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.

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.

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. 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. 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.

The BIBFRAME community as documented on the LoC BIBFRAME website includes a number of test projects that follow some level of BIBFRAME work. 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.

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. 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. 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.

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. 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. 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. 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.

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. 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
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.

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 British 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.
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*. 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.

WorldCat Works is an OCLC service centered on publishing LD about FRBR-esque 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. 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*. 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.

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 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. 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. 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.

**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. 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. 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. 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

19. Ibid., 12.
25. Godby et al., Library Linked Data.
27. Fallgren, “Experimentation with BIBFRAME.”
28. Dean B. Krafft, “Project Description,” Linked Data for

