Author Identifier Analysis: Name Authority Control in Two Institutional Repositories
Poster

Marina Morgan
Florida Southern College, United States
mmorgan@flsouthern.edu

Naomi Eichenlaub
Ryerson University, Canada
neichenl@ryerson.ca

Keywords: persistent identifiers; researcher identifiers; public data; name disambiguation; metadata assessment; LCNAF; Scopus Author Identifier; ORCID; VIAF; ISNI.

Abstract

The aim of this poster is to analyze name authority control in two Institutional Repositories (IRs) to determine the extent to which faculty researchers are represented in researcher identifier databases. A purposive sample of 50 faculty authors from Florida Southern College (FSC) and Ryerson University (RU) were compared against five different authority databases: Library of Congress Name Authority File (LCNAF), Scopus, Open Researcher and Contributor ID (ORCID), Virtual International Authority File (VIAF), and International Standard Name Identifier (ISNI). We first analyzed the results locally, then compared them between the two institutions. Looking at both institutions together, the representation was closest in two and three databases (24% and 22% at FSC and 18% and 20% at RU). This has implications for enhancing local authority data by linking to external identifier authority data to augment institutional repository metadata.

Background

Florida Southern College (FSC) is the oldest private college in the state of Florida with 2,500 FTE and 139 full-time faculty members. The college offers undergraduate, graduate, and post-graduate programs in various disciplines. Ryerson University (RU) is located in the heart of downtown Toronto, Canada. With an FTE of 35,000 across more than 100 undergraduate and graduate programs, Ryerson has close to 900 full-time faculty members including 20 Canada Research Chairs. Both institutions have IRs situated within their libraries that contain faculty research.

Metadata quality in IRs has typically been a challenge owing to a number of circumstances. Designed initially for self-submission of faculty research, metadata workflows in IRs are often lacking the functionality to ensure best practices, particularly in the area of authority control (Salo, D., 2009). Moreover, metadata arrives in repositories from a variety of sources including batch ingests, harvesting, and deposits by staff, students and researchers, making it difficult to enforce consistency (Chapman, J.W., Reynolds, D., & Shreeves, S. A., 2009).

Authority control is the process of identifying headings as access points and ensuring each access point is unique by disambiguating variant headings. Traditionally, authority files were created and maintained in-house in libraries and contained unique character strings to identify authorized subject and name headings. In an online environment, the need for authority control has extended beyond the library catalog to IRs and journal article databases. Name authority control in IRs continues to be difficult to manage, especially in larger repositories, and can impact discovery and retrieval as well as attribution. Furthermore, once variations in name headings are introduced to the repository, they most often must be corrected manually, a very time-consuming process (Salo, 2009).

Introduction

In this digital age, computers require more direction than humans in terms of name disambiguation (Van der Graaf, M. & Waaijers, L., 2014). Therefore authority control, and name authority control in particular, becomes even more critical. As such, a number of web-based author identifier initiatives have emerged such as Scopus Author Identifier, LCNAF, VIAF, ISNI, and ORCID. The online environment of IRs offers an opportunity to harness persistent identifier and linked data initiatives and evolve beyond traditional bibliographic data silos to embrace open author identifiers (Chapman, J. W., Reynolds, D., & Shreeves, S.A., 2009), especially as academic institutions place increasing importance on tracking research outputs. A 2013 Program for Cooperative Cataloguing report acknowledged the need to explore "the role of name authorities and identity data generally in a post-MARC, linked data environment" (PCC, p. 4)

To this end, this study looks at name authority control in two separate IRs in the broader context of research identifiers to determine the extent to which faculty are represented in author identifier databases. It builds on previous studies looking at researcher representation in name authority databases (Sandberg, J. & Jin, Q., 2016; Waugh, L., Tarver, H., Phillips, M.E., 2014) in an effort to understand the scope of name authority representation beyond what is currently included in author name metadata in the IR software platforms at Florida Southern College Roux Library and Ryerson University Library and Archives. As an outcome of this study, the authors hope to explore external options available to enhance name authority in their IRs.

Methodology

The authors selected a purposive sample of faculty researchers in their IRs with the goal of surveying the representation of identifiers available for each researcher. Fifty researchers with content from each IR were selected. The careful selection focused on the researchers with enough content to warrant having identifiers in external authority databases and requiring name authority control since they have multiple entries in the IR. The authors then compared this sample of faculty researchers from each IR across five databases: Library of Congress Name Authority File, Scopus, ORCID, VIAF, and ISNI.

For each institution, we created a spreadsheet containing an entry for each author and separate tabs for LCNAF, Scopus Author Identifier, ORCID, VIAF, and ISNI databases. We then manually searched for each name in each database and added the identifiers to the spreadsheet. We did not count ORCID records that were not public, because it was not possible to disambiguate or confirm researcher identify when no public information was available. Once all the data was gathered, we analyzed the coverage of authors represented across databases for both institutions. To calculate the coverage we used the following formula:

$$f = \frac{\sum r}{y}$$

FIG. 1. Formula used to calculate the authors' representation in five databases.

In this formula, r equals the numbers of author representations, and y equals sample size (50).

Results

Reviewing the FSC results, we found that all but 3 authors were represented in a database, with an overall of 86% represented in Scopus, 36% in VIAF, 30% in LCNAF, 30% in ISNI, and 12% in ORCID. The RU results indicate the researcher sample had the strongest author representation in Scopus with 96%, followed by 76% in ISNI, 62% in ORCID, 62% in VIAF and 40% in LCNAF. All but one of the RU researchers were represented in at least one database.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2018

TABLE 1: Percentage of author names represented in each database

	LCNAF	Scopus	ORCID	VIAF	ISNI
Florida Southern College	30%	86%	12%	36%	30%
Ryerson University	40%	96%	62%	62%	76%

Additionally, analyzing the database representation at FSC, we determined that 6% of the authors were not represented in any database, 36% were represented in one database, 24% represented in two databases, 22% in three databases, 12% in four databases, and there was no author representation in all five databases. At RU, the results indicate that 2% of authors were not found in any of the databases, 6% were represented in one database, 18% in two databases, 20% in three databases, 36% in four databases and 18% in five databases. A complete breakdown of authors' representation is found in table 2.

TABLE 2: Representation of authors in five identifier databases for each institution

	0 DB		1 DB		2 DB		3 DB		4 DB		5 DB	
	FSC	RU	FSC	RU	FSC	RU	FSC	RU	FSC	RU	FSC	RU
No database	3	1										
LCNAF			0	0								
Scopus			18	3								
ORCID			0	0								
VIAF			0	0								
ISNI			0	0								
LCNAF and Scopus					0	0						
LCNAF and ORCID					1	0						
LCNAF and VIAF					0	0						
LCNAF and ISNI					0	0						
Scopus and ORCID					4	5						
Scopus and VIAF					3	1						
Scopus and ISNI					5	3						
ORCID and VIAF					0	0						
ORCID and ISNI					0	0						
VIAF and ISNI					0	0						
LCNAF, Scopus, ORCID							0	0				
LCNAF, Scopus, VIAF							6	0				
LCNAF, Scopus, ISNI							0	0				
LCNAF, ORCID, VIAF							0	0				
LCNAF, ORCID, ISNI							0	0				
Scopus, ORCID, VIAF							0	1				
Scopus, ORCID, ISNI							1	7				
Scopus, VIAF, ISNI							0	2				
LCNAF, VIAF, ISNI							4	0				
ORCID, VIAF, ISNI							0	0				
LCNAF, Scopus, ORCID, VIAF									1	1		
LCNAF, Scopus, ORCID, ISNI									0	0		
LCNAF, Scopus, VIAF, ISNI									5	9		
LCNAF, ORCID, VIAF, ISNI									0	1		
Scopus, ORCID, VIAF, ISNI									0	7		
LCNAF, Scopus, ORCID, VIAF, ISNI											0	9
TOTAL %	6%	2%	36%	6%	24%	18%	22%	20%	12%	36%	0%	18%

The overall state of representation of author identifiers in this study shows that the highest representation of authors at FSC was found in one, two and three databases. At RU, the highest representations of authors was found in four databases, three databases, and two and five databases (tied), respectively. Looking at both institutions together, the representation was closest in two and three databases (24% and 22% at FSC and 18% and 20% at RU). Since the sample selected was purposive, a completely random sample of researchers may have yielded different results for each institution. Moreover, the findings show that while the majority of database permutation results are

115

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2018

comparable between the two institutions, the difference in results from one database (Scopus), four databases (Scopus, ORCID, VIAF, and ISNI), and five databases (LCNAF, Scopus, ORCID, VIAF, and ISNI) is considerable.

As Sandberg & Jin justified, different disciplines have different representation results (Sandberg, J. & Jin, Q., 2016). Science faculty may score lower on book-centric databases such as LCNAF, and higher in Scopus with content driven from serial publications and conference series.

Future Work

Authority control in IRs must become more of a focus, especially as institutions place a higher priority on tracking researcher outputs. To reduce data silos, name authority control in IRs is an opportunity to harness existing external author identifiers such as ORCID, Scopus Author Identifier, ISNI, VIAF, and LCNAF but we need software and platforms that help us take advantage of these, for example through bi-directional updates. An example of this could be ORCID integration with IRs, which serves both to increase the number of researchers with ORCID iDs as well as to match ORCID iDs with institutional affiliation.

In a global research environment, IR managers must prioritize name authority work and advocate for increased system functionality to help manage it, not only IR functionality such as batch edits, global updates and auto-complete (Salo, D., 2009), which do little to enforce authority control, but also to take advantage of linked data resources to maximize interoperability and showcase researcher output.

Conclusions

Authority control in IRs can no longer be limited to manual, in-house cleanup of name headings, but most also leverage the authority data that currently exists in external author identifier sources such as ORCID, Scopus, ISNI, VIAF, and LCNAF, to collectively work together to confirm name identity. The authors determined via a purposive sample that in analyzing the author identifier representation of 100 researchers across two institutions, the results are closest for the two institutions in 2 and 3 databases: 24% and 22% for FSC and 18% and 20% for RU. As a result of these findings, the authors will look at ways of leveraging external name authorities across various external platforms to enhance name authority control in their IRs.

References

Chapman, John W., David Reynolds, and Sarah A. Shreeves. (2009). Repository metadata: Approaches and challenges. Cataloging & Classification Quarterly, 47(3-4), 309-325. 10.1080/01639370902735020

Report for PCC Task Group on the Creation and Function of Name Authorities in a Non-MARC Environment. (2013). http://www.loc.gov/aba/pcc/rda/RDA%20Task%20groups%20and%20charges/ReportPCCTGonNameAuthInA_NonMARC_Environ_FinalReport.pdf

Salo, Dorothea. (2009). Name authority control in institutional repositories. Cataloging & Classification Quarterly, 47(3-4), 249-261. 10.1080/01639370902737232

Sandberg, Jane and Qiang Jin. (2016). How should catalogers provide authority control for journal article authors? Name identifiers in the linked data world. Cataloging & Classification Quarterly, 54(8), 537-552. 10.1080/01639374.2016.1238429

Van der Graaf, Mauritz and Leo Waaijers. (2014). Authority files: breaking out of the library silo to become signposts for research information. http://repository.jisc.ac.uk/6224/1/Authority_files_-_Breaking_out_of_the_library_silo.pdf

Waugh, Laura, Hannah Tarver, and Mark Edward Phillips. (2014). Introducing name authority into an ETD collection. Library management 35 (4 / 5). 10.1108/LM-08-2013-0074