

2016-04

Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata by Seth van Hooland and Ruben Verborgh (review)

Tharani, Karim

The Johns Hopkins University Press

Tharani, K.(2016). Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata by Seth van Hooland and Ruben Verborgh (review). *portal: Libraries and the Academy* 16(2), 444-445. The Johns Hopkins University Press.

<http://hdl.handle.net/10388/7601>

Downloaded from HARVEST, University of Saskatchewan's Repository for Research



PROJECT MUSE®

Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata by Seth van Hooland and Ruben Verborgh (review)

Karim Tharani

portal: Libraries and the Academy, Volume 16, Number 2, April 2016,
pp. 444-445 (Review)

Published by Johns Hopkins University Press



➔ For additional information about this article
<http://muse.jhu.edu/journals/pla/summary/v016/16.2.tharani.html>

ing the institutional role of the academic library in the twenty-first century.

Jennifer Joe

Assistant Professor and Librarian
Western Kentucky University–
Owensboro
Jennifer.joe@wku.edu

Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata, Seth van Hooland and Ruben Verborgh. Chicago: American Library Association, 2014. 224 pages. \$115.00 (paperback). (ISBN 0-8389-1251-6)

Linked Data, a method of publishing and structuring data from different sources so that the data can be interlinked to make them more useful, is perhaps the most tangible and practical branch of the Semantic Web movement. Tim Berners-Lee, the inventor of the World Wide Web, also coined the term Semantic Web, which he defined in a 2001 *Scientific American* article as “an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” So far, Linked Data and the Semantic Web have remained inaccessible to much of the library and information science (LIS) community. Seth van Hooland and Ruben Verborgh make a welcome attempt to introduce these concepts to LIS professionals with *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. In addition to a passion for their subject, both authors have academic and practical technical experience with metadata, the data that describe other data. This handbook aims at lowering barriers for the adoption of Linked Data in the LIS community. Each of the seven chapters includes hands-on

exercises. The subtitle, *How to Clean, Link and Publish Your Metadata*, highlights the organizational structure for this book’s contents as well as a process for enriching and sharing existing institutional metadata using the principles of Linked Data.

Given the technology-laden evolution of Linked Data, the current literature remains highly specialized, even though a wider mobilization of Linked Data depends on its adoption by communities outside the discipline of computer science. The authors make no secret of this critical fact in Chapter 1, defining Linked Data as a “set of best practices for the publication of structured data on the web” rather than a specific technology. (p. 3) Throughout *Linked Data for Libraries, Archives and Museums*, the authors present complementary perspectives from both computer science and LIS. Chapter 2 (“Modeling”) reviews the history and evolution of information modeling, the organization of content so that it can be delivered and reused in a variety of ways, and traces the historical parallels between the computer science and LIS communities. Metadata managers and practitioners will gain an appreciation of Linked Data as an approach for linking metadata across collections, systems, and institutions over the Web in an open and transparent manner. In Chapter 3 (“Cleaning”), which deals with metadata quality, the authors explain, “All metadata is dirty, but you can do something about it.” They argue that the quality of metadata is not only subjective but also contextual and relative to the needs of local users. (p. 71) The premise of Linked Data is to repurpose and share local metadata and make it available to others globally in unknown and unpredictable contexts.

Chapter 4 (“Reconciling”) presents an overview of controlled vocabularies—established lists of standardized terminology, such as library subject headings, for use



in the indexing and retrieval of information—and the evolution and resurgence of controlled vocabularies in the context of Linked Data. The authors stress that “the failure of full-blown ontologies has created a new opportunity for controlled vocabularies in the context of linked data.” (p. 136) LIS metadata practitioners will be elated to learn that computer professionals have finally come around to admitting the value of controlled vocabularies. Chapter 5 (“Enriching”) covers named-entity recognition (NER), the classification of text elements into predefined categories, such as the names of persons, organizations, or locations, and the best practices to enrich lengthy and unstructured data with Linked Data principles. Using Linked Data for library archives and museum metadata, NER services are an efficient way to both identify and disambiguate terms and entities within unstructured data. This chapter includes a discussion of the confusing differences between uniform resource identifiers (URIs), which identify resources, and uniform resource locators (URLs), which help find the resources by giving their network location.

Chapter 6 (“Publishing”) treats the value proposition of Linked Data as an open and cost-effective mechanism to exchange metadata over the Web. The chapter discusses the advantages and disadvantages of markup formats such as HTML (hypertext markup language), JSON (JavaScript Object Notation), and RDF (Resource Description Framework), systems for marking documents to indicate their logical structure, such as paragraphs, and how they should be laid out for electronic transmission and display. In the concluding chapter, the authors warn that the LIS community risks ceding control over much of our collective human heritage and knowledge to commercial interests if

we delay in embracing Linked Data. They appeal to the LIS community to “use their unique potential to stand up and launch a debate on these matters.” (p. 248)

Having read *Linked Data for Libraries, Archives and Museums*, I believe the LIS community would have adopted Linked Data much more rapidly had Tim Berners-Lee called it Linked Metadata. While this book may not provide a clear road map for the utopian Web powered by Linked Data, its message to the leaders, administrators, and practitioners of the library, archives, and museum communities is clear: our meticulous and standards-based institutional metadata codify rich representations and relationships about knowledge resources that must be shared more widely and publicly. Through Linked Data, we have an opportunity to exert our expertise and influence in transforming the current Web of documents into a Web of data. Our profession must start paying attention to Linked (Meta)Data!

Karim Tharani

Information Technology Librarian
University of Saskatchewan
Saskatoon, Canada
karim.tharani@usask.ca

Information Privacy Fundamentals for Librarians and Information Professionals, Cherie L. Givens. Lanham, MD: Rowman & Littlefield, 2015. 127 pages. \$45.00 (ISBN 978-1-4422-2881-8)

Information Privacy Fundamentals for Librarians and Information Professionals raises questions about the proper balance between ensuring security and safeguarding liberty. For librarians and information professionals, a concern for protecting in-