Wikipedia Knows the Value of what the Library Catalog Forgets

By Kris Joseph, krisj@ualberta.ca

The Version of Record of this manuscript has been published and is available in *Cataloging & Classification Quarterly*, April 2019, https://doi.org/10.1080/01639374.2019.1597005

Abstract

Shifting library catalogs from physical to digital has come at a cost. Catalog records no longer leave traces of their own evolution, which is a loss for librarianship. The subjective nature of information classification warrants self-examination, within which we may see the evolution of practice, debates over attribution and relevance, and how culture is reflected in the systems used to describe it. Wikipedia models what is possible: revision histories and discussion pages function as knowledge generators. A list of unanswerable questions for the modern catalog urges us to construct a new, forward-thinking bibliography that allows us to look backward.

Key words: library catalogs, catalog design, historical analysis, revision history, classification systems

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

No potential conflict of interest was reported by the author.

Introduction

The ability to "read" the evolution of library catalog records can contribute significantly to the work of historians, librarians, and other scholars in the social sciences and the humanities. A catalog entry that records its own history can provide insight into how the resource it represents has been contextualized and characterized. Sadly, the current form of the humble library catalog renders such a record irretrievable.

After a discussion of the subjective and occasionally-controversial nature of cataloging, this paper uses the example of Wikipedia to touch on the kinds of analyses that are enabled when information surrogates record their own history. This is contrasted with the limited possibilities arising from the current form of catalog records. A final exploration of questions we cannot answer using today's library catalog implores us to consider its reinvention.

Background

The first law of library science, according to Ranganathan, is that books are for use.[1] Before use, however, there is discovery: a vast collection of books and other materials is useless unless people are empowered to locate items of interest. Accordingly, cataloging is a critical practice within the field of librarianship.

The work of catalogers has been user-centered since the 19th century. Charles Ammi Cutter's basic rules for book cataloging, published in 1876, deliberately placed the human being up front. Cutter ordered that catalogs should enable users to find books according to author, title, subject or character.[2] Thorough descriptions of resources should make them easy to discover, and Cutter's guideline seems straightforward, but a closer look betrays that perception. What do the words *subject* and *character* mean, and from whose perspective should they be applied? In other words: who is the *user*? Surprisingly, even Cutter and Ranganathan differed on the definition and application of these terms. Unsurprisingly, debate on the topic has never subsided.[3]

Some outside the field of Library and Information Studies (LIS) may think of resource descriptions as objective but, as the previously-mentioned debate suggests, they rarely are. Ask ten catalogers to describe the "aboutness" of a book, and you'll likely get eight different responses.[4] Buckland[5] adds complexity to the scope of subjectivity with his consideration of subject classification's temporal character. He outlines tensions inherent in the practice, noting that catalogers must keep one eye trained on the past (to connect new information to existing discourse) and one eye focused on the future (to anticipate the needs and questions of those who seek information). According to Buckland, the work of library classification is fundamentally problematic: subjects are assigned "in the present" but language, context, and culture continue to evolve after the assignment is made. New names emerge and classification boundaries shift. The social acceptability of terms is variable, and terms once thought to be descriptive (simple examples: "negro" and "Rogues and vagabonds see also Gypsies") can become offensive or inappropriate.

In addition to being subjective, though, the work of cataloging is politically situated. This idea is explored in depth by Bowker & Star,[6] who assert that all classification systems—in LIS and elsewhere—bear the marks of differing world views and bureaucratic struggles. As a simple example, they point to a business directory in California that moved *Alcoholics Anonymous* from "rehabilitation" to "emergency services:" as attitudes towards alcoholism have shifted in Western culture, the classification of the illness has followed suit.

Scholars like Bowker, Star, and Buckland urge us to think critically about the definition and application of library classification systems: humanities-focused studies of how these systems are defined and applied have clear value. To shed light on the implementation of catalogs, then, a scholar might examine the evolution of cataloging rules or study the discourse that takes place within the field.[7] Another opportunity, however, lies in the suggestion to read the output of the classification process. These types of analyses can tease out context in philosophical, political, and

social arenas.[8] This suggestion sparks interest and evokes possibilities, but a casual look at the modern catalog reveals that almost none of this kind of reading can be done.

The remainder of this paper argues that the library catalog should be armed with the tools for enhancing historical literacy and suggests mechanisms that can forge these tools. But first: let us explore an example that bemoans our irreversible move to digital cataloging.

An Example in Analog

An invaluable paper from 2012[9] demonstrates how analysis of catalog 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 Testudinata*,[10] which was later published under the title *Tortoises, Terrapins and Turtles, drawn from life* in 1872.[11] Whaite's investigation was enabled by the presence of a tool from a now-bygone era—physical catalog 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 catalogers 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 catalogers can be seen in the traces of catalog 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.[12]

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.

Sadly, the last physical catalog cards were printed in 2015. The OCLC press release that made the announcement also claimed that "digital library networks" can now connect people to the world's knowledge.[13] Though the claim is not false, the loss of the physical catalog evokes more than mere nostalgia. The digital records that have replaced paper cards exhibit far more fragility than their analog predecessors. Not only can local copies of digital catalog records be easily lost, but updates and changes to them leave no trace of what was removed or the rationale for the alteration. This is a critical design flaw, and an existing system can be used to demonstrate what our current digital catalog records lack.

Wikipedia: Leading by Example

Many modern text and source code management systems feature the ability to view and restore previous versions of digital records. An illustration of revision control as a site of qualitative analysis can be drawn with an example from one of the Internet's most popular web sites: Wikipedia.

Zoë Quinn's 2013 indie game, *Depression Quest*, collided with right-wing tweets and media zealotry to create Gamergate in 2014.[14] The debate raged across the Internet, leading to credible real-world threats to the lives of several prominent female gaming industry professionals. No less controversial was the Wikipedia page for "Gamergate Controversy,"[15] which mirrored the greater debate as article editors struggled to document the event with non-contentious framing and language.

Noting that the volume and frequency of edits to the Wikipedia page's content would have made manual analysis almost impossible,[16] Flock *et. al.* relied on data mining algorithms and multiple visualization tools to analyze the evolution of "Gamergate Controversy." These mechanisms filter Wikipedia's detailed change tracking information, presenting information in

ways that are far more accessible for scholars. One tool, WhoVIS[17] (Figures 1a and 1b), creates

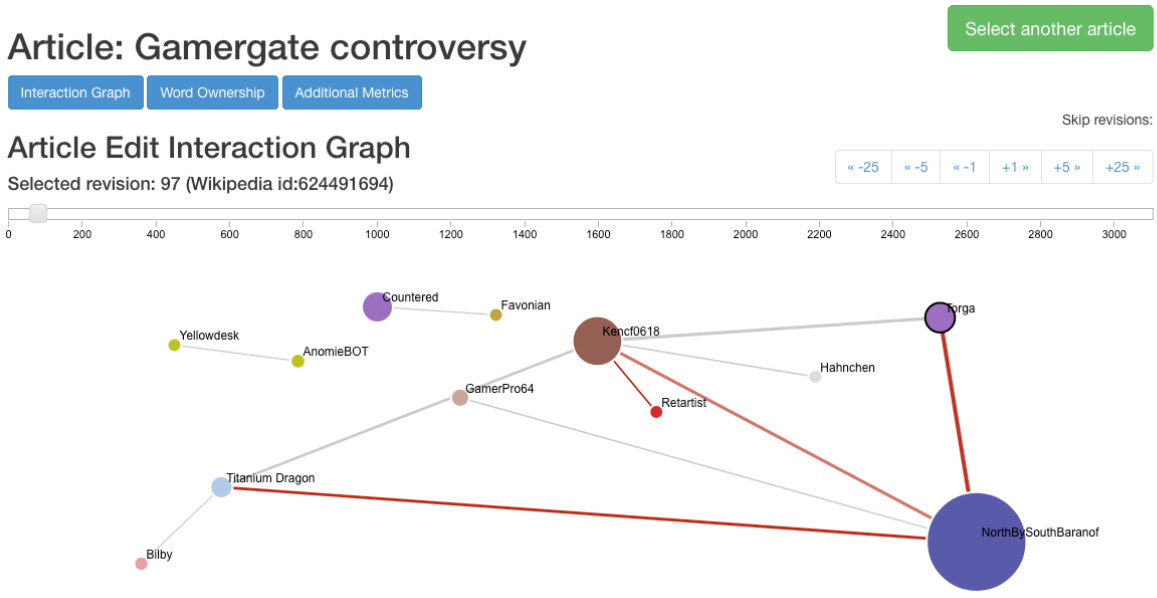

Figure 1a: WhoVIS editor-to-editor interactions for Wikipedia's "Gamergate Controversy" article, summarizing exchanges from revisions 1 to 97.

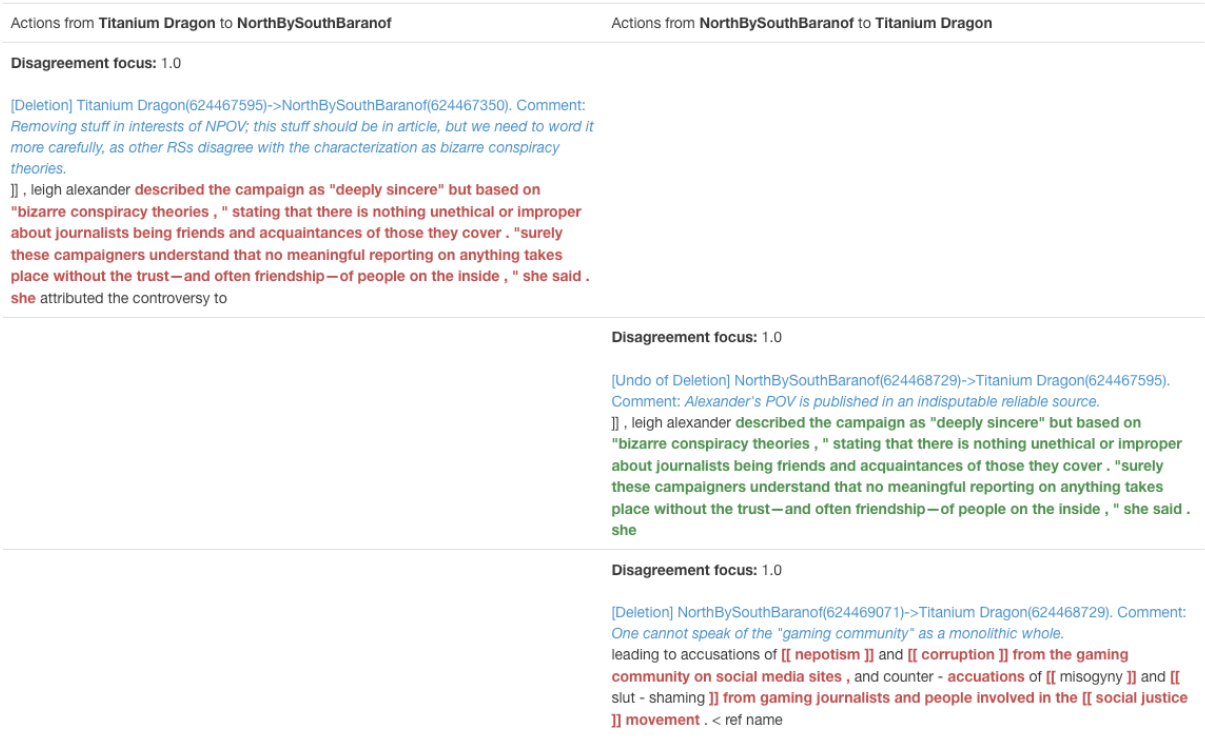

Figure 1b: WhoVIS editor-to-editor interaction details allow users to trace detailed discussions between editors about revisions to Wikipedia articles.

graphs of editor-to-editor interactions, making it easy to trace disagreements over particular pieces of text. By selecting a graph edge, users can see the debate two editors had about a specific word or phrase. The tool also displays the relative volume of edits made by each author. Together, these views help characterize the article's contents by illustrating whose perspectives are debated and whose are dominant. A second tool, Contropedia[18] (Figure 2), uses colored overlays to display portions of text that have been repeatedly edited, deleted, and re-added. A separate dashboard view ranks these portions of text from most controversial to least[19] and displays a histogram of the frequency and volume of edits made to the text. Finally, WhoCOLOR[20] (Figure 3) uses colored overlays to highlight contributions made by individual authors.

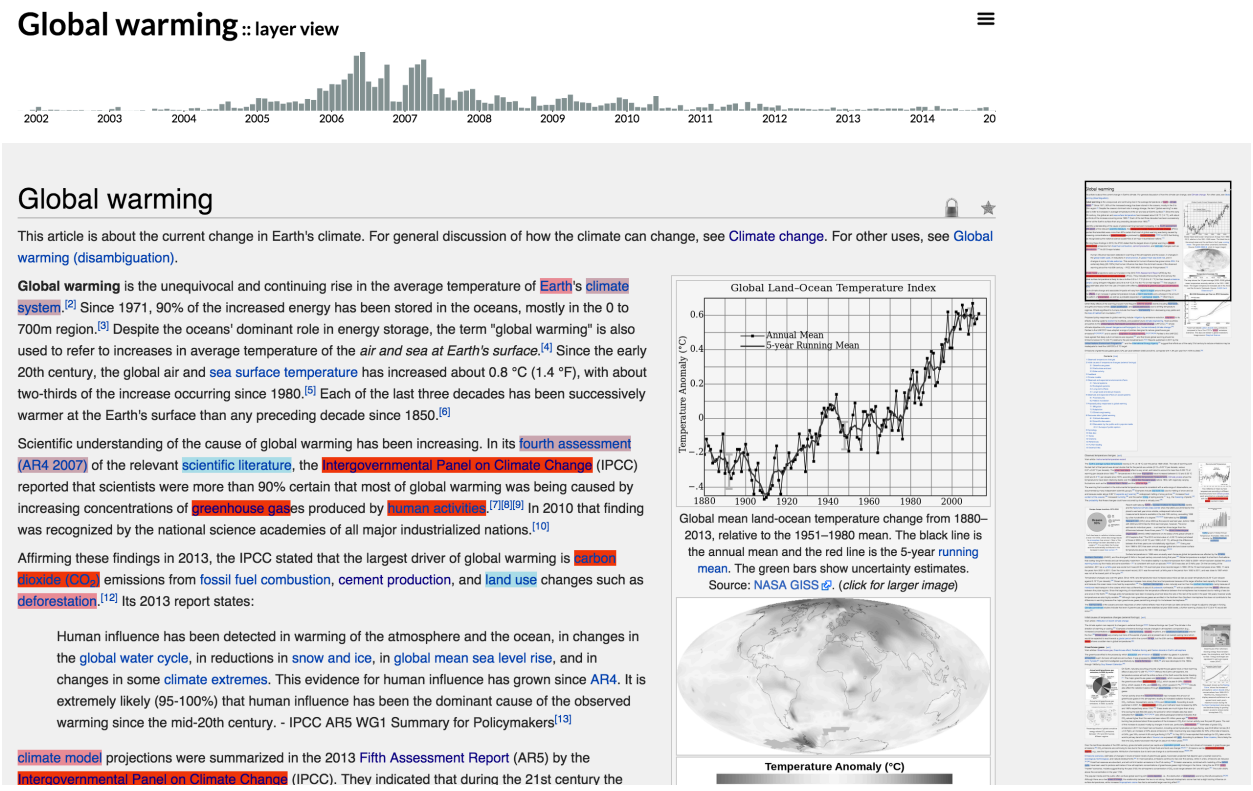

Figure 2: Contropedia illustrates controversial Wikipedia article edits using colored overlays. Clicking on a colored phrase allows users to trace the editor-to-editor debates related to each piece of content.

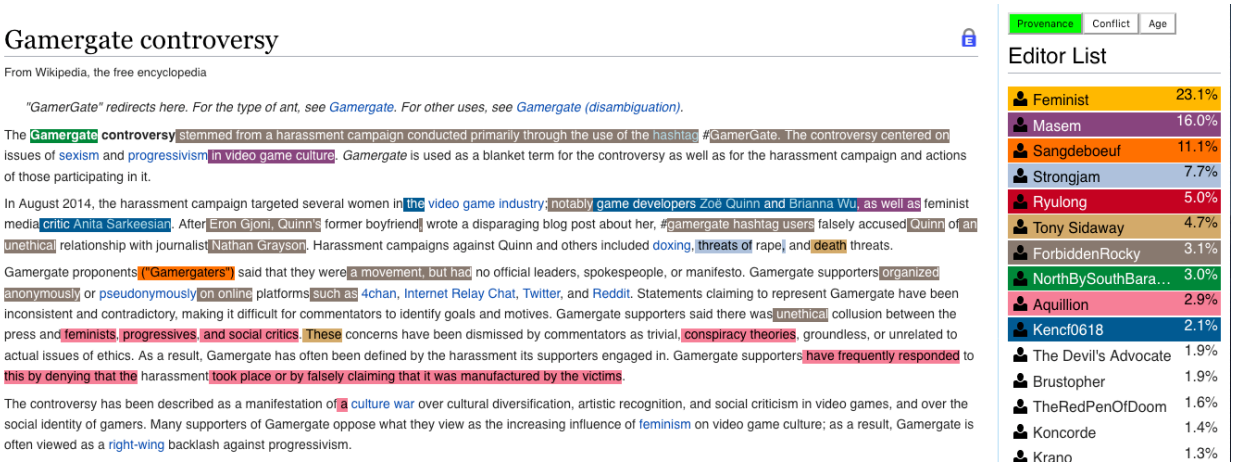

Figure 3: WhoCOLOR highlights contributions to a Wikipedia article by author..

When combined, these tools enabled a close reading of the "Gamergate Controversy" article's mutations over time, allowing researchers to study the interaction of editors who contributed to the article, the content over which they disagreed, and the eventual resolution of those conflicts as the article matured and was moderated.

Wikipedia's usefulness as a self-reflective tool is derived from features that digital library catalogs do not have. These include:

- **View history**: Every Wikipedia article stores a detailed record of previous versions which can be viewed by anyone. Accessible, built-in tools allow for different instances of articles to be compared so that changes over time can be examined. For example, one instance of the word "gamer" had earlier incarnations as "gamer community" and "gamer identity."[21]
- **Talk pages**: each Wikipedia article has an associated discussion page where debates related to the content of an article may take place. In situations where the content of an article is contested, the discussion in these spaces is vigorous and can provide insight into the resolution of differences related to content, tone, use of sources, and more. Like article pages, "talk" pages store revision history.

- **Edit summary**: Each edit to a Wikipedia article requires a brief summary that describes the change being made, adding contextual information through editors' expressed rationale. For example, edits of the phrase *video game culture* on the "Gamergate Controversy" page include reasoning like "those that use the gamer identity are a subset of gaming culture, so it's more accurate to use subculture"[22] and "it's not a movement, but it's unquestionably a group."[23]
- **User pages**: Every Wikipedia contributor is free to create a personal profile page that may include personal interests, information on articles created or edited, awards and achievements, and more. These pages also store edit history.
- **User contributions**: Contributions from each article editor are tracked across Wikipedia, including the number and scope of edits made on every article that they have altered.
- **Open data**: All of Wikipedia's stored data is "open" by default, accessible to anyone with an interest in its content. Of course, this accessibility extends to making contributions to new or existing articles.

The list of features outlined above is not exhaustive. Wikipedia also provides access to its data through open Application Programming Interfaces (APIs); these allow community members to supplement the site with new features that were not envisioned by Wikipedia's designers. For example, thousands of external tools and bots[24] have been created to automate minor edits and corrections, detect and deal with vandalism, and provide contextual links between articles. Visualization tools like WhoCOLOR take advantage of these APIs.[25]

Wikipedia's functionality benefits from its launch in 2001,[26] when revision control systems for text were already well established. By contrast, the first electronic catalog debuted in 1971,[27] nearly commensurate with the first experiments in data networking that would later evolve into the Internet.[28] Had the digital library catalog 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.[29] The economy of computation and storage provided by these tools has made them implicit in the design of modern data management systems, and one of Wikipedia's strengths comes from the features afforded by the way it stores and manages text.

Can the same model of openness and revision tracking be applied to library catalogs? Dare we rebuild the engine from scratch? We should imagine a world in which the library catalog—generally viewed as a secondary source—can serve as a primary source for information about bibliographic debates and the contexts in which they occur. First, however, let's explore what is possible in this domain now.

The open catalog: a positive move

Fortunately, one aspect of Wikipedia's design—openness—has begun to creep into some areas of the digital catalog. Many libraries have opened up their catalog records for public use,[30] enabling some forms of data mining and analysis. Before outlining what can be done with this open data, two pieces of background information are salient.

First, cataloging is highly cooperative. Libraries share their records with one another to reduce the work required to describe resources. In this paradigm, "master" catalog records are drawn from central databases like The Online Computer Library Centre's (OCLC) WorldCat and are then altered or supplemented with local information. The nature of the collaborative cataloging effort also means that catalog records can be modified and resubmitted to the master catalog by libraries that have the appropriate access.[31]

Second, catalog record data has historically been stored in MARC (MAchine-Readable Cataloging) format, which encodes cataloging information into a combination of fixed-length and variable-length fields. The standard was originally created in 1965 as a machine-readable encoding for punch-paper tape,[32] but it has evolved over time. Though there is a gradual push towards RDA (Resource Description and Access) format for bibliographic records, the same field-value paradigm is used, RDA remains tightly integrated with MARC 21,[33] and many libraries still use the MARC format.

The open catalog data that some libraries publish use the MARC format and bear the marks of some of librarianship's cooperative cataloging work. The Harvard University Library catalog[34] is one such source of this data. For the purposes of exploration, roughly three million catalog records were extracted from this collection to explore some possible analyses. Since the work in this section is experimental, based on a subset of the full catalog, it should not be assumed to reflect the true character of the Harvard Library's collection.

Some assessment of catalog quality, defined by measuring the accuracy or completeness of catalog entries, can easily be performed with this type of data; in fact, this type of assessment is already common.[35] Other distant-analysis possibilities for aggregated bibliographic records are possible, however, and a few examples have been visualized.

Figure 4 represents the distribution of Library of Congress top-level subject categories across the selected record set, showing that Philosophy (B), the Social Sciences (H) and Languages and Literature (P) are dominant knowledge categories. Analyses of this type may be used to infer areas of institutional specialization in original cataloging, or to characterize a library's collection based on the shelf classifications of its holdings. In a limited way, this can be done at large scale: Figure 5 splits out the top-level classification of Harvard records according to a selection of libraries listed as a record's original cataloger. This view evinces patterns at a multi-institutional level: note,

for example, that the National Library of Medicine unsurprisingly does most of its original cataloging under the “Medicine” classification.

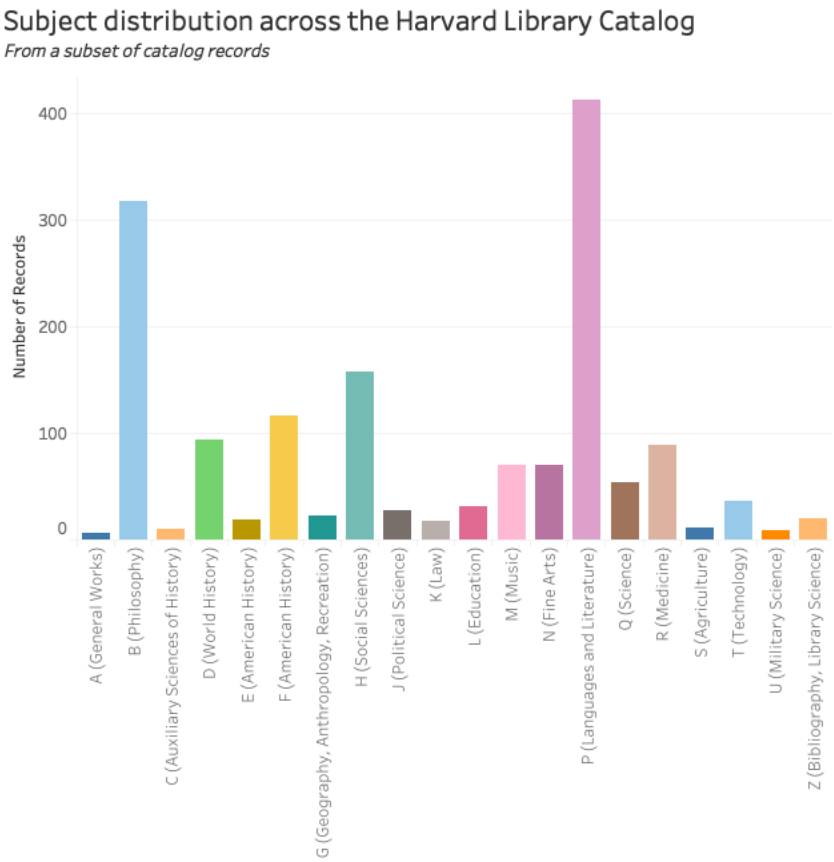

Figure 4: Distribution of top-level LCSH categories across a subset of the Harvard Library catalog's open bibliographic data set.

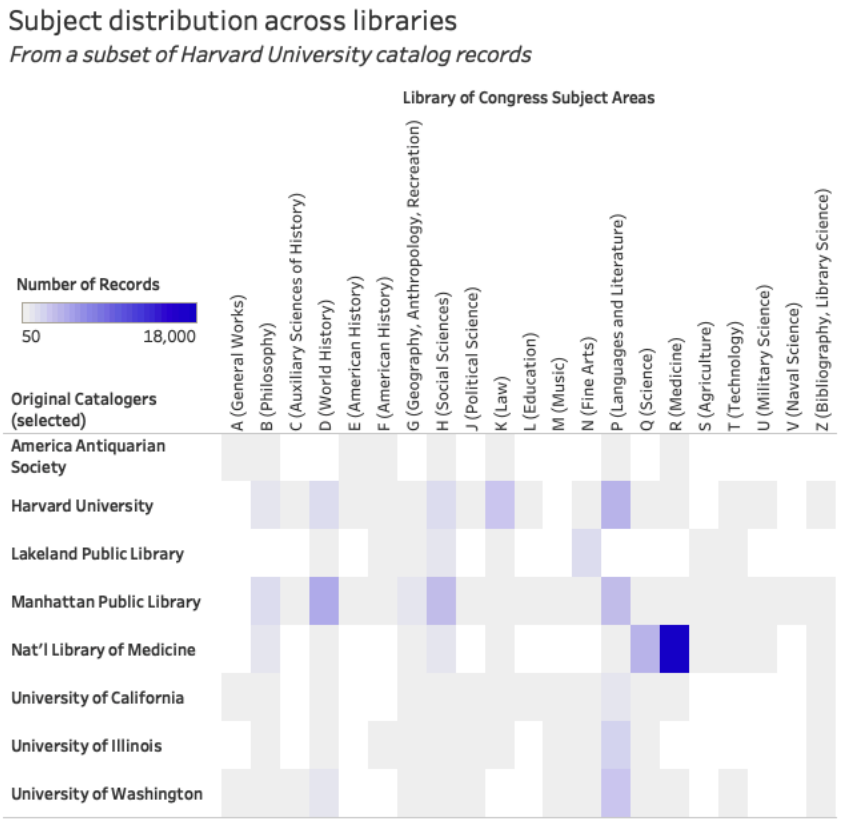

Figure 5: Separating top-level LCSH classifications according to the original cataloger field (MARC 040 subfield a) illuminates patterns of subject specialization within individual library cataloging practices.

The presence of cataloging institution data in MARC records is worth noting, since it is one of the few insights into revision history that the format allows. A library that creates an original catalog 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 cataloger" values. The optional nature of this field makes its use inconsistent, and there is no way to know when a revision was made or the content that was affected, but some inferences about cataloging practices can be drawn from examining this data.

Figure 6 is a filtered graph of data from the 040 field,[36] outlining the connections between institutions that create and modify catalog 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. The dominance of the Library of

Congress (node “dlc”) is clear, as is the prevalence of OCLC-related entities (“oclcq” *et. al.*). Loopbacks (nodes linking back to themselves) are evident for some institutions, meaning that these libraries are editing their own records. This graph only hints at explorations that may be possible: a more rigorous analysis of this type could partially-characterize the nature and distribution of cataloging work, providing insight into the constitution of libraries’ collaborations on resource description. Even more could be done if data were consistent and complete.

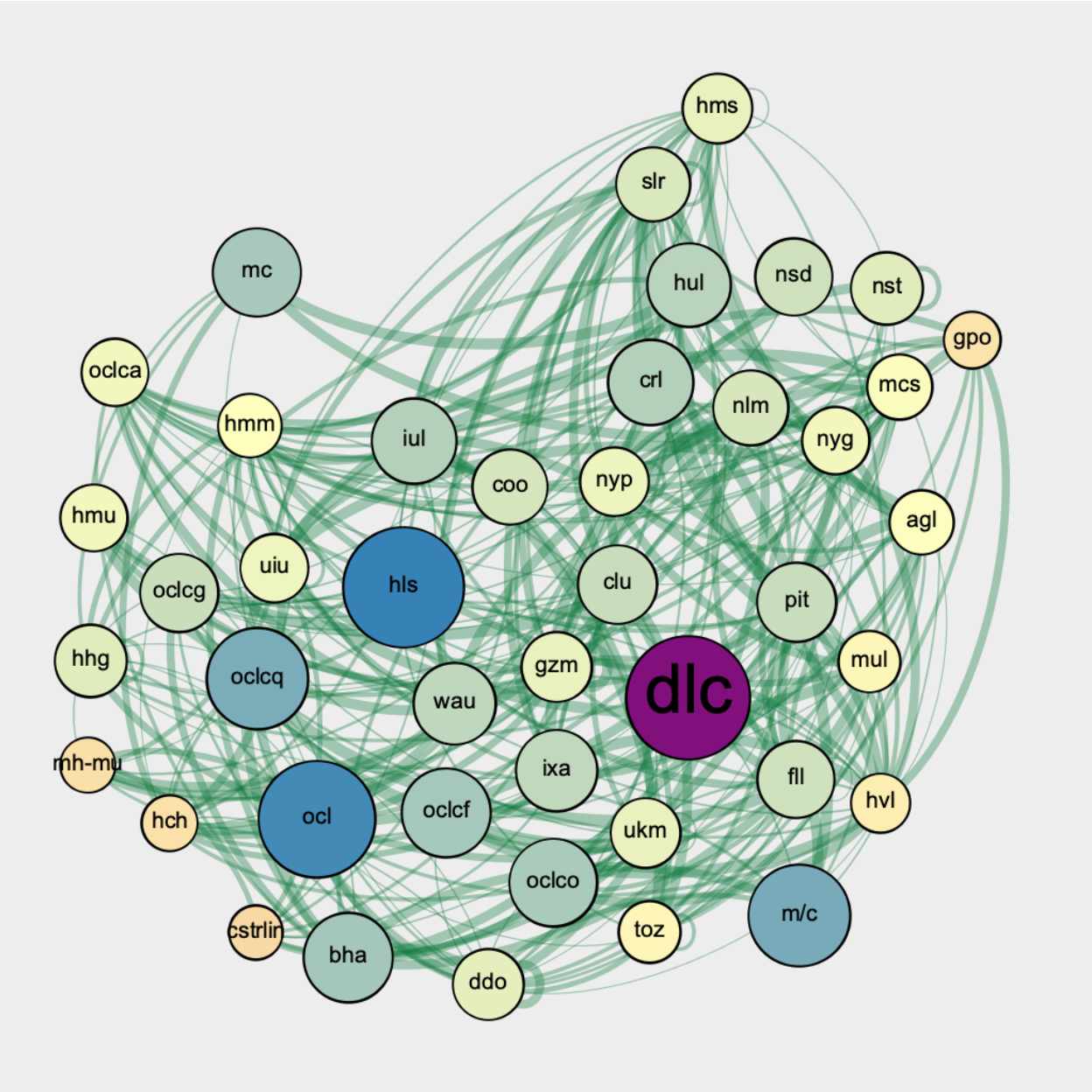

Figure 6: A graph of connections between original and modifying cataloger data from MARC field 040 suggests that, aside from a small number of dominant cataloging organizations, the work of catalogers is distributed quite evenly between cataloging institutions.

People who work with rare books and special collections understand the value of recording provenance for items: this information facilitates rich exploration by scholars of literature, history, and bibliography. However, it would be ideal if this information were tracked implicitly, as it is done on Wikipedia, rather than requesting librarians document changes themselves. An implicit

approach, rather than an optional one, results in the storage of revision information *without* additional labor on the part of catalogers. Rare book librarians would agree that a manual system of change tracking cannot suffice: it has been shown that, outside the scope of rare books and special collections, many librarians do not maintain provenance records for books, and even fewer inscribe this information into catalog records.[37] Imagine what would be possible if full revision history, of the kind offered by Wikipedia, were stored as part of the bibliographic record.

What Could Be

The breadth and complexity of tools created for analysis of Wikipedia matches the scope and scale of system itself: the most popular articles are edited more than 1000 times per month,[38] and the text of English language articles would fill almost 2800 volumes of the Encyclopedia Britannica.[39] Even though WorldCat's collection of more than 440 million bibliographic records is supplemented with new information every two seconds,[40] the wholesale importation of Wikipedia's editing and revision tracking capability may seem an immense effort for minimal benefit. Yet: though the English Wikipedia collection consists of more than 5.8 million articles, less than one tenth of one percent of those have been edited more than 100 times.[41] This pattern is mirrored in the Harvard catalog records examined in the previous section: most are seldom-edited, and the most contentious record's MARC 040 field was altered only seven times. However, the records used in this cursory exploration represent some of the oldest in the Harvard University catalog. Wikipedia's article editing pattern and the "Gamergate Controversy" article suggest that more recent records, related to more controversial subjects, may warrant an investment in new catalog features as a result of the questions they evoke. Take, for example, Bob Woodward's controversial 2018 book, *Fear: Trump in the White House*, whose MARC record has been altered 35 times in its first year.[42] No traces exist for the nature or reasoning of these edits. Imagine the analysis future librarians and historians could perform if individual revisions could be traced using a visual tool

like Contropedia, or if the rationale for each change had been recorded? What value might we extract with a minimal implementation of Wikipedia-like features?

Near the end of Katherine Whaite's article on *A Monograph of the Testudinata*, she writes that "what becomes clear is that the catalogue does not just run alongside the history of the library, or the institution the library is part of, but is part of it and helps to shape it."[43] Her observation invites us to imagine what the catalog, supported by basic revision control features, may teach us about the practice of librarianship and the greater culture in which it is situated. The remainder of this section outlines some of those possibilities.

Cataloging Revisions and Changes to Cataloging Rules

As was noted earlier, the standards and rules for cataloging have evolved over time.[44] 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.[45] Librarians must grapple with the application of "minimal" and "maximum" catalog rules[46] in addition to the shifting sea of standards and formats.[47] Current cataloging mechanisms do not allow us to assess the impact of these shifts within the practice of librarianship. Full revision control for catalog entries would allow us to explore key questions such as these:

1. How has the balance of full and minimal cataloging entries changed over time? Is there a link between this shift and the increasing resource pressures placed on libraries and their funders?
2. How have shifting catalog rules affected the ways in which resources are described?
3. Can historical revisions of catalog records be conflated to provide more robust information about a resource?
4. When and where have librarians "broken the rules" to describe a resource,[48] and how might these intentional fractures inform the evolution of cataloging rules?

5. How have catalog 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:[49] 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)?[50]

Donnelly advocates for flexibility in the assignment of terms by topic or discipline but stresses that there is no framework for debates of this nature among librarians and their communities.[51] One response to this concern is the "talk page" functionality provided by Wikipedia and its ability to record article editors' notes in the "edit summary" field. In other words, the catalog itself can serve as Donnelly's framework.

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. LIS practitioners are constantly embroiled in these debates, and one recent case serves as an example of how far-reaching the debate may become.

The term "illegal aliens," created as a standard subject heading by the Library of Congress in 1980,[52] is controversial. As early as 2012 the Applied Research Center (now known as Race Forward) began a campaign to have the term eliminated,[53] and in 2014 a group of students from Dartmouth College petitioned the Library of Congress to use alternate terms to describe people residing in America illegally. Both groups claimed the existing term was imprecise and offensive.[54]

In March 2016 the Library of Congress agreed and decided to transition to the use of "noncitizens" and "unauthorized immigration,"[55] but this provoked a political backlash. Two months later the House of Representatives passed a funding bill that ordered the Library of Congress to continue the use of the existing term.[56] In a mirror of Wikipedia's coverage of the Gamergate controversy, an examination of the history and use of "illegal aliens" in library catalog records could add context to the debate. A cursory search of WorldCat shows 4,391 titles that have the subject term assigned,[57] but current cataloging systems do not allow us to ask questions like:

1. What terms were used to describe these materials before 1980?
2. How long did it take for the current term to be applied to existing items? Did librarians struggle with its application or refuse to apply it in some cases?
3. What terms have been co-located or cross-referenced with "illegal aliens" and how have those terms changed over time?
4. What other patterns exist for publishers, authors, and libraries that have applied the term to their materials?

The "illegal aliens" example is recent and represents pressure from the general community, but subject classification controversies have also roiled within the LIS profession. Perhaps 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.[58] He keeps a scorecard of his subject classification victories, which include the addition of terms like "makerspaces," "krumping," and "intersexuality."[59] Other scholars have examined the availability of queer subjects,[60] the use of terms like "east indians"[61] and the development of methods to better-describe indigenous people in libraries and archives.[62] The work of Sanford Berman and other scholars should be supported by the ability to examine and analyze the representation of these subjects in the history of the catalog itself.

Conclusion

The library catalog has transitioned from analog to digital and has lost itself along the way. Analyses that were possible with physical catalog cards can no longer be performed, and tools that process digital records leave no traces of the information they add, remove, or update.

But what of it? If the digital revolution has served LIS and its catalog for 40 years without revision history and change tracking, do we need it at all? In one sense, the cataloging tension described by Buckland has been resolved: since we cannot easily look into the past, we can more easily extrapolate the future by iterating on the present. However, this is not economy: it is willful blindness.

Katharine Whaite's 2012 study of physical catalog cards and Flock *et. al.*'s explorations of controversial material on Wikipedia illuminate what is possible. In the traces of our own work as stewards of human knowledge we can see our evolution reflected. The range of questions and challenges facing LIS—which mirror and reverberate the questions faced by society as a whole—demand attention. We need not look far to see a model for the mirror we need, so let's extend what we've done to catalog records to the catalog's engine: delete it and start fresh.

Notes

[1] S. R. Ranganathan, *Theory of Library Catalogue* (London: Madras Library Association, 1938).

[2] These four traits are referred to as the "objects" of the catalog. See Charles A. Cutter, *Rules for a Printed Dictionary Catalog*, 1st ed. (Washington: Government Print Office, 1876), 12.

[3] Birger Hjørland, "Subject (of Documents)," *KNOWLEDGE ORGANIZATION* 44, no. 1 (2017): 56–58, https://doi.org/10.5771/0943-7444-2017-1-55.

[4] 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.

[5] Michael K. Buckland, "Obsolescence in Subject Description," *Journal of Documentation* 68, no. 2 (March 2, 2012): 154–61, https://doi.org/10.1108/00220411211209168.

[6] Geoffrey C. Bowker and Susan Leigh Star, "The Kindness of Strangers: Kinds and Politics in Classification Systems," in *Sorting Things out: Classification and Its Consequences*, Inside Technology (Cambridge, MA: MIT Press, 1999), 53–106.

[7] Seymour Lubetzky, "Development of Cataloging Rules," *Library Trends* 2 (1953): 179–86; Steven A. Knowlton, "Criticism of Cataloging Code Reform, as Seen in the Pages of *Library Resources and Technical Services* (1957–66)," *Library Resources & Technical Services* 53, no. 1 (January 1, 2009): 15–24, https://doi.org/10.5860/lrts.53n1.15; Dorothy Gregor and Carol Mandel, "Cataloging Must Change!," *Library Journal* 116, no. 6 (4/1/1991 1991): 42–47.

[8] Bowker and Star, "The Kindness of Strangers: Kinds and Politics in Classification Systems"; Katharine Whaite, "Finding Value in History: Gaining Knowledge by Examining Historical Practices.," *Catalogue & Index*, no. 169 (December 2012): 25–29.

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

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

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

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

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

[14] Caitlin Dewey, "The Only Guide to Gamergate You Will Ever Need to Read," *The Washington Post*, October 18, 2014, https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read.

[15] "Gamergate Controversy," in *Wikipedia*, November 23, 2018, https://en.wikipedia.org/w/index.php?title=Gamergate_controversy&oldid=870270616.

[16] 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), 1, https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/viewFile/10656/10561.

[17] "WhoVIS," accessed November 24, 2018, https://aifb-ls3-kos.aifb.kit.edu/sites/whovis/index.html.

[18] "Contropedia," accessed November 24, 2018, http://contropedia.net/.

[19] The algorithm for deriving controversy scores can be found in 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), 1–2, https://doi.org/10.1145/2702123.2702436.

[20] "F^2," accessed November 24, 2018, https://f-squared.org/whovisual/.

[21] These examples can be seen in a live demo of Contropedia. See "Contropedia: Gamergate Controversy," Contropedia, accessed December 11, 2018, https://goo.gl/rTi7Fp.

[22] From an edit made by user Hengsheng120 on July 29, 2015; see "Gamergate Controversy: Difference between Revisions," Wikipedia, accessed December 11, 2018, https://goo.gl/e4zGwf.

[23] From an edit made by user MarkBernstein on July 26, 2015; see "Gamergate Controversy: Difference between Revisions," Wikipedia, accessed December 11, 2018, https://goo.gl/YLNsPC.

[24] "Wikipedia:Bots," in *Wikipedia*, October 28, 2018, https://en.wikipedia.org/w/index.php?title=Wikipedia:Bots&oldid=866149225.

[25] Flock et al., "Towards Better Visual Tools for Exploring Wikipedia Article Development --- The Use Case of Gamergate Controversy," 51.

[26] Ted Bedell, "Wikipedia Is Launched," World History Project, accessed December 2, 2018, https://worldhistoryproject.org/2001/1/15/wikipedia-is-launched.

[27] 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.

[28] DARPA, "ARPANET and the Origins of the Internet," Defense Advanced Research Projects Agency, accessed December 2, 2018, https://www.darpa.mil/about-us/timeline/arpanet.

[29] 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.

[30] Internet Archive, "Open Library Data," Open Library, May 3, 2018, para. 1, https://archive.org/details/ol_data.

[31] OCLC, "Instructions and Guidelines for Reporting WorldCat Bibliographic and Authority Record Changes or Duplicates," OCLC Support & Training, May 3, 2018, paras. 2–3, https://www.oclc.org/support/worldwide/en_us/services/worldcat/documentation/records/instruction-and-guidelines.html.

[32] 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.

[33] Seikel and Steele, 332–33.

[34] Harvard Library, "Harvard Library Open Metadata," Harvard Library, 2017, sec. 2, https://emeritus.library.harvard.edu/open-metadata.

[35] Some recent studies in this area include Katrina Fenlon et al., "A Preliminary Evaluation of Hathitrust Metadata: Assessing the Sufficiency of Legacy Records," in *IEEE/ACM Joint Conference on Digital Libraries* (2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL), London, United Kingdom: IEEE, 2014), 317–20, https://doi.org/10.1109/JCDL.2014.6970186; Karen Snow, "Defining, Assessing, and Rethinking Quality Cataloging," *Cataloging & Classification Quarterly* 55, no. 7–8 (November 17, 2017): 438–55, https://doi.org/10.1080/01639374.2017.1350774; Alberto Petrucciani, "Quality of Library Catalogs and Value of (Good) Catalogs," *Cataloging & Classification Quarterly* 53, no. 3–4 (May 19, 2015): 303–13, https://doi.org/10.1080/01639374.2014.1003669.

[36] Only the most active catalog record editors are included in the diagram. The full graph contains more than 7000 nodes, many of whom contribute catalog data on an infrequent basis.

[37] Judith A Overmier and Elaine M Doak, "Provenance Records in Rare Book and Special Collections," *Rare Books & Manuscripts Librarianship* 11, no. 2 (1996): 94–95.

[38] "Database Reports/Most Edited Articles Last Month," in *Wikipedia*, March 4, 2019, https://en.wikipedia.org/w/index.php?title=Wikipedia:Database_reports/Most_edited_articles_last_month&oldid=886051762.

[39] "Size of Wikipedia," in *Wikipedia*, March 1, 2019, https://en.wikipedia.org/w/index.php?title=Wikipedia:Size_of_Wikipedia&oldid=885591528.

[40] OCLC, "Inside WorldCat," OCLC, February 28, 2019, https://www.oclc.org/en/worldcat/inside-worldcat.html.

[41] "Wikipedia:Most Frequently Edited Pages," in *Wikipedia*, March 1, 2019, https://en.wikipedia.org/w/index.php?title=Wikipedia:Most_frequently_edited_pages&oldid=885658754.

[42] "ILink - Fear: Trump in the White House," NEOS Library Consortium Catalogue, accessed March 7, 2019, http://www.library.ualberta.ca/permalink/opac/8345583/WEBSERVER.

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

[44] Elisabeth de Rijk Spanhoff, "Principle Issues: Catalog Paradigms, Old and New," *Cataloging & Classification Quarterly* 35, no. 1–2 (December 2002): 38, https://doi.org/10.1300/J104v35n01_04; Lubetzky, "Development of Cataloging Rules."

[45] Laura Salas-Tull and Jacque Halverson, "Subject Heading Revision: A Comparative Study," *Cataloging & Classification Quarterly* 7, no. 3 (June 4, 1987): 11, https://doi.org/10.1300/J104v07n03_02.

[46] Glenn Patton, "OCLC's Long Association with Less-Than-Full Cataloging," *Technical Services Quarterly* 9, no. 2 (February 12, 1992): 22, https://doi.org/10.1300/J124v09n02_04.

[47] Spanhoff, "Principle Issues," 38.

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

[49] F. K. Donnelly, "Catalogue Wars and Classification Controversies.," *Canadian Library Journal* 43 (August 1986): 246.

[50] 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.

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

[52] Selene Rivera and Steve Padilla, "Library of Congress to Stop Using Term 'Illegal Alien,'" *Los Angeles Times*, April 3, 2016, http://www.latimes.com/nation/la-na-library-congress-alien-20160403-story.html.

[53] Jake Scobey-Thal, "Illegal Alien: A Short History," *Foreign Policy* (blog), August 27, 2014, https://foreignpolicy.com/2014/08/27/illegal-alien-a-short-history/.

[54] Jasmine Aguilera, "Another Word for 'Illegal Alien' at the Library of Congress: Contentious," *The New York Times*, December 21, 2017, sec. U.S., https://www.nytimes.com/2016/07/23/us/another-word-for-illegal-alien-at-the-library-of-congress-contentious.html.

[55] Library of Congress, "Library of Congress to Cancel the Subject Heading 'Illegal Aliens,'" March 22, 2016, 1, https://www.loc.gov/catdir/cpso/illegal-aliens-decision.pdf.

[56] Lisa Peet, "Library of Congress Drops Illegal Alien Subject Heading, Provokes Backlash Legislation," The Library Journal, June 14, 2016, http://www.libraryjournal.com/?detailStory=library-of-congress-drops-illegal-alien-subject-heading-provokes-backlash-legislation.

[57] As of November 20, 2018; see https://www.worldcat.org/search?q=su%3AIllegal+aliens+United+States.&qt=hot_subject

[58] 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): 353–57, https://doi.org/10.1080/01639374.2017.1327468.

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

[60] Matt Johson, "A Hidden History of Queer Subject Access," in *Radical Cataloging: Essays at the Front*, ed. K. R. Roberto (Jefferson, N.C: McFarland & Co, 2008), 22–24.

[61] Paromita Biswas, "Rooted in the Past: Use of 'East Indians' in Library of Congress Subject Headings," *Cataloging & Classification Quarterly* 56, no. 1 (January 2, 2018): 3–6, https://doi.org/10.1080/01639374.2017.1386253.

[62] 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.