Reference Guide: Linked Data for LIS Professionals and Graduate Students

Inna Gogina

LIBR 210 – Reference and Information Services

San José State University

May 11, 2015
Reference Guide: Linked Data for LIS Professionals and Graduate Students

URL: https://linkeddatalibguide.wordpress.com/1

The library, up until now, has been viewed as the place for reliable, authentic information. The catalog linked the users to vetted resources. Can we rethink cataloging to achieve something similar in the world of Google? I hope so.

- Deanna B. Marcum2
  Associate Librarian for Library Services
  Library of Congress

How wonderful it would be, to have a shared online space where people could come from anywhere in the world to search for information on any topic and find everything ever mentioned about the object of their interest. A fantasy? So I thought when this idea of a universal shared knowledge space first crossed my mind about ten years ago, when I was working as an indexer and cataloger at the USC Shoah Foundation Visual History Archive3. The same year, in 2005, Deanna B. Marcum, Associate Librarian for Library Services at the Library of Congress, quoted in the epigraph to this paper, expressed a similar hope. Years have passed since, and while discussing the future of libraries in the LIBR 248 Cataloging and Classification Class of spring 2015 at San José State University, I learned that work has already begun to turn what I thought was a fantasy, into reality. During the first decade of the 21st century, information technology professionals and librarians started driving a revolutionary change to discovery systems and the way we think about bibliographic records, by developing the Linked Data environment and building the Semantic Web. The term, Linked Data, was coined by Tim Berners-Lee, director of the World Wide Web Consortium (W3C), in a design note about the Semantic Web project

1 Originally hosted on http://libraryschool.libguidescms.com/linkeddata-LIS; migrated to the new site in February 2018.
Based on the concept of relationships between various types of data, Linked Data facilitates the creation of the Semantic Web by developing structured metadata and exchange mechanisms for human knowledge in order to make it more consistently discoverable.

Bibliographic Framework (BIBFRAME), an initiative headed by the Library of Congress, paved the way for a Linked Data environment in the library field—the environment which is envisioned to be data-driven, rather than record-based. “A central activity to the Bibliographic Framework Initiative is the development of a new means for capturing and sharing bibliographic data. Included in this activity is pursuing a replacement of the MARC format as the common exchange currency for bibliographic data. This was one recommendation of the 2008 report from the Library of Congress’ Working Group on the Future of Bibliographic Control, *On the Record*[^4], and has been discussed in the community for a number of years,” Library of Congress explains (The Library of Congress, 2011).

Today, one can already see the Linked Data concept in action, whether it is in Google’s enhanced search results, the BBC’s wildlife and music pages, or the Facebook’s Like button (Wood, Zaidman, Ruth, & Hausenblas, 2014, pp. 8-11).

**A Scope Note: Purpose and Audience**

Linked Data is a complex concept that crosses over several theoretical domains, including computing, information technology, and library science. Geared toward the Library and Information Science professionals and graduate students, this reference guide presents personalized recommendations of literature on the topic of Linked Data as they are relevant to library and information professionals. Rather than aiming at being an exhaustive bibliography of everything available on the Linked Data topic, in scholarly literature and other sources, this reference guide is intended to provide the target audience with a selective, annotated list of

useful sources for conducting research on the topic. Designed primarily as an online resource, this guide is not tied to any particular geographic location. It is intended for academic libraries.

Reference Guide Structure

The online *Linked Data for LIS Professionals and Graduate Students* reference guide is available at [https://linkeddatalibguide.wordpress.com/](https://linkeddatalibguide.wordpress.com/). It has the following structure:

**Defining Linked Data**
- Definition of Linked Data
- Guides and Tutorials
- What’s in This Guide
- LibGuide Author

**Readings**
- Books
- Scholarly Articles

**Online Resources**
- Linked Data Projects (LIS)
- Video Presentations
- Websites
- Professional Associations

**Journals, Reference Works & Blogs**
- Journals
- Reference Works
- Blogs

**Discussion Board**

The main sections of the online reference guide can be accessed from the top navigation menu of the reference guide website.

This paper focuses on the methodology of the reference source selection for the Readings section of the guide. In particular, it discusses the criteria for inclusion of the academic books and scholarly articles from peer-reviewed journals, featured in the Readings section, and the challenges encountered in the selection process. The paper provides a bibliography of sources included in the research guide’s Readings, with annotations describing the source and explaining
why each was included. Finally, the paper includes a bibliography of sources consulted but not selected, explaining why the sources were not included.

The author hopes that a range of the sources presented in this guide will help current and future LIS professionals to get a good start in researching the Linked Data topic. The research tips provided in this guide are intended to serve as additional guidance in the research.

**Criteria for selection**

1. **Primary focus/Relevance to the topic.** The sources should focus primarily on the topic of Linked Data in its immediate relevance to the field of Library and Information Science. For example, I made sure that academic books included the authorized Library of Congress Subject Headings main term “Linked data,” as well as one of the broader terms of the main term, “Metadata” or “Semantic Web,” in their bibliographic records’ OCLC WorldCat Subject field.

2. **Authority of the author/Subject expertise.** I selected the sources written by experts in the fields of information technology and LIS, recognized scholars in the subject having an established history of research and writing on the subject of Linked Data.

3. **Time of publication.** As linked data is a relatively new concept, I limited my selection to the resources that were published after the Linked Data term coinage by its author, Tim Berners-Lee, i.e. works published in 2006 or later.

4. **Publication type.** For the Readings section of the reference guide, I chose academic works, in particular books and scholarly articles from peer-reviewed journals.

5. **Originality/Value.** In my selection of sources, I paid attention to the source’s originality/value and attempted avoiding the source redundancy.

6. **Language of publication.** For this guide, I selected sources published in the English language.
Excluded criteria: As Linked Data is a global-scale collaborative effort, and because this reference guide was not intended to be tied to any particular geographic location but rather serve as an electronic resource available online, I decided not to include the country of publication and local availability in the source selection criteria.

Selection of Sources

Challenges

The main challenge was to develop a search strategy that would allow selecting the LIS audience-appropriate and most useful sources from the multitude of materials that have been published in peer-reviewed literature and on the web. The search challenges are described below.

When searching for books written about Linked Data, I first used the Advanced Search in the SJSU DR. Martin Luther King Jr. Library online catalog and searched for “linked data” in the Title, and “book” in the Format field. Surprisingly, none of the search results was related to the topic of Linked Data in LIS. I then moved on to the OCLC WorldCat and using the same approach, search for “linked data” in books, found 3,907 books that had the key phrase “linked data” in the Title field. The problem was that such a large number of the results was too difficult to manage, in order to narrow down to the sources that matched the selection criteria I had set forth for my reference guide.

In the search for scholarly articles, I first tested keyword searching in the SJSU online Catalog for “linked data” in the Quick Search option, located on the online library’s landing page. As the quick search returned quite a large number of results, 142 at the library, and 1,148 via LINK Plus, I decided that I needed to adjust my search strategy.
Search strategy – LCSH approach

In order to identify the search terms for my source research, I decided to look into how the Linked Data topic is classified in the authorized Library of Congress subject headings. In particular, having consulted the main entry of the Library of Congress Linked Data authorized subject heading\(^5\), I chose to include the main term, “Linked data,” as well as its broader terms, in my search. As a result, I identified the following subject headings to use as my search terms:

- LC Subject Heading (main entry):
  - Linked data

- Broader Terms
  - Metadata
  - Semantic Web

Because rather than providing an exhaustive list of all resources on Linked Data, this reference guide was intended to cover the basics of the topic, and in the interest of time, I decided to exclude the variants and related terms of the “Linked data” LCSH from my search:

- Variants
  - Data, Linked
  - Linked open data
  - LOD (Linked data)
  - Open linked data
  - Opendata, Linked

- Related Terms
  - Uniform Resource Identifiers

---

\(^5\) [http://id.loc.gov/authorities/subjects/sh2013002090.html](http://id.loc.gov/authorities/subjects/sh2013002090.html)
Selection process

Armed with the proper search terms, I proceeded to explore the following online catalogs:

- SJSU King Library Catalog
- OCLC WorldCat

In order to identify the academic books for the reference guide, I used the OCLC WorldCat’s Advanced Search and searched for the following combinations: a combination of the “linked data” and “Semantic Web” Subjects OR a combination of “linked data” and “metadata” Subjects, AND the “book” Format AND the 2006-2015 Year range AND English Language. Once I retrieved the search results, I narrowed them down to the “Library Science, Generalities & Reference” topic.

In order to search for scholarly articles, I used the SJSU King Library online Library Literature & Information Science Full Text database. In the database, I used the advanced Boolean search for two combinations: “linked data” AND “Semantic Web,” alternating it with the “linked data” AND “metadata” combination, in the SU Subject fields, and pairing it with the “Academic Journal” Publication Type and “Scholarly (Peer Reviewed) Journals filter.

There was one scholarly article, I must note, for which I used a different selection method. Basically, I found it by a sheer accident. As I was researching library and information science professional associations, I came across the Journal of Library Metadata—a peer-reviewed journal focusing on the importance of metadata in libraries. And, as I was checking the most recent, March 2015, issue, I came across A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies by Silvia B. Southwick—a source that was a perfect match for all of my selection criteria.
Bibliography of Selected Sources

Books


This book covers the basics of the Semantic Web concept. It describes what it is, the components that comprise it, including Linked Data, and the various ways that libraries are engaged in contributing to its development in making library resources and services ever more accessible to end-users.

Matching subjects: Linked data, Semantic Web, Libraries and the Internet
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This eBook describes the progression of the BIBFRAME model developed by the Library of Congress and models based on Schema.org developed by OCLC, while also analyzing the alignment of the two models. Recommendations for closer alignment of the Library of Congress and OCLC linked data models are also discussed.

Matching subjects: Linked data, Metadata--Standards
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This book is a guide to the "how" of making datasets from digital collections of libraries, archives and museums publicly available, downloadable and accessible.

Matching Subjects: Linked data, Metadata
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This book discusses the advances in metadata in libraries, and touches upon relevant work in archives and museums. It examines linked open data through the lens of five building blocks of
metadata: data model, content rules, metadata schema, data serialization, and data exchange. The book features three case studies of current projects: the Europeana digital library, museum, and archive; the Digital Public Library of America; and the BIBFRAME initiative, guided by the Library of Congress.

Matching subjects: Linked data, Semantic Web, Libraries and the Internet
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This book presents the Linked Data model in plain, jargon-free language and offers practical techniques using everyday tools like JavaScript and Python. It explains foundational concepts such as HTTP URIs, the Resource Description Framework (RDF), and the SPARQL query language. Readers will learn to use various Linked Data document formats to create powerful Web applications and emerging Web standards to access, find and query structured data on the Web.

Matching subjects: Linked data, Semantic Web
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)

Scholarly Articles


The author of this article examines the history of a bibliographic record as a conceptual whole that includes all bibliographic information about a resource together in one place. She argues that with the advent of Semantic Web, the traditional conceptualization of bibliographic records affects the affordances and limitations of linked data, as a new “record” model, and calls for reconceptualization of bibliographic description.

Matching subjects: Linked data, Semantic Web
Additional LIS subjects: Cataloging—History, Bibliographic databases
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This is an award winning article by the San José State University alumna, Brighid Gonzales, who won the National Award for Research Paper. According to the SJSU, “the criteria for the award
include relevance to libraries and information technology, timeliness, originality of ideas, persuasiveness of arguments, quality of writing, clarity of presentation, and contribution to the continuing education of the LITA membership” (SJSU School of Information, 2015).

Matching subjects: Linked data, Semantic Web
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This article presents a detailed analysis of a linked data project in academic library, the North Carolina State University (NCSU) Libraries’ Organization Name Linked Data (ONLD) project (http://lib.ncsu.edu/ld/onld). The author shares specific approaches that the project team took to produce and publish their data set, and describes best practices for project planning, data clean up, conversion, and publishing linked data. This article is especially valuable now that more libraries begin publishing and using linked data, it shows them how they can start creating their own linked data sets.

Matching subjects: Linked data, Metadata, Semantic Web
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This article provides a representative overview of literature related to the idea of replacing MARC with a linked-data metadata structure, covering the period from 2002 through the 2012 release of the draft of the proposed bibliographic framework, BIBFRAME. In particular, the author examines key documents leading to the creation of the Library of Congress Bibliographic Framework Transition Initiative, along with some of the critical responses they received, to better understand the chain of ideas shaping BIBFRAME.

Matching subjects: Linked data, Semantic Web
Additional LIS subjects: MARC formats, Library of Congress classification
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


The author discusses the tension between the novelty of the Linked Data concept and the everyday work of the cataloger, which remains mostly unchanged. The originality of this article is in that the author not only calls for reconciling cataloging practice with a linked data future,
but also makes practical recommendations on how to help prepare the way for the emerging information environment.

**Matching subjects**: Linked data, Semantic Web  
**Additional LIS subjects**: Cataloging—Research, Bibliographic databases  
**Criteria met**: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


The author of this paper shares with the readers the challenges faced by the University of Nevada, Las Vegas (UNLV) Linked Open Data (LOD) project team in implementing the transformation of digital-collection metadata into LOD and the solutions that were adopted. This article is of particular value because the library literature lacks best practices for implementing Linked Open Data, as it is a relatively new topic. The author briefly introduces basic concepts of LOD and explains, step-by-step, how to transform digital-collections metadata into LOD, considering the open source technologies as tools for such a transformation. Prior experience by the author in leading a hands-on workshop on LOD indicates that as attendees perform practical tasks for generating triples (statements) from metadata, their understanding of LOD and its potential benefits are consolidated.

**Matching subjects**: Linked Open Data, Metadata, Semantic Web  
**Criteria met**: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)


This article evaluates the Bibliographic Framework (BIBFRAME)—an emerging framework developed by the Library of Congress for bibliographic description based on Linked Data. Using a case study approach, the author examines the BIBFRAME as means for harvesting and sharing bibliographic metadata over the web for libraries. Linked Data leverages the existing web infrastructure and allows linking and sharing of structured data for human and machine consumption. Even though the main search term, “Linked data,” was not included in this article’s Subject field, it was discussed in great detail in the abstract, e.g. “Much like Semantic Web, the goal of Linked Data is to make the web “data aware” and transform the existing web of documents into a web of data.” As the overall discussion revolved around the application of linked data in libraries, I decided that this article did match the Linked Data-oriented primary focus selection criterion and included this article in the selected sources.

**Matching subjects**: Metadata  
**Additional LIS subjects**: Libraries, Access to information
Criteria met: primary focus (1), authority (2), time of publication (3), publication type (4), originality (5), language of publication (6)

Bibliography of Excluded Sources

Below are several examples of the sources that were consulted but not included in this reference guide.

Books


Even though this source matched the primary focus criterion – it had a combination of the Worldcat subjects I had included in my selection criteria, “Linked data” and “Semantic Web,” it did not match the format criterion – this source was a thesis/dissertation, whereas I chose to include only academic books and scholarly articles from peer-reviewed journals.


This book had the following subjects in its OCLC WorldCat bibliographic record: Linked data; Database management; Data structures (Computer science). This book mostly focused on providing a comprehensive overview for database researchers and practitioners about this new publishing paradigm on the web, and showing how the abundance of data published as Linked Data can serve as a fertile ground for database research and experimentation. Even though the “Linked data” was the main subject term I was looking for, this book’s subjects did not include the broader terms, “Metadata” and “Semantic Web” I was looking for, in addition to the “Linked data” term, nor did they include any general LIS-related terms. I therefore decided not to include this book in the reference guide’s selected sources.


The OCLC WorldCat bibliographic record of this book did not include any of the selected terms in subject fields. Focusing primarily on RDA and serials cataloging, and touching upon the relationship between the RDA and linked data only in the epilog, this source did not match the primary focus criterion.


This book did not include the main term I was looking for, “Linked Data,” in the Subject field, and focused mostly on the existing metadata standards and practices, such as and Dublin Core, XML-encoded metadata, MODS (the Metadata Object Description Schema), and VRA Core (the
Visual Resources Association Core Categories). The only subject relevant to this guide’s selection criteria included in this book was “Metadata.” This book, therefore, did not match the primary focus selection criterion.

**Scholarly Articles**


Even though the author of the Linked Data concept, Berners-Lee, was one of this article’s authors, and even though this article discussed the concept I included in my primary focus’ subject terms, “Semantic Web,” I decided against including this article because it was too dated. I only included sources that were published after 2006. This article, therefore, did not match the time of publication criterion.

**Conclusion**

The significance of linked data for library and information environments of the future is threefold. Firstly, by including link data in the online descriptions of library resources, search engines and other information services on the web will help people more easily locate and access the library resources on the Internet. Secondly, linked data allows for more creative web applications based on library metadata. And lastly, it provides additional opportunities for cataloging efficiency, cost savings, and innovation (Online Computer Library Center, 2012).

National libraries and library consortia have already begun converting their bibliographic records into linked data. The projects and case studies discussed in this reference guide’s sources exemplify the value and the challenges of the linked data endeavor in the LIS field. There is still a huge unlocked potential for linked data in libraries’ efforts to provide access to human knowledge. It is my hope that by using this reference guide as a start of the research, LIS professionals and graduate students will gather initial understanding of the Linked Data concept and principles, as well as delve deeper into the understanding of linked data in action and their benefits for the library profession and the learning community at large.
References


http://www.w3.org/DesignIssues/LinkedData.html


[Video recording]. Retrieved from https://www.youtube.com/watch?v=fWfEYcnk8Z8


https://ischool.sjsu.edu/people/community-profile/brighid-gonzales

