**ISNI and traditional authority work**

Amy Armitage\(^{(a)}\), Mary Jane Cuneo\(^{(b)}\), Isabel Quintana\(^{(c)}\), Karen Carlson Young\(^{(d)}\)

\(^{(a)}\)\(^{(b)}\)\(^{(c)}\)\(^{(d)}\) Harvard University

**Contact:** Amy Armitage, amy_armitage@harvard.edu; Mary Jane Cuneo, cuneo@fas.harvard.edu; Isabel Quintana, quintana@fas.harvard.edu; Karen Carlson Young, karen_young@gse.harvard.edu

**Received:** 1 March 2019; **Accepted:** 21 July 2019; **First Published:** 15 January 2020

**ABSTRACT**

This article describes key differences between ISNI (International Standard Name Identifier) and the authority work traditionally performed at libraries. Authority work is concerned with establishing a unique form of name for a person and collocating materials under that form of name. ISNI, on the other hand, is concerned with establishing a unique numerical identifier for each entity, and differentiating distinct entities. The focus of the work becomes identity management rather than the establishment of authorized name forms. This article looks not only at the differences in workflows, but also explains how these theoretical differences can affect the way librarians identify and collocate named entities. The focus is on the future, and how we can best use our skills to ensure that entities are properly differentiated and accessible to our patrons.

**KEYWORDS**

Identity management; ISNI; Authority work.

**CITATION**

This article presents the key differences between traditional library authority work and identity management. In this article, traditional library authority work is exemplified by the Name Authority Cooperative Program (NACO) of the Program for Cooperative Cataloging (PCC). This program is based primarily in the United States, Canada, and the United Kingdom. The program cooperatively adheres to the same cataloguing standards in order to freely share records for both bibliographic and authority entities (“About NACO” 2018). The International Standard Name Identifier (ISNI) is presented as an example of identity management work. ISNI is an international file based primarily in Europe. The comparison reviews differences in provenance, veracity of metadata, corporate body name changes, and duplication. Our focus is on how traditional library authority work and identity management approach disambiguation of entities differently. These differences create different types of databases and perhaps would be most successfully used in different workflows. This article does not address how these identifiers, once constructed, are used by the international community. The article ends with thoughts on how librarians might incorporate identity management strategies as part of their library work.

Since its creation in 2012, the ISNI database has been the subject of various articles published throughout the world. The literature has focused on the nature of ISNI – why and how it was established (Angjeli, 2012, 2014; Gatenby and MacEwan, 2011); and on its use in various settings (Balan, 2017; ISNI Assignments Top 6.5 Million, 2013). However, the process of working in ISNI has not been compared with the process of working with traditional library authority data. This article attempts to fill this gap, by comparing both the theoretical and practical aspects of working in these different environments.

**What is ISNI?**

Let’s face it, finding the right identity on the web can be very difficult at times. These slides, from the ISNI.org website, illustrate the problem. They depict three different people who all have the same name: Michele Smith. Two of them work in the field of music. It might be easy to conflate these identities. The third Michele Smith is an author.
Screen shot for Michele Smith, the singer (ISNI, 2018)

Screen shot for Michele Smith, the musician (ISNI, 2018)
The same problem occurs in our library catalogues. Establishing the correct entity can be a time-consuming and detailed process. Once the correct identity is found, we want to be able to find other materials related to this person. Traditionally, libraries have used a unique form of the person’s name to collocate materials by that person. ISNI takes a different approach, assigning instead a unique number to identify a person. This number can be used in linked data as a bridge identifier to connect this person to other resources by and about the same person, without the need to regularize the form of the name (ISNI website 2018, homepage).

ISNI, or “International Standard Name Identifier,” is an ISO standard, and, as such, works much like the ISBN does to identify books. It is a persistent, unique identifier. ISNIs are primarily assigned for persons, but corporate bodies are included as well. ISNI is supported by the OCLC offices in Leiden, the Netherlands, and has two Quality Teams working at the British Library and the Bibliothèque nationale de France (“International Standard Name Identifier” Wikipedia 2018).

**What are the differences between ISNI and NACO?**

The fundamental difference between ISNI and NACO is that NACO provides authority control, where a unique form of name is created to represent each entity, ISNI, instead, serves identity management, where a unique identifier is assigned to each entity. Both programs attempt to distinguish entities, collecting information about one entity versus another one with a similar name. However, in authority files, the focus is on the name string. NACO training primarily teaches the cataloguing rules for how to formulate a unique name string. In contrast, ISNI focuses on the identity
itself. ISNI practitioners put most of their effort into determining whether one identity is the same as another, or different. In other words, most of an ISNI practitioner’s time is devoted to research. This key difference manifests itself even in how the databases are searched to identify the correct entity. For example, here are the results of a search for John the Baptist in ISNI:

1. Battista, Giuseppe (1610-1675)
2. Johannes Baptist., (currori on:29)
   La figure de Jean-Baptiste dans les lettres françaises [atti del "Seminario Balmes"], Gargnano del Garda, 8-11 giugno 2011
3. Johannes Baptist., (sanctus)
   Sveti Ivan Krstitelj ...
4. John the Baptist, (Musical group)
   John the Baptist, musical group, supermarket graveyard show
5. John the Baptist, [dialogue] / Hywel at the Moon
6. 1 Master of Saint John the Baptist, (German artist, active ca. 1467)
7. 1 Master of the Beheading of Saint John, (German painter, active 1495-1518)
8. 2 Master of the Life of Saint John the Baptist, (Italian painter, active ca. 1325-1350)
   The Feast of Herod and the Beheading of the Baptist
9. 3 Yohane Osuboni, (der Täuber v1-29)
   Mandaean Book of John

To a NACO practitioner, the results can be confusing. The names appear in various languages, and they include artists and musicians. Is the correct entity no. 3, 5, or 9? Clicking on no. 9 brings up this display:

| ISNI: | 0000 0000 8705 0000 |
| Name: | Johannes (der Täuber) Yohane Osuboni |
| Date variant: | v1.29 |
| Location / Nationality: | Israel Jordan |
| Titles: | Mandaean Book of John ابراهيم نورتي |
| Notes: | |
| Sources: | VIAF DNB SWNL WKD |

ISNI record for the identity “Yohane Osuboni” (der Täuber v1-29) in the web client list (ISNI, 2018)

Expanding the “name variant” box produces a long list in many languages, including “John the Baptist (baptizer of Jesus).” So, the name does appear as expected – if one speaks German. This suggests several things: (1) there is no preferred form of name; (2) the ISNI community is broader
than any single library authority file, or all librarians who speak English; and (3) ISNI emphasizes machine actionability. Human eye-readable elements are present, but they are not primary, as an ISNI is designed for automation and for users and contributors worldwide.

The same search in the NACO file yields this result:

![NACO browse search result for "John the Baptist" in OCLC Connexion (OCLC, 2018)](image)

When the NACO record for “John the Baptist” is expanded, this is the result:

![Section of the NACO authority record for "John the Baptist" in OCLC Connexion (OCLC, 2018)](image)

Note the hybrid forms in many of the variant fields (400s). For example, the yellow highlighted field has a German term, “der Täufer,” followed by “Saint” in English. The NACO rules about qualifiers produce this effect. NACO is concerned with the form of the name in the string, regardless of whether the string is the authorized form (MARC 100), or a variant form (MARC 400) (“NACO Training” 2018, Module 1, slide 22).

As suggested above, the NACO and ISNI communities differ in scope. NACO member institutions are libraries, primarily in the US and Great Britain (Frequently asked Questions about Joining the NACO Program 2019, 1). Because the names they contribute come from library resources, they represent primarily authors and their subjects. ISNI member institutions include libraries, too, but also other kinds of organizations in the information supply chain, from all over the world (“How ISNI works” 2019). They may contribute names for the creators of non-literary works, or works that are
not published – creating records that cite, for example: buildings, garden plans, or photograph albums. Furthermore, some ISNI records have no citations at all for related works, as we will see later on. Non-library member organizations bring interests and goals to ISNI that may differ from ours, but they are no less legitimate. ISNI opens the door to a more diverse catalogue of identities, a wider member cohort with points of view to consider, and new audiences with whom to share what we have to offer.

There are also key differences in how metadata is recorded between ISNI and NACO. This is partly due to input workflows. The LC/NACO authority file has been built one record at a time, with care and attention to the PCC cataloguing rules. Cataloguers must check carefully before adding to the file to ensure the record is not a duplicate (“NACO Training” 2018, Module 1, slides 74-76). In contrast, most ISNI records have been batch loaded. ISNI ingests the data first and then improves it. A sophisticated algorithm assigns ISNIs, or not, based on programmed confidence levels, and highlights possible duplicates. A provisional record is one that has been added as a proposal and has not yet been assigned an ISNI identifier, pending further investigation by an ISNI participant. Humans intervene at the end of the process to resolve problems. Humans can also enter new records manually into ISNI (“Data Quality Policy” 2019).

As noted earlier, the NACO and ISNI programs differ significantly in how their contributors are onboarded. Participation in NACO is allowed only after a cataloguer has completed a detailed 5-day training course, and has been reviewed, which usually entails the contribution of more than 100 records (“NACO Training” 2018, Module 8, slides 8-4, 11). The training required for ISNI participation is much shorter, as is the review period, which evaluates the source institution rather than the individual contributor (“PCC ISNI Pilot Home” 2019). An ISNI contributor does not have to be a cataloguer. This is because the form of a name does not depend upon library cataloguing conventions; indeed, the form of the name does not matter. What matters is entering the data, and correctly identifying the entity (Liss 2017, 5, 7).

There is another key difference between ISNI and NACO. Provenance, which means the legitimacy of data based on where it comes from, is important in both NACO and ISNI, but the criteria are different. NACO relies upon “usage” (“Frequently Asked Questions on creating Personal Name Authority Records (NARs) for NACO” 2018): How does the entity represent itself in resources, and where did the name appear? This information is cited, to support or prove the assertions made in an authority record (“Descriptive Cataloging Manual” 2018, section Z1, 670).

To give an example, the authorized heading for Sandro Botticelli is “Botticelli, Sandro, $d 1444 or 1445-1510.” In the authority record for this artist there are statements to support the preferred form of his name. One citation notes that in a book titled ‘Botticelli, painter of Florence’, by H.P. Horne, published in 1980, page 1, states that the artist is commonly called Sandro Botticelli.

In ISNI there is no need to prove the veracity of the metadata. A contributing ISNI institution is called a “source.” ISNI trusts its sources, provided the ISNI algorithm finds no problem when comparing new data with existing data. If more than one source says the same thing, this is even
better. ISNI is interested in which institution added the metadata, and how many institutions corroborate the same metadata (“ISNI Manual” 2019, section 1.5, 5–8).

In the example above, the NACO record for Umberto Eco indicates that he was born on January 5, 1932 and died on February 19, 2016. This information is coded in the 046 field, and the birth date is supported by the last 670 visible in the illustration. The 040 field lists codes for all the libraries that have contributed to this authority record. It readily shows which institutions contributed the metadata, but not which institution contributed what piece of metadata. NACO bases metadata on “literary warrant” (RDA Toolkit 2017, 0.4.3.4), i.e., where the information appeared formally, so each 670 lists a title, year of publication, and sometimes the author (“NACO Training” 2018, Module 1, slides 208-209).

In contrast, ISNI links each piece of metadata with the source institution’s symbol (“Data Contributors” 2019).
The fourth line of this example states that the Bibliothèque nationale de France cites Umberto Eco’s birthday as January 5, 1932. Other institutions list both his birth and death date. The information is repeated, even if it is identical.

NACO and ISNI also differ in how they handle name changes for corporate bodies. NACO requires the creation of a new authority record when a corporate body changes its name; the two authority records are then linked in an earlier/later relationship (RDA Toolkit 2017, 11.2.2.6, 32.1.1.3). In ISNI, a corporate body continues to have the same identifier unless there is a change in the structure of the corporate body, a merge or a split (“ISNI Manual” 2019, section 7, 31). In other words, to warrant a new identifier, the nature of the corporate body must change. Absent that, a name change is just a name variant in ISNI. However, ISNI batch loads data from sources that use mutually incompatible models for handling corporate name changes – mostly, European sources versus US sources. ISNI may then automatically merge records, creating some strange combinations. Recognizing that the result can be problematic, the ISNI Quality Team works to correct these records manually (“ISNI Manual” 2019 section 7, 22–30).

Despite the differences between ISNI and traditional authority work, there are two issues that are key to any good database of entities: (1) the need to differentiate between entities; (2) the need to avoid duplicate records for the same entity. To meet these criteria, identifying information must be provided for each entity.

For example, let’s look at this list of names:

- Quintana, Isabel del Carmen
- Quintana, Isabel
- Quintana, Isabel
- Quintana y Gonzalez, Isabel
The list tells us nothing about who this person, or these people, are, and therefore neither humans nor a computer can know if one, two, or several entities are represented. In NACO, one or more citations provide context by connecting the name with a work (works). It could be a book written by or about this person (“Descriptive Cataloging Manual” 2018, section Z1, 670). For example, in the screen shot below the NACO authority record ascertains that Isabel Quintana is the author of a book titled, Figuras de la experiencia en el fin de siglo, published in 2001.

In contrast, data batch loaded into ISNI may be “sparse” – for example, only a name may be present. In these cases, the computer cannot match the entity with any other entities in the database, and duplicates result. If the sparse records remain provisional, they do not display to the public and so any redundancies may not be of concern. However, if they are upgraded to “assigned” status, they need to be resolved (“Data Quality Policy” 2019).

Duplicates are a major problem for any authority file. In NACO, great care is taken to ensure that duplicates are not created in the first place (“NACO Training” 2018, Module 1, slides 75, 82-83, 94-98).

Yet we find them. This is partly because it is labor-intensive to search every known variant of the name, especially if it is a common name. When cataloguing a book by William Smith, one must look at many records in the current file, including variants such as Will Smith or Bill Smith. The person might also have been established with additional information not present in the book in hand, such as “Smith, William A.” It can be tricky and time-consuming to avoid creating a duplicate.

ISNI resolves as many duplicates as possible automatically and flags others for manual resolution. Because it’s much easier to merge duplicate entities than to separate conflated ones, the matching algorithm is extremely cautious (“Data Quality Policy” 2019), with the result that duplicates are common. How easy is it to merge entities in ISNI? A search for “Michael, George” retrieves a complete record for the British singer who passed away in 2016. It also finds a very brief provisional record. The latter looks like this:

<table>
<thead>
<tr>
<th>010</th>
<th>no2003009259</th>
</tr>
</thead>
<tbody>
<tr>
<td>040</td>
<td>NJP ≠b eng ≠ c NJP</td>
</tr>
<tr>
<td>100</td>
<td>Quintana, Isabel</td>
</tr>
<tr>
<td>670</td>
<td>Figuras de la experiencia en el fin de siglo, 2001: ≠b t.p. (Isabel Quintana)</td>
</tr>
</tbody>
</table>
Section of the provisional ISNI record for “Michael, George” (ISNI, 2018)

A possible match is indicated, and the contributor is asked to compare this record with the fuller record for George Michael. The two records will then be displayed side by side, and one can respond: “Equal,” “Unequal,” or “Don’t know” (“ISNI Web Interface Usage Guidelines” 2019, section 4, 19-20).

Since ISNI logins are at the institution level (“Data Contributors” 2019), the participant’s assertion will be recorded as a statement made by that institution.

So, although one will find more duplicates in ISNI than in NACO, there are efficient ways to deal with the duplicates – once one has determined that they are duplicates.

This case is a typical example of the kind of inquiry sometimes needed to resolve duplicates. There is so little information in the provisional record that it is difficult to know whether these entities are the same. A search for IPDA in ISNI’s list of sources (Ibid) reveals that it stands for the International Performers Database Association. A web search shows that the main objective of the International Performers Database (IPD) is to identify individual performers in audio recordings and audio-visual works (“IPD” 2019). So, these are likely the same person.

In summary, we in the library community have traditionally practiced authority work rather than identity management. However, many of the tasks are similar. We want to have different records set up for unique entities. As librarians it makes sense for us to experiment with ISNI and see where identity management and ISNI workflows could enhance or change our current approach to names. ISNI also provides an opportunity for Anglo-American cataloguers to share data with our colleagues globally, and with representatives of institutions outside of the library sphere who are interested in identities too. Because the identifiers are machine actionable, ISNI is a step toward linked data and discovery of library resources on the web.

**What are some possible next steps?**

The PCC created a Pilot program (“PCC ISNI Pilot Home” 2019) to determine how librarians can better incorporate ISNI into our workflows.

We might experiment with ISNI for local authority files. Many libraries have thousands of legacy local authorities that they cannot add to NACO because they lack the resources to enter the records manually. The PCC created a task group to work on mapping MARC library data to ISNI to facilitate future batch processing (“Pilot Joint Focus Areas & Deliverables” 2019). Perhaps these local authorities can be sent to ISNI, matched against the database, and staff can process ISNI reports that result from batch processing.
We might consider dissertations. Many university libraries have lists of people who have written dissertations; information that in some cases is stored electronically. However, NACO contributors are reluctant to create a personal name authority based on a thesis, since in many cases the form of name as it appears there is not necessarily the form preferred by the author. Creating these names in ISNI allows flexibility in the form of the name, and provides the opportunity to establish the scholar’s affiliation with a university early in his/her career.

What about other names? Perhaps we have a collection of objects, or realia, with associated names which would not fit nicely into traditional authority files. What about local university clubs, such as student organizations? Many of these corporate bodies never issue a document, yet we have information that they exist and about their focus.

In support of the PCC ISNI Pilot, the PCC has appointed an ISNI Training Task Group, charged with developing an ISNI training curriculum for PCC ISNI participants, documenting procedures and workflows for creating and maintaining ISNI records, and using ISNI tools. The task group will work closely with the ISNI Pilot participants to make the training curriculum available to others also interested in participating in ISNI (“ISNI Documentation & Training” 2019).

In short, ISNI presents opportunities to look at named entities differently. It’s critical for librarians to experiment, to determine how best to represent entities using the various data structures available; to explore how we might broaden the community within which we cooperate to do this work; and to meet contemporary information seekers wherever they are, with high-quality metadata for these entities.

References


