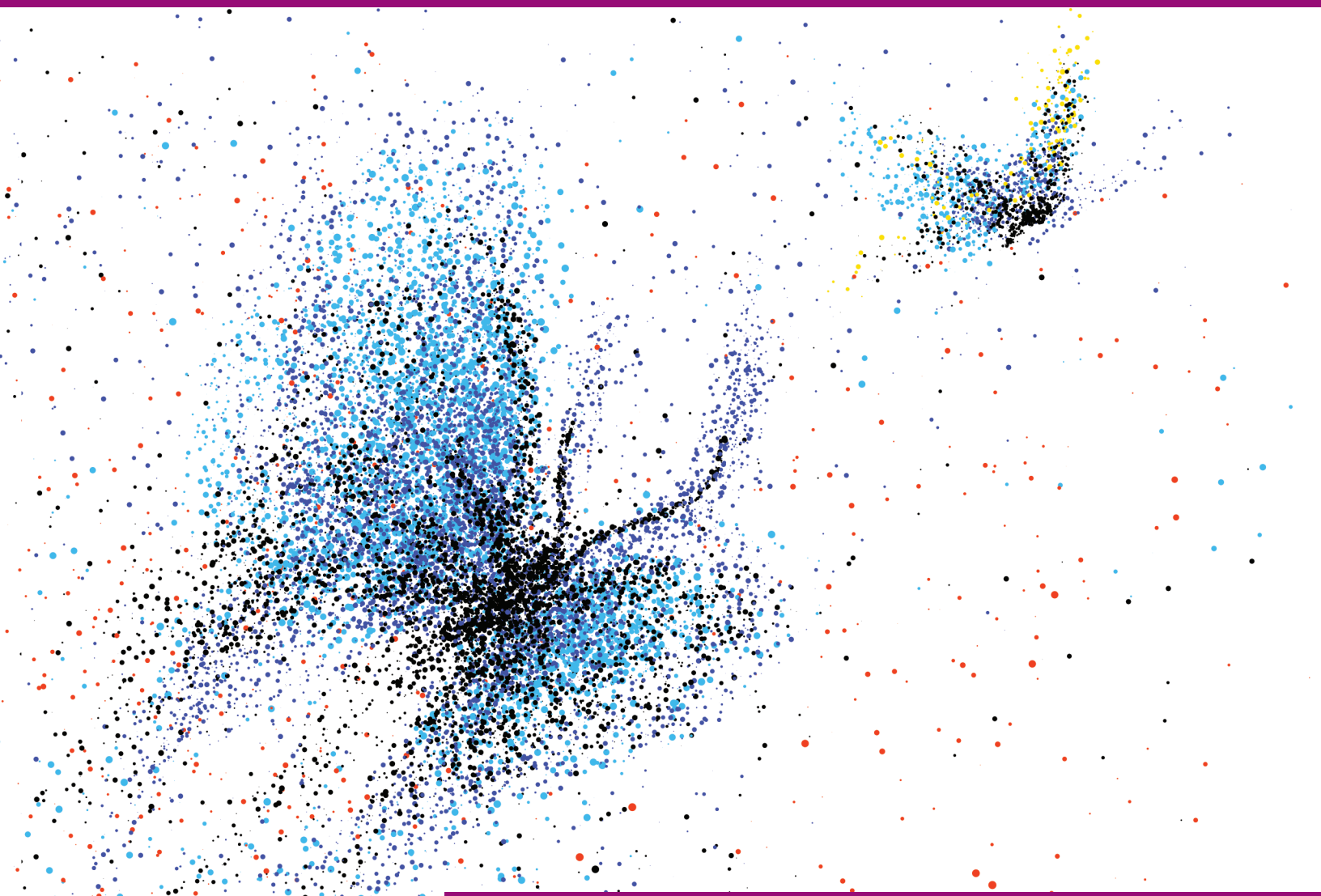
OCLC RESEARCH REPORT



Transitioning to the Next Generation of Metadata

Karen Smith-Yoshimura

OCLC

Transitioning to the Next Generation of Metadata

Karen Smith-Yoshimura
Senior Program Officer

September 2020

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 978-1-55653-167-5
DOI: 110.25333/rqgd-b343
OCLC Control Number: 1197990500

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

Please direct correspondence to:
OCLC Research
oclcresearch@oclc.org

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

CONTENTS

FIGURES

EXECUTIVE SUMMARY

The OCLC Research Library Partners Metadata Managers Focus Group, first established in 1993, is one of the longest-standing groups within the OCLC Research Library Partnership (RLP), a transnational network of research libraries. The Focus Group provides a forum for administrators responsible for creating and managing metadata in their institutions to share information about topics of common concern and to identify metadata management issues. The issues raised by the Focus Group are pursued by OCLC Research in support of the RLP and inform OCLC products and services.

This report, ***Transitioning to the Next Generation of Metadata***, synthesizes six years (2015-2020) of OCLC Research Library Partners Metadata Managers Focus Group discussions and what they may foretell for the "next generation of metadata." The firm belief that metadata underlies all discovery regardless of format, now and in the future, permeates all Focus Group discussions.

Yet metadata is changing. Format-specific metadata management based on curated text strings in bibliographic records understood only by library systems is nearing obsolescence, both conceptually and technically. Innovations in librarianship are exerting pressure on metadata management practices to evolve as librarians are required to provide metadata for far more resources of various types and to collaborate on institutional or multi-institutional projects with fewer staff. This report traces how metadata is evolving and considers the impact this transition may have on library services, posing such questions as:

- **Why is metadata changing?**
- **How is the creation process changing?**
- **How is the metadata itself changing?**
- **What impact will these changes have on future staffing requirements, and how can libraries prepare?**

The future of linked data is tied to the future of metadata: the metadata that libraries, archives, and other cultural heritage institutions have created and will create will provide the context for future linked data innovations as "statements" associated with those links. The impact will be global, affecting how librarians and archivists will describe the inside-out and facilitated collections, inspiring new offerings of "metadata as a service," and influencing future staffing requirements.

Transitioning to the next generation of metadata is an evolving process, intertwined with changing standards, infrastructures, and tools. Together, Focus Group members came to a common understanding of the challenges, shared possible approaches to address them, and inoculated these ideas into other communities that they interact with.

INTRODUCTION

The OCLC Research Library Partners Metadata Managers Focus Group (hereafter referenced as the Focus Group),[1] first established in 1993, is one of the longest-standing groups within the OCLC Research Library Partnership (RLP),[2] a transnational network of research libraries. The Focus Group provides a forum for administrators responsible for creating and managing metadata in their institutions to share information about topics of common concern and to identify metadata management issues. The issues raised by the Focus Group are pursued by OCLC Research in support of the RLP and inform OCLC products and services.

The firm belief that metadata underlies all discovery regardless of format, now and in the future, permeates all Focus Group discussions. Metadata provides the research infrastructure necessary for all libraries' "value delivery systems," fulfilling their community's requests for information and resources. Metadata is crucial for transitioning to next generations of library and discovery systems. Good metadata created today can easily be reused in a linked data environment in the future.[3] As noted in the *British Library's Foundations for the Future*: "Our vision is that by 2023 the Library's collection metadata assets will be unified on a single, sustainable, standards-based infrastructure offering improved options for access, collaboration and open reuse."[4]

Format-specific metadata management based on curated text strings in bibliographic records understood only by library systems is nearing obsolescence, both conceptually and technically.

Format-specific metadata management based on curated text strings in bibliographic records understood only by library systems is nearing obsolescence, both conceptually and technically. Innovations in librarianship are exerting pressure on metadata management practices to evolve as librarians are required to provide metadata for far more resources of various types and to collaborate on institutional or multi-institutional projects with fewer staff. "Traditional methods of metadata generation, management and dissemination," suggests the British Library's Collection Management Strategy, "are not scalable or appropriate to an era of rapid digital change, rising audience expectations and diminishing resources."[5] Focus Group members are eager to unleash the power of metadata in legacy records for different interactions and uses by both machines and end-users in the future. Consistent metadata created according to past rules or standards need to be transformed into new structures.

Why is metadata changing?

Traditional library metadata was and is made by librarians conforming to rules that are mainly used and understood by librarians. It is record-centered, expensive to produce, and has historic size limitations. Metadata is limited in its coverage, notably not including articles within scholarly journals or other scholarly outputs. The infrastructure has been inadequate for managing corrections and enhancements, inducing an emphasis on perfection that has exacerbated the slowness of metadata creation. In short, the metadata could be better, there is not enough of it, and the metadata that does exist is not used widely outside the library domain.

How is the creation process changing?

Metadata is no longer created by library staff alone. Today, publishers, authors, and other interested parties are equally involved in metadata creation. Metadata creation has also been pushed forward in the scholarly life cycle, with publishers creating metadata records earlier than in the traditional cataloging process. Metadata can now be enhanced or corrected by machines or by crowdsourcing.

How is the metadata itself changing?

Machine-readable cataloging (MARC) was created to replicate the metadata traditionally found on library catalog cards. We are transitioning from MARC records to assemblages of well-coded and shareable, linkable components, with an emphasis on references, and we are eliminating anachronistic abbreviations not understood by machines. Instead of relying only on library vocabularies such as subject headings and coded lists, the developing assemblages can accommodate vocabularies created for specific domains, expanding the metadata's potential audiences.

In short, the metadata could be better, there is not enough of it, and the metadata that does exist is not used widely outside the library domain.

The Focus Group's composition has fluctuated over time, and currently comprises representatives from 63 RLP Partners in 11 countries spanning four continents.[6] The group includes both past and incoming chairs of the Program for Cooperative Cataloging (PCC),[7] providing cross-fertilization between the two. Topics for group discussions can be proposed by any Focus Group member and are selected by an eight-member Planning Group (see appendix), who then write "context statements" explaining why the topic is considered timely and important and then develop question sets that delve into the topic. Context statements and question sets are then distributed to all Focus Group members who are given three to five weeks to submit their responses. Compilations of the Focus Group's responses inform face-to-face discussions held in conjunction with the American Library Association conferences[8] and in subsequent virtual meetings.

As the Focus Group facilitator, I have summarized and synthesized these discussions in a series of *OCLC Research Hanging Together Blog* publications.[9] Nearly 40 blog posts on a wide range of metadata-related topics have been published on this forum over the past six years.

The Metadata Managers Focus Group is just one activity within the broader OCLC Research Library Partnership, which is devoted to extensive professional development opportunities for library staff. Focus Group members value their affiliation with the Research Library Partnership as a channel to becoming the "change agents" of future metadata management.[10] Focus Group members' responses to question sets have facilitated intra-institutional discussions and helped metadata managers understand how their institutions' situation compares with peers within the Partnership.

These Focus Group discussions identified a broad range of metadata-related issues, documented in this report. Transitioning to the next generation of metadata is an evolving process, intertwined with changing standards, infrastructures, and tools. Together, Focus Group members came to a common understanding of the challenges, shared possible approaches to address them, and inoculated these ideas into other communities that they interact with.

Collectively, Focus Group members command a wide range of experiences with linked data. The Focus Group's keen interest in linked data implementations sparked the series of OCLC Research's International Linked Data Surveys for Implementers.[11] A subset of Focus Group members have participated in various linked data projects, including the OCLC Research Project Passage and CONTENTdm Linked Data pilot, OCLC's Shared Entity Management Infrastructure, Library of Congress' Bibliographic Framework Initiative (BIBFRAME), the Mellon-funded Linked Data for Production (LD4P) project, the Share-VDE initiative, and the IMLS planning grant Shareable Local Name Authorities, which exposed issues raised by identifier hubs in the linked data environment.[12] In addition, Focus Group members contribute to the PCC task groups addressing aspects of linked data work, including the PCC Task Group on Linked Data Best Practices, Task Group on Identity Management, Task Group on URIs in MARC, and the PCC Linked Data Advisory Committee.[13] This cross-fertilization has prompted the Focus Group to examine issues around the *entities* represented in institutional resources.

This report synthesizes six years (2015-2020) of OCLC Research Library Partners Metadata Managers Focus Group discussions and what they may foretell for the "next generation of metadata." The document is organized in the following sections, each representing an emerging trend identified in the Focus Group's discussions:

- The transition to linked data and identifiers: expanding the use of persistent identifiers as part of the shift from "authority control" to "identity management"
- Describing the "inside-out" and "facilitated" collections: challenges in creating and managing metadata for unique resources created or curated by institutions in various formats and shared with consortia
- Evolution of "metadata as a service": increased involvement with metadata creation beyond the traditional library catalog
- Preparing for future staffing requirements: the changing landscape calls for new skill sets needed by both new professionals entering the field and seasoned catalogers

The document concludes with some observations on the forecasted impact of the next generation of metadata on the wider library community.

The Transition to Linked Data and Identifiers

Linked data offers the ability to take advantage of *structured data* with an emphasis on context. It relies on language-neutral identifiers pointing to objects, with a focus on "things" replacing the "strings" inherent in current authority and catalog records. These identifiers can then be connected to related data, vocabularies, and terms in other languages, disciplines, and domains, including nonlibrary domains. Linked data applications can consume others' contributions and thus free metadata specialists from having to re-describe things already described elsewhere, allowing them instead to focus on providing access to their institutions' unique and distinctive collections. This promises a richer user experience and increased discoverability with more contextual relationships than is possible with our current systems. Furthermore, linked data offers an opportunity to go beyond the library domain by drawing on information about entities from diverse sources.[14]

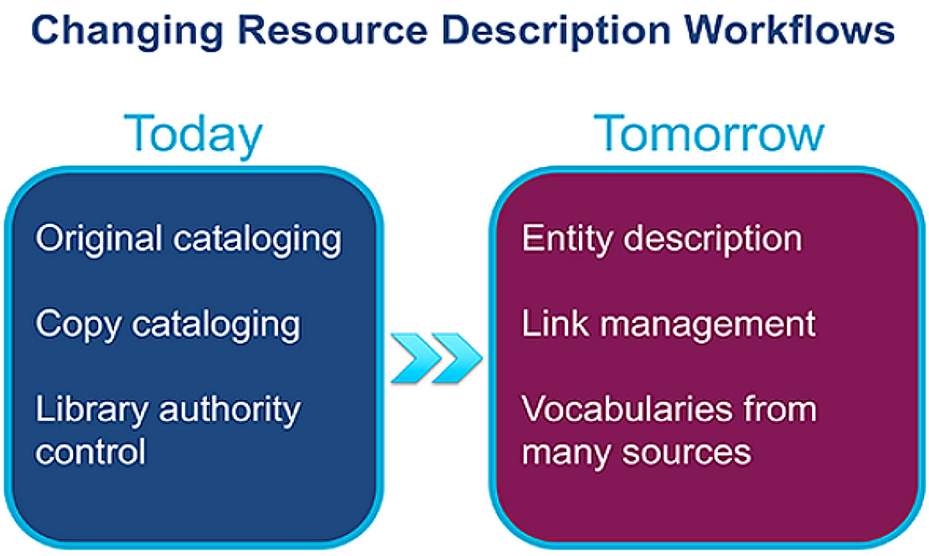


FIGURE 1. "Changing Resource Description Workflows" by OCLC Research[15]

The hope is that linked data will allow libraries to offer new, value-added services that current models cannot support, that outside parties will be able to make better use of library resource descriptions, and that the data will be richer because more parties share in its creation. Moving to a linked data environment portends changes to resource description workflows, as shown in figure 1.

The drive to move metadata operations to linked data depends on the availability of tools, access to linked data sources for reuse, documented best practices on identifiers and the metadata descriptions associated with them ("statements"), and a critical mass of implementations on a network level.

EXPANDING THE USE OF PERSISTENT IDENTIFIERS

The Focus Group discussed the "future-proofing" of cataloging, which refers to the opportunities to unleash the power of metadata in legacy records for different interactions and uses in the future. Persistent identifiers were viewed as crucial to transitioning from current metadata to future applications.[16] Identifiers, in the form of language-neutral alphanumeric strings, serve as a shorthand for assembling the elements required to uniquely describe an object or resource. They can be resolved over networks with specific protocols for finding, identifying, and using that object or resource. In the nonlibrary domain, Social Security and employee numbers are examples of

such identifiers. In the library and academic domains, Focus Group members pointed to ORCID (Open Researcher and Contributor ID)[17] as a "glue" that holds together the four arms of scholarly work: publishing, repository, library catalog, and researchers—but ORCID is limited to only *living* researchers. ORCID is increasingly used in STEM (science, technology, engineering, mathematics) journals for all authors and contributors[18] and included in institutions' Research Information Management systems. ISNI (International Standard Name Identifier)[19] uniquely identifies persons and organizations involved in creative activities used by libraries, publishers, databases, and rights management organizations, and it covers nonliving creators.

> Persistent identifiers were viewed as crucial to transitioning from current metadata to future applications.

Persistent identifiers are used by parties such as Google and HathiTrust for service integration.[20] More institutions are using geospatial coordinates in metadata or URIs (Uniform Resource Identifiers) pointing to geospatial coordinates that support API (Application Programming Interface) calls to GeoNames,[21] enabling map visualizations. Research institutions are also adopting person identifiers such as ORCID to streamline the collection of the institutional research record, usually through a Research Information Management system, as documented in the 2017 OCLC Research Report *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information Management*.[22]

While publishers serve as a key player in the metadata workstream, publisher data does not always meet library requirements. For example, publisher data for monographs usually does not include identifiers. The British Library is working with five UK publishers to add ISNIs[23] to their metadata as a promising proof-of-concept for publishers and libraries working together earlier in the supply chain. The ability to batch load or algorithmically add identifiers in the future is on Focus Group members' wish list.

No single person identifier covers all use cases. Researchers' names have been only partially represented in national name authority files that identify persons both living and dead. A sizable quantity of legacy names are represented only by text strings in bibliographic records. Authority records are created only by institutions involved in the PCC's Name Authority Cooperative Program (NACO)[24] or in national library programs. Even then, authority records are created selectively for certain headings or sometimes only when references are involved. The LC/NACO name authority file contained only 30% of the total names reflected in WorldCat's bibliographic record access points (9 million LC/NACO records compared to the 30 million total names reported on the WorldCat Identities project page as of 2012).[25] By 2020, this percentage decreased to 18%: 11 million LC/NACO authority records compared to 62 million in WorldCat Identities. These statistics illustrate that the number of names represented in bibliographic records are increasing more quickly than those that are under authority control.

Authority files focus on the "preferred form" of a name, which can vary depending on language, discipline, context, and time period. Scholars have objected to the very concept of a "preferred form," as the name may be referred to differently depending on the context.[26] When a name has multiple forms, historians need to know the provenance of each name following the citation

practices commonly used in their field. An identifier linked to different forms of names, each associated with the provenance and context, could resolve this conundrum.

Researcher names are just one example of a need unmet by current identifier systems. Institutions have been minting their own "local identifiers" to meet this need. Use cases for local identifiers include registering all researchers on campus; representing entities that are underrepresented in national authority files such as authors of electronic dissertations and theses, performers, events, local place names, and campus buildings; identifying entities in digital library projects and institutional repositories; reflecting multilingual needs of the community; and supporting "housekeeping" tasks such as recording archival collection titles.[27]

Focus Group members' consistent need to disambiguate names across disciplines and formats spurred creating the OCLC Research working group on Registering Researchers in Authority Files.[28] The need to accurately record researchers' institutional affiliations to reflect the institution's scholarly output, to promote cross-institutional collaborations, and to lead to more successful recruitment and funding led to another working group on Addressing the Challenges with Organizational Identifiers and ISNI,[29] which presented new data modeling of organizations that others could adapt for their own uses. Since then, the Research Organization Registry (ROR) was launched to develop an open, sustainable, usable, and unique identifier for every research organization in the world.[30]

Disambiguating names is the most labor-intensive part of authority work and will still be a prerequisite for assigning unique identifiers. Given the different name identifier systems already in use, libraries need a name reconciliation service. Authority work and algorithms based on text string matching have limits; the results will still need manual expert review. Tapping the expertise in user communities to verify if two identifiers represent the same person may help.

FERMILAB-PUB-14-123-E

Precision measurement of the top-quark mass in lepton+jets final states

V.M. Abazov,[31] B. Abbott,[67] B.S. Acharya,[25] M. Adams,[46] T. Adams,[44] J.P. Agnew,[41] G.D. Alexeev,[31] G. Alkhazov,[35] A. Alton[a,56] A. Askew,[44] S. Atkins,[54] K. Augsten,[7] C. Avila,[5] F. Badaud,[10] L. Bagby,[45] B. Baldin,[45] D.V. Bandurin,[73] S. Banerjee,[25] E. Barberis,[55] P. Baringer,[53] J.F. Bartlett,[45] U. Bassler,[15] V. Bazterra,[46] A. Bean,[53] M. Begalli,[2] L. Bellantoni,[45] S.B. Beri,[23] G. Bernardi,[14] R. Bernhard,[19] I. Bertram,[39] M. Besançon,[15] R. Beuselinck,[40] P.C. Bhat,[45] S. Bhatia,[58] V. Bhatnagar,[23] G. Blazey,[47] S. Blessing,[44] K. Bloom,[59] MI. Boehnlein,[45] D. Boline,[64] E.E. Boos,[33] G. Borissov,[39] M. Borysova[i,38] A. Brandt,[70] O. Brandt,[20] R. Brock,[57] A. Bross,[45] D. Brown,[14] X.B. Bu,[45] M. Buehler,[45] V. Buescher,[21] V. Bunichev,[33] S. Burdin[b,39] C.P. Buszello,[37] E. Camacho-Pérez,[28] B.C.K. Casey,[45] H. Castilla-Valdez,[28] S. Caughron,[57] S. Chakrabarti,[64] K.M. Chan,[51] A. Chandra,[72] E. Chapon,[15] G. Chen,[53] S.W. Cho,[27] S. Choi,[27] B. Choudhary,[24] S. Cihangir,[45] D. Claes,[59] J. Clutter,[53] M. Cooke[k,45] W.E. Cooper,[45] M. Corcoran,[72] F. Couderc,[15] M.-C. Cousinou,[12] D. Cutts,[69] A. Das,[42] G. Davies,[40] S.J. de Jong,[29,30] E. De La Cruz-Burelo,[28] F. Déliot,[15] R. Demina,[63] D. Denisov,[45] S.P. Denisov,[34] S. Desai,[45] C. Deterre[c,20] K. DeVaughan,[59] H.T. Diehl,[45] M. Diesburg,[45] P.F. Ding,[41] A. Dominguez,[59] A. Dubey,[24] L.V. Dudko,[33] A. Duperrin,[12] S. Dutt,[23] M. Eads,[47] D. Edmunds,[57] B. Ellison,[43] V.D. Elvira,[45] Y. Enari,[14] H. Evans,[49] V.N. Evdokimov,[34] A. Fauré,[15] L. Feng,[47] T. Ferbel,[63] F. Fiedler,[21] F. Filthaut,[29,30] W. Fisher,[57] H.E. Fisk,[45] M. Fortner,[47] H. Fox,[39] S. Fuess,[45] P.H. Garbincius,[45] A. Garcia-Bellido,[63] J.A. García-González,[28] V. Gavrilov,[32] W. Geng,[12,57] C.E. Gerber,[46] Y. Gershtein,[60] G. Ginther,[45,63] O. Gogota,[38] G. Golovanov,[31] P.D. Grannis,[64] S. Greder,[16] H. Greenlee,[45] G. Grenier,[17] Ph. Gris,[10] J.-F. Grivaz,[13] A. Grohsjean[c,15] S. Grünendahl,[45] M.W. Grünewald,[26] T. Guillemin,[13] G. Gutierrez,[45] P. Gutierrez,[67] J. Haley,[68] L. Han,[4] K. Harder,[41] A. Harel,[63] J.M. Hauptman,[52] J. Hays,[40] T. Head,[41] T. Hebbeker,[18] D. Hedin,[47] H. Hegab,[68] A.P. Heinson,[43] U. Heintz,[69] C. Hensel,[1] I. Heredia-De La Cruz[d,28] K. Herner,[45] G. Hesketh[f,41] M.D. Hildreth,[51] R. Hirosky,[73] T. Hoang,[44] J.D. Hobbs,[64] B. Hoeneisen,[9] J. Hogan,[72] M. Hohlfeld,[21] J.L. Holzbauer,[58] I. Howley,[70] Z. Hubacek,[7,15] V. Hynek,[7] I. Iashvili,[62] Y. Ilchenko,[71] L. Illingworth,[45] A.S. Ito,[45] S. Jabeen[m,45] M. Jaffré,[13] A. Jayasinghe,[67] M.S. Jeong,[27] R. Jesik,[40] P. Jiang,[4] K. Johns,[42] E. Johnson,[57] M. Johnson,[45] A. Jonckheere,[45] P. Jonsson,[40] J. Joshi,[43] A.W. Jung,[45] A. Juste,[36] E. Kajfasz,[12] D. Karmanov,[33] I. Katsanos,[59] R. Kehoe,[71] S. Kermiche,[12] N. Khalatyan,[45] A. Khanov,[68] L. Kharchilava,[62] Y.N. Kharzheev,[31] I. Kiselevich,[32] J.M. Kohli,[23] A.V. Kozelov,[34] J. Kraus,[58] A. Kumar,[62] M. Kupco,[8] T. Kurča,[17] V.A. Kuzmin,[33] S. Lammers,[49] P. Lebrun,[17] H.S. Lee,[27] S.W. Lee,[52] W.M. Lee,[45] X. Lei,[42] J. Lellouch,[14] D. Li,[14] H. Li,[73] L. Li,[43] Q.Z. Li,[45] J.K. Lim,[27] D. Lincoln,[45] J. Linnemann,[57] V.V. Lipaev,[34] R. Lipton,[45] H. Liu,[71] Y. Liu,[4] A. Lobodenko,[35] M. Lokajicek,[8] R. Lopes de Sa,[64] R. Luna-Garcia[g,28] K. L. Lyon,[45] A.K.A. Maciel,[1] R. Madar,[19] R. Magaña-Villalba,[28] S. Malik,[59] V.L. Malyshev,[31] J. Mansour,[20] J. Martínez-Ortega,[28] R. McCarthy,[64] C.L. McGivern,[41] M.M. Meijer,[29,30] A. Melnitchouk,[45] D. Menezes,[47] P.G. Mercadante,[3] M. Merkin,[33] A. Meyer,[18] J. Meyer[j,20] F. Miconi,[16] N.K. Mondal,[25] M. Mulhearn,[73] E. Nagy,[12] M. Narain,[69] R. Nayyar,[42] H.A. Neal,[56] J.P. Negret,[5] P. Neustroev,[35] H.T. Nguyen,[73] T. Nunnemann,[22] J. Orduna,[72] N. Osman,[12] J. Osta,[51] A. Pal,[70] N. Parashar,[50] V. Parihar,[69] S.K. Park,[27] R. Partridge[e,69] N. Parua,[49] A. Patwa[l,65] B. Penning,[45] M. Perfilov,[33] Y. Peters,[41] K. Petridis,[41] G. Petrillo,[63] P. Pétroff,[13] M.-A. Pleier,[65] V.M. Podstavkov,[45] A.V. Popov,[34] M. Prewitt,[72] D. Price,[41] N. Prokopenko,[34] J. Qian,[56] A. Quadt,[20] B. Quinn,[58] P.N. Ratoff,[39] I. Razumov,[34] I. Ripp-Baudot,[16] F. Rizatdinova,[68] M. Rominsky,[45] A. Ross,[39] C. Royon,[15] P. Rubinov,[45] R. Ruchti,[51] G. Sajot,[11] A. Sánchez-Hernández,[28] M.P. Sanders,[22] A.S. Santos[h,1] G. Savage,[45] M. Savitskyi,[38] L. Sawyer,[54] T. Scanlon,[40] R.D. Schamberger,[64] Y. Scheglov,[35] H. Schellman,[48] C. Schwanenberger,[41] R. Schwienhorst,[57] J. Sekaric,[53] H. Severini,[67] E. Shabalina,[20] V. Shary,[15] S. Shaw,[57] A.A. Shchukin,[34] V. Simak,[7] P. Skubic,[67] P. Slattery,[63] D. Smirnov,[51] G.R. Snow,[59] J. Snow,[66] I. Snyder,[65] S. Söldner-Rembold,[41] L. Sonnenschein,[18] K. Soustruznik,[6] J. Stark,[11] D.A. Stoyanova,[34] M. Strauss,[67] L. Suter,[41] P. Svoisky,[67] M. Titov,[15] V.V. Tokmenin,[31] Y.-T. Tsai,[63] D. Tsybychev,[64] B. Tuchming,[15] C. Tully,[61] L. Uvarov,[35] S. Uvarov,[35] S. Uzunyan,[47] R. Van Kooten,[49] W.M. van Leeuwen,[29] N. Varelas,[46] E.W. Varnes,[42] I.I. A. Vasilyev,[34] A.Y. Verkheev,[31] L.S. Vertogradov,[31] M. Verzocchi,[45] M. Vesterinen,[41] D. Vilanova,[15] P. Vokac,[7] H.D. Wahl,[44] M.H.L.S. Wang,[45] J. Warchol,[51] G. Watts,[74] M. Wayne,[51] J. Weichert,[21] L. Welty-Rieger,[48] M.R.J. Williams,[49] G.W. Wilson,[53] M. Wobisch,[54] D.R. Wood,[55] T.R. Wyatt,[41] Y. Xie,[45] R. Yamada,[45]

Disambiguation is particularly difficult for authors or contributors listed in journal articles, where names are often abbreviated and there may be dozens or even hundreds of contributors. For example, an article in *Physical Review Letters*—Precision Measurement of the Top Quark Mass in Lepton + Jets Final State—has approximately 300 abbreviated author names for a five-page article (figure 2).[31] This exemplifies the different practices among disciplines. By contrast, other objects with many contributors such as feature films and orchestral recordings are usually represented by only a relative handful of the associated names in library legacy metadata.[32] Such differences make creating metadata that is uniform, understandable, and widely reusable a challenge.

FIGURE 2. Some 300 abbreviated author names for a five-page article in *Physical Review Letters*

Abbreviated forms of author names on journal articles make it difficult—and often impossible—to match them to the correct authority form or an identifier, if it exists. Associating ORCIDs with article authors makes it easier to differentiate authors with the same abbreviated forms. Research Information Management (RIM) systems apply identity management for local researchers so that they are correctly associated with the articles they have written. Their articles are displayed as part of their profiles. (See for example, Experts@Minnesota or University of Illinois at Urbana-Champaign's Experts research profiles.)[33] For researcher identity management to work, individuals must create and maintain their own ORCIDs. Institutions have been encouraging their researchers to include an ORCID in their profiles. Researchers have greater incentives to adopt ORCID to meet national and funder requirements such as those of the National Science Foundation and the National Institutes of Health in the United States.[34] Research Information Management Systems harvest metadata from abstract and indexing databases such as Scopus, Web of Science, and PubMed, each of which has its own person identifiers that help with disambiguation; they may also be linked to an author's ORCID. Linked data could access information across many environments, including those in Research Information Systems, but would require accurately linking multiple identifiers for the same person to each other.

Some Focus Group members are performing metadata reconciliation work, such as searching matching terms from linked data sources and adding their URIs in metadata records as a necessary first step toward a linked data environment or as part of metadata enhancement work.[35] Improving the quality of the data improves users' experiences in the short term and will help with the transition to linked data later. Most metadata reconciliation is done on personal names, subjects, and geographic names. Sources used for such reconciliation include OCLC's Virtual International Authority File (VIAF), the Library of Congress's linked data service (id.loc.gov), ISNI, the Getty's Union List of Artists Names (ULAN), Art and Architecture Thesaurus (AAT), and Thesaurus of Geographic Names (TGN), OCLC's Faceted Application of Subject Terminology (FAST), and various national authority files. Selections of the source depend on the trustworthiness of the organization responsible, subject matter, and richness of the information. Such metadata reconciliation work is labor intensive and does not scale well.

Some members of the Focus Group have experimented with obtaining identifiers (persistent URIs from linked data sources) to eventually replace their current reliance on text strings. Institutions concluded that it is more efficient to create URIs in authority records at the outset rather than reconcile them later. The University of Michigan has developed a LCNAF Named Entity Reconciliation program[36] using Google's Open Refine that searches the VIAF file with the VIAF API for matches, looks for Library of Congress source records within a VIAF cluster, and extracts the authorized heading. This results in a dataset pairing the authorized LC Name Authority File heading with the original heading and a link to the URI of the LCNAF linked data service. This service could be modified to bring in the VIAF identifier instead; it gets fair results even though it uses string matching.

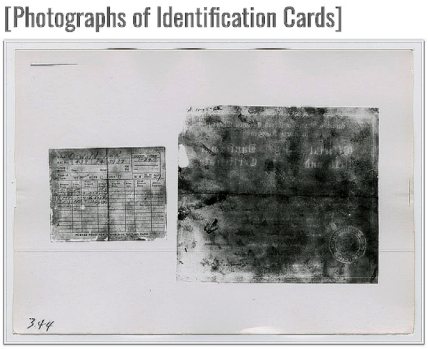
A long list of nonlibrary sources that could enhance current authority data or could be valuable to link to in certain contexts has been identified. Wikidata and Wikipedia led the list. Other sources include: AllMusic, author and fan sites, Discogs, EAC-CPF (Encoded Archival Context for Corporate Bodies, Persons, and Families), EAD (Encoded Archival Description), family trees, GeoNames, GoodReads, IMDb (Internet Movie Database), Internet Archive, Library Thing, LinkedIn, MusicBrainz, ONIX (ONline Information eXchange), Open Library, ORCID, and Scopus ID. The PCC's Task Group on URIs in MARC's document, *Formulating and Obtaining URIs: A Guide to Commonly Used Vocabularies and Reference Sources,*[37] provides valuable guidance for collecting data from these other sources. Wikidata is viewed as an important source for expanding the language range and providing multilingual metadata more easily than with current library systems.[38]

Identifiers for "works" represent a particular challenge, as there is no consensus on what represents a "distinctive work."[39] Local work identifiers cannot be shared or reused. Focus Group members voiced concern that differing interpretations of what a "work" is could hamper the ability to reuse data created elsewhere and look to a central trusted repository like OCLC to publish persistent Work Identifiers that could be used throughout the community.

Identifiers need to be both unchanging over time and independent of where the digital object is or will be stored. For instance, identifiers for data sets such as digital resources and collections in institutional repositories include system-generated IDs, locally minted identifiers, PURL handles, DOIs (Digital Object Identifiers), URIs, URNs, and ARKs (Archival Resource Keys). A few examples of DOI and ARK Identifiers are shown in figure 3. Resources can have both multiple copies and versions that change over time. Institutional repositories used as collaborative spaces can lead to multiple publications from the same data sets, a problem compounded by self-deposits from coauthors at different institutions into different repositories. Furthermore, libraries (as well as funders and national assessment efforts) want to be able to link related pieces (such as preprints, supplementary data, and images) with the publication. Multiple DOIs pointing to the same object pose a problem. Some libraries use DataCite or Crossref to mint and publish unique, long-term identifiers and thus minimize the potential for broken citation links.[40] Ideally, libraries would contribute to a hub for the metadata describing their researchers' data sets regardless of where the data sets are stored.

Examples of Some DOI and ARK Identifiers.

Registration Agency	Examples
DataCite	DOI names for accessing registered research datasets: Sets & Subsets: • Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. *Geological Institute, University of Tokyo*. [doi:10.1594/PANGAEA.726855] Earthquake Event, Authored by Automated System: • Geofon operator (2009): GEOFON event gfz2009kciu (NW Balkan Region) *GeoForschungsZentrum Potsdam(GFZ)*. [doi:10.1594/GFZ.GEOFON.gfz2009kciu]
Multilingual European DOI Registration Agency *m*EDRA	Sampling of resources identified using mEDRA DOI names: • **Journal Article:** Prodi, Romano. "L'industria dopo l'euro", *L'industria-Rivista di economia e politica industriale* 4, 559-566 (2002); [doi:10.1430/8105] • **Monograph:** Attanasio, Piero. *"The use of DOI system in eContent value chain: The case of Casalini Digital Division and mEDRA"*, White Paper (PDF). [doi:10.1392/BC1.0]
Japan Link Center (JaLC)	DOI names for Japanese Journal articles: • **Journal Article:** 竹本 賢太郎, 川東 正美, 久保 信行, 左近 多喜男, 大学におけるWebメールとターミナルサービスの研究, 標準化研究 Vol.7(2009), No.1 p.11-20 [doi:10.11467/isss2003.7.1_11] • **Journal Article:** 川崎 努, 植物における免疫誘導と病原微生物の感染戦略, ライフサイエンス 領域融合レビュー, 2, e008 (2013), [doi:10.7875/leading.author.2.e008]



• **Archival Resource Key**: ark:/67531/metapth346793

Relationships

- [Identification Cards, Photograph #1], DSMA_91-001-photo-344-001, ark:/67531/metapth346554
- [Identification Cards, Photograph #2], DSMA_91-001-photo-344-002, ark:/67531/metapth346846
- [Identification Cards, Photograph #3], DSMA_91-001-photo-344-003, ark:/67531/metapth346928

FIGURE 3. Examples of some DOI (left) and ARK (right) identifiers[41]

MOVING FROM "AUTHORITY CONTROL" TO "IDENTITY MANAGEMENT"

The emphasis in authority work is shifting from construction of text strings to *identity management*—differentiating entities, creating identifiers, and establishing relationships among entities.[42] The intellectual work required to differentiate names is the same for both current authority work and identify management. Focus Group members agree that the future is in identity management and getting away from "managing text strings" as the basis of controlling headings in

bibliographic records.[43] But identity management poses a change in focus, from providing access points in resource descriptions to describing the *entities* in the resource (work, persons, corporate bodies, places, events) and establishing the relationships and links among them.

Identity management poses a change in focus, from providing access points in resource descriptions to describing the ***entities*** in the resource (work, persons, corporate bodies, places, events) and establishing the relationships and links among them.

The transition from "authority control" and "authorized access points" in our legacy systems to identity management requires us to separate identifiers from their associated labels. A unique identifier could be associated with an aggregate of attributes that would enable users to distinguish one entity from another.[44] Ideally, libraries could take advantage of the identifiers and attributes from other, nonlibrary sources. Wikidata, for example, aggregates a variety of identifiers as well as labels in different languages, as shown in figure 4.

FIGURE 4. One Wikidata identifier links to other identifiers and labels in different languages

Providing contextual information is more important than providing one unique label. Labels could differ depending on communities—such as various spellings of names and terms, different languages and writing systems, and different disciplines—without requiring that one form be preferred over another. Label preference becomes localized rather than homogenized for global use.

A key barrier to moving from text strings to identity management is the lack of technology and infrastructure to support it. New tools are needed to index and display information about the entities described with links to the sources of the identifiers. Since multiple identifiers may point to the same entity, tools to reconcile them will also be needed. Some systems index only the controlled access points, which is a problem when dealing with names represented in different languages. Can library systems be reconfigured to deal with *identifiers* as the match point, collocation point, and the key to whatever associated labels are displayed and indexed?[45]

Some Focus Group members are experimenting with Wikidata as another option to assign identifiers for names not represented in authority files, which would broaden the potential pool of contributors.[46] Many libraries are looking toward Wikidata and Wikibase—the software platform underlying Wikidata—to solve some of the long-standing issues faced by technical services departments, archival units, and others.[47] Wikidata/Wikibase are viewed as a possible alternative to traditional authority control and have other potential benefits such as embedded multilingual support and bridging the silos describing the institution's resources. Focus Group members' experimentations with Wikidata and OCLC projects using the Wikibase platform indicate that Wikibase is a plausible framework for realizing linked data implementations. This infrastructure could enable the Focus Group and the wider bibliographic and archival communities to focus on the entities that need to be created, their relationship with each other, and how best they can increase discoverability by end-users.

> Identity management could also bridge the variations of names found in journal articles, scholarly profile services, and library catalogs, transcending these now siloed domains. This bridge is a requirement to fulfill the promises of linked data.

Because Wikidata was originally seeded by drawing data from Wikipedia, representation of books in Wikidata has a focus on "works" and their authors. This focus on works and authors could be viewed as an alternate version of the traditional author/title entries in authority files. Books that are "notable" are more likely to be represented in Wikidata. Recently, an effort to support citations in Wikipedia articles, WikiCite,[48] demonstrates a need to register and support identifiers that make up those citations, including information about a specific edition or document.

One of the most practical—and powerful—aspects of identity management is to reduce the amount of copying/pasting in library metadata workflows when an identifier is stewarded in an external location. Identifiers could provide a bridge between MARC and non-MARC environments and to nonlibrary resources. Librarians would not have to be the experts in all domains.[49] Many resources curated or managed by libraries are not under authority control, such as digital and archival

collections, institutional repositories, and research data. Identifiers could provide links to these resources. Identity management could also bridge the variations of names found in journal articles, scholarly profile services, and library catalogs, transcending these now siloed domains. This bridge is a requirement to fulfill the promises of linked data.

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

Concepts or subject headings are particularly thorny as terminology can differ depending on the time period and discipline. In some cases, terms may be considered pejorative, harmful, or even racist by some communities. Addressing language issues is important as libraries seek to develop relationships and build trust with marginalized communities. The issues around equity, diversity, and inclusion are complex, and the vocabulary used in subject headings is just one aspect, and language-neutral identifiers represent one approach.

The issue of supporting "alternate" subject headings came to the fore when the Library of Congress' initial solution to change the LC subject heading for "Illegal aliens" to "Undocumented immigrants" failed to be implemented. This prompted one Focus Group member to comment, "Being held hostage to a national system slow to change in the face of changing semantics is damaging to libraries, as generally we pride ourselves on being welcoming and inclusive." End-users hold their libraries accountable for what appears in their catalogs. Although LCSH is the *Library of Congress Subject Headings*, it is used worldwide, sometimes losing its context.[50]

Addressing language issues is important as libraries seek to develop relationships and build trust with marginalized communities.

Some see Faceted Application of Subject Terminology (FAST)[51] as a means to engage the community to mitigate the issues that have driven attempts to develop alternate subject headings for LCSH. A subset of the Focus Group has been applying FAST to records that would otherwise lack any subjects. FAST was originally developed by OCLC as a medium between totally-uncontrolled keywords at one end of the spectrum and difficult-to-learn-and-apply precoordinated subject strings at the other end.[52] FAST headings provide an easy transition to a linked data environment, since each FAST heading has a unique identifier. As FAST headings are generated from Library of Congress precoordinated subject headings, they can also include the same terminology that some consider inappropriate or disrespectful.

The recently launched FAST Policy and Outreach Committee[53] represents FAST users to oversee community engagement, term contributions, and procedures and to recommend improvements. Its vision statement reads:

> FAST will be a fully supported, widely adopted and community developed general subject vocabulary derived from LCSH with tools and services that serve the needs of diverse communities and contexts.[54]

Multiple overlapping and sometimes conflicting vocabularies already exist in legacy library data.[55] For example, Focus Group members in New Zealand add terms from the Māori Subject

Headings thesaurus (Ngā Upoko Tukutuku) to the same records as LC subject headings; Focus Group members in Australia add terms authorized in the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) Thesauri.[56] There may be no satisfactory equivalences across languages. Different concepts in national library vocabularies cannot always be mapped unequivocally to English concepts. The multiyear MACS (Multilingual Access to Subjects)[57] built relationships across three subject vocabularies: Library of Congress Subject Headings, the German GND integrated authority file, and the French RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié). It has been a labor-intensive process and is not known to be widely implemented.[58]

A growing percentage of data in institutions' discovery layers comes from non-MARC, nonlibrary sources. Metadata describing universities' research data and materials in Institutional Repositories is usually treated differently—and separately. How should institutions provide normalization and access to the entities described so users do not experience the "collision of name spaces" and ambiguous terms (or terms meaning different things depending on the source)? Synaptica Knowledge Solutions' Ontology Management – Graphite tool[59] to create and manage various types of controlled vocabularies seems promising in this context.

Focus Group members cited examples of established vocabularies or datasets that have become outdated or do not provide for local needs or sensibilities. Slow or unresponsive maintenance models for established vocabularies have tempted some to consider distributed models. High training thresholds to participate in current models have contributed to a desire for alternatives.[60] Linked data could provide the means for local communities to prefer a different label for an established vocabulary's preferred term for a concept or entity. One might reference a local description of a concept or entity not represented—or not represented satisfactorily—in established vocabularies or linked data sources. If these kinds of amendments and additions are made possible in a linked data environment, others could agree (or disagree) with the point of view by linking to the new resource. Such a distributed model for managing both terminology and entity description raises issues around metadata stability expectations, metadata interoperability, and metadata maintenance. How could a distributed model avoid people duplicating work on the same entity or concept? How would a distributed model record the trustworthiness of the contributors, or determine who would be allowed to contribute?

Educational institutions and libraries have undertaken EDI initiatives, and metadata departments have been struggling to support them.

Stability and permanence issues have been highlighted by the numerous vocabularies created for specific projects that, once funding ended, remain frozen in time. As one Focus Group member noted, "Nothing is sadder than a vocabulary that someone invented that was left to go stale." Such examples provide a major reason for librarians wanting to rely on international authority files rather than on local solutions. They also exemplify the value of the Library of Congress taking on the entire cost of creating and maintaining LCSH.

The OCLC Research report on the findings from a 2017 survey of the Research Library Partnership on equity, diversity, and inclusion (EDI)[61] spurred discussions on the complexity of embedding

equity, diversity, and inclusion in controlled vocabularies in library catalogs.[62] Educational institutions and libraries have undertaken EDI initiatives, and metadata departments have been struggling to support them. The excerpt from the EDI survey in figure 5 shows that metadata in library catalogs lags behind other areas in support of the institution's EDI goals and principles.

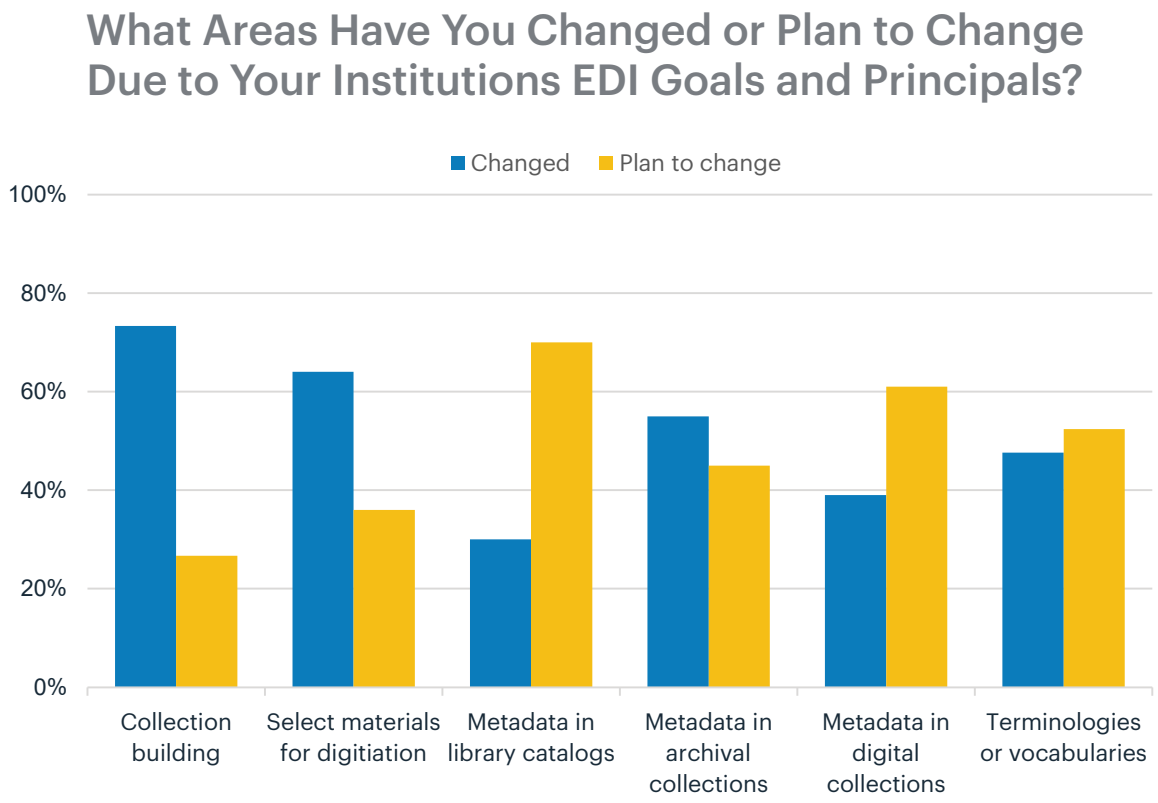


FIGURE 5. Excerpt from the survey results from the 2017 EDI survey of the Research Library Partnership[63]

Focus Group members are eager to provide more detailed subject access than is currently offered by national subject heading systems, such as LCSH, which has more granularity for Western European places than for Southeast Asia and Africa. They see the need to offer more accurate and current terms and replace terms that reflect bias or are considered offensive with more neutral terms.

Challenges that Focus Group members identified in offering more respectful terminology in subject access for users:

- **Discovery:** Using other, less-offensive vocabularies locally can split collections in the library catalog, thus hampering discovery of all relevant materials.
- **Lack of consensus:** Focus Group members doubt that there can ever be complete consensus about any given text string. Terms that may be offensive to one community may not always be clear to others. (For example, "Dissident art" rather than "Non-conformist art.")
- **Speed:** The process of changing standard subject headings can be very slow.
- **Capacity:** Changing headings in existing records can require a massive undertaking. Targeted access point maintenance occurs in the context of access point maintenance generally. For example, the Library of Congress recently changed the heading "Mentally handicapped" to

"People with mental disabilities." Implementing such changes in the catalog can involve a mix of automated, vended, and manual remediation methods, as well as decisions about resource allocation.[64] Some noted it would be less labor-intensive to present a "cultural sensitivity" message as part of the search interface to alert users that terms and annotations they find in a catalog may reflect the creator's attitude or the period in which the item was created and may be considered inappropriate today in some contexts.

- **Sharing**: Local vocabularies cannot be shared with other systems.
- **Maintenance:** Some who have tried to use local vocabularies more suitable for their context and communities found them too burdensome to maintain and abandoned them.
- **Language barriers:** The language of our controlled vocabularies may be exclusive to audiences who do not read that language. The Ohio State University Libraries has tried to address this by developing some non-Latin script equivalents of English subject terms.
- **Classification:** Current classification systems are apt to segregate ethnic groups. Rather than include them as part of an overall concept like history, education, or literature, they tend to be grouped together as one lump. As institutions store more publications off-site, the need to shelve materials together and have just one classification in a record has subsided, but few apply multiple classifications in one record.

Requirements for a distributed system that accommodates multiple vocabularies and could also support EDI converged around the need to support semantic relationships among different vocabularies. Communities of practice need a hub to aggregate and reconcile terms within their own domains. It was noted that different communities of practice might use terms that conflict with others' terminologies or mean different things. The PCC Linked Data Advisory Committee's Linked Data Infrastructure Models: Areas of Focus for PCC Strategies[65] describes high-level functional requirements and a spectrum of models anticipated as cultural heritage institutions adopt linked data as a strategy for data sharing. The model must be both scalable and extensible, with the ability to accommodate the proliferation of new topics and terms symptomatic of the humanities and sciences and facilitate contributions by the researchers themselves. It needs to be flexible enough to coexist with other vocabularies.

Replacing text strings with stable, persistent identifiers would facilitate using different labels depending on context. This would accommodate both different languages and scripts (and different spellings within a language, such as American vs. British English), as well as terms that are more respectful to marginalized communities. The 19 October 2017 OCLC Research Works in Progress webinar on "Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries, and Archives"[66] described the process launched by the Association for Manitoba Archives and the University of Alberta Libraries to examine subject headings and classification schemes and consider how they might be more respectful and inclusive of the experiences of Indigenous peoples.

Expanding vocabularies to include those used in other communities requires building trust relationships. A model of "community contribution" for new terms and community voting could be more inclusive. Libraries' current "consensus environment" excludes a lot of people. Much metadata is currently created according to Western knowledge constructs, and systems have been designed around them. Communicating the history of changes and the provenance of each new or modified term would provide transparency that could contribute to the trustworthiness of the source. The edit history and discussion pages that are part of each Wikidata entity description is a possible model to follow. Requiring provenance as part of a distributed vocabulary model may help in creating an alternative environment that is more equitable, diverse, and inclusive.

LINKED DATA CHALLENGES

Identifiers and vocabularies are just two components required in the transition to linked data. A vital part of describing entities are the associated *statements* made. How will libraries resolve or reconcile conflicts between statements?[67] Different types of inconsistencies may appear than do now with, for example, different birthdates for persons. The provenance of each statement becomes more critical. Even in the current environment, certain sources are more trusted and give catalogers confidence in their accuracy. Libraries often have a list of "preferred sources."[68] OCLC Research explored how libraries might apply Google's "Knowledge Vault" to identify statements that may be more "truthful" than others in the 2015 "Works in Progress Webinar: Looking Inside the Library Knowledge Vault."[69] Focus Group members posited that aggregations such as WorldCat, the Virtual International Authority File (VIAF), and Wikidata may allow the library community to view statements from these sources with more confidence than others. Librarians could share their expertise by establishing the relationships between and among statements from different sources.

But good linked data requires good metadata. Administrators are well aware of the tension between delivering access to library collections in a timely manner and providing good quality description. The metadata descriptions must be full enough to allow libraries to manage their collections and to support accessibility and discoverability for the end-user. Many libraries need to compromise between speed over accuracy, speed over depth, or brevity over nothing. These compromises are reflected by using inadequate vendor records, by creating minimal or less-than-full level descriptions for certain types of resources, and by limiting authority work. Minimal-level cataloging is commonly used as an alternative to leaving materials uncatalogued, often because of large volume of materials and insufficient staff resources.[70] These less-than-full descriptions will result in fewer and less accurate linked data statements.

> Good linked data requires good metadata.

The transition period from legacy cataloging systems reliant on MARC to a new linked data environment with entities and statements has many challenges since both standards and practices are moving targets. It is unclear how libraries will share statements rather than records in a linked data environment. Focus Group members were divided on whether a centralized linked data store would be needed to provide "trustworthy provenance" or whether data should be distributed with peer-to-peer sharing.[71] Different statements might be correct in their own contexts. "Conflicting statements" might represent different world views. Selecting statements based on provenance could be challenging to our principles of equity, diversity, and inclusion.

The Focus Group members wondered how to involve the many vendors that supply or process MARC records in the transition to linked data. In the United Kingdom, the Jisc initiative "Plan M" (where "M" stands for "metadata") seeks to streamline the metadata supply among libraries, publishers, data suppliers, and infrastructure providers.[72] Among the implications cited by stakeholders in the UK's National Bibliographic Knowledgebase (NBK) in Plan M's 10-year vision: "Linked data instances of the NBK will need to be created and maintained requiring convincing business-cases around the impact this could have on research."[73] Working with others in the linked data environment involves people unfamiliar with the library environment, requiring metadata specialists to explain what their needs are in terms nonlibrarians can understand.

Describing "Inside-Out" and "Facilitated" Collections

OCLC Vice President and Chief Strategist, Lorcan Dempsey, refers to the shifting emphasis of libraries to support the creation, curation, and discoverability of institutional resources as the "inside-out collection" (in contrast to the "outside-in collection," in which the library buys or licenses materials from external providers to make them accessible to a local audience). Providing access to a broader range of local, external, and collaborative resources around user needs is the "facilitated collection."[74] Focus Group members' activities have increasingly focused on metadata that will provide access to the resources unique to their institutions as well as those in their consortia or national networks.

All resources collected, created, and curated by libraries require metadata to make them discoverable. However, Focus Group members concentrated on the challenges and issues related to specific formats:

- Archival collections
- Archived websites
- Audio and video collections
- Image collections
- Research data

All these content types can be categorized as belonging to "inside-out" collections and present different challenges. For example, Focus Group members described efforts to retrieve metadata from completely different systems as "super challenging." In addition, many of these resources are not under any authority control. Reconciling access points from various thesauri and metadata mapping work requires technical services expertise and skills.[75] This reconciliation also will be needed in the previously discussed linked data environment.

This section summarizes the discussions on these format types.

ARCHIVAL COLLECTIONS

Archival collections are in many ways the crown jewels of collections as they are unique research resources providing insights into the world across many centuries and places, providing the primary sources for new knowledge creation. Increasing visibility for these collections reaps significant benefits for both scholars and libraries and archives. Archives are, however, complex and present different metadata issues compared with traditional library collections. As institutions turn to ArchiveSpace and other content management systems to provide infrastructures for structured archival metadata, various issues are emerging.[76]

Archives have had more autonomy than libraries within their institutions because they have unique collections with their own population of users, their own metadata standards, and their own systems. While some institutions have integrated archival processing within technical services, most maintain a separate unit. Archivists do not have the tradition of creating authority records and sharing identifiers for the same entity as is common among librarians. They also tend to use the fullest form of a name based on the information found in collections, while librarians focus on "preferred" form found in publications. Even so, a significant shift from artisanal archival approaches to metadata standardization has been occurring.

So how can archivists and librarians better integrate their metadata and name authority practices? The number of personal names in archival collections can be so large that most are uncontrolled and without identifiers. However, the contextual information that archivists provide for person and organization entities could enrich the information provided in authority files—a use case that was explored in the 2017-2018 Project Passage pilot[77] and examined in more detail in 2019-2020 by the OCLC Research Library Partners Archives and Special Collections Linked Data Review Group.[78]

The increased reliance on electronic and digital resources during the COVID-19 pandemic will likely accelerate institutions digitizing their archival and distinctive collections that have been available only in physical form.[79] More metadata may be created from digitized versions of these resources.

ARCHIVED WEBSITES

For some years, 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.[80]

Some archived websites are institutional, theme-based collections supporting a specific research area such as Columbia University's Human Rights, Historic Preservation and Urban Planning, and New York City Religions.[81] National libraries archive websites within their national domain. For example the National Library of Australia's Archived websites (1996-now)[82] collect websites in partnership with cultural institutions around Australia, government websites formerly accessible through the Australian Government Web Archive, and websites from the .au domain collected annually through large scale crawl harvests. These curated collections by subject provide snapshots of Australian cultural and social history. Examples of consortia-based archived websites include the Ivy Plus Libraries Confederation's Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY) and Contemporary Composers Web Archive (CCWA) and the New York Art Resources Consortium (NYARC), which captures dynamic web-based versions of auction catalogs and artist, gallery, and museum websites.[83]

The Focus Group discussed the challenges for creating and managing the metadata needed to enhance machine-harvested metadata from websites. Some of the challenges identified:

- **Type of website matters.** Descriptive metadata requirements may depend on the type of website archived (e.g., transient sites, research data, social media, or organizational sites). Sometimes only the content of the sites is archived when the user experience of the site (its "look-and-feel") is not considered significant.
- **Practices vary.** Some characteristics of websites are not addressed by existing descriptive rules such as RDA (*Resource Description and Access*) and DACS (*Describing Archives: A Content Standard*). Metadata tends to follow bibliographic description traditions or archival practice depending on who creates the metadata.
- **Consider scale and projected use.** Metadata requirements may differ depending on the scale of material being archived and its projected use. For example, digital humanists look at web content as data and analyze it for purposes such as identifying trends, while other users merely need individual pages. The level of metadata granularity (collection, seed/URL, document) may also vary based on anticipated user needs, scale of material being crawled, and available staffing.

- **Update frequency.** Many websites are updated repeatedly, requiring re-crawling when the content has changed. Some types of change can result in capture failures.
- **Multi-institutional websites.** Some websites are archived by multiple institutions. Each may have captured the same site on different dates and with varying crawl specifications. How can they be searched and used in conjunction with one another?

A 2015 survey of the OCLC Research Library Partnership revealed the "lack of descriptive metadata guidelines" as the biggest challenge related to website archiving, leading to the formation of the OCLC Research Library Partnership Web Archiving Metadata Working Group."[84] The challenges that the Focus Group identified were explored in depth by this working group, which issued a report of its recommendations in 2018, *Descriptive Metadata for Web Archiving*.[85]

AUDIO AND VIDEO COLLECTIONS

Focus Group members reported that their institutions had repositories filled with large amounts of audiovisual (A/V) materials, which often represent unique, local collections.[86] However, as Chela Scott Weber states in the publication *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries*, "For decades, A/V materials in our collections were largely either separated from related manuscript material (often shunted away to be dealt with at a later date) or treated at the item level. Both have served to create sizeable backlogs of un-quantified and un-described A/V materials."[87] Much of this audiovisual material urgently requires preservation, digitization, clarification of conditions of use, and description.

In addition, the needed skill sets and stakeholders across institutions are complex. The nature of the management of A/V resources requires knowledge of the use context as well as technical metadata issues, providing a complex environment to think through requirements for description and access. Further, libraries must deal with current time-based media that is either being produced locally as part of research and learning, or streaming media that is being commercially licensed.

Focus Group discussions focused on the A/V resources within archival collections—often in deteriorating formats, in large backlogs, and sometimes requiring rare and expensive equipment to access and assess the files. For locally generated content, institutions prefer that the creators describe their own resources.

Metadata describing the same A/V materials may differ across library, archival, and digital asset management systems.

The overarching challenge was how much effort needs to be invested in describing these A/V materials because they are unique. Institutions have used hierarchical structures to aggregate similar materials with finding aids that are marked up in the Encoded Archival Description standard,[88] which provides useful contextual information for individual items within a specific collection. But often an aggregated approach to description can lack important details about individual items needed for discovery, such as transcribed title and date broadcast. This is a particularly acute issue for legacy data describing recordings from years past. Metadata describing the same A/V materials may differ across library, archival, and digital asset management systems.

Some hope that better discovery layers will alleviate the need to repeat the same information across databases, but presenting the information to users would require using consistent access points across systems. The same will be true in a linked data environment. But the challenge to link between items and the finding aid and to maintain the links over time despite changes in systems will remain.

Metadata for A/V materials needs to include important technical information, such as details about the A/V capture and digitization process like compression, year digitized, the technology used, and file compatibility. This data is critical to ensure perpetual access for such enormous files and mercurial playback formats. Some Focus Group members have implemented PREMIS (Preservation Metadata: Implementation Strategies),[89] the international standard for metadata to support the preservation of digital objects and ensure their long-term usability, for some of their A/V materials.

OCLC Senior Program Officer Chela Scott Weber continues working with the Research Library Partnership on the needs and challenges of managing A/V collections, summarized in *OCLC Research Hanging Together Blog* posts: "Assessing Needs of AV in Special Collections" and "Scale & Risk: Discussing Challenges to Managing A/V Collections in the RLP."[90] A subset of the Focus Group members responded to Weber's 2019 survey to assess the needs of audiovisual materials in special collections within the Research Library Partnership; incorporating A/V collections into archival and digital collections workflows were two of the challenges that most interested respondents, as shown in figure 6.

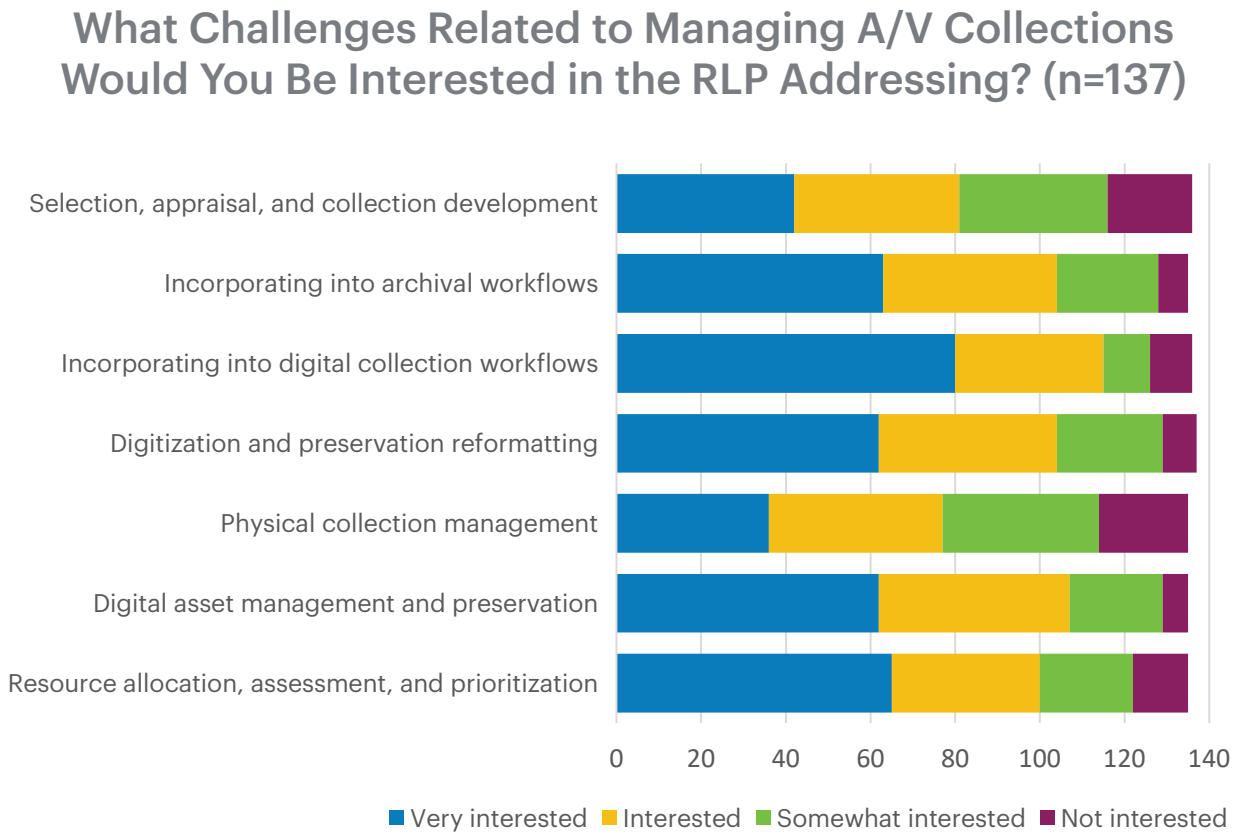


FIGURE 6. Responses to 2019 survey on challenges related to managing A/V collections

IMAGE COLLECTIONS

Focus group members manage a wide variety of image collections presenting 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"). Image collections also offer opportunities for crowdsourcing and interdisciplinary research.[91]

Many libraries describe their digital image resources on the collection level while selectively describing items. As much as possible, enhancements are done in batch. Some do authority work, depending on the quality of the accompanying metadata. Some libraries have disseminated metadata guidelines to help bring more consistency to the data.

Among the challenges discussed by the Focus Group:

- **Variety of systems and schemas:** Image collections created in different parts of the institution such as art or anthropology departments serve different purposes and use different systems and schemas than those used by the library. The metadata often comes in spreadsheets or unstructured accompanying data. Often, the metadata created by other departments requires much editing, massaging, and manual review. The situation is simpler when all digitization is handled through one centralized location and the library does all the metadata creation. Some libraries are using Dublin Core for their image collections' metadata and others are using MODS (Metadata Object Description Schema).[92] Some wrap the metadata records in METS (Metadata Encoding and Transmission Standard),[93] a schema maintained by the Library of Congress designed to express the hierarchical nature of digital library objects, the names and locations of the files that comprise those objects, and the associated metadata. Some suggested that MODS be used in conjunction with MADS (Metadata Authority Description Schema).[94]
- **Duplicate metadata for different objects:** Metadata for a scanned set of drawings may be identical, even though there are slight differences in those drawings. Duplicating the metadata across similar objects is likely due to limited staff. Possibly the faculty or the photographers could add more details.
- **Lack of provenance:** A common challenge is receiving image collections with scanty metadata and with no information regarding their provenance. For example, metadata staff at one institution were given OCR'ed text retrieved by a researcher from HathiTrust. Millions of images lacked the location of the original source material and therefore limited—if not discredited—any further use.
- **Maintaining links between metadata and images:** How should libraries store images and keep them in sync with the metadata? There may be rights issues from relying on a specific platform to maintain links between metadata and images. Where should thumbnails live?
- **Relating multiple views and versions of same object:** Multiple versions of the same object taken over time can be very useful for disciplines like forensics. For example, Brown University decided to describe a "blob" of various images of the same thing in different formats and then describe the specific versions included. This work was done even though there is no system yet that displays relationships among images, such as components of a piece, even when the metadata in records are wrapped and stored in METS.

- **Managing relationships with faculty and curators:** It is important to ensure that faculty feel their needs are met. Collaboration is necessary among holders of the materials, metadata specialists, and developers as all come from different perspectives. The challenge is to support both a specific purpose and groups of people as well as large-scale discovery.
- **Aggregating digital collections:** Institutions have been 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. Targets for sharing this metadata range from tailored topic-based digital discovery services to national and international aggregations such as Google Scholar, HathiTrust, Digital Public Library of America (DPLA), Internet Archive, Trove, and WorldCat to online exhibitions such as Google Arts and Culture or image banks such as Flickr or Unsplash. Such aggregations can help inform an institution's own collection development, as librarians can see their contributions in the context of others' content and identify gaps that they may wish to fill locally.[95]

Aggregators often have different guidelines and input formats. Aggregators' very reasonable contention that they cannot support many variations in submitted metadata conflict with contributors' very reasonable contention that they cannot support the different needs of a wide range of aggregators. Disseminating corrections or updates between the source and the aggregation can be problematic. Information that may have been corrected in the chain leading to incorporation in the aggregation may not be pushed back to the source, so that the same errors must be corrected repeatedly. It is often not clear what data elements have been updated, when, or by whom.

Aggregating images and bringing together different images or versions of the same object was the goal of the 2012-2013 OCLC Research Europeana Innovation Pilots,[96] which developed a method for hierarchically structuring cultural objects at different similarity levels to find "semantic clusters"—those that include terms with a similar meaning. In 2017, OCLC implemented the International Interoperability Image Framework (IIIF)[97] Presentation Manifest protocol in its CONTENTdm digital content management system, an aggregation containing more than 70 million digital records contributed by over 2,500 libraries worldwide. In 2019 OCLC Research developed an IIIF Explorer experimental prototype for testing and evaluation that searches across all the CONTENTdm images using the IIIF Presentation Manifest protocol,[98] as shown in figure 7. Aggregating content across IIIF-compliant systems may facilitate discovery across the plethora of platforms containing digital content mentioned above.

In 2020, OCLC Research launched the CONTENTdm Linked Data Pilot,[99] focused on developing scalable methods and approaches to produce machine-readable representations of entities and relationships and make visible the connections formerly invisible. Existing record-based metadata is being converted to linked data by replacing strings of characters with identifiers from known authority files and local library-defined vocabularies; the resulting graphs of entities and relationships can retrieve contextual information from sources such as GeoNames and Wikidata. This pilot (to be completed by August 2020) is addressing many of the above challenges identified by the Focus Group.

The OCLC ResearchWorks IIIF Explorer Retrieves Images about "Paris Maps" across CONTENTdm Collections

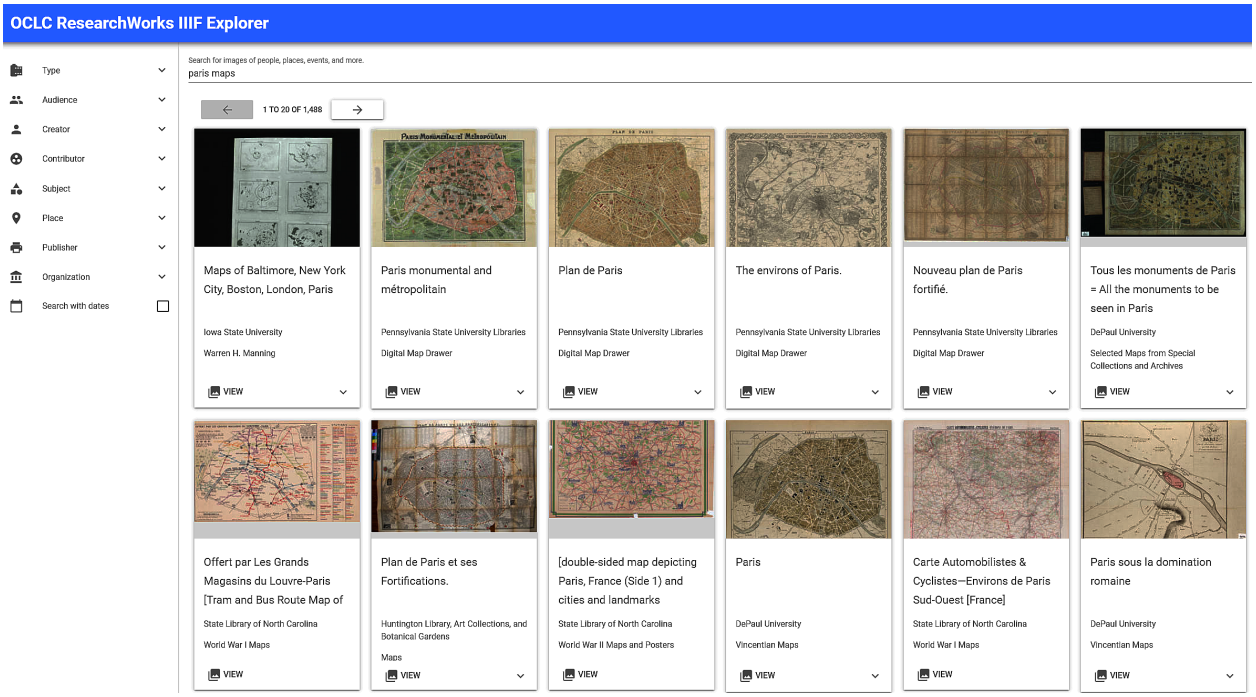


FIGURE 7. The OCLC ResearchWorks IIIF Explorer retrieves images about "Paris Maps" across CONTENTdm collections

RESEARCH DATA

Research funders expect that the research data resulting from research they support will be archived and made available to others. Institutions have allotted more resources to collecting and curating this scholarly resource for reuse within the scholarly record. OCLC Research Scientist Ixchel Faniel's two-part blog entry "Data Management and Curation in 21st Century Archives" (Sept 2015)[100] prompted the discussion among Focus Group members on the metadata needed for research data management.[101] 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.

Metadata is important for both discovery and reuse of datasets. The 2016 OCLC Research report *Building Blocks: Laying the Foundation for a Research Data Management Program* noted:

> Datasets are useful only when they can be understood. Encourage researchers to provide structured information about their data, providing context and meaning and allowing others to find, use and properly cite the data. At minimum, advise researchers to clearly tell the story of how they gathered and used the data and for what purpose. This information is best placed in a readme.txt file that includes project information and project-level metadata, as well as metadata about the data itself (e.g., file names, file formats and software used, title, author, date, funder, copyright holder, description, keywords, observation unit, kind of data, type of data and language).[102]

All four of the of 2017-2018 *The Realities of Research Data Management* series webinars[103] led by OCLC Senior Program Officer Rebecca Bryant mention the importance of metadata. Research information infrastructure calls on many of the key strengths of the library profession. Metadata is fundamental to our complex research environment—beginning with the planning our researchers do before and during the creation of data; to managing the data; then to disseminating the knowledge gained; finally through to understanding the impact, engagement, and the resulting reputation of our home institutions.[104]

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. Faniel highlighted this value in the November 2019 *Next* blog post "Let's Cook Up Some Metadata Consistency":

> [C]ataloging for discovery using terms and definitions that are consistent across repositories is critical, if we want the data and their associated metadata to be discoverable for reuse in any way imaginable. . . . Librarians and archivists can help create consistencies in metadata that build bridges between researchers and repositories, thus greatly increasing the discovery, reuse, and value of their institutions' research investments.[105]

National contexts differ. For example, our Australian colleagues can take advantage of Australia's National Computational Infrastructure for big data and the Australian Data Archive for the social sciences.[106] Canada has launched a national network called Portage for the "shared stewardship of research data."[107]

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.

Some institutions have developed templates to capture metadata in a structured form. Some Focus Group members noted the need to keep such forms as simple as possible as it can be difficult to get researchers to fill them in. All agreed data creators needed to be the main source of metadata. But what will inspire data creators to produce quality metadata? New ways of training and outreach are needed, an area of exploration within Metadata 2020's Research Communications project.[108]

Focus Group members generally agreed on the data elements required to support reuse: licenses, processing steps, tools, data documentation, data definitions, data steward, grant numbers, and geospatial and temporal data (where relevant). Metadata schema used includes Dublin Core, MODS (Metadata Object Description Schema) and DDI (Data Documentation Initiative's metadata standard). The Digital Curation Centre in the UK provides a linked catalog of metadata standards.[109] The Research Data Alliance's Metadata Standards Directory Working Group has set up a community-maintained directory of metadata standards for different disciplines.[110] The disparity of metadata schemas across disciplines represents a hurdle in institutions' discovery layers.

The importance of identifiers for both the research data and the data creator(s) has become more widely acknowledged. DOIs, Handles and ARKs (Archival Resource Key) have been used to provide persistent access to datasets. Identifiers are available at the full data set level and for component parts, and they can be used to track downloads and potentially help measure impact. Both ORCID and ISNI are in use to identify data creators uniquely, and work is continuing on the Research Organizational Registry to address institutional affiliations.

Among the most critical issues identified by Focus Group members is that metadata specialists need to be more involved in the early stages of the research life cycle. Researchers need to understand the importance of metadata in their data management plans. The lack of "metadata governance" across an institution makes integrating workflows between repositories and discovery layers problematic.

Some Focus Group members have started to analyze the metadata requirements for the research data life cycle, not just the final product, asking questions like: Who are the collaborators?[111] How do various projects use different data files? What kind of analysis tools do they use? What are the relationships of data files across a project, between related projects, and to other scholarly output such as related journal articles? Research support services such as those offered at the University of Michigan[112] are being developed to assist researchers during all phases of the research data life cycle, often through collaboration with other campus units.

Among the most critical issues identified by Focus Group members is that metadata specialists need to be more involved in the early stages of the research life cycle. Researchers need to understand the importance of metadata in their data management plans. The lack of "metadata governance" across an institution makes integrating workflows between repositories and discovery layers problematic.

Some libraries have started to provide research data management support in a variety of ways. For example, metadata specialists work with their institutions' Scholarly Communications and Publishing Division which also manages the Institutional Repository. These institutional repositories may have only the "citation" or "metadata-only" records with a link to the full text or data set deposited in a disciplinary repository. "Metadata consultation services" may be provided to advise on the data management plan, which includes appropriate metadata standards and controlled vocabularies, a strategy to effectively organize their data, and an approach that will facilitate reuse of the data years after the research is completed. The OCLC Research *The Realities of Research Data Management* report series classifies metadata support as part of the "expertise" function, and flags some variations in its case studies.[113] At the University of Illinois at Urbana-Champaign, metadata consultants help researchers with metadata regardless of where the research data is deposited; Monash University supports metadata curation only for local deposits.[114]

Communication is key for researchers to understand the importance of metadata throughout the research life cycle. Some universities offer "research sprints" where researchers partner with a team of expert librarians that may include metadata creation, management, analysis, and preservation. The "Shared BigData Gateway for Research Libraries," hosted by Indiana University and partially funded by the Institute of Museum and Library Services, is developing a cloud-based platform to share data and expertise across institutions, including datasets such as records from the US Patent and Trademark Office and the Microsoft Academic Graph.[115]

Curation of research data as part of the evolving scholarly record requires new skill sets, including deeper domain knowledge and experience with data modeling and ontology development. Libraries are investing more effort in becoming part of their faculty's research process and are offering services that help ensure that their research data will be accessible if not also preserved. Good metadata will help guide other researchers to the research data they need for their own projects, and the data creators will have the satisfaction of knowing that their data has benefitted others.[116]

Evolution of "Metadata as a Service"

Metadata underlies the ability to discover all resources in the inside-out and facilitated collections. Focus Group members anticipate more involvement with metadata creation beyond the traditional library catalog and new services that leverage both legacy and future metadata.

METRICS

Library strategic goals often include key phrases such as "foster discovery and use," "enrich the user experience," and "explore new ways to support the whole life cycle of scholarship," all of which is predicated on quality metadata. Usage metrics—such as how frequently items have been borrowed, cited, downloaded, or requested—could be used to build a wide range of library services and activities. Focus Group members identified some possible services: informing collection management decisions about weeding projects and identifying materials for offsite storage; evaluating subscriptions; comparing citations for researchers' publications with what the library is not purchasing; and improving relevancy ranking, personalizing search results, offering recommendation services in the discovery layer, and measuring impact of library usage on research or student success or learning analytics.[117] The University of Minnesota conducted a study to investigate the relationships between first-year undergraduate students' use of the academic library, academic achievement, and retention.[118] The results suggest a strong correlation between using academic library services and resources—particularly database logins, book loans, electronic journal logins, and library workstation logins—and higher grade point averages. In the United Kingdom, the Jisc Library Impact Data Project found a similar correlation.[119]

CONSULTANCY

Metadata's value is demonstrated by integrating it into the fabric of both the library and other units across the campus. For example, metadata specialists can provide "metadata as a service"—consultancy in the earliest stages of both library and research projects.[120] An emerging trend is for digital humanities departments to request advice from metadata specialists on metadata standards and how to use controlled vocabularies. More visibility of this metadata consultant role appears in recent library job postings. In one Metadata Librarian job posting at Cornell,[121] one duty cited was 20% for "metadata outreach and consultation": "Maintains strong working relationships and communicates regularly with staff across Cornell, fostering collaborative efforts between Metadata Services and the greater Cornell community." Georgia Tech is recruiting a metadata librarian who

will "serve as a metadata consultant to larger library projects/initiatives. Work closely with other Library departments, Emory University Libraries, GALILEO, University System of Georgia Libraries, and other partners involved in joint projects."[122]

NEW APPLICATIONS

The shared and consistent use of MARC fields supports new applications. Libraries currently use identifiers in bibliographic records to fetch tables of contents, abstracts, reviews, and cover images and to generate floor maps of where to locate resources in a specific classification range (such as in OCLC's integration with StackMap).[123] Bibliographic metadata is used to populate Digital Asset Management Systems and Institutional Repositories, and with tools such as Tableau and OpenRefine, can enable a richer analysis of collections and a view of collections. MARC metadata is connecting scholars with the bibliographic data for their projects and can generate relationships to related resources with applications such as Yewno.[124] MARC metadata is also being used to inform institutional output measures and affiliation tracking and serves as a source to build organization histories. The provenance implicit in an institution's bibliographic metadata has proven helpful in documenting theft cases. Analyzing catalog data by data mining can also be used to enrich the metadata, such as generating language codes missing in related records or identifying the original titles of translated works. MARC data has also supported generating subject maps to discover relationships otherwise not explicit in the cataloging metadata.[125]

Visualizations represent another type of metadata service. A striking example is from the Auslang national codeathon held in 2019, a collaboration among the National Library of Australia, the Australian Institute of Aboriginal and Torres Strait Islander Studies, Trove, Libraries Australia, and the State and Territory libraries—a national code-a-thon to identify items in Indigenous Australian languages.[126] Figure 8 shows the results, a map indicating the 465 Indigenous languages in the Australian National Bibliographic Database tagged as a result of the code-a-thon, and an example of involving the community to enhance bibliographic metadata.

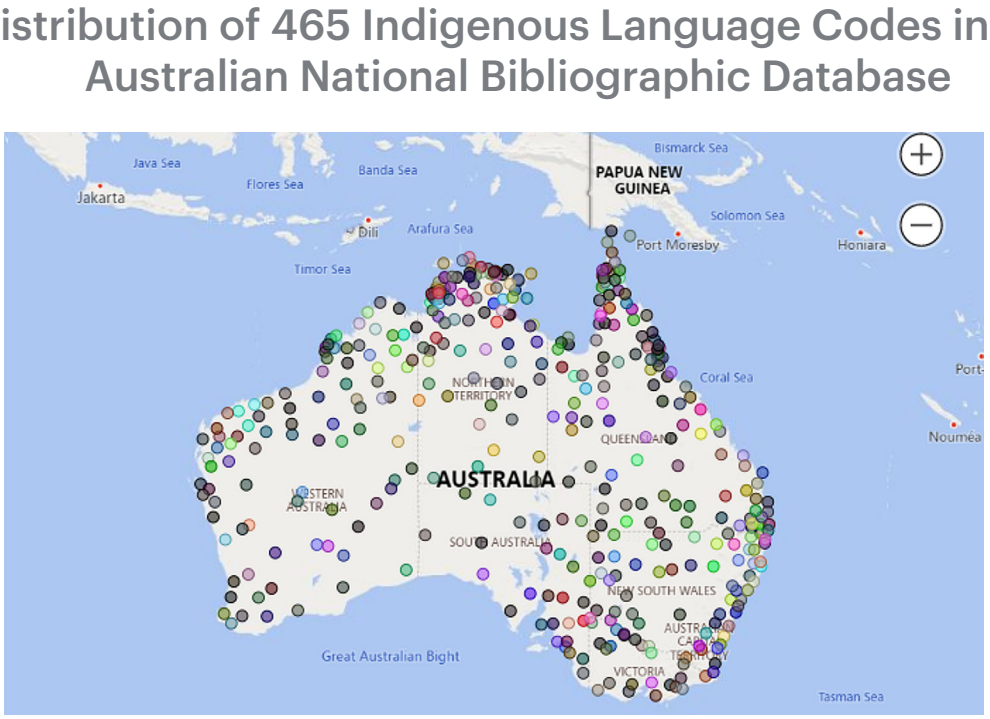


FIGURE 8. Distribution of 465 Indigenous language codes in the Australian National Bibliographic Database

BIBLIOMETRICS

Library metadata is also being used to generate bibliometrics, statistical methods to analyze books, articles, and other publications. Using library metadata for Digital Humanities research projects has much potential. For example, a Library of Congress researcher used bibliographic metadata to trace the history of publishing and copyright; UCLA researchers have used cataloging metadata to track the commercialization of inventions such as insulin.

A novel use of cataloging metadata was by Hachette UK, the United Kingdom's second largest bookseller, which commissioned the Graphic History Company to unlock the histories of all nine of Hachette's publishing houses and weave them into a cohesive story by asking the British Library for every author and book title published by their nine publishing houses spanning 250 years. The British Library provided a list of over 55,000 authors, from which 5,000 of the most prominent were selected to create perhaps the most beautiful example of metadata use: a giant mural spanning eight floors featuring all 5,000 authors in chronological order. (Figure 9 shows one part of the mural; for more images of the mural, see Hachette's River of Authors.)[127]

UK Hatchette's "River of Authors" Generated from the British Library's Catalog Metadata

FIGURE 9. UK Hatchette's "River of Authors" generated from the British Library's catalog metadata

SEMANTIC INDEXING

When controlled vocabularies and thesauri are converted into linked open data and shared publicly, their traditional role of facilitating collection browsing will fade but could find a renewed purpose within web-based knowledge organizations systems (KOS).[128] As Marcia Zeng points out in Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review,

> a KOS vocabulary is more than just the source of values to be used in metadata descriptions: by modeling the underlying semantic structures of domains, KOS act as semantic road maps and make possible a common orientation by indexers and future users, whether human or machine.[129]

Good examples of such repurposing are the Getty Vocabularies that allow browsing of Getty's representation of knowledge and also helps users generate their own SPARQL queries that can be embedded in external applications. Another example is Social Networks and Archival Context (SNAC),[130] which enables browsing of entities and relationships independently of their collections of origins. In such cases, the discovery tool pivots to being person-centric (or family-centric, or topic-centric), rather than (only) collection-centric.

Rather than one "global domain," metadata specialists could provide added value by adding bridges from the metadata in library domain databases to other domains. Wikidata is an example of a platform aggregating entities from different sources and linking to more details in various language Wikipedias. Some institutions have employed Wikimedians in Residence to accelerate this process.

Focus Group members hope that Artificial Intelligence—or at least machine-learning—could mitigate the amount of current manual effort to link names and concepts in research data. Perhaps algorithms could be used to match names based on related metadata or sources, relate topics to each other based on context, disambiguate names based on other metadata available, and analyze datasets to identify possible biases in a collection.[131] A few Research Library Partners participate in Artificial Intelligence for Libraries, Archives & Museums (AI4LAM),[132] an "international, participatory community focused on advancing the use of artificial intelligence in, for and by libraries, archives, and museums."[133] Some high-level recommendations on enhancing descriptions at scale and improving discovery are noted in Thomas Padilla's OCLC Research 2019 position paper *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*.[134]

Preparing for Future Staffing Requirements

The anticipated changes from transitioning to the next generation of metadata will also shift staffing requirements to prepare for the future. Focus Group members identified new skill sets needed for both professionals entering the field as well as seasoned catalogers, driven by the changing information technology landscape and increasing staff attrition. Focus Group members characterized professionals as those who "trail-blaze innovations," which are then routinized for nonprofessionals. These discussions reinforce Padilla's recommendations on investigating core competencies, committing to internal talent, and expanding evidence-based training.[135]

THE CULTURE SHIFT

Focus Group members reported a delicate balance of allocating staff to "traditional cataloging activities" (such as original and copy cataloging, authority work) with more exploratory R&D projects, such as linked data projects, exploring new data models and technologies such as Wikidata, and learning about emerging standards and identifiers. A ***culture shift*** is needed: from pride in production alone to valuing opportunities to learn, explore, and try new approaches to metadata work. Metadata specialists must understand that improving *all* metadata is more important than any individual's productivity numbers. This culture shift requires buy-in from administrators to support training programs for staff to learn new workflows for processing multiple formats and to view metadata specialists as more than just "production machines."

Metadata managers faced with staff reductions while still being expected to maintain production levels must justify allocating staff time for R&D—or "play time"—to explore such questions as: What can we stop doing? What is the one thing you learned that we all need to do more of? What do you need to move forward? What open source software could help us do the work more efficiently? What new methods could enhance discoverability, access, and use of our facilitated collections? Managers must incorporate goals for success that are not based solely on numbers.[136]

A *culture shift* is needed: from pride in production alone to valuing opportunities to learn, explore, and try new approaches to metadata work.

Indications of this culture shift include institutions outsourcing some metadata work or training support staff to create metadata for the "easier stuff" while mandating that catalogers only do what well-trained humans can do. Metadata managers could scope the materials requiring metadata that support staff or students can handle, providing templates where possible. If you remove these tasks, the majority of what remains requires highly skilled metadata specialists with expertise in languages, physical formats, and disambiguating and describing persons, organizations, and other entities.

LEARNING OPPORTUNITIES

To encourage the culture shift among metadata specialists to change their mindsets about how they work and stimulate interest in learning opportunities, Focus Group members have used several approaches:

- Identify who on your team has the aptitude to acquire new skills. At one institution, the staff member shared what she learned and the whole unit became "lively" because she brought her colleagues along. It created appreciation for "continuous learning" and staff presented their activities at national conferences.
- Convene cross-team group discussions to look at problem metadata and come up with solutions, encouraging staff to move forward together. Staff less interested in new skills can pick up some of the production from those learning new skills and producing less.
- Launch "reading clubs" where staff all read an article and respond to three discussion questions to inspire metadata specialists to think about broader metadata issues outside of their daily work.
- Hold weekly group "video-viewing brown-bag lunches" for staff on new developments such as linked data so staff can "watch and learn" together.
- Participate in multi-institutional projects to collaborate with peers to solve problems and cross-pollinate ideas.
- Encourage participation in professional conferences and standards development.

Educating and training catalogers has been at the forefront of many discussions in the metadata community. Both new professionals and seasoned catalogers need new skills to successfully transition to the emerging linked data environment. Catalogers are learning about and experimenting with BIBFRAME while remaining responsible for traditional bibliographic control of collections. Metadata specialists utilize tools for metadata mapping, remediation, and enhancement. They identify and map semantic relationships among assorted taxonomies to make multiple thesauri intelligible to end users. For the more technical aspects of metadata management, competition for talent from other industries has been increasing. This may intensify as metadata becomes more central to various areas of government, nonprofit, and private enterprise.[137]

NEW TOOLS AND SKILLS

The extent of metadata specialists' collaboration with IT or systems staff varies among institutions. Such collaboration is necessary for many reasons, including managing data that is outside the library's control. Some noted that "cultural differences" exist between the professions: developers tend to be more dynamic and focus on quick prototyping and iteration, while librarians focus first on documenting what is needed and are more "schematic." Which is more likely to be successful: teaching metadata specialists IT skills or teaching IT staff metadata principles? The "holy grail" is to recruit someone with an IT background interested in metadata services. Retaining staff with IT skills is difficult—they are in demand for higher-paying jobs in the private sector. Focus Group members' experiences have shown that it is easier for librarians to learn programming skills than it is to hire IT specialists to learn the "technical services mindset." Ideally, Focus Group members would like a few staff who have the technical skills to take batch actions on data, or at least who know how to use the external tools available to automate as many tasks as possible.

For many years, Focus Group members have been using MarcEdit and/or other tools such as OpenRefine, scripts (e.g., Python, Ruby, or Perl), and macros for metadata reconciliation and batch processing.[138] MarcEdit is the most popular tool, and has a large, global, and active user community as indicated in its 2017 Usage Snapshot.[139] Terry Reese, MarcEdit's developer, estimates that about one-third of all users work in non-MARC environments and two-thirds of the most active users are OCLC members. Focus Group members reported that they use MarcEdit for data transformation, enhancing vendor records, building MARC records from spreadsheets, linked data reconciliation, de-duplicating records within a file, merging two or more records into one, Z39.50 harvesting, and reconciling metadata before sending records to other systems. The 2017 release of MarcEdit 7 includes new features such as light weight clustering functionality, providing a powerful way to find relationships between data without introducing a large learning curve. It also has mechanisms that support linked data.[140] Reese has created a series of YouTube tutorials available on his MarcEdit Playlist.[141]

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. Desirable soft skills include problem-solving, effective collaboration, willingness—even eagerness—to try new things, understanding researchers' needs, and advocacy. Although some metadata specialists have always enjoyed experimenting with new approaches, often they lack the time to learn new tools or methodologies while keeping up with their routine work assignments. Libraries should promote metadata as an exciting career option to new professionals in venues such as library schools and ALA's New Members Roundtable. Emphasizing that metadata encompasses much more than library cataloging—entity identification; descriptive standards used in various academic disciplines; describing born-digital, archival, and research data that can interact with the semantic Web—can increase its appeal. As one Focus Group member noted, "We bring order out of a vacuum."[142]

SELF-EDUCATION

Metadata increasingly is being created outside the library by academics and students with minimal training, leading to a need for more catalogers with record maintenance skills. Focus Group members noted the need for technical skills such as simple scripting, data remediation, and identity management to reconcile equivalents across multiple registries. Frequently mentioned sources of instruction include Library Juice Academy, MarcEdit tutorials, LinkedIn Learning (which acquired Lynda.com), Library of Congress Training Webinars, ALCTS Webinars, Code Academy, Software Carpentry, and conferences such as Code4Lib and Mashcat.[143] W3C's Data on the Web Best Practices and Semantic Web for the Working Ontologist were recommended reading.[144] Crucial to the success of such training is the ability to quickly apply what has been learned. If new skills are not applied, people forget what they have learned. Staff feel frustrated when they have invested the time to learn something that they cannot use in their daily work.

Focus Group members have seen a big shift from relying on Library of Congress instructions to self-education from multiple sources. Some approaches mentioned by participants:

- Emphasize continuity of metadata principles when introducing an expanded scope of work.
- Take advantage of the Library Workflow Exchange,[145] a site designed to help librarians share workflows and best practices across institutions, including scripts.
- From the 2017 Electronic Resources and Libraries Conference: "Don't wait; iterate!" In other words, rather than waiting until staff have all the required skills, let them do tasks iteratively, learning as they go, so they are ready for new tasks when the time comes.
- Have small groups of metadata specialists take programming courses together, after which they can continue to meet and discuss ways to apply their new skills to automate routine tasks.
- Encourage staff to participate in events such as OCLC's DevConnect Webinars[146] to learn from libraries using OCLC APIs to enhance their library operations and services.
- Create reading and study groups that include cross-campus or cross-divisional staff.
- Expand the scope of current work to enable metadata specialists to apply their skills to new domains or terminology, such as using Dublin Core for digital collections. Involve staff in digital projects from the conceptual stage to developing project specifications, quality assurance practices and tool selection. As a bonus, this fosters collaborative teamwork relationships.
- Hire graduate students in computer science for short-term tasks such as creating scripts.

ADDRESSING STAFF TURNOVER

Turnover in a professional position within a cataloging or metadata unit now comes with the significant risk that it may be impossible to convince administrators to retain the position in the unit and repost it. This is particularly true when the outgoing incumbent performed a high proportion of "traditional" work, such as original cataloging in MARC. The odds of retaining the position are much greater if careful thought goes into how the position could be reconfigured or re-purposed to meet emerging needs.[147]

Most Focus Group members have had to address varying amounts of turnover, either from retirements or staff leaving for other positions. Half of them needed to reconfigure the positions of outgoing librarians. Looking at what other institutions are advertising helps in creating an attractive position description. Many cataloging positions do not require an MLS degree, so recruiting

professionals has focused on adaptability, aligning new positions with university priorities, and on eagerness to learn and take initiative in areas such as metadata for research output, open access, digital collections, and linked data. Mapping out future strategies and designing ways of making metadata interoperate across systems have been components of recent recruitments. New staff with programming skills are sought after, as they can apply batch techniques to metadata that can compensate for the loss of staff. Using technology in the service of library service helps catalogers "do more with less."

Focus Group members want new staff to be aware of both the shared cataloging community and the overlaps with other cultural heritage organizations such as archives and museums. The library environment keeps evolving, and librarians have had to reflect on their priorities moving forward. Metadata managers need to rethink the roles of metadata specialists beyond "traditional" cataloging work. 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. Many cataloging roles and descriptions may need to be rewritten and retooled. Perhaps the only activities that will perennially remain professional tasks are those like management, scouting new trends, strategizing, participating in new international standards, leading and implementing changes, and thinking about the big picture.

Impact

The next generation of metadata will become even more focused on entities rather than record-based descriptions of an institution's collections. Focus Group members' linked data activities, including their participation in OCLC Research's Project Passage and CONTENTdm Linked Data pilots, contributed to OCLC obtaining Andrew W. Mellon funding for its two-year Shared Entity Management Infrastructure project,[148] launched in January 2020. Eleven of the Shared Entity Management Infrastructure Advisory Group members are also Focus Group members. The project builds on OCLC Research's linked data work, and will provide a *production infrastructure* with persistent, authoritative identifiers for persons and works. It will be largely API-based, allowing librarians to customize their workflows around linked data infrastructure. This infrastructure has long been desired by Focus Group members as it will address many of the challenges documented above around persistent identifiers, especially identifiers for "works."

The next generation of metadata will become even more focused on entities rather than record-based descriptions of an institution's collections.

Authoritative, persistent identifiers provided by the Shared Entity Management Infrastructure will supply the needed language-neutral links to trustworthy sources. The metadata that libraries, archives, and other cultural heritage institutions have created and will create will provide the context for these entities, as "statements" associated with those links. The impact will be global, affecting how librarians and archivists will describe the inside-out and facilitated collections, inspiring new offerings of "metadata as a service," and influencing future staffing requirements.

ACKNOWLEDGMENTS

OCLC Research wishes to thank all Research Library Partners Metadata Managers Focus Group members who have shared their experiences and thoughts summarized here. Additionally, we extend thanks to the dedicated Metadata Managers Planning Group, which initiated the topics and provided the context statements and question sets, the responses to which served as the basis of our discussions. In addition, we particularly appreciate the insightful comments from the following Focus Group members who reviewed an earlier version of this document; their comments improved this synthesis.

- Charlene Chou, New York University
- Suzanne Pilsk, Smithsonian Institution
- Greg Reeve, Brigham Young University
- Alexander Whelan, Columbia University
- Helen K. R. Williams, London School of Economics

I also extend thanks to current and former OCLC colleagues: Rebecca Bryant, Jody DeRidder, Annette Dortmund, Rachel Frick, Janifer Gatenby, Jean Godby, Shane Huddleston, Andrew Pace, Merrilee Proffitt, Nathan Putnam, Stephan Schindehette, and Chela Weber for their careful review of all or parts of earlier versions of this document. Thank you to Erica Melko for her editing, Jeanette McNicol for the design of this report, and JD Shipengrover for the cover artwork.

On a personal note, I have greatly benefited from my interactions with the OCLC Research Partners Metadata Managers Focus Group and have been delighted to play a small part in this transition to the next generation of metadata.

APPENDIX

OCLC Research Library Partners Metadata Managers Planning Group

2015-2020

Planning Group members selected the topics for the OCLC Research Library Partners Metadata Managers discussions, wrote up the context statements why the topic was important and timely, and developed the question sets that Focus Group members responded to. The Planning Group initiators for each topic also reviewed draft summaries that were later posted on the OCLC Research *Hanging Together* blog.

Current Planning Group members are listed in bold; institutional affiliations are given for the time when they served on the Planning Group:

- **Jennifer Baxmeyer, Princeton University**
- Sharon Farnel, University of Alberta
- Steven Folsom, Harvard University and Cornell University
- **Erin Grant, University of Washington**
- Dawn Hale, Johns Hopkins University
- Myung-Ja Han, University of Illinois, Urbana-Champaign
- Kate Harcourt, Columbia University
- Corey Harper, New York University
- **Stephen Hearn, University of Minnesota**
- **Daniel Lovins, Yale University**
- **Roxanne Missingham, Australian National University**
- Chew Chiat Naun, Cornell University and Harvard University
- **Suzanne Pilsk, Smithsonian**
- **John Riemer, University of California, Los Angeles**
- Carlen Ruschoff, University of Maryland
- Philip Schreur, Stanford University
- Jackie Shieh, George Washington University
- **Melanie Wacker, Columbia University**

NOTES

1. OCLC Research Library Partnership Metadata Managers Focus Group. https://www.oclc.org/research/areas/data-science/metadata-managers.html.

2. OCLC Research. "The OCLC Research Library Partnership." https://www.oclc.org/research/partnership.html.

3. Smith-Yoshimura. 2017. "Metadata Advocacy" *Hanging Together: the OCLC Research Blog*, 17 October 2017. https://hangingtogether.org/?p=6282.

4. British Library. 2019. *Foundations for the Future: The British Library's Collection Metadata Strategy 2019-2023. London: British Library.* https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2019-2023.pdf.

5. Ibid, 4.

6. Statistics as of 1 June 2020.

7. Library of Congess. "Program for Cooperative Cataloging." https://www.loc.gov/aba/pcc/.

8. Except for June 2020, when all discussions were held virtually only because of the COVID-19 pandemic.

9. See *Hanging Together: The OCLC Research Blog*, search-category Metadata. https://hangingtogether.org/?cat=81.

10. Benefits from affiliating with the RLP are cited in Smith-Yoshimura. 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.

11. Analyses of the three International Linked Data Surveys for Implementers 2014-2018 and the spreadsheet of all responses to the surveys are available. See OCLC Research. 2020. "Linked Data." International Linked Data Survey. https://www.oclc.org/research/themes/data-science/linkeddata/linked-data-survey.html.

12. Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Holly Tomren, and Craig Thomas. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*. Dublin, OH: OCLC Research. https://doi.org/10.25333/faq3-ax08;

    OCLC Research. 2020. "CONTENTdm Linked Data pilot." https://www.oclc.org/research/themes/data-science/linkeddata/contentdm-linked-data-pilot.html;

    OCLC. 2020. "WorldCat®: OCLC and Linked Data." Shared Entity Management Infrastructure. https://www.oclc.org/en/worldcat/linked-data/shared-entity-management-infrastructure.html;

Library of Congress. "BIBFRAME." Bibliographic Framework Initiative. https://www.loc.gov/bibframe/;

Futornick, Michelle. 2019. "LD4P2 Linked Data for Production: Pathway to Implementation." LS4P2 Project Background and Goals. Lyrasis. Posted 14 January 2019. https://wiki.lyrasis.org/display/LD4P2/LD4P2+Project+Background+and+Goals;

Share-VDE (Share Virtual Discovery Environment). "An Effective Environment for the Use of Linked Data by Libraries." Accessed 17 September 2019. https://www.share-vde.org/sharevde/clusters?l=en;

Casalini, Michele, Chiat Naun Chew, Chad Cluff, Michelle Durocher, Steven Folsom, Paul Frank, Janifer Gatenby, Jean Godby, Jason Kovari, Nancy Lorimer, Clifford Lynch, Peter Murray, Jeremy Myntti, Anna Neatrour, Cory Nimer, Suzanne Pilsk, Daniel Pitti, Isabel Quintana, Jing Wang, and Simeon Warner. 2018. *National Strategy for Shareable Local Name Authorities National Forum: White Paper.* Ithaka, New York: Cornell University Library *eCommons digital repository*. https://hdl.handle.net/1813/56343.

13. Library of Congress. 2019. PCC (Program for Cooperative Cataloging) Task Group on Linked Data Best Practices. 2019. PCC Task Group on Linked Data Best Practices Final Report: Submitted to PCC Policy Committee 12 September 2019. Washington DC: Library of Congress. https://www.loc.gov/aba/pcc/taskgroup/linked-data-best-practices-final-report.pdf;

    Library of Congress. 2018. "Charge for PCC Task Group on Identity Management in NACO," 5. American Bar Association, Program for Cooperative Cataloging, revised 22 May 2018. https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf;

    Library of Congress. 2020 "PCC Task Group on URIs in MARC." Programs of the PCC. Charge. Accessed 19 September 2020. https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html;

    Library of Congress. 2018. "PCC Linked Data Advisory Committee: Linked Data Advisory Committee Charge." PCC Task Groups 2018. Task Groups. Revised 24 July 2018. [Word doc; 28KB]. https://www.loc.gov/aba/pcc/taskgroup/task-groups.html.

14. Smith-Yoshimura, Karen. 2015. "Shift to Linked Data for Production." *OCLC Research Hanging Together Blog,* 13 May 2015. https://hangingtogether.org/?p=5195.

15. OCLC Research. 2020. "LInked Data." Linked Data Overview. https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-overview.html. [All figures CC BY 4.0]

16. Smith-Yoshimura, Karen. 2019. "'Future Proofing' of Cataloging." *OCLC Research Hanging Together Blog,* 10 November 2019 https://hangingtogether.org/?p=7526.

17. ORCID: Connecting Research and Researchers. "What is Orcid." Our Vision. Accessed 19 September 2020. https://orcid.org/about/what-is-orcid/mission.

18. See for example the list of signatories of journal publishers requiring ORCID IDs for authors. ORCID. "ORCID Open Letter - Publishers." Accessed 19 September 2020. https://orcid.org/content/requiring-orcid-publication-workflows-open-letter.

19. ISNI. "What is ISNI." Accessed 19 September 2020. https://isni.org/page/what-is-isni/.

20. HathiTrust is a not-for-profit collaborative of academic and research libraries preserving more than 17 million digitized items. See: HathiTrust Digital Library. "Welcome to HathtiTrust." Accessed 19 September 2020. https://www.hathitrust.org/about.

21. GeoNames. "Browse the Names." Accessed 19 September 2020. https://www.geonames.org/.

22. Bryant, Rebecca, Annette Dortmund, and Constance Malpas. 2017. *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information.* Dublin, OH: OCLC Research. https://doi.org/10.25333/C32K7M.

23. ISNI currently holds 11.02 million identities: 10.26 million individuals (of which 2.91 million are researchers) and 933,039 organizations. Statistics retrieved from ISNI. See ISNI. "Key Statistics." Accessed 5 May 2020. https://isni.org/.

24. Library of Congress. 2020. "NACO – Name Authority Cooperative Program." Documents and Updates. Programs for Cataloging and Acquisitions (PCC). Accessed 19 September 2020. http://www.loc.gov/aba/pcc/naco/index.html.

25. Smith-Yoshimura, Karen. 2015. "Getting identifiers Created for Legacy Names." *Hanging Together: The OCLC Research Blog,* 30 October 2015. https://hangingtogether.org/?p=5463.

26. Smith-Yoshimura, Karen. 2013. "Irreconcilable Differences? Name Authority Control & Humanities Scholarship" *Hanging Together: The OCLC Research Blog,* 27 March 2013. https://hangingtogether.org/?p=2621.

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

28. OCLC Research. 2020. "Registering Researchers in Authority Files." https://www.oclc.org/research/themes/research-collections/registering-researchers.html.

29 Smith-Yoshimura, Karen, Janifer Gatenby, Grace Agnew, Christopher Brown, Kate Byrne, Matt Carruthers, Peter Fletcher, Stephen Hearn, Xiaoli Li, Marina Muilwijk, Chew Chiat Naun, John Riemer, Roderick Sadler, Jing Wang, Glen Wiley, and Kayla Willey. 2016. *Addressing the Challenges with Organizational Identifiers and ISNI.* Dublin, Ohio: OCLC Research. https://doi.org/10.25333/C3FC9Q.

30. Research Organization Registry (ROR). "About." https://ror.org/about/.

31. V. M. Abazov, B. Abbott, B. S. Acharya, M. Adams, T. Adams, J. P. Agnew, G. D. Alexeev et al. (2014) 2020. "Precision Measurement of the Top-Quark Mass in Lepton+jets Final States." (Archived 24 February 2020) *ArXiv.org:* 1501.07912. https://arxiv.org/pdf/1405.1756.

32. Smith-Yoshimura, Karen. 2017. "How Much Metadata Is Practical?" *Hanging Together: The OCLC Research Blog,* 14 November 2017. https://hangingtogether.org/?p=6328.

33. University of Minnesota. 2020. "Experts@Minnesota." Find Profiles. https://experts.umn.edu/en/persons/ or

University of Illinois at Urbana-Champaign. 2020. "Illinois Experts." Find U of I Research, View Scholarly Works, and Discover New Collaborators. https://experts.illinois.edu/.

34. The National Institute of Health (NIH): National Institute of Allergy and Infectious Diseases (NIAID) on 7 April 2020 mandates ORCIDs for training, fellowship, education, or career development awards in FY20. See NIH: NIAID. 2019. "ORCID iD: Required for Some, Encouraged for All." *NIAID Funding News*. Last reviewed 7 August 2019. https://www.niaid.nih.gov/grants-contracts/orcid-id-required-some-encouraged-all;

    See also Lyrasis. 2020. "SciENcv and ORCID to Streamline NIH and NSF Grant Applications." *LyrasisNow* (blog), 8 April 2020. https://lyrasisnow.org/sciencv-and-orcid-to-streamline-nih-and-nsf-grant-applications/.

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

36. Carruthers, Matt. (2014) 2020. *mcarruthers/LCNAF-Named-Entity-Reconciliation*. GitHub Repository. https://github.com/mcarruthers/LCNAF-Named-Entity-Reconciliation.

37. Deliot, Corine, Steven Folsom, Myung-Ja Han, Nancy Lorimer, Terry Reese, and Adam Schiff. 2019. *Formulating and Obtaining URIs: A Guide to Commonly used Vocabularies and Reference Sources*. Library of Congress PCC Task Group on URIs in MARC. https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf.

38. Smith-Yoshimura, Karen. 2019. "New Ways of Using and Enhancing Cataloging and Authority Records." *Hanging Together: The OCLC Research Blog,* 2 April 2019. https://hangingtogether.org/?p=5710.

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

40. DataCite. "Assign DOIs." https://datacite.org/dois.html;

    Wilkinson, Laura J. 2020. "Constructing your DOIs." Crossref: The Crossref Curriculum. Last updated 8 April 2020. https://www.crossref.org/education/member-setup/constructing-your-dois/.

41. See DOI examples in detail from: DOI. 2020. "DOI System Examples." Accessed 20 September 2020. https://www.doi.org/demos.html; and

    See ARK examples in detail from: Department, Dallas (Tex) Police. 1963. "[Photographs of Identification Cards]." Collection. University of North Texas. *The Portal to Texas History digital repository.* https://texashistory.unt.edu/ark:/67531/metapth346793/.

42. "Identity management" here reflects its usage among metadata specialists (See, for example, Library of Congress. 2018. "Charge for PCC Task Group on Identity Management in NACO," 5. American Bar Association, Program for Cooperative Cataloging. Revised 22 May 2018. https://www.loc.gov/aba/pcc/taskgroup/PCC-TG-Identity-Management-in-NACO-rev2018-05-22.pdf.) But the term has other meanings depending on the audience; for example, identity access management, as described in:

    Wikiwand. "Identity Management." https://www.wikiwand.com/en/Identity_management.

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

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

45. Smith-Yoshimura, "Coverage of Identity Management." (See note 43.)

46. Watch the highly-rated Webinar by Andrew Lih and Robert Fernandez. 2018. "Works in Progress Webinar: Introduction to Wikidata for Librarians: Structuring Wikipedia and Beyond." Produced by OCLC Research, 12 June 2018. MP4 video presentation, 1:1:51. https://www.oclc.org/research/events/2018/06-12.html.

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

48. Wikimedia. "WikiCite." Home. https://meta.wikimedia.org/wiki/WikiCite.

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

50. 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.

51. OCLC 2020. "FAST (Faceted Application of Subject Terminology)." https://www.oclc.org/en/fast.html.

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

53 OCLC 2020. "FAST." (See note 51.)

54. OCLC 2020. "FAST (Faceted Application of Subject Terminology)." Heading #3, FAST Policy and Outreach (FPOC) Committee, : https://www.oclc.org/en/fast.html.

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

56. National Library, New Zealand Government. "Ngā Upoko Tukutuku / Māori Subject Headings" http://mshupoko.natlib.govt.nz/mshupoko/;

AIATSIS Pathways: Gateway to the AIATSIS Thesauri. "Pathways." http://www1.aiatsis.gov.au/.

57. Deutsche Nationalbibliothek. 2019. "MACS - Multilingual Access to Subjects." (Archived 13 Jan 2019.) https://web.archive.org/web/20190113003823/https:/www.dnb.de/EN/Wir/Kooperation/MACS/macs_node.html.

58. Smith-Yoshimura, Karen. 2019. "Knowledge Organization Systems." *Hanging Together: The OCLC Research Blog,* 17 March 2019. https://hangingtogether.org/?p=7135.

59. Synaptica. "Ontology Management – Graphite." https://www.synaptica.com/graphite/.

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

61. OCLC Research. 2020. "Equity, Diversity, and Inclusion in the OCLC Research Library Partnership Survey." Overview. Accessed 20 September 2020. https://www.oclc.org/research/areas/community-catalysts/rlp-edi.html.

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

63. Smith-Yoshimura. "Distributed Models." (See note 60.)

64. 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.

65. Baxmeyer, Jennifer, Karen Coyle, Joanna Dyla, MJ Han, Steven Folsom, Phil Schreur, and Tim Thompson. 2017. *Linked Data Infrastructure Models: Areas of Focus for PCC Strategies*. Library of Congress PCC Linked Data Advisory Committee. https://www.loc.gov/aba/pcc/documents/LinkedDataInfrastructureModels.pdf.

66. Bone, Christine, Sharon Farnel, Sheila Laroque, and Brett Lougheed. 2017. "Works in Progress Webinar: Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries and Archives " Produced by OCLC Research, 19 October 2017. MP4 video presentation, 54:35.00. https://www.oclc.org/research/events/2017/10-19.html.

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

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

69. Bruce Washburn and Jeff Mixter, 2018. "Works in Progress Webinar: Looking Inside the Library Knowledge Vault." Produced by OCLC Research, 12 August 2018. MP4 video presentation, 57:45:00. https://www.oclc.org/research/events/2015/08-12.html.

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

71. Smith-Yoshimura, Karen. 2015. "Working in Shared File. "*Hanging Together: The OCLC Research Blog,* 7 April 2015. https://hangingtogether.org/?p=5091.

72. Jisc Library Services. n.d. "What Is 'Plan M'?" Accessed 21 September 2020. https://libraryservices.jiscinvolve.org/wp/2019/12/plan-m/;

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

For more information about the current phase of "Plan M" (May–November 2020), see Grindley, Neil. "Moving Plan M Forwards – We Need Your Help!" *Library Services (PlanM)* (blog), Jisc, 6 May 2020. https://libraryservices.jiscinvolve.org/wp/2020/05/planm_nextphase/.

73. Grindley, Neil. 2019. "Plan M: Definition, Principles and Direction." Jisc. (Word docx.) http://libraryservices.jiscinvolve.org/wp/files/2019/12/Plan-M-Definition-and-Direction-1.docx.

74. Dempsey, Lorcan. 2016. "Library Collections in the Life of the User: Two Directions." *LIBER Quarterly* 26(4): 338–359. http://doi.org/10.18352/lq.10170.

75. Smith-Yoshimura, Karen. 2019. "Presenting Metadata from Different Sources in Discovery Layers. *Hanging Together: The OCLC Research Blog,* 16 April 2019. https://hangingtogether.org/?p=7880.

76. Smith-Yoshimura, Karen. 2017. "Metadata for Archival Collections." *Hanging Together: The OCLC Research Blog,* 30 May 2017. https://hangingtogether.org/?p=5903.

77. Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas, and Holly Tomren. 2019. *Creating Library Linked Data with Wikibase: Lessons Learned from Project Passage*, 49-51. Dublin, OH: OCLC Research. https://doi.org/10.25333/faq3-ax08.

78. The OCLC Research Library Partnership Archives and Special Collections Linked Data Review Group is described at https://www.oclc.org/research/partnership/working-groups/archives-special-collections-linked-data-review.html.

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

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

81. Archive-It. 2008. "Human Rights." Columbia University Libraries Collection. (Archived May 2008). https://archive-it.org/collections/1068;

Archive-It. 2010. "New York City Places and Spaces." Columbia University Libraries Collection. (Archived January 2010). https://archive-it.org/collections/1757;

Archive-It. 2010. "Burke Library New York City Religions." Columbia University Libraries Collection. (Archived May 2010). https://archive-it.org/collections/1945.

82. NLA. "Trove." Archived Websites. Sub Collections. Accessed 20 September 2020. https://trove.nla.gov.au/website.

83. Archive-It. 2014. "Collaborative Architecture, Urbanism, and Sustainability Web Archive (CAUSEWAY)." Ivy Plus Libraries Confederation Collection. (Archived June 2014.) https://archive-it.org/collections/4638;

Archive-It. 2013. "Contemporary Composers Web Archive (CCWA)." Ivy Plus Libraries Confederation Collection. (Archived October 2013.) https://archive-it.org/collections/4019;

NYARC: New York Art Resources Consortium. "Web Archiving." http://www.nyarc.org/content/web-archiving.

84. OCLC Research. 2020. "Web Archiving Metadata Working Group" The Problem, Addressing the Problem, Outputs. https://www.oclc.org/research/themes/research-collections/wam.html.

85. Dooley, Jackie, and Kate Bowers. 2018. *Descriptive Metadata for Web Archiving: Recommendations of the OCLC Research Library Partnership Web Archiving Metadata Working Group*. Dublin, OH: OCLC Research. https://doi.org/10.25333/C3005C.

86. Smith-Yoshimura, Karen. 2018. "Metadata for Audio and Videos." *Hanging Together: The OCLC Research Blog,* 29 October 2018. https://hangingtogether.org/?p=6814.

87. Weber, Chela Scott. 2017. *Research and Learning Agenda for Archives, Special, and Distinctive Collections in Research Libraries*. Dublin, OH: OCLC Research. https://doi.org/10.25333/C3C34F.

88. Library of Congress. "Standards." Encoded Archival Description (EAD) Official Site. Accessed 21 September, 2020. https://www.loc.gov/ead/.

89. Library of Congress. "Standards." Preservation Metadata Maintenance Activity (PREMIS). Accessed 21 September, 2020. https://www.loc.gov/standards/premis/.

90. Weber, Chela Scott. 2019. "Assessing Needs of AV in Special Collections." *Hanging Together: The OCLC Research Blog,* 23 July 2019. https://hangingtogether.org/?p=7405;

Weber, Chela Scott. 2019. "Scale & Risk: Discussing Challenges to Managing A/V Collections in the RLP." *Hanging Together: The OCLC Research Blog,* 1 October 2019. https://hangingtogether.org/?p=7479.

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

92. Library of Congress. "Standards." Metadata Object Description Schema (MODS). Accessed 21 September 2020. http://www.loc.gov/standards/mods/.

93. Ibid.

94. Library of Congress. "Standards." Metadata Authority Description Schema (MADS)." Accessed 21 September 2020. http://www.loc.gov/standards/mads/.

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

96. OCLC Research. 2020. "Europeana Innovation Pilots." Accessed 20 September 2020. http://www.oclc.org/research/themes/data-science/europeana.html?urlm=168921.

97. IIIF (International Image Interoperability Framework): Enabling Richer Access to the World's Images. "Home." Accessed 20 September 2020. https://iiif.io/.

98. OCLC Research. 2020. "OCLC ResearchWorks IIIF Explorer." https://www.oclc.org/research/themes/data-science/iiif/iiifexplorer.html.

99. OCLC Research. 2020. "CONTENTdm Linked Data Pilot." Introduction. https://www.oclc.org/research/themes/data-science/linkeddata/contentdm-linked-data-pilot.html.

100. Smith-Yoshimura, Karen. 2015. "Data Management and Curation in 21st Century Archives – Part 1." 21 September 2015. http://hangingtogether.org/?p=5375.

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

102. Erway, Ricky, Laurence Horton, Amy Nurnberger, Reid Otsuji, and Amy Rushing. 2015. *Building Blocks: Laying the Foundation for a Research Data Management Program*, 8. Dublin, OH: OCLC Research. https://doi.org/10.25333/C39P86.

103. See the OCLC Research Data Management Planning Guide at https://www.oclc.org/research/areas/research-collections/rdm/guide.html.

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

105. Faniel, Ixchel M. 2019. "Let's Cook Up Some Metadata Consistency." *Next* (blog), *OCLC*, 21 November 2019. http://www.oclc.org/blog/main/lets-cook-up-some-metadata-consistency/.

106. NCI (National Computational Infrastructure): Australia. "Home." Accessed 21 September 2020. http://nci.org.au/;

ADA (Australian Data Archive). "Home." Accessed 21 September 2020. https://www.ada.edu.au/.

107. Portage Network. "Home." Accessed 21 September 2020. https://portagenetwork.ca/.

108. Metadata 2020 is a "collaboration advocating richer, connected, reusable, open metadata for all research outputs" (http://www.metadata2020.org/). The Metadata 2020 Researcher Communications project is outlined here: http://www.metadata2020.org/projects/researcher-communications/.

109. Digital Curation Centre. "Disciplinary Metadata." List of Metadata Standards. Accessed 21 September 2020. http://www.dcc.ac.uk/resources/metadata-standards/list.

110. RDA Metadata Directory. "Metadata Standards Directory Working Group." GitHub Repository. Accessed 21 September 2020. http://rd-alliance.github.io/metadata-directory/.

111. NISO is about to make CRediT (Contributor Roles Taxonomy)—which identifies 14 roles describing each contributor's specific contribution to the scholarly output—a standard. CRediT was developed by CASRAI, the Consortia Advancing Standards in Research Administration Information. See CASRAI. "CRediT – Contributor Roles Taxonomy." Accessed 21 September 2020. https://casrai.org/credit/.

112. University of Michigan Library. 2020. "Data Services." http://www.lib.umich.edu/research-data-services.

113. OCLC Research. 2020. "The Realities of Research Data Management." Overview. https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html.

114. Bryant, Rebecca, Brian Lavoie, and Constance Malpas. 2017. *Scoping the University RDM Service Bundle*. The Realities of Research Data Management, Part 2, pp. 16, 21. Dublin, OH: OCLC Research. https://doi.org/10.25333/C3Z039.

115. Indiana University. 2018. "IU will Lead $2 Million Partnership to Expand Access to Research Data: IU Libraries and IU Network Science Institute Are Leading a Public-Private Partnership to Create the Shared BigData Gateway for Research Libraries" *News* at *UI, (Science and Technology.) Indiana University, 18 October 2018*. https://news.iu.edu/stories/2018/10/iu/releases/18-shared-bigdata-gateway-for-research-networks.html;

Microsoft. 2020. "Microsoft Academic Graph." Established 5 June 2015. https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/;

For more details, watch the August 2019 recording of "Democratizing Access to Large Datasets through Shared Infrastructure." See Wittenberg, Jamie, and Valentin Pentchev. "Works in Progress Webinar: Democratizing Access to Large Datasets through Shared Infrastructure." Produced by OCLC Research, 8 August 2019. MP4 video presentation, 58:34:00. https://www.oclc.org/research/events/2019/080819-democratizing-access-large-datasets-shared-infrastructure.html.

116. NISO's Reproducibility Badging and Definitions now out for public comment may also help researchers extend the benefit of their research to others. See "Taxonomy, Definitions, and Recognition Badging Scheme Working Group | NISO Website." n.d. Accessed 22 September 2020. https://www.niso.org/standards-committees/reproducibility-badging.

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

118. Krista M. Soria, Jan Fransen, Shane Nackerud. 2014. "Stacks, Serials, Search Engines, and Students' Success: First-Year Undergraduate Students' Library Use, Academic Achievement, and Retention." *Journal of Academic Librarianship* 40: 84-91. https://doi.org/10.1016/j.acalib.2013.12.002.

119. See Jisc. "Library Impact Data Project (LIDP)." Accessed 21 September 2020. http://www.activitydata.org/LIDP.html.

120. Smith-Yoshimura, Karen. 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.

121. DLF (Digital Library Federation). 2015. "Metadata Librarian, Cornell University Library." *DLF (blog)*, 11 June 2015. https://www.diglib.org/metadata-librarian-cornell-university-library/.

122. Salary.com. (2019) 2020. "Metadata Librarian." Posted by Georgia Tech University 13 November 2019. (Archived 2 September 2020) https://web.archive.org/web/20200903061830/https://www.salary.com/job/gt-library/metadata-librarian/e5644ece-c847-4cfb-994f-c4c80fa81e3d.

123. OCLC. 2020. "Locate Items in the Library with StackMap." https://help.oclc.org/Discovery_and_Reference/WorldCat_Discovery/Search_results/Locate_items_in_the_library_with_StackMap.

124. Yewno: Transforming Information into Knowledge. 2020. "Home." https://www.yewno.com/.

125. Smith-Yoshimura, Karen. 2019. "New Ways of Using and Enhancing Cataloging and Authority Records" *Hanging Together: The OCLC Research Blog,* 2 April 2019. https://hangingtogether.org/?p=7805.

126. National Library of Australia (NLA). "Austlang National Codeathon." Accessed 21 September 2020. https://www.nla.gov.au/our-collections/processing-and-describing-the-collections/Austlang-national-codeathon [Map of Australia. 2020 HERE, Bing, Microsoft Corporation];

NLA. "Trove." Search. Uncover. Australia. Accessed 21 September 2020. https://trove.nla.gov.au/.

127. The Graphic History Company – Hachette UK. "River of Authors." Accessed 21 September 2020. http://theghc.co/project.php?project=hachette-uk-a-river-of-authors.

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

129. Zeng, Marcia Lei, and Philipp Mayr. 2019. "Knowledge Organization Systems (KOS) in the Semantic Web: A Multi-dimensional Review." *International Journal on Digital Libraries* 20: 209-230. https://doi.org/10.1007/s00799-018-0241-2.

130. SNAC (Social Networks and Archival Context). "About SNAC." What is SNAC? https://portal.snaccooperative.org/about.

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

132. AI4LAM (Artificial Intelligence for Libraries, Archives & Museums). Updated 18 May 2020 https://sites.google.com/view/ai4lam/home.

133. AI4LAM's mission is to organize, share, and elevate knowledge about and use of artificial intelligence by libraries, archives, and museums. It was founded in 2018, inspired by the success of the International Image Interoperability Framework (IIIF) in coordinating large scale collaboration on interoperable technology to advance LAMs.

    See AI4LAM. "About." Our Mission. https://sites.google.com/view/ai4lam/about.

134. Padilla, Thomas. 2019. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*. Dublin, OH: OCLC Research. https://doi.org/10.25333/xk7z-9g97.

135. Ibid, 17-19.

136. Smith-Yoshimura, Karen. 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.

137. Smith-Yoshimura, Karen. 2017. "New Skill Sets for Metadata Management." *Hanging Together: The OCLC Research blog,* 17 April 2017. https://hangingtogether.org/?p=5929.

138. Smith-Yoshimura, Karen. 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.

139. Reese, Terry. 2018 "MarcEdit 2017 Usage Information." *Terry's Worklog* (blog), 9 September 2020. http://blog.reeset.net/archives/2572.

140. Reese, Terry. 2020. "Working with Linked Data In MarcEdit." *MarcEdit Development* (blog). Accessed 21 September 2020. https://marcedit.reeset.net/working-with-linked-data-in-marcedit.

141. Reese, Terry. 2018. "MarcEdit Playlist." 139 YouTube videos. Last updated 26 December 2018. https://www.youtube.com/playlist?list=PLrHRsJ91nVFScJLS91SWR5awtFfpewMWg.

142. Smith-Yoshimura, Karen. 2017. "New Skill Sets for Metadata Management." *Hanging Together: The OCLC Research blog,* 17 April 2017. https://hangingtogether.org/?p=5929.

143. "XML and RDF-Based Systems Archives." n.d. *Library Juice Academy* (blog). Accessed 22 September 2020. https://libraryjuiceacademy.com/certificate/xml-and-rdf-based-systems/;

    Reese, Terry. 2013. "Tutorials." YouTube (selected). *MarcEdit Development* (blog). 14 March 2013. http://marcedit.reeset.net/tutorials;

    "Lynda: Online Courses, Classes, Training, Tutorials." n.d. Lynda.com - from LinkedIn Learning. Accessed 22 September 2020. https://www.lynda.com/;

    "Learn to Code - for Free." n.d. Codecademy. Accessed 22 September 2020. https://www.codecademy.com/;

Software Carpentry. "Teaching Basic Lab Skills for Research Computing." Upcoming Workshops. Accessed 22 September 2020. https://software-carpentry.org/.

144. "Data on the Web Best Practices." n.d. Accessed 22 September 2020. https://www.w3.org/TR/dwbp/;

Semantic Web for the Working Ontologist. (2008) 2020. http://workingontologist.org/.

145. Library Workflow Exchange. n.d. "About." Accessed 21 September 2020. http://www.libraryworkflowexchange.org/about/.

146. OCLC Developer Network. 2020. "DevConnect Webinars. https://www.oclc.org/developer/events/devconnect-workshops.en.html.

147. Smith-Yoshimura, Karen. 2019. "Stewardship of Professional FTEs In Metadata Work and Turnover." *Hanging Together: The OCLC Research Blog,* 18 October 2019. https://hangingtogether.org/?p=7580.

148. OCLC. 2020. "WorldCat®: OCLC and Linked Data." Shared Entity Management Infrastructure. https://www.oclc.org/en/worldcat/linked-data/shared-entity-management-infrastructure.html.

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

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

ISBN: 978-1-55653-167-5
DOI: 10.25333/rqgd-b343
RM-PR-216787-WWAE 2009