Transitioning to the Next Generation of Metadata:

An Annotated Bibliography of Selected OCLC Research Hanging Together Blogs

Karen Smith-Yoshimura

OCLC

Transitioning to the Next Generation of Metadata: An Annotated Bibliography of Selected OCLC Research Hanging Together Blogs

Karen Smith-Yoshimura
Senior Program Officer

OCLC

September 2020

ORCID iDs
Karen Smith-Yoshimura ⓘ https://orcid.org/0000-0002-8757-2962

Suggested citation:
Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata: An Annotated Bibliography of Selected OCLC Research Hanging Together Blogs*. Dublin, OH: OCLC Research. https://www.doi.org/10.25333/rq4m-2492.

CONTENTS

The following annotated bibliography provides more details on the *Hanging Together* posts referred to in *Transitioning to the Next Generation of Metadata,** primarily summaries of discussions by the OCLC Research Library Partners Metadata Managers Focus Group in 2015-2020. The order of the sections matches that in the report, with the blog posts listed in reverse chronological order.

INTRODUCTION

Smith-Yoshimura, Karen. 2018. "What Metadata Managers Expect from and Value about the Research Library Partnership." *Hanging Together: The OCLC Research Blog*, 16 April 2018. https://hangingtogether.org/?p=6683.

This post features key benefits of institutional affiliation with the OCLC Research Library Partnership as identified by metadata managers, including professional development and opportunities to do research collaboratively. Metadata managers also expressed the wish for more "practical applications" in future areas of work.

———. 2017. "Metadata Advocacy." *Hanging Together: The OCLC Research Blog*, 17 October 2017. https://hangingtogether.org/?p=6282.

An indispensable skill in these times is the ability to explain in a concise, compelling, and relatable way to administrators and other decision makers why resources should be allocated to creating and maintaining structured metadata. This blog post includes "elevator speeches" to describe metadata's value to support communications with these stakeholders.

LINKED DATA

Smith-Yoshimura, Karen. 2019. "Systematic Reviews of our Metadata." *Hanging Together: The OCLC Research Blog*, 10 April 2019. https://hangingtogether.org/?p=7117.

Librarians and administrators are well aware of the tension that exists between delivering access to our library collections in a timely manner and providing good quality description. The metadata descriptions must be full enough to allow us to manage our collections and to support accessibility and discoverability for the end user. Many libraries need to compromise by using vendor records, by creating minimal or less-than-full level descriptions (according to existing guidelines such as BIBCO or creating their own) for certain types of resources, and by limiting authority work. We need to better understand the impact that these compromises are having on our end users.

———. 2015. "Shift to Linked Data for Production." *Hanging Together: The OCLC Research Blog*, 13 May 2015. https://hangingtogether.org/?p=5195.

Linked data may well be the next common infrastructure both for communicating library data and embedding it into the fabric of the semantic web. Linked data offers an opportunity to go beyond the library domain and authority files to draw on information about entities from diverse sources.

---

* Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata.* Dublin, OH: OCLC Research. https://doi.org/10.25333/rqgd-b343.

———. 2015. "Working in Shared Files." *Hanging Together: The OCLC Research Blog*, 7 April 2015. https://hangingtogether.org/?p=5091.

Sharing files can mitigate the tendency for individual libraries to make redundant local edits. This post summarizes approaches to feed corrected records into the library ecosystem so that individual libraries need not make the same edits. Sharing data in a distributed environment is more complex than sharing data in a centralized file, but the effort the library community has invested in common cataloging standards and practices should facilitate accepting others' records without review or editing. The linked data environment offers a new opportunity to point to the most accurate metadata.

PERSISTENT IDENTIFIERS

Smith-Yoshimura, Karen. 2019. "'Future Proofing' of Cataloging." *Hanging Together: The OCLC Research Blog*, 11 October 2019. https://hangingtogether.org/?p=7526.

Metadata departments need to focus on both current requirements for their metadata in the library catalog or repositories and ensuring that they look ahead to future uses of their metadata in emerging services. Our discussions focused on identifiers, which were viewed as a transition bridge from legacy and current metadata to future applications.

———. 2017. "How Much Metadata is Practical?" *Hanging Together: The OCLC Research Blog*, 24 November 2017. https://hangingtogether.org/?p=6328.

With the increasing availability of online metadata, we are seeing metadata added to discovery environments representing objects of widely varying granularity. Ensuring optimal search results in an environment where metadata from varying sources with differing models of granularity and extensiveness poses challenges for catalogers and metadata managers.

———. 2017. "Use Cases for Local Identifiers." *Hanging Together: The OCLC Research Blog*, 5 May 2017. https://hangingtogether.org/?p=5938.

This post summarizes use cases for maintaining local authority files, the barriers to contributing local authority records to NACO or national authority files, the trade-offs of minting local identifiers, and possible alternative approaches to sharing information about entities gathered currently only on the local level.

———. 2016. "Metadata Reconciliation." *Hanging Together: The OCLC Research Blog*, 28 September 2016. https://hangingtogether.org/?p=5710.

Many libraries are performing metadata reconciliation work, such as searching matching terms from linked data sources and replacing strings in metadata records with URIs or storing URIs, as a necessary first step toward the linked data environment or as part of metadata enhancement work.

———. 2015. "Getting Identifiers Created for Legacy Names." *Hanging Together: The OCLC Research Blog*, 30 October 2015. https://hangingtogether.org/?p=5463.

The LC/NACO name authority file contains only 30% of the total names reflected in WorldCat's bibliographic record access points. The library community has become aware of the importance of getting persistent identifiers created for all names. These identifiers are crucial for the transition to linked data.

————. 2015. "Persistent Identifiers for Local Collections." *Hanging Together: The OCLC Research Blog*, 27 October 2015. https://hangingtogether.org/?p=5445.

Information professionals want to repurpose, present, and connect the data they have created and curated using various standards and practices. Persistent identifiers—unchanging over time and independent of where the object is or will be stored—help collections become accessible globally and can be shared and reused. This blog post summarizes the issues that need to be addressed for identifiers to be more widely adopted.

————. 2013. "Irreconcilable Differences? Name Authority Control and Humanities Scholarship." *Hanging Together: The OCLC Research Blog*, 27 March 2013. https://hangingtogether.org/?p=2621.

This blog post discusses the shared values and differences between humanities scholars and librarians in dealing with names appearing in historical research. Although humanities scholars appreciate the work librarians invest in authority files, they eschew the requirement for a "preferred name" and need to know the provenance of each form of name.

IDENTITY MANAGEMENT

Smith-Yoshimura, Karen. 2020. "Experimentations with Wikidata/Wikibase." *Hanging Together: The OCLC Research Blog*, 18 June 2020. https://hangingtogether.org/?p=8002.

Libraries and archives have been experimenting with Wikidata and Wikibase (the software platform underlying Wikidata) for a variety of purposes, including aggregating a variety of identifiers and labels in different languages as a possible alternative to traditional authority control and to facilitate the transition to linked data. This post summarizes reasons for using Wikidata or Wikibase in a library environment and the barriers to doing so.

————. 2018. "The Coverage of Identity Management Work." *Hanging Together: The OCLC Research Blog*, 8 October 2018. https://hangingtogether.org/?p=6805.

The emphasis in authority work has been shifting from construction of text strings to identity management—differentiating entities, creating identifiers, and establishing relationships between entities.

————. 2017. "Beyond the Authorized Access Point?" *Hanging Together: The OCLC Research Blog*, 10 October 2017. https://hangingtogether.org/?p=6271.

Can librarians continue to insist on using the authorized access point as the primary identifier in bibliographic records? Other communities make use of a variety of identifiers to differentiate entities. We have an opportunity to work with these other communities to establish "same as" relationships among different identifiers and contribute to "identifier hubs" that aggregate multiple identifiers referring to the same entity.

————.2016. "Impact of Identifiers on Authority Workflows." *Hanging Together: The OCLC Research Blog*, 22 March 2016. https://hangingtogether.org/?p=5603.

Using identifiers now to point to "things" rather than relying on text strings will facilitate transforming legacy data into linked data later. This post identifies potential areas of impact and tools.

ADDRESSING THE NEED FOR MULTIPLE VOCABULARIES AND EQUITY, DIVERSITY, AND INCLUSION

Smith-Yoshimura, Karen. 2019. "Strategies for Alternate Subject Headings and Maintaining Subject Headings." *Hanging Together: The OCLC Research Blog*, 29 October 2019. https://hangingtogether.org/?p=7591.

> Institutional commitments to Equity, Diversity, and Inclusion is a driving factor for libraries to use alternate subject headings but applying them only locally has drawbacks. This post discusses driving factors for libraries to use alternate subject headings, the challenges in applying these alternate subject headings only locally, and some strategies for navigating this approach.

———. 2018. "Creating Metadata for Equity, Diversity, and Inclusion." *Hanging Together: The OCLC Research Blog*, 7 November 2018. https://hangingtogether.org/?p=6833.

> Most of our discussion focused on the complexity of embedding equity, diversity, and inclusion in controlled vocabularies used in library catalogs. Consulting the communities reflected in our descriptive metadata and access points would help facilitate moving to terms that are more appropriate and respectful. Our metadata is currently created according to Western knowledge constructs, and our systems have been designed around them.

———. 2018. "Are Distributed Models for Vocabulary Maintenance Viable?" *Hanging Together: The OCLC Research Blog*, 12 April 2018. https://hangingtogether.org/?p=6672.

> In theory, linked data would provide the means for local communities to prefer a different label for an established vocabulary's preferred term for a concept or entity. A distributed model for managing both terminology and entity description raises issues around metadata stability expectations, metadata interoperability, and metadata maintenance.

———. 2017. "Vocabulary Control Data in Discovery Environments." *Hanging Together: The OCLC Research Blog*, 5 October 2017. https://hangingtogether.org/?p=6264.

> Traditional authority control models have relied on left-anchored browsing of alphabetically ordered lists of terms, a model that interposes the controlled terms—preferred, variant, and related—between the searcher and search results. The new world of authority sources in which libraries operate include ORCID and other international registries. Vocabularies designed for left-anchored browsing are a poor fit for current discovery environments oriented toward keyword search and facet term sets pulled directly from displayed search results.

———. 2016. "Faceted Vocabularies." *Hanging Together: The OCLC Research Blog*, 31 October 2016. https://hangingtogether.org/?p=5739.

> This piece discusses the pros and cons of moving from precoordinated subject headings to FAST (Faceted Application of Subject Terminology), which could serve as a "happy medium" between totally-uncontrolled keywords at one end of the spectrum and a difficult-to-learn-and-apply precoordinated subject strings at the other end.

DESCRIBING "INSIDE OUT" AND "FACILITATED" COLLECTIONS

Smith-Yoshimura, Karen 2020. "Presenting Metadata from Different Sources in Discovery Layers." *Hanging Together: The OCLC Research Blog*, 16 April 2020. https://hangingtogether.org/?p=7880.

Most libraries have implemented "discovery layers," which channel metadata from various (but not all) institutional data sources such as library catalogs, archives, institutional repositories, digital collections, and museum exhibitions. This post notes some of the reasons why retrieving metadata describing resources in completely different systems is "super challenging."

———. 2020. "Knowledge Management and Metadata." *Hanging Together: The OCLC Research Blog*, 9 April 2020. https://hangingtogether.org/?p=7845.

Libraries' expertise in metadata standards, identifiers, linked data, and data sharing systems as well as technical systems can be invaluable to the research life cycle. Metadata is fundamental to the complex research environment—beginning with the planning researchers do before and during the creation of data; to disseminating the knowledge gained; to understanding the impact, engagement, and the resulting reputation of home institutions. This post summarizes some of the critical issues in integrating libraries into research workflows.

———. 2018. "Metadata for Audios and Videos." *Hanging Together: The OCLC Research Blog*, 29 October 2018. https://hangingtogether.org/?p=6814.

Our libraries are repositories of large amounts of audiovisual materials, which often represent unique, local collections. Much of this audiovisual material is in dire need of preservation, digitization, clarification of conditions of use, and description. The nature of the management of these resources requires knowledge of the use context and well as technical metadata issues, providing a complex environment to think through programs of description and access.

———. 2017. "Metadata for Archival Collections." *Hanging Together: The OCLC Research Blog*, 30 March 2017. https://hangingtogether.org/?p=5903.

Archival collections are in many ways the jewels in the crown of collections as they are unique research resources and the fodder for creating new research, providing insights into the world across many centuries. Creating visibility for these collections reaps significant benefits for both researchers and libraries/archives. Archives are, however, complex, and present different metadata issues compared to traditional library collections.

———. 2016. "Sharing Digital Collections Workflows." *Hanging Together: The OCLC Research Blog*, 2 November 2016. https://hangingtogether.org/?p=5744.

Institutions are increasingly sharing the metadata for their digital collections with both national and international discovery services. Within individual organizations, librarians create and recreate metadata for digital and digitized resources in a plethora of systems—the library catalog, archive management, digital asset and preservation systems, the institutional repository, research management systems, and external subscription-based repositories.

————. 2016. "Metadata for Research Data Management." *Hanging Together: The OCLC Research Blog*, 18 April 2016. https://hangingtogether.org/?p=5616.

With increasing expectations that research data creation made possible through grant funding will be archived and made available to others, many institutions are becoming aware of the need to collect and curate this new scholarly resource. To maximize the chances that metadata for research data are shareable (that is, sufficiently comparable) and helpful to those considering reusing the data, our communities would benefit from sharing ideas and discussing plans to meet emerging discovery needs.

————. 2016. "Metadata for Archived Websites." *Hanging Together: The OCLC Research Blog*, 14 March 2016. https://hangingtogether.org/?p=5591.

For some years now, archives and libraries have been archiving web resources of scholarly or institutional interest to ensure their continuing access and long-term survival. Some websites are ephemeral or intentionally temporary, such as those created for a specific event. Institutions would like to archive and preserve the content of their websites as part of their historical record. A large majority of web content is harvested by web crawlers, but the metadata generated by harvesting alone is considered insufficient to support discovery.

————. 2015. "Managing Metadata for Image Collections." *Hanging Together: The OCLC Research Blog*, 9 April 2015. https://hangingtogether.org/?p=5130.

Managing a wide variety of image collections presents challenges for metadata management. In some cases, image collections that developed outside the library and its data models need to be integrated with other collections or into new search environments. Depending on the nature of the collection and its users, questions arise concerning identification of works, depiction of entities, chronology, geography, provenance, genre, subjects ("of-ness" and "about-ness"); so do opportunities for crowdsourcing and interdisciplinary research.

EVOLUTION OF "METADATA AS A SERVICE"

Smith-Yoshimura, Karen. 2020. "New Ways of Using and Enhancing Cataloging and Authority Records." *Hanging Together: The OCLC Research Blog*, 2 April 2020. https://hangingtogether.org/?p=7805.

This blog post provides examples of using legacy metadata in new applications, including nonlibrary use cases that leverage that metadata.

————. 2019. "Knowledge Organization Systems." *Hanging Together: The OCLC Research Blog*, 17 April 2019. https://hangingtogether.org/?p=7135.

As controlled vocabularies and thesauri are converted into linked open data and shared publicly, they often separate from their traditional role of facilitating collection browsing and find a renewed purpose as web-based knowledge organizations systems (KOS). Knowledge Organization Systems providing "semantic road maps" would require a major shift from local "collection-centric" systems to "knowledge organizations." The discussion highlighted some of our common aspirations for future systems both for discovery and for metadata management.

———. 2019. "Alternatives to Statistics for Measuring Success and Value of Cataloging." *Hanging Together: The OCLC Research Blog*, 15 April 2019. https://hangingtogether.org/?p=7122.

Traditionally, the most common measure of cataloger productivity statistics on the number of records produced and time spent cataloging. As cataloging and metadata librarians become more involved in other activities that are not easily quantifiable (e.g., participating in linked data or similar projects), the problem of measuring productivity and success becomes more difficult. Our discussions focused on communicating the successes and challenges of metadata specialists with the rest of one's institution and how metadata contributes to the division's and organization's strategic goals.

———. 2015. "Services Built on Usage Metrics." *Hanging Together: The OCLC Research Blog*, 30 September 2015. https://hangingtogether.org/?p=5430.

Possible new kinds of services that rely on nonbibliographic data, usage metrics and data analysis techniques could include collection management, identifying materials for offsite storage, deciding which subscriptions to maintain, comparing citations for researchers' publications with what the library is not purchasing, improving relevancy ranking, personalizing search results, offering recommendation services, and measuring impact of library usage on research or student success.

GEARING UP FOR FUTURE STAFFING REQUIREMENTS

Smith-Yoshimura, Karen. 2020. "Metadata Management in Times of Uncertainty." *Hanging Together: The OCLC Research Blog*, 15 June 2020. https://hangingtogether.org/?p=7998.

The COVID-19 crisis has caused a dramatic change in how libraries deliver services to patrons. The increased importance of online access to collections highlighted gaps in staff skills to be addressed and whole teams had to be upskilled on e-book metadata processes. This was a period when workflows were rethought, and some changes are anticipated to carry over into future metadata workflows post-pandemic. "Teleworking is here to stay."

———. 2019. "Stewardship of Professional FTEs in Metadata Work and Turnover." *Hanging Together: The OCLC Research Blog*, 18 October 2019. https://hangingtogether.org/?p=7580.

Turnover in a professional position within a cataloging or metadata unit now comes with a significant risk that it will be impossible to convince administrators to retain the position in the unit and repost it. Metadata managers have a variety of strategies for dealing with the impact of losing professional librarians. Potential candidates with more flexible skill sets have become more attractive than those with a traditional cataloging background who may not adapt well to working in new environments.

———. 2018. "MarcEdit and Other Tools for Batch Processing and Metadata Reconciliation." *Hanging Together: The OCLC Research Blog*, 26 March 2018. https://hangingtogether.org/?p=6646.

MarcEdit was the most reported tool used for metadata reconciliation and batch-processing as reported by metadata managers. This post summarizes the specific tasks MarcEdit and other tools are used for and why metadata managers value MarcEdit.

———. 2017. "New Skill Sets for Metadata Management." *Hanging Together: The OCLC Research Blog*, 17 April. https://hangingtogether.org/?p=5929.

Today's changing landscape calls for skill sets needed by both new professionals entering the field and seasoned catalogers to successfully transition to the emerging linked data and semantic web environment. Managers want to focus less on specific schema and more on metadata principles that can be applied to a range of different formats and environments. This post summarizes the soft and technical skills needed and opportunities for professional development.

For more information about our work related to digitizing library collections, please visit: **oc.lc/digitizing**

OCLC

6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878
T: +1-614-764-6000
F: +1-614-764-6096
www.oclc.org/research