AN ALCTS MONOGRAPH

LINKED DATA FOR CULTURAL HERITAGE

ED JONES and MICHELE SEIKEL, EDITORS

An imprint of the American Library Association
CHICAGO 2016

www.alastore.ala.org

© 2016 by the American Library Association

Extensive effort has gone into ensuring the reliability of the information in this book; however, the publisher makes no warranty, express or implied, with respect to the material contained herein.

ISBN: 978-0-8389-1439-7 (paper)

Library of Congress Cataloging-in-Publication Data

Names: Jones, Ed, 1951- editor. | Seikel, Michele, editor.
Title: Linked data for cultural heritage / Ed Jones and Michele Seikel, editors.
Description: Chicago : ALA Editions, an imprint of the American Library Association, 2016. | Series: An ALCTS monograph | Includes bibliographical references and index.
Identifiers: LCCN 2016010125 | ISBN 9780838914397 (print : alk. paper)
Subjects: LCSH: Linked data. | Linked data—Case studies. | Semantic Web. | RDF (Document markup language) | BIBFRAME (Conceptual model) | Authority files (Information retrieval) | Machine-readable bibliographic data formats. | Cultural property—Electronic information resources.
Classification: LCC Z666.73.L56 L555 2016 | DDC 025.042/7—dc23
LC record available at https://lccn.loc.gov/2016010125

Cover image ©amgun/Shutterstock, Inc.

Text composition by Dianne M. Rooney in the Adobe Caslon Pro and Archer typefaces.

This paper meets the requirements of ANSI/NISO Z39.48–1992 (Permanence of Paper).

Printed in the United States of America

20 19 18 17 16 5 4 3 2 1

www.alastore.ala.org

CONTENTS

Introduction, by Ed Jones *vii*

One **Linked Open Data and the Cultural Heritage Landscape** **1**
Hilary K. Thorsen and M. Cristina Pattuelli

Two **Making MARC Agnostic** **23**
Transforming the English Short Title Catalogue for the Linked Data Universe
Carl Stahmer

Three **Authority Control for the Web** **41**
Integrating Library Practice with Linked Data
Allison Jai O'Dell

Four **Linked Data Implications for Authority Control and Vocabularies** **55**
An STM Perspective
Iker Huerga and Michael P. Lauruhn

Five **A Division of Labor** **73**
The Role of Schema.org in a Semantic Web Model of Library Resources
Carol Jean Godby

Six **BIBFRAME and Linked Data for Libraries** **105**
Sally H. McCallum

About the Contributors *125*
Index *129*

www.alastore.ala.org

INTRODUCTION

Ed Jones

The Web that we experience is designed for people: documents to be read, music, lectures, and so on to be heard; photos, videos, and so on to be viewed; files (spreadsheets, databases, etc.) to be downloaded; and links to be clicked. But of course the Web is also designed for machines, to the extent that machines read the codes and present the results that make it all work.

Over time the Web has evolved some very effective methods for identifying the different parts of a resource using an ever more elaborate hypertext markup language (HTML) and structuring the ways these parts are presented to us visually using cascading style sheets (CSS). The former will tell your browser "this string of text is a main heading" or "this is a bulleted list," while the latter will tell it "main headings should be presented in dark blue bold italic 12-point Times New Roman type" and "bulleted lists appear in dark green 8-point Arial type, with diamonds representing the bullets." And these behind-the-scene methods have become virtually universal on the Web (and much more elaborate than these examples). But introducing semantics to these documents via linked data is like moving from a black-and-white world to a world of color.

Much can be done when machines can process meaning, but that is not possible when the meaning of the documents, and so on, is not understandable by these machines. A machine may know that such-and-such is the main heading in a document, and that this heading is followed by several paragraphs, with images and videos and links to other documents, but that tells the machine nothing about the *meaning* of these things—their *semantics.* What is in the image? Who is the video about? Linked data and its various web standards enable this machine's "understanding" and allow machines to exploit it.

You can see linked data in action today when you perform a Google search for a popular topic. For example, searching for "hunchback of notre dame book"

www.alastore.ala.org

pulls up a web page with the search results on the left and a panel containing various data about the book on the right. The data has been extracted from a variety of online sources and includes star ratings of the book on websites such as Amazon (with links if you want to order a copy), a summary of the book from Wikipedia, a date of first publication, author, characters, links to various sources for reading the book online, and links to other books by Victor Hugo and other books searched for by people who also searched for the *Hunchback of Notre Dame.* If you click on the link to Victor Hugo, that triggers another search and takes you to search results and a panel of similar data on Hugo, with images, a short description, links to his place of birth, death, and burial, and so on. Clicking on his place of birth (Besançon) pulls up yet another collection, this one including a zoomable map of Besançon, its area and population, current weather, current time, and local points of interest. Clicking on the population brings up a chart created from UN data and an invitation to "explore more" by clicking on a small globe symbol. This in turn expands the chart to explore and compare populations for any place(s) in the world. Google calls this the Knowledge Graph, and while the Google Knowledge Graph makes use of some sophisticated data extraction and categorization techniques as well as linked data, it gives you some idea of the potential of linked data as a technique for establishing semantic relationships across the Web.

While this book will be concerned primarily with the use of linked data in the context of so-called cultural heritage institutions—libraries, archives, and museums (LAM)—you can get an idea of the growing scope of the available datasets by examining the list maintained by the World Wide Web Consortium (W3C) at www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets.

We are still some distance from the world of linked data that Tim Berners-Lee, James Hendler, and Ora Lassila envisaged fifteen years ago when they first proposed a Semantic Web. In their world, one's mobile device would be able to take a verbal request to make a medical appointment and then automatically negotiate with software at the doctor's office to arrive at the absolutely best of all possible appointments, taking into account (and weighting appropriately) a much larger range of relevant variables than you could ever do on your own.[1] Ideally, from a library's point of view, such a world would also enable devices to automatically select the most appropriate copy of a book, article, and so on for you to read, download, or borrow in a physical form, taking into account your preferred format for delivery, your physical location, the institutions where you have borrowing privileges (and where copies are available), commercial

www.alastore.ala.org

services to which you subscribe, and varying restrictions on access, and so on, that attach to the various available copies. Wouldn't it be lovely?

But such a world represents nirvana. For the time being, we must content ourselves with more modest ambitions for linked open data (LOD). First among these is agreement on standards, for without agreed-upon standards, linking becomes much more problematic and labor-intensive.

FIVE-STAR LINKED OPEN DATA

In 2010 Berners-Lee proposed a five-star scheme for rating linked open data in figure I.1.[2]

It can easily be seen how each additional star makes the data more useful as linked open data. The first star is awarded simply for making the data openly available (even if it's just an image scan of a chart or table, for example). The second star is awarded for giving the data a structure (e.g., spreadsheet or database) that can be read by a program. The third star is awarded for doing so in a nonproprietary format (say, comma-separated values [CSV] for a table).

So far, so easy. The resource usage reports many libraries receive from vendors as part of Project COUNTER would typically be awarded three stars.

But now comes the hard part, where machine-readable structured open data becomes linked open data.

The fourth star is awarded for using open standards from the World Wide Web Consortium (W3C) to give semantic structure to the data, specifically using the resource description framework (RDF) to describe the data and providing a SPARQL endpoint to receive and respond to queries (analogous to a database able to respond to a SQL query). RDF provides a basic structure to

★	Available on the web (whatever format) but with an open license, to be open data
★★	Available as machine-readable structured data (e.g., Excel instead of image scan of a table)
★★★	Same as ★★ but in a non-proprietary format (e.g., CSV instead of Excel)
★★★★	All the above, plus, use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★	All the above, plus, link your data to other people's data to provide context

Figure I.1 | **Five-star scheme for rating linked open data**

www.alastore.ala.org

linked data in the form of subject-predicate-object triples (e.g., <thisResource> <hasTitle> <"A la recherché de ma tante perdue">). This basic structure is often elaborated with other vocabularies such as RDFs and OWL (described in more detail in some of our contributions) and various domain-specific vocabularies. For example, the library domain has elaborated various vocabularies for Resource Description and Access (RDA), Functional Requirements of Bibliographic Records (FRBR), and the International Standard Bibliographic Description (ISBD). When your data has achieved four stars, it is in essence "open for business" (though still possibly in want of customers).

Finally, the fifth star is awarded when you take matters into your own hands and begin linking your data to other people's data. At this point, interesting things start happening, like what we experience with those Google Knowledge Graph panels. But creating these links can be very labor-intensive. It's most cost-effective when machines can be called on to do the matching—at least provisionally—as with the virtual international authority file (VIAF), where a commonality of works is used to identify identifiers for the same author in different vocabularies. (For example, it's highly probable that identifier A in vocabulary M represents the same author as identifier B in vocabulary N if both vocabularies identify them as the authors of works X, Y, and Z.)

How many datasets have achieved five stars? A rough idea can be had by periodically viewing the Linked Open Data Cloud at http://lod-cloud.net, which shows datasets clustered in broad categories and linked to one another with varying success (DBpedia and GeoNames are particularly well-connected).

VERY SIMPLE DESCRIPTION OF LINKED DATA

At this point it will be useful to describe, in very broad terms, just what is involved in getting those fourth and fifth stars. A basic understanding of linked data will also help understand the individual contributions to this volume.

In its simplest form, linked data uses the resource description framework and is expressed as three-part statements called triples, each triple consisting of a subject (what the triple is about), a predicate (describing the relationship of the subject to its object), and the object (describing an attribute of the subject or identifying the subject of another triple to which it is related). The subject and predicate are always represented by persistent HTTP URIs (i.e., uniform resource locators [URLs]), where relevant information will be found about them. Ideally, the object is also represented by an HTTP URI, but failing this

www.alastore.ala.org

it may be represented by a literal (such as the transcribed title of a resource). If the object in a triple is represented by an HTTP URI and that HTTP URI is maintained by someone other than you, then you are crossing the border into five-star territory. Congratulations! It is the most difficult border to cross, and many abandon the struggle after a few attempts.

In a given RDF triple, the subject typically represents a resource that you control. For example, subjects in a triplestore representing a library catalog would typically contain HTTP URIs that identify items in the collection (e.g., http://lccn.loc.gov/2013005033).

The predicate, characterizing the relationship between the subject and object, would typically be drawn from a published linked data vocabulary (for example, Schema.org (https://schema.org/docs/schemas.html) or the various vocabularies used for expressing data based on the cataloging standards RDA, FRBR, and ISBD as linked data: www.rdaregistry.info for the RDA vocabularies, http://iflastandards.info/ns/fr/ for the FRBR vocabularies, and http://iflastandards.info/ns/isbd/ for the ISBD vocabularies.

The object may be a literal, especially when it is transcribed data, or it may, like the subject and predicate, be an HTTP URI. This latter case is certainly the preferred one, since linking—that fifth star—becomes problematic without it. The predicate appearing in a triple typically takes a defined value as its object. For example, the predicate RDA carrier type takes values defined at www.rdaregistry.info/termList/RDACarrierType/.

lccn:2013005033 rdam:carrierType rdact:1049

This triple says that the resource identified in the subject (lccn:2013005033), a book called "RDA and Serials Cataloging," has a carrier type (rdam:P3001) of "volume" (rdact:1049). The string of characters (prefix) preceding the colon in each part of the triple is an abbreviated human-readable way of representing the full HTTP URI of the namespace from which the label following the colon is taken. In this case, the subject is taken from the Library of Congress control number namespace, the predicate is taken from the RDA manifestation properties namespace, and the object is taken from the RDA carrier type namespace (a value vocabulary).

While there are only a few legal values for the RDA carrier type, value vocabularies can be quite large. This is especially true in the realm of subject analysis, where an object may be drawn from an extensive vocabulary such as the Library of Congress Subject Headings (http://id.loc.gov/authorities/sub

www.alastore.ala.org

jects) and legal values can run into the thousands. In such cases, the value can link to the thesaurus, and to triples within that thesaurus that link to related values (for instance, broader or narrower terms), and so on. This is where the power of linked data comes into play.

Let us follow one trail (conceivable but still hypothetical at this point, since the links are not all in place):

lccn:2013005033 rdau:author lccn:n2013018475
lccn:n2013018475 skos:exactMatch viaf:105902737
viaf:105902737 skos:exactMatch isni:000000007525185X
isni:000000007525185X skos:exactMatch orcid:0000000279663733
orcid:0000000279663733 rdau:authorOf doi:10.5860/lrts.54n2.77

These five triples take one from the book (the subject of the first triple) to a journal article by the same author (the object of the fifth triple). The intervening triples take one from the identifier of the author in one identification scheme to the corresponding identifier in another (LC NAF, VIAF, ISNI, and finally ORCID) until one has moved from a scheme designed mainly for library materials (LC NAF) to one designed primarily for scholarly articles, research papers, and so on (ORCID).

SPARQL QUERY LANGUAGE

The preceding set of triples can be traversed iteratively by a human searcher using the linked data query language SPARQL or, preferably, by machine with a rather more complex general-purpose SPARQL query built into the discovery system (for example, your library's discovery system or, perhaps, a web search engine) and hidden from the user, a query that, for example, would routinely retrieve all related materials—whether books or scholarly articles—written by an author whenever one clicked on the author's name in a display of bibliographic data.

Of course, this would require that RDF and other linked data standards be implemented much more broadly than they have been at this point. But the potential of linked data for improving the precision and recall of bibliographic searches should be apparent. The crucial element is linking—that fifth star.

www.alastore.ala.org

CHALLENGES

Among the challenges to linked open data is the fact that while linked data is increasingly common both on the open web and within enterprises—linked cells in different Excel spreadsheets are a form of internal linked data—and increasing amounts of data, especially government data, are freely available on the open web, combining the two goals—linking and openness—has presented challenges. For example, most ISSN metadata—essential for identifying and linking continuing resources such as scholarly journals—lies behind a paywall, and its sale helps fund the ISSN Network.

But the presence of paywalls and similar barriers does not necessarily preclude the linking of open and proprietary data, though it may restrict the availability of certain linked data to those with the proper credentials.[3]

Beyond this, recent research on RDF triples extracted from nearly 400,000 pay-level domains (PLD) suggests that widespread adoption of both schema .org and linked open data will be dependent on a number of factors, including

- a direct business incentive such as improved listing in search engine results;
- good documentation with ready-to-adapt examples;
- implementation in widely deployed platforms such as Drupal; and
- use of a flexible standard that adapts to widespread violation.[4]

And finally the old bugbear of incorrect data will always be with us: just because metadata is properly structured doesn't mean it's correct. In this regard, I seem to be forever notifying websites—including those using linked data—that I am not the same person as the basketball player Ed "Too Tall" Jones (born the same year as me). I expect that after I'm gone, my descendants will still need to carry on this activity.

So linked open data remains very much a work in progress, and much of the progress has taken place within the domain of the cultural heritage institutions: libraries, archives, and museums. For an accessible and clear-eyed manual for implementing linked data for these institutions, I strongly recommend the excellent book by Seth van Hooland and Ruben Verborgh listed at the end of this chapter.[5] There is no question that the structure of linked data and the machine inferencing it supports shows great promise; many very large datasets

www.alastore.ala.org

have now been made available as RDF, and the SPARQL query language enables sophisticated queries across datasets. The question is, what will be the "killer app" that breaks linked open data out to the wider world and accelerates its uptake? Will it be an incremental extension of schema.org, the linked data vocabulary supported by the major search engines (such as Google and Bing)? Or will there be a thousand flowers blooming and finally achieving a critical mass, as specialized vocabularies enable the optimal exploitation of a variety of domain-specific data sets? Perhaps it will be a project described in this volume.

OUTLINE

In chapter 1, Hilary Thorsen (Stanford University) and M. Cristina Pattuelli (Pratt Institute) survey the use of linked data in significant projects across the cultural heritage domain, including Europeana and the Digital Public Library of America (DPLA), before proceeding to a more detailed description of Linked Jazz, a research project aimed at using linked data technologies "to uncover meaningful connections between documents and data related to the personal and professional lives of jazz artists" and development of related linked data tools and methods.

In chapter 2, Carl Stahmer (University of California, Davis) describes the migration of the renowned English Short Title Catalog (ESTC) from a MARC environment to one of linked data and the possibilities that migration opens up, especially in terms of involving the broader scholarly community in maintaining and enhancing ESTC metadata.

In chapter 3, Allison Jai O'Dell (University of Florida) reviews and reimagines library thesauri, metadata schemas, and information discovery, looking at how controlled vocabularies integrate library practice with linked data and exploring existing practices that are amenable to linked data, as well as areas for expansion of best practices in a linked data environment.

In chapter 4, Iker Huerga (Signifikance) and Michael Lauruhn (Elsevier Labs) examine linked data and authority control from the perspective of STM publishing, describing the role of authority control, identifiers, and vocabularies, including use of the Web Ontology Language (OWL) to add more formal semantics and the use of the SPARQL query language to create mappings between vocabularies.

In chapter 5, Carol Jean Godby (OCLC) describes OCLC's experiments with Schema.org as the foundation for a model of library resource description expressed as linked data, using 900 million catalog records accessible from WorldCat.org.

www.alastore.ala.org

Godby reports that "OCLC's experiments have shown that Schema.org can be used to define a model . . . which can be expressed in a published standard with institutional backing and potential for widespread adoption."

In chapter 6, Sally McCallum (Library of Congress) relates the development of the Bibliographic Framework Initiative (BIBFRAME) data model, the linked data successor to the data model represented by the MARC 21 formats, describing the fundamental differences between MARC and BIBFRAME. BIBFRAME is designed to be particularly suited as an exchange format for bibliographic data created using Resource Description and Access. The Library of Congress implemented a BIBFRAME pilot in the third quarter of 2015.

As an addendum to the Godby and McCallum contributions (chapters 5 and 6), readers are directed to the executive summary by Godby and Ray Denenberg (Library of Congress) of a technical analysis of the relationship between the LC and OCLC models for library linked data.[6] The technical analysis itself will be released at a later date.

Notes

1. Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American* 284, no. 5 (2001): 34–43.
2. Tim Berners-Lee, "Is Your Linked Open Data 5 Star?" in Linked Data, www.w3.org/DesignIssues/LinkedData.html 2010.
3. Carole Jean Goble et al., "Incorporating Commercial and Private Data into an Open Linked Data Platform for Drug Discovery," *Lecture Notes in Computer Science* 8219 (2013): 65–80. doi: 10.1007/978–3–642–41338–4_5.
4. Heiko Paulheim, "What the Adoption of schema.org Tells about Linked Open Data," *JCEUR Workshop Proceedings* 1362 (2015): 84–90. http://ceur-ws.org/Vol-1362/PROFILES2015_paper6.pdf.
5. Seth van Hooland and Ruben Verborgh, *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata* (Chicago: Neal-Schuman, 2014).
6. Carol Jean Godby and Ray Denenberg, *Common Ground: Exploring Compatibilities between the Linked Data Models of the Library of Congress and OCLC* (Dublin, Ohio: Library of Congress and OCLC Research, 2015), www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015-a4.pdf.

Further Reading

Allemang, Dean, and Jim Hendler. *Semantic Web for the Working Ontologist.* Amsterdam: Morgan Kaufmann, 2011.

Antoniou, Grigoris, and Frank van Harmelen. *A Semantic Web Primer.* 3rd ed. Cambridge, Mass.: MIT Press, 2012.

www.alastore.ala.org

Breitman, Karin K., Marco Antonio Casanova, and Walter Truszkowski. *Semantic Web: Concepts, Technologies and Applications.* London: Springer, 2007. doi: 10.1007/978–1–84628–710–7.

Coyle, Karen. *Linked Data Tools: Connecting on the Web.* Chicago: American Library Association, 2012. doi: 10.5860/ltr.48n4.

Godby, Carol Jean, Shenghui Wang, and Jeffrey K. Mixter. *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resources Description*. San Rafael, CA: Morgan & Claypool Publishers, 2015. doi: 10.2200/S00620ED1V01Y201412WBE012.

Heath, Tom, and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space.* San Rafael, CA: Morgan & Claypool Publishers, 2011. doi: 10.2200/S00334ED1V01Y201102WBE001.

Wood, David, Marsha Zaidman, and Luke Ruth. *Linked Data: Structured Data on the Web.* Shelter Island, NY: Manning, 2014.

www.alastore.ala.org

INDEX

A

AGROVOC thesaurus, 60–61
"Aligning Ontologies with Subsumption and Equivalence Relations in Linked Data" (Zong), 30
Allemang, Dean, 46
American Art Collaborative, 8
Amsterdam Museum, 6–7
Archivio Centrale della Stato, 6
Asadullah, Khusru, 56
author identifiers, linking, 70
authority control. *See also* controlled vocabularies
  data retrieval and information discovery, 47–49
  overview, 41–42, 58–60
  using URIs, 60–62
authority data in MARC, 116
authority records, 43

B

basic triplet structure, 26
Begley, C. G., 56
Berners-Lee, Tim, viii, ix, 31, 105
best practices
  data retrieval and information discovery, 49
  library thesauri, 44–45
  metadata element sets, 46–47
Bib Extend, 46
"BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources," 118–119
BIBFRAME data model
  developing, 112–115
  goals, 114
  issues to be resolved, 118–120
  links in, 118
  MARC data model compared, 117–118
  media in, 118–119
  multiple authority files in, 119
  overview, 46, 116–118
  pilot test for, 120–121
  with RDA mapping, 116–117
  RDF triples and serializations in, 108–110
bibliographic data
  in MARC, 116
  subject area scholars, bibliographic information provided by, 31–33
"Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services," 115
Bibliographic Ontology, 46
Bibliographic Ontology Specification, 68
BiblioGraph.net extension vocabulary, 81–83, 85
Bidney, Marcy, 43
BIO vocabulary, 46
British Library, 75–76, 115
British Library Data Model, 75–76, 77

C

case study on Linked Jazz, 8–20
CDEC (Contemporary Jewish Documentation Center Foundation), 6
challenges to linked open data, xiii–xiv
Clair, Kevin, 43
controlled vocabularies
  cross-references, 45
  currency of terminology, 45
  governance of, 69
  granularity of, 69
  LC-managed vocabularies, use of technology to get broader use of, 106
  library thesauri, 42, 43–45
  literary warrant, 45
  mapping, 65–67

www.alastore.ala.org

controlled vocabularies (cont.)
metadata element sets, 46–47
overview, 33–34, 37–38, 58–60
OWL used to represent, 64–65
quality of, 69
RDF used to represent, 62–63
selecting, 68–69
for special collections and archives, 43–44
specificity of, 69
sustainability of, 69
tagging and, 33–34
value vocabularies, 42–45
Coyle, Karen, 46
cross-references, 45
crowd sourcing in Linked Jazz project, 14–17
CSS (cascading style sheets), vii
currency of terminology, 45
Cutter, Charles Ammi, 49

D
data mashups in Linked Jazz project, 18
data retrieval and information discovery, 47–49
DBpedia, 57, 62
de Melo, Gerard, 76
Denenberg, Ray, 122
Dewey Decimal Classification, 79
Dodds, Leigh, 75
DPLA (Digital Public Library of America), 4
Dublin Core Metadata Initiative (DCMI) Metadata Terms, 67–68, 76
Deutsche Bibliothek, 115
Dunsire, Gordon, 47

E
EDM (Europeana Data Model), 3
Ellis, L. M., 56
Elsevier, 57
EMMeT (Elsevier Merged Medical Taxonomy), 57
equivalency relationships, 45, 65–66
equivalency statement (triplestore data model), 28–30
ESTC21 (English Short Title Catalogue, 21st century), 24, 25, 35–36, 38
ESTC (English Short Title Catalogue)
transformation to linked data
overview of ESTC, 23–24
planning and research, 24
social cataloging, 34–38
tagging and controlled vocabularies, 33–34
triplestore data model, use of, 24–33
Europeana, 3–4
expression (FRBR), 89, 91
extension vocabularies (Schema.org), 80–83

F
FAST (Faceted Application of Subject Terminology), 79, 80
first draft of Schema.org, 76–79
five-star linked open data, ix–x
flexibility of triplestore data model, 27, 31
Flickr, 33
FOAF (Friend of a Friend), 46, 81
folksonomy, 33–34
FRBR Core, 46
FRBR (Functional Requirements of Bibliographic Records), xi, 89, 91–92, 113
FRBR/RDA, 116–117

G
George Washington University, 115
GitHub, 20, 115, 121
Godby, Carol Jean, 73, 85, 88, 89, 92, 125
GoodRelations, 79
Google Knowledge Graph, viii
governance of controlled vocabularies, 69
granularity
of controlled vocabularies, 69
of metadata element sets, 47
Guha, R. V., 77, 78, 79

H
Harper, Corey, 60
Hausenblas, Michael, 34
Hendler, James, viii, 46
Hogan Jazz Archive Photography, 18–19
holdings data in MARC, 116
HTML (hypertext markup language), vii
HTTP (hypertext transfer protocol), 61–62
HTTP URIs, x–xi
Huerga, Iker, 55, 125

I
ID (Library of Congress Linked Data Service), 107, 111–112, 121

www.alastore.ala.org

incorrect data as challenge to linked open data, xiii
inferences, 46–47
information discovery and data retrieval, 47–49
interoperability of metadata element sets, 47
Iorns, Elizabeth, 71
ISBD (International Standard Bibliographic Description), xi
ISNI (International Standard Name Identifier), 70
ISO 2709 structure, 113
Istituto dei Beni Culturali Regione Emilia Romagna, 6
item (FRBR), 91–92

J
Jones, Ed, vii, 125–126

K
Knowledge Graph (Google), viii

L
language of the public and experts, differences in, 83–85
Lassila, Ora, viii
Lauruhn, Michael P., 55, 126
LC (Library of Congress), 106, 107, 115, 119
LC-managed vocabularies, use of technology to get broader use of, 106
LCNAF (Library of Congress Name Authority File), 58–59, 79
LCSH (Library of Congress Subject Headings), xi–xii, 58, 59, 79, 107
Library Linked Data in the Cloud: OCLC's Experiments with New Models of Library Resource Description: Synthesis Lectures in Linked Data and the Semantic Web, 74
Library Linked Data Incubator Group, 75
library metadata as linked data, 73–74
Library of Congress Linked Data Service (ID), 107, 111–112, 121
library thesauri, 42, 43–45
linked data
  challenges to, xiii–xiv
  defined, 105
  five-star rating system for, ix–x
  overview, x–xii, 105–106
  principles, 42, 43, 49
  research, 74–75
linked data for cultural heritage
  Amsterdam Museum, 6–7
  DPLA (Digital Public Library of America), 4
  Europeana, 3–4
  Linked Jazz, 8–20
  overview, 1–3
  projects using, 3–20
  Regesta.exe, 5–6
  Smithsonian Museum of American Art, 7–8
  SNAC (Social Networks and Archival Context) Project, 5
Linked Jazz, 8–20
linked open data. *See* linked data
Linked Open Data Cloud, x
links in BIBFRAME data model, 118
literals (RDF), 86, 110
literary warrant, 45
LODLAM (Linked Open Data in Libraries, Archives, and Museums), 2
LodView, 17–18

M
MADSRDF (Metadata Authority Description Schema in RDF), 111–112
manifestation (FRBR), 89–92
mapping controlled vocabularies, 65–67
MARC data model
  adaptations in, 113
  authority data in, 116
  BIBFRAME data model compared, 117–118
  bibliographic data in, 116
  overview, 25, 58, 115–116
  pilot test comparing MARC and BIBFRAME data and exploring links, 120–121
  with RDA mapping, 116
  replacing MARC format with a data interchange framework to make data more readily available on the Web, 106–107

www.alastore.ala.org

MARC records
for *The Nutcracker and the Mouse King* (Hoffmann), 86–92
for *The Nutcracker* (ballet), 92–98
Marcum, Deanna, 107
MARCXML, 113
McCallum, Sally H., 105, 126
media in BIBFRAME data model, 118–119
Medical Subject Headings (MeSH), 57
metadata
element sets, 46–47
library metadata as linked data, 73–74
schemas, 46–47
Miller, Eric, 115
Mixter, Jeffrey K., 85, 89, 95
modeling division of labor, 83–85
Multilingual WorldCat project, 88, 89
multiple authority files in BIBFRAME data model, 119
multiple namespaces in common data store (triplestore data model), 27–28

N

names
author identifiers, linking, 70
musicians, extracting and mapping proper names of, 9–11
personal names, creating unique, 58–59
place names, creating unique, 58–59
National Library of Medicine, 57, 115
NLP (natural language processing), 14
The Nutcracker and the Mouse King (Hoffmann), 86–92
The Nutcracker (ballet), 92–98

O

OCLC and Schema.org
extension vocabularies, 80–83
first draft of Schema.org, 76–79
linked data research, 74–75
modeling division of labor, 83–85
overview, 73–74
pre-Schema model of linked bibliographic data, 74–76
progress report on, use of *The Nutcracker* as informal, 86–98
published monographs, modeling beyond, 85
publishing and extending Schema.org, 79–80
OCLC (Online Computer Library Center), 115. *See also* OCLC and Schema.org
O'Dell, Allison Jai, 41, 126
On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control (Library of Congress), 73, 106, 107, 122
OWL (Web Ontology Language), 61, 64–65, 112

P

Packard Campus for Audio Visual Conservation, 119
Pattuelli, M. Cristina, 1, 127
paywalls as challenge to linked open data, xiii
personal names
creating unique, 58–59
musicians, extracting and mapping proper names of, 9–11
pilot test for BIBFRAME data model, 120–121
place names, creating unique, 58–59
Pratt Institute's School of Information and Library Science, 8
pre-Schema model of linked bibliographic data, 74–76
Princeton University, 115
Prinz, Florian, 56
profiles, 35
projects using linked open data for cultural heritage, 3–20
PubChem, 57
published monographs, modeling beyond, 85
publishing and extending Schema.org, 79–80
publishing value vocabularies as linked open data, 43–44
Putnam, Hilary, 84

Q

quality of controlled vocabulary, 69

R

Ralston Crawford Collection of Jazz Photography, 18

www.alastore.ala.org

RDA mapping
BIBFRAME data model, 116–117
MARC data model, 116
RDA (Resource Description and Access), xi, 43, 91–92, 113, 119
RDA Vocabularies, 46
RDF (Resource Description Framework)
controlled vocabularies, RDF used to represent, 62–63
graphs, 63, 67
overview, ix–xi
serializations, 108–110
triples, x–xii, 63, 65, 108–110
RDF/XML, 63, 64, 65
Regesta.exe, 5–6
RELOAD (Repository for Linked Open Archival Data), 5–6
resource relationships in RDA, 119
reuse and remixing of data in metadata element sets, 46–47
roles in RDA, 119
Ruddock, Bethan, 46
Ruth, Luke, 34
RWOs (real world objects), 112

S

Schema Bib Extend Community Group, 80, 83
Schema.org. *See also* OCLC and Schema.org Bib Extend project, 46
British Library Data Model compared, 77
extension vocabularies, 80–83
first draft of, 76–79
overview, 73–74
publishing and extending, 79–80
Schlange, Thomas, 56
Science Exchange, 71
search engines, library data appearing in, 48
searching and browsing discovery model, 47–49
Seikel, Michele, 127
semantic search, 48–49
Semantic Web, viii, 33–34
semantics and machines, vii
SKOS (Simple Knowledge Organization System), 46, 75, 112
Smith-Yoshimura, Karen, 88, 92
Smithsonian Museum of American Art, 7–8
SNAC (Social Networks and Archival Context) Project, 5
social cataloging, 34–38
SPARQL CONSTRUCT, 66–67
SPARQL query language, xii, 48, 66–67
special collections and archives, controlled vocabularies for, 43–44
specific entry, 44
specificity of controlled vocabulary, 69
Stahmer, Carl, 23, 127
statements (RDF), 108–111
STM (science, technical, and medical) publishing and research communications
author identifiers, linking, 70
authority control using URIs, 60–62
controlled vocabulary, selecting, 68–69
data sharing, 56
mapping vocabularies, 65–67
overview, 55–57
OWL (Web Ontology Language), 64
reproducible findings, 56
URIs, RDF used to create, 62–63
value vocabularies, 67–68
subject area scholars, bibliographic information provided by, 31–33
subject-predicate-object triples, x–xii, 63, 65, 108–110
sustainability of controlled vocabulary, 69

T

tag clouds, 33
tagging and controlled vocabularies, 33–34
Terse RDF Triple Language (Turtle), 110
Thorsen, Hilary K., 1, 128
Tillett, Barbara, 60
Transcript Analyzer (Linked Jazz project), 9–11
triples, RDF, x–xii, 63, 65, 108–110
triplestore data model
basic triplet structure, 26
equivalency statement, 28–30
in ESTC (English Short Title Catalogue) transformation to linked data, 24–33
flexibility of, 27, 31
multiple namespaces in common data store, 27–28
overview, 26–27

www.alastore.ala.org

triplestore data model (cont.)
subject area scholars, bibliographic information provided by, 31–33
user-enhanced triplestore, 31–33
Tulane University, 18

U
uniform heading, 44
URIs (uniform resource identifiers)
authority control using, 60–62
and Library of Congress Linked Data Service, 107
as object in RDF statement, 110
overview, 42, 60
RDF used to create, 62–63
URLs compared, 62
URLs (uniform resource locators), 60–62
user-enhanced triplestore, 31–33

V
value vocabularies, 42–45, 67–68
van Hooland, Seth, xiii, 28, 34
Vasilevsky, N. A., 56
Verborgh, Ruben, xiii, 28, 34
VIAF (Virtual International Authority File), x, 28, 79, 80

W
Wang, Shenghui, 85, 89
W3C (World Wide Web Consortium), 46, 75, 105
Wikipedia, 57
Wood, David, 34
word meaning as social construct, 84
work (FRBR), 89–92, 117
WorldCat, 79–80

Z
Zaidman, Marsha, 34
Zepheira, 115
Zong, Nansu, 30

www.alastore.ala.org