



Cataloging & Classification Quarterly

ISSN: 0163-9374 (Print) 1544-4554 (Online) Journal homepage: <http://www.tandfonline.com/loi/wccq20>

How Should Catalogers Provide Authority Control for Journal Article Authors? Name Identifiers in the Linked Data World

Jane Sandberg & Qiang Jin

To cite this article: Jane Sandberg & Qiang Jin (2016): How Should Catalogers Provide Authority Control for Journal Article Authors? Name Identifiers in the Linked Data World, *Cataloging & Classification Quarterly*, DOI: [10.1080/01639374.2016.1238429](https://doi.org/10.1080/01639374.2016.1238429)

To link to this article: <http://dx.doi.org/10.1080/01639374.2016.1238429>



Published online: 21 Oct 2016.



Submit your article to this journal [↗](#)



Article views: 28



View related articles [↗](#)



View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at
<http://www.tandfonline.com/action/journalInformation?journalCode=wccq20>

How Should Catalogers Provide Authority Control for Journal Article Authors? Name Identifiers in the Linked Data World

Jane Sandberg ^a and Qiang Jin ^b

^aLibrary, Linn-Benton Community College, Albany, Oregon, USA; ^bContent Access Management, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

ABSTRACT

This article suggests that catalogers can provide authority control for authors of journal articles by linking to external international authority databases. It explores the representation of article authors from three disciplines in four databases: International Standard Name Identifier (ISNI), Open Researcher and Contributor ID (ORCID), Scopus, and Virtual International Authority File (VIAF). VIAF and Scopus are particularly promising databases for journal author names, but we believe that a combination of several name databases holds more promise than relying on a single database. We provide examples of RDF links between bibliographic description and author identifiers, including a partial BIBFRAME 2.0 description

ARTICLE HISTORY

Received August 2016
Revised September 2016
Accepted September 2016

KEYWORDS

Authority control; name authority data; linked data; discovery layers; BIBFRAME; Virtual International Authority File; vocabulary control

Introduction

Traditional authority databases, such as the Library of Congress Name Authority File (LC/NAF), focus on providing authorized name access points for monograph authors in library bibliographic records, rather than journal article authors. This means that users are unable to find all articles by a specific author through the vast proliferation of online journal articles using library tools such as discovery layers. As we move into the linked data environment with several reliable international author identifier databases, we need to start thinking about how catalogers should provide name access points for journal article authors. Our recommendations are informed by a review of relevant literature and a study of how researchers published in three different journals from three different disciplines are represented in major name authority databases.

Background

Authority control is the process of selecting one form of a name and recording it, its alternatives, and the data sources used in the process. It is an important tool

CONTACT Jane Sandberg  sandbej@linnbenton.edu  Linn-Benton Community College, Library, 6500 Pacific Boulevard SW, Albany, OR 97321, USA.

Published with license by Taylor & Francis © Jane Sandberg and Qiang Jin

that boosts recall and precision in the retrieval of information resources. It provides consistency in the form of access points used to identify persons, families, corporate bodies, and subject headings. Without authority control, users can be lost when searching for a particular author with many different forms of a name, or a particular author with a very common name.

Catalogers have been creating name authority records for decades, resulting in huge databases with millions of name authority records, such as LC/NAF. An example may be useful here. The American statesman *Alexander Hamilton* wrote under several pseudonyms, such as *Philo Camillus*, and his name has different forms depending on the language being used, such as the romanized Chinese *Han-mi-erh-teng*, *Ya-li-shan-ta*. Hamilton's authority record in the LC/NAF includes his pseudonyms and variant forms of his names, while also disambiguating him from other authors with the same name by recording his year and place of birth, year of death, occupation, and field of activity.

Discussions of name authority control have historically centered around how catalogers establish the authorized form and variant forms of a person's name following different cataloging rules such as the *Anglo-American Cataloguing Rules* (AACR2) and *Resource Description & Access* (RDA), rather than how they should link bibliographic descriptions to unique author identifiers. Throughout scholarly and professional conversations, there has been an assumption that authority data should work in the background and be relatively invisible to the user. Cutter referenced a "cataloger's author list" which saved the time of the cataloger, rather than the patron. He suggested that entries in this list include "the form of name 'in full' which has been adopted, with a note of the authorities consulted and of their variations."¹ The 1961 Paris Principles reference the need for author disambiguation in their request that catalogers add "a further identifying characteristic" to author headings "if necessary to distinguish the author from others of the same name," but did not take up the question of what roles authority records should play and how they should do so.² When AACR2 was published in 1978 the cataloging community found an entire chapter on how to create a name heading, but silence on any other issues related to name authority control.³

The 2008 Functional Requirements for Authority Data (FRAD), an entity-relationship model developed by the International Federation of Library Associations and Institutions (IFLA), focuses on data regardless of how they may be packaged (e.g., in authority records). FRAD—greatly influenced by concepts from relational database design—frames authority work in terms of entities and relationships: between people and their names; people and their works, manifestations, expressions, and items; and between authority records and other authority records.⁴ FRAD also provides a useful set of criteria for evaluating the usefulness of authority control systems. The FRAD user tasks—basically a set of values that describe how authority data can assist users—are particularly interesting to the authors of this article. FRAD's model has been adopted by RDA, the current cataloging code on authority control. RDA 9.18 is of particular interest to this discussion, as it

establishes a core element called “Identifier for the Person,” which is “uniquely associated with a person, or with a surrogate for a person (e.g., an authority record).”⁵

Catalogers have historically not provided authority control for authors of journal articles for several reasons. It has been partly an economic choice; the sheer scale of articles coming into a library is huge, and creating authority records can be very complex and time-consuming. Journals are frequently added to and dropped from vendor packages, typically without any notification reaching cataloging staff. This has also been a question of control; journal data are almost invariably created by indexing databases or journal publishers. Catalogers are not able to make any changes to these records to better serve their users. Another issue is that there have not been enough trained catalogers to create millions of name authority records for authors of journal articles. Finally, administrators have historically expressed concern that the time-consuming work of authority control does not present a clear return on investment.⁶ However, two general trends in the library world lead us to question this historical exclusion.

The first key trend is that more libraries are directing their patrons to use discovery layers. These interfaces assume that journal articles—not journals themselves—are the “objects of desire.”⁷ They intermingle records for articles with records that have seen more traditional bibliographic control. Sources of these traditionally controlled data—library catalogs and sometimes institutional repositories—may practice authority control, while the huge indices of articles that make up the majority of most search results do not. The FRAD user tasks of Find, Identify, Contextualize, and Justify fall to the end user of these systems, rather than catalogers. When you have found an article by an author in any of the leading discovery layers, it is very difficult to find any other works that they have published within the discovery tool. It is also hard to be sure that two authors in a discovery layer are the same without consulting outside reference sources. This problem is compounded by the incredible rate at which new journal articles are published.⁸ Within discovery layers, it is difficult for users to find articles by the authors they desire.

The second key trend is the emergence of several international databases that provide unique identifiers for authors. When catalogers can create links between pre-existing data sources, rather than spending time to create new authority records, authority control projects become much more feasible. Fortunately, an ever-growing list of institutions provide such “linkable” data sources. The International Standard Name Identifier (ISNI), LC/NAF, Open Researcher and Contributor ID (ORCID), Scopus, Virtual International Authority File (VIAF), and VIVO are examples of sources for author identifiers that use linked data standards to some extent. Some publishers, like the Nature Publishing Group, have also begun to provide identifiers for their journal contributors.⁹

New cataloging tools will encourage catalogers to use unique identifiers to link resources to author data expressed using the Resource Description Framework (RDF), a family of standards that conceptualizes data in subject-predicate-object

expressions. The Bibliographic Framework Initiative (BIBFRAME), an RDF-based ontology developed by the Library of Congress for bibliographic description, actually encourages catalogers to create such links. The Library of Congress' BIBFRAME Editor (BFE) allows catalogers to select personal and corporate names from authority sources, and suggests controlled forms of headings as a cataloger type.¹⁰ Catalogers will have a way to quickly link works to established identities that are expressed in linked data formats. When interfaces such as BFE see maturity, we will have a feasible way to provide authority control for authors who write articles, provided we can identify linked data sources that include the relevant authors and provide sufficient data for us to perform the FRAD user tasks. Given this context, we propose a new, high-impact role for catalogers: use linked data to describe authors of journal articles with authorized name access points. It is time to expand authority control to a new level.

Literature review

The growing interest in library linked data is very important to our current research. Serials librarians are particularly excited by the possibilities of linked data to free bibliographic description from the constraints imposed by our current record-based model. "The linked data model [...] opens up many opportunities for the provision of value-added content to bibliographic descriptions."¹¹

We are particularly interested in the representation of article authors within linked data name databases. In 2015, Panigabutra-Roberts studied the representation of a convenience sample of 55 faculty members at American University in Cairo, Egypt, which is a liberal arts institution with a relatively small research output and a very new Ph.D. program. The study found that over 50% of these faculty were represented in VIAF; with smaller numbers in LC/NAF, ResearchGate, and ISNI; and slightly over 30% in Google Scholar. Different disciplines saw different patterns of representation. Engineering faculty, for example, were not well represented in the "whole book-centric" LC/NAF, but saw much greater representation in VIAF, which included more conference proceedings. Panigabutra-Roberts commented on the self-registered services of ResearchGate and Google Scholar, noting that they are English language-dominant, incomplete, and may not be updated by the researchers. Panigabutra-Roberts also identified ResearchGate as "free but not innocent," noting that its goals align more with profit-seeking than with the open access ethos of library work. Her analysis also highlighted the fact that authors in her sample often romanized their names differently than did the name authority databases she consulted.¹²

In a similar vein, Waugh, Tarver, and Phillips explored the representation of 200 names in their electronic thesis and dissertation collection in 2014. They found that 28% of the names had identifiers in VIAF, 26% in LC/NAF, and only 0.5% in Wikipedia.¹³ The lower rates of VIAF and LC/NAF representation found in this study may reflect that authors of theses and dissertations have shorter publishing

histories than established faculty members, but may also be a function of a different setting or sampling method.

Once they find unique identifiers for authors, a few libraries are embedding those identifiers directly into MARC data. Particularly interesting is George Washington University's project that adds identifiers to its bibliographic records using MarcEdit software. This project located these identifiers in the subfield 0 of several fields, such as the X00, X10, X11, X30, 240, and certain 6XX fields. It broke MARC rules concerning these subfields' format—they are meant to contain a qualifying organization code followed by a control number—to present these identifiers as “fully realized and actionable URIs” which are ready to be part of linked data descriptions.¹⁴ The findings of this project are being investigated by the Program for Cooperative Cataloging (PCC) Task Group on URIs in MARC.¹⁵ However, as Folsom notes, we should not expect to see wide adoption of such practices yet.¹⁶

OCLC has invested a lot of research into representing researchers with identifiers and linked data. Their motivations focus on the needs of universities to track scholarly output, rather than the needs of library end-users to complete the FRAD user tasks. They do, however, provide a very thoughtful analysis of the current state of affairs with author identifiers.¹⁷ Two major OCLC projects: Schema.org Bib Extend and WorldCat Identifiers will have major impacts on how author data are expressed in a linked data environment.

Author name disambiguation is seen as a major unsolved problem by our colleagues in the field of information science. Smalheiser and Torvik describe manual, semi-automated, and automated approaches to the problem, and clearly list the issues inherent in the researcher disambiguation problem. They describe the problem of compiling training data for machine learning approaches, the issue of blocking very unlikely matches to reduce computational cost, and the added challenges that co-authorship present. We agree with Smalheiser and Torvik's assertion that researchers themselves should not be in charge of the disambiguation process, based on their provocative anecdotal evidence that researchers are surprisingly unreliable at identifying their own works. However, we believe that the evidence they present does not rule out manual identification of article authors entirely, as catalogers are very skilled at efficiently making these determinations.^{18,19}

Methodology

Our study sought to identify sources of identifiers suitable for providing authority control for authors of journal articles. We framed this primarily as a question of how likely a source was to include identifiers for a given journal author. Rather than choosing a random sample of authors, we created a sample that intentionally included a set of authors from diverse disciplines and worldwide locations. Our sample includes contributors to the following three journals. *Cataloging & Classification Quarterly*, a library science journal, is published in eight issues a year by a major journal publisher. *Perspectives of New Music*, a music journal, is published

semiannually by an independent corporation. *IEEE Intelligent Systems*, a computer science journal, is published bimonthly by a professional society. Our hypothesis was that the majority of authors of articles in recent volumes of these journals would be represented in name authority databases.

We chose a recent volume of each journal: volume 52 of *Cataloging & Classification Quarterly*, which contains 49 articles by 90 distinct authors; volume 52 of *Perspectives of New Music*, which contains 30 articles by 28 distinct authors; and volume 29 of *IEEE Intelligent Systems*, which contains 40 articles by 173 distinct authors. We created a spreadsheet containing an entry for each author of an article in those volumes, containing article title, digital object identifier (DOI), author affiliation, whether they are the first author listed on the article, and other data useful for disambiguation. We manually searched for each author in the ISNI, ORCID, Scopus, and VIAF databases, and added these identifiers to our spreadsheet.

The decision to search ISNI, ORCID, Scopus, and VIAF was informed by literature review and preliminary searches in several name identifier databases. We selected ISNI because of its impressive size and connections to the library community.²⁰ The British Library was one of the founders of ISNI. The PCC added ISNI identifiers to Name Authority Cooperative Program (NACO) records in the summer of 2015. We searched ISNI in November 2015 and performed a second search in June 2016. We selected ORCID because of its unique approach of relying on authors to manage their own unique identities. ORCID identifiers are assigned from a reserved block of ISNI identifiers for scholarly researchers and administered by a separate organization. Individual researchers can create and claim their own ORCID identifier. The two organizations coordinate their efforts.

We considered an ORCID record to match an author in cases where the forms of their name were exactly the same, or if there were some kind of data to differentiate different authors with the same name. Unfortunately, ORCID entries are overwhelmingly undifferentiated. When more than one author had the same name and no other information provided, we did not include it in our spreadsheet. We searched the ORCID database in November 2015 and performed a second search in March 2016.

We selected Scopus because of its incredible size, with over 60 million records, and because of its good practices with diacritics. When we couldn't find an author using Scopus' author search tools, we searched the title of their articles to find their institutions and Scopus IDs. We searched the Scopus database in November 2015 and performed a second search in March 2016.

We selected VIAF because it includes 35 authority files from around the world and it is representative of traditional authority control methods. We chose not to include LC/NAF for the present study because its records can also be found in VIAF. We searched VIAF in April 2015 and performed a second search in March 2016.

Despite our excitement about the VIVO project, we excluded it from the present study after an unsuccessful initial search of the available VIVO installations. Most

of the journal article authors in our sample did not work at VIVO institutions. This could be because VIVO is a new project, and has not yet been fully implemented at many institutions (only 17 at the time of this writing). Furthermore, many of the authors in the engineering journal work in industry, rather than academia, and would not be represented in any university's VIVO installation.

Similarly underwhelming representation in preliminary searches led us to exclude Frontier Loop from the present study. We also did not include ProQuest's Scholar Universe, because its identifiers are also included in ISNI.

Wikipedia has been suggested as a name authority database.²¹ However, Wikipedia's policies require a significant and well-documented impact for an academic or researcher to be represented in its database, and we considered it unlikely that we would find a large number of article authors in Wikipedia.²²

Once our spreadsheet was complete, we calculated the overlap of coverage for each pair of databases, using the formula established by Bearman et al. in their study of bibliographic database overlap.²³

$$\% \text{ overlap in } A = 100 \times \frac{|A|}{|A \cap B|}$$

In this formula, $|A|$ represents the number of authors represented in database A, and $|A \cap B|$ represents the number of authors represented in both databases A and B. It is helpful to read these percentages as "the percentage of authors in database A who can also be found in database B." We also ran a chi-squared test to check for a significant difference between representation of first authors and non-first authors.

Results

We found 290 unique authors identifiers (all but one of the authors in our sample) in Scopus, 111 unique authors in VIAF, 95 unique authors in ISNI, and 42 unique authors in ORCID. All authors in our sample were represented in at least one database. One hundred and forty-two authors were found in at least one of the openly licensed databases (ISNI, ORCID, and VIAF). More specific details about numbers of authors represented in various databases can be found in [Table 1](#).

While the four databases complemented each other very nicely, there was a great deal of overlap. In fact, ORCID and either ISNI or VIAF could have been left out completely, and we still would have found that every author was represented. Setting aside Scopus, with its nearly comprehensive coverage, we see that ISNI and VIAF are close cousins. 86.3% of the authors we found in ISNI could also be found in VIAF, and 73.9% of those found in VIAF were also represented in ISNI (see [Table 2](#)).

Each journal had the same ranking of author representation, with Scopus providing near-comprehensive coverage, followed by VIAF, ISNI, and finally ORCID offering the smallest number of author identifiers. However, the degree of this pattern was not uniform across all journals. For example, VIAF contained a

Table 1. Representation of journal article authors in each database.

	Represented in one DB	Represented in 2 DBs	Represented in 3 DBs	Represented in all 4 DBs
ISNI	0			
Scopus	149			
ORCID	0			
VIAF	0			
ISNI and Scopus		7		
ISNI and ORCID		0		
ISNI and VIAF		1		
Scopus and ORCID		18		
Scopus and VIAF		25		
ORCID and VIAF		0		
ISNI, Scopus, and ORCID			3	
ISNI, Scopus, and VIAF			63	
ISNI, ORCID, and VIAF			0	
Scopus, ORCID, and VIAF			7	
ISNI, Scopus, ORCID, and VIAF				18
Total authors = 291	149 (51.2%)	51 (17.5%)	73 (25.1%)	18 (6.2%)

noteworthy 82.1% of the authors of *Perspectives of New Music*, with many of their records containing references to musical works they had composed, rather than monographs. Only 26.6% of the authors in *IEEE Intelligent Systems* were represented in VIAF, which is higher than their representation in ISNI or ORCID, but still an unsatisfactory level. Authors for *Cataloging & Classification Quarterly* were much more likely to be represented by ORCID identifiers than were authors for the other journals. A complete breakdown of representation by journal can be found in [Table 3](#).

Fifty-nine percent of first authors were represented in multiple databases whereas only 43% of non-first authors were represented in multiple databases (see [Table 4](#)). The difference in representations is significant, $\chi^2(1, N = 291) = 7.5233$, $p = 0.006091$.

Discussion

Representation in ISNI

ISNI includes rich data for disambiguation, including affiliations, works, and dissertation information. However, it appears that these rich data are only available in its Web interface, not its XML or RDF formats. When ISNI can add more of these

Table 2. Overlap percentage based on Bearman et al.'s formula. Service A = Service of row.

	ISNI	Scopus	ORCID	VIAF
ISNI	—	98.9%	22.1%	86.3%
Scopus	32.4%	—	14.5%	37.9%
ORCID	50.0%	100.0%	—	59.5%
VIAF	73.9%	99.1%	80.2%	—

Table 3. Percentage of names represented in name databases by journal.

	ISNI	ORCID	Scopus	VIAF
Cataloging and Classification Quarterly	41.1%	25.6%	100.0%	50.0%
IEEE Intelligent Systems	22.0%	12.1%	100.0%	26.6%
Perspectives on New Music	58.6%	6.9%	96.5%	82.1%

rich data to its RDF representations, we believe that it would be a suitable database for providing name authority control for journal article authors.

Representation in ORCID

ORCID had the smallest representation of the databases we consulted for this project. Our ORCID searches were complicated by a preponderance of identifiers that were attached to non-unique names and contained no other distinguishing information. In fact, at the time of this writing, only 499,518 of the 2,433,434 ORCID identifiers were attached to any works.²⁴ When no works or biographical data are connected to an ORCID identifier, the ORCID interface returns the string “No public information available,” which was a common sight as we searched in this database.

Authors for *Cataloging & Classification Quarterly* were much more likely to be represented by ORCID identifiers than other authors were. This may point to a greater familiarity with ORCID among catalogers than among engineers or musicians. It may also be related to the journals’ manuscript submission processes. To manage submissions, *IEEE Intelligent Systems* and *Cataloging & Classification Quarterly* both use ScholarOne ManuscriptsTM—software from Thomson Reuters. In 2012, ScholarOne ManuscriptsTM added a feature that prompts authors to create or link to their ORCID identifiers during the submission process. Both *IEEE Intelligent Systems* and *Cataloging & Classification Quarterly* have that feature enabled.²⁵ Another possible explanation is related to differences in how authors publish their work. Many of the authors for *Perspectives of New Music* are known for their compositions and performances as much as their academic writing and ORCID is more often associated with papers rather than compositions and performances.

ORCID is a self-managed database. Besides the problems with researchers not updating their profile, ORCID is also less reliable than more traditional name authority databases.

We believe that ORCID would be useful as a supplementary database for providing name authority control for journal article authors, but should be used in

Table 4. Representation of first authors and non-first authors in multiple name identifier databases.

	Multiple databases	Single database
1st author	64	44
Not 1st author	78	105

conjunction with other, more complete databases, particularly in disciplines without heavy ORCID representation.

Representation in Scopus

Scopus was the most comprehensive source for author identifiers in our study by far. However, Scopus is a licensed database that many small and medium-sized institutions do not subscribe to. Linking to a Scopus ID at these smaller institutions will still help to identify specific authors, but will not allow access to data that can serve to contextualize authors or clarify their relationships. Scopus proved to have a near comprehensive representation of journal article authors, but it may not be a suitable source for identifying creators of other creative works, particularly artistic ones. The only author in our sample not represented by a Scopus identifier contributed a poem, rather than a typical academic article, to the journal in question.

Scopus also has an unfortunate tendency to assign multiple identifiers to the same person. One author in our sample received three Scopus IDs because their articles were published in three different journals. Authors can also manage their Scopus IDs, including merging duplicate identifiers.

Scopus includes affiliation data for some authors, which is incredibly helpful for disambiguating journal authors. However, when authors change affiliations, only the newer affiliation or older affiliation tends to be listed. The affiliation data also contributes to the multiple identifier problem: an affiliation might be represented in two different ways (e.g., *National University of Defense Technology, School of Electronic Science and Engineering, Changsha* vs. *National University of Defense Technology, Changsha, China*), resulting in two different Scopus identifiers.

Scopus continues to grow; a recent blog post indicated that it will be adding monographs to its database as well.²⁶ We believe that Scopus would be a suitable database for providing name authority control for journal article authors, especially for institutions with a Scopus subscription.

Representation in VIAF

VIAF had the second highest representation of journal article authors in our sample. VIAF's metadata are very rich, including data from several national libraries. It also represents the smallest break from traditional authority files.

We must note a major complication in linking to VIAF: its fluctuating identifiers. As more data are added from participating libraries, clusters of authority records may coalesce or split, leading to some fluctuation in the VIAF identifier of certain authority records.²⁷ We believe that, despite this challenge, VIAF would be a suitable database for providing name authority control for journal article authors.

General issues

Several issues complicated our findings. In this section, we will list these issues, which range from authority control challenges to limits on the current study's generalizability to methodological considerations.

We had great difficulty identifying authority records for authors with Chinese names, even though one of us is a native speaker of Chinese. We saw pairs of different names that both have the same romanized form. Some pairs of authors had the same name and also work at the same university, complicating our disambiguation work.

The interdisciplinary nature of *IEEE Intelligent Systems* also complicated matters. This journal's authors were philosophers, psychologists, computer scientists, information scientists, biomedical researchers, and engineers. This caused disambiguation trouble in VIAF and ISNI, where an author might be differentiated with the names of monographs they had authored. If these monographs were in a field such as psychology or philosophy, we often needed to look at the author's online CV to resolve the identity question.

IEEE Intelligent Systems also saw lower representation from the non-Scopus databases because of its greater number of coauthors (with up to nine authors contributing to some articles). Unfortunately, non-first authors are not as likely to be represented in multiple databases as their colleagues. This has unfortunate ramifications for authors who write in highly collaborative fields.

We also missed an opportunity by not searching arXiv and Google Scholar. These two author name databases could have potentially included several more authors who contributed to *IEEE Intelligent Systems*, because of its strong computer science presence.

Other factors limited the generalizability of our research. First of all, we did not use randomization in any way when we created our sample. Furthermore, we chose recent journal issues containing articles by contemporary researchers. Representation of these authors might be very different for older journal articles. For example, the ORCID registry went live in 2012 and so authors who were active in the 1980s are unlikely to be represented in its database.

Perspectives of New Music had a smaller number of articles and distinct authors than the other two journals. Additionally, the volume we consulted included a special issue. This sample may have been too small to provide a precise estimate of the proportion of recent *Perspectives of New Music* authors with identifiers in the databases we searched.

A final flaw is apparent in our methods: we performed no check for inter-searcher consistency. Prior to searching, we had only a basic discussion about how we would determine matches in authority databases, and relied primarily on our cataloger's judgement, rather than a specific procedure, for searching. We suggest that future studies of this type that use human searchers employ more formal controls for inter-searcher consistency.

```

@prefix dc: <http://purl.org/dc/elements/1.1/> .

@prefix lcnaf: <http://id.loc.gov/authorities/names/> .

<doi:10.1080/01639374.2015.1009671> dc:creator lcnaf:no2013089082 .

```

Figure 1. A very simple link between a bibliographic description and an authority record expressed in RDF Turtle.

Implications for practice

Author data with unique identifiers, when expressed in a language such as RDF, can radically transform authority control. Currently, a cataloger might find an authority record and record the authorized form of the name in a bibliographic record, then periodically check it to make sure there have been no changes. Many libraries hire vendors to do this authority maintenance work instead.

If we link to an external linked data source, this process becomes as simple as writing a single line of RDF and relying on the external authority agency to maintain the authority record. For example, if we wanted to express that a particular article is written by a particular author, we can type a few short lines in RDF Turtle (see [Figure 1](#)).

In this admittedly simplified example, the Library of Congress maintains the data related to this person, so there is no need for a cataloger to update the authorities on their local system. However, if they feel more data are necessary to identify or contextualize the person or justify their selection of heading, they could add more information using an ontology such as BIBFRAME.

The example presented in [Figure 2](#) includes LC/NAF and ISNI identifiers for the same person. The two sources refer to the author with slightly different strings: one includes a birth year; the other does not. Since the two authority records are linked to the same `bf:Contribution`, they are assumed to refer to the same “agent and role with respect to the resource being described,” not to two people with similar names.^{28,29}

This example also included `rdfs:label` objects for the authorized form of the person’s name. In this example it is important to note that a library would still have to concern itself with updating the headings in the `rdfs:label` objects when authors’ names or cataloging practices change. However, this type of description is still easier to update than a traditional local authority record, since it includes a URI that a script could use to retrieve the most current heading from a name authority database.

This example also used the friend of a friend (foaf) vocabulary to add details about the author’s research interests and affiliation. Specifically, the foaf lines tell us that Dr. Lilley is affiliated with an organization with a web page identified by `<http://www.massey.ac.nz>` and is interested in the topic described by the LCSH heading `sh95010034`, namely “Maori (New Zealand people) and libraries.”

```

@prefix bf: <http://bibframe.org/vocab/> .
@prefix isni: <http://isni.org/isni/> .
@prefix lcnaf: <http://id.loc.gov/authorities/names/> .
@prefix lcsh: <http://id.loc.gov/authorities/subjects/>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rel: <http://id.loc.gov/vocabulary/relators/> .

_:article a bf:Work;
  bf:contributor _:author .

_:author a bf:Contribution;
  bf:role rel:aut ;
  bf:agent lcnaf:no2013089082 ;
  bf:agent isni:0000000418419694 .

lcnaf:no2013089082 rdfs:label "Lilley, Spencer, 1963-" ;
  a bf:Person ;
  foaf:interest lcsh:sh95010034 ;
  foaf:workplaceHomepage <http://www.massey.ac.nz/> .

isni:0000000418419694 rdfs:label "Lilley, Spencer".

```

Figure 2. A sample BIBFRAME 2.0 link between a bibliographic description and two authority records, presented in the Turtle serialization.

An RDF-based discovery layer that allows catalogers to link journal article authors to meaningful name identifier sources may not be far off. OCLC's WorldCat Discovery Service API provides bibliographic data in the Turtle RDF format, including many linked data URIs for entities, which opens up opportunities for discovery tools that allow catalogers to add authority control.

Conclusion

This research explored the representation of article authors in four authority identifier databases.

The growing prominence of linked library data has caused our profession to be increasingly interested in authority control and external data sources that have previously not been used in library work. Our research suggests that three of these data sources—ISNI, Scopus, and VIAF—would be suitable as a standalone database of names for journal article authors, and that a combination of ISNI, ORCID, and VIAF provides a reasonable level of representation for these authors. All of these databases continue to grow in size and usefulness, so we recommend future study to assess quality and adoption.

We envision discovery layers that use linked data ontologies such as BIBFRAME to describe articles. If we start seeing these discovery layers, we recommend that they link to external authority databases to provide authority control for the articles our users desire. We also recommend that further research be done on the suitability of ISNI, ORCID, Scopus, VIAF, and other sources for name information to provide authority control for other areas of bibliographic description that are not well represented in LC/NAF, such as archival or thesis and dissertation collections.

If we are wise in our implementation of the identifier model of name authority control, our users will once again have the power to find, identify, and contextualize authors and justify those assertions using library tools.

ORCID

Jane Sandberg  <http://orcid.org/0000-0001-7038-778X>

Qiang Jin  <http://orcid.org/0000-0002-9383-5650>

Notes

1. Charles Ammi Cutter, *Rules for a Dictionary Catalog* (Washington, D.C.: US Government Printing Office, 1891), 133.
2. “I.F.L.A. International Conference on Cataloguing Principles, Paris, October 9-18, 1961,” *Library Resources and Technical Services* 6, no. 2 (1961): 165.
3. “Chapter 22: Headings for Persons,” American Library Association, *Anglo-American Cataloguing Rules*, 2nd ed. (Chicago: American Library Association, 1978).
4. IFLA Working Group on Functional Requirements and Numbering of Authority Records, 2009, *Functional Requirements for Authority Data: A Conceptual Model*.
5. RDA Toolkit (2013), <http://access.rdatoolkit.org/> (accessed August 6, 2016)
6. Susan K. Burke and Jay Shorten, “Name Authority Work Today,” *Library Resources & Technical Services* 54, no. 1 (2011): 4–20.
7. “PCC Strategic Themes 2014,” <https://www.loc.gov/aba/pcc/documents/Strategic-Themes-2015.pdf> (accessed August 2, 2016).
8. Bo-Christer Björk, Annikki Roos, and Mari Lauri, “Global Annual Volume of Peer Reviewed Scholarly Articles and the Share Available via Different Open Access optionsm,” in *Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 -*

Proceedings of the 12th International Conference on Electronic Publishing, Toronto, Canada, 25-27 June 2008, ed. Leslie Chan and Susanna Mornati, 78-186. http://elpub.scix.net/cgi-bin/works/Show?178_elpub2008

9. "Contributors Dataset," Nature Publishing Group, <http://www.nature.com/ontologies/datasets/contributors> (accessed November 19, 2015).
10. "BIBFRAME Editor," Library of Congress, last modified October 26, 2015, <http://bibframe.org/tools/editor>.
11. Marlene van Ballegooye and Juliya Borie, "From Record-Bound to Boundless: FRBR, Linked Data, and New Possibilities for Serials Cataloging," *The Serials Librarian* 66, no. 1-4 (2014): 76-87.
12. Anchalee Panigabutra-Roberts, "Researchers' Identity Management in the 21st Century Networked World" (presented at Hamburg, Germany, 2015).
13. Laura Waugh, Hannah Tarver, and Mark Edward Phillips, "Introducing Name Authority into an ETD Collection," *Library Management* 35, no. 4/5 (2014): 271-283.
14. Jackie Shieh and Terry Reese, "The Importance of Identifiers in the New Web Environment and Using the Uniform Resource Identifier (URI) in Subfield Zero (\$0): A Small Step That Is Actually a Big Step," *Journal of Library Metadata* 15, no. 3-4 (2015): 208-226, <http://dx.doi.org/10.1080/19386389.2015.1099981>
15. "PCC Task Group on URIs in MARC," <https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html> (accessed July 13, 2016)
16. Steven Folsom, "URIs you Can Expect to Find in Library Records (in 8 Minutes or Less)," last modified February 23, 2015, <https://wiki.duraspace.org/download/attachments/68060801/LD4LUC2.pptx>.
17. Karen Smith-Yoshimura, Micah Altman, Michael Conlon, Ana Lupe Cristán, Laura Dawson, Joanne Dunham, Thom Hickey, et al., *Registering Researchers in Authority Files* (Dublin, OH: OCLC, 2014), <http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf>.
18. Neil R. Smalheiser and Vetele I. Torvik, "Author name disambiguation," *Annual Review of Information Science and Technology* 43, no. 1 (2009): 1-43.
19. For a beautiful taxonomy of automated author name disambiguation methods, see Anderson A. Ferreira, Marcos André Gonçalves, and Alberto HF Laender, "A Brief Survey of Automatic Methods for Author Name Disambiguation." *ACM Sigmod Record* 41, no. 2 (2012): 15-26.
20. Anila Angeli, Andrew Mac Ewan, and Vincent Boulet, "ISNI and VIAF—Transforming Ways of Trustfully Consolidating Identities," (paper presented at IFLA 2014, Lyon, France, 2014), <http://library.ifla.org/985/1/086-angeli-en.pdf>.
21. Jesús Jiménez Pelayo, "Wikipedia Como Vocabulario Controlado: ¿Está Superado el Control de Autoridades Tradicional?," *El profesional de la información* 18, no. 2 (2009): 188-201.
22. "Wikipedia: Notability (academics)," *Wikipedia*, https://en.wikipedia.org/wiki/Wikipedia:Notability_%28academics%29 (accessed July 6, 2016)
23. Toni Carbo Bearman and William A. Kunberger, *A Study of Coverage Overlap Among Fourteen Major Science and Technology Abstracting and Indexing Services* (Philadelphia: NFAIS, 1976).
24. "ORCID Statistics," <https://orcid.org/statistics> (accessed January 14, 2016)
25. "v4.10 Release Notes," *ScholarOne Manuscripts*, <https://mchelp.manuscriptcentral.com/releasenotes/4.10.pdf> (issued September 2012)
26. "Books, Books and More Books: Scopus' Titles Expansion Program," <https://blog.scopus.com/posts/books-books-and-more-books-scopus-titles-expansion-program> (accessed August 8, 2016)

27. Thomas Hickey and Jenny A. Toves, "Managing Ambiguity in VIAF," *D-Lib Magazine* 20, no. 7/8 (2014). <http://dx.doi.org/10.1045/july2014-hickey>.
28. "Bibframe 2.0 Vocabulary," <http://id.loc.gov/ontologies/bibframe.html> (accessed August 1, 2016)
29. Thanks to Ray Denenberg and Robert Sanderson for their help with this example provided over the BIBFRAME email list.