

Mitigating Bias in Metadata

A Use Case Using Homosaurus Linked Data

Juliet L. Hardesty and Allison Nolan

ABSTRACT

Controlled vocabularies used in cultural heritage organizations (galleries, libraries, archives, and museums) are a helpful way to standardize terminology but can also result in misrepresentation or exclusion of systemically marginalized groups. Library of Congress Subject Headings (LCSH) is one example of a widely used yet problematic controlled vocabulary for subject headings. In some cases, systemically marginalized groups are creating controlled vocabularies that better reflect their terminology. When a widely used vocabulary like LCSH and a controlled vocabulary from a marginalized community are both available as linked data, it is possible to incorporate the terminology from the marginalized community as an overlay or replacement for outdated or absent terms from more widely used vocabularies. This paper provides a use case for examining how the Homosaurus, an LGBTQ+ linked data controlled vocabulary, can provide an augmented and updated search experience to mitigate bias within a system that only uses LCSH for subject headings.

INTRODUCTION

Controlled vocabularies are a vital part of how individuals and communities are understood and discussed in scholarly discourse and research. Controlled vocabularies are also a way to standardize terminology and allow items to be grouped by common subjects for easier discovery and access points. While larger, more universally recognized vocabularies like the Library of Congress Subject Headings (LCSH) exist, they are often slow to be updated and they reflect a largely white, heterosexual, cisgender, male, Christian-centric point of view.¹ When the terminology used to define a systemically marginalized group is determined by those outside of the group, often the terms are outdated or reflect a biased perspective.² The prevalence and continued use of outdated metadata and vocabularies in discovery systems creates a cycle of biased search practices that can be difficult to break without the help of information professionals and outside resources. Controlled vocabularies that have been created by or have the input of marginalized communities tend to be more inclusive and up to date. Unfortunately, these vocabularies often are not known to the public or to researchers not well versed in metadata practices. Providing access to controlled vocabularies created by marginalized communities and linking them to existing vocabularies such as LCSH can help make the search process more representative of the people who are using discovery systems and can connect them to resources that better represent themselves and their needs in a complex information world.

LCSH terms are available as linked data, a format that enables online machine-readable connections between concepts and terms, and there needs to be an effort to make systems using LCSH terms more inclusive and representative of marginalized communities. The project described in this article built and gathered feedback on a proof-of-concept JavaScript application to show how defined connections between vocabularies can be used to provide alternative and

Juliet L. Hardesty (jlhardes@iu.edu) is Metadata Analyst, Indiana University. **Allison Nolan** (anolan147@gmail.com) is Library and Information Science Graduate Student, Indiana University. © 2021.

often enhanced access to library catalog resources. In this instance, Simple Knowledge Organization System (SKOS) relationships link LCSH subject terms to the Homosaurus linked data vocabulary, an “international linked data vocabulary of LGBTQ terms that supports improved access to LGBTQ resources within cultural institutions.”³ SKOS is “a common data model [from the W3C] for sharing and linking knowledge organization systems via the Web.”⁴ This project uses `skos:exactMatch` relationships defined by the Homosaurus to enable researchers to use Homosaurus terms to search a library catalog and retrieve relevant results based on the connected LCSH terms that are already in the catalog record.⁵ Subject searches are conducted when the Homosaurus term and the LCSH term match exactly, since the LCSH term’s presence in the library catalog record indicates a specific grouping of records could have this subject term applied. If the Homosaurus term does not match exactly to the LCSH term, a keyword search is conducted using the Homosaurus term to retrieve library catalog results where the Homosaurus term appears in any indexed field in the catalog record, including creator-supplied title and abstract information. Using a vocabulary like the Homosaurus this way helps to connect researchers to resources that more accurately reflect systemically marginalized communities and potentially more accurately reflects the researchers themselves. By providing connections for users that they would otherwise have difficulty finding without the help of a librarian or other information professional, projects such as this one hope to combat the cycle of biased metadata and biased research practices that has dominated academic research.

LITERATURE REVIEW

Students in higher education who identify as members of systemically marginalized communities can continue to experience marginalization within higher educational institutions, and the academic library setting is no exception. Brook, Ellenwood, and Lazzaro provide analysis of multiple studies showing the effect of mostly white staffing in academic libraries, the impact this can have on reference services provided to patrons from marginalized communities, and the overwhelming and intimidating spaces in sizable academic libraries that can be “compounded for students who already feel that they do not belong on campus on the basis of their race.”⁶ When considering how this experience impacts using an online library catalog or digital repository system for conducting research, these same students can find themselves not well represented.⁷ Additionally, crossing disciplines to capture intersectionalities of an identity can be complicated by narrow controlled vocabulary terms which compound problems that already make interdisciplinary research difficult.⁸ Drabinski proposes that the library catalog should be treated as a biased text that requires critical thinking to understand.⁹ Subject headings from authorities such as the Library of Congress will never be unbiased as attitudes, perspectives, and identities change over time. It is therefore important to leverage information literacy competency standards put forward by the Association of College & Research Libraries and teach students how to critically engage the library catalog as another information source. Library instruction is one way to ease the challenges faced by marginalized researchers in higher education, helping researchers effectively use a system like a library catalog that incorporates biased subject headings. However, with interdisciplinary research, materials are often dispersed across information systems and physical locations, and there is still the challenge to identify and locate everything relevant to the research topic.¹⁰ Using available fields within the library catalog record itself (the 590 in MARC, for example) can identify cross-disciplinary resources. Examples are provided by Hogan for Black LGBTQ resources and Latina lesbian literature.¹¹ What all of these efforts seem to point to is what Hannah Buckland proposes: changing the framing of catalog records from “aboutness” to

“fromness,” providing “culturally-responsive metadata” that J. L. Colbert recognizes can create an “equitable subject access” experience that “center[s] the information needs and information seeking behaviors of those whom our systems disenfranchise.”¹² These changes can often only be implemented locally due to language variation and localized community relevance; but Colbert then considers how linked open data might prove useful to combine or relate different subject or community vocabularies. “When we decenter the idea that for every concept there is one controlled term to describe it, we allow the play of seemingly opposite ways of thinking. . . . A linked open data catalog allows libraries to complement, replace, or even reject the standards that have been decided for us and our patrons.”¹³

Librarians and archivists have suggested and tried other methods to mitigate the impact of systemic marginalization. These efforts go beyond the use of controlled vocabularies in the creation of catalog records. One of the earliest and most significant examples of this is Dorothy Porter’s work in organizing the collections she managed at Howard University. Up to that point in the 1930s and 1940s, Dewey Decimal Classification (DDC) was used to organize works on the shelf. Many libraries of the time were predominantly white institutions and Dorothy Porter remembered them using DDC to shelve anything by a Black author or about the Black experience under the DDC heading for colonization (325) or slavery (326).¹⁴ Porter instead organized her collections based on subject matter, genre, and author, categorizing the work based on what it was about rather than the race of the author or the race of any people mentioned in the work. This subtle yet fundamental shift shows the real impact that libraries have on access to collections for their audiences.

Hope A. Olson and Dennis Ward created a proof-of-concept Microsoft Access database interface connecting Mary Ellen Capek’s *A Women’s Thesaurus* to the Dewey Decimal Classification scheme to offer an end user interface for searching a DDC system using the thesaurus terminology. The idea, initially from Joan Mitchell (then editor of DDC), was to develop “a means of making DDC accessible from the point of view of a marginalized knowledge domain—in particular, creating a means of browsing DDC from a feminist/women’s studies perspective.”¹⁵ Variables were defined from characteristics of different classifications to enable a systematic match to thesaurus terms.

Dorothy Berry’s work at University of Minnesota Libraries to gather and digitize African American-related materials from across archival collections for aggregating in *Umbra Search African American History* shows an option for pulling a collection from other collections and highlighting what would otherwise remain marginalized items from marginalized communities.¹⁶ Discovering these materials required searching with a variety of terms used over time to refer to African Americans. Adding collection level context at the folder level for these materials allows aggregation without losing original place and context, while at the same time centering the marginalized communities represented in these materials by gathering them from these various and marginalized original locations.

Archives for Black Lives in Philadelphia is “a loose association of archivists, librarians, and allied professionals in the Philadelphia and Delaware Valley area responding to the issues raised by the Black Lives Matter movement.” Within this group, the Anti-Racist Description Working Group has compiled an annotated bibliography and metadata recommendations to address racist and anti-Black archival description.¹⁷ The recommendations focus on the Black community but can be applied more broadly when describing records by and about any marginalized community. The

recommendations include decentering “neutrality” and “objectivity” for “respect” and “care,” particularly when deciding on controlled vocabulary terms to use in archival description. Specific recommendations to use “terminology that Black people use to describe themselves,” to recognize that this “terminology changes over time, so description will be an iterative process,” and to consult “alternative cataloging schemes created by the subjects of the records being described when and if they are available” provide an approach that looks for descriptive terms from within the community and moves away from terms applied to a community by others.¹⁸ Paying attention to the controlled vocabularies applied to archival description helps to change the narrative and the power structure of the historical record, centering those who have been marginalized and oppressed and increasing discoverability and access to their stories and perspectives. Allowing for changes in controlled vocabulary terms keeps systems flexible enough to accommodate changes in a community’s terminology over time. Linked data relationships can connect term changes for more comprehensive searching while also identifying the current controlled vocabulary term to use.

The Lavender Library, Archives, and Cultural Exchange (LLACE) community archives in Sacramento, California is an archive for a marginalized community.¹⁹ In developing archival and circulating library collections that serve the queer community, the library collections use a thesaurus of queer terms from Dee Michel for classification and the archival collections use subject headings from Michel’s thesaurus along with LCSH.²⁰ The focus, again, begins with the community being served and recognizes that widely used controlled vocabularies like LCSH do not serve these collections or communities well. Starting with a community-specific vocabulary and then connecting LCSH terms centers the collections and community first and then makes connections to the larger library and archives community possible.

Other efforts have used alternatives or supplements to common vocabularies and schemes. The Xwi7x̱wa Library’s use of the Brian Deer Classification System at the University of British Columbia incorporates names and terminology from the First Nations community to better represent that community beyond what something like Library of Congress Classification provides. Using accurate names of nations and peoples, according to the head librarian, Ann Doyle, helps create identity among users of the collection and “shapes the research and types of questions that people ask.”²¹ The National Indian Law Library began cataloging using local terminology only. As it moved records online and sought to be more discoverable and cooperative with other libraries, this local terminology was synchronized with LCSH and specialized terms for federal Indian law and tribal law were kept as a supplement.²² Doing this work is not only about changing terms on catalog records but also learning and making connections with communities who have been marginalized by these systems. Farnel et al. explain the process of decolonizing both the library catalog and digital collections description at University of Alberta Libraries through investigation, analysis, partnering with other institutions doing this work and, most importantly, reaching out to Indigenous communities represented in these records to engage and learn about the most appropriate terminology to use.²³

Different methods and attempts to center the marginalized in cataloging and collection description show it is possible and essential to voice the concerns of those least represented in order to have the most impact on all researchers using these resources. Widely used controlled vocabularies like LCSH continue to be a major way to aggregate collections and provide common access points. Groups like the Association for Library Collections and Technical Services’

Cataloging and Metadata Management Section Subject Analysis Committee continue to work to change terms in these vocabularies to provide better and more accurate representation for systemically marginalized communities, but the process is slow and will likely never be enough.²⁴ Incorporating vocabularies from systemically marginalized communities for use either on the cataloging/description side or for researchers to use for search and discovery offers possibilities for more inclusive experiences that center marginalized voices and expand the options for research questions to ask and answer.

METHODOLOGY

To test this idea that connections provided between a systemically marginalized community's controlled vocabulary and a more generalized vocabulary like LCSH could be helpful, a proof-of-concept information retrieval aid was conceived. The idea was to create a lightweight JavaScript application that could use a select set of terms from the Homosaurus (<http://homosaurus.org>), an LGBTQ+ vocabulary originally created by IHLIA LGBT Heritage (<https://www.ihlia.nl/?lang=en>) and now also used in its linked data form by the Digital Transgender Archive (<https://www.digitaltransgenderarchive.net>), to connect to LCSH terms and provide search links against a library catalog (IUCAT, <https://iucats.iu.edu>, Indiana University's online library catalog) that uses LCSH for subject headings.

Homosaurus version 1 was used initially and did not identify connections to LCSH terms. Analysis of Homosaurus terms against LCSH terms suggested some connections could be made and for initial construction of the proof-of-concept application these were used, but with the recognition that these connections were not coming from the community vocabulary. This was a problem since the point in mitigating bias is to use the community's definitions and any outside interpretations are necessarily not going to reflect the community's intentions. As the application concept continued to form and the initial term comparison work continued, Homosaurus version 2 was released containing explicit connections to LCSH terms, using `skos:exactMatch` for mapping those connections. Those connections in version 2 are not expressed as linked data but are provided in the vocabulary's site for each term. The proof-of-concept work switched to using select terms from Homosaurus version 2 in order to make use of the LCSH connections now being provided by the community.²⁵

The proof-of-concept application used the select set of Homosaurus version 2 terms downloaded as JSON-LD and added in the LCSH terms using the supplied `skos:exactMatch` relationship. The user interface provided visual connections from the selected Homosaurus term to its narrower, broader, and related terms within Homosaurus. Any exact matches to LCSH terms and any Use For terms Homosaurus indicated should be replaced by this term were provided together. The visual layout for the application is directly influenced by the IHLIA LGBT Heritage collections browse interface.²⁶ In IHLIA's system, after searching for a term ("love," for example), the interface provides broader, narrower, related, and used for terms as suggestions for other ways to discover items in these collections in a visually connected bubble layout surrounding the search term. Those connections are linked and can be used to navigate IHLIA's controlled vocabulary, which also happens to be powered by a local non-linked data form of the Homosaurus vocabulary. In the proof-of-concept application, for terms where there is an LCSH exact match, the LCSH term was used for the connection to search IUCAT and was only revealed on screen if the Exact Match (LCSH) bubble was clicked by the user (see fig. 1).

The screenshot displays the Homosaurus v2 terms interface. On the left, a vertical list of terms includes: Asexual people, Demisexual people, Fetishism, Gay community, Gay men, Gender dysphoria, Gender expression, Gender identity, Gender roles, Heteronormativity, Heterosexuality, Homosexuals, Lesbianism, Lesbians, Masochism, Non-binary people, Sexual minorities, Sexual orientation, Sexual preference, Sexuality, Transgender people, Transgenderism, and Two-spirit people. A 'Click here to reset.' link is located at the top right of the interface. The main content area shows search results for the term 'transgenderism'. It features a central 'Preferred:' section with the term 'Transgenderism' in blue. To the left is a 'Narrower:' section with the URL 'http://homosaurus.org/v2/transsexualism'. To the right is a 'Broader:' section with the URL 'http://homosaurus.org/v2/medicalisation'. Above the 'Preferred:' section is an 'Exact Match (LCSH):' section with the term 'Gender nonconformity' in blue and a 'Use For:' label. Below the 'Preferred:' section is a 'Related:' section listing several related terms and their corresponding URLs: 'http://homosaurus.org/v2/gender', 'Gender expression', 'Gender identity', 'http://homosaurus.org/v2/intersex', 'http://homosaurus.org/v2/transgenderInformationCentres', 'http://homosaurus.org/v2/transgenderMovement', 'Transgender people', 'http://homosaurus.org/v2/transgenderStudies', and 'http://homosaurus.org/v2/LGBTI'.

Figure 1. Information retrieval aid showing the Homosaurus term “transgenderism” linked to search IUCAT. Exact Match (LCSH) shows the LCSH term “gender nonconformity” (also linked to search in IUCAT) along with narrower, broader, and related Homosaurus terms.

The initial proof-of-concept information retrieval aid JavaScript application was shared with and tested by Olivia Adams, a graduate student at Indiana University working as the library coordinator for the LGBTQ+ Culture Center Library at Indiana University (<https://lgbtq.indiana.edu/programs-services/library/index.html>). This library has adapted the LLACE classification system, the shelving organizational scheme developed by the Lavender Library in Sacramento, California (<http://lavenderlibrary.com>), for organizing its own physical collection of resources. The LGBTQ+ Culture Center Library also has its own online library catalog that makes use of an established local list of tags for items included in that system (<https://www.librarycat.org/lib/iuglbtlibrary/>). The information retrieval aid application was first presented to the LGBTQ+ Culture Center library coordinator for general impressions and feedback. Additionally, specific tasks were proposed. Please note that the proposed tasks use a vocabulary term as an example that is offensive and outdated. The results of this testing, along with feedback from the Homosaurus Editorial Board, clarified the need to change the information retrieval aid to supply this additional contextual information (available in Homosaurus as a description for the term). The tasks presented for trying the information retrieval aid were the following:

- You want to find resources at IU about transgenderism. What do you think of the resources that IUCAT is offering through this information retrieval aid?
- How do the Homosaurus terms you are seeing here compare to the LLACE classification terms or the tags/subjects you use in the LGBTQ+ Library catalog?
- What is the importance of transparency for the LCSH terms in relation to community values (for terms that are different and only shown in the hidden section right now)?

“Transgenderism” is a term Homosaurus connects to LCSH’s term “gender nonconformity” with an exact match relationship (<http://homosaurus.org/v2/transgenderism>). To provide results for answering the first question, the proof-of-concept information retrieval aid interface showed the Homosaurus term with a linked search in IUCAT that provided results using the LCSH term as a subject search.²⁷ The second question was asked to get a sense of the relevance of the Homosaurus terms to the collections organized and housed in the LGBTQ+ Culture Center Library. The third question about the importance of transparency for the LCSH terms in relation to community values was meant to investigate how a system like this proof-of-concept information retrieval aid might be used by the community of researchers and patrons using the Culture Center’s library, and if the mechanism to mask the LCSH term in favor of the Homosaurus term is useful or not.

The code for this JavaScript web application in its current state is available on GitHub at <https://github.com/jlhardes/metadataBias>. The initial proof-of-concept application was developed by Justina Kaiser, at the time an Information and Library Science graduate student at Indiana University. The current code is a fork of her project, also available on GitHub (<https://github.com/juskaise/metadataBias>).

DISCUSSION

Sharing this proof-of-concept information retrieval aid using Homosaurus terms with the LGBTQ+ Culture Center librarian revealed the importance of usability testing and being receptive to a community’s needs. An introduction and explanation of the controlled vocabulary and the community it represents was a recommended addition since the term list presented was not initially easily identified. Additionally, the interface terminology of narrower/related/broader/exact match/use for is familiar in the library world but not necessarily for the casual user. This terminology is still in use by the information retrieval aid but is under review for updated labels that are easier to understand. This initial version kept any Use For terms hidden unless the user clicked on that bubble in the interface to see them. The reasoning was to give more emphasis to the Homosaurus term and to keep any potentially derogatory or harmful terms still in use by LCSH out of the way of researchers (even though the searches conducted against the catalog might need to use those terms if no other linked data connection is available). Feedback here was helpful: hiding terms that Homosaurus does not recommend might hinder discovering results if the researcher wants to search on a term that is no longer used by the community or is considered derogatory or harmful. This is a useful lesson in that covering up the past is not helpful to those in a marginalized community who have experience with that marginalization or those trying to learn about the past experiences of a marginalized community. Also, being able to find all relevant resources can mean a variety of terms (both current in the community and no longer current) might be necessary. The Homosaurus Editorial Board also explained that Use For terms are sometimes slang terms and are not always considered derogatory. This information is helpful in figuring out how to present LCSH terms in the interface

in the context of the Homosaurus terms. Additionally, moving Use For terms next to Related Terms connected these sets of terms better than placing Use For terms with Exact Match terms.

Further feedback from the Homosaurus Editorial Board regarding the example term used for testing showed the terms and their connections to other terms do not supply enough information to express the full meaning of the term within the community. Without supplying the Homosaurus description for the term “transgenderism” (“Pathologizing term often used in the medicalization of transgender people; use only in historical context,” see <http://homosaurus.org/v2/transgenderism>), the term can come across in the information retrieval aid as a preferred term from the community when, in fact, it is not. This was a critical update needed for the information retrieval aid to be effective as a research tool.

In using the proof-of-concept interface to search against IUCAT, it was noted by the LGBTQ+ Culture Center librarian that using the LCSH term to conduct a subject search against the catalog might not produce useful results if the Homosaurus term is not an actual exact match to the LCSH term. In this case the Homosaurus term should be searched in the catalog as a keyword instead of a subject, so the search is conducted on all indexed fields in the catalog record. In the example tried for the term “transgenderism” the skos:exactMatch relationship is defined as the LCSH term “gender nonconformity” (see fig. 1). Even though the relationship is identified in Homosaurus as an exact match, searching for “gender nonconformity” as a subject term in the catalog (267 results) and “transgenderism” as a keyword in the catalog (289 results) arrives at different result sets with different types of entries (see figs. 2 and 3). Use For terms, while not always representative of the community providing the vocabulary, do have possible historical relevance if present in supplied information (such as a title) and can be connected to the catalog via keyword searching as well.

There is an importance to revealing these differences within the library catalog and providing results that reflect the terms used by the community. The library’s applied terminology via subjects organizes a different set of resources compared to searching for terminology available via titles or other information supplied by authors and creators. When considering who is part of a community and who is not in this scenario, there are benefits to trying to work around or in addition to the library’s applied organizational scheme. Subject searching in the catalog provides another view (and set of results) for those familiar with the community’s terminology. Those approaching a research topic from outside of a community are able to learn more about how to find resources most effectively, moving from the catalog’s terminology to the community’s terminology.

After trying the proof-of-concept information retrieval aid, the LGBTQ+ Culture Center librarian provided feedback that this could be useful for people new to studying the LGBTQ+ community and unfamiliar with the community’s terminology. With an introduction and explanation of the controlled vocabulary in place and an easy-to-follow interface to guide users through the vocabulary terms, effective searches against the catalog that also reveal terminology used by the community and differences between that terminology and the catalog’s terminology can be both educational and useful for research.

The screenshot shows the IUCAT search interface. At the top, the Indiana University logo and name are visible. Below that, the IUCAT logo and navigation links for Locations, History, Folders (1), and Help are present. A search bar contains the subject "Gender nonconformity". Below the search bar, there are options for "Online Only" and "Choose Search Type" with links for Advanced Search, Browse, Reserves, and Music Search. The search results section is titled "Search Results" and shows "You searched for: Subject > 'Gender nonconformity'". It includes navigation links like "« Previous | 1 - 25 of 267 | Next »" and options for "Sort by Relevance", "25 per page", and "Add checked to folder". On the left, there is a "Limit your search" sidebar with "Newly Added Titles" and a "Format" dropdown menu showing counts for Book (150), E-Book (85), Online Resource (26), Video & Film (25), HathiTrust Emergency Access (6), Journal, Periodical (6), and Manuscript (1). The main results list two items:

- 1. A-Z OF GENDER AND SEXUALITY : from ace to ze**
Author: POTTS, MORGAN
Published: [S.I.] : JESSICA KINGSLEY, 2018
Language: English
Format: E-Book
Copies At: Bloomington
- 2. Omnigender : a trans-religious approach**
Author: Mollenkott, Virginia R.
Published: Cleveland : Pilgrim Press, 2007
Edition: Rev. and expanded ed.
Language: English

Figure 2. Searching Indiana University’s online library catalog (IUCAT) for the LCSH term “Gender nonconformity” as subject shows 267 results.

The screenshot shows the Indiana University IUCAT online library catalog interface. At the top, the IU logo and 'INDIANA UNIVERSITY' are visible. Below that, the 'IUCAT' logo is prominent, followed by navigation links for 'Locations', 'History', 'Folders (1)', and 'Help'. A 'Sign in' button is in the top right. The search bar contains the keyword 'transgenderism' and a search icon. Below the search bar, there are options for 'All Fields' and 'Online Only', along with search type filters: 'Advanced Search', 'Browse', 'Reserves', and 'Music Search'. The search results section is titled 'Search Results' and shows 'You searched for: Keyword > "transgenderism"'. It indicates 289 results and provides navigation options like 'Previous', 'Next', and 'Sort by Relevance'. On the left, there are filters for 'Limit your search' and 'Newly Added Titles'. The main results list two items, both titled '1. Bisexuality and transgenderism interSEXions of the others'. The first result is an E-Book published in Binghamton, NY by Harrington Park Press in 2003, available in Southeast, Bloomington, East, Northwest, Indianapolis, Kokomo, South Bend, and Columbus. The second result is a Book published in Binghamton, NY by Kinsey Institute in 2003.

Figure 3. Searching Indiana University’s online library catalog for the Homosaurus term “transgenderism” as keyword shows 289 results.

One of the largest obstacles to connecting marginalized communities to reliable, representative controlled vocabularies is the lack of controlled vocabularies that are readily available as linked data. Unless an individual or organization has made the effort to establish connections between a community’s vocabulary and LCSH, the representative vocabularies stand alone and remain difficult to discover or use. The proof-of-concept testing of this project illustrates not only the need for connections to community-created controlled vocabularies, but also that having access to those vocabularies can result in more accurate and effective searches and usage of catalog resources. Although vocabularies like LCSH contain outdated terms, having access to a variety of terms that are acceptable at different points in a community’s history can be useful for researchers who may not be as informed about certain systemically marginalized communities and whether certain terms have been completely eliminated, reclaimed, or replaced by more accurate terminology.

Efforts to mitigate bias in metadata via linked data are representative of a larger effort to correct a long-standing issue in libraries and other fields where the voices and perspectives of marginalized individuals have been overshadowed by the voices and needs of the majority. In addition to working to update large, generalized vocabularies and trying to incorporate these voices and

perspectives, this change in method is meant to add those voices and center their importance. By linking community-created vocabularies and placing them front and center in the search process, metadata can become a tool with which to center the voices of marginalized communities and move toward a more equitable method of searching, finding, and using resources.

CONCLUSION

The information retrieval aid is still progressing beyond a proof-of-concept but it has seen significant updates since its initial implementation. Figure 1 shows the initial proof-of-concept that was tested. Introductory information has been added to explain the Homosaurus vocabulary and the information retrieval aid tool itself. More terms are available (although still not the full set of Homosaurus version 2 terms) and the term list in JSON-LD is being used to automatically populate the term list in the interface. If available, the term description is provided for more complete context. Additionally, no terms are hidden in the bubble navigation and Use For is located with Related Terms now. Future work for this project includes incorporating the full list of Homosaurus terms; reconsidering the category names (narrower/related/broader/exact match/use for) to determine if there are better labels to use for these categories that will be easier to understand for a general research audience; and testing the tool with researchers new to LGBTQ+ terminology as well as those more knowledgeable about the LGBTQ+ community and its terminology and history. Additional areas of work that welcome investigation include automating the term list generated for use with the information retrieval aid (via API calls, for example) to help reflect any changes or updates made to the community vocabulary over time; the technical implications of connecting this information retrieval aid to a search engine beyond Indiana University's online library catalog; and using this tool with controlled vocabularies from other systemically marginalized communities, such as the BC First Nations Subject Headings, the Glossary of Disability Terms from the North Carolina Council on Developmental Disabilities, or Atria: Women's Thesaurus from the Institute on Gender Equality and Women's History.²⁸ What difference does it make to use a different search engine that incorporates LCSH terms? Likewise, is it possible to connect other linked data (or non-linked data) controlled vocabularies from systemically marginalized communities and is that effective for retrieving information and improving research outcomes? The work so far shows the possibility of centering systemically marginalized voices by using the system more effectively, making linked data work to connect and update the terminology and search terms available for research.

ACKNOWLEDGEMENTS

The authors would like to thank the LGBTQ+ Culture Center librarian at Indiana University for Spring 2020, Olivia Adams, for her helpful review and feedback of the initial proof-of-concept information retrieval aid. We would also like to thank Brian M. Watson, Editorial Board member of Homosaurus.org, for their help with using Homosaurus version 2 terms and the Homosaurus Editorial Board, particularly K. J. Rawson, for reviewing and supplying article feedback. The authors also acknowledge the work of Justina Kaiser who created the initial code behind the information retrieval aid.

ENDNOTES

- ¹ Hope A. Olson, "Mapping Beyond Dewey's Boundaries: Constructing Classificatory Space for Marginalized Knowledge Domains," *Library Trends* 47, no. 2 (Fall 1998): 238.
- ² The term "systemically marginalized group," used recently by Dr. Nicki Washington from Duke University at the September 3, 2020, Indiana University Center for Women in Technology talk, "Bring a Folding Chair': Understanding and Addressing Issues of Race in the Context of STEM," was revealing to the authors as a better term to use than "historically marginalized communities." This is significant in that it emphasizes the continued oppression and marginalization of these communities, rather than viewing these communities' struggles as something of the past that has been overcome/surmounted.
- ³ "Mission, History, Editorial Board," Homosaurus Vocabulary Site, accessed March 2, 2021, <http://homosaurus.org/about>.
- ⁴ "SKOS Simple Knowledge Organization System Reference," W3, published August 18, 2009, <https://www.w3.org/TR/skos-reference/>.
- ⁵ "skos:exactMatch," SKOS Simple Knowledge Organization System Namespace Document—HTML Variant, 18 August 2009 Recommendation Edition, W3, last modified August 6, 2011, <https://www.w3.org/2009/08/skos-reference/skos.html#exactMatch>.
- ⁶ Freeda Brook, David Ellenwood, and Althea Eannace Lazzaro, "In Pursuit of Antiracist Social Justice: Denaturalizing Whiteness in the Academic Library," *Library Trends* 64, no. 2 (Fall 2015): 259, <https://muse.jhu.edu/article/610078>.
- ⁷ Holly Tomren, "Classification, Bias, and American Indian Materials" (San Jose State University, 2003), <http://ailasacc.pbworks.com/f/BiasClassification2004.pdf>.
- ⁸ Amelia Koford, "How Disability Studies Scholars Interact with Subject Headings," *Cataloging & Classification Quarterly* 52, no. 4 (2014), <https://doi.org/10/gf542p>.
- ⁹ Emily Drabinski, "Queering the Catalog: Queer Theory and the Politics of Correction," *Library Quarterly: Information, Community, Policy* 83, no. 2 (April 2013), <https://www.jstor.org/stable/10.1086/669547>.
- ¹⁰ Sara A. Howard and Steven A. Knowlton, "Browsing through Bias: The Library of Congress Classification and Subject Headings for African American Studies and LGBTQIA Studies," *Library Trends* 67, no. 1 (Summer 2018), <https://doi.org/10.1353/lib.2018.0026>.
- ¹¹ Kristen Hogan, "'Breaking Secrets' in the Catalog: Proposing the Black Queer Studies Collection at the University of Texas at Austin," *Progressive Librarian* 34 (2010), http://www.progressivelibrariansguild.org/PL/PL34_35/050.pdf.
- ¹² J. L. Colbert [ <https://orcid.org/0000-0001-5733-5168>], "Patron-Driven Subject Access: How Librarians Can Mitigate That 'Power to Name'," *In the Library with the Lead Pipe*, November 15, 2017, <http://www.inthelibrarywiththeleadpipe.org/2017/patron-driven-subject-access-how-librarians-can-mitigate-that-power-to-name/>.

- ¹³ J. L. Colbert, "Patron-Driven Subject Access."
- ¹⁴ Avril Johnson Madison and Dorothy Porter Wesley, "Dorothy Burnett Porter Wesley: Enterprising Steward of Black Culture," *Public Historian* 17, no. 1 (Winter 1995): 25, <https://www.jstor.org/stable/3378349>; Janet Sims-Wood, *Dorothy Porter Wesley at Howard University: Building a Legacy of Black History* (Charleston, SC: The History Press, 2014), 39; Zita Cristina Nunes, "Cataloging Black Knowledge: How Dorothy Porter Assembled and Organized a Premier Africana Research Collection," *Perspectives on History: The News Magazine of the American Historical Association* (November 20, 2018), <https://www.historians.org/publications-and-directories/perspectives-on-history/december-2018/cataloging-black-knowledge-how-dorothy-porter-assembled-and-organized-a-premier-africana-research-collection>.
- ¹⁵ Hope A. Olson and Dennis B. Ward, "Feminist Locales in Dewey's Landscape: Mapping a Marginalized Knowledge Domain," in *Knowledge Organization for Information Retrieval: Proceedings of the Sixth International Study Conference on Classification Research* (The Hague, Netherlands: International Federation for Information Documentation, 1997), 129.
- ¹⁶ Dorothy Berry, "Digitizing and Enhancing Description Across Collections to Make African American Materials More Discoverable on Umbra Search African American History," The Design for Diversity Learning Toolkit, Northeastern University Libraries, August 2, 2018, <https://des4div.library.northeastern.edu/digitizing-and-enhancing-description-across-collections-to-make-african-american-materials-more-discoverable-on-umbra-search-african-american-history/>.
- ¹⁷ Alexis A. Antracoli et al., *Anti-Racist Description Resources* (Philadelphia, PA: Archives for Black Lives in Philadelphia, 2019), i, https://archivesforblacklives.files.wordpress.com/2019/10/ardr_final.pdf.
- ¹⁸ Antracoli et al., "Anti-Racist Description Resources," 5.
- ¹⁹ Diana K. Wakimoto, Debra L. Hansen, and Christine Bruce, "The Case of LLACE: Challenges, Triumphs, and Lessons of a Community Archives," *American Archivist* 76, no. 2 (Fall/Winter 2013), <http://www.jstor.org/stable/43490362>.
- ²⁰ According to the article, "the word *queer* is used throughout this article as the most general, over-arching term to describe communities and individuals who support LLACE and make it possible." Diana K. Wakimoto et al., "Case of LLACE," 439; Dee Michel, ed., *Gay Studies Thesaurus*, rev. ed. (Urbana, IL, 1990).
- ²¹ Catelynne Sahadath, "Classifying the Margins: Using Alternative Classification Schemes to Empower Diverse and Marginalized Users," *Feliciter* 59, no. 3 (June 2013): 16.
- ²² Monica Martens, "Creating a Supplemental Thesaurus to LCSH for a Specialized Collection: The Experience of the National Indian Law Library," *Law Library Journal* 98, no. 2 (Spring 2006).

- ²³ Sharon Farnel et al., “Rethinking Representation: Indigenous Peoples and Contexts at the University of Alberta Libraries,” *International Journal of Information, Diversity, & Inclusion* 2, no. 3 (2018), <https://doi.org/10.33137/ijidi.v2i3.32190>.
- ²⁴ ALCTS is a division of the American Library Association—<http://www.ala.org/alcts/mgrps/camms/cmtes/ats-ccssac>; SAC Working Group, “Report of the SAC Working Group on Alternatives to LCSH ‘Illegal Aliens,’” American Library Association Institutional Repository, submitted June 19, 2020, https://alair.ala.org/bitstream/handle/11213/14582/SAC20-AC_report_SAC-Working-Group-on-Alternatives-to-LCSH-Illegal-aliens.pdf.
- ²⁵ This is a moment to acknowledge the work of several Homosaurus Editorial Board members, including Brian M. Watson, who is studying and working with linked data at University of British Columbia; Chloe Noland from American Jewish University; and Walter “Cat” Walker from the William H. Hannon Library and ONE National Gay and Lesbian Archives. There was never a request to add these LCSH term connections, but the timing was incredibly helpful, and the effort greatly appreciated.
- ²⁶ Example search for term “love” that results in browsable terms in a visual interface: <https://www.ihlia.nl/search/index.jsp?q%3Asearch=love&q%3Azoekterm.row1.field3=&lang=en>.
- ²⁷ “Gender nonconformity,” (search results, IUCAT, Indiana University, accessed March 2, 2021), https://iucat.iu.edu/?utf8=%26%2310004%3B&search_field=subject&q=Gender+nonconformity.
- ²⁸ *BC First Nations Subject Headings* (Vancouver, BC: Xwi7xwa Library First Nations House of Learning, March 2, 2009), <http://branchxwi7xwa.sites.olt.ubc.ca/files/2011/09/bcfn.pdf>; “Glossary of Disability Terms,” North Carolina Council on Developmental Disabilities, accessed March 8, 2021, <https://nccdd.org/welcome/glossary-and-terms/category/glossary-of-disability-terms>; “Search in the Women’s Thesaurus,” Atria—Institute on gender equality and women’s history, accessed March 8, 2021, <https://institute-genderequality.org/library-archive/collection/thesaurus>.