

Brief on the Future of Non-MARC Authority

by the Indiana University PCC Non-MARC Authorities Issues Group

November 1, 2013

The Indiana University PCC Non-MARC Authorities Issues Group was formed in response to the Program for Cooperative Cataloging's (PCC) call for "... a well-articulated plan as to how [a non-MARC authorities] future could work based on [ideas from the PCC community]." Of particular interest were ideas that could serve as a bridge "between an ideal world and what [catalogers] can do now to create [that world]."

In this brief paper, the group proposes many changes that are **not** supported by current systems. Furthermore, the group did not dismiss ideas merely because implementation seemed resource-intensive. This was deliberate and in the spirit of the group's charge.

Premises

In discussing possible futures for non-MARC authority data, the group built ideas upon a foundation consisting of the following premises:

Linked data is here to stay. The corporate world, search engines, social networking sites, international governments, *and libraries abroad* have already embraced and implemented linked data. Linked data isn't a pie in the sky; it's a viable means of conducting business.

A move to linked data will require a complete remodeling of both bibliographic and authority data. While linked data developments, such as BIBFRAME, should inform future cataloging policy decisions, the group acknowledges that the BIBFRAME data model is, and will continue to be for at least the next few years, a work in progress. The group also acknowledges that the BIBFRAME model is many years away from being published as a formal schema.

Undifferentiated personal name authorities pose a significant barrier to a linked data implementation of library data. The group applauds PCC efforts to engage the community in the discussion of possible solutions and looks forward to implementing a workable solution as quickly as possible. The group supports halting the creation of new undifferentiated records.

End-user interface designers and cataloging policy makers are locked into a chicken-and-egg conundrum. Discovery systems aren't built to support faceted browsing for enriched authority data; therefore, cataloging policy doesn't encourage enrichment of authority data. The vast majority of LCNAF records lack enriched, machine-readable data; therefore, discovery layer providers don't prioritize development of new faceted browsing functionality. The group isn't sure what is to be done about this problem; however, it seems as though the onus rests on the library community to ensure that data is available in a rich, machine processable form.

Current cataloging interfaces are outmoded and should be retired as soon as possible in favor of sensibly designed metadata curation interfaces. Very few communities expect professionals to manipulate data in a native serialization format (e.g., MARC21). In a future in which library data is to be valued and widely used in an open linked environment, the ability to provide authoritative, timely data is of chief importance. Cataloging and metadata professionals cannot afford to risk

inaccuracies and validation errors for the sake of working natively in MARC, XML, Turtle, JSON, N-triples, etc.

Reimagining Discovery

While no one serving on the group is a developer by trade, we enjoyed imagining a future in which discovery systems leverage authority data to power faceted searching and browsing. Authority data enriched with information such as chronological time period, geographic place, occupation, activity, and associated group(s) would support queries such as:

- German illustrators active in the nineteenth century
- Female funk musicians active in Dayton, OH

A linked data environment might make efficient use of authority data in conjunction with bibliographic data. While difficult for everyone in the group to imagine a recordless data world, the group did imagine “landing pages” for personal, corporate, or family entities (see also WorldCat Identities) that contain links to available resources.

Reimagining data in this way solves a common problem in describing digitized still images. Catalogers are often frustrated that they are not able to add certain access points that are known to be of interest to users because those access points aren’t the subject of the images. For instance, an image might be of a group of people taken at a social event. All of the people pictured may happen to be Capitol Records recording artists; however, adding the authorized access point to the record as a subject or contributor is inappropriate. If catalogers added “Capitol Records, Inc.” to the associated group field in the authority records of each of the artists pictured, that relationship may be searchable in a linked data environment.

Policy Paths

In discussing possible directions of future authority work, we began to see two distinct trends. Both trends add value to the authority file in a machine processable manner and both trends might support faceted searching and browsing in order to enhance users’ discovery experience:

- Policies that promote **robust, granular data encoding practices** through the consistent population of fields/elements with machine readable values whenever these values are readily available
- Policies that focus on **adding meaningful linkages** to LCNAF, non-LCNAF authorities, and to related resources

We envisioned these trends as two congruous avenues that may guide the nature of authority work in the future.

Authority Work as a Granular Data Enrichment Endeavor

Historically, disambiguating information about a name was included in subfield b of MARC Authority 670 citations. However, the *MARC21 Format for Authority Data* provides numerous fields for expressing data typically captured in 670 fields in a more machine processable way. The example in [Appendix A](#) demonstrates how an existing MARC authority record might have been encoded for more data granularity. Semantically meaningful elements such as occupation and birthplace become data points that may be leveraged for faceted discovery.

With the advent of linked library data entering non-English speaking, non-Latin script environments, it may be necessary to reevaluate how we record parallel vernacular fields. Currently, vernacular appears in both bibliographic and authority records. The group wondered how catalogers might code the language and script of data appearing in a field (for authorized access points in particular but perhaps for variants as well). With language and script identified, users could be served different languages/scripts corresponding to a web browser setting, an IP address, or a toggle on the discovery interface.

A successful transformation of library data to linked data (BIBFRAME or otherwise) will hinge upon how relationships between FRBR Group 1 and Group 2 entities have been encoded. The use of relator terms or codes in the 1XX and 7XX of bibliographic records will be useful in defining the predicate of triple statements. It is unclear to the group how authority data might help make relationships between resources and creators/contributors more clear; however, the document [On BIBFRAME Authority](#) proposes a few interesting options on creator/contributor roles.

Authority Work as a Linking Endeavor

The group unanimously supports adding non-LCNAF standard identifiers to LCNAF records and will watch with great interest for the result of the PCC Policy Committee discussion concerning the addition of ISNIs to LCNAF records. The group also saw value in adding VIAF identifiers to LCNAF records, particularly when a VIAF record is cited in a 67X field of an authority record anyway.

The group proposes providing linkages to other LCNAF records in the See Also From Tracing fields (5XX). Pointing directly to another record by supplying an identifier in the subfield 0 of a 5XX for an early/late form of name, pseudonym, etc. will aid in machine processing and discovery. The group also wonders if the creation of a subfield 2 for 5XX fields would aid in linked data transformations to come. Currently, the MARC Organization Codes/Standard Identifier Source Codes are contained within parenthesis immediately preceding the identifier in the subfield 0. This makes the 5XX subfield 0 field difficult to check for validity or manipulate for discovery.

Linking endeavors may also extend to how citations are encoded in authority records. By including citations in the 672 or 673 fields, the citations themselves become links to resources associated with the named person, corporation, or family. Examples of how 672 fields might be used are shown in [Appendix A](#). What the group was not clear on was whether it was a wise to abandon the 670 altogether in favor of 672 or 673 fields. Taking authority record to bibliographic record linking one step further, the group noticed that the German national library is already providing authority record identifiers in the access points of the bibliographic records they create. The group found this to be compelling and is interested in a further study of this practice.

Conclusion

While the transition from MARC won't be easy, there are encoding practices we can adopt now that will make the transition easier. As the group worked through various issues, it became glaringly obvious that we'll need some very slick tools to help us curate metadata in a linked data future. GUI-style autosuggest/autocomplete fields, dropdown lists for controlled values, and robust field validation were all items that appeared on our metadata tool wish list. Having the proper tools enables catalogers to focus on the intellectual activities associated with metadata creation, curation, and strategy that cannot

be automated. In addition to forward-thinking policies and standards, our community is in need of a strong advocate for metadata curation systems.

Indiana University PCC Non-MARC Authorities Issues Group

James Castrataro

Carl Horne

Jennifer A. Liss (chair)

Ronda L. Sewald

Rachel Wheeler

Appendix A: Example

In Figure 1, rich data exists in free-text MARC 670 fields. This data, meant in part to disambiguate the author, may also be leveraged for discovery.

ARN		5176122					
<u>Rec stat</u>	n	Entered	20000303	Replaced	20000314033712.0		
<u>Type</u>	z	<u>Upd status</u>	a	<u>Enc lvl</u>	n	<u>Source</u>	c
<u>Roman</u>	■	<u>Ref status</u>	a	<u>Mod rec</u>		<u>Name use</u>	a
<u>Govt agn</u>	■	<u>Auth status</u>	a	<u>Subj</u>	a	<u>Subj use</u>	a
<u>Series</u>	n	<u>Auth/ref</u>	a	<u>Geo subd</u>	n	<u>Ser use</u>	b
<u>Ser num</u>	n	<u>Name</u>	a	<u>Subdiv tp</u>	n	<u>Rules</u>	c

010		nr 00007317
040		MWA #b eng #c MWA
100	1	Hardy, Jeremiah Pearson, #d 1800-1887
400	1	Hardy, J. P. #q (Jeremiah Pearson), #d 1800-1887
670		Williamson, W.D. The history of the State of Maine, 1839: #b v. 2, front. (J.P. Hardy pinxt.)
670		Groce, G.C. The N.Y. Hist. Soc. dict. of artists in Amer., 1957 #b (Hardy, Jeremiah Pearson, portrait, miniature, genre, animal, and still-life painter; born Pelham, N.H., October 22, 1800; moved with his parents to Hampden, Me. in 1811; went to Boston to study ca. 1822; returned to Hampden ca. 1828, living there and in Bangor until his death, February 9, 1887)

Figure 1. The name authority record as it existed on October 28, 2013.

Figure 2 shows the same name authority record data encoded in semantically meaningful MARC fields. Where possible, the fields are populated with values from controlled vocabularies. In addition, each 670 is translated into a 672 field, which allows for the addition of bibliographic record identifiers in the subfield w. Though the 672 field is not valid for use in OCLC, the adoption of this field will aid in the creation of meaningful linkages between creators/contributors and resources via the bibliographic record control numbers (subfield w).

ARN 5176122						
Save File 45						
Rec stat	n	Entered	20000303	Replaced	20000314033712.0	
Type	z	Upd status	a	Enc lvl	n	Source c
Roman	■	Ref status	a	Mod rec		Name use a
Govt agn	■	Auth status	a	Subj	a	Subj use a
Series	n	Auth/ref	a	Geo subd	n	Ser use b
Ser num	n	Name	a	Subdiv tp	n	Rules z
010		nr	00007317			
040		MWA #b eng #e rda #c MWA				
024	7	0000000067036885 #2 isni				
024	7	6148211 #2 viaf				
046		#f 18001022 #g 18870209				
100	1	Hardy, Jeremiah Pearson, #d 1800-1887				
370		Pelham (N.H.) #b Bangor (Me.) #2 naf				
374		Painters #2 lcsb				
375		male				
400	1	Hardy, J. P. #q (Jeremiah Pearson), #d 1800-1887				
670		Williamson, W.D. The history of the State of Maine, 1839: #b v. 2, front. (J.P. Hardy pinxt.)				
670		Groce, G.C. The N.Y. Hist. Soc. dict. of artists in Amer., 1957 #b (Hardy, Jeremiah Pearson, portrait, miniature, genre, animal, and still-life painter; born Pelham, N.H., October 22, 1800; moved with his parents to Hampden, Me. in 1811; went to Boston to study ca. 1822; returned to Hampden ca. 1828, living there and in Bangor until his death, February 9, 1887)				
672	4	The history of the state of Maine #f 1839 #w (OCoLC)647142 #w (DLC) 26002506				
672	4	The New-York Historical Society's dictionary of artists in America #f 1957 #w (OCoLC)510658 #w (DLC) 57006338				

Figure 2. The name authority record edited for data granularity.