Review of E. T. Mitchell's *Library Linked Data (Library Technology Reports, Vol. 52)*

Dallas Long

Associate Dean, Illinois State University

Abstract: Review of Mitchell, E.T. (2016). *Library Technology Reports: Vol. 52: Library linked data: Early activity & development. Library Technology Reports 52(1).* Chicago, IL: American Library Association.

Keywords: linked data, digital content, metadata, BIBFRAME, SPARQ, scripting languages, book review

The premise of linked data is that users will create new knowledge by exploring connections between related pieces of information held by more than one library or institution. In order to make the data held in their catalogs more broadly accessible and shareable, library, archives, and museum (LAM) communities must develop new tools, standards, and vocabularies that provide context to data in ways that will help search engines process and retrieve search results. In the July 2013 issue of *Library Technology Reports*, Mitchell explored the research and practice related to metadata and the emerging information systems then under development by the LAM communities. In this report, Mitchell reviews the state of linked data adoption by the LAM communities and pays special attention to the vocabularies, tools, and systems these communities are rapidly building.

Mitchell's experience as Associate University Librarian for Digital Initiatives and Associate CIO at the University of California at Berkeley provides him a unique perspective into the opportunities and challenges that nurture or stymie the adoption of linked data workflows, services, and technologies in the LAM communities. In the succinct yet comprehensive style common to most *Library Technology Reports*, Mitchell's report is clearly written and easily digestible. This review of the current state of linked data activity over the past several years will be most useful to information professionals who are not intimately involved in linked data work themselves but want to stay informed of the current trends. Information technologists, cataloging and metadata librarians, and digital content creators will find this report particularly helpful. However, Mitchell assumes his readership has at least a basic grasp of linked data and some knowledge of the major players at work in this still emerging area. His report is not a primer for those with no previous familiarity with linked data.

The report is organized into four chapters. The first chapter addresses the broad state of linked data adoption by examining the findings of a 2014 survey conducted by OCLC staff. The survey identified 172 currently in-development projects employing substantial linked data components. The primary reasons institutions were engaged in linked data projects were to expose their data to a larger audience on the web and to support future research endeavors. The survey described 76 of the projects and found that the types of data most commonly published as linked data included descriptive metadata, bibliographic data, authority files, statistical data, ontologies and controlled vocabularies, geographic data, data about museum objects, and encoded archival descriptions or finding aids. The projects used linked data to enrich their own bibliographic descriptions, automate authority control,

enable dataset discovery, and populate auto-suggests in search bars, among other applications.

In addition to the OCLC survey, Mitchell recommends interested readers follow the annual updates of the ALCTS Technical Services Directors of Large Research Libraries Interest Group for additional information about specific projects. He notes the 15 public reports from June 2015 showed a range of projects and other ongoing research interests, such as efforts to improve or normalize vocabulary creation and the use of APIs as stopgap measures when linked data is not fully available.

In the second chapter, Mitchell selected specific large-scale projects that represent advanced momentum and might be considered models for the LAM communities, including BIBFRAME and BIBFRAME Lite, the Digital Public Library of America, Europeana Digital Library, several projects by the British Library, the British Museum, and WorldCat.org and other WorldCat works. He considered three questions as he evaluated each project: What is the overall goal and focus of the platform? How does this platform situate itself in the context of other information systems? What unanswered questions does this platform raise? Although all most of the projects share similar goals of serving as portals of discovery to unique resources across multiple institutions, Mitchell concludes that rights management issues complicate large-scale projects. Unclear rights or multi-layered rights impede the organizations' abilities to share content online and to promote the integration of their resources into teaching and research initiatives, who are the ultimate consumers of such resources. However, government agencies and private funders are taking interest and are now supporting efforts to explore right management issues further. Additionally, Mitchell finds the nature and goals of the projects suggest multiple broad trends are emerging in linked data research and initiatives, including the interoperability across metadata standards, the increasing need to bring URI minting services together in order to ensure vocabulary adoption is manageable, and an increasing interest in using Social Psychological Answers to Real-world Questions (or SPARQ) end-points as aspects of data publishing.

Mitchell explores in the third chapter the trends concerning the specific tools, vocabularies, and systems used by the projects he reviewed in the second chapter. He notes a growing number of common scripting languages such as JavaScript Object Notation for Linked Data (or JSON-LD) are growing in popularity for linked data (LD) frameworks in part because of its syntax and in part because of its ease of use. Mitchell describes many more languages and vocabulary-aware tools, and he concludes that there are simply too many in use to list comprehensively or to evaluate effectively. The diversity of resources make research issues such as interoperability all the more pressing.

Mitchell wraps up his report by posing important questions for the LAM communities to consider. He notes the central challenge for LAM communities is defining use cases that will truly have value to researchers. As the number of projects have multiplied since the publication of Mitchell's first report, he notes the wealth of published information has focused on technical issues. Instead, Mitchell encourages greater attention and focus on whether projects meet a social good. Additionally, he notes the need for the LAM communities to demonstrate whether their projects have impact. Ultimately, Mitchell concludes the amount of LD activity is extensive and moving forward but in many different directions. Linked data seems promising but it is still not yet in a place where users are connecting data and transforming it to something else in substantial ways. Still, it is a development that bears watching carefully.