Exploring Models for Shared Identity Management at a Global Scale: The work of the PCC Task Group on Identity Management in NACO

Authors: Erin Stalberg, John Riemer, Andrew MacEwan, Jennifer A. Liss, Violeta Ilik, Stephen Hearn, Jean Godby, Paul Frank, Michelle Durocher, Amber Billey

Exploring Models for Shared Identity Management at a Global Scale: The work of the PCC

Task Group on Identity Management in NACO

Running head short title: Identity Management

Abstract

The paper discusses the efforts of the PCC Task Group on Identity Management in NACO to explore and advance identity management activities. The Task Group's work serves as a recent example of how the Program for Cooperative Cataloging has engaged the metadata community to incubate practical solutions to the perennial capacity issues that libraries face in creating and maintaining authority and bibliographic data. The Task Group's multi-pronged charge compelled several outputs including research, collaboration, education, and publication. This paper outlines the Task Group's labors within the context of the decades of innovations carried out under the banner of the PCC.

Keywords: identity management, authority control, bibliographic maintenance, linked library data

Introduction

As cultural heritage communities have advanced the development of Semantic Web-ready standards and technologies in the last decade, libraries have become increasingly interested in minting and linking identifiers to create better pathways for globally shared entity management. Identity management activities have much in common with traditional library authority control in the identification and disambiguation of named entities through the Name Authority Cooperative (NACO) of the Library of Congress and the Program for Cooperative Cataloging (PCC). This paper specifically discusses the work of the PCC Task Group on Identity Management in NACO as described in Strategic Direction 3 of the PCC's *Vision, Mission, and Strategic Directions, 2015-2017*: "[to p]rovide leadership for the shift in authority control from an approach primarily based on creating text strings to one focused on managing identities and entities"[1][2][3] and continued into the fourth item of the 2*018-2021 PCC Strategic Directions*: "[a]ccelerate the movement toward ubiquitous creation and identity management at the network level.[4] Owing to its long history with and investment in authority control, standards development, software applications, and training, the PCC is well-positioned to be a leader in providing frameworks for coordination, education, and experimentation in the realm of identity management.

Over the past several years, the PCC has explored a variety of strategies for encouraging libraries to pursue identity management endeavors:

- The PCC Advisory Committee on Initiatives (2011-2014) introduced the concept of *NACO Lite*, which was intended to lower the barrier for participation in authority work within the context of the existing NACO environment.
- The PCC Task Group on URIs in MARC convened in 2015 to develop strategies for managing identifiers in our existing systems, as libraries began attempting to expose MARC on a large scale as linked data.

- The PCC Task Group on Identity Management in NACO was charged in 2016 to investigate and lead coordinated PCC discussion on transitioning the community from authority control to identity management.

- In 2017-2018, the PCC launched a pilot to test the principles of identity management within the International Standard Name Identifier (ISNI) environment: minting, sharing, and linking identifiers, and creating metadata enrichment lifecycles that enable distribution of identity management activities.

- And the community has recently begun to examine the viability of contributing library data to Wikidata and minting identifiers within the Wikibase platform.

While there is still much work to do, this paper shares the efforts of the PCC Task Group in Identity Management in NACO to date and proposes areas for further experimentation.

The work of the PCC Task Group on Identity Management in NACO builds in large part on earlier discussions of the PCC Advisory Committee on Initiatives (2011-2014). The Advisory Committee aimed to enable greater contributions from digital library projects and, potentially, non-PCC participants into shared authority files. In their work, the Advisory Committee identified several barriers to participation in the NACO program -- including extensive training and review requirements, outdatedness of relevant documentation, and the inability or impracticality of contributing new entities to central, shared files -- all of which result in a duplication of effort and artificially constrain the number of librarians who might participate in creating and maintaining authority data in an environment growing increasingly complex and connected everyday.[56]

In the early days of its work, the PCC Task Group on Identity Management in NACO developed definitions of the two key terms: authority control and identity management. These definitions were intended to establish a baseline understanding upon which the PCC would build out its role in leadership and education for the community. The definitions are:

Authority control manages access to entities by authorizing a specific form of name or other term for access use and bibliographic file maintenance. A distinguishing feature of authority control is its focus on a preferred, unique, human-friendly access point.[7]

Identity management operates by associating a registered identifier with characterizing data which specify a single identity or identified entity. It is distinguished by its focus on differentiating entities through the use of identifiers. Differentiation of entity names is of secondary importance and may be accomplished by assigning numbering to or concatenating identifying information with instances of the same name in presentation contexts.

The Task Group's subsequent research revealed nine use cases demonstrating opportunities for extending library authority control into the space of identity management. The use cases, which stemmed from libraries, archives, museums, galleries, and other stakeholders, were described at length in several educational sessions given by Task Group members, including a presentation delivered at the 2017 ALA Midwinter Meeting.[8]

The PCC Task Group on Identity Management in NACO then began to seek opportunities for experimentation, taking the identified use cases and trying to situate them in the context of a possible new identity management environment for the PCC. Connecting back to the efforts of the PCC Advisory Committee on Initiatives, this new Task Group revisited the Advisory Committee's earlier work on an idea called *NACO Lite*.

NACO Lite: vision, definitions and implications

The phrase *NACO Lite*, first developed by the PCC Advisory Committee on Initiatives, has now been in use for over five years in the literature[9]. It has been variously used to refer to

streamlined procedures for contributing authority data and to broadened participation in that activity. A 2015 comparison showed that the Library of Congress Name Authority File (LCNAF) coverage of access points in OCLC WorldCat bibliographic records was only about 33%: 10 million authority records and about 30 million entries in the WorldCat Identities project.[10] The currently unfolding shift in emphasis from heading construction to identity management raises hopes of reducing the delta between entities supported by controlling structures (identifiers and/or authority records) and those that are not. The PCC Task Group on Identity Management in NACO aimed to advance mechanisms for achieving the earlier Advisory Committee's goal.

The Task Group considered several strategies for implementing a version of the imagined *NACO Lite* program where broader participation in authority work could be facilitated. The earliest approach envisioned adding another tier to the existing LCNAF to accommodate contributions from outside of the traditional PCC and MARC communities. This would inevitably require developing conventions for when and how a record in the second tier would get elevated to the first tier, as well as some type of combined indexing. As Task Group discussions evolved, however, it seemed possible to have more participants contributing to identity management activity using one of the existing files directly if the emphasis of authority work shifted to the identifier. The ability to discern which access points and existing identifiers should be connected and when a new identifier should be established are skills readily found in the library community, well beyond just NACO-trained catalogers.

In 2016, an idea emerged in the Task Group of collaborating with the ISNI community -- offering a potential new setting to execute the vision of *NACO Lite* beyond simply the LCNAF. Both ISNI[11] and ORCID[12] are organizations actively addressing aspects of unambiguously identifying people and parties on the web. ORCID, however, is restricted to individual researchers and does not share the underlying paradigm in which libraries complete identity management work

on behalf of individuals and organizations. In contrast, ISNI does largely share the library paradigm and presented desirable features such as batch matching and batch loading, fewer data elements with "justification" requirements, and no requirement for a unique text string. The PCC Task Group for Identity Management in NACO's inquiry into ISNI workflows hearkened new opportunities for partnership.

PCC ISNI pilot

ISNI (International Standard Name Identifier) is an ISO standard in the family of the ISSN and ISBN used to uniquely identify persons and organizations. In 2016, at the recommendation of the Task Group for Identity Management in NACO, the PCC leadership authorized exploration of a PCC umbrella membership in ISNI in order to further identity management experimentation within the organization. By the spring of 2017, a cohort of early-adopter institutions volunteered to learn ISNI and pave a path for ISNI adoption within the PCC.

Thirteen institutions pooled resources to fund the initial cost of ISNI membership for the PCC and the group began working with ISNI tools and services to gain experience creating ISNIs. With access to the ISNI platform, the pilot institutions pursued a range of institutionally-defined projects and explored the difference between NACO work and identifier creation, supporting many different use cases.

As the pilot group gained experience with ISNI, their questions illuminated areas in need of clarity for when PCC members transition in large scale toward identifier creation. Among the findings is that working with ISNI proved to be nearly as time consuming as creating records for LCNAF, given the issues and problems with the tools and, while NACO-trained catalogers had extensive skills that they applied to ISNI creation, they were also accustomed to the more precise rules associated with traditional authority control as compared with the type of data

policies associated with ISNI.[13] Even with training and documentation developed to guide PCC members in their future ISNI contributions, it became clear that certain mental models will need to be adjusted for the community to move to identity management. In addition, expanded PCC participation in ISNI requires a revenue model not currently in place for the PCC's volunteer organization. Member pricing for the PCC and alternative financial models for libraries are under active discussion with the ISNI International Agency. A recent ISNI Summit meeting at the Bibliothèque Nationale de France led to the formation of a library sector steering group (with representation from the PCC) that will further explore the place of libraries within the ISNI community.

PCC URIs in MARC

Another effort of the PCC was the Task Group on URIs in MARC, convened in 2015 to increase the prevalence of URIs in MARC bibliographic and authority data and, thereby, facilitate the transition from current standards to linked data. According to the charge:

> "Globally unique identifiers, given in the form of URIs, are a key requirement for the effective dissemination of data on the Semantic Web. Traditional library cataloging practices have not favored the explicit provision of identifiers and this has been an obstacle to the library community's ability to exploit linked data technologies."[14]

By adding URIs to MARC, this group aimed to realize three benefits. First, the investment in existing standards can be leveraged if conversions to BIBFRAME and other linked data ontologies are able to preserve as much detail as possible. Second, URIs in MARC records provide a glimpse of what a linked data future can offer before the transformation is fully realized in the library community. Finally, the addition of URIs to bibliographic data democratizes the transformation because catalogers can participate in the creation of linked data simply by performing tasks in their current jobs using current technologies.

To lay the groundwork for future innovations, the PCC Task Group for URIs in MARC developed policies and best practices for the use of linked data identifiers, defined requirements for tool development, conducted surveys and pilot studies, and proposed modifications to the MARC bibliographic and authority standards.[15]

The proposals approved by the MARC Advisory Committee to capture URIs in MARC $0 and $1 subfields in bibliographic and authority records (MARC Proposal Nos. 2019-03[16] and 2017-08[17]) represent the most significant connection between the PCC task groups on URIs in MARC and Identity Management in NACO. These two subfields encode the essential difference between authority control and identity management, as defined by the two task groups. In the example below, discussed in MARC Proposal 2017-08, the $0 URI cites the source of library authority control for the heading 'Obama, Michelle' shown in the $a subfield. In contrast, the newly defined $1 URI resolves to a wealth of biographical information accessible from Wikidata, DBpedia, and other resources maintained outside the library community about the woman named Michelle Obama, who served as the First Lady of the United States from 2009-2016.

100 1# $a Obama, Michelle, $d 1964- $0 http://id.loc.gov/authorities/names/n2008054754 $1 http://viaf.org/viaf/81404344

The implications of this distinction are being discussed in various contexts and in research and production environments throughout the library community in North America and Europe. To further exploration in the PCC, the Policy Committee authorized a URIs in MARC pilot in August 2019 to be led by the PCC Task Group on Identity Management in NACO. As with the ISNI pilot, the launch of this initiative gives the PCC community opportunity to gain concrete experience

and uncover issues emerging in practice, in addition to the more abstract linked data

discussions that are ongoing in the community.

This pilot will test and develop PCC best practices for use of subfield $0 and subfield $1 in both

bibliographic and authority records. The work will address, but not be limited to, the:

- Use of Real World Object URIs in subfield $1 vs traditional library identifiers

- Use of identifiers from traditional as well as nontraditional sources

- How to record alternative identifiers from multiple sources

- Preference among vocabulary sources

- Use of the 024 field in NACO authority records.

In consultation with the PCC Task Group on Identity Management in NACO, pilot participants

will experiment with one or more of these focus areas and share their findings. By

demonstrating the integration of Semantic Web URIs as an alternative to controlled terms for

managing the relationship of entities to library resources, the pilot will mark another step toward

the use of linked data in library catalogs.

Upcoming Area for Research: a Wikidata Pilot

The PCC Task Group on Identity Management in NACO continues to incubate practical

solutions for identity management in libraries and push to execute the spirit of *NACO Lite*. In

addition to ISNI and the URIs in MARC pilots described above, the Task Group is now turning

attention to Wikidata. Wikidata provides both a platform that instantly creates linked data in a

collaborative environment *and* that is freely available and open, unlike the other registries

explored to date.[18]

Within all the Wikipedia content, much of it is structured data and the underlying infrastructure

for that data is managed through Wikidata. Wikidata, like Wikipedia, is hosted by the Wikimedia

Foundation on the Wikibase platform and it "is intended to provide a common source of data

which can be used by Wikimedia projects such as Wikipedia and by anyone else, under a public

domain license"[19].

In their recent article, Allison-Cassin and Scott discuss how they successfully used Wikidata as

a low-barrier method to mint URIs for Juno award-winning Canadian musicians.[20] They note

how easy it was to use the platform and how quickly they saw results. With their newly minted

URIs, Scott built a proof-of-concept extension to the Evergreen ILS that harvests metadata from

Wikipedia about the musicians in order to provide contextual information in their records.

To expand on the work of Allison-Cassin and Scott and other librarians experimenting

with Wikidata, the PCC Task Group on Identity Management in NACO is now organizing a

Wikidata pilot project. To gauge member interest and to determine possible parameters for a

pilot, the PCC sponsored a Wikidata workshop immediately following the annual PCC

Operations Committee meeting in May 2019. Participants created Wikidata identities for

persons and corporate bodies associated with their institutions -- which was one of the use

cases identified by the Task Group at the start of this work. The Wikidata workshop was a

success, with over 75 participants at the live session, and many more participants joining

remotely through live streaming of the event, and it demonstrated that the use cases framing

could prove a fruitful path forward to guide the community in this transition.

Conclusion

The shift towards using identifiers as the primary means of controlling and differentiating entities

offers numerous benefits -- from more actionable metadata and greater ease of production to

greater coverage of the entities contained in resource descriptions, and alignment of library efforts with those of other viable partners. In its multi-pronged efforts to facilitate transition in this space, the Task Group on Identity Management in NACO takes its educational role to heart by surfacing issues in need of attention and resolution in order for identity management to be adopted as the primary focus of NACO members. The research into and experimentation with ISNI and Wikidata -- while simultaneously encouraging a growth in the number of URIs in existing MARC records -- leverages the PCC's historical role in leadership and learning to expand the possibilities and partnerships for how libraries might now achieve our goals around the identification and connection of named entities found in library data.

---

[1] Program for Cooperative Cataloging, *Vision, Mission, and Strategic Directions January 2015-December 2017 Revised: November 20, 2015* (2015), p. 4-5. Accessed August 28, 2019. https://www.loc.gov/aba/pcc/about/PCC-Strategic-Plan-2015-2017.pdf

[2] Program for Cooperative Cataloging, *PCC Identity Management in NACO Task Group Charge* (2016). Accessed August 28, 2019. https://www.loc.gov/aba/pcc/documents/Identity-management-NACO-PCC-TG.pdf

[3] Program for Cooperative Cataloging, *PCC Identity Management in NACO Task Group Updated Charge* (2018). Accessed August 28, 2019. https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf

[4] Program for Cooperative Cataloging, *PCC Strategic Directions January 2018-December 2021* (2018) p. 5-6. Accessed August 28, 2019. http://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-20182021.pdf

[5] Program for Cooperative Cataloging, *PCC Advisory Committee on Initiatives Charge* (2011). Accessed August 28, 2019. http://www.loc.gov/aba/pcc/resources/Adv-Initiatives.doc

[6] Program for Cooperative Cataloging, PCC Advisory Committee on Initiatives, *Name Authorities in Transition: Implications for the PCC* (2014). Accessed August 28, 2019. https://www.loc.gov/aba/pcc/documents/PCC-white-paper-ACI.pdf

[7] In some cases, authority control establishes a single name to represent more than one identified entity (e.g., undifferentiated personal names in LCNAF, some animal names and geographic names in LCSH), but there is evidence that this practice is waning. Current LC-PCC policy prohibits creating or updating undifferentiated personal name authorities (Library of Congress, 2016). Library of Congress*, Descriptive Cataloging Manual, Z1, 008/32*, (Washington, D.C. : Library of Congress, 2016). Accessed August 28, 2019. https://www.loc.gov/catdir/cpso/dcmz1.pdf

[8] Riemer, John, Ilik, Violeta, and PCC Task Group on Identity Management in NACO. "Authority Control' see (also) 'Identity Management'." LITA/ALCTS Authority Control Interest Group. American Library Association Midwinter Meeting, Atlanta, GA, January 22, 2017. http://connect.ala.org/node/261280

[9] MacEwan, 2014 and Durocher, Cuneo, Riel, Moody, & Eslao, 2016 are two of the two oldest documented uses of the phrase: MacEwan, Andrew. "NACO and the future of authority control: why the BL is working with ISNI." PCC Policy Committee presentation, November 3, 2014, p. 7 of handout. Accessed August 28, 2019. https://www.loc.gov/aba/pcc/documents/ISNI-NACO-PoCo-2014.pdf and Durocher, Michelle, Cuneo, M.J., Riel, Steven, Moody, Honor, & Fernsebner Eslao, Christine. "NACO Lite?: Reimagining name authority work as identity management." ALA Midwinter PCC Participants' Meeting, January 10, 2016. Accessed August 28, 2019. https://www.loc.gov/aba/pcc/reports/ALA_Midwinter2016_PCCParticipantsMtg.pptx

[10] Smith-Yoshimura, Karen. "Getting identifiers created for legacy names." Hanging Together: the OCLC Research blog, October 30, 2015. Accessed August 28, 2019. http://hangingtogether.org/?p=5463

[11] ISNI. *ISNI website.* Accessed August 28, 2019. http://www.isni.org/

[12] ORCID. *ORCID website.* Accessed August 28, 2019. https://orcid.org/

[13] Fletcher P., Dagher, I., Long, C., & Mak, L. "PCC ISNI pilot: experiments in identity Management." LITA/ALCTS Authority Control Interest Group. American Library Association Annual Meeting, New Orleans, LA, June 24, 2018. Accessed August 28, 2019. https://wiki.duraspace.org/display/PCCISNI/PCC+ISNI+presentations?preview=/96999238/101782660/ALA_ACIG_PCCISNI_UnifiedSlides_June2018.pptx

[14] Program for Cooperative Cataloging. *PCC Task Group on URIs in MARC.* (2016). Accessed August 28, 2019. https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html

[15] Program for Cooperative Cataloging, *PCC Task Group on URIs in MARC* (2016). Accessed August 28, 2019. https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html

[16] Library of Congress, *MARC proposal no.2019-03: Defining Subfields $0 and $1 to Capture URIs in Field 024 of the MARC 21 Authority Format* (2017). Accessed August 28, 2019. https://www.loc.gov/marc/mac/2019/2019-03.html

[17] Library of Congress, *MARC proposal no.2017-08: Use of Subfields $0 and $1 to Capture Uniform Resource Identifiers (URIs) in the MARC 21 Formats* (2017). Accessed August 28, 2019. https://www.loc.gov/marc/mac/2017/2017-08.html

[18] Most catalogers working in MARC and creating authorities in the LCNAF do so using a subscription membership to OCLC which, like ISNI, is proprietary and requires a significant financial commitment from libraries. The LCNAF also does not mint actionable URIs on the fly. Once a record is created in the LCNAF, an index cycle run by the Library of Congress pushes the metadata to their Linked Data Service where a stable URI is published and maintained.

[19] Wikidata. *Wikidata website.* Accessed August 28, 2019. https://en.wikipedia.org/wiki/Wikidata

[20] Allison-Cassin, Stacy & Scott, Dan. "Wikidata: a platform for your library's linked open data." *Code4Lib Journal*, 40. (2018). Accessed August 28, 2019. https://journal.code4lib.org/articles/13424