

## Authors and Authorities in Post-RDA Library Systems: A Case Study

**Heather Lea Moulaison**

The iSchool at the University of Missouri, Columbia, USA.



Copyright © 2014 by Heather Lea Moulaison. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

---

### Abstract:

*Persons are complex, and their representation in library authority records is becoming increasingly complex through the addition of attributes under RDA rules. This case study, using a longitudinal approach, examines attributes in authority records from the Merlin Cluster of academic libraries at both six months (i.e. September 2013) and one year (i.e. April 2014), after the official adoption of RDA to assess trends in attribute field usage.*

*This study revealed that after one year of RDA cataloging, 87.58 percent had no attributes at all, but that percentage was down -2.43% from the percentage at six months. Almost 8% of authority records had at least one attribute after a year; and almost 5% had two or more attributes. In terms of the attributes recorded, dates recorded in the 046 of authority records were the most common with 7.1 percent of records including them after six months of RDA cataloging and nearly 10 percent of authority records including them after one year of RDA. The 678 (Biographical or Historical Data) and the 375 (Gender) were the next most common attributes supplied.*

*Overall, attributes supplied tend to be sparse and focus on English-speakers, males, and those associated with universities in some way. Although this analysis investigates the metadata quality indicator of completeness for attributes in post-RDA authority records, in the process of investigating aspects of the demographics of persons in the authority file, a few informal observations came forth.*

**Keywords:** personal name authority records, metadata quality, RDA, case studies

---

### Introduction

Resource Description and Access (RDA)-based library cataloging records, including authority records, are designed to capture information about entities in the bibliographic universe as never before. Library catalogs have historically focused on providing information about materials in the collections to assist ultimately with retrieval. These materials, however, do not exist independently of the people they were created by and are about. The *Functional Requirements for Bibliographic Records* (FRBR) lists *find* as the first user task supported by the FRBR conceptual model, followed by *identify*, *select*, and *obtain* (FRBR, 2009). A truly robust *find* functionality would permit library catalog users to focus their searches not only on

library materials and their attributes, but also on the attributes of and relationships between related persons. This in turn would allow for retrieval based on criteria hitherto unthinkable. Being able to pull a set of items written in Chinese by women dentists who are also poets remains a dream given the current library systems, but one that could potentially be a reality in the very near future.

Persons are complex; it is a vast understatement to state that people are much more complex than the library metadata that describes them. Persons are represented in two ways in library systems: through the personal name character string used as an access point or *identifier* in the bibliographic record, and through the information housed in the authority record. RDA defines the identifier for the person as “a character string uniquely associated with a person, or with a surrogate for a person (e.g. an authority record). The identifier serves to differentiate that person from other persons” (RDA Toolkit, 2014) and to assist with the finding function of catalogs. Attributes of persons, along with the identifier and justification for the creation of the identifier, are recorded in RDA authority records, records that historically have not contributed to retrieval in a robust way (Yee, 2005). Attributes in authority records help identify persons and have the potential to help differentiate them as well. Personal name authority records based on RDA are now permitting more information, such as gender, occupation, associated place, language and others (RDA Toolkit, 2014), to be recorded as attributes and that data is designed to be usable into the future.

There is an interconnection between the character string for persons, their attributes, and the role of authorities in the library system of the future. As the nature of personal name identifiers evolves, data must be present in the form of attributes to assist with the user task of *identifying*. Although with RDA, it is now possible to add additional information about individuals beyond that which was permitted using the *Anglo-American Cataloguing Rules*, second edition (AACR2), supplemental content is not necessarily being provided in RDA-based authority records. What attributes are actually being included in authority records and what conclusions can be drawn from this? The potential opportunities afforded the end-user by the addition of personal data are great, especially if character strings in personal name identifiers can be simplified and differentiated.

### **FRAD and RDA: New Models for Access**

*Functional Requirements for Bibliographic Data* (FRAD) (2013) supports FRBR (2009) in identifying the kinds of personal name attributes that support organization and retrieval within the bibliographic universe. FRAD identifies fourteen attributes for persons, eleven of which are in addition to the attributes already identified in FRBR. The three attributes of persons identified in FRBR are dates, title, and other information associated with the person. The additional attributes identified in FRAD are gender, place of birth, place of death, country, place of residence, affiliation, address, language, field of activity, profession, and biography/history. Some of these attributes have a relationship to the person entity that they describe (FRAD, 2013). These attributes support the FRAD user tasks of *find*, *identify*, *contextualize*, and *justify*. Of these, end-users of library systems will be most interested in the first three: find, identify, and contextualize. In a full-text system, there is the potential for any attribute to assist with the finding function. Person attributes also assist with identification. Dates associated with a person and biography are considered to contextualize as well (FRAD, 2013).

Additionally, FRAD (2013) identifies a series of potential relationships between persons and the character strings that represent them as they pertain to library materials. In access points,

the person is primarily known by the name chosen by librarians according to RDA instructions. The character string, like other metadata, is an effective knowledge organization tool, but is nonetheless an artificial product of library metadata creation.

Resource Description and Access (RDA) is the content standard currently in use in the United States. The Program for Cooperative Cataloging (PCC) announced its *PCC Day One for RDA Authority Records* would coincide with the Library of Congress’s official adoption of RDA on March 31, 2013 (“PCC Day One,” 2011). Additional MARC fields were added to authority records to support the new data that can be recorded. Table 1 shows MARC authority fields that encode attribute information; all but the 678 (Biographical or Historical Data) are new with RDA (MARC 21 encoding, 2012). The 678 is being revived with the intention of displaying it to patrons (“MARC 21 Encoding,” 2012; “RDA Cataloger Training,” 2013).

Table 1: Person-Related Fields in MARC for Authority based on RDA ((R)=Repeatable (NR)=Nonrepeatable) (“MARC 21,” 2013)

<b>MARC For Authority Field Code</b>	<b>Description</b>
<b>046</b>	Special Coded Dates (R)
<b>370</b>	Associated Place (R)
<b>372</b>	Field of Activity (R)
<b>373</b>	Associated Group (R)
<b>374</b>	Occupation (R)
<b>375</b>	Gender (R)
<b>376</b>	Family Information (R)
<b>377</b>	Associated Language (R)
<b>378</b>	Fuller Form of Personal Name (R)
<b>678</b>	Biographical or Historical Data

Because RDA governs bibliographic records as well as access points and authorities, personal name access points and the underlying authority records underwent mass changes at the time of RDA’s adoption (c.f. “RDA Cataloger Training,” 2013; “PCC Day One,” 2012). These changes did not, however, provide for the mass copy of information from the access points to the attributes fields of authority records. All attributes have evidently been supplied manually by catalogers.

#### *Record Content Contribution*

Currently, the Library of Congress and all PPC (<http://www.loc.gov/aba/pcc/>) members are creating RDA bibliographic and authority records. Name Authority Cooperative Program (NACO) (<http://www.loc.gov/aba/pcc/naco/index.html>) members have been trained to provide RDA authority records, and documentation for the new RDA fields is at their disposal on the NACO website. A number of NACO funnels including funnels with themes relating to the fine arts like Art NACO Funnel, NACO-AV, and NACO-Music, are very active. Although NACO is an open program, only libraries with adequate budgets, staff, and time are in a position to consider it for their best and brightest librarians; alternatively, individual librarians need to undertake NACO training on their own and join a funnel if they

wish to participate. In both cases, once trained, catalogers must have the time available to create the new records or to update existing ones. If creating new authority records, these catalogers will be involved in cataloging workflows of materials that are not held by the Library of Congress; the materials they are cataloging are unique, otherwise there would be cataloging copy available. NACO-trained catalogers are, therefore, a somewhat exclusive group working largely in academic libraries with the support to enable them to participate in this prestigious national program.

#### *Metadata Quality: Completeness*

One of the measurement indicators of metadata quality is completeness (Zeng & Qin, 2009). Completeness covers the “range of descriptive detail provided by individual records as well as the overall distribution of various levels of records in a metadata database or a repository” (Zeng & Qin, 2009, p. 254). Within individual records, therefore, are all of the fields filled out? Having complete data in a single record is of little use in terms of recall if other records in the collection have incomplete data. In a collection such as a library’s or a consortium’s holdings, are all of the fields in all of the records filled out completely?

Completeness among records in a collection is a worthy goal. High quality and complete authority records in a single institution, once those records are mingled with records of inferior quality, may actually be seen as negatively influencing, at least temporarily, the quality of the overall collection of records. The dictum “the future is longer than the past” has been seen as a base value in cataloging (Miller, 2007), and has guided cataloging agencies striking out with RDA, especially in the beginning. Ideally, all authority records, though the active interaction of PCC libraries, will be updated and completed in time. At present, though, it is unclear exactly how many records have been updated and what the overall implications are for quality.

#### **Personal Name Authority Records**

Authority records have preferably always indicated one person (principle of unique headings). At times, though, the person’s characteristics are not sufficiently evident to the cataloger to enable a person’s information to be recorded as a unique record due to similarities in the character strings that comprise names. When two or more individuals are designated by the same character string, a single *undifferentiated* authority record is created. RDA defines undifferentiated name indicators as “A categorization indicating that the core elements recorded are insufficient to differentiate between two or more persons with the same name” (“RDA Toolkit,” 2014, sect. Undifferentiated Name Indicator). In 2013, a white paper authored by Reimer and Schreur (2013) recommended that undifferentiated name authority records be broken up. Semantic web technologies, including the Virtual International Authority File (VIAF) require unique and addressable authority records. Even if the character strings in the indicator are identical, the underlying authority records need to be unique for different persons.

#### *MARC as Encoding Scheme*

The context in which library data is found, namely, integrated library systems (ILSs) currently are adapted for use with the MARC encoding scheme. Given the library’s interest in the semantic web and linked data environments, the Library of Congress is investigating ways to move forward with library encoding in a way that will allow relationships to be made evident. The BIBFRAME (<http://www.loc.gov/marc/transition/pdf/marclid-report-11-21-2012.pdf>) initiative (Kroeger, 2013) is currently being put forth as the primary way in which

cataloging records will be encoded. Until the scheme is mature, however, authority records will need to continue to be created in MARC.

### Personal Name Identifiers

Personal name identifiers in library records contain character strings representing the last name of the person, the first name, and other information used to differentiate the character string from others at present or in the future. Table 2 shows the most popular subfields appearing in access points that relate to persons. Elements shown to end-users include titles, dates, relator terms or codes, qualifiers that spell out the fuller form of a name. Control numbers and linkages are potentially useful for systems but generally are not displayed.

Table 2: Bibliographic Records Personal Name Access Points (100 and 700) Subfields  
((R)=Repeatable (NR)=Nonrepeatable)

Subfield code	Description
‡a	Personal name (NR)*
‡c	Titles and other words associated with a name (R)**
‡d	Dates associated with a name (NR)**
‡e	Relator term (R)***
‡q	Fuller form of name (NR)**
‡0	Authority record control number (R)***
‡4	Relator code (R)***
‡6	Linkage (NR)***

Key: Mandatory\*, Required if applicable\*\*, Optional\*\*\*

Access points used in libraries have been criticized for the additional data they contain. How relevant is it for users to see birth and death dates as a means of differentiating authors? Booksellers like Amazon (<http://amazon.com>) do not include terms to differentiate authors in the access point for the name, for example. “Traditional authority control focuses on expressing identity by means of unique headings [...]. The unique name heading for a person was what ensured that bibliographic records which cited that person as a creator or contributor or subject could be gathered under one heading” (“Report for PCC,” 2013, p. 13).

With extraneous data in the character strings for access points, mapping between library controlled vocabularies and other systems is more challenging. DBpedia (<http://dbpedia.org/About>), the linked data version of Wikipedia, does not use dates to differentiate persons. Systems like the Internet Archive’s Open Library (<https://openlibrary.org/>) that pull from library resources as well as linked data datasets can have a difficult time with automatic record matching. This is not surprising since the library world initially experienced a set of challenges matching library authority records from different international systems in VIAF, the Virtual International Authority File (Bennett, Hengel-Dittrich, O’Neill, & Tillett, 2006).

#### *Personal Name Identifiers in WorldCat: Analysis*

Persons who are creators (authors, actors, etc.) are represented by access points in MARC bibliographic records in fields 100 and 700. Access points in bibliographic records are comprised of one or more subfields. Trends in the complexity of the access points can be analyzed based on fields used in bibliographic records. The *MARC Usage in Worldcat* project

(<http://oclc.org/research/activities/marcusage.html>) published by OCLC Research, makes data about WorldCat’s holdings available. Through the usage data for the 100 and 700 fields, the use of subfields in WorldCat records for personal names of creators can be evaluated. Dates, fuller forms of names (i.e. qualifiers) and titles have continued to be added, but not at a faster rate, over the course of the 13-month period under study. Qualifiers were added less frequently to controlled access points for personal names, as their percentage of use in access points fell over the course of the year. See Table 3 for data from OCLC Research, in a tabular form, for before and after the adoption of RDA. Authority records, as demonstrated in the previous section are changing, but so are the forms of access points being created.

Table 3: Complexity of Author Identifiers (i.e. MARC bibliographic 100 and 700) in OCLC’s WorldCat

‡	Description <sup>1</sup>	January 2013 occurrences	Jan 2013 % <sup>2</sup>	Jan 2014 occurrences	Jan 2014 % <sup>2</sup>	% change Jan 2013-Jan 2014
<b>a</b>	Personal name (NR)	291,478,411	100.00%	311,607,120	100.00%	0.00%
<b>d</b>	Dates associated with a name (NR)	85,309,924	29.27%	91,269,104	29.29%	+0.02%
<b>4</b>	Relator code (R)	43,604,446	14.96%	48,798,542	15.66%	+0.70%
<b>0</b>	Authority record control number (R)	25,855,327	8.87%	29,363,041	9.42%	+0.55%
<b>e</b>	Relator term (R)	19,551,901	6.71%	23,745,659	7.62%	+0.91%
<b>6</b>	Linkage (NR)	15,406,166	5.29%	18,011,782	5.78%	+0.49%
<b>q</b>	Fuller form of name (NR)	10,308,492	3.54%	10,837,508	3.48%	-0.06%
<b>c</b>	Titles and other words associated with a name (R)	8,785,193	3.01%	9,509,560	3.05%	+0.04%

### Current Study

The literature has not investigated the extent to which attributes data is being supplied in authority records for persons and how complete any additions are. Without complete metadata, future initiatives’ ability to recall full sets of items based on attributes will be questionable.

The present study, based on a longitudinal case study of authority records in a United States university library cluster, explores and analyses completeness of attribute information

<sup>1</sup> Description taken from <http://www.loc.gov/marc/bibliographic/bd100.html>

<sup>2</sup> For each, it was assumed that 100% of personal name identifiers contained the required, non-repeatable subfield a.

supplied in personal name authority records. Additionally, it investigates future options given the current state.

## Method

The current research study examines attributes in the set of authority records in an academic library consortium located in the state of Missouri, United States (see Appendix I). This case study, using a longitudinal approach, examines authority record attributes at both six months (i.e. September 2013) and one year (i.e. April 2014), after the official adoption of RDA to assess trends in attribute field usage.

On September 31, 2013, 6 months after the official adoption of RDA by the Library of Congress, attributes fields from personal name authority records (n=1,156,315) were pulled as four delimited text files. The first three files (with 300,000, 300,000, and 299,999 records respectively) were able to be analyzed further using Microsoft Excel. The fourth file, containing 256,317 records, was eye-readable but was corrupted and not able to be manipulated in Excel. To ensure that the fourth file was not statically different from the first three, 21,058 records (8.2%) were drawn at random from the corrupted file and were manually examined to make sure data analyzed was representative of the other three file's data. Because there were not statistically significant differences, an analysis of the usable data files was considered representative.

On April 7, 2014, after one year of RDA implementation, the same set of attributes fields from all personal name authority records for the Merlin cluster (n= 1,274,618) were again pulled. The records were pulled as thirteen text files and all were analyzed using Excel. Results below are indicated on a by-field or by-record basis as appropriate, meaning that if repeated fields were provided in a single record or if content was repeated within a given field, it was counted once for the record or the field. Additionally, this study adopted a *take what you see* policy of analysis: if content was entered incorrectly or in the wrong field, it was nonetheless counted as entered. Due to the size of the corpus, if 30 of the 1,274,618 records incorrectly entered content in the 375 Gender field, the content was nonetheless counted as an entry in the 375 field.

## Results

### *Completeness of Attributes at the Consortium Level*

Dates recorded in the 046 of authority records are the most common attributes included in the authority records with 7.1 percent of records including them after six months of RDA cataloging and nearly 10 percent of authority records including them after one year of RDA. Data appeared in the 678 next most frequently after the 046; the 678 is a field that had been used in the past and that is being used again. Although the overall trend this data shows is a slight reduction in the amount of 678 content, this trend could be because of inconsistencies in the dataset examined in September, 2013. Further analysis is needed in assessing this field in particular because of its previous use in cataloging. All of the new RDA 3xx fields saw an increase in their use over the period of study. See Table 4 for the results in tabular form.

Table 4: Field Use in MERLIN Authority Records for Persons

MARC Field Code	Description of Element	6 months: % of sample (N=899,999)	1 year: % of total (N=1,274,618)	Change
046	Special Coded Dates (R)	7.10%	9.62%	+2.52%
678	Biographical or Historical Data	2.84%	2.31%	-0.53%

<b>375</b>	Gender (R)	1.86%	3.39%	+1.51%
<b>374</b>	Occupation (R)	1.78%	3.12%	+1.34%
<b>370</b>	Associated Place (R)	1.49%	2.74%	+1.25%
<b>377</b>	Associated Language (R)	1.33%	2.44%	+1.11%
<b>378</b>	Fuller Form of Personal Name (R)	1.07%	1.45%	+0.38%
<b>372</b>	Field of Activity (R)	0.90%	1.69%	+0.79%
<b>373</b>	Associated Group (R)	0.89%	1.70%	+0.81%

### *Completeness of Attributes at the Record Level*

Attributes are given in authority records in the corresponding MARC field. To gauge the completeness of the records, any time data was included in a field, the field was counted; repeated fields were not taken into consideration. At six months, the vast majority (90%) of records has no attributes, with only 7% having additional attribute content in one MARC field, and only 3 percent having more than two attributes. At one year, more attributes were in the authority records overall. The percentage of records without attributes had dropped by 2.43 percent; only 88 percent (n=1116259) of records had no attributes. Almost 8 percent of authority records had at least one attribute after a year; and almost 5 percent had two or more attributes. See Table 5 for a more complete breakdown of the number of attributes in records.

Table 5: Number of Fields in Authority Records (per Attribute)

<b>No. of attributes by MARC field</b>	<b>6 months: % of total</b>	<b>1 year: % of total</b>	<b>Change in percentages</b>
<b>0</b>	90.01%	87.58%	-2.43%
<b>1</b>	7.05%	7.63%	+0.58%
<b>2</b>	0.89%	1.15%	+0.26%
<b>3</b>	0.28%	0.53%	+0.25%
<b>4</b>	0.49%	0.82%	+0.33%
<b>5</b>	0.50%	0.88%	+0.38%
<b>6</b>	0.45%	0.80%	+0.35%
<b>7</b>	0.26%	0.49%	+0.23%
<b>8</b>	0.08%	0.13%	+0.05%
<b>9</b>	0.01%	0.01%	+0%

Given the commonness of the 046 (9.62%, n=122676) in authority records in the study, and given the fact that 9.31% (n=118696) of records studied contain between one and three attributes, it is reasonable to assume that when a record contains an attribute, that attribute overwhelmingly is the 046 field. This is not surprising, as copying content from the †d of the identifier to the 046 of the authority record should be a straightforward procedure for catalogers if that data is already present, and it was, as noted, in nearly 30% of WorldCat access points.

### *Person Demographics Attributes*

Although a full content analysis of the attributes in the MARC records was not the goal of this study, a limited number of descriptive statistics provide some basis for future

considerations. Of the second and subsequently more complete set of records pulled in April 2014, 375 Gender and 377 Language attributes were assessed (see Tables 6 and 7). Males (80%, n=34515) were overwhelmingly represented in the supplied metadata. According to RDA 9.14.1.1., *Language* “of the person is a language a person uses when writing for publication, broadcasting, etc.” (RDA Toolkit, 2014). Authors who write in English (73%, n=22666), potentially among other languages, were also very well-represented. Table 7 shows a sampling of languages indicated in 377 fields. Because the field is repeatable, more than one language may be supplied in a single authority record.

Table 6: Gender as Recorded in MERLIN Authority Records

<b>375 Gender entries</b>	<b>Number of records with text (N=43077)</b>	<b>Of records with 375, % with this text</b>
male	34515	80.12%
female	8544	19.83%
listing both male and female	(-12)	(-0.03%)
presumed cataloging errors	30	0.07%
total	43077	100.00%

Table 7: Selected Languages as Recorded in MERLIN Authority Records

<b>377 Language entries</b>	<b>Number of records with text (N=31149)</b>	<b>Of records with 377, % with this text</b>
eng	22666	72.77%
fre	2066	6.63%
ger	1961	6.30%
spa	1505	4.83%
rus	1032	3.31%
lat	610	1.96%
chi	420	1.35%
jpn	269	0.86%
hin	117	0.38%
swe	92	0.30%
ice	11	0.04%

There are limitations involved in this kind of analysis of attributes in records, not the least of which is the fact that there is no way of knowing how representative this sample is of persons in the authority file of this academic library consortium’s cluster. Are academic collections in the United States so heavily skewed in favor of males and writers who use English, both potentially in terms of authors and subjects? If so, is there a discussion of the consequences in order? Questions of representativeness aside, the metadata supplied is for a group of

individuals strikingly unrepresentative of international demographics in scholarship over time.

As part of the investigation in to the demographics of persons, limited information about Occupations was also assessed. Occupations are generally text strings based on Library of Congress Subject Headings and are subject to all of the inconsistencies of free-text fields such as being mis-entered and containing typographical errors. Eight somewhat diverse occupations were sampled from among the 374 Occupation fields (N=39772) in search of a discernable pattern. Around 15 percent (n=6517) of 374 Occupation entries included the term *Author* or *Authors*. The activity of NACO members cataloging non-textual materials was evident in the creation of 374 fields relating to the performing arts such as *Actor\**, *Pianist\**, and *Musician\**. See Table 8 for more information.

Table 8: Selected Occupations as Recorded in MERLIN Authority Records

<b>374 Occupation entries</b>	<b>Number of records with text (N=39772)</b>	<b>Of authority records with 374 field(s), % with this text</b>
Author*	6517	16.39%
Professor*	2885	7.25%
Actor*	1962	4.93%
Poet*	1826	4.59%
Historian*	1724	4.33%
Lawyer*	1320	3.32%
Pianist*	1130	2.84%
Musician*	1109	2.79%

A very limited number of Associated Places (MARC 373) were also investigated with some success. Associated Places contain free-text entries for institutions that are or are based on corporate body access points. Of 373 fields containing data (N=21620), over half contained the term *University* (n=11040) and over 10% contained the term *College* (n=2517) implying that the information supplied in the 373 is primarily entered for educated people; if the 374 is indicative, then these people are not all professors, but may have been students. See Table 9 for a tabular representation of this data.

Table 9: Selected Associated Place Terms as Recorded in MERLIN Authority Records

<b>373 Associated Places entries</b>	<b>Number of records with text (N=21620)</b>	<b>Of records with 373, % with this text</b>
University	11040	51.06%
College	2517	11.64%

## Discussion

The potential opportunities afforded the end-user by the addition of personal data are great. Yet, to provide systems meets the needs of users, more than just English-language authors, males, and the educated, will need to be the object of attention. If we consider completeness an indicator of metadata quality, then the data being maintained on persons is not of high quality, neither in the completeness of data in individual records studied nor in the completeness of data in given fields across this collection.

Constraints that might be keeping personal data on authors from entering shared authority records include issues relating to individual library operations and collections. Catalogers with the NACO authorizations to create RDA authority records for persons are not able to keep up with the manual data creation scenario RDA requires. Based on the results presented above, we can infer that catalogers are encoding additional information for males with some relationships to higher education, potentially to their own universities where they work. Catalogers also may be supplying additional content for persons who are part of their fine arts workflows. This shot-gun approach to providing metadata is supplying library systems with only a sliver of information on a very small fraction of persons in the bibliographic universe. In library systems of the future, authority records will not only differentiate individuals, but will also allow for grouping by shared attributes if and only if the attributes are supplied. The richness and complexity of information about persons has the potential to assist with relationships and linkages in future systems such as the proposed BIBFRAME-friendly library, and that information must be complete if it is to be used to its potential.

The inclusion of attributes in authority records will alleviate the need for additional text beyond the name of the person to appear in the personal name identifier, adhering to the suggestions of the PCC Task Group on the Creation and Function of Name Authorities in a Non-MARC Environment (Report for PCC, 2013). The complex identifiers noted in WorldCat can be streamlined through the systematic addition of attributes to the authority record. Of the nearly 30% (29.29%, n=91,269,104) of personal name identifiers in WorldCat with dates, all of these can be moved or copied to the attributes area of the authority record. Although the study reported here does not examine the authority records stored in WorldCat, the recommendation could be carried out first as a pilot in a smaller corpus. The same approach to supplying attributes holds true for fuller forms of names generally listed as qualifiers (in 3.51% of identifiers, n=10,837,508) and titles and other words associated with a name (in 3.06% of identifiers, n=9,509,560).

One limitation of the present study is the fact that undifferentiated name headings are not permitted to contain person-specific fields since, by definition, the persons represented by the identifier are not known. It was not possible to ascertain the number of undifferentiated name headings based on the data that was pulled for this study. At some future point, it will become necessary to differentiate these headings, splitting apart the undifferentiated records to create as many authority records as necessary. These identifiers, therefore, will not necessarily have person-related data recorded right away in the underlying authority records. Future study may reveal when and how attribute data for persons is supplied in authority records, especially if the name is undifferentiated. If the author is neither prolific nor famous, it is difficult to imagine that much more content will be added to the authority record, effectively excluding her work from future person-attribute based searches unless strategic action is taken.

### **Future work**

If the catalog is to provide equitable access based on attributes, it needs to include attributes for all, equally. More persons than English-speakers, males, and the educated, will need to be the object of attention. In the future, the ideal library system will have access to complete sets of metadata that will assist uniformly with all of the user tasks identified in FRAD (2013): find, identify, contextualize, and justify. For this to happen, libraries must find a way to leverage data freely available on the semantic web. Inputting content on an ad-hoc basis by a select group of highly trained librarians cannot and will not suffice. As the PCC remarks, "It is integrating external data with one's own more than simply exposing one's own data that,

when engaged in by all parties, will enable more fluid navigation and leveraging of the richness of the linked data environment for enhanced discovery” (PCC Day One, 2013, p. 10). Authority data is unique and precious, but given the current aspirations, librarians need to resolve to relinquish full and complete control over the contents of authority record attributes and resolve to work through content from curated resources such as Wikipedia via DBPedia.

The vision presented here is not inconsistent with the BIBFRAME vision, where, as explained by the PCC, “Authorities are not designed to compete or replace existing authority efforts but rather provide a common, light weight abstraction layer over various different Web based authority efforts to make them even more effective” (“PCC Day One,” 2013, p. 13). The ISNI project has similar aspirations (MacEwan, Angjeli, & Gatenby, 2014). A similar example is evident in the author pages of Wikipedia. Author Chris Anderson’s entry ([http://en.wikipedia.org/wiki/Chris\\_Anderson\\_%28writer%29](http://en.wikipedia.org/wiki/Chris_Anderson_%28writer%29)) includes a section of authority records in library systems for the differentiated author. Human-readable (and machine-readable) links appear to his authority record in WorldCat, VIAF, LC[NAF], ISNI: 0000 GND, SUDOC, BNF, and NDL.

It is distinctly possible that not all persons associated with content in library collections have a presence on the semantic web; these persons will require additional attention by catalogers, but this work should be undertaken strategically. The amount of work done by catalogers will ideally be inversely proportional to the amount of data already in existence on the semantic web.

Additionally, attributes content provided in authority records should be supplied in a way that is machine readable. Free-text entries represent a challenge for systematic search and retrieval due to their inconsistent nature. Semantic-web enabled identifiers and the possibilities they represent should be investigated as a possible solution to the problems and vagaries of free-text entries that are merely text strings, not *data*.

## **Conclusion**

This study revealed that after one year of RDA cataloging, low-hanging fruit such as dates are the most commonly added attribute to authority records. Other attributes supplied tend to be sparse, and focus on English-speakers, males, and those associated with universities in some way. Although this analysis investigates the metadata quality indicator of completeness for attributes in post-RDA authority records, in the process of investigating aspects of the demographics of persons in the authority file, a few informal observations came forth. It seems that librarians are not necessarily creating records representing the breadth of persons having the potential to be searched. It also logically seems that librarians from well-funded, large academic institutions are creating the bulk of the records, potentially skewing the kinds of persons described. These observations should be investigated more thoroughly in subsequent research.

RDA is paving the way for a richer and more performant future of library systems in regards to the way that persons are included along with resources. Additional changes that are required for the attributes about authors to be included in metadata records will not be inconsequential. Because RDA is moving in the right direction, librarians will need to think creatively about how best to carry out these new tasks while balancing precision with volume. The promise of the semantic web must be investigated as systems and practices are reconsidered.

## Appendix I

### MERLIN Consortium libraries

- Missouri University of Science and Technology
- University of Missouri (MU)
- MU Law
- University of Missouri - Kansas City
- UMKC Law
- University of Missouri - St. Louis
- University of Missouri Archives and WHMC

## Acknowledgments

An enormous debt of gratitude is owed to Felicity Dykas, Head of Digital Services at MU's Ellis Library. The author would also like to thank Roy Tennant and the team at OCLC Research for making their MARC data available for reuse.

## References

- Bennett, R., Hengel-Dittrich, C., O'Neill, E., & Tillett, B. (2006). VIAF (Virtual International Authority File): Linking Die Deutsche Bibliothek and Library of Congress name authority files. In *World Library and Information Congress: 72nd IFLA General Conference and Council*. Retrieved from:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.5328&rep=rep1&type=pdf>
- Functional Requirements for Bibliographic Data (FRAD)*. (2008, 2013 rev.). Retrieved from  
[http://www.ifla.org/files/assets/cataloguing/frad/frad\\_2013.pdf](http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf)
- Functional Requirements for Bibliographic Records (FRBR): Final report*. (1997, 2009 rev.). International Federation of Library Associations and Institutions. Retrieved from  
[http://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf)
- Kroeger, A. (2013). The road to BIBFRAME. *Cataloging & Classification Quarterly*, 51(8), 873-890.
- MacEwan, A. Angjeli, A., & Gatenby, J. (2014). The International Standard Name Identifier (ISNI): The evolving future of name authority control. *Cataloging & Classification Quarterly*, 51(1-3): 55-71.
- MARC 21 encoding to accommodate new RDA elements 046 and 3XX in NARs and SARs. (2012, October). Retrieved from  
[http://loc.gov/aba/pcc/rda/PCC%20RDA%20guidelines/RDA%20in%20NARs-SARs\\_PCC.pdf](http://loc.gov/aba/pcc/rda/PCC%20RDA%20guidelines/RDA%20in%20NARs-SARs_PCC.pdf)
- MARC 21 format for authority data*. (1999, rev. September 2013). Library of Congress Network Development and MARC Standards Office. Update No. 1 (October 2001) through Update No. 17 (September 2013). Retrieved from  
<http://www.loc.gov/marc/authority/ecadhome.html>
- MARC Usage in WorldCat*. (2014). OCLC Research. Retrieved October 2013 and April 2014 from <http://experimental.worldcat.org/marcusage/>
- Miller, D. (2007). Basic values and the future of cataloging. *ALCTS*. Retrieved from  
<http://www.ala.org/alcts/ianda/bibcontrol/07futcat>
- PCC Day One for RDA Authority Records. (2011, Dec. 20; updated 2012, Mar. 23). Retrieved from

<http://www.loc.gov/aba/pcc/rda/PCC%20Day%20One/PCC%20Day%20One%20for%20RDA%20Authority%20Records.docx>.

- RDA Cataloger Training (2013, Feb. 18). *Indiana University Libraries*. Retrieved from <http://www.indiana.edu/~libslav/RemoteRDATraining/IULtraining2-authorityrecords.doc>
- RDA Toolkit: Resource description and access. (2014). American Library Association, Canadian Library Association, and CILIP: Chartered Institute of Library and Information Professionals. Retrieved from <http://www.rdatoolkit.org/>
- Report for PCC task group on the creation and function of name authorities in a non-MARC environment. (April 5, 2013). Retrieved from [http://www.loc.gov/aba/pcc/rda/RDA%20Task%20groups%20and%20charges/ReportPCC%20GonNameAuthInA\\_NonMARC\\_Environ\\_FinalReport.pdf](http://www.loc.gov/aba/pcc/rda/RDA%20Task%20groups%20and%20charges/ReportPCC%20GonNameAuthInA_NonMARC_Environ_FinalReport.pdf)
- Riemer, J. J. & Schreur, P. E. (2012, February 14). The future of undifferentiated personal name authority records and other implications for PCC authority work. End notes updated May 8, 2012. Retrieved from <http://www.loc.gov/aba/pcc/Undiff%20Personal%20NARs%20Discussion%20Paper%20March%202012.doc>
- Yee, M. M. (2005). FRBRization: A method for turning online public finding lists into online public catalogs. *Information Technology & Libraries*, 24(2), 77-95.
- Zeng, M. L., & Qin, J. (2009). *Metadata*. New York: Neal-Schuman.