in the "Design Framework" chapter, but is achieved in parts of the design with the implementation of IPFS. This is due to the IPFS architecture, which builds on the earlier, peer-to-peer BitTorrent file distribution system. Taken all together, these platforms make Catagora available to anyone with an Internet connection and web browser.

Political decentralization is more challenging to achieve with Catagora. It is not possible to *guarantee* that political forces will not bend Catagora's organization towards centralization, mirroring current shared-cataloguing solutions, but the decisions made for the system's initial design do not *preclude* this as an eventuality. In other words, the system

[253] This model is outlined beginning on page 39.

design presented in this thesis avoids political centralization by refusing to specify content moderation or approval mechanisms that prescribe it. Though some future work is required to minimize the impact of content vandalism and abuse, creating and editing catalogue records in Catagora has intentionally been left open to all contributors in this design. This "open edit" model mirrors Wikipedia, whose implementation was also completely free of programmatic moderation mechanisms for its first five years.[254] Content moderation processes should evolve over time with the input and participation of the development and user community. This design choice also addresses Benkler's *moral systems* design lever: the sustainability and growth of the system is likely to be encouraged through community's ability to define social norms and processes in a collective manner, using the system's own communication tools.

Public participation

The ownership of library catalogues has traditionally been in the sphere of librarians, with professionals undertaking the work of characterizing and describing resources for library collections. The popularity of community networking sites on the internet and the move towards crowd-sourced organization and content management (as

---

[254] One of the earliest proposals for a free and open online encyclopedia came from Richard Stallman in late 2000; Wikipedia appeared shortly thereafter. Wikipedia creator Jimmy Wales relied entirely on self-policing of content—even allowing anonymous creation and editing of articles—until public controversies over article content resulted in the addition of a user registration requirement in 2005. One year later, the site introduced "time delays" for article edits, as well as "semi-protected" and "protected" status settings for articles. See Richard Stallman, "The Free Universal Encyclopedia and Learning Resource," December 18, 2000, https://www.gnu.org/encyclopedia/anencyc.txt; Daniel Terdiman, "Growing Pains for Wikipedia," CNET, December 7, 2005, https://www.cnet.com/news/growing-pains-for-wikipedia/; Katie Hafner, "Growing Wikipedia Refines Its 'Anyone Can Edit' Policy," *The New York Times*, June 17, 2006, sec. Technology, https://www.nytimes.com/2006/06/17/technology/17wiki.html; "History of Wikipedia," in *Wikipedia*, accessed February 24, 2019, https://en.wikipedia.org/wiki/History_of_Wikipedia; EditMe, "Wikipedia's Content Moderation: Proof That Wikis Work," August 28, 2009, https://www.editme.com/wikipediaflaggedrevisions.

used on sites like DeviantArt) has led to user-originated content tagging and the creation of folksonomies that augment resource descriptions.[255] Online platforms like WorldCat and Bibliocommons have also embraced this shift by adding user-defined tags, reviews and ratings to item descriptions, but these are supplements to existing records:[256] mechanisms for direct public input into the content of authoritative catalogue records are nonexistent.

There is precedent for involving a user community in the development of library resource descriptions: work with local communities in Canada is leading to the development of more accurate descriptions for Indigenous material;[257] and a case study in North Carolina demonstrates the value of community consultation in the assignment of subject classifications and bibliographic descriptions.[258]

Catagora proposes to take these steps to the furthest extent possible by providing edit access to anyone who would like to contribute to the work of cataloguers. Like Wikipedia, which allows contributors to make changes to almost any page, Catagora allows users to provide input into the development of resource descriptions through direct interaction with the catalogue itself. This is an intentional shift from the current cataloguing paradigm, where resource descriptions are available for public access, but not

---

[255] Richard Gartner, "Democratizing Metadata," in *Metadata: Shaping Knowledge from Antiquity to the Semantic Web* (New York, NY: Springer Berlin Heidelberg, 2016), 97–106.

[256] BiblioCommons, "BiblioCore," BiblioCommons, 2018, https://www.bibliocommons.com/products/bibliocore; OCLC, "WorldCat Help - Item Details," 2015, https://www.oclc.org/support/help/worldcat/Content/Itemdetails/item_details.htm.

[257] Sharon Farnel, "Making Meaning Together: Decolonizing Descriptions in Local Digitized Collections" (University of Alberta Libraries, 2018), https://doi.org/10.7939/R31G0J933; Denise Koufogiannakis et al., "Decolonizing Description: Changing Metadata in Response to the Truth and Reconciliation Commission" (University of Alberta Libraries, 2017), https://doi.org/10.7939/R3MS3KF68.

[258] Pattuelli, "Modeling a Domain Ontology for Cultural Heritage Resources."

for direct public contribution. The model allows resource descriptions to be improved by users, who may update catalogue entries to incorporate their expertise, their classifications, and their own language; it also echoes a broader shift, among public institutions, towards directly-engaging the public for the benefit of the public good.[259]

For the sake of a prototype, at least, the barriers to participation in cataloguing work are no different for members of the public than they are for library professionals. In a real-world context—much like the evolution of Wikipedia—processes must evolve for content moderation, reversion of abusive edits, etc., but these systems should be allowed to develop with the full participation of the user community.[260]

Commitment to openness

Apart from its emphasis on public access and participation, the principle of openness has been a focus of Catagora's design. Following Pomerantz and Peek's six-faceted exploration of definitions of open,[261] the system is intended to embrace several of the dimensions named in their work.

First, the catalogue is ***open for use***: access to existing data is provided at no cost (since there are no Ethereum transactions associated with reading data from the blockchain) and may be freely downloaded or copied for any purpose. As noted in the previous section, "use" in this context is bidirectional since records may be updated by anyone who can bear the cost of an Ethereum transaction.

---

[259] Nina Simon, Lauren Benetua, and Shelley Bernstein, "OF/BY/FOR ALL," 2018, https://www.ofbyforall.org/; Nina Simon, *The Art of Relevance* (Santa Cruz, CA: Museum 2.0, 2016).

[260] See footnote 254.

[261] Pomerantz and Peek, "Fifty Shades of Open."

Second, though not all legal jurisdictions provide copyright protection for data and databases, Catagora data is contributed with the requirement that it is granted a Creative Commons Zero (CC0) licence.[262] This explicitly frees catalogue records from all copyright restrictions, maximizing ***usage rights*** for anyone who has access to its data. Licensing in this manner also ***enables openness***, since it encourages downstream evolutions and adaptations of catalogue data to preserve openness as a key principle. The source code for Catagora is also open, extending public access rights to the system's engine as well.

The open source nature of Catagora's code represents one of two ways in which the system touches on another facet of openness: ***transparency***. The system's source code is intended to be released under a GNU General Public License (GPL),[263] aligning with the principles outlined in the Open Source Initiative's Open Source Definition.[264] This maximizes the availability and utility of Catagora as free[265] software: the GNU GPL requires that any modified or derivative implementations are distributed under the same licence, and provides protection for free downstream use by limiting the application of patents or technological protection measures that would affect anyone else's ability to modify, study or share the code. The second way that transparency is enabled is with the use of a public

---

[262] Creative Commons, "CC0 1.0 Universal," accessed March 6, 2019, https://creativecommons.org/publicdomain/zero/1.0/.

[263] Open Source Initiative, "GNU General Public License Version 3.0," accessed February 23, 2019, https://opensource.org/licenses/GPL-3.0.

[264] Open Source Initiative, "The Open Source Definition," March 22, 2007, https://opensource.org/osd.

[265] The term "free" is as defined by the Free Software Foundation's commitment to user freedom, and is phrased as "the freedom to share, study and modify." See Free Software Foundation, "What Is Free Software and Why Is It so Important for Society?," 2019, https://www.fsf.org/about/what-is-free-software.

blockchain architecture. The decentralized ledger, implemented through the Ethereum platform, renders all of the catalogue's data—including complete, authoritative records of every change made to every catalogue entry, as well as the methods used to implement those changes—open for public exploration, inspection, and analysis.

Finally, the system has been designed for ***open participation*** by its user community. Incentives for participation are built in to the system using the design levers outlined in the Design Framework chapter,[266] and are discussed in more detail in the next section.

Benkler's design levers for participation and sustainability

Catagora's reputation system is intended to encourage participation through an implementation of social information production "design levers" described by Yochai Benkler.[267] The concept of moral systems has been touched upon briefly in the discussion of the design's decentralization features,[268] but this section explicitly addresses how Benkler's other design levers are implemented in Catagora's design.

Reputation

The system has been designed to track four parameters that can establish and foster the reputation of cataloguing participants. First, each act of participation in the improvement of catalogue data is awarded a point. For example, points are provided for up- or down-voting another cataloguer's contributions (which helps serve the quality of the catalogue), for editing a catalogue record, or for participating in the discussion related to

---

[266] See "Open source software and social information production" on page 68.

[267] *Ibid*.

[268] See pages 108-109.

the catalogue record. Second, all cataloguers are given a "creation" value equal to the number of catalogue entries created by the user. Third, other users of the system have the ability to upvote or downvote each cataloguer's contributions, resulting in a "user-based reputation" value that may be positive or negative. Since these systems can be abused,[269] the fourth parameter is a content reputation score that is determined algorithmically. The proposed method for this determination is a chronological approach that weights contributions according to the survival of each contributor's text and edits.[270]

Together, these values can be used to establish the overall reputation of a cataloguer, and can be used in two ways: first, as a social incentive for participation in the system; and second, as a mechanism for detecting and managing abuse or vandalism of the catalogue's contents. For a prototype, these values need not be combined or processed in any way; however, a full implementation of the system could award differing numbers of points for each kind of contribution, with a total scores combined and processed with other parameters to arrive at final user- and content-reputation scores. Blockchain-based systems such as Steemit use this approach, processing raw contribution score data to arrive at a final, logarithmically-scaled reputation score.[271]

---

269 Adler and de Alfaro, "A Content-Driven Reputation System for the Wikipedia," 262.

270 This method was selected for its low computational weight (which is of critical importance on the Ethereum platform) and its lack of reliance on external information as input. See Adler and de Alfaro, 264–66.

271 arcange, "What Is Steemit Reputation and How Does It Work?," Steemit, January 7, 2017, https://steemit.com/steemit/@arcange/what-is-steemit-reputation-and-how-does-it-works; digitalnotvir, "How Reputation Scores Are Calculated - the Details Explained with Simple Math," Steemit, August 24, 2016, https://steemit.com/steemit/@digitalnotvir/how-reputation-scores-are-calculated-the-details-explained-with-simple-math.

Communication

The implementation of Wikipedia-like "talk" pages for each catalogue entry permits discussion between cataloguers on the content of any record, no matter who the cataloguers are or where they are physically located. This design feature addresses an issue raised in work by F. K. Donnelly, and was referenced in my discussion of Ranganathan's fourth law ("save the time of the reader") in Chapter 2: the decentralized nature of shared cataloguing efforts is not supported by a communications framework for resolving debates about resource descriptions.[272] In addition to providing a direct mechanism for cataloguers to communicate with one another, the open nature of Catagora's system design leaves the path open for future forms of communication and information synthesis (like data mining and visualization tools) that can provide insights across the whole catalogue or subsets of its contents.

Moral systems

Apart from the fact that Catagora's source code and contents are open and accessible, the system's implementation on the Ethereum blockchain renders all of its transactions and state changes transparent and open for public inspection. This transparency is key to Benkler's assertion that the fairness of a system is governed by the fairness of its processes.[273] It also lays groundwork for the community construction of social norms, which will arise over time as cataloguers participate in discussions about how the system's governance should evolve. An intentional decision has been made to

---

[272] Donnelly, "Catalogue Wars and Classification Controversies.," 247.

[273] Benkler, *The Penguin and the Leviathan*, 133–40.

refrain from the wholesale imposition of behavioural incentives or access controls using code, embodying a constructivist (rather than determinist) approach to the design of a new cataloguing technology. The lack of access controls mirrors the evolution of Wikipedia, which emphasized participatory reputational systems and crowd-sourced content moderation;[274] these systems were only introduced after five years of fully-open operation, with the input and participation of the user community.

Reward and punishment

Underpinning the behavioural incentive system for Catagora is the nature of the Ethereum blockchain itself, which imposes both a time delay and a small cost for each transaction. The time delay is determined by the block processing time (15 seconds on average), and each change to data on the Ethereum blockchain (including an update to a catalogue entry and a vote on another cataloguer's contributions) incurs a median transaction cost of about three cents.[275] A cost is also incurred for the storage of each new catalogue record; the design articulated in this chapter represents a per-record storage cost of about 51 cents.[276]

---

[274] Paul B. de Laat, "From Open-Source Software to Wikipedia: 'Backgrounding' Trust by Collective Monitoring and Reputation Tracking," *Ethics and Information Technology* 16, no. 2 (June 2014): 157–69, https://doi.org/10.1007/s10676-014-9342-9.

[275] As of January 2019, the average cost of an Ethereum transaction was 6 cents, with a median value of 3 cents. Though the average cost per transaction has varied slightly, the median cost has remained consistent over time. For current information and historic graphs, see BitInfoCharts, "Ethereum / Ether (ETH) Statistics," BitInfoCharts, accessed February 4, 2019, https://bitinfocharts.com/.

[276] Including two 32-byte file pointers and associated data, most catalogue records should be well under 100 bytes. The 51-cent figure is derived from Ethereum's storage cost of 0.032 ETH per kilobyte (1024 bytes), and a January 2019 cost of $165 CAD per ETH.

Though modeling and analysis of these incentives is beyond the scope of this thesis, it is hoped that these costs address Benkler's "reward and punishment" design lever by creating a barrier against abuse of the system that does not also choke the efforts of the majority of cataloguers. The time delay and transaction costs reduce the ability of human actors to "spam" the system with negative or positive votes on other people's contributions.

Though not explored in this design, additional limits to avoid abuse could be examined using the contribution and reputation scores stored as part of user profile information: these include a limit on the number of contribution votes per day, or limits on editing and contributions made by users with low or dropping content reputation scores. These limits are easily implemented in the Ethereum layer by checking the frequency and type of user contributions against a set of limits or cooldown durations.

Revision History for Catalogue Entries

Perhaps the most unique and paradigm-shifting feature of Catagora, created implicitly by the use of blockchain technology, is the ability to store and retrieve provenance and revision history information for all catalogue records. No known library cataloguing system can make this claim, and the implications of the feature warrant in-depth discussion.

The first electronic catalogue debuted in 1971,[277] nearly commensurate with the first experiments in data networking that would later evolve into the Internet.[278] Had the

[277] Kilgour, "Report to the Committee of Librarians of the Ohio College Association," 1.

[278] DARPA, "ARPANET and the Origins of the Internet," Defense Advanced Research Projects Agency, accessed December 1, 2018, https://www.darpa.mil/about-us/timeline/arpanet.

digital library catalogue been created later, in the age of open source UNIX systems and multi-site, collaborative software development, its design might have taken advantage of text-processing and change management tools that are now commonplace in systems that manage digital data and source code. The creation of tools like *diff* enabled a revolution in revision control for software systems, saving space and processing time by storing a list of changes to a file, rather than multiple copies of the file itself.[279] The economy of computation and storage provided by these tools has made them implicit in the design of modern data management systems.

One significant advantage of blockchain-based systems for library catalogues is the introduction of record-level revision history for every entry in the database. Each iteration of a catalogue entry in Catagora is tagged with a timestamp, the Ethereum account address of the editor, and a complete record of all changes made (including an editing summary and references to discourse related to the change that may have been made in an associated discussion page). This aligns Catagora with the practices of contemporary information management systems in other domains, like source code repositories, and opens resource descriptions to in-depth analysis and support for historiography.

An invaluable paper from 2012[280] demonstrates how analysis of catalogue records can shed light on old contexts and controversies. In it, Katharine Whaite examines the evolution of descriptions and classifications of Thomas Bell's *A Monograph of the*

---

[279] J W Hunt and M D McIlroy, "An Algorithm for Differential File Comparison," 1976, https://nanohub.org/infrastructure/rappture/export/3582/trunk/gui/src/diff.pdf.

[280] Katharine Whaite, "Finding Value in History: Gaining Knowledge by Examining Historical Practices.," *Catalogue & Index*, no. 169 (December 2012): 25–29.

Testudinata,[281] which was later published under the title *Tortoises, Terrapins and Turtles, drawn from life* in 1872.[282] Whaite's investigation was enabled by the presence of a tool from a now-bygone era—physical catalogue cards—allowing her to examine everything from typed content notes to handwritten updates on the cards' margins.

Her analysis illustrates variations in how lithographic plates and other aspects of authorship were attributed, how cataloguers focused on different aspects of the work to better-suit their audiences (scientists, for example, as opposed to members of the general public), and how the books were eventually cross-referenced as variant expressions of the same work. The subjective activity of cataloguers can be seen in the traces of catalogue records, and Whaite's analysis draws a picture of how these workers told the story of the book in the ways they described it. Summarizing, Whaite writes: "Investigating a catalogue as one would a text can be extremely rewarding, and provide information about users, librarians, collections, libraries and institutions, as well as shed light on how those entities interact with each other."[283] Put another way, the ability to see a resource reflected in its surrogates is to give that resource perspective: on itself, on its context, and on the culture that interprets it.

Though the Catagora design does not prescribe or specify the tools required to explore and illuminate revision history information, all the requisite data is in place

---

[281] Thomas Bell, *A Monograph of the Testudinata. [Plates, with Descriptive Letterpress.] Pt. 1-8.* (Samuel Highley: London, 1832).

[282] James de Carle Sowerby and Edward Lear, *Tortoises, Terrapins, and Turtles Drawn from Life,* (London, Paris, and Frankfort: H. Sotheran, J. Baer & co., 1872).

[283] Whaite, "Finding Value in History: Gaining Knowledge by Examining Historical Practices.," 29.

because it is recorded implicitly as part of the cataloguing mechanism. The possibilities this opens for bibliographic research are promising, spanning the following categories and questions: evolutions in cataloguing practice, controversies over content and authorship, and issues with subject-based classification. Each of these topics warrants additional exploration.

Cataloguing Revisions and Changes to Cataloguing Rules

The standards and rules for cataloguing have evolved over time.[284] In addition to reflecting shifts in how we think about information, these changes have been driven by increased pressure to minimize the cost of developing descriptions for resources. Librarians must grapple with the application of "minimal" and "maximum" catalogue rules[285] in addition to the shifting sea of standards and formats.[286] Current cataloguing mechanisms do not allow us to assess the impact of these shifts within the practice of librarianship. Full revision control for catalogue entries allows exploration of key questions like:

1. How has the balance of full and minimal cataloguing entries changed over time? Is there a link between this shift and the increasing resource pressures placed on libraries and their funders?

---

[284] Elisabeth de Rijk Spanhoff, "Principle Issues: Catalog Paradigms, Old and New," *Cataloging & Classification Quarterly* 35, no. 1–2 (December 2002): 37–59, https://doi.org/10.1300/J104v35n01_04; Seymour Lubetzky, "Development of Cataloging Rules," *Library Trends* 2 (1953): 179–86.

[285] Laura Salas-Tull and Jacque Halverson, "Subject Heading Revision: A Comparative Study," *Cataloging & Classification Quarterly* 7, no. 3 (June 4, 1987): 3–12, https://doi.org/10.1300/J104v07n03_02.

[286] Patton, "OCLC's Long Association with Less-Than-Full Cataloging," 22.

2. How have shifting catalogue rules affected the ways in which resources are described?
3. Can historical revisions of catalogue records be conflated to provide more robust information about a resource?
4. When and where have librarians "broken the rules" to describe a resource,[287] and how might these intentional fractures inform the evolution of cataloguing rules?
5. How have catalogue records themselves shifted (i.e. have full records been stripped, to minimize them, or is there a trend towards iteratively adding to records over time)?

Controversies over Content or Authorship

There are many situations where authorship or subject classification for works has been altered as a result of new information. F. K. Donnelly cites numerous examples:[288] how should we re-classify material that is discovered to be a forgery (such as *The Protocols of the Elders of Zion*) or a false memoir (such as James Frey's *A Million Little Pieces*)? What of work that has been scientifically-debated (such as resources about creationism) or affected by a contested representation of identity (such as the works of Joseph Boydon)?[289]

---

[287] Whaite, "Finding Value in History: Gaining Knowledge by Examining Historical Practices.," 28.

[288] Donnelly, "Catalogue Wars and Classification Controversies."

[289] Tanya Talaga, "Joseph Boyden's Identity Crisis Opens up Questions on Who Is Part of a Community," *Toronto Star*, January 14, 2017, https://www.thestar.com/news/canada/2017/01/14/joseph-boydens-identity-crisis-opens-up-questions-on-who-is-part-of-a-community.html; Ian Austen, "Voice for Native Canadians Defends Claim to Be One," *New York Times*, January 14, 2017.

Issues with Subject Classification

Under Library of Congress classification systems (and related ones, like Sears, MeSH, etc.) predetermined, controlled-vocabulary terms are used to describe materials by subject. As language, culture, and context shift, terms may become inaccurate or offensive. Current cataloguing systems do not allow us to ask questions like:

1. What historical terms were used to describe these materials that are now governed by new classifications?
2. How long does it take for updated terms to propagate through existing items? Do librarians struggle with application of new classifications, or refuse to apply them in some cases?
3. What terms have been co-located or cross-referenced with previously-used terms, and how have those terms changed over time?
4. What other patterns exist for publishers, authors, and libraries that have applied specific terms to their materials?

The most prominent activist in this area is Sanford Berman, who has engaged in a career-long fight with the Library of Congress over subject terms that he feels do not reflect contemporary English usage.[290] He keeps a scorecard of his subject classification victories, which include the addition of terms like “makerspaces,” “krumping,” and

[290] When speaking in public, Berman loved to hold up a lightbulb and ask his audiences to identify it. After receiving a unanimous response, he would point out that the Library of Congress’ preferred term was “Electric Lamp – Incandescent.” For a general overview of Berman’s ideology, see Sanford Berman and Tina Gross, “Expand, Humanize, Simplify: An Interview with Sandy Berman,” *Cataloging & Classification Quarterly* 55, no. 6 (August 18, 2017): 347–60, https://doi.org/10.1080/01639374.2017.1327468.

"intersexuality."[291] Other scholars have examined the availability of queer subjects, the use of terms like "east indians" and the development of methods to better-describe indigenous people in libraries and archives.[292] The ongoing interrogation of subject-based classification practices can be supported by using Catagora's implicit storage of revision history and rationale to provide insight into the historical context and representation of information resources.

A new engine for the catalogue

A reinterpretation of Ranganathan's *Five Laws of Library Science*, shaped by the design framework outlined in Chapter 4, has resulted in a system design that aims to nudge contemporary cataloguing practice into a closer alignment with the vision Ranganathan articulated almost a century ago.[293] A new engine, built on a foundation of blockchain technology and supplemented by the implementation of critical lessons from social information production systems, may serve to improve the work of cataloguers, disrupt the tendency towards outsourcing and centralization as a means of cataloguing cost reduction, enable crowd-sourced discussion and evolution of catalogue records, and provide better precision for library users. However, the deliberate decision to embrace a social constructivist approach to technology, embodied in the facets of the Catagora design that are dependent on human adoption and participation, is an acknowledgment that software

[291] Sanford Berman, "Personal LCSH Scorecard," July 2016, https://www.dropbox.com/s/78oqo5igs3u9i0h/sbsh-scorecard-july2016.pdf?dl=0.

[292] Farnel, "Making Meaning Together"; Koufogiannakis et al., "Decolonizing Description."

[293] Ranganathan, *The Five Laws of Library Science*.

design alone cannot prescriptively or independently disrupt the practice of resource description. The human element is as unpredictable as it is critical and will likely govern the ultimate success or failure of a public implementation of Catagora. In other words, the engine of the machine can be updated or replaced, but the utility of the machine is determined by the acceptance and support of its operators. These factors are among the topics covered in the next chapter.

6. Discussion and Next Steps

Returning to Ranganathan

In the Chapter 2 I interrogated Ranganathan's 1931 book, *The Five Laws of Library Science*, to reflect on contemporary cataloguing practices and assess how they measure up to librarianship's canonical foundations. By way of summary, it is worth examining how the design of a shared cataloguing system like Catagora may address (or possibly exacerbate) the challenges outlined in my introduction.

Four of Ranganathan's five laws align the work of librarians with the needs of readers, and ultimately the bridge between an information resource and its audience is the finding aid used to connect them. In large libraries with sizable staffing levels, cataloguers are often excluded from the sphere of patrons, working out of office spaces that physically separate those professionals from the users they serve. Smaller libraries that cannot dedicate staff to cataloguing work exacerbate this separation even further, relying instead on records supplied by third-party services like OCLC's WorldCat. Error rates in catalogue records are typically low[294] and are commonly related to typographical errors and adherence to cataloguing standards, but it is far more difficult for cataloguers to assess whether or not resources have been described using terms that patrons themselves would use in searches. Separation (physical or logical) from patrons cannot possibly improve this gap. An opportunity exists to shorten this distance by allowing cataloguers and patrons to

[294] A recent study at Western University found the error rate in outsourced cataloguing records to be less than 5% overall; this rate is in line with other studies noted in the literature review. See Claire Doran and Cheryl Martin, "Measuring Success in Outsourced Cataloging: A Data-Driven Investigation," *Cataloging & Classification Quarterly* 55, no. 5 (July 4, 2017): 307–17, https://doi.org/10.1080/01639374.2017.1317309.

interact directly concerning the catalogue records themselves. Catagora's commitment to public participation is a possible step in that direction.

The Catagora design provides public access to the revision of catalogue records, as well as the opportunity to discuss the nature of those edits. It is not difficult to envision a future where OPACs that integrate and display Catagora records also offer direct user access to revisions and feedback mechanisms. Opening up a catalogue in this way has benefits and risks. It is likely that errors in records will increase as changes are made that do not conform to bibliographic standards; in addition, subject classifications and resource descriptions may be subject to abuse, vandalism, or activism. On a more positive note, however, patrons with expertise in a subject area will be able to augment records with more robust information, creators of library resources will have the ability to self-catalogue to make their online work discoverable, and information seekers may update resource descriptions to reflect their own use of language. All of these activities have the potential to render information easier to locate by others. In current practice, cataloguers must describe material in a way that anticipates future needs and future searches, hoping to mirror the evolving demands of user communities. Is this more efficient than allowing users an opportunity to do so for themselves? Ranganathan's suggestion that we "save the time of the reader" (his fourth law) can be activated by giving readers the opportunity to tell the catalogue—not librarians, as intermediaries—how resources might be described.

Catagora does not explicitly address Ranganathan's first law—"books are for use"—in a manner that existing catalogues cannot, but the open access mandate inherent in its design explicitly enables downstream uses and adaptations of its data or even the system itself. From its source code foundation, authored under a GPL General Public License,[295] to its catalogue records, which are explicitly rendered free of copyright restrictions, the design enables openness by consciously and deliberately making every component a part of the intellectual commons. An external service may incorporate the decentralized catalogue at will in an application, or it may make copies of records in order to reuse or repurpose them. Each of these choices is intended to help make resources more discoverable, which is a foundational step in ensuring that the accessibility of information resources is maximized. In short, the open access ideal advocated by Ranganathan in the 20th century has been articulated as an Open Access mandate for catalogues in the 21st century.

When Ranganathan observed that libraries are growing organisms, he referred to both the scale and evolution of the institution. The scope of libraries has vastly increased in these aspects, and the common response of institutions has been to focus on economization with the centralization and outsourcing of library services and functions. Decentralization has traditionally been commensurate with inefficiency, but the evolution of technologies like blockchains have promised to usurp this relationship with introduction of scalable efficiency.[296] Unfortunately, the experience of Bitcoin, Ethereum, and similar initiatives has

---

[295] Open Source Initiative, "GNU General Public License Version 3.0."

[296] Challenges with blockchain scalability—especially related to transaction capacity and storage demands—continue to be an area of evolution and development. For Ethereum-related approaches to these challenges, see Hunter Hillman, "The Case for Ethereum Scalability," *Connext* (blog), January 18, 2019, https://medium.com/connext/the-case-for-ethereum-scalability-d2a8035f880f; Josh Stark, "Making Sense of Ethereum's Layer 2 Scaling Solutions: State Channels, Plasma, and Truebit," *Medium* (blog), February 12, 2018, https://medium.com/l4-media/making-sense-of-ethereums-layer-2-scaling-solutions-state-channels-plasma-and-truebit-22cb40dcc2f4.

shown us that centralization is an almost intractable foe. For example, once-distributed Bitcoin mining operations have coalesced into large pools of processing power that threaten to take control of the network.[297] Examples like this refute the proponents of technological determinism, reinforcing the idea that human interaction ultimately governs the evolution of technology. Accordingly, it is unlikely that a decentralized system design can single-handedly resist the forces of central control, and it will not be possible to see if Catagora is different until it is tested *in situ*. That assessment requires a pilot project or implementation that is beyond the scope of this thesis.

Such an experiment should take place, however, and the ecosystem of shared library cataloguing may be an environment in which decentralized administration and technology can be more easily sustained. The source of this belief is the inherently-decentralized nature of cataloguing work, which stands in contrast to many other blockchain experiments. Bitcoin's attempt to replace centralized monetary systems[298] has not usurped the habits of people who have only known governance under such a system, and the result has tended to be a drift towards centralization. By contrast, the work of cataloguers has historically been a decentralized exercise and has drifted towards a centralized model with the creation of collectives like OCLC, supported by the evolution of information and

---

[297] Adem Efe Gencer et al., "Decentralization in Bitcoin and Ethereum Networks," *ArXiv:1801.03998 [Cs]*, January 11, 2018, http://arxiv.org/abs/1801.03998; Alireza Beikverdi and JooSeok Song, "Trend of Centralization in Bitcoin's Distributed Network," in *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)* (2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Takamatsu: IEEE, 2015), 1–6, https://doi.org/10.1109/SNPD.2015.7176229.

[298] Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 1.

communications technology and its ability to minimize transaction costs. The design of Catagora offers the suggestion that the practice can, with adequate support and at reasonable cost, return the practice of shared cataloguing to its inherently-decentralized roots.

Envisioning a Catagora implementation: lessons from other initiatives

The catalogue design presented in this thesis is a starting point, but a true exploration of how such a design might compare to existing shared cataloguing systems warrants an experimental implementation and further study. To foster such work in the future, it is worth offering recommendations for a project that may extend the design to include social incentives and supports that cannot be addressed by source code alone. How may the implementation of Benkler's design levers (framing, authenticity, empathy, and solidarity) be evaluated and improved? By exploring projects that have attempted to serve the public good, can we derive common elements might inform the design of a Catagora implementation project?

An analysis of other examples of librarian-driven, technology-enabled projects may serve to provide recommendations. Two boundaries, aligned with the form and purpose of Catagora, have been applied to the exploration presented in this section. First, selected examples focus on the dominance of information and communications technologies (ICT) in the development and evolution of library services, as suggested by literature citing technology as a critical factor in services that contribute to the good of the public.[299]

---

[299] Shannon Crawford Barniskis, "Access and Express: Professional Perspectives on Public Library Makerspaces and Intellectual Freedom," *Public Library Quarterly* 35, no. 2 (April 2, 2016): 103–25,

Second, each of the chosen examples reflect an extension of the practice of librarianship outside the walls of the traditional institution. This choice is supported by recent debates about the role of libraries in the information age, several of which suggest the tradition of library-as-edifice is outdated.[300]

Selection of project examples

Cases for analysis were selected using a purposive sampling method.[301] Within the boundary conditions presented in the previous section, two selection criteria were applied in the search for related projects:

1. The creation of a single system or tool, versus the creation of a technology or tool that libraries can implement independently.
2. Coordination by a central authority or organization versus independent, uncoordinated participation.

Applying this logic, the following four projects were chosen:

---

https://doi.org/10.1080/01616846.2016.1198644; Paul T. Jaeger et al., "Democracy, Neutrality, and Value Demonstration in the Age of Austerity," *The Library Quarterly* 83, no. 4 (October 2013): 368–82, https://doi.org/10.1086/671910; Agnes Mainka et al., "Public Libraries in the Knowledge Society: Core Services of Libraries in Informational World Cities," *Libri: International Journal of Libraries & Information Services* 63, no. 4 (November 2013): 295–319, https://doi.org/10.1515/libri-2013-0024; Leif Kajberg, "Revisiting the Concept of the Political Library in the World of Web 2.0 Technologies," *Progressive Librarian*, no. 36/37 (Summer/Fall2011 2011): 30–41.

[300] R. David Lankes, "Libraries Are Obsolete," *OLA Quarterly* 18, no. 2 (2012): 12–17, https://doi.org/10.7710/1093-7374.1354; John McTernan, "Don't Mourn the Loss of Libraries – the Internet Has Made Them Obsolete," *The Telegraph*, March 29, 2016, https://www.telegraph.co.uk/opinion/2016/03/29/dont-mourn-the-loss-of-libraries--the-internet-has-made-them-obs/; Thu-Huong Ha, "Forbes Deleted a Deeply Misinformed Op-Ed Arguing Amazon Should Replace Libraries," Quartz, July 23, 2018, https://qz.com/1334123/forbes-deleted-an-op-ed-arguing-that-amazon-should-replace-libraries/.

[301] Victor Jupp, "Purposive Sampling," in *The SAGE Dictionary of Social Research Methods* (London, United Kingdom: SAGE Publications, Ltd, 2006), https://doi.org/10.4135/9780857020116; "Purposive Sampling," in *The SAGE Encyclopedia of Qualitative Research Methods*, by Lisa Given (Thousand Oaks, CA: SAGE Publications, Inc., 2008), https://doi.org/10.4135/9781412963909.n349.

- The HathiTrust Digital Library: single system, central coordination.
- The Internet Archive's Archive-It service: single system, independent coordination.
- The Public Knowledge Project's (PKP) Open Journal Systems software: independent implementations, central coordination.
- The Library Freedom Project's (LFP) Tor node program: independent implementations, independent coordination.

Content analysis of documentation and related academic literature for each of these projects—even though none of the selected projects have been exhaustively-documented or studied—results in a useful sketch of each project's history, milestones, challenges and successes.

Descriptions of the selected projects

The HathiTrust Digital Library was created and launched in 2008. The project began when the University of Michigan proposed a shared, centralized digital archive.[302] The HathiTrust partnership now consists of more than 135 institutions and six consortia, [303] united in the goal of creating a co-owned digital archive of materials created from existing print collections.[304]

---

302 Alissa Centivany, "The Dark History of HathiTrust," 2017, 2357, https://doi.org/10.24251/HICSS.2017.285.

303 HathiTrust Digital Library, "Member Community," accessed November 20, 2018, https://www.hathitrust.org/community.

304 Jeremy York, "This Library Never Forgets: Preservation, Cooperation, and the Making of HathiTrust Digital Library," in *Archiving 2009 Final Program and Proceedings* (Archiving, Society for Imaging Science and Technology, 2009), 6; HathiTrust Digital Library, "Mission and Goals," accessed November 20, 2018, https://www.hathitrust.org/mission_goals.

In January 2004 the Internet Archive released their open source web-crawling tool, *Heritrix*, and the subscription-based Archive-It service debuted in early 2005.[305] The service allows subscribers to curate, crawl and archive collections of websites, which are then stored under open access principles in the Internet Archive's data centres. Archive-It is part of a larger suite of projects under the Internet Archive umbrella.[306]

Open Journal Systems (OJS) software is maintained by the larger Public Knowledge Project (PKP) initiative. The software was created to encourage scholars to publish in online, open access publications. Making the package free and open source nearly eliminated in-house software management and development costs for online journal publication, lowering access barriers for scholarly communication.[307]

Finally, in July 2015 the Library Freedom Project (LFP) partnered with The Tor Project to promote the deployment of Tor exit relays in libraries.[308] Tor—or "The Onion Router"—is free, open source software (FOSS) designed to protect privacy on networks by relaying traffic through a randomly-selected series of encrypted server connections.[309] The system depends on volunteer-run servers, and LFP felt that librarianship's alignment with

---

[305] "Archive-It - Web Archiving Services for Libraries and Archives," accessed November 2, 2018, https://archive-it.org/.

[306] Internet Archive, "Internet Archive: Projects," accessed November 30, 2018, https://archive.org/projects/.

[307] John Willinsky, "Open Journal Systems: An Example of Open Source Software for Journal Management and Publishing," ed. Scott P. Muir, *Library Hi Tech* 23, no. 4 (December 2005): 506–8, https://doi.org/10.1108/07378830510636300.

[308] Library Freedom Project, "Tor Exit Relays in Libraries: A New LFP Project," July 28, 2015, https://libraryfreedomproject.org/torexitpilotphase1/.

[309] The Tor Project, Inc., "Tor Project: Overview," accessed November 21, 2018, https://www.torproject.org/about/overview.html.en.

values of intellectual freedom and privacy would make libraries an excellent site for Tor servers and for education on how to use Tor.[310]

Success factors for the selected projects

It is difficult to arrive at a standard definition of "success" that applies to all four projects: first, the nature of their work varies greatly; and second, it is impossible to account for external factors that did not leave traces in the limited volume of materials that could be located. Only the HathiTrust project articulates an over-arching goal for its work,[311] but the efforts of all four projects are ongoing. Archive-It will exist as long as institutions are willing to pay for their subscriptions; LFP's Tor node program has wound down but the organization has created resources for libraries interested in the technology.[312] The number of Open Journal Systems software installations appears to have peaked in 2016,[313] but evincing the reasons for this is beyond the scope of this exploration. Though all four projects have attained some form of success, momentum has been strongest for HathiTrust and Open Journal Systems.

Based on information available for this limited analysis, common themes related to the outcomes of all four projects are as follows: cost, alignment with the ideology of librarianship, and strategic use of partnerships. These themes align helpfully with Catagora's design framework and design.

---

[310] Library Freedom Project, "Tor Exit Relays in Libraries," para. 3.

[311] HathiTrust Digital Library, "Mission and Goals."

[312] Library Freedom Project, "Curriculum for Teaching All about Tor," accessed November 28, 2018, https://libraryfreedomproject.org/allabouttor/.

[313] Public Knowledge Project, "OJS Stats," accessed November 28, 2018, https://pkp.sfu.ca/ojs/ojs-usage/ojs-stats/.

The most interesting cost-related finding is this: the realization of cost savings is less important than their *perception* (or, perhaps more specifically, the stated *intent* to realize them). Users of Archive-It, Open Journal Systems and HathiTrust all stated a desire to reduce costs, suggesting that that this concern must be addressed to obtain institutional buy-in for projects of this kind. However, little data could be found to suggest cost reductions actually materialized for the projects under analysis. A 2010 OJS study found that 51% of respondents' journals were breaking even and 28% were operating at a deficit but did not ask respondents to compare their pre-OJS and post-OJS cost figures.[314] Ithaka S+R's 2011 HathiTrust survey was the only case where cost savings information was reported; however, survey respondents stated that the project's mission and sense of collaboration trumped cost as drivers for participation.[315]

As suggested by comments in Ithaka S+R's 2011 report on the HathiTrust project, an intentional and explicit alignment with the ideology of librarianship is central to success. This conforms to Benkler's descriptions of framing, empathy and solidarity as critical design levers for sustainable social information production projects. HathiTrust, OJS and the Internet Archive all claim to serve the public good by enhancing the openness and accessibility of library resources, scholarly publications, and web archives. Though services of this kind are intended to serve the public good, and in alignment with Brafman and

---

[314] Brian D Edgar and John Willinsky, "A Survey of Scholarly Journals Using Open Journal Systems," *Scholarly and Research Communication* 1, no. 2 (June 14, 2010): 15, https://doi.org/10.22230/src.2010v1n2a24.

[315] Ithaka S+R, "Briefing Paper on Progress and Opportunities for HathiTrust," July 15, 2011, 5, https://www.hathitrust.org/documents/hathitrust-3year-review-2011.pdf.

Beckstrom's model for successful decentralized organizations,[316] the *adoption* of these services depends on a project champion's ability to convince librarians that their work is simpatico with the core values of librarianship.

It is clear that successful projects seek early alignment with existing consortia and associations and are typically seeded by grants or direct funding. The strongest evidence for this is in the *absence* of such elements for the Library Freedom Project (LFP) and its minimal success in comparison with the other three projects. There is evidence that LFP has learned its lesson in this area: it has pivoted into an education centre with a focus on training librarians to be local advocates for privacy technologies like Tor relays, and won federal funding in 2018 as a result of this shift.[317] It is likely no coincidence that the best-known and widest-reaching of the four analyzed projects—HathiTrust and Open Journal Systems—place a strong focus on partnerships and institutional collaborations.

Recommendations for a Catagora implementation project

In light of the exploratory nature of this analysis, what advice can we take on strategies for planning and implementing a Catagora-based pilot project? The examples covered in this section provide guiding principles that might read as follows:

1. Focus on Catagora's alignment with the core values of librarianship, including access, democracy, diversity, and the public good.[318]

---

[316] Brafman and Beckstrom, *The Starfish and the Spider*, 87–100.

[317] Library Freedom Project, "Library Freedom Institute," accessed November 28, 2018, https://libraryfreedomproject.org/lfi/.

[318] American Library Association, "Core Values of Librarianship."

2. Aim to keep cost savings in mind and rationalize these savings as much as possible as part of the implementation design. Acknowledge that externalities are difficult to predict or track, but build cost measurement instruments into the implementation.
3. Disseminate information about the project, and provide support for distributed efforts, by identifying a single point of contact with enough resources to handle or delegate requests for action.
4. Select strategic partners *within the field of librarianship* who may benefit from the use of the tool and who can advocate on its behalf. The cursory analysis of MARC 040 records presented in "Decentralization" on page 108 suggests that targeted approaches to a small number of major cataloguing institutions may prove fruitful.
5. The Catagora implementation requires the relentless advocacy of a key individual who can serve as a catalyst for the creation and evolution of the catalogue.

Despite these recommendations, and like the software design of Catagora itself, no formula can guarantee success. The features inherent in the design of the system offer great promise, however, and so it is hoped that the recommendations presented in this section may serve as guidance for an implementation that can further-realize the goals of Catagora's conceptual design. Possibilities for future research are provided in the next section.

Next Steps

The design presented in this thesis is conceptual, and outstanding questions about its value and effectiveness cannot be addressed without some form of implementation and evaluation. Four areas of future interest, all of which pose additional questions, are presented here: comparing functionality to existing catalogues, evaluating of the cost of Catagora's implementation and operation, addressing vandalism and content moderation, and assessing the effectiveness of the social incentives in the design. These topics provide a wide range of possible next steps for work on Catagora.

Comparisons to existing catalogues

Catagora's untested design leaves lingering questions about the relevance of the system to Ranganathan's second and third laws.[319] Since the heart of a shared catalogue is the information submitted and curated by disparate participants, and the value of that information cannot be gauged until it is discoverable and usable, it remains to be seen whether or not such a system has the ability to connect information resources and their users in a way that existing cataloguing systems cannot.

Comparing Catagora to existing catalogue designs is challenging since current implementations are typically integrated into larger library automation frameworks, like Integrated Library Systems (ILS), which include modules for material acquisition, circulation management, and patron services.[320] One approach would be to model Catagora separately as an independent database of specific content, such as Open Educational Resources, with an interface similar to those provided by contemporary OPACs. This configuration would enable testing of catalogue record creation mechanisms, evaluation of

---

[319] As a refresher, the second and third laws are "every reader his book" and "every book its reader." See Chapter 2.

[320] Lopata, "Integrated Library Systems. ERIC Digest."

record quality, and the effectiveness of interface-based invitations to improve existing resource descriptions or contribute new ones.

Cost determinations

Estimates for the cost of Ethereum node maintenance and data storage are provided in Chapter 5, but the actual costs of implementing and operating Catagora need to be measured and compared to existing systems. There are three factors that must be considered as part of "actual" cost measurements:

1. Data storage costs are "global" on the Ethereum platform but are paid on a transaction-by-transaction basis when records are created or updated. Like the catalogue itself, these costs will be distributed among its participants.
2. A library that runs a full Ethereum node will be participating in transaction processing and validation and will be able to generate revenue as part of this activity. Moreover, the nature and scope of this revenue will vary over time as the global Ethereum network evolves and shifts away from energy-wasting proof-of-work validation mechanisms towards proof-of-stake alternatives.
3. Ethereum nodes may be deployed on institutionally-owned hardware or under the management of cloud-based service operators. These approaches involve unexamined trade-offs between capital investment and operational overhead costs.

Each of these cost factors is worthy of exploration on its own, but their *combination*, whose optimum formulation is unknown, will serve as the ultimate basis for comparison against existing forms of catalogue creation and maintenance.

Vandalism detection and content monitoring

Abuse of content on Wikipedia has been well-studied and is almost certain to appear in relation to controversial Catagora entries because of the intended open edit policy. The design articulated in Chapter 5 does not address this concern, apart from suggesting that the community should participate in the development of moderation systems and that these systems must be addressed in a future phase of the project. Thankfully, the Wikipedia experience suggests some possible paths forward.

Research has shown that about 5% of Wikipedia edits are the result of vandalism,[321] and that more than 90% of the abuse comes from anonymous users.[322] Moreover, preliminary testing of the content-based reputation system developed by Adler and de Alfaro showed that their six-parameter system had good ability to predict poor-quality edits based on reputation scores.[323] Apart from the possibility that the user community will develop automated approaches to vandalism detection,[324] two methods for dealing with record abuse could be implemented and tested based on Wikipedia research:

1. Rely on content-based reputation mechanisms to predict edits that are likely-abusive and flag those transactions for moderation by highly-reputed

---

[321] Lakshmish Ramaswamy et al., "A Content-Context-Centric Approach for Detecting Vandalism in Wikipedia," in *9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (9th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2013), Austin, TX, 2013), 115.

[322] Songrit Maneewongvatana and Suthathip Maneewongvatana, "A Combined Approach to Minimizing Vandalisms on Wikipedia," in *2010 10th International Symposium on Communications and Information Technologies* (2010 10th International Symposium on Communications and Information Technologies (ISCIT), Tokyo, Japan: IEEE, 2010), 398, https://doi.org/10.1109/ISCIT.2010.5664872.

[323] Adler and de Alfaro, "A Content-Driven Reputation System for the Wikipedia," 268–70.

[324] Maneewongvatana and Maneewongvatana, "A Combined Approach to Minimizing Vandalisms on Wikipedia," 395.

contributors. Implementing this approach under the Adler and de Alfaro model requires that all new contributors are assumed to have low reliability until a body of constructive contributions establishes a healthy overall reputation score.

2. Remove the ability for anonymous Catagora users to contribute or make edits to catalogue entries. This can be meaningfully achieved without creating additional privacy-related issues: for example, a security-focused Self-Sovereign Identity (SSI) solution like Sovrin[325]—which is also blockchain-based—can serve to ensure cataloguers are unique and non-impersonating. The system uses external, automated "agents" as validators, providing cryptographic proof-of-identity tokens that can be stored as part of a Catagora user profile. Institutional affiliations for contributors can likewise be validated and stored in this way. Authenticated cataloguers who demonstrate patterns of vandalism can be delayed or blocked from vandalizing entries.

Effectiveness of incentives

The implementation of many of Benkler's social information production design levers, as outlined in Chapter 5, will produce a multi-variate system (including cost- and reputation-related factors this section has already touched upon) that cannot be examined without a live Catagora environment to serve as a basis for observation. Their effectiveness is further complicated with the suggested use of human-centred elements, like moral

---

[325] Sovrin Foundation, "Sovrin," Sovrin, 2019, https://sovrin.org/.

framing, to encourage sustainable, intrinsically-motivated participation in the work of distributed, shared cataloguing.

7. Conclusion

The shared catalogue design presented in this thesis is based on the open source, decentralized Ethereum platform and is grounded in principles of open use, transparency, and public participation. Its elements are derived from an in-depth analysis of current cataloguing challenges, and its ability to address those challenges marks it as a significant contribution to nascent explorations of blockchain technology in the field of librarianship. While this work was created in response to three research questions, the areas of future interest outlined in the previous section highlight the inability of Catagora's conceptual design, alone, to address all three questions completely.

A decentralized, collaborative library catalogue clearly has the potential to improve cataloguing's adherence to Ranganathan's *Five Laws of Library Science*; in fact, the ability to decentralize shared cataloguing work effectively seems to be intrinsically-tied to a decision to make the technology more transparent and open to public participation. The design of Catagora and its commitment to the principles of openness represent an opportunity to bring members of the public into more intimate contact with the finding aids of the library. Eliminating the friction between the catalogue, its creators, and its users with a welcoming approach to contribution and discourse can serve to enhance the quality of catalogue records and make them more responsive to the needs and search preferences of information seekers. Moreover, blockchain technology affords this functionality at a cost that can be kept reasonable with a mindful implementation of the Ethereum platform, supplemented with the Interplanetary File System (IPFS) as a closely-connected data storage partner.

The ability for shared cataloguing practices to be truly decentralized is somewhat limited from a technical perspective and cannot be guaranteed from a social perspective. On the technical side, the nature of blockchain technology (which depends upon the integrity of a single, linked chain of transactional data) renders it incapable of full logical decentralization, but this failing is mitigated somewhat by the torrent-like nature of the IPFS system used to store the catalogue's data and user interface. Though architectural decentralization is achievable from a technical perspective with widespread distribution of "full" Catagora nodes across the network, the ability to attain and maintain decentralization of this type depends heavily on human-centred decisions about implementation including the location of network nodes, the number of them, and the mechanisms employed to administer and maintain them. The architectural factor of decentralization is deeply-connected to the factor of political decentralization, but political decentralization is the most difficult to predetermine with the design articulated in this thesis. An ideal scenario would be one where the technology is eagerly adopted, on an independent basis, by libraries who wish to implement it, but lessons from existing large-scale initiatives like the HathiTrust Project suggest that this may not be possible without central coordination of some kind. Though the act of decentralization involves the work of giving up central control, there is nothing in the design of this system that does not prevent a politically-centralized implementation. Indeed, the unique affordances of blockchain-based catalogues, such as community collaboration and full revision history for catalogue records, may serve as an enticing incentive for existing, centralized cataloguing service providers like OCLC.

The encouraging results of this exploration suggest that cataloguing systems powered by decentralized ledgers hold great potential for the field of librarianship. The architecture of blockchain-based systems and their implicit ability to track the history and provenance of stored data enables a historiographic analysis of cataloguing records and processes that is unequalled by any current shared-cataloguing platform. However, the most critical aspect of cataloguing work—the contributions and participation of the public and of cataloguers themselves, as well as the utility provided to users of catalogues—cannot be examined without further exploration, including a “live” implementation intended to assess how the design articulated in this thesis translates into its envisioned benefits. A promising opportunity lies in a centrally-coordinated implementation project that targets the participation of institutions where cataloguing work is a focus. On the basis of this work, and the new questions it poses, it is hoped that such a project will be undertaken.

References

1Kosmos. "Costs Of Storing Data On The Blockchain." 1Kosmos BlockID, May 1, 2018. https://onekosmos.com/blog/cost-of-storing-data-on-the-blockchain/.

Adler, B. Thomas, and Luca de Alfaro. "A Content-Driven Reputation System for the Wikipedia." In *Proceedings of the 16th International Conference on World Wide Web - WWW '07*, 261. Banff, Alberta, Canada: ACM Press, 2007. https://doi.org/10.1145/1242572.1242608.

Allison, DeeAnn. "Information Portals: The Next Generation Catalog." *Journal of Web Librarianship* 4, no. 4 (November 30, 2010): 375–89. https://doi.org/10.1080/19322909.2010.507972.

American Library Association. "Core Values of Librarianship." Text. Advocacy, Legislation & Issues, July 26, 2006. http://www.ala.org/advocacy/intfreedom/corevalues.

Anthony, Denise, Sean W. Smith, and Timothy Williamson. "Reputation and Reliability in Collective Goods: The Case of the Online Encyclopedia Wikipedia." *Rationality and Society* 21, no. 3 (August 2009): 283–306. https://doi.org/10.1177/1043463109336804.

Apache Software Foundation. "Apache Hadoop," 2018. https://hadoop.apache.org/.

arcange. "What Is Steemit Reputation and How Does It Work?" Steemit, January 7, 2017. https://steemit.com/steemit/@arcange/what-is-steemit-reputation-and-how-does-it-works.

"Archive-It - Web Archiving Services for Libraries and Archives." Accessed November 2, 2018. https://archive-it.org/.

Arthur, W. Brian. "Positive Feedbacks in the Economy." *Scientific American* 262, no. 2 (February 1990): 92–99. https://doi.org/10.1038/scientificamerican0290-92.

Association of Research Libraries. "Summary of Antitrust Lawsuit: SkyRiver & Innovative Interfaces v. OCLC," January 1, 2010. http://www.arl.org/news/arl-news/2356-summary-of-antitrust-lawsuit-skyriver-innovative-interfaces-v-oclc.

Atkinson, Hugh C. "A Brief for the Other Side." *Journal of Academic Librarianship* 9, no. 4 (1983): 200–201.

Atkinson, Robert D., and Michael Lind. "Is Big Business Really That Bad?" *The Atlantic*, April 2018. https://www.theatlantic.com/magazine/archive/2018/04/learning-to-love-big-business/554096/.

Atzori, Marcella. "Blockchain Technology and Decentralized Governance: Is the State Still Necessary?" *Journal of Governance and Regulation* 6, no. 1 (2017). https://doi.org/10.22495/jgr_v6_i1_p5.

Austen, Ian. “Voice for Native Canadians Defends Claim to Be One.” *New York Times*. January 14, 2017.

Ayers, Sheila. “The Outsourcing of Cataloging: The Effect on Libraries.” *Current Studies in Librarianship* 27, no. 1/2 (2003): 17–28.

Banush, David. “Cooperative Cataloging at the Intersection of Tradition and Transformation: Possible Futures for the Program for Cooperative Cataloging.” *Cataloging & Classification Quarterly* 48, no. 2–3 (February 10, 2010): 247–57. https://doi.org/10.1080/01639370903535742.

Barney, Darin David. *The Network Society*. Key Concepts. Cambridge, UK: Polity, 2010.

Barniskis, Shannon Crawford. “Access and Express: Professional Perspectives on Public Library Makerspaces and Intellectual Freedom.” *Public Library Quarterly* 35, no. 2 (April 2, 2016): 103–25. https://doi.org/10.1080/01616846.2016.1198644.

Barton, Joshua, and Lucas Mak. “Old Hopes, New Possibilities: Next-Generation Catalogues and the Centralization of Access.” *Library Trends* 61, no. 1 (2012): 83–106. https://doi.org/10.1353/lib.2012.0030.

Bates, Marcia J. “Designing Online Catalog Subject Access To Meet User Needs.” In *55th IFLA Council and General Conference*, 40–24 to 40–26. Paris, France: IFLA Division of Bibliographic Control/Section on Classification and Indexing, 1989.

Beall, Jeffrey. “OCLC: A Review.” In *Radical Cataloging: Essays at the Front*, edited by K. R. Roberto, 85–93. Jefferson, N.C: McFarland & Co, 2008.

Beikverdi, Alireza, and JooSeok Song. “Trend of Centralization in Bitcoin’s Distributed Network.” In *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 1–6. Takamatsu: IEEE, 2015. https://doi.org/10.1109/SNPD.2015.7176229.

Bell, Thomas. *A Monograph of the Testudinata. [Plates, with Descriptive Letterpress.] Pt. 1-8.* Samuel Highley: London, 1832.

Benet, Juan. “IPFS - Content Addressed, Versioned, P2P File System,” April 1, 2015. https://github.com/ipfs/papers/raw/master/ipfs-cap2pfs/ipfs-p2p-file-system.pdf.

Benkler, Yochai. “Coase’s Penguin, or, Linux and ‘The Nature of the Firm.’” *The Yale Law Journal* 112, no. 3 (December 2002): 369. https://doi.org/10.2307/1562247.

———. *The Penguin and the Leviathan: The Triumph of Cooperation over Self-Interest*. 1st ed. New York, NY: Crown Business, 2011.

———. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press, 2006.

Berman, Sanford. "Personal LCSH Scorecard," July 2016. https://www.dropbox.com/s/78oqo5igs3u9i0h/sbsh-scorecard-july2016.pdf?dl=0.

Berman, Sanford, and Tina Gross. "Expand, Humanize, Simplify: An Interview with Sandy Berman." *Cataloging & Classification Quarterly* 55, no. 6 (August 18, 2017): 347–60. https://doi.org/10.1080/01639374.2017.1327468.

Bianchini, Carlo, and Mauro Guerrini. "A Turning Point for Catalogs: Ranganathan's Possible Point of View." *Cataloging & Classification Quarterly* 53, no. 3–4 (May 19, 2015): 341–51. https://doi.org/10.1080/01639374.2014.968273.

BiblioCommons. "A Better Online Catalog Is Just the Beginning." BiblioCommons, 2017. https://www.bibliocommons.com/about/.

———. "BiblioCore." BiblioCommons, 2018. https://www.bibliocommons.com/products/bibliocore.

Bishop, David F. "OCLC Copyright: A Threat to Sharing." *Journal of Academic Librarianship* 11, no. 4 (September 1985): 202.

Bishop, William Warner. "A Decade of Library Progress in America" 66 (December 1904): 131–38.

BitInfoCharts. "Ethereum / Ether (ETH) Statistics." BitInfoCharts. Accessed February 4, 2019. https://bitinfocharts.com/.

BitTorrent.org. "What Is BitTorrent?," 2015. http://www.bittorrent.org/introduction.html.

BlockChannel. "A Crash Course in Mechanism Design for Cryptoeconomic Applications." *Medium* (blog), October 17, 2017. https://medium.com/blockchannel/a-crash-course-in-mechanism-design-for-cryptoeconomic-applications-a9f06ab6a976.

Bloomfield, Brian P., and Rod Coombs. "Information Technology, Control and Power: The Centralization and Decentralization Debate Revisited." *Journal of Management Studies* 29, no. 4 (July 1992): 459–459. https://doi.org/10.1111/j.1467-6486.1992.tb00674.x.

BOINC. "Volunteer Computing," 2014. https://boinc.berkeley.edu/trac/wiki/VolunteerComputing.

Bolin, Mary K. "Make a Quick Decision in (Almost) All Cases: Our Perennial Crisis in Cataloging." *Journal of Academic Librarianship* 16, no. 6 (1991): 357–61.

Borra, Erik, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. "Societal Controversies in Wikipedia Articles." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 193–96. Seoul, Republic of Korea: ACM, 2015. https://doi.org/10.1145/2702123.2702436.

Bourne, Charles P., and Trudi Bellardo Hahn. *A History of Online Information Services, 1963-1976*. Cambridge, MA: MIT Press, 2003.

Bowles, Samuel. *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. The Castle Lectures in Ethics, Politics, and Economics. New Haven, CT; London, UK: Yale University Press, 2016.

Brafman, Ori, and Rod A Beckstrom. *The Starfish and the Spider: The Unstoppable Power of Leaderless Organizations*. New York, NY: Portfolio, 2014.

Brandom, Russell. "Amazon Web Services Starts Blocking Domain-Fronting, Following Google's Lead." The Verge, April 30, 2018. https://www.theverge.com/2018/4/30/17304782/amazon-domain-fronting-google-discontinued.

Brophy, Jan, and David Bawden. "Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources." *Aslib Proceedings* 57, no. 6 (December 2005): 498–512. https://doi.org/10.1108/00012530510634235.

Brown, Rowland C. W. "OCLC, Copyright, and Access to Information: Some Thoughts." *Journal of Academic Librarianship* 11, no. 4 (September 1985): 197–98.

Buckland, Michael Keeble. *Redesigning Library Services: A Manifesto*. Chicago: American Library Association, 1992.

Buterin, Vitalik. "EIP-2: Homestead Hard-Fork Changes." Ethereum Improvement Proposals, November 15, 2015. http://eips.ethereum.org/EIPS/eip-2.

———. "The Meaning of Decentralization." *Vitalik Buterin* (blog), February 6, 2017. https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274.

Buterin, Vitalik, and Virgil Griffith. "Casper the Friendly Finality Gadget." *ArXiv.Org*, October 25, 2017. http://arxiv.org/abs/1710.09437.

Calhoun, Karen, Joanne Cantrell, Peggy Gallagher, and Janet Hawk. "Online Catalogs: What Users and Librarians Want." Dublin, OH: OCLC, 2009. https://web.archive.org/web/20110101163633/http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf.

Carr, Patrick L. "Reimagining the Library as a Technology: An Analysis of Ranganathan's Five Laws of Library Science within the Social Construction of Technology Framework." *The Library Quarterly* 84, no. 2 (April 2014): 152–64. https://doi.org/10.1086/675355.

Centivany, Alissa. "The Dark History of HathiTrust," 2017. https://doi.org/10.24251/HICSS.2017.285.

Chan, Lois Mai. "Inter-Indexer Consistency in Subject Cataloging." *Information Technology & Libraries* 8, no. 4 (December 1989): 349–58.

Coase, R. H. *The Firm, the Market, and the Law*. Chicago, IL: University of Chicago Press, 1988.

ConsenSys. "The Inside Story of the CryptoKitties Congestion Crisis." ConsenSys Media, February 20, 2018. https://media.consensys.net/the-inside-story-of-the-cryptokitties-congestion-crisis-499b35d119cc.

"Contracts: Events." Solidity 0.5.3 documentation, February 11, 2019. https://solidity.readthedocs.io/en/v0.5.3/contracts.html#events.

"Contropedia." Accessed November 23, 2018. http://contropedia.net/.

Conway, Melvin. "How Do Committees Invent?" *Datamation* 14, no. 4 (1968): 28–31.

Costello, Jean. "An Inflection Point for American Public Libraries." *In the Library with the Lead Pipe*, 2009. http://www.inthelibrarywiththeleadpipe.org/2009/an-inflection-point-for-american-public-libraries/.

Crain, Brian Fabian. "Glen Weyl: Radical Markets – Uprooting Capitalism and Democracy for a Just Society." Accessed September 5, 2018. https://soundcloud.com/epicenterbitcoin/eb-251.

Creative Commons. "CC0 1.0 Universal." Accessed March 6, 2019. https://creativecommons.org/publicdomain/zero/1.0/.

———. "Creative Commons: When We Share, Everyone Wins." Creative Commons, 2018. https://creativecommons.org/.

Cronin, Christopher, Mary S. Laskowski, Ellen K. W. Mueller, and Beth E. Snyder. "Strength in Numbers: Building a Consortial Cooperative Cataloging Partnership." *Library Resources & Technical Services* 61, no. 2 (April 2017): 102–16.

Cutter, Charles A. *Rules for a Printed Dictionary Catalog*. 1st ed. Washington: Government Print Office, 1876.

Dannen, Chris. *Introducing Ethereum and Solidity: Foundations of Cryptocurrency and Blockchain Programming for Beginners.* [Place of publication not identified]: Apress, 2017.

DARPA. "ARPANET and the Origins of the Internet." Defense Advanced Research Projects Agency. Accessed December 1, 2018. https://www.darpa.mil/about-us/timeline/arpanet.

De Gennaro, Richard. "From Monopoly to Competition: The Changing Library Network Scene." *Library Journal* 104 (January 1, 1979): 1215–17.

digitalnotvir. "How Reputation Scores Are Calculated - the Details Explained with Simple Math." Steemit, August 24, 2016. https://steemit.com/steemit/@digitalnotvir/how-reputation-scores-are-calculated-the-details-explained-with-simple-math.

Doctorow, Cory, Amanda Palmer, and Neil Gaiman. *Information Doesn't Want to Be Free: Laws for the Internet Age*. San Francisco, CA: McSweeney's, 2014.

Donnelly, F. K. "Catalogue Wars and Classification Controversies." *Canadian Library Journal* 43 (August 1986): 245–47.

Doran, Claire, and Cheryl Martin. "Measuring Success in Outsourced Cataloging: A Data-Driven Investigation." *Cataloging & Classification Quarterly* 55, no. 5 (July 4, 2017): 307–17. https://doi.org/10.1080/01639374.2017.1317309.

———. "Measuring Success in Outsourced Cataloging: A Data-Driven Investigation." *Cataloging & Classification Quarterly* 55, no. 5 (July 4, 2017): 307–17. https://doi.org/10.1080/01639374.2017.1317309.

Dunkle, Clare B. "Outsourcing the Catalog Department: A Meditation Inspired by the Business and Library Literature." *The Journal of Academic Librarianship* 22, no. 1 (January 1996): 33–44. https://doi.org/10.1016/S0099-1333(96)90032-4.

Dworkin, Morris J. "SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions." National Institute of Standards and Technology, July 2015. https://doi.org/10.6028/NIST.FIPS.202.

Edgar, Brian D, and John Willinsky. "A Survey of Scholarly Journals Using Open Journal Systems." *Scholarly and Research Communication* 1, no. 2 (June 14, 2010). https://doi.org/10.22230/src.2010v1n2a24.

EditMe. "Wikipedia's Content Moderation: Proof That Wikis Work," August 28, 2009. https://www.editme.com/wikipediaflaggedrevisions.

Eisenberg, Nancy, and Janet Strayer, eds. *Empathy and Its Development*. Cambridge Studies in Social and Emotional Development. Cambridge, NY: Cambridge University Press, 1987.

Enis, Matt. "SJSU-Led Team Explores Blockchain." *Library Journal* 143, no. 1 (January 2018): 16–20.

Ensign, David. "Copyright Corner: Blockchain and Copyright." *Kentucky Libraries* 82, no. 3 (2018): 4–5.

Ethereum. "Solidity — Solidity 0.4.21 Documentation," 2017. https://solidity.readthedocs.io/en/v0.4.21/#.

Ethereum Foundation. *Devcon2: Ethereum in 25 Minutes*. Ethereum Developer Conference, 2016 September 19-21. Shanghai, China, 2016. https://www.youtube.com/watch?v=66SaEDzlmP4.

Ethereum JavaScript API. JavaScript. 2014. Reprint, ethereum, 2019. https://github.com/ethereum/web3.js.

Ethereum.org. “About the Ethereum Foundation.” Ethereum.org, 2018. https://www.ethereum.org/foundation.

Etherscan. “Ethereum ChainData Size Growth - Fast Sync.” Etherscan. Accessed February 4, 2019. https://etherscan.io/chart2/chaindatasizefast.

———. “Ethereum Gas Price Tracker.” Etherscan: The Ethereum Block Explorer, 2019. https://etherscan.io/gasTracker.

Falkon, Samuel. “The Story of the DAO — Its History and Consequences.” *Medium* (blog), December 24, 2017. https://medium.com/swlh/the-story-of-the-dao-its-history-and-consequences-71e6a8a551ee.

Farnel, Sharon. “Making Meaning Together: Decolonizing Descriptions in Local Digitized Collections.” University of Alberta Libraries, 2018. https://doi.org/10.7939/R31G0J933.

Findlay, Cassie. “Participatory Cultures, Trust Technologies and Decentralisation: Innovation Opportunities for Recordkeeping.” *Archives and Manuscripts* 45, no. 3 (September 2, 2017): 176–90. https://doi.org/10.1080/01576895.2017.1366864.

Fisher, Julian H. “Scholarly Publishing Re-Invented: Real Costs and Real Freedoms.” *The Journal of Electronic Publishing* 11, no. 2 (May 30, 2008). https://doi.org/10.3998/3336451.0011.204.

Flock, Fabian, David Laniado, Felix Stadthaus, and Maribel Acosta. “Towards Better Visual Tools for Exploring Wikipedia Article Development --- The Use Case of Gamergate Controversy.” In *AAAI Workshop - Technical Report*, 48–55. Oxford, UK: AI Access Foundation, 2015. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/viewFile/10656/10561.

Forrester, Amy. “Barriers to Open Access Publishing: Views from the Library Literature.” *Publications* 3, no. 4 (September 3, 2015): 190–210. https://doi.org/10.3390/publications3030190.

Frederick, Donna E. “Metadata Specialists in Transition: From MARC Cataloging to Linked Data and BIBFRAME (Data Deluge Column).” *Library Hi Tech News* 33, no. 4 (June 6, 2016): 1–5. https://doi.org/10.1108/LHTN-03-2016-0015.

Frederick, Donna Ellen. “Library Data: What Is It and What Changes Do Libraries Need to Make? (The Data Deluge Column).” *Library Hi Tech News* 34, no. 8 (October 2, 2017): 1–7. https://doi.org/10.1108/LHTN-06-2017-0044.

Free Software Foundation. “What Is Free Software and Why Is It so Important for Society?,” 2019. https://www.fsf.org/about/what-is-free-software.

Gartner, Richard. “Democratizing Metadata.” In *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, 97–106. New York, NY: Springer Berlin Heidelberg, 2016.

Geist, Michael. "Rethinking IP in the TPP: Canadian Government Plays Key Role in Suspending Unbalanced Patent and Copyright Rules." *Michael Geist* (blog), November 11, 2017. http://www.michaelgeist.ca/2017/11/rethinking-ip-in-the-tpp/.

Gencer, Adem Efe, Soumya Basu, Ittay Eyal, Robbert van Renesse, and Emin Gün Sirer. "Decentralization in Bitcoin and Ethereum Networks." *ArXiv:1801.03998 [Cs]*, January 11, 2018. http://arxiv.org/abs/1801.03998.

George, Joey F., and John L. King. "Examining the Computing and Centralization Debate." *Communications of the ACM* 34, no. 7 (July 1, 1991): 62–72. https://doi.org/10.1145/105783.105796.

Ghapanchi, Amir Hossein, Aybuke Aurum, and Graham Low. "A Taxonomy for Measuring the Success of Open Source Software Projects." *First Monday* 16, no. 8 (July 28, 2011). https://doi.org/10.5210/fm.v16i8.3558.

Ginsberg, Debbie. "Blockchain: What It Is, How It's Being Used, and What It Means for the Future of Law Libraries." *AALL Spectrum* 22, no. 1 (2017): 36–39.

Gordon, Rich. "Six Things to Know about Successful Open-Source Software." Northwestern University Knight Lab, July 24, 2013. https://knightlab.northwestern.edu/2013/07/24/six-lessons-on-success-and-failure-for-open-source-software/.

Gorman, Michael. "Implementing Changes in Cataloging Rules." *Library Journal* 112, no. 3 (1987): 110–12.

Gregor, Dorothy, and Carol Mandel. "Cataloging Must Change!" *Library Journal* 116, no. 6 (4/1/1991 1991): 42–47.

Ha, Thu-Huong. "Forbes Deleted a Deeply Misinformed Op-Ed Arguing Amazon Should Replace Libraries." Quartz, July 23, 2018. https://qz.com/1334123/forbes-deleted-an-op-ed-arguing-that-amazon-should-replace-libraries/.

haadcode. *Peer-to-Peer Databases for the Decentralized Web.* JavaScript. 2015. Reprint, OrbitDB, 2019. https://github.com/orbitdb/orbit-db.

Hafner, Katie. "Growing Wikipedia Refines Its 'Anyone Can Edit' Policy." *The New York Times*, June 17, 2006, sec. Technology. https://www.nytimes.com/2006/06/17/technology/17wiki.html.

Hargraves, Ian. "Controversies of Information Discovery." *Knowledge, Technology & Policy* 20, no. 2 (2007): 83.

Harvard Library. "Harvard Library Open Metadata." Harvard Library, 2017. https://emeritus.library.harvard.edu/open-metadata.

HathiTrust Digital Library. "Member Community." Accessed November 20, 2018. https://www.hathitrust.org/community.

———. “Mission and Goals.” Accessed November 20, 2018. https://www.hathitrust.org/mission_goals.

Herther, Nancy K. “Blockchain Technology in the Library.” *Online Searcher*, October 2018.

Hillman, Hunter. “The Case for Ethereum Scalability.” *Connext* (blog), January 18, 2019. https://medium.com/connext/the-case-for-ethereum-scalability-d2a8035f880f.

“History of Wikipedia.” In *Wikipedia*. Accessed February 24, 2019. https://en.wikipedia.org/wiki/History_of_Wikipedia.

Hoffman, Martin L. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge, MA: Cambridge University Press, 2007.

Holley, Edward G. “Reaction to ‘A Brief...’” *Journal of Academic Librarianship* 9, no. 4 (1983): 201–2.

Houk, Gary R. “OCLC Speaks out on Record Nabbing.” *Library Collections, Acquisitions, & Technical Services* 27, no. 3 (2003): 277–79.

Howley, Brendan. “The Razor’s Edge: Blockchain, Ledger Legerdemain, and the Public Library.” *Information Today*, November 2016.

Hoy, Matthew B. “An Introduction to the Blockchain and Its Implications for Libraries and Medicine.” *Medical Reference Services Quarterly* 36, no. 3 (July 3, 2017): 273–79. https://doi.org/10.1080/02763869.2017.1332261.

Hunt, J W, and M D McIlroy. “An Algorithm for Differential File Comparison,” 1976. https://nanohub.org/infrastructure/rappture/export/3582/trunk/gui/src/diff.pdf.

Infura. “Your Access to the Ethereum Network.” Infura - Scalable Blockchain Infrastructure, 2019. https://infura.io.

Internet Archive. “Internet Archive: Projects.” Accessed November 30, 2018. https://archive.org/projects/.

———. “Open Library Data.” Open Library, May 3, 2018. https://archive.org/details/ol_data.

IPFS Implementation in JavaScript. JavaScript. 2014. Reprint, IPFS, 2019. https://github.com/ipfs/js-ipfs.

Ithaka S+R. “Briefing Paper on Progress and Opportunities for HathiTrust,” July 15, 2011. https://www.hathitrust.org/documents/hathitrust-3year-review-2011.pdf.

Jaeger, Paul T., Ursula Gorham, John Carlo Bertot, and Lindsay C. Sarin. “Democracy, Neutrality, and Value Demonstration in the Age of Austerity.” *The Library Quarterly* 83, no. 4 (October 2013): 368–82. https://doi.org/10.1086/671910.

Jisc. “OpenDoar: Directory of Open Access Repositories.” Accessed December 14, 2018. http://v2.sherpa.ac.uk/opendoar/.

Jordan, Jay. “OCLC 1998-2008: Weaving Libraries into the Web.” *Journal of Library Administration* 49, no. 7 (October 2009): 727–62.

Jupp, Victor. “Purposive Sampling.” In *The SAGE Dictionary of Social Research Methods*. London, United Kingdom: SAGE Publications, Ltd, 2006. https://doi.org/10.4135/9780857020116.

Kajberg, Leif. “Revisiting the Concept of the Political Library in the World of Web 2.0 Technologies.” *Progressive Librarian*, no. 36/37 (Summer/Fall2011 2011): 30–41.

Kascus, Marie, Dawn Hale, and Association for Library Collections & Technical Services, eds. *Outsourcing Cataloging, Authority Work, and Physical Processing: A Checklist of Considerations*. Chicago: American Library Association, 1995.

Kasireddy, Preethi. “How Does Ethereum Work, Anyway?” *Preethi Kasireddy* (blog), September 27, 2017. https://medium.com/@preethikasireddy/how-does-ethereum-work-anyway-22d1df506369.

Kennedy, Ian. “Infura Manages Most of the Ethereum (ETH) Network’s Nodes, but This Could Affect Decentralization.” Crypto Daily Gazette, July 30, 2018. https://cryptodailygazette.com/2018/07/30/infura-manages-most-of-the-ethereum-eth-networks-nodes-but-this-could-affect-the-decentralization/.

Kent, Allen, Harold Lancour, and William Z. Nasri, eds. *Encyclopedia of Library and Information Science*. Vol. 45. 73 vols. New York, NY: M. Dekker, 1968.

Kilgour, Frederick G. “Computer-Based Systems, A New Dimension to Library Cooperation.” *College & Research Libraries* 34, no. 2 (March 1, 1973): 137–43. https://doi.org/10.5860/crl_34_02_137.

———. “Report to the Committee of Librarians of the Ohio College Association.” In *Collected Papers of Frederick G. Kilgour, OCLC Years*, edited by Lois L. Yoakam, 1. Dublin, OH: OCLC Online Computer Library Center, 1984.

———. “The Economic Goal of Library Automation.” *College & Research Libraries* 30, no. 4 (July 1, 1969): 307–11. https://doi.org/10.5860/crl_30_04_307.

King, Sunny, and Scott Nadal. “PPCoin: Peer-to-Peer Crypto-Currency with Proof-of-Stake,” August 19, 2012. Internet Archive. https://archive.org/details/PPCoinPaper.

Koch, Theodore Wesley. *On University Libraries*. Evanston, IL: Northwestern University, 1924. https://archive.org/embed/onuniversitylibr00kochuoft.

Kortekaas, Simone, and Bianca Kramer. “Thinking the Unthinkable - Doing Away with the Library Catalogue.” *Insights: The UKSG Journal* 27, no. 3 (November 2014): 244–48.

Koufogiannakis, Denise, Ann Carr-Wiggin, Sheila Laroque, Ian Bigelow, Kayla Lar-Son, Debbie Feisst, and Sharon Farnel. “Decolonizing Description: Changing Metadata in Response to the Truth and Reconciliation Commission.” University of Alberta Libraries, 2017. https://doi.org/10.7939/R3MS3KF68.

Krishnamurthy, Sandeep. “Cave or Community? An Empirical Examination of 100 Mature Open Source Projects.” *First Monday* 7, no. 6 (2002). https://firstmonday.org/ojs/index.php/fm/article/view/1477/1392.

Kuhlthau, Carol Collier. *Seeking Meaning: A Process Approach to Library and Information Services*. 2nd ed. Westport, Conn: Libraries Unlimited, 2004.

Laat, Paul B. de. “From Open-Source Software to Wikipedia: ‘Backgrounding’ Trust by Collective Monitoring and Reputation Tracking.” *Ethics and Information Technology* 16, no. 2 (June 2014): 157–69. https://doi.org/10.1007/s10676-014-9342-9.

Labazova, Olga, Tobias Dehling, and Ali Sunyaev. “From Hype to Reality: A Taxonomy of Blockchain Applications.” In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 4555–64. Honolulu, HI, 2019. https://hdl.handle.net/10125/59893.

Lankes, R. David. “Libraries Are Obsolete.” *OLA Quarterly* 18, no. 2 (2012): 12–17. https://doi.org/10.7710/1093-7374.1354.

Lemieux, Victoria Louise. “Trusting Records: Is Blockchain Technology the Answer?” *Records Management Journal* 26, no. 2 (July 18, 2016): 110–39. https://doi.org/10.1108/RMJ-12-2015-0042.

Leonard, Lawrence E. “Inter-Indexer Consistency Studies, 1954-1975: A Review of the Literature and Summary of Study Results.” *Occasional Papers (University of Illinois at Urbana-Champagne. Graduate School of Library Science)*, no. 131 (December 1977): 1–54.

Library Freedom Project. “Curriculum for Teaching All about Tor.” Accessed November 28, 2018. https://libraryfreedomproject.org/allabouttor/.

———. “Library Freedom Institute.” Accessed November 28, 2018. https://libraryfreedomproject.org/lfi/.

———. “Tor Exit Relays in Libraries: A New LFP Project,” July 28, 2015. https://libraryfreedomproject.org/torexitpilotphase1/.

Library of Congress. “BIBFRAME - Bibliographic Framework Initiative.” Accessed April 17, 2019. https://www.loc.gov/bibframe/.

———, ed. *The Card Catalog: Books, Cards, and Literary Treasures*. San Francisco, CA: Chronicle Books, 2017.

Little, Will. "A Primer on Proof-of-Stake and Why It Matters for the Future of Blockchains." Hacker Noon, May 14, 2018. https://hackernoon.com/a-primer-on-proof-of-stake-and-why-it-matters-for-the-future-of-blockchains-48764373d4b1.

Lopata, Cynthia. "Integrated Library Systems. ERIC Digest." ERIC Clearinghouse on Information Resources, April 1995. https://eric.ed.gov/?id=ED381179.

Lubetzky, Seymour. "Development of Cataloging Rules." *Library Trends* 2 (1953): 179–86.

Lynn, Barry C. *Cornered: The New Monopoly Capitalism and the Economics of Destruction*. Hoboken, NJ: John Wiley & Sons, 2010.

Mainka, Agnes, Sarah1 Hartmann, Lisa1 Orszullok, Isabella1 Peters, Anika1 Stallmann, and Wolfgang G.1 Stock. "Public Libraries in the Knowledge Society: Core Services of Libraries in Informational World Cities." *Libri: International Journal of Libraries & Information Services* 63, no. 4 (November 2013): 295–319. https://doi.org/10.1515/libri-2013-0024.

Malmsten, Martin. "Making a Library Catalogue Part of the Semantic Web." In *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 146–50. Berlin, Germany: Universitätsverlag Göttingen, 2008. http://www.oapen.org/download?type=document&docid=610315#page=162.

Maneewongvatana, Songrit, and Suthathip Maneewongvatana. "A Combined Approach to Minimizing Vandalisms on Wikipedia." In *2010 10th International Symposium on Communications and Information Technologies*, 395–400. Tokyo, Japan: IEEE, 2010. https://doi.org/10.1109/ISCIT.2010.5664872.

Mann, Thomas. "'Cataloging Must Change!' And Indexer Consistency Studies: Misreading the Evidence at Our Peril." *Cataloging & Classification Quarterly* 23, no. 3–4 (March 28, 1997): 3–45. https://doi.org/10.1300/J104v23n03_02.

Markey, Karen. "The Online Library Catalog: Paradise Lost and Paradise Regained?" *D-Lib Magazine* 13, no. 1/2 (January 2007). https://doi.org/10.1045/january2007-markey.

Marschak, Thomas. "On the Comparison of Centralized and Decentralized Economies." *The American Economic Review* 59, no. 2 (1969): 525–32.

McTernan, John. "Don't Mourn the Loss of Libraries – the Internet Has Made Them Obsolete." *The Telegraph*, March 29, 2016. https://www.telegraph.co.uk/opinion/2016/03/29/dont-mourn-the-loss-of-libraries--the-internet-has-made-them-obs/.

MetaMask. "MetaMask." Accessed February 4, 2019. https://metamask.io/.

Miksa, Shawne D. "You Need My Metadata: Demonstrating the Value of Library Cataloging." *Journal of Library Metadata* 8, no. 1 (April 9, 2008): 23–36. https://doi.org/10.1300/J517v08n01_03.

Miller, Robert A. “Centralization versus Decentralization.” *ALA Bulletin* 33, no. 2 (February 1939): 75-79,134-135.

Mimno, David, Gregory Crane, and Alison Jones. “Hierarchical Catalog Records: Implementing a FRBR Catalog.” *D-Lib Magazine* 11, no. 10 (October 2005). https://doi.org/10.1045/october2005-crane.

MIT Technology Review. “In Blockchain We Trust.” *MIT Technology Review* (blog), April 9, 2018. https://medium.com/mit-technology-review/in-blockchain-we-trust-1cafe1c914b2.

Möller, Johanna, and M. Bjørn Von Rimscha. “(De)Centralization of the Global Informational Ecosystem.” *Media and Communication* 5, no. 3 (September 22, 2017): 37. https://doi.org/10.17645/mac.v5i3.1067.

Mugridge, Rebecca, ed. *Cooperative Cataloging: Shared Effort for the Benefit of All*. London, UK: Routledge, 2012.

Nakamoto, Satoshi. “Bitcoin: A Peer-to-Peer Electronic Cash System,” October 31, 2008. https://bitcoin.org/bitcoin.pdf.

NEO Team. “NEO Smart Economy,” 2019. https://neo.org/.

OCLC. “Mergers and Acquisitions,” 2017. https://www.oclc.org/en/about/finance/mergers.html.

———. “OCLC Prints Last Library Catalog Cards.” OCLC.org, October 1, 2015. https://www.oclc.org/en/news/releases/2015/201529dublin.html.

———. “OCLC Technology.” OCLC, February 8, 2019. https://www.oclc.org/en/technology.html.

———. “Policy for Use and Transfer of WorldCat Records [Archived],” November 19, 2008. https://web.archive.org/web/20081203043912/http://www.oclc.org:80/worldcat/catalog/policy/recordusepolicy.pdf.

———. “What Is WorldCat?” WorldCat.org, 2019. https://www.worldcat.org/whatis/default.jsp.

———. “WorldCat Help - Item Details,” 2015. https://www.oclc.org/support/help/worldcat/Content/Itemdetails/item_details.htm.

———. “WorldCat Search API (Web Service).” WorldCat, 2019. https://www.worldcat.org/affiliate/tools?atype=wcapi.

Oder, Norman. “BiblioCommons:. A Breakthrough?” *Library Journal* 133, no. 13 (August 15, 2008): 14–14.

Open Access Max-Planck-Gesellschaft. “Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities,” October 22, 2003. https://openaccess.mpg.de/Berlin-Declaration.

Open Archives Initiative. "Open Archives Initiative Protocol for Metadata Harvesting." Accessed December 14, 2018. https://www.openarchives.org/pmh/.

Open Source Initiative. "GNU General Public License Version 3.0." Accessed February 23, 2019. https://opensource.org/licenses/GPL-3.0.

———. "The Open Source Definition," March 22, 2007. https://opensource.org/osd.

Osborn, Andrew D. "The Crisis in Cataloging." *The Library Quarterly: Information, Community, Policy* 11, no. 4 (1941): 393–411.

Ostrom, Elinor, James Walker, and Roy Gardner. "Covenants with and without a Sword: Self-Governance Is Possible." *American Political Science Review* 86, no. 02 (June 1992): 404–17. https://doi.org/10.2307/1964229.

Palau, Albert. "Analyzing the Hardware Requirements to Be an Ethereum Full Validated Node." *Medium* (blog), September 24, 2018. https://medium.com/coinmonks/analyzing-the-hardware-requirements-to-be-an-ethereum-full-validated-node-dc064f167902.

Pãnescu, Adrian-Tudor, and Vasile Manta. "Smart Contracts for Research Data Rights Management over the Ethereum Blockchain Network." *Science & Technology Libraries* 37, no. 3 (July 3, 2018): 235–45. https://doi.org/10.1080/0194262X.2018.1474838.

Patton, Glenn. "OCLC's Long Association with Less-Than-Full Cataloging." *Technical Services Quarterly* 9, no. 2 (February 12, 1992): 21–30. https://doi.org/10.1300/J124v09n02_04.

Pattuelli, M. Cristina. "Modeling a Domain Ontology for Cultural Heritage Resources: A User-Centered Approach." *Journal of the American Society for Information Science and Technology* 62, no. 2 (February 2011): 314–42. https://doi.org/10.1002/asi.21453.

Peer-to-Peer Databases for the Decentralized Web. JavaScript. 2015. Reprint, OrbitDB, 2019. https://github.com/orbitdb/orbit-db.

Pink, Daniel H. *Drive: The Surprising Truth about What Motivates Us*. New York, NY: Riverhead, 2011.

Pomerantz, Jeffrey, and Robin Peek. "Fifty Shades of Open." *First Monday* 21, no. 5 (April 12, 2016). https://doi.org/10.5210/fm.v21i5.6360.

Posner, Eric A, and E. Glen Weyl. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton, NJ: Princeton University Press, 2018.

Price, Gary. "III Drops OCLC Suit, Will Absorb SkyRiver." LJ infoDOCKET, March 4, 2013. http://www.infodocket.com/2013/03/04/innovative-interfaces-integrates-all-skyriver-services-and-withdraws-antitrust-lawsuit-against-oclc/.

ProQuest. "Shelf Ready Services." ProQuest. Accessed May 24, 2019. https://www.proquest.com/products-services/Shelf-Ready-Services.html.

Public Knowledge Project. “OJS Stats.” Accessed November 28, 2018. https://pkp.sfu.ca/ojs/ojs-usage/ojs-stats/.

“Purposive Sampling.” In *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE Publications, Inc., 2008. https://doi.org/10.4135/9781412963909.n349.

Rahman, H Kalilur, and J Dominic. “An Analytical Study of Online Public Access Catalogues in Comparison with Features of Amazon and Google: A Checklist Approach.” *Asian Journal of Information Science & Technology* 2, no. 1 (2012): 17–23.

Ramaswamy, Lakshmish, Raga Sowmya Tummalapenta, Kang Li, and Calton Pu. “A Content-Context-Centric Approach for Detecting Vandalism in Wikipedia.” In *9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 115–22. Austin, TX, 2013.

Ranganathan, S. R. *Colon Classification*. 6th ed. New Delhi, India: Ess Ess Publications, 2007.

———. *The Five Laws of Library Science*. Madras: The Madras Library Association, 1931. https://babel.hathitrust.org/cgi/pt?id=mdp.39015073883822.

———. *Theory of Library Catalogue*. London: Madras Library Association, 1938.

Ray, James. “Sharding Introduction R&D Compendium.” Ethereum Wiki, June 13, 2018. https://github.com/ethereum/wiki/wiki/Sharding-introduction-R&D-compendium.

Raymond, Eric S. “The Cathedral and the Bazaar,” May 5, 2000. https://web.archive.org/web/19990224193551/http://www.tuxedo.org:80/~esr/writings/cathedral-bazaar/cathedral-bazaar.ps.

———. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Rev. ed. Cambridge, MA: O’Reilly, 2001.

Rosenblatt, Bill. “Can Blockchain Disrupt the E-Book Market? Two Startups Will Find Out.” Forbes. Accessed August 30, 2018. https://www.forbes.com/sites/billrosenblatt/2018/08/18/can-blockchains-disrupt-the-e-book-market-two-startups-will-find-out/.

Rushkoff, Douglas. *Program or Be Programmed: Ten Commands for a Digital Age*. Edited by Leland Purvis. Berkeley, CA: Counterpoint, 2011.

Ryan, Richard M., and Edward L. Deci. “Self-Regulation and the Problem of Human Autonomy: Does Psychology Need Choice, Self-Determination, and Will?” *Journal of Personality* 74, no. 6 (December 2006): 1557–86. https://doi.org/10.1111/j.1467-6494.2006.00420.x.

Salas-Tull, Laura, and Jacque Halverson. “Subject Heading Revision: A Comparative Study.” *Cataloging & Classification Quarterly* 7, no. 3 (June 4, 1987): 3–12. https://doi.org/10.1300/J104v07n03_02.

San Jose State University. “Blockchains for the Information Profession.” Blockchains for Libraries. Accessed December 21, 2018. https://ischoolblogs.sjsu.edu/blockchains/.

Sardan, Thibaut. “What Is a Light Client and Why You Should Care?” Blockchain Infrastructure for the Decentralised Web, July 26, 2018. https://www.parity.io/what-is-a-light-client/.

Schuitema, Joan E. “The Future of Cooperative Cataloging: Curve, Fork, or Impasse?” *Cataloging & Classification Quarterly* 48, no. 2–3 (February 10, 2010): 258–70. https://doi.org/10.1080/01639370903536088.

Schultz-Jones, Barbara, Karen Snow, Shawne Miksa, and Richard L. Hasenyager. “Historical and Current Implications of Cataloguing Quality for Next-Generation Catalogues.” *Library Trends* 61, no. 1 (2012): 49–82. https://doi.org/10.1353/lib.2012.0028.

Schweik, Charles M, and Robert C English. *Internet Success: A Study of Open-Source Software Commons*. Cambridge, MA: MIT Press, 2012.

Seikel, Michele, and Thomas Steele. “How MARC Has Changed: The History of the Format and Its Forthcoming Relationship to RDA.” *Technical Services Quarterly* 28, no. 3 (May 19, 2011): 322–34. https://doi.org/10.1080/07317131.2011.574519.

Seiler, Lauren, and Thomas Surprenant. “When We Get the Libraries We Want, Will We Want the Libraries We Get?” *Wilson Library Bulletin* 65, no. 10 (June 1991): 29–31, 152,157.

Sen, Rijurekha, Sohaib Ahmad, Amreesh Phokeer, Zaid Ahmed Farooq, Ihsan Ayyub Qazi, David Choffnes, and Krishna P. Gummadi. “Inside the Walled Garden: Deconstructing Facebook’s Free Basics Program.” *ACM SIGCOMM Computer Communication Review* 47, no. 5 (October 25, 2017): 12–24. https://doi.org/10.1145/3155055.3155058.

Simon, Nina. *The Art of Relevance*. Santa Cruz, CA: Museum 2.0, 2016.

Simon, Nina, Lauren Benetua, and Shelley Bernstein. “OF/BY/FOR ALL,” 2018. https://www.ofbyforall.org/.

SJSU iSchool. “Blockchains for the Information Profession.” Blockchains for Libraries, 2017. https://ischoolblogs.sjsu.edu/blockchains/.

Smith, Vicki Toy. “Outsourcing Cataloging: An Evaluation.” *Journal of Educational Media & Library Sciences* 34, no. 4 (1997): 380–95.

Sovrin Foundation. “Sovrin.” Sovrin, 2019. https://sovrin.org/.

Sowerby, James de Carle, and Edward Lear. *Tortoises, Terrapins, and Turtles Drawn from Life,*. London, Paris, and Frankfort: H. Sotheran, J. Baer & co., 1872.

Spanhoff, Elisabeth de Rijk. “Principle Issues: Catalog Paradigms, Old and New.” *Cataloging & Classification Quarterly* 35, no. 1–2 (December 2002): 37–59. https://doi.org/10.1300/J104v35n01_04.

Stallman, Richard. "The Free Universal Encyclopedia and Learning Resource," December 18, 2000. https://www.gnu.org/encyclopedia/anencyc.txt.

Stanford University. "LOCKSS." Accessed January 17, 2019. https://www.lockss.org/.

Stark, Josh. "Making Sense of 'Cryptoeconomics.'" *L4 Media* (blog), November 16, 2017. https://medium.com/l4-media/making-sense-of-cryptoeconomics-5edea77e4e8d.

———. "Making Sense of Ethereum's Layer 2 Scaling Solutions: State Channels, Plasma, and Truebit." *Medium* (blog), February 12, 2018. https://medium.com/l4-media/making-sense-of-ethereums-layer-2-scaling-solutions-state-channels-plasma-and-truebit-22cb40dcc2f4.

Stinchcombe, Kai. "Blockchain Is Not Only Crappy Technology but a Bad Vision for the Future." *Medium* (blog), April 5, 2018. https://medium.com/@kaistinchcombe/decentralized-and-trustless-crypto-paradise-is-actually-a-medieval-hellhole-c1ca122efdec.

———. "Ten Years in, Nobody Has Come up with a Use for Blockchain." Hacker Noon, December 22, 2017. https://hackernoon.com/ten-years-in-nobody-has-come-up-with-a-use-case-for-blockchain-ee98c180100.

Szilágyi, Péter. "Geth 1.8 – Iceberg." Ethereum Blog, February 14, 2018. https://blog.ethereum.org/2018/02/14/geth-1-8-iceberg%c2%b9/.

Tai, Li-Cheng. "The History of the GNU General Public License," July 4, 2001. https://www.free-soft.org/gpl_history/.

Talaga, Tanya. "Joseph Boyden's Identity Crisis Opens up Questions on Who Is Part of a Community." *Toronto Star*, January 14, 2017. https://www.thestar.com/news/canada/2017/01/14/joseph-boydens-identity-crisis-opens-up-questions-on-who-is-part-of-a-community.html.

Techopedia. "What Is a Cryptographic Hash Function?" Techopedia.com, 2019. https://www.techopedia.com/definition/27410/cryptographic-hash-function.

Tennant, Roy. "Library Catalogs: The Wrong Solution." *Library Journal* 128, no. 3 (February 15, 2003): 28.

Terdiman, Daniel. "Growing Pains for Wikipedia." CNET, December 7, 2005. https://www.cnet.com/news/growing-pains-for-wikipedia/.

Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Rev. and expanded ed. New York, NY: Penguin Books, 2009.

The Tor Project, Inc. "Tor Project: Overview." Accessed November 21, 2018. https://www.torproject.org/about/overview.html.en.

Thornton, John. "Blockchain on the Books: If Trust and Transactions Are a Major Part of an Organisation's Work, It Could Benefit from Secure Shared Records Created Using Blockchain Technology." *Public Finance*, no. 9 (September 2017): 43–43.

Todd, Ryan. "Ethereum Essentials: Node Nuances – The Block." *CnbCrypto* (blog), January 23, 2019. https://cnbcrypto.com/2019/01/ethereum-essentials-node-nuances-the-block/.

Tomaino, Nick. "The Slow Death of the Firm." The Control, October 21, 2017. https://thecontrol.co/the-slow-death-of-the-firm-1bd6cc81286b.

Trail of Bits Blog. "Contract Upgrade Anti-Patterns." *Trail of Bits Blog* (blog), September 5, 2018. https://blog.trailofbits.com/2018/09/05/contract-upgrade-anti-patterns/.

Trosow, Samuel E. "The Commodification of Information and the Public Good." *Progressive Librarian*, no. 43 (2014): 17–29.

University of Michigan. "University of Michigan MARC Catalog Records." Internet Archive, 2010. https://archive.org/details/UniversityOfMichiganMarcCatalogRecords.

University of Toronto. "University of Toronto Catalog." Internet Archive, February 11, 2019. https://archive.org/details/marc_university_of_toronto.

"Use for Blockchain in Libraries." *Blockchains for Libraries* (blog), September 13, 2017. https://ischoolblogs.sjsu.edu/blockchains/blockchains-applied/applications/.

Wang, Hsiao-Wei. "Ethereum Sharding: Overview and Finality." *Hsiao-Wei Wang* (blog), December 27, 2017. https://medium.com/@icebearhww/ethereum-sharding-and-finality-65248951f649.

Whaite, Katharine. "Finding Value in History: Gaining Knowledge by Examining Historical Practices." *Catalogue & Index*, no. 169 (December 2012): 25–29.

Willinsky, John. "Open Journal Systems: An Example of Open Source Software for Journal Management and Publishing." Edited by Scott P. Muir. *Library Hi Tech* 23, no. 4 (December 2005): 504–19. https://doi.org/10.1108/07378830510636300.

Wood, Gavin. "Ethereum: A Secure Decentralised Generalised Transaction Ledger," April 7, 2018. https://github.com/ethereum/yellowpaper.

Yang, Sharon Q., and Melissa A. Hofmann. "The Next Generation Library Catalog: A Comparative Study of the OPACs of Koha, Evergreen, and Voyager." *Information Technology and Libraries* 29, no. 3 (September 1, 2010): 141–50. https://doi.org/10.6017/ital.v29i3.3139.

York, Jeremy. "This Library Never Forgets: Preservation, Cooperation, and the Making of HathiTrust Digital Library." In *Archiving 2009 Final Program and Proceedings*, 5–10. Society for Imaging Science and Technology, 2009.

Zheng, Zibin, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends." In *2017 IEEE International Congress on Big Data*, 557–64. Honolulu, HI: IEEE, 2017. https://doi.org/10.1109/BigDataCongress.2017.85.

Zukerfeld, Mariano. "The Tale of the Snake and the Elephant: Intellectual Property Expansion under Informational Capitalism." *The Information Society* 33, no. 5 (October 20, 2017): 243–60. https://doi.org/10.1080/01972243.2017.1354107.