

## The PCC ISNI Pilot: Exploring Identity Management on a Global, Collaborative Scale

Michelle Durocher, Iman Dagher, Violeta Ilik, Chris Evin Long, Jeanette A. Norris, Isabel del Carmen Quintana & Manon Th  roux


To cite this article: Michelle Durocher, Iman Dagher, Violeta Ilik, Chris Evin Long, Jeanette A. Norris, Isabel del Carmen Quintana & Manon Th  roux (2020): The PCC ISNI Pilot: Exploring Identity Management on a Global, Collaborative Scale, *Cataloging & Classification Quarterly*, DOI: [10.1080/01639374.2020.1713952](https://doi.org/10.1080/01639374.2020.1713952)

To link to this article: <https://doi.org/10.1080/01639374.2020.1713952>



Published online: 30 Jan 2020.



[Submit your article to this journal](#) 



Article views: 139



[View related articles](#) 



[View Crossmark data](#) 



## The PCC ISNI Pilot: Exploring Identity Management on a Global, Collaborative Scale

Michelle Durocher<sup>a</sup> , Iman Dagher<sup>b</sup> , Violeta Ilik<sup>c,d</sup> , Chris Evin Long<sup>e</sup> , Jeanette A. Norris<sup>f</sup> , Isabel del Carmen Quintana<sup>a</sup> , and Manon Th  roux<sup>g</sup> 

<sup>a</sup>Harvard University, Cambridge, Massachusetts, USA; <sup>b</sup>UCLA, Los Angeles, California, USA; <sup>c</sup>Stony Brook University, Stony Brook, New York, USA; <sup>d</sup>Columbia University, New York, New York, USA; <sup>e</sup>University of Colorado Boulder, Boulder, Colorado, USA; <sup>f</sup>Brown University, Providence, Rhode Island, USA; <sup>g</sup>Library of Congress, Washington, DC, USA

### ABSTRACT

The Program for Cooperative Cataloging (PCC) conducted a multi-institution Pilot project to contribute metadata to the International Standard Name Identifier (ISNI) database. In addition to describing the goals, structure, and activities undertaken in the Pilot, this article examines some of the experiences of PCC metadata practitioners as they sought to incorporate ISNIs into their metadata workflows and pursue a wide range of ISNI use cases. Lastly, the article outlines next steps that flowed from the Pilot as the PCC has parlayed its newly developing identity management expertise into additional PCC program explorations.

### ARTICLE HISTORY

Received December 2019  
Revised January 2020  
Accepted January 2020

### KEYWORDS

Authority control; experimental; cataloging research; cooperative cataloging; interoperability; metadata; catalogers; identifiers

The Program for Cooperative Cataloging (PCC) set a bold, new direction when it released a document in January 2015 that it called *Vision, Mission, and Strategic Directions: January 2015-December 2017*.<sup>1</sup> Among its signature aspirations was a call to the PCC membership to play a leadership role in the library community's transition to linked data, particularly "leading the transition to a new set of practices that promote the use of unique identifiers."<sup>2</sup> While the PCC urged this internal shift in practice from text string-based authority control to one focused on identifiers, it also directed attention outside itself to seek collaborations with expert communities outside of libraries and to lower barriers to participation in the PCC.

In November of 2016, the PCC Policy Committee (PoCo) endorsed pursuing an umbrella membership with the International Standard Name Identifier (ISNI).<sup>3</sup> As an ISO standard identifier with wide adoption by many library partner organizations, such as publishers, music and film industries, and a wide and increasing number of global national libraries, ISNI was chosen as an appropriate platform to explore identity management in a global, shared environment because of its stability and

prominence. PoCo leadership agreed that a PCC ISNI Pilot project would be the best way to put its strategic directions into action and to learn about the needs and issues that would arise for libraries, ISNI, and the PCC when metadata practitioners began incorporating ISNIs into their metadata workflows. With ISNI's scope being personal and organizational names, PoCo anticipated PCC institutions being able to pursue a wide range of ISNI use cases.

In order to be most inclusive in drawing together participants for the Pilot, interest was solicited as an open call on the PCC List in February 2017. The invitation made clear that this early-adopter group would receive support from a peer network rather than the typical PCC model of extensive and comprehensive documentation, training, and guidelines; however, the group was enticed by the opportunity to shape a new shared path for libraries to create and manage identifiers for persons and organizations. Those who stepped forward to participate also agreed to collectively share the cost of setting up the Registration Agency membership in ISNI as well as establishing a small pool of funds to address any other expenses that could be encountered, such as the cost for batch processing and dedicated scripting that was anticipated. After lengthy contract review and discussions with the ISNI-International Agency (ISNI-IA) executive director, the PCC ISNI membership was established in July 2017, with the Pilot getting underway immediately with a webinar-style kickoff meeting.

The Pilot participants were comprised of twelve academic libraries: Brown University; University of Chicago; University of Colorado Boulder; Columbia University; Cornell University; Michigan State University; Stanford University; Stony Brook University; Texas A&M University; University of California Davis; University of California, Los Angeles (UCLA); and University of Wisconsin-Madison. Coordination for the Pilot was provided by a team of metadata staff at Harvard Library. Harvard staff had developed expertise working in the ISNI systems since joining ISNI independently in 2015, and they were enthusiastic to offer their guidance to train colleagues new to ISNI. Along with the invaluable partnership of the ISNI technical team in the OCLC Leiden office and the ISNI Quality Team at the British Library, Harvard developed extensive self-service documentation for learning the ISNI data structures, best practices, and practitioner interfaces. With generous support from Duraspace (now Lyris), a Pilot wiki<sup>4</sup> was set up where all the documentation and training resources were hosted.

At the start of the Pilot, remote meetings were offered regularly to help the group gain mastery of ISNI data elements, best practices, and tools. Two electronic mailing lists were set up, one for general Pilot updates and announcements, and another to create a safe space for metadata practitioners to ask each other questions or share detailed information about what was working

best for them. While the lists were fairly low-traffic, some teams shared their project milestones as encouragement to others and proudly announced their “ah ha!” moments as they overcame obstacles in their learning process.

In addition to the institutional goals associated with each Pilot participant, they each also volunteered to contribute to one or more of five joint focus areas: ISNI Application Programmers Interface (API) review and testing, batch processing, documentation and training, ISNI maintenance best practices, and workflow and tools evaluation.<sup>5</sup> Each subgroup defined their own ways of working together and divided the work according to their skills and interests.

By late Fall 2017, a first quarter check-in survey was sent to Pilot participants to gauge their progress in learning ISNI and their level of comfort in the systems so far. The initial results revealed some pain-points regarding the self-service documentation and lack of formal training. New training videos had recently been posted to the Pilot wiki and many were optimistic those would meet their needs; those who had already viewed them were enthusiastic about the help they provided. There were also several comments that referenced the “wild West” feeling that Name Authority Cooperative Project (NACO) participants felt as they navigated the differences between the heavily prescriptive practices associated with traditional name authority work, compared with the few formal rules and open-endedness when establishing an identifier in ISNI. Some could see ahead that scaling up future work in ISNI to the broader PCC community would require defining what would constitute a rich ISNI entry to meet PCC standards. Additional detail about practitioner experiences are described colorfully below.

### **Pilot institution experiences**

In this section, participants from five of the institutions involved in the PCC ISNI Pilot describe their experiences from a mix of perspectives. Some provide more of a high-level institutional view; others offer reflections on a more personal level. Their accounts of their experiences are presented in alphabetical order by institution name.

#### **Brown University Library—Jeanette A. Norris**

Brown University Library’s ISNI Pilot project involved enhancing, correcting, and creating ISNI records for the university’s 209 academic centers and departments. We chose to work on these academic units because we use them in three different services managed by the library: the faculty profile system, the digital repository, and the catalog. Each of these systems approach the access points for university departments differently. We were able to invest in the project because we have a strategic interest in improving

data about Brown in international databases and in exploring options for improving the interoperability of data in systems that we manage.

Five staff members participated in the project, including a mix of experienced catalogers and staff who were not catalogers but had experience managing MARC record loads and authority file maintenance. Previous experience changed how each person approached the creation of ISNI records, but it did not affect any individual's ability to learn how to effectively create and edit records. All of the participants reported having enjoyed the project and found value in exploring new methods of creating and managing entity data. The team found that 133 of the 209 academic units were already represented in the ISNI database. Ringgold, an organization primarily interested in the financial side of scholarly communications, contributed many of those existing records. The finding that many of the records were not created by other libraries highlights the value of engaging with communities outside of libraries for the creation and management of data that has shared relevance. Additionally, thirty-three of the departments had duplicated records that we were able to merge and correct.

Through the project, we found distinct advantages to the ISNI model for organizations in that it allows us to track the name variations over time and use single identifiers to represent them. Brown University's academic units are relatively fluid. Names change frequently, even when the essence of the department does not change, and the organizations have a variety of relationships with other organizations on campus over time. While tracking individual name changes has benefits, we do not use the same model in all of our systems. Having an identifier that represents the unit across systems is important to creating meaningful links between systems.

Brown University Library is beginning to explore opportunities to build a service that can reuse ISNI data to link together information in our repository, faculty profile system, and the catalog without having to change much of the existing data in any of those individual systems. This project includes working with programmers who are interested in using the data to create efficiencies in their work, librarians interested in exploring the integration of data to improve the discoverability of resources, and university archivists interested in contributing their knowledge of the history of the organization to improve discoverability of archival collections. We are still experimenting, and we are excited about the opportunities afforded by further exploration and the expansion of our knowledge and expertise.

### **Harvard Library – Isabel del Carmen Quintana**

I very much appreciated the ability to take part in the PCC ISNI Pilot. As a NACO contributor for over twenty-five years, I had strong conceptions

about the value of identity management and how to structure a workable authority file. Participating in this project helped me see identity management in a different way. While my grounding in NACO work certainly helped me understand the metadata, being able to work in ISNI expanded my thoughts on identity management and the value of authority files.

My work in the Pilot involved searching Harvard faculty from two departments in both the LC/NACO Authority File and the ISNI database, evaluating whether the faculty names were correctly represented in the files, cleaning up the data, and creating new records as needed. I worked in both the ISNI web interface and the ISNI client, but preferred the web interface. I very much appreciated the amount of data stored in ISNI and how well it was represented to the user via a simple search.

I discovered several practical advantages to using ISNI. For example, when doing authority work for an author who writes in native vernacular characters as well as in English, NACO guidelines require me to choose the predominant form of name, which can be time-consuming. The name is easier to create in ISNI where I do not have to worry about a “preferred form.” Because my institution’s OPAC links variant forms of names from the authority record to our bibliographic records, our patrons do not gain anything by the choice of a “preferred form.” I also found that ISNI sometimes has a long list of works associated with an entity, which can be very helpful for choosing the correct entity to assign to a work during cataloging, especially when the person writes on various topics – for example, a mathematician who also writes poetry. In my NACO work, I have started using MARC 672 fields to add other titles associated with an author. I see the value in this field in a way that I did not before I started working in ISNI.

My work in ISNI has also shown me how difficult and important it is to provide patrons with solid identity management. Since ISNI data is primarily batch loaded and then cleaned up by librarians, the ISNI database has a lot of duplicates. Despite our advances in technology, it still takes human intervention to perform identity management and sort through persons with similar or identical names. During the Pilot, I was struck by the importance of this work and the enormity of the task.

The project was both humbling and deeply rewarding for me. I learned a lot during the Pilot – not only in terms of developing skills necessary to work in a new system, but also in terms of theories of identity management. Since working on the Pilot, I have asked myself often: Would this identity be better served in NACO or ISNI? Perhaps most importantly, I have changed some practices in my daily NACO work, because this project has helped me to focus on patron needs and the best ways to perform identity management in an age when there are a plethora of identities on the web.

### **Stony Brook University Libraries – Violeta Ilik**

Stony Brook University Libraries joined the PCC ISNI Pilot with the intention of creating ISNI records for the university's colleges and departments. The project team began its work by performing an environmental scan to evaluate databases containing information on the university's organizational structure. The team discovered that our institutional repository, built on DSpace infrastructure, was the best data source for our project. The repository manager provided the team with a spreadsheet containing data on various categories of corporate bodies within the university. The team decided to assign ISNIs for the university's schools, colleges, centers, institutes, and departments, but not for its offices, programs, and stores. The four team members created or updated a total of ninety ISNI records during the project. All team members were comfortable working in the ISNI web portal; only one team member used the ISNI client to create records. The documentation on how to use the web portal and the client was provided in a common wiki space. All questions that came up were first shared among the Stony Brook team members and if answers were not readily available the team reached out to the other PCC ISNI cohort members. The team lead performed the final quality check for all records that were added to the production database.

Major takeaways from our work on the PCC ISNI Pilot involved the clear differences between using NACO and ISNI. One similarity that was observed is that, in both, metadata specialists should strive to have only one record for an identity. The majority of the team was clearly more accustomed to working on NACO records and it took time to get accustomed to working on the ISNI database. Everyone expressed deep gratitude for the opportunity to participate in the Pilot and learn how to use the ISNI web portal.

### **UCLA Library – Iman Dagher**

From my perspective as a NACO cataloger, getting involved with the PCC ISNI Pilot was a rewarding experience. It provided an opportunity to explore working with an identity management system, to describe and control entities outside the NACO environment, and to shift my focus from standardized text strings to unique identifiers. Although the goal remained identifying, differentiating, and linking between entities, working on the ISNI project required adjusting my train of thought and my pattern of work.

PCC catalogers spend a good amount of time and effort selecting unique text strings – reflecting the common usage of entities – for use as authorized and variant access points in accordance with cataloging instructions.

In the ISNI environment, uniqueness and standardization of text strings is not mandatory, so the first challenge is abandoning rigid rules and thinking outside of *Resource Description and Access* (RDA) and NACO instructions. With ISNI, catalogers are directed to loosely construct their preferred forms and accept forms entered by other institutions. There can be multiple preferred names (which the ISNI database records in MARC-like 7XX fields) in various forms, scripts, etc., all equally valid and not necessarily stable, along with variant names of one's choosing (4XX fields).

Checking the LC/NACO authority file for conflicts is no longer part of the workflow, although a thorough search in the ISNI database is necessary to avoid duplication of entries. Identities can be differentiated by means other than string matching. ISNI relies on sets of elements (date of birth, affiliated institution, language, associated country, etc.) and relationships (cited works, coauthors, organizations, etc.) to facilitate matching and disambiguation and prevent duplication. These elements resemble RDA entity attributes (MARC 046 and 3XX fields) but they are not rule-dependent and can be taken from controlled lists (ISO codes) or chosen from lists of terms available in the system. NACO catalogers also work meticulously to justify their choices of authorized access points, variants, and elements by citing sources in a structured manner. With ISNI, such justification is not required. Although NACO and ISNI are both collaborative forums, NACO catalogers have the privilege of changing others' contributions while one cannot alter other institutions' data in ISNI. Finally, the NACO program requires a great deal of training and a review period to reach independence. Learning and training in ISNI, on the other hand, is a simple and straightforward process. Even a student assistant without any knowledge of cataloging techniques may assist in creating ISNIs for public identities needed for a library's projects.

### **University of Colorado Boulder Libraries – Chris Evin Long**

Our project involved creating or revising ISNIs for the faculty and academic units in our College of Engineering and Applied Sciences. We thought this was a big enough department for us to get a sense of the work and the issues we might encounter, but small enough to be something we could complete. We found a ready partner in our campus's Faculty Affairs Department, which supplied us with information about the faculty and the units in the college – about 250 faculty and 25 units. It would have taken us a good bit of time to gather this information on our own.

We decided not to create NACO records concurrently with this project (largely because of time constraints), so we were able to assemble a team with a mix of staff, including some who did not have NACO training. The



biggest challenge for the NACO-trained staff was suppressing our “NACO brains” and not spending time on tasks that do not apply in ISNI. We did not have to worry about creating unique headings, or determining preferred forms of names, or finding information to justify every assertion we made in the ISNI record, as we do with NACO. At the outset we often found ourselves wanting to put a NACO lens on ISNI, but eventually we gave ourselves permission to enjoy the greater freedom in the ISNI approach. We completed our project in about three months. Something that surprised me about the results is that about half the work was maintenance – mostly modifying existing ISNIs to include our organization information and merging duplicate records for our faculty.

Post-Pilot, I assessed what we might do with ISNIs in the future and decided that we should create or enhance ISNIs for all the faculty and units on our campus. I believe, and Faculty Affairs seems to agree, that this will help our campus harvest its research productivity. The faculty information we get from Faculty Affairs now includes ORCID, and we have retrospectively added ORCID to our engineering faculty’s ISNIs, a function that was not available during our initial project. Due to the size of our campus and staffing constraints, our ability to create or modify ISNIs for all faculty and units is predicated on the success of a batch loader, the status of which is uncertain to me at this time. We also intend to create and revise ISNIs in tandem with a wiki project we are initiating involving women poets of the Romantic Period. ISNIs will be created and added to the Wikidata entities we create for this project. Our commitment to sustaining our ISNI work is reflected in the fact that we have added ISNI job duties to a new position we recently created, which involves database integrity and identity management functions.

### **Concluding the Pilot, lessons learned and PCC next steps**

The PCC ISNI Pilot concluded at the end of calendar year 2018. Throughout the duration of the eighteen-month Pilot, members of PCC and ISNI-IA leadership had extensive discussions about Pilot experiences. The PCC and ISNI-IA shared a keen interest in learning about the training and tooling needs of metadata practitioners, the effectiveness of incorporating ISNI into efficient, production-level workflows, and the ideas from the Pilot group to address ambiguity in metadata practices that emerged as anticipated in this first PCC foray into identity management. With the goal of addressing library needs more directly and providing an avenue for library-sector shared priorities, the ISNI-IA formed an ISNI Library-Sector Steering Group in August 2019. A PCC representative has been appointed to this group, which also includes several European national libraries.

One area that posed a significant challenge for practitioners was setting data integrity expectations appropriate to the ISNI database when shifting mindsets from the LC/NACO file. For example, with the ISNI database being primarily built through large-scale batch loading from a wide range of library and non-library data sources, metadata practitioners encountered less consistency in data entry and completeness in the ISNI database compared with their experiences in the manually curated environment of the LC/NACO file. While many existing ISNIs were found for the entities being searched and fewer than anticipated, original ISNIs needed to be created from scratch, metadata practitioners often had to merge duplicates and enrich the data to more effectively disambiguate an identity. This experience points to a changing role for catalogers who will be working in files like ISNI, where we can anticipate that human expertise will supplement and complement the role of batch loading in order to meet the large-scale need for identifiers in library and non-library sectors.

In the Pilot work undertaken as joint focus areas, some work proved more fruitful than others in terms of immediately applicable recommendations; however, in all cases, practitioners gained deeper knowledge of the benefits and challenges associated with the discrete ISNI function or service they were analyzing:

- The ISNI API analysis revealed a need for more extensive and up-to-date documentation for both the public SRU API as well as the member AtomPub API. Gaining access to the latter was cumbersome and difficult, given that it required registering static IP addresses which some participants did not have. The group hopes an API key can be implemented as a more flexible solution.
- The Batch Processing subgroup created a detailed MARC 21 to ISNI XML schema mapping with the intent that libraries' extensive MARC legacy data could be piped to ISNI through an efficient self-service batch process or by utilizing existing community tools such as MarcEdit or OpenRefine that might incorporate their mapping. The group's extensive knowledge of MARC led to a high-quality reference document that is being used now for ISNI technical developments being spearheaded by the British Library.
- The PCC Standing Committee on Training created an ISNI Training task group that has developed additional detailed documentation for future library ISNI adopters, such as the recent Linked Data for Production (LD4P) ISNI cohort,<sup>6</sup> who have chosen to learn ISNI practices. Refinements and new training content continue to be posted on the PCC ISNI Pilot wiki.
- Recommendations for enhancements to ISNI tools were gathered through the Pilot and will be channeled through the new ISNI Library-

Sector Steering Group mentioned above. One workflow that was exercised preliminarily by a subset of Pilot participants toward the end of the Pilot was URIs in MARC experimentation. With changes to the MARC format that were approved and implemented in mid-2018, some Pilot participants jumped on the opportunity to gain experience adding ISNI URIs in their MARC cataloging. This effort will be built upon as in December 2019 the PCC formally launched a URIs in MARC Pilot project that will incorporate ISNIs and a wide range of additional data sources into bibliographic and authority data. Over thirty institutions have volunteered to participate.

- ISNI maintenance best practices emerged as a complex issue in the Pilot. Despite enthusiasm in the group to contribute to ISNI data integrity, the practitioner interfaces for maintenance were frustrating and non-intuitive to use, so this led to a crisis of confidence that libraries would have the resources to devote to workflows that were seen as inefficient and outmoded. The hope is for improved tools in this area.

Additional analysis will need to be done within the PCC to determine the ongoing role of PCC members within ISNI and how those activities may relate to the NACO program. A project is underway between the Library of Congress and the technical team in the ISNI Leiden office to load ISNIs to the NACO file, which will increase the interoperability between the files and present greater options for tooling for NACO-ISNI workflows. As part of its updated charge, the PCC Task Group on Identity Management in NACO will be reviewing the “NACO Lite” concept that has been discussed informally in the community for several years but not yet codified in practice. A range of identifier workflows would be in scope for consideration.

As PCC leadership evaluates the ongoing role of ISNI in formal PCC programs, the PCC ISNI Pilot and discussions that flowed from it have excited considerable interest in launching a similar identity management Pilot for experimentation in Wikidata. In a survey to Pilot participants toward the end of the ISNI Pilot in 2018, over thirty Pilot participants who responded to the survey expressed interest in exploring Wikidata and bringing a comparative viewpoint from their experience working in ISNI. The PCC Task Group on Identity Management in NACO is planning to launch a PCC Wikidata Pilot in 2020.

As the Pilot was nearing completion, PoCo voted unanimously in November 2018 to renew the 2019 PCC ISNI membership in order to pursue expanded access to ISNI by additional PCC members beyond the Pilot group; the PCC LD4P2 cohort was selected for eligibility. PoCo voted to renew the annual membership again in December 2019, which

demonstrates an ongoing understanding of mutual benefit and an interest in continued collaboration between the PCC and the ISNI-IA, and yet complex issues for PoCo relate to funding. After the ISNI seed money from the Pilot is exhausted, there is no revenue stream from which PCC can allocate funds to sponsor membership renewal, nor any independent PCC financial infrastructure outside the Library of Congress. Additionally, with the PCC being a completely volunteer organization without payment of dues for membership, PCC does not have legal status to truly undertake contractual obligations on behalf of its membership. These issues have been discussed in PoCo, and there is commitment to seek solutions given their bearing on the future of the PCC ISNI partnership.

These issues notwithstanding, forged through the PCC ISNI Pilot has been a true appreciation for the value that each organization brings to the other. The PCC ISNI Pilot was a milestone in PCC history as the first initiative that enabled the PCC to explore a new path for creating internationally recognized identifiers for individuals and organizations outside of the PCC NACO program. It moved the PCC toward its goal of establishing identity management as a workflow alternative to traditional text string-based authority control and has demonstrated the value for the PCC in exploring collaborations outside of familiar library contexts.

## ORCID

Michelle Durocher  <http://orcid.org/0000-0001-5431-1873>  
 Iman Dagher  <http://orcid.org/0000-0002-1793-8578>  
 Violeta Ilik  <http://orcid.org/0000-0003-2588-3084>  
 Chris Evin Long  <http://orcid.org/0000-0003-3953-7940>  
 Isabel del Carmen Quintana  <http://orcid.org/0000-0002-8868-3365>

## Notes

1. “Program for Cooperative Cataloging: Vision, Mission, and Strategic Directions: January 2015–December 2017,” last updated November 20, 2015, <https://www.loc.gov/aba/pcc/about/PCC-Strategic-Plan-2015-2017.pdf>
2. Ibid., 4.
3. For more information about ISNI, see the forthcoming article: A. Armitage, M. J. Cuneo, I. Quintana, K. Carlson Young, “ISNI and Traditional Authority Work,” *JLIS.it* 11, no. 1 (January 2020): 150–62, <http://dx.doi.org/10.4403/jlis.it-12554>
4. “PCC ISNI Pilot Home,” <https://wiki.lyrasis.org/display/PCCISNI/PCC+ISNI+Pilot+Home> (accessed December 13, 2019).
5. “Pilot 2017-18 Joint Focus Areas & Deliverables,” <https://wiki.lyrasis.org/x/2542BQ> (accessed December 13, 2019).
6. “LD4P ISNI Cohort,” <https://wiki.lyrasis.org/x/nYDXBg> (accessed December 13, 2019).