

Local name authorities and linked data: The landscape

By: [Anna R. Craft](#)

Craft, A. R. (2019). Local name authorities and linked data: The landscape. *Serials Review*, 45(1). 61-65. <https://doi.org/10.1080/00987913.2019.1613736>

This is an Accepted Manuscript of an article published by Taylor & Francis in *Serials Review* on 23 May 2019, available online:

<http://www.tandfonline.com/10.1080/00987913.2019.1613736>.

*****© 2019 The Author. Reprinted with permission. No further reproduction is authorized without written permission from Taylor & Francis. This version of the document is not the version of record. *****

Abstract:

Name authority control—the process of selecting and maintaining a set of headings or access points for named entities—brings consistency to names in online systems and can help users locate needed items and identify related materials. Linked data provides the opportunity to extend name authority data sets and make them available to wider audiences. This column seeks to provide an overview of the landscape of local name authority projects using linked data, along with practices, challenges, and other considerations in this area.

Keywords: digital libraries | linked data | metadata | name authorities | special collections

Article:

Introduction

Name authorities work currently exists at a complicated intersection. Many libraries are grappling with constraints on staff time and expertise available to devote to name authority work. Simultaneously, increasing numbers of resources—primary source materials in digital collections and scholarship in institutional repositories, among others—are being made available online and are not always associated with established name authority files. Linked data—a means to potentially make such names available and usable in a wider context—remains a topic of interest and importance for many librarians. This column seeks to provide an overview of the landscape of local name authority projects using linked data, along with practices, challenges, and other considerations in this area.

Name authority control in library work

Name authority control—the process of selecting and maintaining a set of headings or access points for named entities—is recognized as a best practice across many library communities, as it supports users in finding and identifying resources (International Federation of Library Associations, 2017). Neatrour and Myntti (2017) delineate ways that name authorities contribute

value to library catalogs. Use of authorized name headings makes it easier to search for known names, provides predictability in searching and consistency in records, organizes and collocates information, provides efficiencies for creators of metadata, facilitates data maintenance over time, reduces errors in name data sets, differentiates similar names, and provides context for names. Many library personnel who work in cataloging and related fields already have expertise in authority control practices and principles.

The predominant shared name authority file in use in the library community in the United States is the Library of Congress Name Authority File (LCNAF)—also known as the NACO Authority File—which holds over 8 million records (Library of Congress, n.d.c). NACO (the Name Authority Cooperative Program) oversees this data file, which includes records for personal, corporate, and jurisdictional names, as well as uniform titles and series titles. To maintain the integrity of the authority file, NACO provides member institutions with training on the group's standards and guidelines (Library of Congress, n.d.a). Records in the LCNAF are available not only to NACO members but are freely accessible online through a web interface provided by the Library of Congress. These records are also available as linked data (Library of Congress Authorities, 2018). NACO membership is open to any institution. While there is no initial or annual fee to join, member institutions must pay for staff training and costs associated with hosting a trainer. Membership in NACO also requires that institutions meet annual goals related to the creation or editing of authority records unless they are part of a geographic or subject-based NACO funnel project that brings together multiple libraries (Library of Congress, n.d.b).

But many libraries and other cultural heritage institutions are not NACO members, and many institutions—NACO and non-NACO—hold name data sets that fully or in part are not represented in LCNAF or other extant authority files. These data are often associated with local archival collections or other primary resource materials that may be held in digital collections or institutional repositories. In some cases, these projects lead libraries to the development of local name thesauri, which may include names of persons, buildings, university departments, schools, churches, corporations, hospitals, and other entities (Carlson & Seely, 2017; Dragon, 2009). In some cases, these local thesauri may contain significant amounts of information that is not available in other authority files (Carlson & Seely, 2017).

While local name thesauri may serve the specific needs of an individual institution, they limit the ability of researchers to find associated materials across multiple institutions. And in cases where multiple libraries hold materials related to the same entities, local practices may cause duplicative work across those libraries. Recent years have seen increasing interest and action around the exploration of sharing local name authorities using linked data, as the existence of these local name data sets offers opportunities for data extension and reuse across institutions (Chiat Naun et al., 2018).

The National Strategy for Shareable Local Name Authorities (SLNA) National Forum

The importance of exploration and development in this area has already been recognized by leaders in the field. Cornell University Library, partnering with the Library of Congress, OCLC, the Program for Cooperative Cataloging, the ORCID organization, the Coalition for Networked Information, the Social Networks and Archival Context Cooperative (SNAC), the BIBFLOW

project, Stanford University Library, and Harvard University Library, received a 2016 Institute of Museum and Library Services (IMLS) National Leadership Grant to hold a national forum on issues concerning local name authorities. This project, the National Strategy for Shareable Local Name Authorities National Forum, produced a white paper and a draft reference model, both of which are freely available online (Cornell University, n.d.). The white paper provides detailed information on the project and its outcomes, as well as technical details, a review of existing name authority aggregators, and case studies of workflows from data creators and data aggregators, some of which are using linked data protocols and others that are not (Chiat Naun et al., 2018). The SLNA Reference Model “provides a framework for understanding the relationships among the entities managing and sharing authority data in a global ecosystem” (Wang, 2018). In addition to the white paper and reference model, the IMLS Shareable Authorities Forum site provides further information about the project, including meeting summaries, a background paper, and an inventory of related work (Chiat Naun et al., 2018).

Regional, local, and subject-based name authority projects using linked data

Several institutions and groups within the profession are already sharing local name authorities using linked data. This column looks at four of these projects. The Western Name Authority File and the Colorado Local Authorities Project bring together multiple institutions with regional name authority data. The Linked Jazz Name Directory is subject-specific within the realm of jazz artist names. The Organization Linked Data Project at North Carolina State University (NCSU) focuses on organization names from one institution. Following is a brief description of each.

Western Name Authority File (WNAF)

Like the SLNA National Forum, the Western Name Authority File (WNAF) is an IMLS-funded, multi-institution project. WNAF includes personal and corporate names drawn from partner institutions: the University of Utah, Utah State University, Brigham Young University, the University of Denver, the University of Oregon, the University of Nevada–Reno, the Utah State Archives, and the Utah Department of Heritage and Arts. As of this writing, the file contains 50,153 personal names and 10,562 corporate names and is searchable through a dedicated website where the name vocabulary is published using Omeka S (<https://omeka.org/s/>) (J. Willard Marriott Library, n.d.). Separate from the searchable vocabulary is a project website with fuller information about participants and project workflows. The project website also provides a tool kit and steps for others pursuing this type of effort (Western Name Authority Project, n.d.).

Colorado Local Authorities Project (CoLA)

The Colorado Local Authorities Project (CoLA) “aims to provide publishing and access services for biographical and historical information about people, organizations, and families of interest to cultural institutions in Colorado and Wyoming” (Colorado Local Authorities Project, n.d.). In a 2015 article, Crowe and Clair discuss planning and technical details for this project. Additional documentation is available via GitHub (Colorado Local Authorities Project, n.d.).

Linked Jazz Name Directory

The Linked Jazz Name Directory was created to support the Linked Jazz project (Pattuelli, 2012). This effort brings together content from the Hamilton College Jazz Archive; Rutgers Institute for Jazz Studies Archives; Smithsonian Jazz Oral Histories; The University of California, Los Angeles's Central Avenue Sounds Series; and the University of Michigan's Nathaniel C. Standifer Video Archive of Oral History (Linked Jazz, n.d.). Pattuelli (2012) explained some of the reasons for the creation of the Linked Jazz Name Directory: Existing library authority files, such as LCNAF and VIAF (Virtual International Authority File), have limited coverage of names within the domain of jazz music and do not offer the ability to filter out the names of jazz artists; and while domain-specific data sets exist, they are not open or available as linked data. Instead of relying on an internal or organizational data set, the project used DBpedia (<https://wiki.dbpedia.org/>) to construct its data set (Pattuelli, 2012).

North Carolina State University's (NCSU) Organization Linked Data (ONLD)

NCSU's Organization Linked Data (ONLD) Project provides an example of a single-institution project with a very targeted focus. Built using name data from their Organization Name Authority tool, which incorporates names used in acquisitions and licensing, Hanson (2014) emphasized that the ONLD data set is acquisitions oriented and not a traditional bibliographic name authority file. Data are available via Resource Description Framework Extensible Markup Language (RDF-XML), N-Triples, N3/Turtle, and JavaScript Object Notation Linked Data (JSON-LD) (NCSU Libraries, 2018).

Name authority linked data workflows and practices

For those who are embarking on or considering undertaking linked data name authority projects, the literature provides guidelines, best practices, and workflows to consider. This column does not attempt to supply in-depth details; readers should refer to source material for such. One potential starting point for background reading is the W3C Working Group Note ‘Best Practices for Publishing Linked Data,’ edited by Hyland, Atemezing, and Villazón-Terrazas (2014). The paper describes 10 best practices for developing and delivering linked data:

1. Prepare stakeholders
2. Select a data set
3. Model the data
4. Specify an appropriate license
5. Use Good URIs for linked data
6. Use standard vocabularies
7. Convert data
8. Provide machine access to data
9. Announce new data sets
10. Recognize the social contract

While the authors' focus is government data, these steps are applicable in other contexts.

The SLNA white paper (Chiat Naun et al., 2018) and the WNAF tool kit (Western Name Authority Project, n.d.) are also meant to support those considering or developing linked data

name authority projects. In addition to providing case studies and technical details, the SLNA white paper offers consideration of a Minimum Viable Product Specification (MVP) approach to name data, with respect to data, technical, and business requirements. It also puts forward considerations in the realms of data provider obligations, workflows, and data reconciliation (Chiat Naun et al., 2018). The WNAF tool kit breaks down the process, providing five steps toward developing a linked data name project: gathering and evaluating metadata, making a vocabulary publishing plan; reconciling, reviewing, and cleaning metadata; publishing the vocabulary; and sharing, assessing, and evaluating the project. Pages devoted to project steps include examples, questions to consider, use cases and case studies, and information about useful tools (Western Name Authority Project, n.d.).

Further reading might include Hanson's 2014 article, "A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project." Hanson divided the ONLD project into five phases: modeling, cleanup, enhancing, converting, and publishing. In addition to describing these stages, he provides technical details and suggests project management practices that may help the project run more smoothly. He advocates for considering what use cases exist for the data set, developing a shared vision among the project team, and taking a realistic look at the size and complexity of the source data in relation to the time and staffing resources that are available for the project. He also reiterates the importance of data cleanup (Hanson, 2014).

Challenges and other considerations

The literature also lays out challenges that can impact the development and creation of linked data name sets. Crowe and Clair (2015) point out that some name data sets exist only in somewhat closed formats—such as PDFs (Portable Document Format) and spreadsheets—which may add to the complexity of accessing and working with the data. Data quality can present a significant challenge, especially when attempting to disambiguate names and match names to entries in existing vocabularies (Chiat Naun et al., 2018). As Dragon (2009) notes, scalability can be a critical concern for local name authority projects even without the inclusion of linked data, as name authority work can be extremely time-consuming. To take advantage of linked data capabilities, locally created name data sets must be linked with existing data sets, such as LCNAF and VIAF, and projects that rely on manual matching of local records with existing authorities will likely not be scalable (Hanson, 2014). For published linked data projects, there are also questions of continuing updates and maintenance of the data set, as well as ongoing assessment. To learn more about the applications of their data sets, owners and creators of linked data sets may want to consider allowing for or encouraging communications from users of those data sets (Hanson, 2014).

Conclusion

The implementation of linked data for name authorities is a natural fit in the library context. Many library personnel already have significant experience creating and disambiguating name authorities, and important local name authority data is held at institutions of all sizes. Exposing and connecting these data create opportunities for discovery as well as the potential to draw valuable connections between resources and collections. Linked data offer the possibility of

enabling the extension and reuse of local name authority data while also supporting consistency in the creation of metadata, potentially helping institutions better manage, maintain, and understand their collections while also assisting users in identifying needed materials and collocating resources across institutions. As understanding of and support for linked data grow in the profession, it is likely that more institutions will look to undertake linked data projects. Initiatives described here, such as the SLNA National Forum and local and regional projects, are helping spread awareness, guidelines, and tools for others.

References

- Carlson, S., & Seely, A. (2017). Using OpenRefine's reconciliation to validate local authority headings. *Cataloging & Classification Quarterly*, 55(1), 1–11.
doi:10.1080/01639374.2016.1245693
- Chiat Naun, C., Kovari, J., Casalini, M., Cluff, C., Durocher, M., Folsom, S., ... Warner, S. (2018). *National Strategy for Shareable Local Name Authorities National Forum: White Paper*. Retrieved from <http://hdl.handle.net/1813/56343>
- Colorado Local Authorities Project. (n.d.). *About the project*. Retrieved from <https://github.com/jackflaps/cola>
- Cornell University. (n.d.). *IMLS Shareable Authorities Forum*. Retrieved from <https://confluence.cornell.edu/display/sharedauth/IMLS+Shareable+Authorities+Forum+Home>
- Crowe, K., & Clair, K. (2015). Developing a tool for publishing linked local authority data. *Journal of Library Metadata*, 15(3–4), 227–240.
doi:10.1080/19386389.2015.1099993
- Dragon, P. M. (2009). Name authority control in local digitization projects and the Eastern North Carolina Postcard Collection. *Library Resources & Technical Services*, 53(3), 185–196.
doi:10.5860/lrts.53n3.185
- Hanson, E. M. (2014). A beginner's guide to creating library linked data: Lessons from NCSU's Organization Name Linked Data Project. *Serials Review*, 40(4), 251–258.
doi:10.1080/00987913.2014.975887
- Hyland, B., Atemezing, G., & Villazón-Terrazas, B. (2014). *Best practices for publishing linked data*. Retrieved from <http://www.w3.org/TR/ld-bp/>
- International Federation of Library Associations (2017). *Best practice for national bibliographic agencies in a digital age*. Retrieved from <https://www.ifla.org/node/7858>
- Library of Congress. (n.d.a). *About NACO*. Retrieved from <https://www.loc.gov/aba/pcc/naco/about.html>
- Library of Congress. (n.d.b). *Frequently asked questions about joining the NACO program*. Retrieved from <https://www.loc.gov/aba/pcc/naco/nacoprogfaq.html>
- Library of Congress. (n.d.c). *Library of Congress names*. Retrieved from <http://id.loc.gov/authorities/names.html>

- Library of Congress Authorities (2018). *Frequently asked questions*. Retrieved from <https://authorities.loc.gov/help/auth-faq.htm#2>
- Linked Jazz. (n.d.). *Linked Jazz*. Retrieved from <https://linkedjazz.org/about-the-project/>
- Neatrour, A., & Myntti, J. (2017). *The Western Name Authority File: Improving access for digital collections and archives [Google slides]*. Retrieved from Western Name Authority File Project
Website: <https://docs.google.com/presentation/d/1Bq8tK2UwCLZAdFzr95FHrEdmcDEcOMVmPfVIqBsg7g0/edit?usp=sharing>
- NCSU Libraries (2018). *Organization name linked data*. Retrieved from <https://www.lib.ncsu.edu/ld/onld/>
- Pattuelli, M. C. (2012). Personal name vocabularies as linked open data: A case study of jazz artist names. *Journal of Information Science*, 38(6), 558–565.
doi:10.1177/0165551512455989
- Wang, J. (2018). *Shareable local name authority reference model (draft)*. Retrieved from <http://hdl.handle.net/1813/56582>
- Western Name Authority Project. (n.d.). *Western Name Authority File project: Toolkit*. Retrieved from <https://sites.google.com/site/westernnameauthorityfile/toolkit>
- Willard Marriott Library, J. (n.d.). *Western Name Authority File*. Retrieved from <https://exhibits.lib.utah.edu/s/wnaf/page/welcome>