

Illinois State University

ISU ReD: Research and eData

Faculty and Staff Publications – Milner Library

Milner Library

8-28-2019

Applying Library of Congress Demographic Group Characteristics for Creators

Eric Willey

Illinois State University, emwille@ilstu.edu

Angela Yon

Illinois State University, ayon@ilstu.edu

Follow this and additional works at: <https://ir.library.illinoisstate.edu/fpml>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Eric Willey & Angela Yon (2019) Applying Library of Congress Demographic Group Characteristics for Creators, *Cataloging & Classification Quarterly*, 57:6, 349-368, DOI: 10.1080/01639374.2019.1654054

This Article is brought to you for free and open access by the Milner Library at ISU ReD: Research and eData. It has been accepted for inclusion in Faculty and Staff Publications – Milner Library by an authorized administrator of ISU ReD: Research and eData. For more information, please contact ISURed@ilstu.edu.

Applying Library of Congress Demographic Group Characteristics for Creators

by Eric Willey & Angela Yon

This research project explores how to ethically, efficiently, and accurately add demographic terms for African American authors to catalog records. A Department of History graduate student located evidence of self-identification as African American for authors, and catalogers added the demographic term to catalog records and are in the process of creating work level Name Title Authority Records for titles written by those authors. This work is the beginning of systematically enhancing catalog data with author characteristics to support making such information available in faceted results.

<https://doi.org/10.1080/01639374.2019.1654054>

Introduction

The Library of Congress began exploring a controlled vocabulary describing demographic groups in 2013, and in 2015 387 initial terms were approved as Library of Congress Demographic Group Terms (LCDGTs).¹ This rose to 1,172 approved terms in January of 2018.² Creation of a controlled vocabulary for describing demographic groups immediately raises the question of how those terms can be added to catalog records accurately, efficiently, and ethically. Early attempts to use the Gender (MARC 375) field in Name Authority Records (NARs) highlighted ethical difficulties in adding demographic terms to NARs, especially for historically marginalized and oppressed groups. Given the number of existing catalog records that would require editing to achieve a critical mass of terms available as faceted search results it follows that research into an efficient means of locating accurate

¹ Library of Congress, Working Group on Full Implementation of the Library of Congress Faceted Vocabularies, “A Brave New (Faceted) World: Towards Full Implementation of Library of Congress Faceted Vocabularies,” 2017, <http://hdl.handle.net/11213/8146>

² Library of Congress, Policy and Standards Division, “Introduction to Library of Congress Demographic Group Terms,” 2018 edition, <https://www.loc.gov/aba/publications/FreeLCDGT/2018%20LCDGT%20intro.pdf>

demographic data that can be applied efficiently will be valuable. Adding demographic group terms allows patrons to filter their search results based on these criteria.

Efforts are already underway to enhance existing Library of Congress Subject Headings (LCSH) for music by using the new Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT), the Library of Congress Medium of Performance Thesaurus (LCMPT), and the Library of Congress Demographic Group Terms (LCDGT) vocabularies. The project found the implementation of demographic group terms to be a “challenging area” and specifically noted that outside of obvious cases LCDGT is far from comprehensive; furthermore, even when a group is described it may not fit the scope of the intended audience or creator/contributor.³ While the African Americans writer project was not limited by lack of terms, it did find similar issues with the use of the Creator/Contributor Characteristics (MARC 386) field, particularly for works with more than one creator or contributor.

RDA Toolkit defines a creator as “An agent responsible for the creation of a work. Creators include agents who are jointly responsible for the creation of a work either performing the same role, such as in a collaboration between two writers, or performing different roles, such as in a collaboration between a composer and a lyricist.”⁴ However, while monographs and sound recordings can have multiple creators under this rule, RDA specifies that for serials, “A person or family is considered to be the creator of a serial if it is responsible for the serial *as a whole*, not an individual issue or a few issues.” [emphasis added].⁵ Individuals responsible for issues or smaller portions of a serial would

³ Casey A. Mullin, “An Amicable Divorce: Programmatic Derivation of Faceted Data from Library of Congress Subject Headings for Music,” *Cataloging & Classification Quarterly*, <https://doi.org/10.1080/01639374.2018.1516709>

⁴ RDA Steering Committee, Glossary: C, RDA Toolkit, <http://access.rdatoolkit.org/> (accessed April 19, 2019)

⁵ RDA Steering Committee, 19.2.1.1.3: Persons or Families Considered to be Creators of Serials, RDA Toolkit, <http://access.rdatoolkit.org/> (accessed April 19, 2019)

likely be listed as contributors under RDA as “An agent contributing to an expression. Contributors include editors, translators, arrangers of music, performers, etc.”⁶ From a cataloging standpoint this distinction between creators and contributors is irrelevant when adding demographic data to the Creator/Contributor Characteristics (MARC 386) field as the field does not distinguish between creators and contributors.

The *Demographic Group Terms Manual* provided by Library of Congress also allows for the inclusion of multiple Creator/Contributor Characteristics (MARC 386) fields in the same record.⁷ From a technical perspective this solves the issue of multiple creators or contributors. Simply add terms for as many creators or for as many demographic terms that apply to a single creator as are deemed relevant to the work. However, the question remains of whether an individual would be considered a creator or contributor by a patron looking at search results. In order to determine this threshold and what constitutes best practice user studies will be necessary, and are a rich vein for future research.

These discussions are also occurring as catalogs are including linked data in search results. This project initially attempted to link Wikipedia pages with Library of Congress NARs by adding URIs so as to create 5 star linked open data as described by Tim Berners Lee.⁸ While a temporary moratorium issued by the Program for Cooperative Cataloging (PCC) on adding Other Standard Identifier (MARC 024) fields halted this portion of the project, the goal remains to be taken up later.⁹ This will link the

⁶ RDA Steering Committee, Glossary: C, RDA Toolkit, <http://access.rdatoolkit.org/> (accessed April 19, 2019)

⁷ Library of Congress, Policy and Standards Division, *Demographic Group Terms Manual*, <http://www.loc.gov/aba/publications/FreeLCDGT/freelcdgt.html> (accessed Jan. 14, 2019)

⁸ “5 ★ Open Data,” <http://5stardata.info/en/> (accessed Jan. 14, 2019)

⁹ Paul Frank, “024 (Other Standard Identifier) data in NACO records: Temporary moratorium,”

PCCLIST@LISTSERV.LOC.GOV, 13 September 2018, <https://listserv.loc.gov/cgi-bin/wa?A2=ind1809&L=PCCLIST&P=38986> (accessed April 19, 2019).

Wikipedia pages and NARs reciprocally to help users to discover related data. Persistent URIs were used to avoid the danger of broken links. The creation of 5 star linked data also reinforces efforts by catalogs to pull information from Wikipedia or Wikidata into search results (such as VuFind 4.1's incorporation of information from Wikipedia articles), and follows the recommendations by OCLC Research to "integrate researchers' external IDs within library applications and services as appropriate."¹⁰

Similar to this present study's awareness of the linked data goal of linking out to other's data, Ferriter discussed the benefits of integrating Wikidata at the Library of Congress. Ferriter wanted to include links in the Library of Congress authority records out to Wikidata. One of the possible functions for Wikidata is to be a center for institutional identifiers. Ferriter's research shows that libraries and cultural institutions are growing increasingly interested in Wikidata, and that connections between knowledge systems can be built by adding identifiers in NARs out to Wikidata.¹¹

Literature review

This project, funded through an Illinois State University (ISU) Research Grant, hypothesizes that many of the names on the Wikipedia "List of African American Writers" are individuals who satisfy Library of Congress criteria for the African Americans demographic group term. The LCDGT manual states that self-identification is the highest (but not exclusive) criteria of proof for membership in a

¹⁰ Karen Smith-Yoshimura, Micah Altman, Michael Conlon, Ana Lupe Cristán, Laura Dawson, Joanne Dunham, Thom Hickey, Daniel Hook, Wolfram Horstmann, Andrew MacEwan, Philip Schreur, Laura Smart, Melanie Wacker, and Saskia Woutersen, *Registering Researchers in Authority Files* (Dublin, Ohio: OCLC Research, 2014), <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf>

¹¹ Margaret Ferriter, "Integrating Wikidata at the Library of Congress," *The Signal* (blog), accessed May 23, 2019, <https://blogs.loc.gov/thesignal/2019/05/integrating-wikidata-at-the-library-of-congress/> .

demographic group.¹² Wikipedia provides criteria for editors describing who is considered African American, but it is difficult to determine to what extent this criteria is followed by those who place people on the list of African American writers.¹³ To evaluate how many individuals on the Wikipedia list satisfy Library of Congress criteria, a graduate student searched for evidence of self-identification as African American for people on the list with items in the local library catalog. Research shows that of the 258 writers who were examined eighty-four percent (217/258) self-identified as African American (or an equivalent term), and there was evidence that could be used to infer that an additional twelve percent (30/258) identified as African American. Circumstantial evidence that an additional two percent (4/258) could be considered African American was found, and evidence was found that three percent (7/258) identified as or would likely be considered an ethnicity other than African American. There was not sufficient time or evidence to determine which category an additional twenty-six writers would have fallen under, and they were not counted in these percentages.

Faceted classification and the faceted approach are important and useful methods in information organization and information retrieval. Broughton presented an examination of the faceted classification as the basis of all methods of information retrieval. Faceted classification has become an important concept for display on the web and its usefulness was appreciated early in automated

¹² For a discussion of the self-identification principle see Jonathan Furner, "Dewey Deracialized: A Critical Race-Theoretic Perspective," *Knowledge Organization* 34, no. 3 (2007): 144-168, <http://works.bepress.com/furner/14/> (accessed May 10, 2019); Jonathan Furner, "Interrogating 'Identity': A Philosophical Approach to an Enduring Issue in Knowledge Organization," *Knowledge Organization* 36, no. 1 (2009): 3-16; Julianne Beall, "Racially Mixed People, DDC Table 5 Ethnic and National Groups, and MARC 21 Bibliographic Format Field 083," *Cataloging & Classification Quarterly*, 47, no. 7: 657-670, DOI:10.1080/01639370903112005

¹³ Wikipedia contributors, "African Americans," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/African_Americans#Who_is_African_American? (accessed August 24, 2018).

information retrieval. According to Broughton, the development of faceted classification and facet theory in the 1930s is credited to Ranganathan, an Indian mathematician and librarian. Many subsequent scholars have contributed and expanded to the understanding of facets. Facets are properties or characteristics of a class or specific subject, and faceted classification allows items to be classified according to those characteristics and properties.¹⁴ While faceted classification is not new, progress has occurred when user studies research was applied with facet theory to create tools that allow users to perform a task in an engaging and meaningful way.¹⁵

What is especially important for online retrieval is that individual facets can be accessed and retrieved separately or in multiple combinations. In an effort to improve the effectiveness of catalogs to help users find information more efficiently, Niu conducted research on the facet feature and faceted search in library catalogs. Niu defined faceted search as “the application of classic facet theory in the online digital environment. It is the combination of free, unstructured text search, with faceted navigation.”¹⁶

Broughton summarized the many varied functions of the faceted system.¹⁷ The faceted system has brought advanced benefits to library catalogs, as a tool for browsing, for navigation, for retrieval, and in query formulation and modification.¹⁸ One of the main strengths of the faceted search in the

¹⁴ Xi Niu, "Faceted Search in Library Catalogs," in *New Directions in Information Organization*, eds. Jung-ran Park, Lynne C. Howarth, and Amanda Spink (Bingley: Emerald Group Publishing Limited, 2014), 173-208, [http://dx.doi.org/10.1108/S1876-0562\(2013\)0000007013](http://dx.doi.org/10.1108/S1876-0562(2013)0000007013)

¹⁵ Vanda Broughton, "The need for a faceted classification as the basis of all methods of information retrieval," *Aslib Proceedings: New Information Perspectives* 58, no. 1/2 (2006): 65, <https://doi.org/10.1108/00012530610648671>

¹⁶ Niu 2014, 174-176.

¹⁷ Broughton, 49-72.

¹⁸ Broughton, 69.

library catalog is that it encourages user participation with the system. Faceted search interfaces combine keyword searches and browsing. An example of a faceted search interface is one that shows results for keyword queries but also organizes results by categories. The meaningful categories lend themselves to guided learning and information finding. Niu defined the main traits of faceted search as “the interactive filtering among multiple dimensions of information.” Users may filter the result set by selecting attributes of different features of a subject, giving them the opportunity to assess and narrow the result set to their needs.¹⁹

Faceted library catalogs have gained popularity in many academic and public libraries. From a study published in 2011, out of a sample of 100 academic libraries, 78 were with facet-based catalogs. Open source catalogs often developed by librarians, in addition to commercial ones, claim facet searching as one of their leading features.²⁰ Today, as much as the faceted approach is regularly used for products on commercial websites it is also evident in a majority of library catalogs.²¹

Facets have been shown to play an important role in discovery. Kules et al. found that their research participants (six undergraduates and eleven graduate students) spend about 50 seconds looking at results, 25 seconds looking at facets, and about 6 seconds looking at the query itself.²² Results from Niu’s dissertation research indicated that most searchers understand the concept of facets naturally and without much training. Most participants in the study (40 users at the University of North Carolina at Chapel Hill Library and the Phoenix Public Library) preferred faceted search over non-faceted search

¹⁹ Niu 2014, 186.

²⁰ Niu 2014, 192.

²¹ Broughton, 49.

²² Bill Kules, Robert Capra, Matthew Banta, and Tito Sierra, "What Do Exploratory Searchers Look at in a Faceted Search Interface?" in JCDL '09 Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (New York, ACM, 2009), 313, <http://dx.doi.org/10.1145/1555400.1555452>

and valued this feature to aid in narrowing their search.²³ However, faceted searches were viewed as supplemental compared to text searches.²⁴ Only a small group employed the faceted search. However, the number increased when browsing facets were built-in the search. The study revealed the faceted catalog did improve search accuracy, but there was no reduction in the search time. Participants applied facets for more difficult tasks that required more investigation into the search.²⁵ People rated the faceted interface higher over non-faceted for its intuitiveness, ease and quickness. Based on observation in the study, facets supported searches in five ways: “to append specific information, to exclude unwanted information, to see what is available, to get an idea for query terms and to solve a ‘cold start’ problem by browsing.”²⁶ While the present study did not seek to determine how useful LCDGTs might be in discovery, subject librarians at the investigating institution did indicate they sometimes receive reference questions related to locating works by members of certain demographic groups. African Americans and child composers were offered as examples of demographic groups that had been included in reference requests. The research conducted by Kules et al. and Niu supports the notion that faceted search results improve discoverability for patrons.

Research has shown positive benefits to including faceted search results. Jody Condit Fagan performed a literature review of usability studies of faceted browsing, and found that including the ability to search facet results (among other benefits) facilitated efficient retrieval, helped avoid dead

²³ Xi Niu, “Beyond text queries and ranked lists: faceted search in library catalogs” (PhD. diss., University of North Carolina at Chapel Hill, 2012), 153-154, <https://cdr.lib.unc.edu/concern/dissertations/h128nf68h> (accessed July 30, 2019)

²⁴ Niu, 2012, iii.

²⁵ Niu 2014, 199.

²⁶ Niu 2012, 130.

ends, and allowed users to retrieve more results and retrieve results more quickly.²⁷ While the studies Fagan reviewed do not necessarily prove conclusively that facets point to success in library catalogs, faceted searching appears to have been adopted by many libraries as standard practice.²⁸ Other studies show that users prefer and retrieve results more quickly in catalogs that support faceted searching (compared to searches that do not include facets in results) in catalogs.²⁹ Results also show that users incorporate facets while searching both public library and university library catalogs.³⁰ Online public access catalog (OPAC) studies indicate users take advantage of facets if they are displayed during the search process. Facets help users refine a search, narrow a query, create a new query, and browse an entire information collection. LC Classification is the most dominant facet used, next is Subject: Topic, and then Library, Format, Author, and Subject: Genre.³¹ Other attributes that contribute to faceted search are FAST headings, demographic group terms, and genre and form terms.

According to a study by Niu, even though faceted search displays some authority data to users, information barriers still exist. The vocabulary revealed to the user in the subject facet is from a controlled vocabulary term in the bibliographic records. If the user's own vocabulary and the vocabulary of bibliographic records do not match, or if the bibliographic record vocabulary is missing from the facet, an appropriate record based on the search may not be retrieved. Users' vocabulary is

²⁷ Jody Condit Fagan, " Usability Studies of Faceted Browsing: A Literature Review," *Information Technology and Libraries* 29, no. 2 (2010): 62, <https://doi.org/10.6017/ital.v29i2.3144>

²⁸ Fagan, 65.

²⁹ Sarah Ramdeen and Bradley M. Hemminger. "A Tale of Two Interfaces: How Facets Affect the Library Catalog Search Experience," *Journal of the American Society for Information Science and Technology* 63, no. 4 (2012): 702-715, <http://dx.doi.org/10.1002/asi.21689>

³⁰ Xi Niu, Xi and Bradley M. Hemminger, "Beyond Text Querying and Ranking List: How People are Searching through Faceted Catalogs in Two Library Environments," *Proceedings of the American Society for Information Science and Technology* 47, no. 1 (2010): 1-9, <https://doi.org/10.1002/meet.14504701294>

³¹ Niu 2014, 197.

varied and often not very flexible. It is difficult for users to fix searches using synonyms. Because of the inability to apply synonyms, they are not able to apply faceted search effectively to pass this information barrier. Another issue that may cause an information barrier is the arrangement of the collection. The search through a library catalog is through MARC records, which are surrogates of the collection. If these surrogates contain typos, missing or incorrect information, faceted search will prove unsuccessful for the user, regardless of a highly innovative faceted system and interface.³² Another problem that occurs with faceted searching is the large number of results that can be retrieved for certain demographic categories.³³ Similarly, a large number of facets themselves can confuse searchers and may not be constructive to the user. Niu's research recognized that participants rarely used some facets, such as the author facet and the MeSH facet.³⁴

In adding demographic data catalogers must also respect creator's wishes. In the past, missteps have been made in this regard. An extensive and well-reasoned critique of RDA 9.7 (which instructed catalogers to record gender when identifying persons), and Library of Congress instructions that limited gender options to male, female, and unknown demonstrates how cataloging practice can directly contradict an individual's identity.³⁵ Other research shows that 39 of 60 records analyzed includes information that could identify an individual as trans.³⁶ While the present research focuses on race and

³² Niu 2014, 200-201.

³³ Chew Chiat Naun, "Next Generation OPACs: A Cataloging Viewpoint," *Cataloging & Classification Quarterly* 48, no. 4 (2010): 330-342, <https://doi.org/10.1080/01639370903437709>

³⁴ Niu 2014, 201-202.

³⁵ Amber Billey, Emily Drabinski, and K.R. Roberto, "What's Gender Got to Do With It? A Critique of RDA Rule 9.7," *University Libraries Faculty and Staff Publications* 19, <https://scholarworks.uvm.edu/libfacpub/19>

³⁶ Kelly J. Thompson, "More Than a Name: A Content Analysis of Name Authority Records for Authors Who Self-Identify as Trans," *Library Resources & Technical Services* 60 (3), 2016: 140-155, <http://dx.doi.org/10.5860/lrts.60n3.140>

ethnicity, the lessons on respecting an individual's identity and privacy presented in these articles are well taken.

Methods

Illinois State University (ISU) is primarily an undergraduate institution with 17,150 undergraduates and 2,370 graduate students enrolled as of spring 2019. Within the university, there are twenty-one master's and nine doctoral programs encompassing disciplines across the university. The university consists of seven colleges made up of Applied Science and Technology, Arts and Sciences, Business, Education, Fine Arts, Nursing, and Milner Library.³⁷ ISU also has an African American Studies minor, and enrollment numbers show an increase in African American students on campus. Milner Library holds over 1.5 million volumes and serves the entire campus.

Library of Congress Demographic Group Terms (LCDGT) was selected as the controlled vocabulary for the project as it was relatively new, and research into methods and the cost of applying the terms would be important for implementation. Three criteria in order of relevance are recommended by the LCDGT manual before a demographic group term is applied to a work: self-identification as a member of that group, or strong evidence (for example, attendance at an all-black university) that would allow a reasonable person to infer a person was a member of a group, or scholarly consensus. In this case our researcher attempted to find evidence of self-identification as an African American, the highest level of proof, although other forms of evidence were included if they were found and self-identification as a group member could not be verified.

³⁷ Illinois State University, "Planning, Research, and Policy Analysis," 2018 spring semester, https://prpa.illinoisstate.edu/data_center/student/Spring%202018%20Enrollment%20Report%20Updated.pdf (accessed April 19, 2019).

The project group specifically wanted to explore adding demographic group terms that patrons would find useful in search results. After discussions with subject librarians African Americans were selected as the demographic group for the project. Subject librarians stated that library patrons sometimes requested titles by African Americans, and would find it useful to be able to locate works specifically by members of that group in faceted search results. There was also an existing list of African American writers in Wikipedia with a significant but not overwhelming number of entries (381 names), and as the list was specifically composed of writers it was likely that there would be existing bibliographic records for their works and they would have NARs. This would limit the amount of original cataloging and NAR creation required by the project group.

Wikipedia was selected as a source to provide an initial list of names as it allows links to NARs through the Wikipedia Authority Template, and provides a list that can be easily accessed by anyone with an internet connection that wished to undertake a similar project. Furthermore, some local catalogs (VuFind 4.0 for example) are already integrating Wikipedia entries into their search results. OCLC has also explored collaborative partnerships between Wikipedia and libraries.

A high priority of the project group was to make sure demographic group terms were added to records ethically and responsibly. This was accomplished by submitting a grant request to hire a graduate student or advanced undergraduate student with appropriate subject expertise to consult on the project and conduct original research to determine if authors self-identified as African Americans. The grant included funds for a student worker for 118 hours at a salary of \$12 per hour, or \$1416 total. Co-investigators consulted with Milner Library History Subject Librarian, Vanette Schwartz, and were referred to Department of History Professor Toure Reed who recommended three students with subject expertise in African American History. All three were strong candidates, and graduate student Trumaine Mitchell was hired due to his extensive background in conducting original research and knowledge of African American history. An initial meeting was held to explain the project and purpose

of controlled vocabularies and demographic group terms, but no technical training in cataloging was provided (MARC, ISBD punctuation, etc.) as the graduate student would be generating citations and evidence from historical sources to be used by catalogers, and not creating or manipulating catalog records himself.

The project had five steps. First, the graduate student subject expert added the Uniform Resource Identifiers (URIs) for NARs in the Wikidata pages for writers where they were absent, and also added the Authority Control Template to the writers' Wikipedia pages if needed. This began the process of creating five-star linked data between NARs and Wikipedia pages, and familiarized the graduate student with Wikipedia and NARs. In the second step the graduate student searched for evidence that individuals on the Wikipedia list of African American Writers with titles in the local catalog self-identified as African American, and created a citation for the evidence he found. For the third part of the project cataloging librarians added demographic group terms to bibliographic records in the local catalog and to their corresponding WorldCat records. In the fourth step, citations showing self-identification were included in a Source Data Found (MARC 670) field and a Biographical or Historical Data (MARC 678) field in the author's Name Authority Record. URIs pointing back to the Wikidata page and Wikidata numbers for the author were also included in Other Standard Identifiers (MARC 024) fields in the NARs (until the moratorium by PCC on adding the MARC 024 field to NARs took effect). In the fifth and final step Name Title Authority Records were created with the demographic terms in Creator/Contributor Characteristics (MARC 386) fields and the citation used for justification in a Source Data Found (MARC 670) field. At the time of writing (28 May 2019), steps one through three are complete and steps four and five are ongoing.

In preparation for the first step the names on the Wikipedia list of African American writers were copied and pasted from the web page into an Excel sheet.³⁸ Working from this list the graduate student examined entries in the Wikipedia List of African American Writers and if it had not already been done linked them to their corresponding Name Authority Record using the Wikidata page and Wikipedia Authority Control Template. This was done to begin the process of creating 5-star linked data and provide the graduate student with a more robust work experience by including editing Wikipedia and working with NARs.

Figure 1: Wikidata form with NAR LCCN field displayed.

Figure 2: Authority Control Template with link to NAR LCCN in Wikipedia page.

After Wikidata and Wikipedia pages were edited to include URIs to NARs the graduate student began the second step of the project, research to see if a person self-identified as African American. The graduate student consulted with the History subject librarian on potentially useful reference sources for the project, and developed his own research methodology. This initially involved a Google Books keyword search for the author's name and "African American" or an equivalent historical term depending on when the author lived. He was initially searching for interviews with the writers where they might speak about their experiences and their identity as an African American. The graduate student's second preferred source was published correspondence, diaries, or other primary source material created by the authors. Finally, the graduate student searched for archival holdings of the author's papers if no other source could be found. When evidence of self-identification as an African American was found he created a citation which was added to an Excel spread sheet, and brief (one

³⁸ Wikipedia contributors, "List of African-American writers," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=List_of_African-American_writers&oldid=855683364 (accessed August 24, 2018).

sentence) biographies including the words “African American” were created to allow for discovery by keyword searching. When evidence of self-identification as an African American could not be found, he included evidence which could be used to infer a person’s membership in the African American community, circumstantial evidence which might indicate a person belonged in that group, or evidence showing that a person identified as an ethnicity other than African American.

While the subject expert was initially asked to examine the entire list of writers and search for evidence of their self-identification as African American, it quickly became evident there were not sufficient funds for hourly wages in the grant budget to complete this task. In order to maximize local impact, student worker Josey Allison searched each name on the list in the local catalog, identified which authors had titles in the local catalog, and added the authorized form of their name to a second column in the spreadsheet. The graduate student then focused on searching for evidence of self-identification for the writers who had created materials in the local catalog.

With the list of citations as justification, the catalogers then began the third step of the project to add demographic group terms to bibliographic records. While evidence of self-identification was not found for all authors in local holdings, the graduate student researcher stated that as a subject expert he felt all names on the list qualified as African Americans, with the possible exception of Marcus Garvey. Entries on the initial list of authorized forms of names for writers with titles in the local catalog were entered individually into an SQL script written by Library Information Technology Services Lab Manager and Voyager Specialist Nancy Boulware. This script returned a list of titles, formats, publication dates, OCLC numbers, and Voyager bibliographic record numbers for each authorized form of name corresponding to an African American writer in local catalog holdings.³⁹ This was done for

³⁹ Nancy Boulware, “Find Bibs Associated with Name Authority Heading,” CARLI: Description of a Library’s Collection, <https://www.carli.illinois.edu/products-services/i-share/reports/secure/sql-desc#isuBibNAR> (accessed May 10, 2019).

Main Entry--Personal Name (MARC 100) and then a separate list was generated for Added Entry--Personal Name (MARC 700) fields. Each list was then examined for entries that might include a writer's authorized for of name but for which they might not be considered a creator. This presented many more interesting cases than was initially expected.

Figure 3: Creator/Contributor Characteristics (MARC 386) fields in a bibliographic record.

Records where best judgment indicated that the item in question would likely not be considered to have been created by the writer were removed from the list, and the records were then pulled by their OCLC numbers. Duplicate records were removed, demographic group terms were added with MarcEdit, and the records were reloaded. Using this method 3,053 bibliographic records in the local catalog and the same number in Worldcat had demographic group terms added. This is the latest portion of the project to be fully completed.

Steps four and five are currently ongoing. In phase four the citation generated by the graduate student is being formatted and added as a Source Data Found (MARC 670) field to the Name Authority Record. A brief free text Biographical or Historical Data (MARC 678) field is also being added to NARs identifying the author as an African American. This allows for the NARs to be located with keyword searching. Staff intended to add MARC 024 (Other Standard Identifiers) fields with the Wikidata number and concept URI, completing the circle with the link to the LCCN already in the Wikipedia page to create 5-star linked data; however, a moratorium by the Program for Cooperative Cataloging (PCC) on adding 024 fields to NARs has caused a temporary halt in this work.⁴⁰

Figure 4: Name Authority Record with Other Standard Identifier (MARC 024) (added pre-moratorium), Source Data Found (MARC 670), and Biographical or Historical Data (MARC 678) fields added.

⁴⁰ Paul Frank, "024 (Other Standard Identifier) data in NACO records (accessed April 19, 2019).

The fifth step is being performed concurrently with the fourth, and is in progress. In this step work level Name Title Authority Records are being created for works by authors in local catalog holdings. This Name Title Authority Record includes the same citation showing self-identification as an African American in Source Data Found (MARC 670) fields, as well as Creator/Contributor Characteristics (386) fields for African Americans and Americans. Finally, when it can be easily determined Date of Creation (046 \$k) fields are being added to the Name Title Authority Record. While catalogs do not currently use authority records to provide Headings-General Information (MARC 3xx) fields in local search results, it is hoped that this will serve as an impetus for further creation and future use. This work is ongoing at the time of this writing, and represents a considerably larger time investment than adding terms to bibliographic records.

Figure 5: Work level Name Authority Title record.

Results

There were 381 names on the Wikipedia list when the project began, and of the 258 entries that were researched, the graduate student worker found evidence that 84% (217/258) self-identified as African American. Evidence that would prompt someone to infer that an additional 12% (30/258) individuals were African American was also found. Circumstantial evidence suggesting (but not rising to the previous level) that 2% (4/258) additional people were African American was also found. Evidence that 3.0% (7/258) of the individuals were members of an ethnic demographic group other than African American was discovered. There were an additional 26 names on the list for that evidence could not be found for, either because it was not available or due to time constraints. These 26 names were not included in the total or any category when calculating the above statistics.⁴¹

⁴¹ Results available upon request, and will be placed on the institutional repository when citations have been converted to MARC 670 format.

When asked if there were any names on the Wikipedia list he as a subject expert felt should not be included, the graduate student subject expert indicated he was uncertain about Marcus Garvey; however, he was uncertain if Garvey qualified as an American citizen, not as an individual of Sub-Saharan African descent.⁴² With evidence showing that 84% of the individuals examined self-identified as African American and 12% could be inferred to be African American (96% total), the hypothesis that many of the names on the Wikipedia “List of African American Writers” are individuals who satisfy Library of Congress criteria for the African Americans demographic group term is supported.

Using citations generated from this research as justification 3,053 local catalog and Worldcat bibliographic records have had demographic data added in Creator/Contributor Characteristics (MARC 386) fields. Work level Name Title Authority Records with demographic information are also being created. The graduate student researcher was paid \$12 per hour for 118 hours, therefore the research for each individual author citation cost an average of \$5.73 (\$1416/247 citations). Some citations resulted in demographic information being added to considerably more bibliographic or Name Title Authority Records than others. This depended on how many titles an author had in the local catalog, and institutions wishing to maximize their research to effect as many records as possible may wish to prioritize research into authors with more titles in their local catalog.

Discussion

While the support for the hypothesis is encouraging, multiple difficulties in using Wikipedia lists exclusively to identify candidates for demographic terms were identified. Wikipedia lists are not

⁴² Garvey was born in Jamaica, and lived in the United States for several years. He applied for United States citizenship; however, he was convicted of mail fraud and deported without receiving his official citizenship.

always complete, and cataloging practice may cause gaps in locating titles by authors through authorized forms of names. Both result in incomplete application of demographic group terms to records. Many works are collaborative in nature, and can create difficulties in determining if it is appropriate to assign a demographic group term to a title that was created by individuals from multiple ethnic groups. Finally, the vocabulary itself is a work in progress and precise terms may not be available for all groups.

The first of these difficulties is that the Wikipedia list of African American writers is incomplete. Michelle Obama, Benjamin Griffith Brawley, and Phillip Hayes Dean are all described as African Americans on their Wikipedia pages, and had titles in local holdings, but were not on the list of African American writers when the project began. For this reason it may be wise to consult directories compiled by subject experts, and edit Wikipedia lists in advance before undertaking similar projects. This sort of editing project may be suitable for students or volunteers, or a Wikipedia edit-a-thon.

Cataloging practice may also cause gaps in identifying creators. In phase one, working from the complete list of African American writers from the Wikipedia page (not the subsequent shorter list of writers with titles in local holdings) the graduate student added ninety-six URIs for NARs in Wikipedia pages. Many of these had NARs but had not been linked yet and were easy to resolve; however, at least one case proved to be interesting. Clarissa Minnie Thompson Allen had no Name Authority Record, although she had written a novel. Further examination showed that this novel had been serialized in newspapers, not published as a stand-alone monograph. Therefore it was never cataloged as a monograph, and Allen did not receive an NAR.

Fifteen names on the Wikipedia List of African American Writers do not have NARs, either because their works are not cataloged by libraries at the author level (Minnie Allen Thompson, for example, and Corey J. Hodges who is a columnist), they were not authors (basketball referee Lonnie Dixon was included in the list for unknown reasons, possibly in error), they self-publish and their

works are not cataloged by libraries in Worldcat (Jeffrey Daniels and Larry D. Alexander), or simply because no library has established an NAR (Christopher C. Bell). Because the local catalog does not contain works by these authors, there was no item in hand and NARs have not been created. While this is a small subset of the total list, it is worth considering that if marginalized groups turn to non-traditional publishing methods any approach that relies heavily on NARs may omit their works. This in turn may make them even more difficult to find if researchers assume all members of a group can be retrieved through browsing an author characteristics facet.

Collaborative works have often presented difficulty for catalogers in defining who should be considered a creator, and this project was no exception. While the Creator/Contributor Characteristics (MARC 386) field allows for inclusion of creators and contributors, what metadata would be most useful to patrons was in collaborative works was, in some instances, a judgment call. The author of a book would clearly be considered a creator, but many writers had portions or excerpts of their work included in documentaries and it could be very difficult to determine if they would be considered a creator or contributor of that particular work by patrons interpreting search results. Editors of works also presented a difficult to judge case, and were generally not considered creators for the purposes of this project. Overall, while there were some borderline judgment calls, the demographic group terms were applied where it was judged that an individual or individuals with evidence of self-identification as an African American would likely be considered the creator of a work by a patron.

Perhaps the most interesting case was that of President Barack Obama. The authorized form of his name was included as an Added Entry--Personal Name (MARC 700) field on over one-thousand records in the local catalog. While it was clear that he would be considered the creator of works such as *Dreams from My Father*, *The Audacity of Hope*, and *Of Thee I Sing*, it seemed less likely that patrons would find retrieving works with his name attached in his capacity as President of the United States based on his status as an African American to be useful. There were a large number of reports and other

documents generated by the Executive Branch of the government in the catalog, but it was difficult to consider these as being created by President Obama. Speeches were another area of uncertainty. While it was often clear he delivered a speech, it could not be easily ascertained if he wrote it or not. In these cases, and because they were delivered in his capacity as President of the United States, he was not listed as a creator. This was a judgment call, and future application of demographic terms would benefit from user studies concerning what they expect from such search results and guidance from Library of Congress.

It should be noted that racial and ethnic demographic terms cannot be placed directly in NARs at this time. While there are fields for Occupation (MARC 374) and Gender (MARC 375) there is no field for racial and ethnic demographic terms. This reflects the policy that demographic terms describe an *item's creator*, but not the *creator* directly. That is to say a heading for African Americans is intended to convey that the item was created by a member of the African American community, not that the creator was African American. While this may seem to be splitting a fine hair, it provides for a level of granularity in describing works. If a person is a professional football player from 2005-2010, are works created by them before 2005 created by a football player? What about works after 2010 that draw on their experiences, but were created when they were no longer an active player? Adding such demographic information to NARs could filter it down to all works by a person, whereas placing it in Name Title Authority or bibliographic records affords more granular control over search results and improves accuracy of retrieval for patrons.

Regarding how to add terms, there are benefits and drawbacks to each of the approaches undertaken in this project. Assigning terms directly to bibliographic records provided more immediate impact and required less staff time per record; however, it only assigned the term to one bibliographic record at a time (two if records in Worldcat and their equivalents in the local catalog are considered as separate records). Records in Worldcat are more widely available, but there are often multiple records

for each work by an author. This leads to a fragmented situation where some records for an author's work will include the demographic term, but others won't. It also requires each institution to add the fields themselves, either by re-importing OCLC records, or modifying local records themselves.

Adding demographic terms to work level Name Title Authority Records might prove more efficient in the future, if catalogs are able to leverage the 3xx fields in NARs and Name Title Authority Records for local discovery. Creation of Name Title Authority Records has proven more time consuming than the bulk addition to bibliographic records described above, but may provide sufficient benefits in the future to justify the time investment. Ideally, work level Name Title Authority Records with demographic terms would replace the need to add demographic terms to bibliographic records. However, there is no immediate benefit to the creation of Name Title Authority Records beyond showing it is possible, and providing examples for those interested in the work or test cases for individuals who might be interested in developing tools to automate the process. One critical advantage that is offered by this method is that the Source Data Found (MARC 670) field justifying the demographic group term is included in the record. By including the citation in the record libraries and researchers can examine it and determine if they feel it is sufficient. While the applicability of some terms may rarely if ever be disputed, for edge or controversial cases this can be valuable information.

Both of these approaches also rely on libraries to have correct authorized forms of names in their bibliographic records. If libraries have not been doing authority maintenance, a cleanup project may be necessary before undertaking a similar project and adding terms to bibliographic records. In the event that catalogs do begin to use 3xx fields from NARs that will also require accurate authorized forms of names in bibliographic records.

Other issues arise with the terminology of the controlled vocabulary itself. The graduate student noted that many African Americans might describe themselves as "black" even though they consider themselves African Americans. Other authors, such as those born in the West Indies, might refer to

themselves as “black” without identifying as African American. LCDGT does list black and African American as related terms; however, even with evidence of self-identification it may be difficult to determine precisely which group an individual is self-identifying as for certain groups. The graduate student researcher also noted that bi-racial authors present an interesting case. While he estimated that around ninety percent of bi-racial authors probably self-identify as African American, it is another area for catalogers to consider carefully.

As pointed out by University of Washington Libraries Principal Cataloger Adam Schiff, certain demographic terms also imply other terms. To include the demographic term African American but not American would likely be confusing to users in search results, since if the term African American applies to an individual, the term American would also likely apply. These terms are also from different categories, with African American categorized as an ethnic term and American as a national term.⁴³ There is no scope note for the demographic term African Americans, but all three sources state some variation of the criteria that it is a person of African or black African ancestry who is an American. Only one source (LCSH) specifies that the person be an American citizen.⁴⁴ There is also no scope note for the demographic term American, and the single source note (from Merriam-Webster) states “American: def. 3: a native or inhabitant of the U.S. : a U.S. citizen.” In general there was no reason to doubt that writers who were found to self-identify as African Americans were also Americans, with the previously discussed exception of Marcus Garvey.

⁴³ Adam L. Schiff, “Applying Library of Congress Faceted Vocabularies,” OLAC 2017 Conference, October 28, 2017, slides 108-112, <https://faculty.washington.edu/aschiff/OLAC2017-ApplyingLCFacetedVocab.pptx> (accessed June 10, 2019).

⁴⁴ Library of Congress, “African Americans,” LC Demographic Group Terms (LCDGT), <http://id.loc.gov/authorities/demographicTerms/dg2015060362.html> (accessed April 19, 2019)

This does however raise the issue that the term African American itself is a blend of ethnic and national characteristics, and not necessarily consistent across a person's entire life. While some demographic groups are much more fluid, there is still the possibility that a person of black African descent born in Jamaica (for example), might create a work and be considered black but not African American by the LCDGT. If they then moved to the United States, obtained citizenship, and created a second work that would be considered by LCDGT to be a work created by an African American. A third work created while they lived in the United States but before obtaining citizenship might occupy in a gray area if they were black, African American, or both terms.

Conclusion and Recommendations

Implementation of the LCDGT controlled vocabularies provides several opportunities for research and discussion going forward. User studies to determine if search and retrieval of resources by patrons is improved through use of demographic terms would be helpful in evaluating if the expenditure of staff time is warranted. A more efficient, automated means of adding terms to Name Title Authority Records (possibly similar to Gary Strawn's invaluable work with the Authority Toolkit) could lower the barriers to creating Name Title Authority Records with demographic information specific to works. As linked data becomes more common further software improvements, such as alerting catalogers to possible demographic group terms for inclusion in the Creator/Contributor Characteristics (MARC 386) field of bibliographic records based on Main Entry (MARC 1XX) fields in the record might be feasible and useful.

These improvements would be especially helpful in adding terms quickly and achieving the critical mass necessary for catalogs to display the information. As practice develops, further guidelines and documentation will be especially helpful in resolving unusual and edge cases. Catalogers may wish to consider if acquiring proof of self-identification is necessary for all demographic group terms, or if subject expert consensus is sufficient. For edge cases, as discussed above, the Name Title Authority

Records provide a way to include citations that provide evidence for why a decision was made, but may not be necessary in all cases. If that is true, a hybrid approach may be beneficial, with terms added to bibliographic records in bulk based on scholarly consensus, and Name Title Authority Records that can provide the evidence used to include a demographic group term.

None of these issues are insurmountable, and the LCDGT represents an exciting opportunity to provide additional access points for users. Given the sheer number of demographic groups a person may belong to in their lifetime it also represents a potential deluge of metadata into records and search results. Challenges are certainly abundant, but as our campuses grow more diverse we can aid our students, faculty, and staff in locating works by creators whose life experiences mirror their own. If our patrons are searching for materials by the demographic characteristics of the creators, our catalog needs to provide that functionality. And as the library catalog is the reflection of our users' intellectual life, we can reflect the diversity of our users in the collection and its records.

Acknowledgements

The authors wish to thank Sandra K. Roe, Merrilee Proffitt, Paul Frank, Janice Young, Adam L. Schiff, Jay Shorten, and everyone else who has generously provided their expertise and wisdom during this project.