lita

New Top Technologies Every Librarian Needs to Know

A LITA Guide

Edited by
Kenneth J. Varnum

ALA Neal-Schuman
CHICAGO 2019

KENNETH J. VARNUM is the senior program manager at the University of Michigan Library. In this role, Ken is responsible for the library's discovery systems, document delivery and link resolver interfaces, and the library's evolving and emerging analytics infrastructure. He received his master's degrees from the University of Michigan's School of Information and Center for Russian and East European Studies and his bachelor of arts from Grinnell College. Over his two decades working with public-facing technology in academic, corporate, and special libraries, he has gained a deep appreciation and understanding of the need to tailor systems and interfaces to the local user base. A frequent speaker and author, Ken presents and writes about discovery systems, library analytics, and technology. In addition to numerous articles and chapters, he wrote *Drupal in Libraries* (2012), compiled the *LITA Guide The Top Technologies Every Librarian Needs to Know* and *Exploring Discovery: The Front Door to Your Library's Licensed and Digitized Content* (2016), and edited Lorcan Dempsey's *The Network Reshapes the Library* (2014). Contact Ken through his website at www.varnum.org or on Twitter at @varnum.

ISBNs
978-0-8389-1782-4 (paper)
978-0-8389-1803-6 (PDF)
978-0-8389-1805-0 (ePub)
978-0-8389-1804-3 (Kindle)

Library of Congress Cataloging-in-Publication Data

Names: Varnum, Kenneth J., 1967– editor.
Title: New top technologies every librarian needs to know / edited by Kenneth J. Varnum.
Description: Chicago : ALA Neal-Schuman, An imprint of the American Library Association, 2019. | Series: A LITA guide | Includes index.
Identifiers: LCCN 2018043367 | ISBN 9780838917824 (print : alk. paper) | ISBN 9780838918036 (pdf) | ISBN 9780838918050 (epub) | ISBN 9780838918043 (kindle)
Subjects: LCSH: Libraries—Information technology. | Libraries and the Internet. | Digital libraries.
Classification: LCC Z678.9 .N455 2019 | DDC 027.00285—dc23 LC record available at https://lccn.loc.gov/2018043367

Book design in Berkeley and Avenir. Cover image © Adobe Stock.

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

Printed in the United States of America

23 22 21 20 19 5 4 3 2 1

Contents

Preface

The original *Top Technologies Every Librarian Needs to Know* was published in 2014. Similar to what this volume's authors were asked to do, the authors of the 2014 edition were tasked with thinking about technologies that were relatively new then (either new in general or perhaps somewhat well-established but still novel to libraries) and forecast what libraries would be like in three to five years when that technology had become part of the landscape. Now that we have reached that five-year horizon (2018) from when the first edition's authors were writing (in 2013), it seems a good time to review the predictions from the first set of chapters and, with a different group of authors, take a gaze into 2018's near-term future with a new set.

A LOOK BACK

The original book focused on these broad trends: content and technology convergence, augmented reality, cloud-based library systems, discovery systems, web services, text mining, large-scale digital libraries, and open hardware. Some of these technologies are still being adopted, whereas others have become firmly entrenched in the routine of libraries. For example, text mining, large-scale digital libraries, and cloud-based library systems have become core to what many institutions do. The race to the cloud for many back-end (and user-facing) services continues apace. Discovery layers on top of whatever suite of resources a library offers its clientele are still common; the desire to provide access to the breadth of a library's content is still strong—but with increasing recognition that one-size-fits-all discovery interfaces do not necessarily meet all the needs of all the users.

Other technologies are still on the rise, with potentials constantly being explored. Augmented and virtual reality technologies described in 2014 are now becoming technologically possible as the tools to support them become cheaper. Content and technology convergence, in which users bring their own desires for how they wish to consume the content they seek and libraries can offer multiple formats to meet those needs, is still actively developing. Open hardware and the Internet of Things are still tantalizingly at hand but not quite commonplace.

A LOOK AHEAD

Like the original volume, chapters in *New Top Technologies Every Librarian Needs to Know* are structured around a rough topical outline, but authors were free to adopt a style they wanted. In each of this book's chapters, you will find a description of a technology or trend as it exists in the world in early 2018; how it is being used in libraries, archives, or museums by its early adopters; and what the future will look like if, as the authors envision, the technology is widely adopted in the next three to five years.

The book is organized into four parts, each looking at a broad swathe of library activity: "Data," "Services," "Repositories and Access," and "Interoperability." Part 1 explores linked open data, the Internet of Things, and the criticality of web archiving for the medium- and long-term preservation of the scholarly record. Part 2 explores ways that libraries are beginning to enhance their offerings to their particular patron base. The chapters cover patron privacy protection, data and data visualization, and virtual reality.

Part 3 explores how the information age will continue to reshape the ways libraries support scholarship. The chapters discuss digital humanities, digital repositories, and digital publishing. Finally, part 4 covers standards (IIIF and LTI), interoperability of content and tools, and the ever-expanding functionality of mobile devices for access.

Kenneth J. Varnum
Ann Arbor, Michigan
January 2019

PART I

Data

Linked Open Data in Libraries

Cliff Landis

Most library users will interact with linked data about your library before they ever interact with the library itself. If a user does a web search on Google for your library's name, they will see a summary of the library's vital information displayed in a "knowledge panel" to the right of their search results. This box displays information pulled from Google's proprietary Knowledge Graph, a specialized database of aggregated facts.

If your user prefers Wikipedia as a starting point, a quick search for the library will pull up a Wikipedia page that includes an "infobox" that again is populated with linked data facts. These were initially pulled to Wikidata, the linked open-data knowledgebase used to tie together facts across Wikimedia projects and languages. Under the menu of every Wikipedia page is a link to the Wikidata item for that page. On that Wikidata page, you can add more details about your library or correct erroneous information (with proper secondary source citations, of course). Unlike

Google's Knowledge Graph (linked proprietary data), Wikidata facts can be edited and reused under a Creative Commons license (linked *open* data).

But linked open data can do much more than display facts in handy little boxes. It also connects content across systems, opens up content for reuse, disambiguates similar concepts, creates relationships among data sets, and makes semantic searching possible. At its most basic, linked open data is a group of technologies and standards that enables

- writing factual statements in a machine-readable format,
- linking those factual statements to one another, and
- publishing those statements on the web with an open license for anyone to access.

For example, we know that *Harry Potter and the Deathly Hallows* is a fantasy book written by J. K. Rowling and published on July 21, 2007. That sentence can be broken up into simple three-part statements, or triples, consisting of a subject, predicate, and object:

Harry Potter and the Deathly Hallows -> has format -> book
Harry Potter and the Deathly Hallows -> has literary genre -> fantasy literature

Harry Potter and the Deathly Hallows -> was written by -> J. K. Rowling

Harry Potter and the Deathly Hallows -> has publication date -> July 21, 2007

Each fact triple describes how two concepts are connected by a relationship. Triples are stored in specialized databases variously called graphs, triplestores, or knowledge bases. But we don't need to limit these facts to bibliographic information; most information can be expressed in this three-part format:

J. K. Rowling -> has place of birth -> Yate

Yate -> has county -> Gloucestershire

Gloucestershire -> has country -> England

J. K. Rowling -> graduated college from -> University of Exeter

With these triples linked together, we can quickly answer very specific questions like "Which fantasy authors born in Gloucestershire County in England went to University of Exeter?" A semantic query like this uses the meaning of relationships between concepts to provide a results list. You can try out your own semantic searches like this one using examples on Wikidata's Query Service (https://query.wikidata.org).

The example triples above are human readable, but for them to be machine readable, each part must be represented by a web address (a Uniform Resource Identifier, or URI)—in this case from DBpedia, one of the earliest large-scale triplestores.

Subject	Predicate	Object
Harry Potter and the Deathly Hallows	Has literary genre	Fantasy literature
http://dbpedia.org/page/Harry_Potter_and_the_Deathly_Hallows	http://dbpedia.org/ontology/literaryGenre	http://dbpedia.org/page/Fantasy_literature

The subject and object are both things or concepts (often called entities) represented by their own pages, whereas the predicate is stored in an ontology, a set of definitions for different types of relationships. Collections of these URIs (both entities and ontologies) are called vocabularies. Because each URI stands on its own, you can combine entities and relationships from vocabularies across the web to connect information—that's what puts the *linked* in linked open data and makes it possible to create meaning across different triple databases. It's easy to show that an entity in one database is the same as an entity in a different database using the "sameAs" predicate from the Web Ontology Language (OWL) vocabulary.

Subject	Predicate	Object
Harry Potter and the Deathly Hallows (in DBpedia)	Is the same as	*Harry Potter and the Deathly Hallows* (in Wikidata)
http://dbpedia.org/page/Harry_Potter_and_the_Deathly_Hallows	www.w3.org/2002/07/owl#sameAs	www.wikidata.org/wiki/Q46758

These equivalence connections can be made manually by humans (such as editing Wikidata) or matched in large batches automatically using data tools like OpenRefine (http://openrefine.org). Either way, connections must be validated by humans because we can differentiate nuances of meaning that are lost on computers. Making these "sameAs" connections between databases makes it possible to connect the same concept across multiple languages and across multiple sites, dissolving information silos among different systems. This decentralized approach to metadata management also means that openly publishing locally rich metadata records can potentially help users accessing completely different systems.

Of course, to break down these metadata silos, the information has to be accessible and free to use—this is why the *open* in linked open data is so important.

By releasing metadata and datasets with open licenses, other users can transform and reuse the data, making new connections possible. Tim Berners-Lee suggests a five-star scheme for linked open data, and it begins with open licensing (http://5stardata.info/en/). This orientation toward openness and providing access aligns well with the ethos of libraries and archives.

But as librarians and archivists know, maintaining open access to information isn't a one-shot effort. Records that are enriched with linked open data must be maintained by metadata experts to ensure that the connections persist over time. This is one of the commitments that cultural heritage institutions make when they create persistent online resources of any kind; linked data URIs, just like digital object identifiers (DOIs), need to be maintained to ensure that connections among systems don't break.

One barrier to linked open data adoption by librarians is the alphabet soup of technologies and standards that make it all possible. There are free online glossaries describing the multitude of linked data and semantic web terms, but for the purpose of this chapter, only a few need to be defined. First is the Resource Description Framework (RDF), a set of specifications from the World Wide Web Consortium that explains what a triple is and how triples should be connected and queried. RDF was originally designed as a data model for metadata of web-based resources but has grown to encompass all types of metadata. It is paired with the Extensible Markup Language (XML) to create RDF/XML, the first standard format for writing linked data triples. XML is a versatile language for writing machine-readable content and is used widely in libraries and archives for publishing and exchanging metadata.

Although RDF/XML was the frontrunner for implementing linked data, other ways of expressing RDF triples have been created, including RDFa, N-Triples, N-Quads, Turtle, TriG, and JSON-LD (www.w3.org/TR/rdf11-primer/). Notably, JavaScript Object Notation for Linked Data (JSON-LD) has seen wider adoption in recent years because it is more concise and human readable than RDF/XML. Once linked data is published, it is accessed remotely using the SPARQL query language. (SPARQL is a recursive acronym that means "SPARQL Protocol and RDF Query Language" and is pronounced *sparkle*.) You can see SPARQL in action by using the Wikidata Query Service mentioned above.

Technical jargon aside, it is important to point out that linked data is fundamentally changing how people will access information in the future. To date, users have had to sift through dozens or hundreds of documents retrieved in keyword searches to find the nuggets of information that they need. However, users in the near future will have rich facets to search and additional context to discover and visualize information. For example, history students studying Renaissance dragomans (diplomatic interpreter-translators) can now browse by the dragomans' languages or professional titles thanks to the linked open data records in the Dragoman Renaissance Research Platform from the University of Toronto Scarborough (http://dragomans.digitalscholarship.utsc.utoronto.ca). This platform also provides visualizations of the geographic distributions of dragomans and a timeline of their careers, but because the data are openly accessible, users can access the database to create their own visualizations. Although coding visualizations from scratch is too difficult most users, tools are becoming available that allow users to explore linked open data and create their own visualizations with little technical knowledge. For example, RelFinder (www.visualdataweb.org/relfinder.php) shows connections between different linked open data points in DBpedia. A quick search of J. K. Rowling and Stephen King shows that both authors are influenced by author Michael J. Sullivan, while a search for gothic fiction and fantasy literature shows a cluster of books and authors that bridge those genera.

Linked data promises to unlock information from its container and break it down into interlinked facts, allowing users to discover and interact with resources in new ways. But like all new technologies, linked data will come with its own challenges and problems and will amplify older problems that librarians have always dealt with (e.g., verifiability, contextualization, interpretation, and politicization of information). However, even with these problems, linked data will reshape the landscape of information seeking and retrieval. Thankfully, librarians and archivists are hard at work staying abreast of this change and approaching these challenges head-on.

LARGE-SCALE LINKED DATA IN LIBRARYLAND

Large university libraries, national libraries, and library consortia have been the frontrunners in laying the groundwork for library linked data in two ways: by developing data models and schema for representing resource information as linked data and by publishing and aggregating resource metadata as linked data.

Library resource description is undergoing a dramatic transformation as new models, standards, and schemata are produced. As these tools are being tested in the field, they are evolving to meet the practical needs of the library community. For example, the Bibliographic Framework Initiative (BIBFRAME) from the Library of Congress aims to create a new alternative to the MARC 21 formats and, more broadly, to expand the way that resource description happens in the networked environment of linked data (www.loc.gov/bibframe/). After its initial pilot in 2015, experiments and community feedback led to further development, resulting in BIBFRAME 2.0 being released in 2016.[1] Experiments with both versions are available to explore, like BIBFRAME 1.0 implemented at the University of Illinois at Urbana–Champaign Library (http://sif.library.illinois.edu/bibframe/) and BIBFRAME 2.0 at the US Army Corps of Engineers Research and Development Center Library (http://engineerradcc.library.link/). BIBFRAME is building on the momentum created by Resource Description and Access (RDA), the international cataloging standard designed to replace AACR2 (www.rda-rsc.org).[2] RDA was created with linked data principles in mind and published in 2010, and in 2013 it was adopted by several national libraries. Support for both old and new tools will be needed for years to come, as transitions will happen slowly and carefully.

In both of the above BIBFRAME examples, bibliographic records are enriched with information from external RDF vocabularies. Some of these are smaller vocabularies, such as licenses from Creative Commons Rights Expression Language (http://creativecommons.org/ns) and metadata classes from DCMI Metadata Terms (http://purl.org/dc/terms). Larger vocabularies are now being provided by national and international organizations, such as authority data from the Virtual International Authority File (http://viaf.org) and Library of Congress Linked Data Service (www.dublincore.org/documents/dcmi-terms/). VIAF aims to provide linked data connections among authority records at national libraries, providing a global source for disambiguating person, organization, and place names. While most contributing members are national libraries, you can also find records from other linked data initiatives such as Wikidata and the International Standard Name Identifier International Agency (http://isni.org). The connections among these international metadata initiatives mean that a metadata expert can correct an error in one system (such as disambiguating the two Robert W. Woodruff Libraries in Atlanta, Georgia) and see that correction cascade to catalogs around the world as new linked open data connections are made. This type of "long-tail" effort can leverage local knowledge to enrich and repair records on a global scale.

The scope of the Library of Congress Linked Data Service is broader than VIAF, as it encompasses both general authority and subject terms as well as subject-specific terms for areas like children's literature, art, and music. Additionally, the Library of Congress provides vocabularies for preservation metadata, languages, and formats. This has been no small task, since building the infrastructure for linked open data requires publishing machine-readable metadata alongside the human-readable metadata that already exists for millions of records.

Linked open data is being leveraged to improve access and discovery at national and international scales. Europeana Collections (http://europeana.eu) is a search engine portal that aggregates more than fifty-four million metadata records from across Europe. Aggregating this number of records from a variety of sources creates challenges, as the metadata from different libraries needs to be mapped to a central data model (the Europeana Data Model, or EDM). A similar project in the United States, the Digital Public Library of America (https://dp.la/) has created its own Metadata Application Profile (MAP) modeled on Europeana's EDM. DPLA and Europeana have also collaborated to create RightsStatements.org, a linked data website with a set of twelve standardized rights statements that can be freely used to declare the copyright and reuse status of digital cultural heritage objects. By mapping these data models to an increasing number of metadata standards, combined with standardized rights statement, Europeana and DPLA are opening the door for a truly global search of cultural heritage resources. However, to take part in this opportunity for increased discovery and access, libraries have to dedicate time and energy to cleaning up resource metadata and making it ready for a global audience. As stated in DPLA's MAP (http://dp.la/info/map), "Consider how your data will work in a global context next to the data of thousands of other institutions in DPLA. Will John Brown in Australia understand that your geographic location 'Washington' is different than the State of Washington or Washington County in Wisconsin?" In practice, metadata experts at each institution will need to make sure that their local metadata records align to the larger metadata profile (i.e., the title field contains titles and not dates). Additionally, metadata experts will need to review linked open data URIs in their records (both manually created and automatically generated ones) to ensure that ambiguities are avoided. This extra work will pay off by making the library's records more globally findable, but it will also allow those records to be included in linked open data apps.

Because many users find library resources through public search engines, it is important for libraries and library consortia to create connections to the linked data efforts of large technology companies. Schema.org and the Open Graph protocol (http://ogp.me) are two examples of linked data projects initiated by tech giants that have since been released as open source schema. For consortial aggregation like Europeana Collections and DPLA, it is important to have resource metadata mapped to these larger efforts so that search engine results can point to library resources.[3] Otherwise, there is a risk of libraries creating larger collective information silos rather than breaking those silos down.

Unfortunately, large general schemata like Schema.org often do not begin with the depth necessary to fully describe library resource metadata. The Online Computer Library Center (OCLC) addressed this problem by building on the work of Schema .org to create an extension, BiblioGraph.net. This extension included library-centric vocabulary terms such as *agent*, *microform*, *newspaper*, *publicationSeries*, and *thesis*—terms that are more descriptive than "schema:Person" or "schema:CreativeWork." Schema.org has since added a bibliographic extension into its schemata, which pairs well with OCLC's larger effort to include linked data on their bibliographic records in WorldCat. Since 2012, users have been able to view linked data in WorldCat by clicking on "Linked Data" at the bottom of each bibliographic record. Meanwhile, machine-readable linked data have been published at the same URI with different formats available at different extensions (i.e., http://worldcat.org/oclc/155131850.rdf, http://worldcat.org/oclc/155131850.jsonld). This parallel publishing allows for both human- and machine-readable metadata at the same address.[4] All of these efforts are the latest in the long history of libraries working to improve access by normalizing, enhancing, and connecting library resource metadata. Although the medium of cataloging has changed from print catalogs to cards, to electronic records to linked open data, the aim has been the same—namely, to make information accessible to users. Linked open data takes this to a new level by publishing individual components of a bibliographic record in a machine-readable format, making new forms of user discovery and collection analysis possible.

EXTRA-SPECIAL COLLECTIONS

In the information ecosystem, bibliographic metadata is just the tip of the iceberg. Having copies of the same book in multiple libraries means that metadata about that book could be shared to improve access. But what happens when resources are truly unique? How do we make cultural heritage materials more discoverable? Archivists are familiar with this barrier to access, but experiments with linked open data in archives and special collections are showing the potential for new methods of discovery—ones that shine light on the relationships and contexts implicit within archival materials.

At the University of Nevada, Las Vegas (UNLV), Cory Lampert and Silvia Southwick led a team to add linked data to the UNLV University Libraries' Southern Nevada Jewish Heritage Project. This linked data initiative led to the creation of the

Jewish Heritage Project Navigator, a pilot search and browse interface (http://lod.library.unlv.edu/nav/jhp/). By using external relationship vocabularies to connect people, organizations, subjects, and materials, the UNLV team was able to create an interface that allows users to browse the collections in a nonlinear way. For example, connections between family and community members are described using predicates like "parentOf," "childOf," "neighborOf," and so on. This is in turn connected to materials, creating a rich context for users to explore the project's digital collection. This type of exploratory interface will become much more common in the next five years as libraries and archives seek to enhance their collections. To date, relevance-ranked keyword searches were the best researchers could hope for. But with the advent of linked open data in libraries and archives, researchers will have immediate access to additional contexts that can save both time and energy when exploring digital collections.

The Linked Jazz project (http://linkedjazz.org) takes this same work a step further by using linked open data to connect the relationships in the jazz community described in oral histories. These relationships come alive in the Linked Jazz Network Visualization Tool, where users can explore how various jazz artists, producers, and educators connect (http://linkedjazz.org/network/). This visualization is possible because Cristina Pattuelli and her team of researchers took transcripts from fifty-four oral histories and identified more than nine thousand entity names through automated and human analysis. This list of names was then connected to other linked open data vocabularies like DBpedia, VIAF, and MusicBrainz (http://musicbrainz.org). However, because some of the people mentioned in the oral histories were not in any of the other vocabularies, the Linked Jazz team had to publish (a.k.a. "mint") their own linked open data URIs to represent those people.[5] As more archival institutions mint their own URIs and connect them, it will become possible to see relationships within collections, across collections, and even across archival institutions.

This last kind of connection is particularly needed, as researchers often have difficulty locating materials about the same person or organization held at separate archives. The Social Networks and Archival Context (SNAC) Cooperative aims to help solve this problem through its website (http://n2t.net/ark:/99166/w6gk06z2). For example, 318 different library collections have materials about the influential and prolific activist W. E. B. Du Bois (http://snaccooperative.org/view/67874817). By harvesting metadata records at archival institutions, the SNAC website makes it possible to identify records connected to Du Bois. Like libraries, linked open data in archives and special collections is made possible because of metadata standards and schemata. Archives have been using XML to create Encoded Archival

Description (EAD) finding aids since 1998, and more recent creations like SNAC make use of the Encoded Archival Context-Corporate Bodies, Persons and Families (EAC-CPF) standard.[6] By publishing standardized authority records for entities, new linked open data connections—and therefore new methods of discovery—become possible.

THE FUTURE: MOVING FROM CODE TO TOOLS

Publishing and integrating linked open data is the future trend for cultural heritage organizations, but until recently, taking advantage of these standards required planning a discrete project and hiring programmers to do the coding. But library vendors have also been paying attention to emerging standards and schemata, so tools are now being designed with linked data standards in mind. It will take the availability of both open source and proprietary linked open data library tools before broad adoption can happen.

Archives and special collections have seen the popular digital exhibit tool Omeka launch a new, completely rewritten version of the tool in late 2017 dubbed Omeka S (http://omeka.org/s). This version includes the ability to integrate with external vocabularies, create custom vocabularies, publish linked data URIs, and use templates designed to work with DPLA's MAP data model. There is a plug-in for an International Image Interoperability Framework (IIIF) server to assist in processing and sharing images (see chapter 13 for more on this framework). Additionally, Omeka S administrators can install plug-ins to connect to other popular tools like Fedora 4 and DSpace.

Institutional repositories had a linked data overhaul in 2015, when DSpace version 5 came out with linked data capabilities, and that trend has continued with version 6 (http://dspace.org). It is important for scholars, researchers, and academic librarians to make research and its associated metadata as widely available as possible. By including RDF metadata and a SPARQL endpoint, DSpace provides researchers with the ability to increase the findability and impact of their research while also opening up repositories for large-scale data analysis.

Digital repository vendors are also adding linked data capabilities to their tools. Since Fedora is the backbone of several digital repository programs, its inclusion of linked data capabilities with the Fedora 4 release had a large impact on future development of tools like Islandora (http://islandora.ca/CLAW) and Samvera (http://samvera.org). Fedora 4 also supports the Portland Common Data Model (PCDM), a model for sharing content in digital repositories and institutional repositories (http://pcdm.org).[7]

The standards and technologies at the core of linked open data are being integrated into an increasing number of library schemata and software; as these tools become more available, the entire landscape of information seeking and retrieval will change.

SCI-FI SEARCH

Through the early research into linked open data by the projects mentioned earlier, we're already beginning to see how linked open data will impact libraries. The discovery and contextualization of information are being reshaped, and new possibilities for research are becoming broadly available.

In the near future, resource discovery will take on new dimensions for users and researchers alike. Context clues will be scattered throughout bibliographic records as libraries pull in biographical information from public databases to tell students about the backgrounds of authors. Underrepresented populations will get new levels of exposure, as users will be able to narrow down authors and subjects by new facets enabled by richer metadata provided through linked open data, like with Stacy Allison-Cassin and Dan Scott's pilot project to improve visibility for Canadian musicians.[8] We can imagine other examples where students will be able to limit search results by an author's gender, allowing women's studies and queer studies students to quickly locate research by women or nonbinary authors. With large sets of archival metadata online, historians will be able to study changes in the patterns of human communication and behavior as publications, correspondence, and photography open for analysis in aggregate.

For public domain materials, there is no need to limit linked open data to bibliographic records. Text mining and machine learning can be put to work to unlock tabular data from old books and archival manuscripts and publish them online as linked open data. Further in the future, we could see historical weather patterns published from the daily diaries of manuscript collections, helping us study the direction of climate change. Social justice advocates could query historical demographic information en masse, revealing patterns of disenfranchisement that have previously been hidden from view in scattered tables and maps. Special collections and archives will be publishing their own locally controlled vocabularies and ontologies and connecting them across institutions. This work will require metadata experts to wrangle messy data sets, connect them, and maintain those connections over time. But because of this effort, the lives of historical figures who have previously been passed over will find new light as authority records are connected across the world.

The next five years will be both exciting and a tad chaotic for libraries. Linked open data has slowly matured over the seventeen years since the concept was first introduced. But now that linked data is entering mainstream use in libraries, we will need to contend with larger problems that have been easy to overlook in the past. Disambiguation will require renewed urgency and care as information professionals connect resources together. The importance of metadata standardization will only increase as we link our data together, and gray areas of interpretation and bias in description will need refreshed scrutiny. As libraries start minting new URIs to represent resources, we must make a long-term commitment to preventing link rot (see chapter 3 for more). The library community will need to balance data models that describe library resources in depth while also connecting to outside search engines; we must accept that some tools will work and others will go the way of dodo birds and Betamax tapes. Within all these challenges, however, are opportunities to further express our core value of helping people connect with the information they need.

Experiments with linked open data in libraries are giving us glimpses of a tantalizing future where users have increased access and context for interacting with information. But even more, building the infrastructure of library linked open data will make information available for remixing and reusing in ways that we can't yet imagine. Those new uses will be discovered not only by librarians and archivists but by the users we serve, making linked open data a top technology that every librarian needs to know.

NOTES

1. Sally H. McCallum, "BIBFRAME Development," *JLIS.it* 8 (September 2017): 71–85, https://doi.org/10.4403/jlis.it-12415.
2. Karen Coyle, "Mistakes Have Been Made," filmed November 24, 2015, at Semantic Web in Libraries, Hamburg, Germany, www.youtube.com/watch?v=d0CMuxZsAIY, 37:30.
3. Richard Wallis, Antoine Isaac, Valentine Charles, and Hugo Manguinhas, "Recommendations for the Application of Schema.Org to Aggregated Cultural Heritage Metadata to Increase Relevance and Visibility to Search Engines: The

Case of Europeana," *Code4Lib Journal* 36 (April 2017), http://journal.code4lib.org/articles/12330.

4. Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter, *Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description*, Synthesis Lectures on the Semantic Web: Theory and Technology, vol. 4, no. 5 (Williston, VT: Morgan & Claypool, 2015), https://doi.org/10.2200/S00620ED1V01Y201412WBE012.
5. M. Cristina Pattuelli, Karen Hwang, and Matthew Miller, "Accidental Discovery, Intentional Inquiry: Leveraging Linked Data to Uncover the Women of Jazz," *Digital Scholarship in the Humanities* 32, no. 4 (2017): 918–24, https://doi.org/10.1093/llc/fqw047.
6. Hilary K. Thorsen and M. Cristina Pattuelli, "Linked Open Data and the Cultural Heritage Landscape," in *Linked Data for Cultural Heritage*, ed. Ed Jones and Michele Seikel (Chicago: ALA Editions, 2016), 5.
7. Erik T. Mitchell, "Library Linked Data: Early Activity and Development," *Library Technology Reports* 52, no. 1 (January 2016): 24.
8. Stacy Allison-Cassin and Dan Scott, "Wikidata: A Platform for Your Library's Linked Open Data," *Code4Lib Journal* 40 (May 2018), http://journal.code4lib.org/articles/13424.

Index

B

C

F

G

M

N

O

P

Q

R

S

T

U