ALA American Library Association

LIBRARY LINKED DATA

EARLY ACTIVITY & DEVELOPMENT

Erik T. Mitchell

Library Technology Reports
Expert Guides to Library Systems and Services

JAN 2016
Vol. 52 / No. 1
ISSN 0024-2586

Library Technology

R E P O R T S

Expert Guides to Library Systems and Services

Library Linked Data: Early Activity and Development

Erik T. Mitchell

ALA TechSource
alatechsource.org

American Library Association

Library Technology
REPORTS

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

Volume 52, Number 1

Library Linked Data: Early Activity and Development

ISBN: 978-0-8389-5968-8

American Library Association

50 East Huron St.
Chicago, IL 60611-2795 USA
alatechsource.org
800-545-2433, ext. 4299
312-944-6780
312-280-5275 (fax)

Advertising Representative

Patrick Hogan
phogan@ala.org
312-280-3240

Editor

Patrick Hogan
phogan@ala.org
312-280-3240

Copy Editor

Judith Lauber

Production

Tim Clifford and Alison Elms

Cover Design

Alejandra Diaz

Library Technology Reports (ISSN 0024-2586) is published eight times a year (January, March, April, June, July, September, October, and December) by American Library Association, 50 E. Huron St., Chicago, IL 60611. It is managed by ALA TechSource, a unit of the publishing department of ALA. Periodical postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to Library Technology Reports, 50 E. Huron St., Chicago, IL 60611.

ALA TechSource
alatechsource.org

About the Author

Erik T. Mitchell, PhD, is associate university librarian and associate CIO at the University of California, Berkeley. In addition to focusing his work on information technology adoption and use in libraries, Mitchell studies metadata issues and professional development in library and information science. He is the author of *Cloud-Based Services for Your Library, Metadata Standards and Web Services in Libraries, Archives, and Museums* and is a columnist for *Technical Services Quarterly*. He holds a doctorate in information and library science from the University of North Carolina at Chapel Hill, a master's degree in library science from the University of South Carolina, and a bachelor's degree in literature from Lenoir-Rhyne University.

Abstract

Erik T. Mitchell wrote *Library Technology Reports* (vol. 50, no. 5), "Library Linked Data: Research and Adoption," published in July 2013. This report revisits the adoption of Linked Data by libraries, archives, and museums, identifying current trends, challenges, and opportunities in the field. By looking at services and research-related large-scale projects, such as BIBFRAME and DPLA, the report describes a trajectory of adoption. It looks at the vocabularies, schemas, standards, and technologies forming the foundation of Linked Data as well as policies and practices influencing the community.

Contents

The Current State of Linked Data in Libraries, Archives, and Museums

Since the last issue of *Library Technology Reports* (*LTR*) on Linked Data (LD) in July 2013, the library, archive, and museum (LAM) communities have put considerable work into developing new LD tools, standards and published vocabularies, as well as explored new use cases and applications. In 2013, there was already a range of LD systems in production, and in the past two years, the number of systems has grown steadily. Alongside this growth and experimentation, the discussion of Linked Data and Linked Open Data (LOD) has explored the nuanced differences between schemas such as BIBFRAME and BIBFRAME Lite, has explored the expansion of vocabularies and technologies, and has expanded around themes of technology adoption, LD literacy, evolution of standards and schemas, case studies in adoption, and studies of value and impact.

The 2013 *LTR* issue on LD used a largely technical lens to explore these issues, as there were many unanswered questions about how LAM organizations might apply emerging LD concepts in their metadata and information systems. In studying three important LD platforms (Europeana, OAI-PMH, and DPLA) and in devoting a chapter to exploring the fundamentals of LD, that issue sought to capture the state of adoption and technology use across the LAM community. This update on LD adoption takes a different approach by exploring at a broader level the issues, trends, and LD programs that are shaping our community perspectives. In order to do this, chapter 1 of this issue considers the broad state of LD adoption. Chapter 2 examines projects, services, and research efforts with a goal of better understanding the overall trajectory of adoption. Chapter 3 takes a more detailed look at the vocabularies, schemas, standards, and technologies that are forming the foundation of LD, and chapter 4 considers the policies and practices that are influencing the community and considers next steps that may hold promise in the LAM community.

In order to paint a picture of current efforts and adoption in Linked Data as well as to project the potential future of LD efforts, this issue draws on surveys of LD adoption, updates from national and international project teams, and selective exploration of technical topics that are emerging as new concepts in LD and are likely to influence LD adoption in the coming year. Just as with the 2013 issue, this update serves two purposes. First, it seeks to collect project reports and literature to synthesize ideas and trends as well as inform perspectives on the current state of LD adoption. Second, this issue seeks to capture and document current thinking and practice in LD, recognizing that, at this point, LD has become part of the central discourse in LAM communities, influencing the education and operating principles of the information professions.

The State of Linked Data Adoption

This section examines the findings of a 2014 survey on LD adoption, considers technical developments around LD in LAM contexts specifically, considers how projects and standards are evolving, and discusses broadly the visibility and maturity of projects.

Survey Results from LD Adoption

In 2014, OCLC staff conducted a survey on LD adoption, a survey that is being repeated for 2015. The analyzed results from the 2014 survey are captured in a series of blog posts on the site hangingtogether.org and provide a substantial window into the state of LD deployment in LAM institutions.[1] The survey surfaced 172 projects, of which 76 included substantial description. Of those 76 projects, over a third (27) were in development. The larger, in terms of metadata transformed, projects included OCLC's WorldCat.org, Library of Congress's (LoC) id.loc.gov service, and the British Library's British National Bibliography.[2] General descriptions of selected projects are available in the second blog post as well as the raw data from the survey.[3] A revised survey closed in August 2015 and results, although not available at the time of this writing, should be available on the OCLC Linked Data Research web page by the date of publication.

OCLC Linked Data Research
www.oclc.org/research/themes/data-science/linkeddata.html

One interesting area of analysis from the 2014 survey focused on intended use cases and overall purpose of a LD project. Common use cases cited included "enrich[ing] bibliographic metadata or descriptions," "interlinking," "as a reference source and to . . . harmonize data from multiple sources," "[to] automate authority control," "[to] enrich an application."[4] In addition, the most common reasons for creating an LD service were to publish data more widely and to demonstrate potential use cases and impact.[5] In addition, the Linked Data for Libraries (LD4L) group has gathered a set of use cases to inform their work.[6] These use cases have been clustered into six main areas including "Bibliographic + Curation" data, "Bibliographic + Person" data, "Leveraging external data including authorities," "Leveraging the deeper graph," "Leveraging usage data," and "Three-site services" (e.g., enabling a user to combine data from multiple sources).

Although the analyzed data from the survey showed that a wide range of vocabularies were used in the projects reported, there was also a strong cluster around just a few published vocabularies. According to Smith-Yoshimura, the most commonly used LD data sources were id.loc.gov, DBpedia, GeoNames, and VIAF.[7] Data in the projects analyzed was often bibliographic or descriptive in nature. As captured in the analysis by Smith-Yoshimura, the most common organizational schemas used were Simple Knowledge Organization System (SKOS), Friend of a Friend (FOAF), Dublin Core and Dublin Core terms, and Schema.org.[8] In addition to this short list of highly used vocabularies and schemas, the data shows a much longer list of all of the vocabularies cited in the results.

The analyzed results of the survey indicated that Resource Description Framework (RDF) serialized in the eXtensible Markup Language (XML) was commonly used, as was RDF serialized in JavaScript Object Notation (JSON) and Terse RDF Triple Language (Turtle).[9] Advice from implementers, the content of the sixth blog post on the LD survey, presents a range of perspectives on project management, project scope, and possible technologies and standards to use in development.[10] One sentiment captured in the results is the importance of publishing "useful" data. This sentiment is part of the LOD building blocks popularized by Berners-Lee, especially the rule "When someone looks up a URI, provide useful information, using the standards."[11] This notion, although seemingly obvious, has become part of subsequent recommendations around the creation of LD. For example, the CIDOC Conceptual Reference Model Special Interest Group (CRM-SIG) has codified this sentiment in a series of guidelines for creating and publishing LOD.[12] Of equal importance but with less guidance is the issue of data licensing. The referenced CIDOC recommendation focuses largely on technical issues and does not mention licensing recommendations. Somewhat surprisingly, in the OCLC survey results, there was a range of approaches to licensing of data, including many Creative Commons CC0 licenses but also Open Data Commons (ODC) and noncommercial use licenses.[13] Such variation in licensing may not be a substantial issue, but it does add a level of complexity when considering what uses an organization can make of published data.

A related policy question surfaced in this survey is how LAM institutions should approach LD production or adoption. It appears that despite the transition to Linked Data for large-scale and core services such as the transformation of library MARC platforms and the migration of EAD finding aids, the community has not yet distilled a set of activities or systems into an "easy-to-implement" platform or adoption approach. Indeed, LD efforts might still be categorized as existing in the startup phase of a technology adoption hype cycle given the variation in standards, tools, approaches, and perceived benefits documented in survey results and published literature. At the same time, however, LD services have expanded to a point where they may soon reach critical mass in enabling widespread use in the LAM community. This is demonstrated in part by the continued growth of LD adopters and test programs that are working with data that would impact a large number of libraries and archives. It is also indicated by the growth of the number of triples published by these services, showing that the automation and refinement tools needed are reaching a level of maturity and that successive LD projects have more to build on.

Activities across US Libraries

Another useful source of information about developments and projects in LD is the annual updates of research libraries in conjunction with the American Library Association (ALA) ALCTS Technical Services Directors of Large Research Libraries Interest Group.[14] The fifteen public reports from June 2015 show a range of LD efforts in these libraries. For example, many institutions are pursuing education for staff via the Library Juice Academy certificate program (http://libraryjuiceacademy.com) or the Zepheira LibHub early adopters training (http://zepheria.com/solutions/library/training). Many of the reports indicate that institutions have approached LD from an exploration and research perspective (e.g., formation of a project team; establishing broad goals; working with available tools and standards to explore impact in the local environment). Trends in these reports included exploring how to leverage LD and LD URIs in discovery systems generally and potentially in local catalog applications.

Within this research thread there are a number of specific projects. As a partner in the LD4L project, Cornell has been active in an ontology group and working to set up a Vitro instance for LD cataloging.[15] The Library of Congress reported its multifaceted work in BIBFRAME, providing a window into the development and testing of this schema. The report indicates that LoC is using the MarkLogic platform for development of BIBFRAME and leveraging the vocabularies at the LoC Linked Data Service Authorities and Vocabularies web page. It is projecting a test of this platform for late summer and early fall of 2015, the goal of which is to explore the application of BIBFRAME and these vocabularies in a real-world setting.[16] Likewise, the National Library of Medicine (NLM) has undertaken considerable testing and development with LD, as reported elsewhere in this issue. This work includes releasing Medical Subject Headings (MeSH) as RDF. This data is being made available as annually updated downloadable files.[17] Although much of the work in LD in the LAM community comes from bibliographic roots there is evidence of a growing interest in other data sources and applications. For example, in addition to traditional resource-based metadata, some institutions are working with ORCID identifiers as a way to better capture research productivity for faculty and graduate students.

VIVO
http://vivoweb.org

LoC Linked Data Service: Authorities and Vocabularies
http://id.loc.gov

Medical Subject Headings (MeSH)
http://id.nlm.nih.gov/mesh

In addition to containing specific project information on LD, there are several projects that seem poised to benefit from advances in LD. Migration of libraries from either older versions of their ILS or to a new open-source ILS platform (e.g., the Open Library Environment) was mentioned in a number of these reports, either as an accomplishment in 2015 or as an upcoming project in 2016. Likewise, the deployment or enhancement of discovery platforms remained a central activity. One trend, tangentially related to LD, was the publication of digital objects with open-access licenses. The University of Pennsylvania, for example, released OPenn, a resource focused on making cultural heritage materials available under Creative Commons licenses.[18] With a similar goal, the University of Michigan released the Special Collections Image Bank with the goal of capturing digitized images and making them available under the appropriate license.[19] These released products suggest potential paths of new development in LD, particularly the potential of these open digital platforms to enable more extensive discovery and reuse of resources and metadata.

OPenn
http://openn.library.upenn.edu

University of Michigan Special Collections Image Bank
http://quod.lib.umich.edu/s/sclib

Linked Data Trends: Technical, Application, and Visibility

Technical Developments in LD Adoption

In the past two years, the LD community has continued to focus on RDF and has increased its use of JSON serializations of RDF. Several important standards have seen increasing adoption, including the final specification of HTML5 and the definition of the RDF 1.1 standard in 2014.[20] HTML5 provides enhanced support for geolocation services, application cache and local data, server sent events (i.e., automatic updates from the server to the client), and support for web worker application programming interfaces (APIs; e.g., JavaScript running in the background of the client application). These interactivity tools are enabling the development of a new generation of interaction and data-rich web services and allow the web client to make extensive use of published open data. Similarly, the RDF 1.1 standard expands the utility of RDF by adding much-needed support for RDF datasets, a collection of RDF graphs, expansion of data types, and new definitions for handling of internationalized resource identifiers (IRIs) and literals.[21]

The RDF 1.1 primer explores these concepts in more detail, in addition to providing an overview of emerging serialization languages including TriG, N-Quads, and JSON-LD.[22] Each of these serialization techniques provides expanded support for named graphs, TriG extending Turtle to add this functionality and N-Quads extending N-Triples. JSON-LD, like JSON in general, has been an emerging and popular serialization platform for several years. At the same time, the increased emphasis on JSON-LD is not without controversy in the LD community. JSON has been praised for being a lightweight, platform-integrated approach but also criticized for not supporting the complex models and relationships that can be expressed in XML.[23] At the time of this writing, JSON-LD's inclusion of new keywords (e.g., @graph) has helped provide more robust support for the representation of RDF in JSON. In addition, as any casual user of LD applications in LAM contexts will observe, JSON-LD is increasingly common, featured in a number of LD enabled services including DPLA's API. Given the increasing use of JSON and JSON-LD, it is likely that the LD community would benefit from the further support of JavaScript and server integration coming from the HTML5 community.

In addition to efforts in the LD community to transform bibliographic and other metadata services and data stores (e.g., BIBFRAME, BIBFRAME Lite, Schema.org), there is considerable work being done to leverage LD to develop new products and services. Jason Clark and Scott Young, for example, recently explored the use of JSON-LD in creating and structuring e-book content.[24] Their work drew on several of the perceived benefits of LD creation, including search engine optimization, connection with social media networks, and connection to other resources through links and content integration. On the theme of service integration through structured and linked metadata, Suzanna Conrad explored the use of Google Analytics to study use of DSpace metadata fields.[25] Finding that the tag manager tool in Google Analytics was a good fit for tracking metadata fields in DSpace, Conrad pointed to an analytical application of data linking, even if the tools discussed do not surface metadata in a conventional LD platform.

Another important area of work in LD is the application of existing tools to improve the quality of data. Although not necessarily focused on generating LD, the increase in use of these tools is important to the long-term viability of data cleanup and normalization. Donnelley, for example, used a combination of Python and OpenRefine tools to clean up and normalize zip code information.[26] Such a task is often one of many steps that occur prior to the publication of data and is particularly important in the generation of unique pointer information such as zip code data. This article in particular provides useful instructions in the detailed work required for such a task.

Coming from a different perspective, Bianchini and Willer explored the role of historic library standards such as International Standard Bibliographic Description (ISBD), asking how the concepts in ISBD fit with Semantic Web needs.[27] Their article explored a notion that is common in other areas of research around metadata standards: that our older vocabularies and approaches are not always easily mapped onto new technologies and use cases. In particular, Bianchini and Willer explored the shifting notion of resource from ISBD to the concept of a resource in RDF. Dunsire conducted a parallel analysis of ISBD and ISBD punctuation, finding similar challenges in employing this standard in semantic contexts without some level of modification.[28] These two works focusing on standard alignment with an emphasis on the role of older standards in new LD settings are representative of larger discussions in the LD community. The ALA Metadata Standards group, for example, has also debated the perceived value of ISBD in LD settings and recently drafted a series of guidelines for assessing metadata standards to help shape this discussion at a broader level.[29]

Although much of the LD focus of the LAM community is on transformation of bibliographic and collection (e.g., MARC and EAD) schemas, there is also interest in authorities and translation of LD schemas to new domains. The electronic thesis and dissertation (ETD) community, for example, has looked at some level at the influence of LD models on connecting ETD repositories and enabling new scholars to enjoy more visibility on the web.[30] Likewise, emerging researcher ID platforms such as ORCID, ResearcherID, arXiv, Author Claim, and Scopus Author ID are pushing more communities toward LD-related discussions through the thread of name disambiguation and author-based graphs. The emergence of scholar identifiers in LD standards focused on earlier stages in an academic's career could have considerable impact in increasing awareness around LD issues (e.g., disambiguation, persistent identifiers, open data, and metadata) in the broader research community. The extent to which the maturity of the tools and the abilities of researchers and practitioners are at a state to support widespread adoption is yet to be seen, but such advances bode well for the broad appeal of LD and other Semantic Web technologies

ORCID
http://orcid.org

ResearcherID
www.researcherid.com

Focused more closely on enterprise tools and projects, a growing area of research seeks to advance understanding of potential systems based on services provided by DPLA, Europeana, and WorldCat. One example of this is Péter Király's work implementing translation services for queries with the goal of enabling a user to query terms across multiple languages simultaneously.[31] In addition to work focused on exploring adaptive ways of using LD via APIs, other efforts continue on vocabulary improvement and publishing. Toves and Hickey recently documented expanded algorithms for processing dates in VIAF, demonstrating that the new approach has led to considerable improvements in normalization in the dataset.[32] In a similar thread, some libraries are branching into their own targeted vocabulary creation. Hanson documented North Carolina State University's efforts to develop an LD dataset of organization names.[33] This project, having been in production for many years, is used to manage name information in library information systems and is also part of the Global Open Knowledgebase (GOKb). Each of these vocabularies represents highly impactful projects occurring at different scales in the LAM community.

NCSU Libraries, Organization Name Linked Data
www.lib.ncsu.edu/ld/onld

Global Open Knowledgebase
http://gokb.org

Occurring somewhat in contrast to these efforts to generate more LD or improve LD quality, there is also a strong thread of research around the use of APIs. Perhaps ironically, APIs are usually seen as a stopgap measure that is required when LD is not available, but in many cases they are the tools that enable the creation of LD in the first place. Reese, for example, completed an in-depth introduction to tools, techniques, and output associated with the WorldCat API.[34] Similarly, Nugraha introduced MariaDB, a replacement open-source server similar to MySQL and Sphyinx, a full-text search platform that works in concert with relational databases.[35] While such work is more related to rather than directly connected with LD work, advances in the tools and techniques from work like this are important to laying the groundwork and making better use of available information systems.

Evolution of Projects and Standards

In the past year, the Library of Congress and OCLC have completed a report comparing their two approaches to LD creation,[36] while other efforts have spawned BIBFRAME Lite, Zepheria's extended BIBFRAME vocabularies, or have defined alternative approaches to exploring a BIBFRAME implementation, such as the NLM work on this topic.[37] Although BIBFRAME, Schema.org, BIBFRAME Lite, and other similar standards tend to be at the center of LD discussions for libraries, a number of other standards are emerging that are designed with LD principles in mind. Encoded Archival Description 3 (EAD3), for example, is building in new elements to make better use of Encoded Archival Context—Corporate Bodies, Persons and Families (EAC-CPF) as well as Uniform Resource Identifiers (URIs) from other sources.[38] Likewise, a World Wide Web Consortium (W3C) community group has been formed to explore how to extend the Schema.org standard to include better descriptive metadata for digital and physical archives.[39]

BIBFRAME Lite
http://bibfra.me

NLM's efforts to test bibliographic LD schemas as documented in its June 2015 update surfaced test records that followed the BIBFRAME Lite vocabulary where possible, using more granular schemas where necessary.[40] Per Fallgren's update, the NLM effort largely sought to map BIBFRAME Lite to Resource Description and Access's (RDA) RDF vocabulary, but vocabulary definitions were also drawn from LoC's BIBFRAME vocabulary, MODS RDF, Schema.org, and W3C. One justification offered for this approach is the concern that many efforts are focusing on MARC and BIBFRAME alignment, rather than on designing a vocabulary that is oriented toward a broader range of resources. Alongside these efforts, LoC has continued to advance work on BIBFRAME, launching testing platforms, refining test applications, and contributing to an expansive discussion on BIBFRAME schema issues in the community. The BIBFRAME model has been documented in a series of releases including vocabularies, relationship models, and suggested nonbibliographic applications.[41] Although LoC established a release of BIBFRAME in the summer of 2015, it also continues to refine the standard through a series of proposals.

Outside of the LAM community, LOD has been increasingly adopted to enable better search engine optimization (SEO) and to surface knowledge cards and "rich snippets" in search results and Google's Knowledge Graph.[42] In 2015, the W3C released a specification for a Linked Data platform that defines a set of systems and system integrations to enable the creation and publication of Linked Data.[43] In commercial environments, APIs appear to continue to take precedence over openly published LD. Amazon, for example, preferences APIs to surface catalog data and enable functional integration. Services such as Alexa (the tool behind Amazon's Echo room system), Marketplace (its tool to publish data on the Amazon catalog), and Mechanical Turk (a system to enable crowdsourced processing of information) all follow an API over LD model.[44]

Wikipedia: Knowledge Graph
https://en.wikipedia.org/wiki/Knowledge_Graph

Geographic and location-based services including mapping, way finding, and navigation are seeing increasing system integration, but largely through API-based services such as Map APIs, Bluetooth beacon technology, and push-to-mobile interaction techniques. Bluetooth beacons are a good example of the complex relationships that are developing between location-aware services, embedded technology, and the trend toward sensor-based networks, according to Gruman.[45] These sensors trigger actions in applications based on proximity and can transmit details about the environment, including temperature and time. They can correspondingly log access, provide small bits of information to devices, and help devices triangulate the location of a user in a space by using the proximity information from multiple sensors.

Bluetooth beacons are part of a larger development around the "Internet of Things" (IoT) community in that they can provide description and location information for physical items. Internet-based cameras, Wi-Fi-enabled household products (e.g., televisions, refrigerators, thermostats), and Internet-connected locks and access systems are each contributing to the growing presence of Internet-connected and data-generating devices. As these devices become more common and as their use grows, there is an increasing need to help users bring together these devices and the information they create into a cohesive network that is capable of sharing data as well as inferring new information from shared data among devices.

Ermilov and Auer suggest, for example, that Internet-connected television services could be connected to LD publishers such as DBpedia and IMDb at the client or individual level, enabling a user to actively select content to connect (e.g., a TV guide and IMDb ratings; actor lists and DBpedia entries) from his or her own device rather than working through a centralized service provider that had pre-integrated those services.[46] At the moment, most IoT technologies work within a specific ecosystem, making it difficult to develop generalized information networks, but some tools, such as Bluetooth beacons, are being designed to work across a range of applications rather than simply within a single application.

LOD Visibility

Within the LAM community, LOD is a commonly discussed topic that tends to have a shared set of values (e.g., make data open, enable reuse, support new uses of data). These values are common in other academic communities, including researchers dedicated to open scholarship and reproducibility as well as creators of data in certain domains. The US government website http://data.gov, for example, now provides access to over 150,000 datasets, although in many cases these datasets are serialized in HTML, PDF, and other noncomputational document formats. In addition, while the Data.gov site makes items available through a faceted discovery platform, it does not seek to act as an authoritative location for the data and as such does not publish persistent URLs (PURLs). In many cases, however, the data is provided with authorship and license information, two important elements in creating open, if not linked, data.

While LOD is highly visible in the LAM community and is increasingly referenced, by concept if not name, in reproducibility and data publishing communities, it has yet to enjoy widespread understanding or popularization in the press. In fact, searching the web for news stories on Linked Data surfaces more articles from 2000 to 2009, when news companies like the *New York Times* began publishing data as LD, than more recent articles. LD continues to attract funding, however—for example, from the Mellon Foundation, a supporter of the LD4L project; from the Institute for Museum and Library Services (IMLS) in its support for BIBFLOW and the Linked Data for Professional Education programs; and from a range of libraries, archives, and museums that use internal funding to experiment with LD.

New York Times: Linked Open Data (Beta)
http://data.nytimes.com

Outside of these funded areas and LAM-focused research threads, whether or not LD and LOD need to enjoy greater visibility in the research community is a topic of debate. Digital humanities programs and communities may be most likely to benefit from LOD experimentation in data publishing as newly published datasets hold the potential to directly drive new threads of research. Likewise, the reproducibility and data science communities could be strong contributors to the evolving practice of LOD in LAM institutions through the development of tools and methods that could be applied to other research domains. The related but as yet unresolved question around visibility is whether or not LD has reached critical mass in the LAM community to ensure further adoption and transformation. The overall lack of visibility of the role and impact of LD does not help address this issue, although the commitment of large-scale organizations is still heavily influencing how organizations perceive the importance of LD.

Maturity of Vocabularies

The OCLC survey of adoption reviewed earlier in this chapter indicated that LAM institutions are beginning to agree on a series of vocabularies, even if there are areas of ambiguity in how the vocabularies are used or differences of opinion in which vocabularies should be used. One key set of vocabularies that are part of this discussion are BIBFRAME and BIBFRAME Lite and the vocabularies associated with LoC (e.g., Name Authority File, Subject Authority File), as well as the VIAF. The investment in these vocabularies in non-LD formats may ensure that the LD versions enjoy adoption, and in fact they are featured in BIBFRAME and BIBFRAME Lite schemas. How much consensus exists around the higher-level schemas, particularly as framed in the discussion of web visibility, has yet to be seen.

Another important discussion in the LD community centers on the proper fit of vocabularies with different communities of practice. Although BIBFRAME was designed to be a resource-agnostic vocabulary, it has a way to go before it will enjoy broad adoption. As might be expected, the geographic information system (GIS) community has branched out to create its own vocabularies and vocabulary-publishing platform in GeoNames. The discussion around appropriate fit dovetails with related conversations about the perceived value of LD work in general (e.g., how should LAM institutions balance the need for generalized LD models that encourage interoperability with external community members against the need for highly granular internally focused standards)?

Conclusion

Chapter 1 of this issue has served as an overview of the state of LD adoption and sought to catch the reader up from the July 2013 issue of *Library Technology Reports* on Linked Data. This chapter focused in part on the survey completed in 2014 on LD adoption across the LAM community and expanded on identified themes through literature review and exploration of developments in LAM communities.

An original goal of this issue was to gather together the various projects and initiatives underway in the LAM community. As the author engaged in research and studied the results of the 2014 OCLC survey, it became apparent that the LD community has become too large to study comprehensively in a detailed way. With that in mind, the author is glad to see a revised version of the LD adoption survey being conducted and expects that the results of that survey will be informative for those seeking best practices and guidance on how to launch their own LD projects. Given the fact that the survey results will come shortly after the publication of this issue, it makes sense to focus this work on broad trends and technologies rather than on specific projects and use cases.

In chapters 2 and 3, this issue skims the surface of LD adoption in order to identify representative trends and activities that are currently important in the LD LAM community. Recognizing that these project examples and their importance are situated in the larger context of the web and of the growing use of the Internet of Things and in the broader questions around value and impact, chapter 4 seeks to study the "so what?" questions around LD innovation and adoption.

Notes

1. Karen Smith-Yoshimura, "Linked Data Survey Results 1—Who's Doing It (Updated)," *hangingtogether.org* (blog), OCLC Research, August 28, 2014, last updated September 4, 2014, http://hangingtogether.org/?p=4137; Karen Smith-Yoshimura, "Linked Data Survey Results 2: Examples in Production (Updated)," *hangingtogether.org* (blog), OCLC Research, August 29, 2014, last updated September 4, 2014, http://hangingtogether.org/?p=4147; Karen Smith-Yoshimura, "Linked Data Survey Results 3—Why and What Institutions Are Consuming (Updated)," *hangingtogether.org* (blog), OCLC Research, September 1, 2014, last updated September 4, 2014, http://hangingtogether.org/?p=4155; Karen Smith-Yoshimura, "Linked Data Survey Results 4—Why and What Institutions Are Publishing (Updated)," *hangingtogether.org* (blog), OCLC Research, September 3, 2014, last updated September 4, 2014, http://hangingtogether.org/?p=4167; Karen Smith-Yoshimura, "Linked Data Survey Results 5—Technical Details," *hangingtogether.org* (blog), OCLC Research, September 5, 2014, http://hangingtogether.org/?p=4256; Karen Smith-Yoshimura, "Linked Data Survey Results 6—Advice from the Implementers," *hangingtogether.org* (blog), OCLC Research, September 8, 2014, http://hangingtogether.org/?p=4284.
2. Smith-Yoshimura, "Linked Data Survey Results 1."
3. Smith-Yoshimura, "Linked Data Survey Results 2"; Karen Smith-Yoshimura, "Results of Linked Data Survey for Implementers," Excel file, September 5, 2014, OCLC Research, www.oclc.org/content/dam/research/activities/linkeddata/oclc-research-linked-data-implementers-survey-2014.xlsx.
4. Smith-Yoshimura, "Linked Data Survey Results 3."
5. Smith-Yoshimura, "Linked Data Survey Results 4."
6. Simeon Warner, "LD4L Use Cases," *Linked Data for Libraries* wiki, last modified by Tom Cramer, May 7, 2015, http://wiki.duraspace.org/display/ld4l/LD4L+Use+Cases.

7. Smith-Yoshimura, "Linked Data Survey Results 3."
8. Smith-Yoshimura, "Linked Data Survey Results 4."
9. Smith-Yoshimura, "Linked Data Survey Results 5."
10. Smith-Yoshimura, "Linked Data Survey Results 6."
11. Tim Berners-Lee, "Linked Data," W3C, last updated June 18, 2009, www.w3.org/DesignIssues/LinkedData.html.
12. Nick Crofts, Martin Doerr, and Mika Nyman, "Call for Comments—Linked Open Data Recommendation for Museums," International Council of Museums, accessed July 24, 2015, www.cidoc-crm.org/URIs_and_Linked_Open_Data.html.
13. Smith-Yoshimura, "Linked Data Survey Results 4."
14. Jennifer Marill, "Round Robin Reports—Annual 2015," ALA Connect, June 12, 2015, http://connect.ala.org/node/240751.
15. Jim LeBlanc, "Report from Cornell," June 17, 2015, "Round Robin Reports—Annual 2015," ALA Connect, http://connect.ala.org/node/240751.
16. Jennifer Marill, "Annual 2015 Report from LC," June 18, 2015, "Round Robin Reports—Annual 2015," ALA Connect, http://connect.ala.org/node/240751.
17. Jennifer Marill, "Annual 2015 Report from NLM," June 18, 2015, "Round Robin Reports—Annual 2015," ALA Connect, http://connect.ala.org/node/240751.
18. Beth Camden, "Report from Penn," June 19, 2015, "Round Robin Reports—Annual 2015," ALA Connect, http://connect.ala.org/node/240751.
19. Bryan Skib, "Report from Michigan," June 23, 2015, "Round Robin Reports—Annual 2015," ALA Connect, http://connect.ala.org/node/240751.
20. Ian Hickson, Robin Berjon, Steve Faulkner, Travis Leithead, Erika Doyle, Navara Edward O'Connor, and Silvia Pfeiffer, eds., "HTML5: A Vocabulary and Associated APIs for HTML and XHTML," W3C Recommendation, October 28, 2014, www.w3.org/TR/html5; Richard Cyganiak, David Wood, and Markus Lanthaler, "RDF 1.1 Concepts and Abstract Syntax," W3C Recommendation, February 25, 2014, www.w3.org/TR/rdf11-concepts.
21. David Wood, ed., "What's New in RDF 1.1," W3C Working Group Note, February 25, 2014, www.w3.org/TR/rdf11-new.
22. Frank Manola, Eric Miller, and Brian McBride, eds., "RDF 1.1 Primer," W3C Working Group Note, February 25, 2014, www.w3.org/TR/2014/NOTE-rdf11-primer-20140225.
23. Erik Wilde, "JSON or RDF? Just Decide," *dretblog*, February 10, 2015, http://dret.typepad.com/dretblog/2015/02/json-or-rdf-just-decide.html.
24. Jason A. Clark and Scott W. H. Young, "Building a Better Book in the Browser (Using Semantic Web Technologies and HTML5)," *Code4Lib Journal,* no. 29 (July 15, 2015), http://journal.code4lib.org/articles/10668.
25. Suzanna Conrad, "Using Google Tag Manager and Google Analytics to Track DSpace Metadata Fields as Custom Dimensions," *Code4Lib Journal*, no. 27 (January 21, 2015), http://journal.code4lib.org/articles/10311.
26. Frank Donnelly, "Processing Government Data: ZIP Codes, Python, and OpenRefine," *Code4Lib Journal*, no. 25 (July 21, 2014), http://journal.code4lib.org/articles/9652.
27. Carlo Bianchini and Mirna Willer, "ISBD Resource and Its Description in the Context of the Semantic Web," *Cataloging & Classification Quarterly* 52, no. 8 (2014): 869–87, http://dx.doi.org/10.1080/01639374.2014.946167.
28. Gordon Dunsire, "The Role of ISBD in the Linked Data Environment," *Cataloging & Classification Quarterly* 52, no. 8 (2014): 855–68, http://dx.doi.org/10.1080/01639374.2014.933753.
29. Jennifer Liss, "DRAFT Checklist for Evaluating Metadata Standards," *Metaware.Buzz* (blog), ALA Metadata Standards Committee, January 20, 2015, http://metaware.buzz/2015/01/20/draft-checklist-for-evaluating-metadata-standards.
30. Lucas Mak, Devin Higgins, Aaron Collie, and Shawn Nicholson, "Enabling and Integrating ETD Repositories through Linked Data," *Library Management* 35, no. 4/5 (2014): 284–92, http://dx.doi.org/10.1108/LM-08-2013-0075.
31. Péter Király, "Query Translation in Europeana," *Code4Lib Journal*, no. 27 (January 21, 2015), http://journal.code4lib.org/articles/10285.
32. Jenny A. Toves and Thomas B. Hickey, "Parsing and Matching Dates in VIAF," *Code4Lib Journal*, no. 26 (October 21, 2014), http://journal.code4lib.org/articles/9607.
33. Eric M. Hanson, "A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project," *Serials Review* 40, no. 4 (2014): 251–28, http://dx.doi.org/10.1080/00987913.2014.975887.
34. Terry Reese, "Opening the Door : A First Look at the OCLC WorldCat Metadata API," *Code4Lib Journal*, no. 25 (July 21, 2014), http://journal.code4lib.org/articles/9863.
35. Arie Nugraha, "Indexing Bibliographic Database Content Using MariaDB and Sphinx Search Server," *Code4Lib Journal*, no. 25 (July 21, 2014), http://journal.code4lib.org/articles/9793.
36. Carol Jean Godby and Ray Denenberg, *Common Ground: Exploring Compatibilities between the Linked Data Models of the Library of Congress and OCLC* (Washington, DC: Library of Congress and Dublin, OH: OCLC Research, January 2015), www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf.
37. Nancy Fallgren, "Experimentation with BIBFRAME at the National Library of Medicine," GitHub, last updated June 24, 2015, https://github.com/fallgrennj/BIBFRAME-NLM.
38. "EAD3 Gamma Release," Society of American Archivists, accessed July 24, 2015, www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/ead3-gamma-release.
39. Schema Architypes Community Group, W3C Community and Business Groups, accessed September 12, 2015, https://www.w3.org/community/architypes.
40. Marill, "Annual 2015 Report from NLM"; example records can be found as links on the NLM BIBFRAME

experimentation report: Fallgren, “Experimentation with BIBFRAME.”
41. Library of Congress, “BIBFRAME Model and Vocabulary,” accessed September 12, 2015, www.loc.gov/bibframe/docs/index.html.
42. “Rich Snippets (Microdata, Microformats, RDFa, and Data Highlighter),” Google, Webmaster Tools Help, accessed March 11, 2013, http://support.google.com/webmasters/bin/answer.py?hl=en&answer=99170 (page now discontinued).
43. Steve Speicher, John Arwe, and Ashok Malhotra, eds., “Linked Data Platform 1.0,” W3C Recommendation, February 26, 2015, www.w3.org/TR/2015/REC-ldp-20150226.
44. “What’s Available?” Amazon Offerings for Developers, accessed September 12, 2015, https://developer.amazon.com.
45. Galen Gruman, “What You Need to Know about Using Bluetooth Beacons,” *Smart User* (blog), InfoWorld, July 22, 2014, www.infoworld.com/article/2608498/mobile-apps/what-you-need-to-know-about-using-bluetooth-beacons.html.
46. Timofey Ermilov and Sören Auer, “Enabling Linked Data Access to the Internet of Things,” in *Proceedings: iiWAS2013: 15th International Conference on Information Integration and Web-Based Applications and Services*, ed. Edgar Weippl, Maria Indrawan-Santiago, Matthias Steinbauer, Gabriele Kotsis, and Ismail Khalil, 300–308 (New York: Association for Computing Machinery, 2013), http://dx.doi.org/10.1145/2539150.2539157.

Chapter 2

Projects, Programs, and Research Initiatives

Chapter 2 examines representative projects, programs, and research initiatives in the LD community. In doing so, the goal of this chapter is to identify and illustrate trends and themes across LD adoption and innovation rather than to capture every project or program. While chapter 1 explored the broad themes and trends, chapter 2 explores some detailed use cases that illustrate the trends in chapter 1. Chapter 2 concludes with a discussion of the projects, their shared features, and goals.

The July 2013 Linked Data issue of *Library Technology Reports* considered the technical design around the metadata standards and their associated systems and platform contents. In the past two years, there have been updates to the metadata schemas and content of the systems, but by far the more interesting questions that have emerged are focused on how these platforms are being used and what part of the LD ecosystem they are seeking to fill. For this reason, the 2015 LD update focuses on broader policy and adoption questions, as opposed to technical and functional questions. In addition, in order to get a broader sampling of perspectives, the systems surveyed in this issue are selected from a broader, if not representative, range of platforms. This range includes large-scale production systems as well as niche, domain-centric, and experimental platforms.

In order to consistently evaluate these platforms, the review of projects, programs, and research initiatives explores the following questions for each system or service reviewed:

- What is the overall goal and focus of the platform?
- How does this platform situate itself in the context of other information systems?
- What gaps or unanswered questions does this platform raise?

Although it is fair to characterize the systems discussed as library-, archive-, museum-, or gallery-focused, the success of these systems is not based on their functional alignment but rather on their ability to interoperate with data sources and contribute new LD to the web. Therefore, while these groupings may be mentioned, they are not a categorizing focus of this issue.

National Projects and Programs

In order to better understand how LD issues and advances are playing out in large-scale collaboratives, this section explores selected projects including BIBFRAME, BIBFRAME Lite, Europeana, British Library and British Museum programs, and advances in OCLC's Linked Data projects. It is clear that this is not a representative or comprehensive selection. At the same time, these projects represent considerable efforts and momentum in the LD LAM community.

Developments in BIBFRAME and BIBFRAME Lite

Although there is a wide range of applications in the BIBFRAME and BIBFRAME Lite community, this issue clusters these applications to some extent, given the overlap in goal and focus. As a whole, the work across BIBFRAME-related projects is focused on transforming existing bibliographic metadata or creating new descriptive metadata following BIBFRAME or

BIBFRAME Lite standards. Given the complexity of BIBFRAME, this issue does not dive deeply into its structure. More information on the vocabulary is on the website of the Bibliographic Framework Initiative. The website includes a definition of the properties, classes, and relationships in the BIBFRAME vocabulary as defined by the Library of Congress.

> *Bibliographic Framework Initiative*
> http://bibframe.org

The BIBFRAME Lite and related vocabularies are available at the BIBFRAME Vocabulary Navigator and include four categories of vocabulary elements (i.e., Lite, Library, Relation, and Rare Materials) that define differences between different levels of complexity in the BIBFRAME Lite vocabulary. The BIBFRAME Lite vocabulary defines equivalence relationships with BIBFRAME, Schema.org, SKOS, and Dublin Core, although not every defined Lite class has an equivalence relationship. The BIBFRAME Lite and related vocabulary set is made available under a Creative Commons International 4.0 (i.e., share, adapt, any use, but with attribution) license.

> *BIBFRAME Vocabulary Navigator*
> http://bibfra.me

Although the BIBFRAME and BIBFRAME Lite projects are largely centered around vocabulary development, they are mentioned in the context of a program because of the broader community engagement and tool development activities surrounding them. Similarly, OCLC's use of the Schema.org vocabulary set is discussed later in this chapter in part because of its larger context around metadata migration and use.

The BIBFRAME initiative was well developed in 2013 and received in-depth consideration in the previous *LTR* issue on LD. In the past two years, LoC has engaged more testing organizations and had a plan to further test BIBFRAME in the fall 2015.[1] One of the more public testers has been the National Library of Medicine, which has created its own documentation around BIBFRAME use cases and potential applications. In the summer of 2015, NLM published the results of its further testing with the BIBFRAME Lite and related vocabularies.[2] These vocabularies include BIBFRAME Lite, BIBFRAME+Library, BIBFRAME+Relation, RDA RDF, and MODS RDF. Each of the BIBFRAME vocabularies is from Zepheria's BIBFRAME efforts, rather than the core LoC-managed BIBFRAME vocabulary. It is difficult to gather from the literature what the underlying efforts are that led to the creation of these two parallel vocabularies that are employing the same name, if not the same namespace.

The NLM update reported on efforts to apply these vocabularies to metadata creation activities. In its testing report, NLM is careful to point out that it did not convert data from a MARC record but rather generated new metadata according to RDA principles. This may be a confusing point given the direction that libraries will likely take in creating LD (i.e., in deriving records from MARC), but as Fallgren points out, there is an overriding concern that basing too much work on MARC at this point risks making and following assumptions about how data should be structured based on historic rather than forward-looking data.[3]

The BIBFRAME community as documented on the LoC BIBFRAME website includes a number of test projects that follow some level of BIBFRAME work.[4] The University of Illinois at Urbana-Champaign, for example, is converting 300,000 e-books from MARC to BIBFRAME and providing a search interface to support discovery of those e-books. Following e-learning integration, the University College London Department of Information Studies is developing a BIBFRAME dataset as an Open Educational Resource (OER). Such a step may help with future integration activities from library databases into learning management systems. In the past year, projects from Columbia, NLM, Princeton, George Washington University, and the Music Library Association (MLA) have all sought to explore different use cases around BIBFRAME. The MLA project is documented on the *CMC BIBFRAME Task Force Blog*, a site that contains a range of updates and posts related to BIBFRAME developments and reports.

> *University of Illinois: Search BIBFRAME Works and Instances*
> http://sif.library.illinois.edu/bibframe/search.php?utf8=%E2%9C%93
>
> *CMC BIBFRAME Task Force Blog*
> www.musiclibraryassoc.org/blogpost/1230658/CMC-BIBFRAME-Task-Force-blog

One of the key issues highlighted in the task force blog and prevalent elsewhere in discussions is the question of how far BIBFRAME should go in attempting to be a complete vocabulary. The discussion is well framed by Vermeij, Adams, and McFall, who explored the tension between the need for standardization to support widespread adoption and the value in leveraging the standards relevant to specific communities.[5] Their blog post also observed that gaps remain in the BIBFRAME vocabulary, the example given being the lack of a vocabulary for sound carriers. The work of the MLA testing group highlights a number of other concerns with BIBFRAME largely, but not always, centered on cases associated with music-type resources and issues.

One area the MLA blog devotes considerable attention to is the testing of conversion tools produced by LoC and Zepheria. These tools are the main source of conversion functions and, according to the findings of the MLA, still have room for improvement. One challenge faced by users of the tools and potential experimenters is the technical expertise needed to download and install these tools. In addition to the tools offered by LoC and Zepheria, Zepheria has built a set of open-source conversation tools collectively called pybibframe. Pybibframe can convert MARCXML to Versa, RDF/XML, or RDF/Turtle. The Versa model is described as a model for web resources and relationships.[6] More information about Versa is available in the Versa GitHub documentation pages. Other tools designed to facilitate conversion of MARC data to LD include the LoC tool suite, which includes a series of conversion, searching, and editing tools. Several of these tools are also available on hosted demonstration sites.

pybibframe
https://github.com/zepheira/pybibframe

Versa
https://github.com/uogbuji/versa

Versa documentation pages
https://github.com/uogbuji/versa/blob/master/doc/index.md

Library of Congress BIBFRAME Tools and Downloads
www.loc.gov/bibframe/tools

The information made available on blogs and websites about the BIBFRAME and BIBFRAME Lite initiatives leaves many questions unanswered about the coming evolution and potential rollout of these vocabularies. A considerable complication is the lack of definition of the differences between these two seemingly competing instances of the BIBFRAME concept and the related lack of symmetry around the conversion and editing tools associated with the standards. Libraries and librarians seeking to better understand the overall direction of BIBFRAME and BIBFRAME Lite are well served by paying attention to related projects, such as BIBFLOW, LD4L, NLM testing, and other testing sites.

Digital Public Library of America (DPLA)

The Digital Public Library of America (DPLA) launched in 2013 after a brief planning period from 2011 to 2013. Upon launch, the DPLA published a metadata application profile (MAP) that filled a role similar to the Europeana Data Model (EDM) in that it was oriented toward normalization and co-indexing of data. The 2013 *LTR* issue on Linked Data explored the DPLA MAP version 3.1 in detail.[7] This specification was updated in 2015 to version 4.0, although the API for DPLA is still based on the 3.1 model to provide backward compatibility.[8] The DPLA also surfaces the entire database of harvested records in a bulk download format. In the two years since DPLA launched, considerable investment has gone into expanding the database of gathered materials as well as developing new public-facing services and expanding the developer API.

In the past two years, the DPLA has grown to include over 10 million objects from twenty-seven partners. In 2015, it released a strategic plan that emphasized continued technical development, sustained outreach to new partners, and development of a plan for sustainability.[9] DPLA has framed its program as consisting of three facets: a portal for discovery, a platform to support application development, and a public option for accessing scholarship. DPLA sees its service hubs model (i.e., partner organizations that act as intermediaries for individual contributors), such as the North Carolina Digital Heritage Center, as a top priority. As the strategic plan points out, there is more work required to fully realize the vision of Linked Data use in DPLA.

A challenge highlighted by the DPLA is the wide variation in rights statements and the impact that a wide variation in rights has on a user's abilities to make use of resources. Although no concrete outcomes have been announced, the DPLA did receive funding from the John S. and James L. Knight Foundation to explore this issue further.[10]

Europeana Digital Library

While WorldCat.org may represent the largest published collection of LD derived from bibliographic metadata, Europeana may be the largest example of LD published through large-scale gathering and normalization of data. With nearly 150 providers and providing metadata and discovery services for more than 44 million records, Europeana provides researchers and institutions with a new and more highly scaled mechanism for surfacing digital collections.[11] The 2013 *LTR* issue focused on a deep exploration of the EDM, and it appears that over the past two years that model has been fairly stable. The most recent EDM schema, version 5.2.6, was released in late 2014, but it appears to have refined, rather than rewritten, the schema that was deployed in 2013. The Europeana schema draws on a range of vocabularies, including RDF and RDFS, OAI-ORE, SKOS, Dublin Core, the W3C Data Catalog Vocabulary, and the Creative Commons vocabulary.[12]

Although a wholly separate entity, the European Library is a major contributor to Europeana and provides access to a dataset of over 82 million bibliographic records under a Creative Commons CC0 1.0 license.[13] The data is available under an OpenSearch API as well as a robust API that outputs data in XML, JSON, and RDF/XML via the Europeana Library LD model. The OpenSearch API provides faceted search support and access to thumbnail previews.[14] As with the DPLA platform, the European Library API can support the development of new search and display platforms. For example, a search of the word *cats* returns 28,845 results, presented twenty results at a time, with facets such as year, country of publication, creator, publisher, catalog record links, and TEL URIs. The European Library database contains 20 million LD records from the Research Libraries UK (RLUK), consisting of records from thirty-four libraries. Vocabularies linked to using the RLUK include VIAF (Virtual International Authority File), GeoNames, LCSH, LCC, data.bnf.fr, Gemeinsame Normdatei, Dewey Decimal Classification (DDC), ISO639-2 Languages, and MARC Countries.[15] This dataset is available in whole as well as through API access.

Register for a European Library API Key
www.theeuropeanlibrary.org/tel4/register

The issues highlighted in Europeana publications include a need to better manage rights issues by allowing institutions to share content online[16] and to promote more integration of resources into educational settings, as well as the establishment of rights that support this type of integration.[17] Like the DPLA, Europeana is launching a strategic plan in 2015.[18] The plan shares a goal similar to that of the DPLA, to enhance the organization's current ability to gather data and store it, to make the data available to end users through discovery and access services, and to make the data available to more sophisticated users via a service platform. The three associated priorities for these services are to improve data, make the data open, and create value for members.[19] In addition, the strategic plan addresses financial sustainability and governance in more detail.

The British Library and British Museum Efforts

The British Library has a history of leading in LD projects, having been an early adopter of the metadata publishing technique. One of the highest profile projects in the British Library around LD is the British National Bibliography (BNB), which consists of metadata records from resources published in the United Kingdom and Republic of Ireland. These collections are available under a CC0 1.0 license in N-Triples, RDF/XML, and Turtle formats as well as CSV formats oriented toward researchers, Z39.50 access for MARC, and SPARQL endpoints.[20] The BNB consists of a range of vocabularies including the Bibliographic Ontology, Biographical Ontology, British Library Terms, Dublin Core, Event Ontology, FOAF, OWL 2, RDF Schema, and RDA.[21] The BNB takes a more nuanced approach to rights and open data than some other projects in that it retains the ability to license data for particular uses.

The British Museum Semantic Web Collection (SWC) provides LD via a SPARQL endpoint with complete coverage of the museum's online collection. Like some other models, the SWC conforms to the CIDOC CRM to enable interoperability with cultural heritage collections. The collection consists of over 2 million objects.[22] The platform is driven by OntoText, a commercial, hosted graph database and semantic tool suite.

OntoText GraphDB
http://ontotext.com/products/ontotext-graphdb

There are an increasing number of LD services in production in the LAM community, and these selected examples are by no means representative. Other highly developed platforms not explored in this issue include the CEDAR census project and the Yale Center for British Art's Linked Data Service. Collectively, there appears to be growing maturity in the selection of vocabularies and representation of data through APIs and SPARQL endpoints. Projects like BNB, The European Library (TEL), and Europeana all provide data through a range of access points, for example, and with varying levels of access and security. TEL, for example, requires registration to access the API, while BNB provides its data openly but with a specific filter (e.g., open data but not linked, via downloadable snapshots, via SPARQL endpoints). The range of approaches may be a sign as much of the different goals of the institutions as it is a sign of the differences in software tools that are available. In chapter 3, we explore several of these tools and ask how each type of tool can be used to help generate LD.

CEDAR
www.cedar-project.nl

Linked Open Data, Yale Center for British Art
http://britishart.yale.edu/collections/using-collections/technology/linked-open-data

WorldCat.org and WorldCat Works

WorldCat and WorldCat Works are both LD applications that rely on LD following the Schema.org standard. WorldCat.org contains approximately 300 million records, making it one of the largest, if not the largest, LAM-related LD projects in production. The Schema.org standard defines a vocabulary that OCLC augments with the VIAF vocabulary, classification vocabularies (e.g., id.loc.gov), Metadata Authority Description System (MADS), and a library-specific vocabulary extension for Schema.org. A complete exposition of OCLC's use of vocabularies and RDFa to surface bibliographic metadata in WorldCat.org is available in *Library Linked Data in the Cloud.*[23] Although Schema.org does not have bibliographic-specific metadata at the level needed for full granular representation of MARC data, OCLC is pursuing an extended bibliographic data standard within Schema .org in the form of a W3C community forum called Bib Extend. Although this community is in its early stages and has yet to set working goals and objectives, the stated mission of the group, generally speaking, is to extend the Schema.org standard to provide better representation of bibliographic data by seeking consensus around ideas.

Full Hierarchy, Schema.org
https://schema.org/docs/full.html

Experimental "Library" Extension Vocabulary for Use with Schema.org
http://purl.org/library

Schema.org Bib Extend Community Group
https://www.w3.org/community/schemabibex

WorldCat Works is an OCLC service centered on publishing LD about FRBResque work sets, expressed in Schema.org using the schema:CreativeWork and schema:Product elements. The Works service is browsable in the OCLC Linked Data Explorer via selected examples, although it is not clear exactly how this service will mature.[24] WorldCat Works IDs are available within the Linked Data published alongside any given resource in WorldCat under the element schema:exampleOfWork (e.g., schema:exampleOfWork http://worldcat.org/entity/work/id/52960). The URI that is the value of this element can be used to identify all associated instances of a work through the Schema.org element workExample. This approach to the representation of FRBR relationships using Schema.org elements is a different path from that taken in other FRBR models suggested in the past. Although the author was not able to locate definitive documentation on the algorithms used to generate work identifiers, more information on techniques being employed in OCLC research is available in chapter 4 of *Library Linked Data in the Cloud.*[25]

OCLC's focus on supporting a web-facing serialization technique for LD as opposed to transforming internal systems first is markedly different from the two related BIBFRAME efforts. Although there have been shared publications discussing the complementary nature of the efforts, it does appear that the work is taking OCLC's metadata in a different direction.[26]

Research Efforts and Initiatives

While much work around LD for LAM communities is focused on growing a community of practitioners and converted data, a similarly long list of projects focuses on asking research questions and exploring new potential use cases of LD. Funding for these projects comes from governmental agencies including the National Endowment for the Humanities (NEH) and the Institute for Museum and Library Services (IMLS), as well as private funders including the Andrew W. Mellon Foundation. High-profile projects in the LAM community include the BIBFLOW project, an IMLS-funded project led by the University of California, Davis, and Linked Data for Libraries (LD4L), a Mellon-funded partnership between Cornell, Harvard, and Stanford libraries.

BIBFLOW
https://www.lib.ucdavis.edu/bibflow

Linked Data for Libraries (LD4L)
https://wiki.duraspace.org/pages/viewpage.action?pageId=41354028

BIBFLOW is exploring technical services workflows using updated standards and user needs as a starting point. One product in the pipeline for the BIBFLOW project is the adaptation of the Open Library Environment (OLE) to incorporate RDF data and support resource description using LD augmented metadata. BIBFLOW's collaboration with Zepheria and the NLM on BIBFRAME Lite is documented in the NLM

BIBFRAME testing update by Nancy Fallgren.[27] As of spring 2015, efforts within the BIBFLOW project included developing a graph-based integration with the OLE, studying cataloging interfaces and needs, and mapping metadata to LD bibliographic standards.

Like the BIBFLOW project, the LD4L community has explored the adaptation of existing vocabularies to create an appropriate LD vocabulary. The overarching goal of LD4L was to create SIRSIS, an LD platform and ontology.[28] In the past two years, the project has produced use cases, code for metadata transformation, and tools to integrate with the Hydra platform. More products from the LD4L project are available on its GitHub site. The community has generated tools to convert data to LD, including a tool called marc2linkeddata. In addition to converting existing MARC data to an LD format, the program will do entity resolution for selected authorities. The LD4L project has developed a robust documentation site on the DuraSpace site that includes overviews of past work in LD as well as detailed documentation on other efforts. The LD4L community has identified several use cases that may add useful context for LAM institutions seeking potential avenues of adoption. These use cases include building virtual collections, tagging scholarly resources, expanding search around author and work connections, searching within geographic data, enriching data via external vocabularies (e.g., GIS, subject, person), using authorities for higher quality data creation, identifying related works, cross-site searching, and combining data for analytics.[29]

GitHub, Linked Data for Libraries Project
https://github.com/ld4l

marc2linkeddata
https://github.com/ld4l/marc2linkeddata

Linked Data for Libraries, Previous Partner LD Work
https://wiki.duraspace.org/display/ld4l/Previous+Partner+LD+Work

There are a number of grant projects dedicated to the generation of datasets and vocabularies based on LD principles. Global Open Knowledgebase (GOKb), for example, is a Mellon Foundation–funded project connected with the Kuali OLE project, as well as JISC collections.[30] While not explicitly published in LD, the platform has an OpenRefine extension to enable reconciliation of data and the insertion of URIs for organization data.[31] The Encoded Archival Context—Corporate bodies, Persons and Families (EAC-CPF) project and Social Networks and Archival Context (SNAC) are two projects driven by the archive community that seek to provide more specificity around name authorities and the other information that is included in records.[32] SNAC was initially supported by the NEH and has continued work in partnership with IMLS and the Andrew W. Mellon Foundation.

Discussion and Conclusion

As the recap of projects indicates, there have been advances in technology and standards development in the past two years, but also larger efforts around collaboration and discussion of policy, governance, and funding issues. In particular, as the LoC effort continues alongside other community and commercial efforts, there are new questions to ask about the appropriate home and standards body for LAM metadata.

In the technical sphere, the advances of technology do not appear to have had dramatic influence on the direction of projects. The RDF/XML standards that have existed since the mid-2000s continue to be the preferred data publishing platform, and the approaches for publishing LD have not changed considerably in the past few years. The release of RDF 1.1 does offer new relationship and vocabulary elements for standards to take advantage of, but as yet the projects reviewed do not appear to have done so. An emphasis on triplestores, interoperable vocabularies, and SPARQL endpoints continues to captivate the LD community, while service providers also focus on data serialization for search engine optimization and data exchange formats.

As yet there is no cloud-based open source LD data exchange service, although efforts by some vendors are pushing in that direction. The BIBFLOW project in particular is exploring various approaches to making data available by adopting the OLE platform to store triples and links data while also pulling in vocabularies and unique data from other systems.

Broad trends noted in reviewing the projects, workshop proceedings, and literature include these:

- an increasing interest in offering SPARQL endpoints as part of data publishing
- the distinction between discovery (end-user), access/service (developer/professional), and policy/rights (legal) perspectives in LD services
- the increasing need to bring together URI minting services and ensure that vocabulary adoption is done in a manageable way
- the discussion around comprehensive versus distributed standards
- the value of peer-to-peer metadata sharing and linking versus large or centralized sharing
- reconciliation and interoperability across metadata standards

These broad topics and issues are important, particularly as the discussion around LD centers more on national and international initiatives and as organizations attempt to come to terms with questions around how they would actually implement LD solutions. Across this chapter, the focus on programs, projects, and funded initiatives has shaped our exploration toward broader policy issues in LD. In chapter 3, we turn our attention to the development of vocabularies and tools to better understand how the building blocks of LD in LAM institutions are coming along.

Notes

1. Library of Congress, "BIBFRAME Implementation Register," accessed September 12, 2015, www.loc.gov/bibframe/implementation/register.html; "Fifteen Minute Hour bfmockup.docx," Google Docs, accessed September 12, 2015, https://docs.google.com/document/d/1N9Xrjc52_Z2VlHVuutvERUl7NW9PzwvV3iRAyjnoQj0/edit.
2. Nancy Fallgren, NLM BIBRAME Update, NLM Technical Bulletin 404 (May/June 2015): e13, https://www.nlm.nih.gov/pubs/techbull/mj15/mj15_bibframe.html.
3. Nancy Fallgren, "Experimentation with BIBFRAME at the National Library of Medicine," GitHub, last updated June 24, 2015, https://github.com/fallgrennj/BIBFRAME-NLM.
4. Library of Congress, "BIBFRAME Implementation Register."
5. Hermine Vermeij, Anne Adams, and Lisa McFall, "Report 3.2. BIBFRAME Extensions vs. External Vocabularies," *CMC BIBFRAME Task Force Blog*, Music Library Association, January 15, 2015, www.musiclibraryassoc.org/blogpost/1230658/206561/Report-3-2-BIBFRAME-Extensions-vs-External-Vocabularies.
6. "uogbuji," "Versa," GitHub, last updated September 18, 2015, https://github.com/uogbuji/versa.
7. Erik T. Mitchell, "Library Linked Data: Research and Adoption," *Library Technology Reports* 49, no. 5 (July 2013).
8. *An Introduction to the DPLA Metadata Model,* Digital Public Library of America, March 5, 2015, http://dp.la/info/wp-content/uploads/2015/03/Intro_to_DPLA_metadata_model.pdf.
9. *Digital Public Library of America: Strategic Plan: 2015 through 2017*, 2015, http://dp.la/info/wp-content/uploads/2015/01/DPLA-StrategicPlan_2015-2017-Jan7.pdf.
10. "Digital Public Library of America Wins Knight News Challenge Award, Receives $300,000 to Develop Simplified Rights Structure for Digital Materials alongside International Partners," *DPLA Updates* (blog), Digital Public Library of America, June 23, 2014, http://dp.la/info/2014/06/23/dpla-wins-knight-news-challenge-award.
11. "Providers," Europeana, accessed September 12, 2015, www.europeana.eu/portal/europeana-providers.html.
12. Europeana, *Definition of the Europeana Data Model*, v5.2.6 (Europeana, December 17, 2014), http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM%20Definition%20v5.2.6_01032015.pdf.
13. "About the European Library Datasets," European Library website, accessed July 28, 2015, www.theeuropeanlibrary.org/tel4/access.
14. DeWitt Clinton with Joel Tesler, Michael Fagan, Joe Gregorio, Aaron Sauve, and James Snell, "Specifications » OpenSearch » 1.1 » Draft 5," OpenSearch.org, accessed September 12, 2015, www.opensearch.org/Specifications/OpenSearch/1.1.
15. European Library, *Research Libraries UK Linked Open Data at the European Library* (The Hague, Netherlands: National Library of the Netherlands, April 2014) www.rluk.ac.uk/wp-content/uploads/2014/04/Research-Libraries-UK-Linked-Open-Data-at-The-European-Library.pdf.
16. "European Parliament Demands Copyright Rules That Allow Cultural Heritage Institutions to Share Collections Online," Europeana Pro, June 23, 2015, http://pro.europeana.eu/blogpost/eu-parliament-in-favour-of-copyright-rules-better-fit-for-a-digital-age.
17. Europeana Foundation, *Europeana for Education and Learning: Policy Recommendations*, May 2015, http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana for Education and Learning Policy Recommendations.pdf.
18. Europeana, *"We Transform the World with Culture": Europeana Strategy 2015–2020,* 2015, http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana Strategy 2020.pdf.
19. Ibid., 12.
20. http://bnb.data.bl.uk/sparql.
21. "Documentation," British National Bibliography website, accessed September 11, 2015, http://bnb.data.bl.uk/docs.
22. "Help," British Museum website, accessed July 28, 2015, http://collection.britishmuseum.org/help.html.
23. Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter, *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*, Synthesis Lectures on the Semantic Web: Theory and Technology (Morgan & Claypool, April 2015), http://dx.doi.org/10.2200/S00620ED1V01Y201412WBE012.
24. The OCLC Linked Data explorer documentation is at "How to Explore WorldCat Linked Data," OCLC Developer Network, accessed September 12, 2015, https://www.oclc.org/developer/develop/linked-data/linked-data-exploration.en.html, and an example of the WorldCat Works result "Gandhi, An Autobiography: The Story of My Experiments with Truth," WorldCat Linked Data Explorer, accessed September 12, 2015, http://experiment.worldcat.org/entity/work/data/1151002411.
25. Godby et al., *Library Linked Data*.
26. Carol Jean Godby and Ray Denenberg, *Common Ground: Exploring Compatibilities between the Linked Data Models of the Library of Congress and OCLC* (Washington, DC: Library of Congress and Dublin, OH: OCLC Research, January 2015), www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf.
27. Fallgren, "Experimentation with BIBFRAME."
28. Dean B. Krafft, "Project Description," *Linked Data for*

Libraries wiki, February 4, 2014, https://wiki.duraspace.org/display/ld4l/Project+Description.
29. Simeon Warner, "LD4L Use Cases," *Linked Data for Libraries* wiki, last modified by Tom Cramer, May 7, 2015, http://wiki.duraspace.org/display/ld4l/LD4L+Use+Cases.
30. "Global Open Knowledgebase Receives Additional Mellon Funding to Pioneer Community-Sourced Management of Digital Content for Education and Research," news release, North Carolina State University Libraries, January 22, 2015, http://news.lib.ncsu.edu/blog/2015/01/22/global-open-knowledgebase-receives-additional-mellon-funding-to-pioneer-community-sourced-management-of-digital-content-for-education-and-research.
31. Knowledge Integration, Ltd., "gokp-phase1," GitHub, last updated July 13, 2015, https://github.com/k-int/gokb-phase1/wiki/GOKb-Refine-Extensions.
32. Society of American Archivists, "Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF)," January 2011, http://www2.archivists.org/groups/technical-subcommittee-on-eac-cpf/encoded-archival-context-corporate-bodies-persons-and-families-eac-cpf; SNAC: Social Networks and Archival Context homepage, Institute for Advanced Technology in the Humanities, University of Virginia, 2013, http://socialarchive.iath.virginia.edu.

Applied Systems, Vocabularies, and Standards

In the previous *LTR* issue on LD (July 2013), one of the compelling comments in the NISO community forum indicated that the important work in metadata and LD should focus on "mapping not migration."[1] The notion that the future of bibliographic or other types of metadata would involve the ability to round-trip metadata rather than a wholesale adoption of Linked Data models and vocabularies is not entirely in sync with some of the efforts we have seen in the review of projects that have taken shape since 2013. The research in the field around LD for the past two years has focused largely on surveys of adoption and specific technical works focused on defining best practices and proof-of-concept services. As the exploration of example projects and research initiatives in chapter 2 indicates, the LD LAM community is reaching a level of maturity that may be shaping next steps in LD adoption toward production systems and permanent migration.

Chapter 3 explores trends around specific tools, vocabularies, systems, and approaches employed by the projects mentioned in chapter 2. While the space allotted limits this section to providing pointer and brief descriptive information, the chapter seeks to provide references to literature and project approaches that may provide sufficient detail for organizations seeking to get started in their own LD projects. Readers seeking a more in-depth understanding of how to approach Linked Data projects would be well served by spending time with one of the growing sets of implementation guides. These include *Linked Data* by Wood, Zaidman, and Ruth and "The Joy of Data" by Hyland and Wood, in addition to a range of other resources.[2]

As explored in chapter 1, the growing list of LD adopters is laying important groundwork for those taking on LD creation next by developing tools and approaches as well as establishing more robust vocabularies to draw on. In chapter 3, we explore a few representative vocabularies and some tools that are increasingly used in LD projects.

Vocabularies and Schemas

The LAM community has largely centered on RDF and RDFS as a main representation data model for LD but varies in its choice of serializations (e.g., RDF/XML, RDFa, JSON-LD, Turtle). RDF/XML remains popular, but N-Triples, Turtle, RDFa, and especially JSON-LD are growing in popularity. New serialization standards, such as Versa, continue to emerge but do not appear to have widespread adoption. JSON-LD's increasing use in the LD community is notable in part because of its lightweight syntax but also because of its ease of use in programming languages. In fact, over the past two years, more programming languages have built libraries to make use of JSON-LD, and a more robust vocabulary has been developed within the standard to support lossless encoding of RDF. More information on JSON-LD is available at the JSON for Linking Data website, including a demonstration site. One common application of JSON-LD is to use the data in a framework such as AngularJS, a JavaScript-based development framework primarily oriented at using HTML to express web applications. AngularJS has been used by the British Museum, for example, to deploy a SPARQL search demonstration.

JSON for Linking Data
http://json-ld.org

British Museum AngularJS SPARQL Demo
http://collection.britishmuseum.org/angularsparqldemo/#

As more projects advance around LD standards, there are a growing number of vocabulary-aware tools built into common scripting languages that are lowering barriers to adoption. Python includes libraries like RDFLib, a library for working with RDF, and Django-RDF, a Django-based RDF engine. Other tools include html5lib, an HTML library for publishing data; Apache Jena and Fuseki, an in-memory database for processing RDF; and Callimachus, a Linked Data management system or an application server for Linked Data.

Django-RDF
https://code.google.com/p/django-rdf

Html5lib
https://github.com/html5lib

Apache Jena and Fuseki
http://jena.apache.org/index.html

Callimachus
http://callimachusproject.org

In addition to RDF, common organizing vocabularies include RDFS, OWL, and SKOS, within which FOAF, GeoNames, Dublin Core, and MODS are vocabularies commonly implemented. In several cases, these vocabularies are implemented in more comprehensive Semantic Web services such as sameAs.org, a service to support disambiguation and URI identification of data; DataHub, a site for publishing datasets; and DBpedia, a Linked Data platform for *Wikipedia* data. Another popular source for discovering datasets is Wikidata, an LD platform for collecting structured data that is also used in other Wikimedia projects.

sameAs.org
http://sameas.org

DataHub
http://thedatahub.org

DBpedia
http://dbpedia.org

DataHub: Datasets
http://datahub.io/dataset

Wikidata
https://www.wikidata.org

Of all of the vocabularies that are of interest to the LAM community, BIBFRAME and BIBFRAME Lite are certainly among the most discussed. The BIBFRAME Lite vocabulary is available online and includes four base terms: Work, Instance, Authority, and Event. These terms mirror those in BIBFRAME but do not entirely overlap with BIBFRAME vocabulary meanings. The BIBFRAME Lite site includes interoperability maps showing the overlap and interoperability with other LD schemas, including Schema.org and BIBFRAME. The author found, in his research about the status of LD adoption and services, that there is a wealth of resources that document the structure and application of these vocabularies. As a result, this issue of *LTR* does not attempt to replicate this information.

BIBFRAME Lite vocabulary
http://bibfra.me/view/lite

A vocabulary that is becoming more common in the LAM community is BiblioGraph.net, an extension to Schema.org designed to add bibliographic-specific content to Schema.org. As the Schema.org vocabulary matures, it is developing methods for representing videos and music in ways that allow computers to embed the media in web pages as well as capturing and promoting events. Such new structured data elements in the Schema.org vocabulary pose opportunities for LAM institutions to embed not only descriptive metadata centered on resources but also actual media and activity information in their sites. Another vocabulary related to Schema.org practices is called GoodRelations. GoodRelations provides a semantic structure for dealing with product data, sales locations, and other commercially focused concepts.

BiblioGraph.net
http://bibliograph.net/docs/bgn_releases.html

Schema.org: TV and Movie Watch Actions
https://developers.google.com/structured-data/actions/watch-movies

Schema.org: Event Markup
https://developers.google.com/structured-data/events/venues

GoodRelations wiki
http://wiki.goodrelations-vocabulary.org

In the cultural heritage community, a more established cultural heritage vocabulary, Lightweight Information Describing Objects (LIDO), has seen many adopters. Tsalapati as well as Van Keer, for example, studied the migration of LIDO using the CIDOC CRM model.[3] The CRM model is a conceptual model that defines semantic relationships for cultural heritage resources. CIDOC continues to enjoy adoption across a range of communities. The FRBRoo model represents FRBR relationships using the CRM model. Likewise, PRESSoo extends FRBRoo for serials and other continuations.

LIDO: XML Schema for Contributing Content to Cultural Heritage Repositories
www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html

CIDOC: FRBRoo Introduction
www.cidoc-crm.org/frbr_inro.html

PRESSoo
www.issn.org/the-centre-and-the-network/our-partners-and-projects/pressoo

Portland Common Data Model

A commonly mentioned schema around LAM applications of LD is the emerging Portland Common Data Model (PCDM). The PCDM is growing out of the digital asset management system (DAMS) community in particular to serve Hydra-based systems but with a focus on supporting other RDF and Fedora-based services as well. PCDM is primarily focused on structural and administrative metadata and includes provisions for access control. As with many current data models, PCDM draws heavily on Dublin Core, RDF, FOAF, Internet Assigned Numbers Authority (IANA), and other related vocabularies. At its core, PCDM implements collections and objects that are subclasses of Object Reuse and Exchange (ORE) vocabularies. The PCDM also includes an access control notion that provides a granular rights-granting platform that includes read, write, append, and control methods. The PCDM is under development and is envisioned as an important part of the Fedora 4 deployment in the LAM community. More developments are expected in this area.

Portland Common Data Model
https://github.com/duraspace/pcdm/wiki

Linked Data Platform 1.0 Specification

The Linked Data Platform (LDP) 1.0 specification, released in December of 2014, defines a standardized method of interaction for LD applications. The LDP refers to resources that have relationships via containers and that can be manipulated through web standard behaviors (e.g., get, post, put, patch, delete, options head) and returns data in a prescribed way using Turtle and JSON-LD.[4] The LDP specification is published as a working group recommendation at this point, meaning that it is not yet endorsed as a specification by the W3C. The goal of LDP is to define a standard set of application behaviors and response formats. This would be a useful next step in standardizing LD applications. In addition, the fact that the LDP standard focuses on tracking direct and indirect relationships between resources and containers of resources means that the data model that it employs may be a good fit for LAM institutions seeking to create LD applications. Fedora 4 has adopted the LDP model with these goals in mind and uses the LDP specification to inform its implementation of create, read, update, and delete (CRUD) functions.[5]

FRBR Library Reference Model

The Functional Requirements for Bibliographic Records (FRBR) model has been in development and discussion since the 1990s, with Functional Requirements for Authority Data (FRAD) and Functional Requirements for Subject Authority Data (FRSAD) having been defined more recently. The IFLA FRBR working group has recently undertaken the consolidation of these three models to create the FRBR Library Reference Model (FRBR-LRM). This model incorporates authority and subject authority relationships without modifying the core works, expressions, manifestations, and items (WEMI) model that has guided FRBR. In combining the models, the user task Explore is drawn in from FRSAD but is also expanded to include the FRAD task Conceptualize.[6] Although this model is in early draft form and slated to be reviewed in 2016, it is worth noting that IFLA as well as other organizations are exploring how to manage the WEMI and other FRBResque relationships that are at the core of many of the LD-focused user tasks that the LAM community imagines will be impactful.

Linked Data Services

The building blocks of Linked Data platforms commonly employ an ingest and reconciliation service, a data storage platform, a SPARQL endpoint, and, in many cases, some sort of more user-focused discovery platform. The Yale Center for British Art, for example, harvests data using OAI-PMH using LIDO, indexes data using Apache Solr, provides data via an API service, and supports discovery and interaction through VuFind, websites, and other application plugins.[7] In contrast, the British Museum collection relies on a unified platform called OntoText to provide indexing and SPARQL services. OntoText provides a service called Self-Service Semantic Suite (S4), which provides a set of semantic and text analysis tools that stores output in an RDF graph database running as a database-as-a-service. S4 integrates with other knowledge graph platforms such as GeoNames, DBpedia, and Freebase.[8]

The survey of LD vocabularies in use from the systems and projects reviewed surfaced a wide range of vocabularies for LAM and other applications. As with the survey of projects and systems, the vocabularies and tools in use are too numerous to catalog comprehensively. Many of the sources used for this issue, including the OCLC survey results; websites including Linked Data and Schema.org; the BIBFRAME implementation register; the Linked Data incubator group; and research articles cited in this issue are good sources for exploring the vocabularies in use in the LAM LD community.

OCLC survey results
www.oclc.org/content/dam/research/activities/linkeddata/oclc-research-linked-data-implementers-survey-2014.xlsx

Linked Data
http://linkeddata.org

Schema.org
http://schema.org

BIBFRAME implementation register
www.loc.gov/bibframe/implementation/register.html

Linked Data incubator group
www.w3.org/2005/Incubator/lld

Tools and Systems

There has been considerable growth in available tools to convert metadata to LD, in systems to serve LD, and in applications to query LD over the last few years. Tools already well known in the LAM community, including MarcEdit, OpenRefine, and RIMMF3, all provide LD-related editing functions. SPARQL command-line tools such as ARQ are increasingly common in the literature, and there is a wide range of triplestores available to store RDF data. For interested readers, two good sources of LD-related tools include the series of OCLC surveys (see chapter 1) and survey articles on *Wikipedia* and the W3C. For the reader looking for quick suggestions, a survey of the OCLC results indicates that Dydra, OpenLink Virtuoso, Jena, SESAME, and AllegroGraph are all common tools. Increasingly, there are cloud-based services available to support RDF triplestores, including Dydra. There is another set of tools focused on providing support for viewing LD data. These viewers include rdf:SynopsViz, Tabulator, OpenLink Data Explorer, and a range of other viewers. The W3C site on Semantic Web tools remains an up-to-date catalog of tools as well as standards and best practices.

Wikipedia: Triplestore
https://en.wikipedia.org/wiki/Triplestore

W3C: Large TripleStores
www.w3.org/wiki/LargeTripleStores

rdf:SynopsViz
http://synopsviz.imis.athena-innovation.gr

W3C Semantic Web wiki, Category:Tool
https://www.w3.org/2001/sw/wiki/Category:Tool

Tabulator
https://github.com/linkeddata/tabulator

OpenLink Data Explorer
http://ode.openlinksw.com

LAM-specific tools in the LD community tend to center on a specific vocabulary or use. The BIBFRAME editor and other tools made available by LoC and Zepheria, for example, provide support for working with BIBFRAME and related metadata but are not appropriate for more generalized work. Other tools common in the LAM community, such as ArchivesSpace, do not include built-in editor support that is LD-focused but are designed around principles of linking and can make use of APIs and data integration and export tools that are useful in the LD community. Just as there was value in tools that sought to automatically catalog web pages or extract metadata from structured HTML, there is an emerging set of tools dedicated to harvesting and transformation of LD in web pages. One such tool is the RDF Translator developed by Alex Stolz. This tool supports input via RDFa, Microdata, XML, N3, NT, and JSON-LD and translates that output to RDFa, microdata, pretty-xml, XML, N3, NT, and JSON-LD formats. The service is built on a Python library (RDFLib) and also uses pyRdfa, pyMicrodata, and rdflib-jsonld libraries. As this issue finds in many cases, Python and Python-related libraries are becoming a common platform for LD work across LAM and other institutions.

RDF Translator
http://rdf-translator.appspot.com

A similar tool that facilitates working with JSON-LD data is the JSON-LD Playground. Similar to the RDF Translator, the JSON-LD Playground tool provides different serializations of JSON-LD data, including translation into N-Quads and multiple forms of JSON data. While the focus of this issue is on LD metadata, another area of interest is RDF and LD visualization tools. Tools commonly used in the community include Gephi and Tableau. Ontology-specific visualization tools, such as the WebVOWL platform, provide the ability to visualize FOAF and other ontologies (http://vowl.visualdataweb.org/webvowl.html). In addition to these client-based tools, web-based tools such as Node.js, D3.js, and MongoDB are increasingly common in helping to display LD relationships.

JSON-LD Playground
http://json-ld.org/playground/index.html

WebVOWL
http://vowl.visualdataweb.org/webvowl.html

As LD platforms mature, more "comprehensive" or end-to-end tools are becoming available. One system that is featured in Wood et al.'s *Linked Data* is the Callimachus project, an LD ingest, hosting, and publishing platform.[9] This platform includes template systems for web publication, allowing authors to create Semantic Web applications. The platform adheres to each of the five building principles of LD (i.e., open on the web, machine-readable, non-proprietary, RDF-based, linked). The publishers of Callimachus compare it to content management systems (CMSs), differentiating it from these platforms in that Callimachus primarily manages structured data. Another similar tool, Graphity, provides a unified data publishing platform that includes an LD client, publishing platform, and processing engine. Like other tools, Graphity is available under an open-source license, although a commercial provider (GraphityHQ) provides commercial services. Another such tool, Arches, is a cultural heritage inventory and management system. Although this platform was not necessarily designed around LD principles, there are an increasing number of use cases related to how this platform is making use of LD, including one connected with the city of Los Angeles, California.[10]

Graphity
https://github.com/Graphity

Issues in LD Translation

Enhancing Data via LD

A common use case for LD is the use of vocabularies and authorities to create metadata with more obvious community value. While the LAM community as a whole appears to agree on this goal and the value of the work, there is still much work to do in creating the tools that enable widespread normalization. Johnson and Estlund suggested a number of potential outcomes from LD processes, including removal of "noise," normalized presentation, assignment of URIs for curated objects, and migration from legacy metadata to new LD vocabularies.[11] By removing "noise," Johnson and Estlund mean "eliminating valueless metadata entries" such as elements without content or values that essentially say "unknown." One application of this idea of URI resolution has been documented by Klein and Kyrios.[12] The project matched VIAF records against *Wikipedia* entries using the Pywikipediabot framework, a Python-based *Wikipedia* framework. Starting with VIAF clusters with a *Wikipedia* link, associated *Wikipedia* pages were scanned for content. One of the primary outcomes of this work is the notion that the VIAF bot may be a model for application with other types of data. It successfully connected VIAF data and *Wikipedia* pages at the "hundreds of thousands" of pages level.

The generation of LD through automated text and metadata analysis is an area where research is advancing the integration of tools, including text analysis, natural language processing (NLP), and connection with existing authority vocabularies. Pattuelli et al., for example, developed a Python-based platform to match DBpedia URIs and LoC Name Authority File (NAF) records as well as applying named-entity recognition using the Natural Language Toolkit (NLTK) platform.[13] Similarly, some libraries are using programs to bring LD into discovery platforms. For example, Hatop designed a platform to create a Solr index using LD sources.[14]

Natural Language Toolkit (NLTK 3.0) Documentation
http://nltk.org

Conversion Strategies

The conversion of metadata to LD is one of the more complex topics in the LD community, often complicated by issues of scale and diversity of metadata as well as the fact that LAM institutions have not yet settled on new systems, meaning that LD systems often contain secondary or derivative instances of metadata. Two strategies in particular around conversion, iterative (i.e., retransforming metadata as new features and requirements are integrated) and cumulative (i.e., building on previously transformed metadata) are commonly used. OCLC, for example, combines data from production and experimental processes to enhance MARC records and publish new data as Linked Data using a cumulative process. OCLC's new model for representing Works is motivated by FRBR concepts and algorithms but follows its own set of relationships to express the creative work.[15] This identifier is represented via RDFa as well as via the OCLC xID service.[16] In contrast, the LoC BIBFRAME tools encourage iterative transformation through the regular incorporation of enhancements that require the complete retransformation of all data.

Although the next clear step, particularly in the bibliographic arena, is to get to a level of system and schema maturity to move away from older systems and standards, it appears that this is still an aspiration rather than a realized goal for most projects. The Oslo Public Library's transformation to LD is an example of one project that has reached that goal, moving away from its old ILS to LD metadata using the Koha ILS in early 2015.[17] The Oslo Public Library was an early innovator in RDF and LD research, having developed MARC2RDF in 2011 as well as experimenting with LD-based services.

marc2rdf
https://github.com/digibib/marc2rdf

Conclusion

Chapter 3 has explored the systems, vocabularies, and standards in use in the LAM community to generate or make use of LD and has explored key issues in LD generation—options for enhancing LD as well as approaches to conversion of existing metadata to LD. Given the number of state of adoption reports that have been completed in recent years as well as the upcoming release of new survey results on adoption, this report did not seek to provide a comprehensive listing of tools, standards, and services. Rather, this chapter focused on example tools and standards and identified themes and trends in more depth. In chapter 4, we consider several of these themes in more detail and consider what the coming year might hold in LD exploration and adoption.

Notes

1. Erik T. Mitchell, "Library Linked Data: Research and Adoption," *Library Technology Reports* 49, no. 5 (July 2013), 46.
2. David Wood, Marsha Zaidman, and Luke Ruth, with Michael Hausenblas, *Linked Data: Structured Data on the Web* (Shelter Island, NY: Manning Publications, 2014); Bernadette Hyland and David Wood, "The Joy of Data: A Cookbook for Publishing Linked Government Data on the Web," in *Linking Government Data*, edited by David Wood (Cham, Switzerland: Springer International Publishing, 2011), 3–26. See also Eric M. Hanson, "A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project," *Serials Review* 40, no. 4 (2014): 251–58, http://dx.doi.org/10.1080/00987913.2014.975887; Bernadette Hyland, Ghislain Atemezing, and Boris Villazón-Terrazas, "Best Practices for Publishing Linked Data," W3C Working Group Note, January 9, 2014, www.w3.org/TR/ld-bp; Martin Malmsten, "Exposing Library Data as Linked Data," Web Technology Laboratory, Ferdowsi University of Mashhad, 2008, http://wtlab.um.ac.ir/images/e-library/linked_data/other/Exposing%20Library%20Data%20as%20Linked%20Data.pdf; Silvia. B. Southwick, "A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies," *Journal of Library Metadata* 15, no. 1 (2015): 1–35, http://dx.doi.org/10.1080/19386389.2015.1007009.
3. Eleni Tsalapati, Nikolaos Simou, Nasos Drosopoulos, and Regine Stein, "Evolving LIDO Based Aggregations into Linked Data" (paper presented at CIDOC 2012, Helsinki, Finland, June 10–14, 2012), http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2012/simou.pdf; Ellen Van Keer, "Moving from Cross-Collection Integration to Explorations of Linked Data Practices in the Library of Antiquity at the Royal Museums of Art and History, Brussels," in *Current Practice in Linked Open Data for the Ancient World*, ISAW Papers 7, edited by Thomas Elliott, Sebastian Heath, and John Muccigrosso (New York: Institute for the Study of the Ancient World, 2014), 5–8, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/vankeer.
4. Nandana Mihindukulasooriya and Roger Menday, eds., "Linked Data Platform 1.0 Primer," W3C Editor's Draft, September 11, 2015, https://dvcs.w3.org/hg/ldpwg/raw-file/default/ldp-primer/ldp-primer.html.
5. Greg Jansen, "Linked Data Platform (Slides)," last updated March 23, 2015, https://wiki.duraspace.org/pages/viewpage.action?pageId=68065495.
6. Pat Riva and Maja Žumer, "Introducing the FRBR Library Reference Model" (paper presented at IFLA World Library and Information Congress, Cape Town, South Africa, August 16–20, 2015), http://library.ifla.org/1084.
7. "In Depth," Yale Center for British Art, accessed July 28, 2015, http://britishart.yale.edu/collections/using-collections/technology/in-depth.
8. "British Museum Semantic Web Collection Online," British Museum website, accessed July 28, 2015, http://collection.britishmuseum.org.
9. Wood et al., *Linked Data*.
10. Annabel Lee Enriquez, "HistoricPlacesLA.org, Powered by Arches v3.0, Launches in Los Angeles," Arches, March 4, 2015, http://archesproject.org/historicplacesla-org-powered-by-arches-v3-0-launches-in-los-angeles.

11. Thomas Johnson and Karen Estlund, “Recipes for Enhancing Digital Collections with Linked Data,” *Code4Lib Journal*, no. 23 (January 17, 2014), http://journal.code4lib.org/articles/9214.
12. Maximilian Klein and Alex Kyrios, “VIAFbot and the Integration of Library Data on Wikipedia,” *Code4Lib Journal*, no. 22 (October 14, 2013), http://journal.code4lib.org/articles/8964.
13. M. Cristina Pattuelli, Matt Miller, Leanora Lange, Sean Fitzell, and Carolyn Li-Madeo, “Crafting Linked Open Data for Cultural Heritage: Mapping and Curation Tools for the Linked Jazz Project,” *Code4Lib Journal*, no. 21 (July 15, 2013), http://journal.code4lib.org/articles/8670.
14. Götz Hatop, “Integrating Linked Data into Discovery,” *Code4Lib Journal*, no. 21 (July 15, 2013), http://journal.code4lib.org/articles/8526.
15. Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter, *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*, Synthesis Lectures on the Semantic Web: Theory and Technology (Morgan & Claypool, April 2015), 64, http://dx.doi.org/10.2200/S00620ED1V01Y201412WBE012.
16. “WorldCat Work Descriptions,” OCLC Developer Network, accessed September 14, 2015, https://www.oclc.org/developer/develop/linked-data/worldcat-entities/worldcat-work-entity.en.html.
17. Asgeir Rekkavik, “RDF Linked Data as New Cataloguing Format at Oslo Public Library” *SCATNews*, Newsletter of the Standing Committee of the IFLA Cataloguing Section, no. 41 (June 2014): 13–16, www.ifla.org/files/assets/cataloguing/scatn/scat-news-41.pdf.

The Evolving Direction of LD Research and Practice

Emerging Issues in LD

In chapter 1, the author posed the question, "How should LAM institutions balance the need for generalized LD models that encourage interoperability with external community members against the need for highly granular internally focused standards?" This question is one example of a continuing discussion in the LAM community that exemplifies the current state of adoption of LD. Although the community as a whole is moving in the same direction, many paths are being taken, and clearly not all of these paths will arrive at the same place.

At the same time, while there are still many technical issues to resolve in LD adoption, the LAM community has made considerable progress in the past two years in building proof-of-concept tools, production vocabularies, and LOD-enabled services that demonstrate how data can be transformative in supporting information services rather than simply being useful. In chapters 2 and 3, this issue examined projects, tools, services, and vocabularies in more detail. The tools, vocabularies, and programs reviewed in these chapters are being informed by philosophical perspectives in the LAM community, including the value of data openness, the importance of standards and approaches for defining and maintaining standards, and approaches to system development.

Data Openness

The common practice that LAM communities forged around open-source development and licensing is now influencing how we approach making data open. In fact, while LAM institutions are choosing different open-use licenses, there is much shared practice around the creation and dissemination of open data. There are exceptions, however, as many libraries have bibliographic data from outside suppliers without having the ability to make that data available to their users under an open license. Likewise, some institutions have data policy rules that make publishing data as open data difficult. One such policy issue is often the ability to allow others to make commercial use of published data. Perhaps a much larger issue, however, is the fact that libraries are creating less metadata than they used to and are licensing much more of it from outside suppliers, meaning that the ability to drive the discussion around open metadata is being limited. This is a simple reality given the shift of information institutions to the web and the widespread licensing of electronic content. In fact, metadata generation in general is an area that requires serious consideration as information institutions and the information communities that serve them ask questions about how to afford to create metadata for the newly published information objects.

The overall lack of data openness and transparency is an influential factor in the library discovery service market. Although there is an open discovery initiative led by NISO, there is no real momentum yet behind the notion that LAM institutions should be able to make this data openly available or that data can be separated from the discovery systems that provide access to it. This creates an unfortunate circumstance in which libraries in particular are purchasing metadata multiple times and in multiple information systems. At the same time, libraries are seeking out cloud providers to make use of and manage this new metadata and must find viable commercial models to ensure that system producers are incentivized to provide the desired services. It is entirely feasible that LAM institutions should consider opting out of licensed metadata and select publishers and vendors that produce metadata in a consistent format for open use. In fact, many publishers already build metadata for the web and are directing users to their own discovery portals, often with the purpose of selling access to licensed content that may be available through a user's institutional affiliation. This practice is having a considerable negative impact on communities of valid users whose use of the web to find resources is not supported with the systems and services required to allow them to make use of the license fees their institutions have paid.

In a legal context, the 2014 court decisions regarding uses of digitized book data by HathiTrust and Google indicated that nonconsumptive use of digitized full text falls under fair use.[1] These decisions support the efforts of LAM institutions to make new uses of copyrighted and noncopyrighted resources in new ways, with a particular emphasis on using contextualized data to support discovery and research. The related discussion about whether or not metadata is copyrightable is an important one in the LD community.[2] The DPLA took a stand in 2013 that "the vast amount of metadata is not copyrightable."[3] Such a stance is appealing in LD circles as it simplifies or removes issues associated with reusing data and making your own data available.

While many LAM institutions are turning to Creative Commons (CC) licenses that support reuse with or without attribution, reuse by commercial or noncommercial entities, and derivative or original form use only, there is no true consensus on how to ensure that data licenses are consistent and easy to apply in an automated fashion. For example, while many libraries use CC, OCLC makes use of the Open Data Commons (ODC) licenses. The ODC makes three licenses available, an "attribution" license (ODC-By), a public domain license (PDDL), and an "attribution and Share-Alike" license (ODbL). The key difference between ODC-By and ODbL is that the "Share-Alike" license allows you to adapt a dataset and rerelease it as long as you use the same license. In fact, some suggest that metadata should in fact be in the public domain and not made available via a data license, the key impact being that data licenses are in themselves restrictive and can lead to improper attribution.[4]

Open Data Commons
http://opendatacommons.org

Standards Compatibility

A similarly large issue related to LD is the issue of standards adoption and cross-community compatibility. As LAM projects are moving forward, the organizations are making highly impactful decisions about vocabularies to use, required granularity of selected approaches, and potential reuse purposes of published datasets. Without widespread agreement over how these vocabularies exchange standards should operate, LAM institutions may find themselves in a difficult-to-navigate mixed-metadata world. One such confusing area that has arisen in the past few years is the use of the BIBFRAME Lite name by Zepheria to represent an alternative to the BIBFRAME vocabulary. The reuse of the name is introducing some confusion into an already complex discussion around related standards.

Although there is yet to be a singular approach around metadata schemas, more consensus is emerging around serialization of LD. While LAM institutions are using a range of serialization standards, including RDFa, RDF/XML, Turtle, N-Triples, and JSON-LD (i.e., the predominating serialization formats for LAM LD), the stability of the RDF data model across these serialization standards as well as the growth in transformation tools, has meant that this is not as complex an issue as one might think. In fact, in the past two years, JSON-LD has grown as a standard that is more robust and appears to have a preference among the LD LAM developer community, even though it is not as granular as RDF/XML. The inclusion of JSON-LD in the RDF 1.1 specification was a signal that the issues with specificity and granularity in this serialization have largely been addressed.

Lack of Supporting Systems

It is fair to say that that LOD LAM applications are still in a "roll your own" phase of development. LAM institutions that seek to deploy LD applications are often exploring technical platforms and making localized decisions about the best systems to select. While systems do not need to be identical—in fact, it is advantageous for them to not be identical—the fact that LAM institutions are still having to select triplestores, SPARQL engines, indexing platforms, and other services means that there is still a relatively high bar for institutions to cross in taking up LD projects. A later section for this chapter explores some of the systems in use in common projects and seeks to identify some selected systems that appear to be bringing the various LD publishing tools together (e.g., triplestore, SPARQL endpoint, index, discovery interface, and creation interface).

Another area of system development that is also very much in focus is the extent of vendor support for LD applications. Library system vendors have taken different approaches over the past two years in developing the next generation of information systems. At ALA 2015, many ILS vendors expressed support for BIBFRAME and spoke to broad roadmaps around adoption. Chapter 2 explored how some research projects focused on transforming bibliographic data are making use of existing systems, particularly open-source platforms. At the same time, there does not yet appear to be a comprehensive turnkey solution for libraries seeking to create and publish LD. On the systems front, it appears that more progress has been made in the archival and museum communities. Similar challenges still exist in these communities, although the information systems they use, such as ArchivesSpace, CollectionSpace, Fedora, DSpace, and other related tools, are already aligned around metadata standards that can be easily converted to LD for publication.

Whether or not getting to the turnkey level is necessary to see LOD adoption grow is a fair question, but it is clear that libraries are investing in LOD as a way to drive down costs as well as increase value. It is not feasible or sustainable for LOD systems to ultimately cost more than their current metadata publishing counterparts (e.g., Integrated Library Systems, Digital Asset Management systems), but it is likely that this is the reality for early adopters who need to invest in both traditional and new LD systems simultaneously.

Important Questions in the Linked Data Community

How Have Standards Evolved over the Last Two Years?

One of the key difficulties in creating LD and making it available is in defining the use cases that make sense and will have value to the community. Publishing data in some serialization of RDF is not especially useful or interesting if it does not capitalize on links to other datasets or provide new opportunities for computational analysis of data. As the LD community has grown through experimentation and project developments in the past few years, more best-case examples of how to create and publish metadata have been explored and reported. Perhaps the clearest expression of these principles is in a working group report titled "Best Practices for Publishing Linked Data." This guide surfaces ten steps for publishing Linked Data, reproduced in the list below.

1. Prepare Stakeholders
2. Select a Dataset
3. Model the Data
4. Specify an Appropriate License
5. The Role of "Good URIs" for Linked Data
6. Standard Vocabularies
7. Convert Data to Linked Data
8. Provide Machine Access to Data
9. Announce to the Public
10. Social Contract of a Linked Data Publisher.[5]

Of these ten steps, five of them focus on policy and social-good issues rather than purely technical issues or topics. This document, as well as many others, cites Hyland and Wood's work on creating Linked Data from a technical perspective,[6] and as a result, a more policy-focused document is a useful and somewhat unique contribution to the Linked Data publishing space. Although the author will not replicate the core recommendations of the document here, many key items are worth highlighting—the need for documentation using self-descriptive techniques, the importance of persistence, and the importance of supporting multiple languages.

What Are Communities of Practice Saying about the Direction of Linked Data, and How Have the Issues around LD LAM Changed in the Past Two Years?

Overall, the focus of the library community remains on a conversion to LD, and we have seen considerable development efforts to make the conversion of data as well as the creation of new data possible. In fact, new projects, such as LD4L and BIBFLOW, point to future potential production systems that may advance LD work. At the same time, libraries are challenged to demonstrate impact and prove that they have capacity. In a summer workshop held at UC Berkeley, for example, the common discussions around capacity building paralleled discussions around innovation and new projects. It is clear from the state of the projects that libraries undertaking LD efforts now must be prepared to continually convert data and to reconvert data to capitalize on new areas of development and granularity.

The state of adoption across libraries of all sizes remains limited although the tools are becoming more available and metadata standards are becoming more resolved and manageable. Whether or not simpler systems are the correct next step remains to be seen, but after several years of development it appears to be a necessary step.

With these forward steps, particularly via projects led by OCLC and the Library of Congress and grant-funded initiatives, the LAM community is pointing toward a robust future for LD. At the same time, it is also worth remembering that the community as a whole has yet to see transformative impacts from LD generation that resonate for all organization types and sizes. The goals of web visibility, research reuse, and granular preservation remain important, and it is clear that LAM institutions are driving their systems toward these purposes. Whether or not that will have a real impact in the research community remains to be seen.

What Role Do We Expect Large-Scale Projects to Play in Linked Data?

This is a difficult question to answer given the grassroots approach to LD projects in the LAM community at the moment. Traditionally, central players in the LD space, including LoC, OCLC, and NLM, are being complemented by players such as Europeana, the British Library, and multi-institutional cooperatives such as LD4L. A foundational discussion that is occurring among these groups centers on community alignment—especially how LAM institutions can make their data align with other communities of practice. OCLC, for example, has recently begun exploring the notion of a "Knowledge Vault" for libraries, a concept built on Google's work in knowledge graphs.[7] Likewise, companies such as Zepheria and its LibHub initiative continue to have a strong influence on the direction of the community, and there are a number of examples of secondary uses of metadata to create field trips in Google's mobile field trip tool to support visualization services on top of DPLA harvested data and to publish new vocabularies that aim to turn LAM data into LD.

Google: Customizing Your Knowledge Graph
https://developers.google.com/structured-data/customize/overview

Field Trip
https://www.fieldtripper.com

It should not be surprising that as organizations like DPLA and Europeana develop, that issues of sustainability and governance become important. The fact that both of these organizations included these issues in their strategic plans indicates how interesting it is timing-wise and how pressing the topic of the value of these organizations is for LAM institutions in their related countries. In fact, one of the key issues surrounding efforts of LD publishing is how to ensure that the LD that is published remains available via the published URIs over time.

The Europeana-proposed funding model is interesting in its detailed exploration of customer groups and benefit analysis.[8] The groups include end users, cultural institutions and their associated member states, project funders, and creative industries. The projected cost of Europeana during the next three years is anticipated to be €10 million annually, or approximately $10.8 million (US). While this is not an insurmountable funding challenge, gathering this level of funding for other national initiatives will likely be a focus in the coming years.

How Will LD Influence Cataloger Work and Notions of Value Moving Forward?

Seeman and Goddard explored the pressing question "what now" in relation to guiding catalogers in the creation of metadata as these LD standards are evolving.[9] Observing that much of the core work of cataloging (e.g., authority control, access point assignment, disambiguation) remains philosophically, if not functionally, the same, they suggested that this work, taken along with commonsense approaches, may make capacity for forward progress. It goes without saying that in a community driven by process and standards, the long-term discussions around a set of emerging but fluid standards without action does not serve the community well.

In fact, the LAM community as a whole has yet to tackle the true early adopter problem. Given the high level of collaboration and interoperability developed throughout the preceding century among libraries in the sharing of metadata and cooperative resource sharing, it may be that there is recognition that the stakes for early adopters are high. One such technique that is being suggested is embedding URIs in traditional MARC records. Interestingly, this notion was discussed in a 2010 LoC brief.[10]

A question related to value is whether or not LAM metadata, when transformed into LD, becomes something more than it was as unlinked metadata. Does the creation of LD, for example, make the metadata a "first class" research object? Does the publishing of LD create new streams of research or support new research methods? The fact that some institutions are publishing datasets in a more complete form points to the idea that this is possible, yet LAM metadata has typically focused on resource description and object management, areas of information that do not necessarily lend themselves to expansive research questions.

Current Education Opportunities

Challenges around bringing library staff up to speed on new approaches in metadata creation and management continue to impact the community. Some institutions have reported using the Juice Academy series, particularly the XML program. In addition, the Educational Curriculum for the Usage of Linked Data (EUCLID) project publishes a comprehensive textbook focused on Linked Data creation and use. In fact, this issue is as pressing for LIS schools as it is for practicing professionals. As a result, there is likely to be more restructuring of LIS curricula in the coming years as traditional work in resource description shifts and new concepts and skills are needed to work with LD technologies.

Library Juice Academy
http://libraryjuiceacademy.com

EUCLID
http://euclid-project.eu

Conclusion

This issue has explored current practice and emerging trends in LD LAM projects and activities and has considered some of the broad questions and topics of future exploration. In doing research for this issue, the author found that in the past two years considerable research and publication had occurred documenting specific technical projects, applications, vocabularies, and community best practices. In fact, the amount of literature and activity in this area is large enough to defy concise analysis. If anything, the exploration of trends, projects, and topics indicates that while the LAM community may be moving in a common direction, we are doing so in a number of parallel, if not identical, paths.

Notes

1. Ian Chant, "Appeals Court Upholds Wins for Fair Use in HathiTrust Case," *Library Journal*, June 12, 2014, http://lj.libraryjournal.com/2014/06/litigation/appeals-court-upholds-wins-for-fair-use-in-hathitrust-case.
2. Karen Coyle, "Metadata and Copyright: Peer to Peer Review," *Library Journal*, February 28, 2013, http://lj.libraryjournal.com/2013/02/opinion/peer-to-peer-review/metadata-and-copyright-peer-to-peer-review.
3. Ibid.
4. Timothy Vollmer, "Library Catalog Metadata: Open Licensing or Public Domain," Creative Commons, August 14, 2012, http://creativecommons.org/tag/open-data-commons-attribution-license.
5. Bernadette Hyland, Ghislain Atemezing, and Boris Villazón-Terrazas, "Best Practices for Publishing Linked Data," W3C Working Group Note, January 9, 2014, www.w3.org/TR/ld-bp.
6. Bernadette Hyland and David Wood, "The Joy of Data: A Cookbook for Publishing Linked Government Data on the Web," In *Linking Government Data*, ed. David Wood (Cham, Switzerland: Springer International Publishing, 2011), 3–26.
7. Merrilee Proffitt, Bruce Washburn, Diane Vizine-Goetz, and Roy Tennant, "OCLC Research Update" (presentation at ALA Annual Conference and Exhibition, San Francisco, CA, June 25–30, 2015), www.slideshare.net/oclcr/oclc-research-update-ala-annual-2015?from_action=save.
8. "Europeana Strategy 2020: Network & Sustainability (Draft)," Europeana, May 30, 2014, http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana Strategy Network Sustainability.pdf.
9. Dean Seeman and Lisa Goddard, "Preparing the Way: Creating Future Compatible Cataloging Data in a Transitional Environment," *Cataloging & Classification Quarterly* 53, no. 3/4 (2015): 331–40, http://dx.doi.org/10.1080/01639374.2014.946573.
10. RDA/MARC Working Group, "Encoding URIs for Controlled Values in MARC Records," MARC Discussion Paper No. 2010-DP02, MARC Standards, Library of Congress, December 14, 2009, www.loc.gov/marc/marbi/2010/2010-dp02.html.

Notes

Notes

Library Technology REPORTS

Upcoming Issues	
February/ March 52:2	**Learning Management Systems: Tools for Embedded Librarianship** by John Burke and Beth Tumbleson
April 52:3	**Trends: Accessibility, Ecosystems, Content Creation** by Nicole Hennig
May/June 52:4	**Privacy in Library Automation Products** by Marshall Breeding