NACO, Authority Control, and Identity Management: Evolving Strategies for a Changing Name Authorities Landscape

Presented on December 6, 2017

The Impetus for a Change from Authority Control to Identity Management ............... p. 2
presented by John J Riemer

PCC Initiatives in Identity Management ......................................................... p. 16
presented by Michelle Durocher

Part II. Authority Control in Straitened Circumstances ................................. p. 27
presented by Joseph Nicholson

Webinar hosted by ALCTS, Association for Library Collections and Technical Services
The Impetus for a Change from Authority Control to Identity Management

NACO, Authority Control, and Identity Management: Evolving Strategies for a Changing Name Authorities Landscape

John J Riemer

December 6, 2017
Environmental Factors Driving Changes

• Linked Data requires differentiation and an identifier for each entity
  • 9M LCNAF records but 30M WorldCat Identities (30% coverage)*
  • Digital library project work—Authorities a luxury?
  • Numerous researchers & organizational units to track

• Multiple files exist & we could be making use of them
  • Do I really need to replicate data from other files in the LCNAF?
  • Can’t I quote an ISNI number in the subfield $0 of an access point?

• Other files have batch matching capabilities

Environmental Factors Driving Changes (cont’d)

• In 2012, the PCC reassessed the future of undifferentiated personal name authority records

• By 2015 LC & PCC had ended their use

• LCCN Permalink: https://lccn.loc.gov/n99282107
• 010 __ |a n99282107
• 040 __ |a DLC |b eng |c DLC |d UPB |d DLC
• 100 1__ |a Jones, Amy
• 667 __ |a THIS 1XX FIELD CANNOT BE USED UNDER RDA UNTIL THIS UNDIFFERENTIATED RECORD HAS BEEN HANDLED FOLLOWING THE GUIDELINES IN DCM Z1 008/32
• 670 __ |a [Illustrator of A child’s garden of verses]
• 670 __ |a A child’s garden of verses, 1946.
• 670 __ |a [Co-author of Parliamentary practices in devolved parliaments]
• 670 __ |a Parliamentary practices in devolved parliaments, [1998]: |b t.p. (Amy Jones, research asst.)
• 670 __ |a [Co-author of What to say when-- for Latter-day Saints]
• 670 __ |a What to say when-- for Latter-day Saints, c2007: |b t.p. (Amy Jones)
Environmental Factors Driving Changes (cont’d)

• Does the differentiating information all need to go in one field?
Environmental Factors Driving Changes (cont’d)

<table>
<thead>
<tr>
<th>#</th>
<th>Bib Records</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Adams, John, 1930-1999</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Adams, John, 1933-</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Adams, John, 1935-</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Adams, John, 1936-</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>Adams, John, 1938-</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Adams, John, 1938 Aug. 13-</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>Adams, John, 1938 August 13-</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>Adams, John, 1939-</td>
</tr>
</tbody>
</table>

Photography, Violist, Economics, Geography, Law
Environmental Factors Driving Changes (cont’d)

• A need to reconsider membership criteria within NACO
  • 9K OCLC members with cataloging authorizations – 700 NACO members
  • What are the essential skills needed: heading construction vs. differentiation?
  • Who will create identifiers for your institution’s researchers, if not you?
Environmental Factors Driving Changes (cont’d)

• Potential new uses of library authority data
  • Harvesting citations of scholarly output
  • Accreditation efforts for an academic department, school, or university
Program for Cooperative Cataloging response

In the PCC Vision, Mission and Strategic Directions document, 2015-2017, Strategic Direction 3 is:

Provide leadership for the shift in authority control from an approach primarily based on creating text strings to one focused on managing identities and entities.

3.1 Establish a task group to develop best practices for the use of authority data sources beyond the LC Name Authority File.

3.2 Partner with other agencies, project partners, and communities ... to explore models for and to build the infrastructures necessary to host the cooperative management of identities.

3.4 Investigate options and develop a plan to expand community participation in the creation of identifiers and authority data.
Definitions

**Authority control**

manages access to entities by authorizing a specific form of name or other term for access use.
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 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.
What difference it makes

Authority Control--Reluctance to establish the name of an ETD author, since it is unlikely to be the form of name seen in books or articles written subsequently.

Identity Management--Strong desire to get an identifier registered at the outset of one’s career, to facilitate accurate, efficient gathering of researcher output.
What difference it makes

Authority Control--Reluctance to supply death date information, for fear of imposing “Bibliographic File Maintenance” burden on all the institutions with copies of records.

Identity Management--Incentive to record, or point to, any additional information of interest or value to researchers.
What difference it makes

Authority Control--Expectation that other parties (LC, PCC members) will establish the names associated with published material; a willingness to wait.

Identity Management--Realization that no one else is likely to create identifiers for researchers and organizational units tied to your institution; *you* will need to do it.
What difference it makes

Authority Control--Mostly confined to libraries, vendors supporting them, some museums & archives.

Identity Management--Expanded identifier “supply chain” includes non-library partners such as publishers, rights management agencies.
PCC Initiatives in Identity Management

Michelle Durocher

NACO, Authority Control, and Identity Management: Evolving Strategies for a Changing Name Authorities Landscape

December 6, 2017
Webinar hosted by ALCTS, Association for Library Collections and Technical Services
Identity Management... where do we begin?
Identifiers for use in libraries need to be...

- Stable
- Persistent
- Discoverable
- Reusable / shareable
- High (enough) quality

... and ideally have a way for engagement by metadata practitioners
Why ISNI?

• ISNI = International Standard Name Identifier
• ISO global standard in the family of identifiers like ISBN and ISSN
• Persistent unique identifying number for persons and organizations (researchers, inventors, writers, artists, performers, publishers, etc.)
  – i.e. the full range of needs for all kinds of libraries
• Cross-domain standard used by libraries, publishers, databases, rights management organizations, etc.
# PCC ISNI Pilot participants

<table>
<thead>
<tr>
<th>University of Colorado-Boulder</th>
<th>Columbia University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell University</td>
<td>Michigan State University</td>
</tr>
<tr>
<td>Stanford University</td>
<td>Stony Brook University</td>
</tr>
<tr>
<td>Texas A&amp;M</td>
<td>UC Davis</td>
</tr>
<tr>
<td>UCLA</td>
<td>University of Wisconsin-Madison</td>
</tr>
</tbody>
</table>

*with Harvard Library contributing coordination with ISNI and OCLC*
PCC ISNI pilot began in late July 2017

Pilot libraries charged to:

• work within the ISNI systems, examining through use the functionality and gaps of existing tools
• establish batch workflow processes that lower overhead for all and can be adopted as common approaches within the PCC umbrella membership once in broad production
• document pain points from the pilot to inform ISNI technical & policy roadmap
• identify PCC training, documentation & policy needs
PCC ISNI Pilot Home

Created by Bill Branan, last modified by Michelle Durocher on Aug 25, 2017

Scope: Overview of the PCC ISNI Pilot

Creator: Michelle Durocher

Goals and Objectives
Pilot Participation
Commitments
Workflow Activities
Sample Use Cases and Workflows
Background on the PCC

Goals and Objectives

In initiating an ISNI Registration Agency Membership agreement, th
“NACO Lite” & Opening Up the PCC

• What are essential data elements for “authority data” in an identity management context?...
  – How does that differ when considering the needs of humans and the needs of machine matching?

• How do we achieve greater coverage of entities receiving identifiers?
Possible Venues for Identity Management

- LCNAF
- ISNI (PCC pilot project)
- Wikidata

- Text string uniqueness requirements
- Preferred labels & Multilingual structure
- Batch searching capabilities
- Ability to serve as a reconcilable data source (OpenRefine & endpoints)
Issues Arising in Identity Management Work

• How rich does data need to be for an entity... initially?
• Is it necessary to designate data as preliminary/provisional in? (How?)
• Does embryonic authority data that might begin as Identity Management work need to assert RDA compliance?
• How best to reflect existence of the same entity in multiple files?
  • Full incorporation of one file’s data into the other
  • Merely pointing to it from one file (e.g. MARC 024 fields in LCNAF)
  • ‘Same As’ assertions, aggregated
• What will support searching across multiple files simultaneously?
Your feedback is very important!

• The PCC Identity Management in NACO Task Group would like your input to help us prioritize our next steps on NACO Lite:

• Please fill out our very brief survey:
  
  https://goo.gl/forms/WrlyDLkgCllLI753
NACO, Authority Control, and Identity Management
Evolving Strategies for a Changing Name Authorities Landscape

Part II
Authority Control in Straitened Circumstances

Joseph Nicholson
J. Murrey Atkins Library
UNC Charlotte
- 700 organizations participating

- Many NACO members are larger libraries with robust staffing

- Smaller libraries may find it difficult to contribute

- Slow, rigorous process for creating records discourages participation?

Yearly Contribution Requirements

ARL libraries: 200 records

Non-ARL libraries: 100 records
Setting the Scene

J. Murrey Atkins Library
UNC Charlotte

- Large institution
  - 2,000,000 volumes
  - 84 faculty and staff
  - 28,000 students

- Reorganization of technical services that depleted cataloging staff

- Migration to new ILS and discovery system with limited authority control functionalities

- Need for authority control in both digital projects and in catalog

- NACO member since early 2014

- Severe shortage of trained NACO personnel
NACO Program Participants at UNC Charlotte

1 paraprofessional with some catalog maintenance experience

1 undergraduate student worker

1 Atkins Fellow intern

1 librarian (myself)

3-4 digital programs, oral histories staff and several graduate student interns
**Local Authority Control Efforts at UNCC**

- Digital projects in CONTENTdm and Islandora repository
- Access points for names created in spreadsheets for digital projects by student assistants and staff
- Much of repository content harvested for MARC records in catalog
- Could labor on local authority work be parlayed into NACO records?
MYTHS?

- Only professional cataloging librarians can create authority records.

- Authority work can only be done in a bibliographic utility like OCLC Connexion.

- Authority work is solitary, non-collaborative work.

- Everyone participating in NACO workflows should be at same basic skill level.
Rethinking NACO Workflows

- Streams of authority data rather than finished records—some of it quite messy; some of it in MARC format, some not

- Participants with very different skill levels

- Collaborative model of producing records rather than one cataloger-one record approach
ETD Names
Paraprofessional, undergraduate student assistant

Names from books workflow
Intern, staff, me

Names from manuscripts
Staff

Digital project names
Undergraduate student assistant

Oral history names
Oral history project staff
“Just Enough” Training Philosophy

- PowerPoint slides with pointers for each MARC field and examples
- Written documentation on choosing most common form of name, etc.
- Some instructions incorporated into tools used for gathering data
- Lots of reviewing and critiquing
### OCLC Connexion Constant Data

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>0x00</td>
</tr>
<tr>
<td>Name</td>
<td>Library Science</td>
</tr>
</tbody>
</table>

**NACO Records Produced**

**Spreadsheets with Field of Activity Data**

<table>
<thead>
<tr>
<th>Degree Name</th>
<th>LCSH for 372</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthropology</td>
<td>Anthropology $2 lsh</td>
</tr>
<tr>
<td>Applied ethics and philosophy</td>
<td>Applied ethics &amp; Philosophy $2 lsh</td>
</tr>
<tr>
<td>Applied mathematics</td>
<td>Applied mathematics $2 lsh</td>
</tr>
<tr>
<td>Architecture</td>
<td>Architecture $2 lsh</td>
</tr>
<tr>
<td>Bioinformatics and computational biology</td>
<td>Bioinformatics &amp; Computational biology $2 lsh</td>
</tr>
<tr>
<td>Biology</td>
<td>Biology $2 lsh</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Chemistry $2 lsh</td>
</tr>
<tr>
<td>Civil and environmental engineering</td>
<td>Civil engineering &amp; Environmental engineering $2 lsh</td>
</tr>
<tr>
<td>Civil engineering</td>
<td>Civil engineering $2 lsh</td>
</tr>
<tr>
<td>Clinical and community psychology</td>
<td>Clinical psychology &amp; Community psychology $2 lsh</td>
</tr>
<tr>
<td>Clinical exercise physiology</td>
<td>Clinical exercise physiology $2 lsh</td>
</tr>
<tr>
<td>Communication studies</td>
<td>Communication $2 lsh</td>
</tr>
<tr>
<td>Computer science</td>
<td>Computer science $2 lsh</td>
</tr>
<tr>
<td>Construction and facilities management</td>
<td>Facility management &amp; Building-Planning $2 lsh</td>
</tr>
</tbody>
</table>

**Simple Tools**

*Example of simple tools used in the process.*
- Spreadsheets
- OpenRefine
- MARCXML Templates
- MARCXML Template
- MARCEdit

• XSLT

More Complex Tools
1. Staff fill out form for new name encountered in oral history interviews

Instructions can be entered in form

Fields: 046, 100, 370, 372, 373, 374, 375, 377, 400, 670
NACO Records Via Form Data Procedure

2. Data from form is automatically transferred to a Google sheet

<table>
<thead>
<tr>
<th>Birth And/Or Death Dates</th>
<th>Place of Birth</th>
<th>Place of Death</th>
<th>Country of Birth</th>
<th>Place of Residence 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1928-1962</td>
<td>China Grove (N.C.)</td>
<td>Belmont (Gaston County)</td>
<td>United States</td>
<td>Belmont (Gaston County)</td>
</tr>
<tr>
<td>1920-1944</td>
<td>Saint Louis (Mo.)</td>
<td>Charlotte (N.C.)</td>
<td>United States</td>
<td>Saint Louis (Mo.)</td>
</tr>
<tr>
<td>1916-1944</td>
<td>Augusta (Ga.)</td>
<td>Charlotte (N.C.)</td>
<td>United States</td>
<td>Charlotte (N.C.)</td>
</tr>
<tr>
<td>1927</td>
<td>Charlotte (N.C.)</td>
<td></td>
<td>United States</td>
<td>Charlotte (N.C.)</td>
</tr>
<tr>
<td>1923-1945</td>
<td>Union County (N.C.)</td>
<td>Union County (N.C.)</td>
<td>United States</td>
<td>Union County (N.C.)</td>
</tr>
<tr>
<td>1927-1945</td>
<td>Mecklenburg County (N.C)</td>
<td>Charlotte (N.C.)</td>
<td>United States</td>
<td>Charlotte (N.C.)</td>
</tr>
<tr>
<td>1927</td>
<td>Charlotte (N.C.)</td>
<td></td>
<td>United States</td>
<td>Charlotte (N.C.)</td>
</tr>
<tr>
<td>1946</td>
<td>Charlotte (N.C.)</td>
<td></td>
<td>United States</td>
<td>Charlotte (N.C.)</td>
</tr>
<tr>
<td>1900-1953</td>
<td>Cumberland County (N.C)</td>
<td>Charlotte (N.C.)</td>
<td>United States</td>
<td>Charlotte (N.C.)</td>
</tr>
</tbody>
</table>

3. Spreadsheet is imported into OpenRefine
4. Prior to this step, MARCXML template created in Notepad
NACO Records Via Form Data Procedure

2. Spreadsheet imported into OpenRefine for cleaning, transformed into MARCXML, exported as XML file

<table>
<thead>
<tr>
<th>Birth And Or Death Dates</th>
<th>Place Of Birth</th>
<th>Place Of Death</th>
<th>Country Of Birth</th>
<th>Place Of Residence 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1918-2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

XSLT strips out empty elements
NACO Records Via Form Data Procedure

5. File transformed into MARC in MARCEdit

6. File of records is imported into OCLC
7. Records edited and revised—and finally submitted
Local Authority Data → NACO Records

Unruly local authorities Google Docs spreadsheet for digital project

Spreadsheet downloaded, cleaned in OpenRefine, outfitted with dropdowns

OpenRefine transformation yields MARC records
Intensive review and reworking of records

Many corrections

Steep learning curve for participants

Very different skill levels = different training approaches required

Setting up process is labor-intensive

Too many personal name records

Infelicitous Aspects of the UNORTHODOX Tactics
Successes?

- 1000+ records submitted so far
- 80-90 using recent form approach
- 130 from digital projects in pipeline

Aspirations

- Training several of the oral history staff to do NACO work
- More automation of the process
- Find time to devote to more records of my own
- Hiring another staff member!

Results and Future Approaches
Other Recommendations for Easing Burden of NACO Membership

1. Join a NACO funnel project

https://www.loc.gov/aba/pcc/naco/nacofunnel.html
Other Recommendations Continued

2. Use Authority Toolkit – courtesy of Gary Strawn

What the toolkit will help you do:

- create a new authority record from an access field in a bibliographic record, and then help you enhance the preliminary authority record.
- enhance an existing authority record
- extract one element from an undifferentiated personal name authority record (this task combines the modification of an existing authority record with the creation of a new authority record), and then help you enhance the preliminary authority record for the extracted identity

http://files.library.northwestern.edu/public/oclc/documentation/
2. Selectively Create NACO “Lite” or Lighter Records

Not Core: Personal Names

* Gender
* Place of Birth
* Place of Death
* Country Associated with Person
* Place of Residence
* Affiliation
* Language
* Field of Activity
• Profession or Occupation (unless needed to distinguish)

• Several RDA elements not “core”—others core only if needed to distinguish
• Many names do not merit exhaustively detailed records
• Possible approach: apply full battery of fields only in special cases?
New Model of Authority Control Needed?

- If identifiers will be the differentiating mechanism for identities, how much labor should be spent on creating unique text strings?
- Slowness and complexity of authority record creation process = many names remain uncontrolled
- Method more aligned with linked data principles and will synchronize with work of other identity communities—e.g. ISNI—needed?
Questions?
Image Credits 1

Slide 2


Slide 3


Slide 4


Slide 6

Slide 7
Rabich, Dietmar. River Tiber in Rome (Italy) at night. 2013. Photograph. Wikimedia Commons.

Slide 8


Mosaic seascape Musei Capitolini. 1888. Photograph. Wikimedia Commons.

Slide 9

Slide 10

Slide 11

Slide 19
Image Credits 4

Slide 20


Slide 21


Slide 24