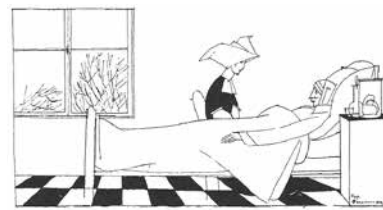


Current Trends and Opportunities in Health Sciences Library Metadata

by **Joelen Pastva** (Head, Collection Management and Metadata Services, Galter Health Sciences Library & Learning Center) <joelen.pastva@northwestern.edu>

and **Tony Olson** (Cataloging Librarian, Galter Health Sciences Library & Learning Center) <ajolson@northwestern.edu>



The efficient management and discovery of library resources have always been of concern to catalogers and metadata librarians working in health sciences libraries, but the past several years have changed many of the systems and workflows employed to do so and created opportunities for applying existing skillsets to new challenges. This article examines how the dominance of electronic resources in the health sciences has shifted cataloging workflows and priorities. It also examines efforts currently underway to bring cataloging practices and standards into better alignment with modern web standards. Finally, it identifies new roles for metadata librarians and catalogers that have emerged in recent years in health sciences libraries that leverage existing skills and library metadata for new initiatives and collaborative opportunities that reach beyond the borders of traditional technical services activities.

Workflows

The growing footprint of electronic resources in library collections has necessitated changes in the way those collections are managed by catalogers and metadata librarians. A 2017 *Library Journal* study revealed that 88% of library collections spending in North America is toward electronic-only or electronic/print combination products.¹ Health sciences collections tend more toward journals than monographs, and electronic formats have had an especially large impact on journals. For example, over the last five years, electronic formats accounted for 99% of **Galter Health Sciences Library & Learning Center's** collections spending. Gone are the days of physical carts of new arrivals waiting to be cataloged.

Although print backlogs have nearly disappeared, different kinds of cataloging backlogs have sprung up in their place that require new skills and workflows. Batch record uploads have edged out individual title-by-title cataloging and become the norm, requiring catalogers to rely on tools such as MarcEdit, Excel, OpenRefine, and even command line approaches for high-level metadata analysis and cleaning. After resources are cataloged, they require ongoing attention to assure access is maintained, subscription coverage is reflected, and platform changes are handled. Although this is commonly viewed as the realm of electronic resources librarians, the work of navigating the library catalog, updating MARC records, troubleshooting linking problems, and tracking down title changes lends itself to catalogers and metadata librarians. The management of electronic resources is a never-ending and highly collaborative process.

Library systems have also evolved to better integrate the workflows associated with e-resource management. For example, **Galter Library** uses **Ex Libris's** Alma platform, which utilizes electronic collections and portfolios for managing e-resource package, coverage, and linking information, allowing for improved integrations with traditional bibliographic metadata. Alma also offers the Community Zone of shared records, electronic collections, and portfolios for easy access to shared records and centralized management of e-resources. Although the completeness and currency of many records leaves much to be desired, the concept of globally shared records incorporating vendor updates in the ILS has dramatically altered e-resource workflows. Whole packages with corresponding MARC records and linking and coverage information can be activated for discovery in the catalog with the click of a button, and in some cases removed just as easily. Although enhancements to records in the Community Zone can be undertaken, core metadata is often viewed as "good enough" to allow for the discovery of resources.

In place of the cataloging duties replaced by the availability of records in shared environments, catalogers have shifted focus to other projects. Many libraries have begun prioritizing their unique physical and electronic collections for metadata work. Catalogers also spend time identifying and rectifying gaps in the shared catalog and resolving higher-level cataloging problems in areas such as legacy catalog records, serials title changes, and authority work. Cataloging work and database maintenance are interdependent, and the continuous improvement of library metadata is only growing in importance as libraries work to make resources discoverable to broader audiences via aggregators, external web search engines and the Semantic Web.

Linked Data

Initially the World Wide Web was developed to link documents. The Semantic Web advances this concept by linking the data and information that resides in the documents and identifying the relationships among them. Hence, the use of the phrase "Linked Data" to describe how the Semantic Web works.² The Semantic Web also contains datasets, including library catalogs and authority files such as VIAF, LC/Names, MeSH, LCSH, etc. Furthermore, the Semantic Web provides links between the data elements (i.e., entities) that reside in these documents and datasets. If libraries are to participate fully in the Semantic Web, they must use the technologies that support it along with metadata schemas that are able to manage linked data.³

In moving toward the Semantic Web, the library community (including health sciences libraries) hopes to replace their current metadata standard, MARC, with a linked data-based schema. For libraries MARC has been the standard for library cataloging and metadata creation for the past 50 years, and it has served the community very well. With the developments in computer and web technologies over the past 30 years, the environment in which libraries operate has changed significantly.⁴ Within this new environment the limitations and inadequacies of MARC have become obvious. MARC does a good job of enabling communication between humans, but it does not enable effective communication among modern computers, which is what optimizes the discovery and exchange in the new World Wide and Semantic Web environment.

There are several projects that the library community has begun in order to incorporate linked data into their catalogs and authority files, and to transition to a linked data system and metadata schema.⁵ One major project is the **Library of Congress's** BIBFRAME project, which was begun in 2012.⁶ The goal of this project is to replace the MARC metadata schema with a linked data model, BIBFRAME. In contrast to MARC, in which all the data describing a library resource is aggregated in a single catalog record, the BIBFRAME model consists of three bibliographic entities (Work, Instance, and Item); the relationships between these entities; and relationships to other entities related to the bibliographic entities such as subjects, agents (authors, contributors, publishers), events, classification systems, etc. These relationships will be provided by the **Library of Congress** and other linked data services.⁷ BIBFRAME is based on the Resource Description Framework (RDF), a Semantic Web standard used for managing linked data. The current version of BIBFRAME is 2.0, which LC has made available for testing and implementation by libraries and vendors.

The **National Library of Medicine (NLM)** is also playing an important role investigating linked data for libraries. In addition to participating in the BIBFRAME project, they have initiated several of their own projects.⁸ These include converting MeSH into a faceted subject system; developing a linked data service for MeSH; and adding unique resource identifiers (URIs) to MeSH headings in bibliographic records, which link them to NLM's linked data service.⁹ All of these are aimed at making legacy library data more machine actionable and ready for a transition to a linked data environment.

continued on page 17

New Roles

Outside of library resource management, there are other aspects of the health sciences library environment that have led to new roles for catalogers and metadata librarians. The **National Institutes of Health** implemented the first open access mandate by a large funding body in the United States in 2008, and in 2013 the **Office of Science and Technology** mandated all federal agencies with budgets over \$100 million develop plans for making the results of federally funded research publicly available. It is now the norm for many funding agencies to require data management plans as a component of grant applications. Because many health sciences researchers rely on grant funding, the mandates that have emerged over the past ten years have forced researchers to invest more in the long-term preservation of and access to their data sets.

Often overlooked in the data life cycle, metadata plays a crucial role in data preservation and discovery. Poorly constructed filenames can make file contents inscrutable, and a lack of supporting documentation can potentially render a dataset unusable over time. Metadata librarians are well situated to apply their expertise in standardized vocabularies and description to these problems. They can assist researchers in identifying and applying appropriate file naming conventions, creating supplementary README files with administrative and technical metadata, and using established schema such as the Data Documentation Initiative (DDI) to fully describe data in repositories. These practices will ensure that datasets can be discovered, preserved, and reused according to open access policies, funder mandates, and research reproducibility efforts, which are increasingly important for publishing and validating scientific discoveries.

Research impact evaluation is another service that is emphasized in health sciences libraries that has created new collaborative opportunities for metadata librarians. Researchers are frequently called upon to demonstrate the impact of their research and work for promotion and tenure. The **NIH** also frequently requires researchers to submit biographical sketches to demonstrate their qualifications for a project when applying for funding. Library services built around research impact are often situated in reference departments to take advantage

of the direct relationships liaison librarians have with faculty members and departments. However, metadata librarians and catalogers are increasingly involved in data gathering and analysis activities because of their unique skillset. Their contributions can range from constructing advanced queries for discovery tools or databases, downloading citation data, performing data cleanup, and visualizing datasets. These activities take advantage of an advanced knowledge of database structures, search techniques, and metadata analysis that are leveraged regularly in the performance of normal library cataloging work.

Several enterprise-level platforms for research information management have also been developed to assist with tracking researcher productivity. These not only benefit individual researchers, but also their affiliated institutions which are highly interested in examining and showcasing the productivity of their faculty. Commercial systems include Pure, Converis, Activity Insight, and Symplectic Elements, with open source options available through VIVO, Profiles, and Opus. Even though the implementation and maintenance of such systems does not always happen at the library level, the library is often a crucial partner given the way that these systems incorporate data from library-subscribed citation databases. Metadata librarians can contribute their knowledge of ontologies, controlled vocabularies, and citation metadata, which are all essential to configuring faculty profiles that accurately reflect updated publication information. Information from a variety of streams are often matched based on identifiers, and metadata librarians are excellent collaborators for navigating this landscape to ensure matches are accurate.

The health sciences landscape thus creates challenges and opportunities for catalogers and metadata librarians. The dominance of electronic resources and evolving library systems have changed cataloging priorities and workflows. Evolving standards in information consumption and exchange have forced libraries to investigate new ways to structure library metadata. And the unique scholarly landscape in the health sciences has created opportunities in data and research information management which utilize the skills of catalogers and metadata librarians in new ways. Metadata impacts so many areas, and presents so much potential, that the field is ripe for innovation and remains an integral part of health sciences library services. 🌱

Endnotes

1. **Stephen Bosch, and Kittie Henderson**, "New World, Same Model," *Library Journal* 142, no. 7 (2017): 40-45, (<https://lj.libraryjournal.com/2017/04/publishing/new-world-same-model-periodicals-price-survey-2017>) (accessed May 22, 2018).
2. **Tim Berners-Lee, James Hendler, and Ora Lassila**, "The Semantic Web," *Scientific American* 284, no. 5 (May 2001): 35-43, (<https://www.nature.com/scientificamerican/journal/v284/n5/pdf/scientificamerican0501-34.pdf>) (accessed May 29, 2018).
3. **Karen Coyle**, "Linked Data Tools: Connecting to the Web," *Library Technology Reports*, no. 4 (May/June 2012): 6-42, (<https://journals.ala.org/index.php/ltr/issue/view/183>) (accessed May 29, 2018).
4. **Karen Coyle**, "RDA Vocabularies for a Twenty-First-Century Data Environment," *Library Technology Reports*, no. 2 (February/March 2010): 5-36, (<https://journals.ala.org/index.php/ltr/issue/view/177>) (accessed June 4, 2018).
5. **Erik T. Mitchell**, "Linked Library Data: Early Activity & Development," *Library Technology Reports*, 52 no. 4 (January 2016): 5-33, (<https://journals.ala.org/index.php/ltr/issue/view/534>) (accessed May 29, 2018).
6. **Library of Congress. Bibliographic Framework Initiative**. (Washington, D.C.: Library of Congress) (<https://www.loc.gov/bibframe/>) (accessed May 29, 2018).
7. **Library of Congress. Linked Data Service**. (Washington, D.C.: Library of Congress) (<https://id.loc.gov/about>) (accessed May 29, 2018).
8. **Nancy Fallgren**, "NLM BIBFRAME Update," *NLM Technical Bulletin*, 404 (May/June 2015) (https://www.nlm.nih.gov/pubs/techbull/mj15/mj15_bibframe.html) (accessed May 29, 2018).
9. **Diane Boehr**, "Discontinuing Distribution of Cataloging Bibliographic Records with Artificially Reconstructed Subject Strings: Comment by August 31, 2015," *NLM Technical Bulletin*, 404 (May/June 2015) (https://www.nlm.nih.gov/pubs/techbull/mj15/mj15_cataloging_unstringing_survey.html) (accessed May 29, 2018); National Library of Medicine (U.S.). Medical Subject Headings RDF. (Bethesda, Md.: National Library of Medicine) (<https://id.nlm.nih.gov/mesh/>) (accessed May 29, 2018); Diane Boehr, "Adding MeSH URIs to NLM Catalog Records," *NLM Technical Bulletin*, 415 (March/April 2017) (https://www.nlm.nih.gov/pubs/techbull/ma17/ma17_adding_uris_mesh_2_nlm_catalog.html) (accessed May 29, 2018).