

ARL White Paper on Wikidata

Opportunities and Recommendations

April 18, 2019



About ARL

The [Association of Research Libraries](#) (ARL) is a nonprofit membership organization of libraries and archives in major public and private universities, federal government agencies, and large public institutions in the US and Canada. ARL advances research, learning, and scholarly communication, fosters the open exchange of ideas and expertise, promotes equity and diversity, and pursues advocacy and public policy efforts that reflect the values of the library, scholarly, and higher education communities. ARL forges partnerships and catalyzes the collective efforts of research libraries to enable knowledge creation and to achieve enduring and barrier-free access to information.¹

ARL Task Force on Wikimedia and Linked Open Data

Stacy Allison-Cassin, York University

Alison Armstrong, Ohio State University

Phoebe Ayers, MIT; former trustee, Wikimedia Foundation

Tom Cramer, Stanford University

Mark Custer, Beinecke Rare Book & Manuscript Library, Yale University

Mairelys Lemus-Rojas, Indiana University–Purdue University

Indianapolis (IUPUI)

Sally McCallum, Library of Congress

Merrilee Proffitt, OCLC Research

Mark A. Puente, Association of Research Libraries

Judy Ruttenberg, Association of Research Libraries

Alex Stinson, Wikimedia Foundation

Contributors and Editors

The ARL Task Force on Wikimedia and Linked Open Data made a draft of this paper available for public comment from November 19 through November 30, 2018. Many people (librarians, Wikimedians, and others) provided valuable and constructive feedback including the following:

1. Asking for clarity on the framing of the document (was it meant to inform or to advocate, and to which audience or audiences?)
2. Pointing out known cultural challenges within Wikipedia and questioning the extent to which those issues are present in the Wikidata community
3. Pointing out known structural challenges within libraries with respect to how authoritative metadata is created and propagated, and the workflow implications of embracing Wikidata
4. Challenging whether there is reciprocity and mutual benefit in strengthening the Wikimedia-library relationship, in particular when time is a scarce resource
5. Questioning the use of a centralized wiki-database for structured data, with respect to both quality control and preservation in the library context
6. Reflecting on how the growth and practice of Wikidata, and the white paper's presentation of it, relates to linked open data in libraries writ large

While the Task Force on Wikimedia and Linked Open Data provides responses to some of the critiques and thorny issues, some will simply not be resolved in this paper. Some of the feedback provides the basis for a research agenda we hope will be taken up by readers.

ARL convened the task force and wrote this white paper to inform its membership about GLAM (galleries, libraries, archives, and museums) activity in Wikidata and highlight opportunities for research library involvement, particularly in community-based collections, community-owned infrastructure, and collective collections. The task force included colleagues who do not work in ARL member institutions, and the white paper covers activity well outside ARL libraries, including smaller libraries, museums, and scholarly communities. Many in the international research community, including in libraries, are focused on community-owned infrastructure² and robust metadata³ to facilitate open scholarship practices,⁴ and this paper takes a close look at Wikidata and Wikibase through that lens—as a public good worthy of

examination and support. The task force, the public comments, and the structures upon which the white paper is built are all volunteer-based. ARL is grateful to the volunteer effort that is the Wikimedia community.

External contributors generously fixed references, created items in Wikidata, added or improved examples and use cases, suggested new sections, and copy-edited for sense and clarity. This draft addresses the critiques, and benefits enormously from the contributions.

Thanks to Paul Burley, Karen Coyle, Teri Embrey, Beat Estermann, Steven Folsom, Violet Fox, Valeria De Francesca, Michelle Futornick, R. Benjamin Gorham, Stephen Hearn, Regine Heberline, Evelin Heidel, Andy Mabbett, Luca Martinelli, Monica McCormick, Daniel Mietchen, Jere Odell, Daniel Poulter, Lane Rasberry, Charles Riley, Amanda Rust, Dorothea Salo, Robert Sanderson, Dan Scott, Christina Spurgin, Andrew Su, Sara Thomas, Ruth Tillman, Simeon Warner, and several anonymous contributors.

Recommendations in this white paper are from the task force and are intended to inform and provide information for working with Wikidata if desired.

Table of Contents

About ARL	2
ARL Task Force on Wikimedia and Linked Open Data	2
Contributors and Editors	2
Executive Summary	6
Recommendations	8
Introduction	10
A Brief History and Introduction to Wikidata	17
Wikidata and Wikibase Lower Barriers to LOD, Help Scale Its Applications	25
The Culture and Policy Environment of Wikidata	26
Wikidata and Wikibase Applications	27
Authority Data and Using Wikidata as a Linking Hub	27
Notable Examples in the Research Library Community	30
Enhanced Data in Library Catalogs	30
Archival and Special Collections Discovery	32
How Can an Institution Get Started Adding Bibliographic Data to Wikidata?	34
Wikidata in the Landscape of Scholarly Communication	35
Scholarly Communication Exemplars	36
How Can an Institution Get Started Adding Scholarly Data to Wikidata?	37
Wikibase and Infrastructure for Linked Open Data	38
Wikibase Exemplars	39
How Can an Institution Get Started Implementing Wikibase?	40
Diversity, Equity, and Inclusion	40
How Can an Institution Get Started Improving the Diversity, Equity, and Inclusion of Wikidata?	43
Community Outreach and Training	43
Community Outreach Exemplars	44
How Can an Institution Get Started Doing Outreach on Wikidata?	44
Conclusion	45
Endnotes	47

Executive Summary

ARL and the Wikimedia Foundation have been in conversation about collaboration and mission alignment since 2015. Representatives from the two organizations held a summit at the International Federation of Library Associations and Institutions (IFLA) conference in 2016 in Columbus, Ohio, and subsequently contributed to several white papers⁵ exploring additional opportunities to work together. The IFLA summit surfaced the following principal areas of interest:

1. Using linked open data (LOD) to describe and connect resources, to mutually enrich Wikimedia and library discovery sources
2. Establishing learning communities for Wikimedians in libraries, cultural heritage, and research institutions
3. Librarians addressing the gaps in content and the cultural barriers inherent in Wikipedia, and conversely, using Wikimedia projects to help address cultural barriers in traditional library and archival practice

ARL charged a Task Force on Wikimedia and Linked Open Data in mid-2018 to focus on areas 1 and 3: linked open data and diversity & inclusion. In practice, this meant (1) focusing on [Wikidata](#)⁶ as a potential repository for libraries' linked open data, and (2) that a significant use case driving the formation of the task force was a mutual interest between libraries⁷ and the Wikimedia community⁸ in creating culturally competent descriptive metadata in collaboration with communities whose lives, collections, and relationships are being described. When the task force was convened, it became clear that [Wikibase](#)⁹ (the infrastructure) and Wikidata (the community and the knowledge base) should also be explored explicitly in the context of ARL's stated commitment to equitable and barrier-free scholarly communication.

Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation, the nonprofit organization that also hosts

Wikipedia and several other wiki-based knowledge projects. Wikidata serves as a central knowledge base for Wikimedia projects as well as being a freely available open database of linked open data for other projects. The task force was asked to consider issues, challenges, educational needs, and resource requirements for libraries to participate in Wikidata (and its software, Wikibase) in their uses and applications of linked open data to advance and enrich discovery of locally curated collections on the global web. Task force members collectively articulated the following goals for a white paper:

- Elevating the visibility of linked open data in research libraries and in general, and Wikidata/Wikibase in particular, as part of a global, openly licensed, linked data network for public good
- Providing an in-depth and actionable introduction to Wikidata for the library community, including a method to get started—like Wikipedia, Wikidata is both easy to use and complex to master.
- Modeling the role of librarians as contributors as well as users of a global, largely volunteer open knowledge network, by encouraging contribution of library services, workflows, tools, and software to the Wikidata community—this contribution can and should include research, development, and experimentation.
- Identifying barriers for librarians and library staff to contribute, on a regular basis, to Wikidata/Wikibase—these barriers may include, among others, lack of sufficient time, skill, staff capacity, or institutional policy environment.
- Creating a call for a community of practice around a metadata infrastructure for knowledge equity and diversity, with a focus on developing an understanding of the culture and skills of contributing to a knowledge commons like Wikidata
- Providing example use cases for how libraries and other cultural heritage institutions are already using Wikidata, which might inspire other ways to work with the project, for example, [Wikidata: WikiProject Open Access¹⁰](#)

While ARL convened this task force to inform its membership as the principal audience, the recommendations in this white paper apply to libraries well beyond the Association. The recommendations are meant for individual librarians wishing to explore the Wikidata community, and for library departments and organizations wishing to consider structural commitments to this open data source in the advancement of their work.

Contributors to this paper during the public comment period raised serious concerns about library participation (both individual and organizational) in a community with no formal governance or leadership, where systemic social power dynamics can reign and where sustainability and persistence are at risk. These concerns, acknowledged and addressed to the extent practicable, might provide a fruitful research agenda for further conversation. Tangibly, one reviewer asked whether we can advise on the **best** way to proceed with Wikidata involvement, which we look forward to exploring with more data.

In the meantime, this paper's recommendations are based upon use cases in cultural heritage documentation, open scholarly communication, archival and bibliographic discovery—in particular through collaborative description with communities represented in our collections.

Recommendations

Work on Wikidata is primarily contributed by individual volunteer contributors, who have organized their editing and data-modeling efforts in thematic projects, such as cultural heritage.¹¹ Contributing to these projects provides a way to interlink Wikidata to sources of library data. Wikidata's rise in linked open data communities also provides an opportunity for libraries to get involved in contributing to modeling and data efforts on a larger scale.¹²

For individual librarians:

- Use and experiment with Wikidata, for example:
 - Contribute local name authorities to Wikidata, particularly for underrepresented creators and organizations.
 - Add institutional holdings to existing Wikidata items using the “archives at”¹³ property.¹⁴
 - Create items for faculty in an institution.
- Explore and experiment with Wikidata editing tools such as Mix’n’match, batch uploading, and database dumps.¹⁵
- Create a “hub of hubs” for authority controls, metadata vocabularies, and other data sources, to facilitate the connection between existing external metadata sources and Wikidata.
- Get involved in the greater Wikimedia community by holding edit-a-thons and workshops, participating in discussions on email lists and in social media channels, and by joining the Wikimedia and Libraries User Group.¹⁶
- Advocate within your research communities and organizations for open, compatible licensing of data sets so that they can be incorporated into Wikidata.¹⁷

For library leadership and organizations:

- Give staff time to experiment and contribute to Wikidata, including by determining tasks that can be added to existing positions and workflows, or incorporating Wikidata participation into existing incentive and reward structures.
- Expand capacity with Wikimedians in Residence or fellowships.
- Inform and advocate with your patrons/scholars/research community to use LOD for their research projects that involve data/data sets.
- Make data sets and scholarship from existing institutional projects visible on Wikidata as part of a global network of knowledge. Large-scale cooperative projects like Social Networks and Archival Context (SNAC),¹⁸ and VIAF, the Virtual

International Authority File,¹⁹ for example, have added identifiers to Wikidata.

- Provide linked data support to researchers, academics, and other patrons wishing to expand the context of their own research and data or to develop web applications representing knowledge from their field.
- Engage scholars and communities working in underrepresented knowledge areas to help extend existing sets of knowledge in Wikidata.
- Explore and advocate for the use of Wikidata identifiers (“Q IDs”) or equivalent uniform resource identifiers (URIs) in library and archival systems, repositories, and platforms.
- Consider the use of Wikibase as a LOD store for local identifiers and authority-like data.

Wikidata is a largely volunteer community that leverages passion and enthusiasm. If through participation and experimentation libraries come to rely upon the infrastructure, they will want to assess their contributions to its growth and development within the context of their other official commitments and expenditures to Wikidata as a public good.

Introduction

In 2017, librarians at York University advanced a compelling use case for Wikidata when they began a project with ARL to work in partnership with Indigenous communities to create inclusive, culturally competent metadata about Indigenous communities and collections in Canada.²⁰ The work is part of a response to the national calls to action issued in the 2015 final report of Canada’s Truth and Reconciliation Commission. Museums and archives, as sites of public memory, were called out as crucial spaces to advance reconciliation. The York project leaders chose a linked open data approach in order to share their work—which would involve original research and consultation with Indigenous communities to identify self-determined

names, places, and relationships in their collections—as widely as possible to the extent desired by the communities. Linked data facilitates semantic interoperability with related materials. York chose Wikidata as the place to store that original structured data because it is open, non-proprietary, flexible, community-oriented, and growing in adoption and use internationally. Finally, by contributing to the metadata hub underlying Wikipedia, the York project would increase the visibility and representation of Indigenous people and collections in Wikipedia in a way that addresses what the Truth and Reconciliation Commission called “ways that have excluded or marginalized Aboriginal peoples’ cultural perspectives and historical experience.”²¹

Much of the conversation between ARL and Wikimedia at their 2016 summit centered around diversity and inclusion challenges within the Wikimedia community, which are reflected in the content of Wikipedia and often grounded in the self-fulfilling prophecy of “notability.”²² Several readers of this paper pointed out similar inequities in Wikidata, with respect to data about people, for example, but others noted that the different notability criteria, and the community of editors, distinguishes Wikidata as a place of greater opportunity than Wikipedia for inclusion of communities. Within the Wiki community writ large, librarians who are active Wikimedians are also among the community’s staunchest critics, and serve as advocates for other library professionals to remediate some of its social problems.

Wikidata is a multilingual project, supporting labeling and items in hundreds of languages, all stored in a single repository. A community of approximately 18,000 active editors from around the world work to create, enhance, and validate Wikidata’s data. Many of these editors also have experience as editors of Wikipedia or other Wikimedia projects, but many specialize in Wikidata, and the project has developed a distinct editorial community of its own, which also attracts contributors from outside the Wikimedia community.

Since the founding of Wikipedia in 2001, individual librarians have been involved in the project as volunteer contributors, but in the last several years, as the site has grown in prominence and reach, libraries have significantly increased their engagement and participation in Wikipedia. Recognizing Wikipedia as a highly-used source for their communities, librarians organize edit-a-thons,²³ and institutions are rewriting job descriptions to encourage participation in Wikimedia projects. Some have hired Wikimedians in Residence²⁴ or use graduate student interns or fellows to support representation of particular subjects from a diversity and equity perspective, and to enhance the visibility of their unique collections. Ease of use and familiarity of Wikipedia to library users has made it a successful platform around which to organize specific communities of interest and leverage their enthusiasm—in music, art and feminism, history, and more.²⁵ Similarly, some expert communities, such as medicine, have organized both on- and offline edit-a-thons to ensure the quality and accuracy of Wikipedia articles.²⁶

Libraries employ various strategies to make their collections discoverable and accessible on the web outside of traditional library discovery platforms. Such efforts have included making links to unique collections and local scholarship available in Wikipedia and other open knowledge platforms. Within the library community, a move to openly licensed metadata, open citations, and linked open data are hallmarks of a shift toward more open scholarship and infrastructure. In addition to advancing a broad mission to contribute to the public good, enhance the reputation of their institutions, and contribute to the connectedness of their scholars, such efforts have also been tied to community engagement, broadly conceived. Increasing exposure of library collections in Wikipedia has also been a critical part of advancing a diversity and equity agenda by helping to fill and address known gaps.

Linked open data (LOD) is openly sourced, commonly formatted, and interrelated. LOD breaks apart the bibliographic record into its

component parts (for example, dates, works, creators), leverages authority files and uniform resource identifiers (URIs), highlights the relationships among those components, and aims to transcend individual databases and put “information where people are looking for it—on the web.”²⁷ LOD provides the potential for greater interlinking between library collections regardless of where the collections are physically housed or virtually hosted, and enables machines to connect related items across platforms. Deploying LOD applications in libraries has been a complex task involving developing the entirety of the infrastructure needed to create it. This complexity has primarily restricted LOD activity to large, well-resourced institutions, often with external financial support.

“Linked open data is openly sourced, commonly formatted, and interrelated. It breaks apart the bibliographic record into its component parts...leverages authority files and uniform resource identifiers, highlights the relationships among those components, and aims to transcend individual databases and put ‘information where people are looking for it—on the web.’”

Wikidata offers software and an application framework in Wikibase, as well as user-friendly editing tools,²⁸ that put experimentation and implementation of linked open data within reach of more libraries. According to a 2018 survey of international linked data implementers in libraries, Wikidata has become “the #5 ranked data source consumed by linked data projects/services.”²⁹ Many in the international research community, including in libraries, are focused on community-owned infrastructure³⁰ and robust metadata³¹ to facilitate open scholarship practices,³² and this white paper takes a close look at Wikidata and Wikibase through that lens.

Through Wikidata, libraries can use their expertise in the creation of structured data and resource description in an open, reusable, and globally interoperable environment. And as a crowd-sourced and open effort, Wikidata—a community and a knowledge base—both challenges libraries’ traditional practices around authority control and at the same

time presents opportunities to promote core library values of diversity, equity, and inclusion in the collaborative creation of descriptive metadata.

Providing background on the Wikidata community and knowledge base, this white paper includes use cases exploring projects and initiatives in the broader cultural heritage community, including:

- Using Wikidata and Wikibase to expand the scope of applications of linked open data that have until now been domain or academic-subject specific
- Establishing the relationship between Wikidata and authority files
- Increasing the visibility of institutions, content, people, events, and more in Wikipedia through structured data in Wikidata
- Using Wikidata for bibliographic/archival description and discovery
- Using Wikidata within an emerging system of open scholarly communication and scholarly infrastructure
- Deploying Wikibase as infrastructure for linked open data (including open, authoritative data not open to edit)
- Wikibase and Wikidata as infrastructure for FAIR (Findable, Accessable, Interoperable, Reusable) data³³
- Librarians learning from and influencing the Wikidata community by contributing to documentation and tool development that will reduce the barriers to participation

Most active LOD projects in libraries have been concentrated in large institutions in the US and Europe—the kind of organizations with dedicated information technology and systems departments, and staff with capacity to invest in retooling basic infrastructure and data models. The Task Force encourages library leaders to take a broad-based view of how their organizations might discuss and implement its recommendations, convening staff across their organizations to see Wikidata as part of:

- Enhancing metadata and discovery
- Advocating for open bibliographic data and open metadata exchange
- Promoting greater visibility and reach of institutional research and scholarship (the inside-out library) through its open citation corpus
- Sustainability of the open approach for libraries

From the earliest discussions of the semantic web, librarians along with other academic communities have been looking for practical pathways to connect dispersed knowledge on the internet. Linked data uses data “triples” (subject-predicate-object) that connect two different concepts (subject and object) with a predefined concept relationship (a predicate). With these relationships, a growing number of items become connected into an expanding graph of interrelated concepts, which can be traversed, and in turn represented, through computational processes and human exploration. Navigation of these growing networks of concepts allows for a serendipitous series of relationships. Queries of these concepts can lead to much more human answers to questions like “Which authors were born in a specific town?” or “Which authors of papers in our collection were PhD candidates at the same university?”

But the technical approach to creating linked data is challenging. The findings of OCLC’s 2015 survey of linked data implementers in the library world indicated that the principal motivations for publishing linked data were learning and experimentation, exposing data to a larger audience on the web, and improving discovery of local content by search engines. Among the most cited barriers to publishing linked data were “steep learning curve for staff,” “little documentation or advice on how to build the systems,” “lack of tools,” “restrictive or unclear licenses,” and “ascertaining who owns the data.”³⁴

But by 2018, Wikidata rose from the 15th most used linked data source (9% of surveyed projects) in 2015 to the 5th most used linked

data source (41% of surveyed projects).³⁵ Because it was designed to support hundreds of languages, wide-scale human contribution, and the scalability required for supporting Wikipedia, Wikidata offers one of the better human-and-machine-interactive platforms for creating interdisciplinary linked data. In particular:

- Wikidata has an established global community of stakeholders among national libraries (including the National Library of the Netherlands, National Library of Sweden, British Library, National Library of Finland, National Library of Wales, German National Library, National Library Service of Italy, and others).
- Wikidata has investments from academic libraries (including the Linked Data for Libraries project, a collaboration variously involving the libraries of Columbia, Cornell, Harvard, Princeton, and Stanford, and the Library of Congress,³⁶ and from research universities throughout Europe), other GLAM (galleries, libraries, archives, and museums) institutions (The Metropolitan Museum of Art, the Pritzker Military Museum & Library, and the Smithsonian Institution), and broader academic and professional communities working in linked data.
- The experience of creating linked data within Wikidata is very human-centered, while still adopting most of the technical best practices from the broader linked data community. This means that teaching people how to create linked data requires fewer technical skills and can be achieved in less time using Wikidata compared to other platforms.
- Because of its broad scope and open community, Wikidata allows for a much more diverse, multidisciplinary approach to linked data creation than many other linked data and authority projects that are bounded by institutional membership and participation.
- Wikimedia has a long-term commitment to, and track record of, creating reliable knowledge on the internet.³⁷

Though still-emerging platforms with many open questions about their role in the linked data community, both Wikidata and Wikibase

have demonstrated substantial potential for changing the practice and the environment for practical implementation of linked open data in libraries.³⁸ By consuming and contributing to Wikidata, libraries will also have a vested interest in its scalability and persistence.

A Brief History and Introduction to Wikidata

Wikidata is a knowledge base of structured linked data. Like the Wikimedia Foundation's other projects, Wikidata is a wiki, and is editable both by individuals and by machines (using automatic programs that can make a series of specified changes, commonly known as "robots" or "bots"). Wikidata is published under the Creative Commons Public Domain Dedication 1.0, which means that the data contained in it is free to copy, modify, and distribute. It is a multilingual project, supporting labeling of items in hundreds of languages, all stored in a single site.

Wikidata was developed to connect and serve other Wikimedia projects, including Wikipedia (which has 292 active language editions),³⁹ Wikimedia Commons (Wikimedia's media repository, which as of February 2019 includes over 52 million freely licensed photos, videos, and other multimedia files),⁴⁰ Wikisource (a collection of primary source and historical documents),⁴¹ and more. For instance, a single Wikidata item can provide links to Wikipedia articles about that item in various language editions of Wikipedia, connect to photos in Wikimedia Commons about the same topic, link to a definition on Wiktionary, and, if it is an item about a historical text, link to the full text on Wikisource. These inter-wiki links are dynamically updated: if a new Wikipedia article is written in a new language, the link will be added to Wikidata and it will be available from all other Wikipedia editions.

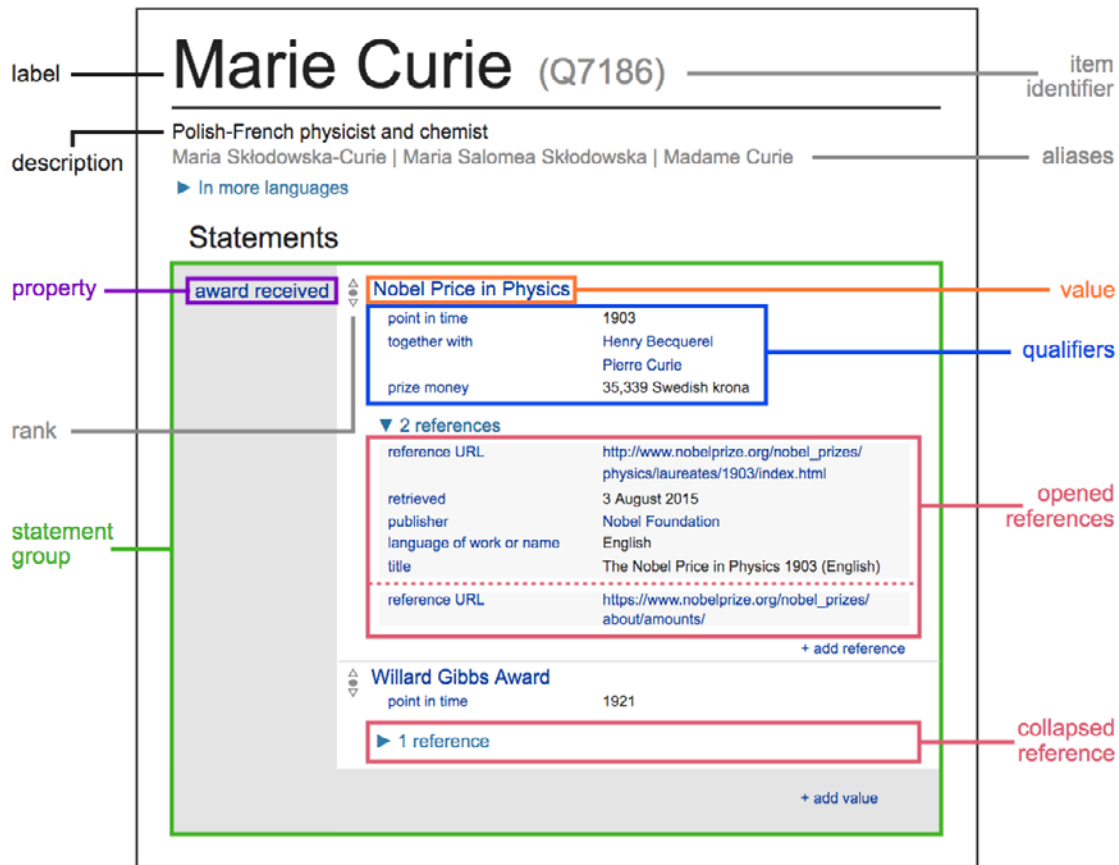


Figure 1. Graphic representing Wikidata’s data model with a statement group that includes opened and collapsed references, <http://w.wiki/32q>.

Wikidata’s structured data is designed to summarize, augment, and update the content of Wikipedia articles in a number of ways, including through dynamically populating “infoboxes,” the summary boxes of data that sit on the side of many Wikipedia articles. Wikidata provides a central place to update data that can change (such as population figures), which ensures consistency across all Wikipedia editions and provides users with the most up-to-date information on the subject. Though this is not yet being used across all infoboxes or languages, Wikimedia’s volunteer developer and editor communities are rapidly building tools to make using Wikidata easier. Other uses for this rich central repository of structured data will continue to emerge both in Wikipedia and in other projects. Further, the use of structured data through Wikidata on Wikipedia articles directly impacts the wider web. For example, Google uses Wikidata and Wikipedia in its

knowledge panels, the summarized claims to the right of a set of search results.

The first use for Wikidata was populating and supporting the connection between articles on different language editions of Wikipedia. For example, Wikidata acts as a central hub linking the article in the French-language edition of Wikipedia about “the French Revolution” with the article in the English Wikipedia and the Swahili Wikipedia, along with the 138 other language Wikipedias that have an article about this concept. If an article is created in a new language

“Further, the use of structured data through Wikidata on Wikipedia articles directly impacts the wider web. For example, Google uses Wikidata and Wikipedia in its knowledge panels, the summarized claims to the right of a set of search results.”

edition of Wikipedia (for instance, if someone starts an article about the French Revolution on the Igbo Wikipedia, which as of 2018 does not have one), the link to that article can be added to the Wikidata item, and the link to Igbo will be automatically populated across the other editions of Wikipedia.

As support for new data properties in Wikidata and coverage for Wikipedia projects expanded, an increasing number of other data sets and sources of information became the foundation for adding more and more topics and properties on Wikidata, including matching sets of external “identifiers,” which includes URIs, name authorities, and controlled vocabularies from other data sources around the web. Wikidata includes one item for every topic that has a Wikipedia article in any language (including meta-items like Wikipedia categories); but it also now includes millions of items for topics that do not have an associated Wikipedia article that meet Wikidata’s simpler notability requirements. Topics in Wikidata may never become Wikipedia articles, either because the topic is missing in Wikipedia or because Wikipedia editors have not found the topic notable enough for a stand-alone Wikipedia article. For example, Wikidata projects were developed to create items and contribute data related to paintings,

monuments, and journal articles (discussed further below). Each of these subcommunities contributes to an ever growing body of knowledge created as linked open data.

The image shows the Wikidata 'Create a new item' page. At the top left is the Wikidata logo. A search bar is located at the top right. The main heading is 'Create a new item'. Below this, there are instructions: 'Please make sure that your item complies with the notability policy.' A note states: 'Note: Don't create an item for your userpage. It will be deleted.' Another instruction says: 'Make sure the item does not already exist! (If you make a mistake, you can request your item's deletion here.)' Further instructions state: 'You should create a label and a description for all new items. The first letter of your label should be capitalized only if it is a proper noun, and your description should not be phrased as a sentence.' A disclaimer follows: 'By clicking "Create", you agree to the terms of use, and you irrevocably agree to release your contribution under the Creative Commons CC0 License.' The form itself has four input fields: 'Language' (a dropdown menu with 'en' selected), 'Label' (with placeholder text 'enter a label in English'), 'Description' (with placeholder text 'enter a description in English'), and 'Aliases, pipe-separated:' (with placeholder text 'enter some aliases in English'). A blue 'Create' button is at the bottom of the form.

Figure 2. Wikidata item creation page. Creating an item in Wikidata can be done by anyone, through either the item creation page, or tools that allow for batch item creation.

Entities are the elements of the Wikidata knowledge base, which can be either items or properties. Entities act similar to terms in a controlled vocabulary. Both item and property entities have their own section of the project,⁴² and each entity has a Wikidata page. They are assigned unique identifiers, using the letter “Q” for items and “P” for properties, followed by unique numbers. All entities can have labels (preferred name of the entity), descriptions (short description of the entity), and may have aliases (alternative names) in multiple languages.

Items contain data describing a single topic, object, or concept.⁴³ Items contain a series of statements in the form of a property-value pair, which can be further enriched by the use of qualifiers. For instance, the following statement in Wikidata describes the fact that the novel *The Able McLaughlins* won the Pulitzer Prize for Fiction. The statement

is further described using qualifiers to include when the prize was awarded and the name of the winner:

[The Able McLaughlins](#) (item) → [award received](#) (property) → [Pulitzer Prize for Fiction](#) (value) → [point in time](#) (qualifier) → 1924 (value) → [winner](#) (qualifier) → [Margaret Wilson](#) (value)

The claim made in each individual property-value statement can be supported with references. An item can also have sitelinks, which are interwiki links (links to other Wikimedia projects). Items and their statements can be created and edited by anyone.⁴⁴

Properties, on the other hand, are more controlled since they require a community review process. Properties are the equivalent to metadata elements/fields since they are used to record values. Properties in Wikidata are equivalent to predicates in general linked data terminology. A property describes a relationship between the item and another item, an identifier, or a literal value (free text string, date, etc.). As of February 2019, there are over 6,000 properties, both general and domain-specific, that are supported on Wikidata.⁴⁵

Properties that relate to existing or well-used metadata standards will be supported by the community—so getting support for new properties needed by libraries should be straightforward. However, at the time of writing, properties equivalent to common standards used in both libraries and other communities are inconsistent—often the first step for

A commonly used example of a property describing a relationship between two items is **instance of (P31)**, for example, The Able McLaughlins (Q7712123) → instance of (P31) → book (Q571). An example of a property recording that an item has an associated identifier is Library of Congress authority ID (P244), for example, Margaret Wilson (Q6760032) → **Library of Congress authority ID (P244)** → n86800547. Properties that express the relationship of an item to a literal value include date of birth (P569) and reference URL (P854).

projects working with Wikidata is identifying the alignment of standard data models with the Wikidata model.

The values associated with properties may be restricted to only certain allowed values, such as other Wikidata items, or restricted to certain formatting. For example, any value entered as a Library of Congress authority ID is expected to be a “string of 1 or 2 lowercase letters, a 2- or 4-digit year and a sequence of 6 digits.” This is typically expressed with a format as a regular expression statement on the property page.⁴⁶

Pacific Coast Architecture Database ID [edit | edit source]

Under discussion

Description	identifiers for people, buildings and firms related to architecture on the West Coast of the United States
Data type	External identifier
Domain	human (Q5), building (Q41176), business (Q4830453)
Allowed values	(person firm building)\d+
Example 1	John C. Pelton Jr. (Q55508624) → person/1364
Example 2	Geisel Library (Q962291) → building/1153
Example 3	Gensler (Q3100834) → firm/19
Source	http://pcad.lib.washington.edu
Planned use	add to person, building and firm items
Number of IDs in source	35,313
Expected completeness	eventually complete (Q21873974)
Formatter URL	http://pcad.lib.washington.edu/\$1/

Motivation [edit | edit source]

This is a fairly extensive architecture database maintained at the University of Washington Libraries. It seems useful to be able to bring in these IDs, espe buildings and firms that are less well-covered by other sources. [Certainstars \(talk\)](#) 20:48, 24 July 2018 (UTC)

Discussion [edit | edit source]

- **Support** [David \(talk\)](#) 14:34, 25 July 2018 (UTC)
- **Support** --[ChristianSW \(talk\)](#) 09:32, 26 July 2018 (UTC)
- **Comment** You can't have 3 different formatter URL's for one property. Are the numeric ID's (1364 etc from your examples) unique across the differ better formatter URL that works for any type? Otherwise you will need to prefix the numeric values with the type - I've edited your examples to do this different properties here. [ArthurPSmith \(talk\)](#) 17:21, 26 July 2018 (UTC)
 - Yes, I was initially unsure how best to handle that. The numeric IDs are not unique across types, so I think prefixing the numeric value with type, as seems tidier than creating three properties. Thank you! [Certainstars \(talk\)](#) 18:49, 26 July 2018 (UTC)
- **Support** [Jneubert \(talk\)](#) 14:34, 28 July 2018 (UTC)
- **Comment.** Creating 3 properties would be better for constraints. [Thierry Caro \(talk\)](#) 10:29, 31 July 2018 (UTC)

Figure 3. Property proposal for an architecture database maintained by the University of Washington Libraries. Most conversations about authorities are easy to advance in conversation with the Wikidata community, and property proposals are valuable opportunities for feedback and conversation about how to best restructure the property for it to work well in Wikidata's existing structure.

Wikidata's software, Wikibase, includes a number of features that make both Wikidata and linked data projects flexible. The software has a human-editable interface—encouraging individual contribution, and

allowing contribution from people with limited knowledge of linked-data standards—as well as a batch-editing API, which can be accessed with several generic batch uploading tools, including QuickStatements⁴⁷ and OpenRefine,⁴⁸ as well as custom-written bot scripts. The software supports multilinguality, allowing for simple translation of labels; and it has built-in quality-control features and permissions. Wikibase has been used in medical libraries and other GLAM projects. For example, OCLC led a pilot project to explore the application of Wikibase for linked open data in the library workflow.⁴⁹ Similarly, a number of new domain-specific linked data projects are either adopting or seriously considering adopting the Wikibase software for linked data development.⁵⁰ Because of this increasing demand for third-party installations of Wikibase, the software has been packaged with a number of utilities (including the batch uploading tool QuickStatements, and the SPARQL querying tool), for installation by server administrators, making the software accessible for creating new linked data projects.⁵¹

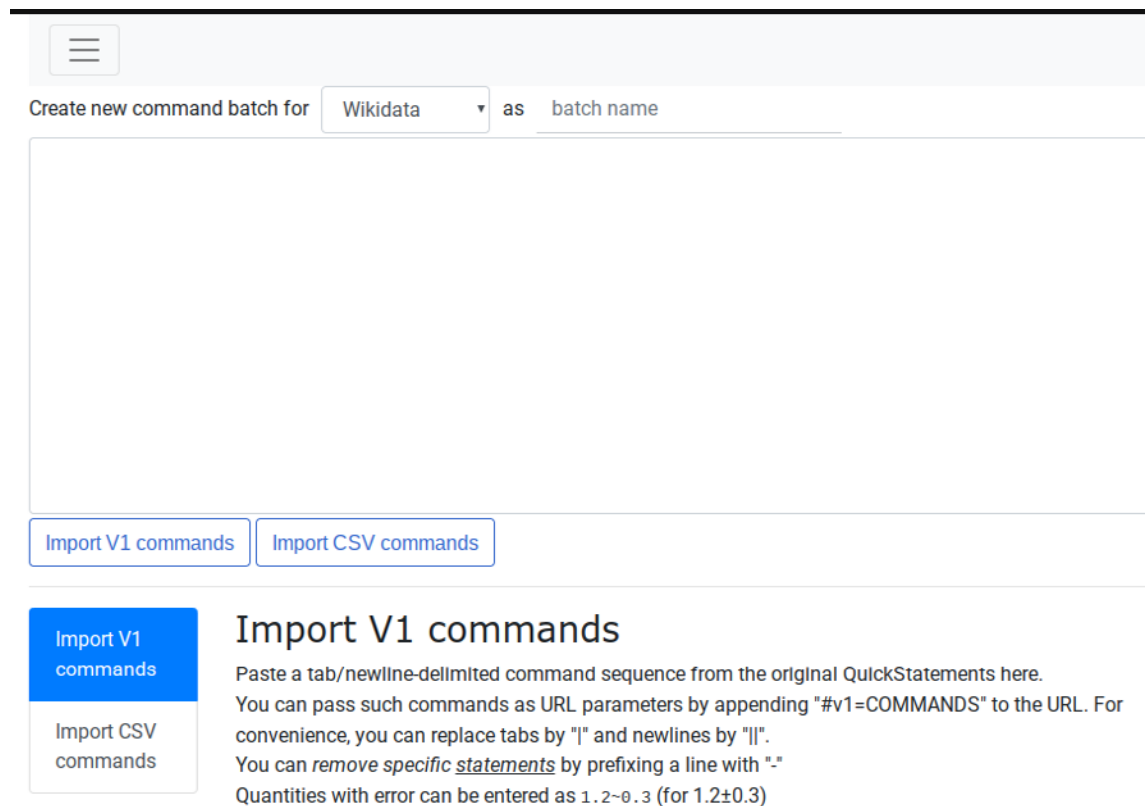


Figure 4. QuickStatements tool for creating and/or editing Wikidata items in batches.

QuickStatements allows a user to batch edit Wikidata or Wikibase by importing multiple editing commands (or statements) in a delimited text format. There are tools for generating such statements including OpenRefine, CSV files, and Zotero. The interface takes either tab/newline or comma-delimited data for batch upload into Wikidata and Wikibase. QuickStatements can be installed on a server alongside independent deployments of Wikibase.

All	Year	Work	Author - Wikidata test	Publisher	Prize Category	
<input type="checkbox"/>	81.	1931	Years of Grace	Margaret Ayer Barnes <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Margaret Ayer Barnes (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Houghton	Novel
<input type="checkbox"/>	82.	1930	Laughing Boy	Oliver Lafarge <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Houghton	Novel
<input type="checkbox"/>	83.	1929	Scarlet Sister Mary	Julia Peterkin <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Julia Peterkin (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Bobbs	Novel
<input type="checkbox"/>	84.	1928	The Bridge of San Luis Rey	Thornton Wilder <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Thornton Wilder (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Boni	Novel
<input type="checkbox"/>	85.	1927	Early Autumn	Louis Bromfield <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Louis Bromfield (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Stokes	Novel
<input type="checkbox"/>	86.	1926	Arrowsmith	Sinclair Lewis <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Sinclair Lewis (81) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Harcourt	Novel
<input type="checkbox"/>	87.	1925	So Big	Edna Ferber Choose new match	Doubleday	Novel
<input type="checkbox"/>	88.	1924	The Able McLaughlins	Margaret Wilson <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Margaret Wilson (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Margaret Wilson (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Margaret Wilson (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Margaret Wilson (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Margaret Wilson (71) <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match	Harper	Novel
<input type="checkbox"/>	89.	1923	One of Ours	Willa Cather Choose new match		
<input type="checkbox"/>	90.	1922	Alice Adams	Booth Tarkington <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Booth Tarkington <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new item Search for match		

Match this Cell Match All Identical Cells Cancel

 Margaret Wilson (Q6760032)
American novelist

Figure 5. Wikidata reconciliation service in OpenRefine.

The Wikidata reconciliation service as part of OpenRefine allows for rapid matching between strings and Wikidata items in a data set, before

uploading. The OpenRefine tool is being generalized for use on other Wikibase installations.

Wikidata and Wikibase Lower Barriers to LOD, Help Scale Its Applications

Some past attempts to do linked data at scale and across domains have run into practical limitations. DBpedia,⁵² a linked data set based on Wikipedia (and recently including some Wikidata data) and created by a small group of academics has been published inconsistently, and requires parsing information from individual Wikipedias. Freebase, a Google-supported effort to create linked data, failed due to lack of a crowdsourcing community and investment. On the other hand, closed data sets like the Getty Vocabularies and Library of Congress Name Authorities are editorially created authority systems focusing on items held in museum or library collections, which means they represent a small fraction of the world's culture and introduce layers of complexity to contribution. These limitations have largely precluded smaller institutions and independent contributors from participating, creating barriers for marginalized knowledge to enter the data set.

Wikidata and Wikibase directly address many of the challenges created by other linked data environments:

- it provides interfaces that allow for both strong human and bot-driven contribution;⁵³
- it maintains a precise history of changes;
- the data model supports the addition of citation and/or attribution for each data point;
- the Wikidata community extends the work and knowledge created by the community of contributors that make Wikipedia and Wikimedia while also encouraging the inclusion of other open licensed data sets; and

- by being radically open, following the “anyone can edit” model of contribution, it encourages the participation of a wider range of users.

Additionally, the Wikimedia community has continued to create tools, tutorials, and games to assist participants in acquiring the confidence and expertise to participate fully in the linked data creation process across skill levels.⁵⁴ Wikibase promotes these same principles, supporting rapidly changing and growing data structures. This allows for the creation of linked data projects which aren’t limited by the availability of technical skills—such as programming and data management—enabling broader participation by those with more expertise in the content itself.

The Culture and Policy Environment of Wikidata

The Wikidata community is as important an asset as the repository of data itself. The Wikidata project grows out of many of the same values that guide contributors in other Wikimedia projects. For example, Wikidata users embraced the importance of “verifiable”⁵⁵ content within the data, building on the Wikipedia concept that in an ideal world, all facts presented could be substantiated through another source. Data quality is tied heavily to the idea that an original source provides the authority behind the information, which in practice differs from the Wikipedia practice of relying on authority from secondary or other published sources of information. Additionally, Wikidata allows for items to be created from a single external source—working around the barrier for inclusion of new marginalized topics in English Wikipedia, which requires representation of a topic in multiple sources.

Similarly, the consensus-driven approach to create policies, norms, and practices on Wikipedia has been brought in large part to Wikidata. For example, property creation (and thus the ontology of Wikidata) emerges as community members propose and then endorse the

creation of a property. Similarly, as properties are not used, or data items are left unconnected to the larger graph, the community may delete them. If on the other hand, items from niche or lesser known data sets are connected to properties in the larger graph, they can find a home in Wikidata where they might not in other name or authority systems.⁵⁶ These values, practices, and norms of interaction that happen on Wikimedia projects require patience and learning how the community works. At the same time, at least for now, the Wikidata community is more open to newcomers than some of the Wikimedia sister projects.

Wikidata and Wikibase Applications

There are many different opportunities for libraries to contribute to Wikidata as a public good and to benefit their organizations. Some may consider using Wikibase as a platform for deploying a linked data store. This next section will highlight several thematic areas of interest to libraries and give examples of existing projects and initiatives to demonstrate the kinds of projects which lend themselves to Wikidata or Wikibase.

Authority Data and Using Wikidata as a Linking Hub

Contributing to open knowledge projects, such as Wikidata, aligns with the mission and values of libraries in their value-driven commitment to contribute to open culture. Authority data in the form of names (personal, corporate, or jurisdictional) are part of the backbone of functional bibliographic metadata. Authority data form the most highly structured and standardized metadata within bibliographic records, aiding with disambiguation. These data are the most readily usable and linkable as linked data on the open web.⁵⁷ Implementing authority data in the form of uniform resource identifiers (URIs) connected to collections is a powerful way to link to related collections through Wikidata, as well as opening the possibility of enriching library bibliographic systems with external data sources.

To fully take advantage of these possibilities and to more fully participate in the linked data environment, modifications must be made to standards applied to bibliographic records. These changes include allowing for the use of alternate external data sources such as Wikidata, ORCID, and ISNI to establish and link to names. These changes would enable libraries to more fully link into the cloud, and reduce resource barriers caused by requirements of the Library of Congress's Name Authority Cooperative Program (NACO). Further efforts should be made to engage in the creation of links between data stores and collections, to create fully actionable linked data.

The process of minting and verifying name authorities in the library community has often been constrained by editorial processes. For example, NACO requires a large resource commitment in both staff time and financial support, and contributions can only be made through a NACO hub, such as OCLC or SkyRiver. A more collaborative and open approach to the creation of structured data could allow libraries to concentrate efforts on unique collections, as well as choose the most appropriate mechanism for the minting of URIs for names. Furthermore, libraries could focus on creating and hooking into a network of names, rather than on a one-to-one relationship with the Library of Congress.

Wikidata works in concert with open platforms (whether on Wikidata, in appropriate relevant Wikibases, or on other dynamic LOD platforms with liberal contribution policies) to develop linked data, and connect to existing sources of that data. It also enables greater openness and opportunities for the creation, sharing, and reuse of bibliographic and authority-like data. Bibliographic data is part of a network of data sources and libraries should continue to advocate for reducing resource barriers to data creation processes and for making data as reusable and open as possible. The authors see particular benefits in areas related to unique and special collections and smaller institutions that might lie outside traditional library workflows.

Libraries' participation in the creation and enhancement of Wikidata items related to their unique collections is an area of great potential. Providing a presence in Wikidata for creators of archival collections (corporate body, person, or family responsible for the creation/ compilation of the works), and people and organizations represented in those collections, may help facilitate discoverability of the materials as well as their connection to materials in other repositories.

At the same time, the data in Wikidata is connected to external data sources, many of which were created through national libraries. Making these connections not only enriches the data, but it also aids in disambiguating concepts. Already large sets of identifiers from such data sources as the Library of Congress Name Authority File (LCNAF), the German National Library GND, and the National Library of France SUDOC have been extensively matched. The interconnection between Wikidata and trusted external data sources and schemas⁵⁸ bolsters Wikidata as a trustworthy conveyor of data.

“...Wikidata has the potential to be an important part of the linked data ecosystem for authorities.”

However, an area that will need further development is finding ways to increase the ease and speed of creating items in Wikidata. Even with the use of existing tools it is cumbersome to create items with a full set of associated properties. Community development of application profiles containing practical, base element sets for common library-related materials would aid in the utility of Wikidata and its place within library workflows. Furthermore, it would help to ensure a consistent application of data across users in libraries and other cultural heritage institutions.

A variety of projects described below demonstrate that Wikidata has the potential to be an important part of the linked data ecosystem for authorities. Smaller institutions with limited resources could benefit from the free, community-driven, easy-to-use Wikidata interface to

provide a presence for underrepresented subjects in the knowledge base.⁵⁹

Notable Examples in the Research Library Community

Enhanced Data in Library Catalogs

The University of Wisconsin–Madison Libraries developed a linked data project called BibCard to build knowledge cards for library data. BibCard brings identifiers related to authors from external sources (such as VIAF, Library of Congress Name Authority File, Wikidata, DBpedia, and the Getty Vocabularies) into the linked data instance of their library catalog.⁶⁰ An example of this integration can be seen in the entry for Indiana University–Purdue University Indianapolis (IUPUI) professor Una Osili’s work *Does Female Schooling Reduce Fertility? Evidence from Nigeria*. (See Figure 6.) The Wikidata entry for Osili was created as part of the IUPUI initiative to bring faculty members to Wikidata, mentioned later in this white paper. This shows how a library’s contribution to Wikidata can benefit other institutions in the enhancement of their bibliographic catalog records.

Data from other sources

Una Osili [Collapse](#)

Brief Biography

Abstract from Wikidata: professor of economics and philanthropic studies

<http://www.wikidata.org/entity/Q32979516>

Worked in: Indianapolis

2 Alma Maters

Source: [Wikidata](#)

Northwestern University

Harvard University

Identities on The Web

<http://id.loc.gov/authorities/names/n2004144107>

<http://viaf.org/viaf/43025147>

<http://www.wikidata.org/entity/Q32979516>

Figure 6. Una Osili’s information in the University of Wisconsin–Madison Libraries catalog, where Wikidata is used as one of the external sources for information to expand the author’s biographical data.

A prototype developed at Laurentian University demonstrates the integration of Wikidata into the institution’s Evergreen-based catalog.⁶¹ (See Figure 7.) In this project, identifier data is pulled from Wikidata to display contextual links for musical artists.

Laurentian University
Université Laurentienne

Library Home Electronic Journals (A-Z) Databases Archives Course Reserves Contact Us

Search the Catalog [Advanced Search](#) [Browse the Catalog](#)

Search: Type: Keyword Format: All Formats Library: Laurentian Univers

The spirit of radio [sound recording] : greatest hits, 1974-1987 / Rush
[Rush](#). (Performer). [Lee, Geddy](#). (Instrumentalist). [Lifeson, Alex](#). (Instrumentalist). [Peart, Neil](#). (Instrumentalist).

Rush
 Canadian rock band
 Web site: <http://www.rush.com/>
[Discography \(Musicbrainz\)](#)
[Tour dates \(Songkick\)](#)
 Twitter: [rushtheband](#)
 Facebook: [rushtheband](#)
 Edit on Wikidata: [Q203871](#)

Figure 7. Snippet of a catalog record in the Laurentian University Library ILS (Integrated Library System), Evergreen, where data pulled from Wikidata is used to augment the record.

A similar feature has been rolled out by library technology provider Zepheira, which is creating “About Author” tools for entities represented in the collections of public libraries across the United States.⁶² Once linked data entities are matched against Wikidata, a much larger community of reusers can take advantage of that data for discovery applications.

Archival and Special Collections Discovery

SNAC (Social Networks and Archival Context) provides the archival community with a central hub where data about archival creators (corporations, people, and families) are maintained. In SNAC, a record is defined as a “constellation” and each constellation is assigned a unique identifier. During the creation of the SNAC database, many of these constellations were linked to library authority records as well as Wikipedia. SNAC has also benefited from those links by pulling associated images from Wikimedia Commons into its interface when a SNAC constellation is linked to a Wikipedia entry. More recently, the SNAC constellations were matched with existing Wikidata items,

providing an additional data source offering more information on a given concept. This work was done in two parts: a wiki editor affiliated with a library institution proposed a new property in Wikidata to accommodate SNAC identifiers; and another wiki editor created a procedure to match SNAC constellations with existing Wikidata items. The proposed property, SNAC Ark ID,⁶³ underwent a community review process and was later added to the knowledge base. With the new property in place, the matching of constellations to existing Wikidata items resulted in the addition of about 128,000 SNAC IDs to the knowledge base. Since that time, the property has continued to be used by Wikidata editors, with additional matches and corrections being undertaken.

Europeana, the EU digital platform for cultural heritage, has called on libraries and other cultural heritage institutions to add their identifiers and vocabularies to Wikidata as a way for Europeana to pull them into their system. Wikidata was identified as a priority area for Europeana, and having more libraries using Wikidata as a linking hub strengthens the overall structure of semantic information about cultural heritage collections.⁶⁴

The Smithsonian Institution is running an Open Data Pilot that focuses on contributing open data about the institutional collections to Wikidata. Because the Smithsonian is a multidisciplinary organization, the collection and management of the identities within their content management platforms has considerable challenges—including, for example, reconciling identities and data models across different disciplines. As part of this pilot for regularizing and building a model for impact around linked open data, the Smithsonian is developing two data sets: one in the sciences (Natural History Specimens) and one in the humanities (Artworks).⁶⁵ Some of this experimentation will be paired with the Smithsonian's American Women's History Initiative as a series of digital innovation initiatives that work across institutional boundaries within the Smithsonian.

In a corollary project, Yale University Library, through the Wikidata for Digital Preservation project,⁶⁶ is using Wikidata to store technical information about software preservation. Wikidata provides a way to model and create collaborative, responsive data that allows the software preservation community to reduce redundancies and maximize the sharing of work. Because Wikidata is a stable, collaborative project with a great deal of flexibility, it reduces barriers to collaboration and documentation that might make this work otherwise challenging.

How Can an Institution Get Started Adding Bibliographic Data to Wikidata?

- Host workshops on Wikidata.⁶⁷
- Make micro-contributions by using existing external tools (for example through the Wikidata Distributed Game⁶⁸ or through the Mix'n'match tool).⁶⁹
- Add archival holdings to existing Wikidata items using the “archives at” property.⁷⁰
- Create items for creators of archival collections and link them to your institution.
- Add missing descriptions to existing items in your language of preference.⁷¹
- Particularly for institutions where ORCID⁷² are widely used, and/or VIVO is deployed, create items for individual faculty members at your institution (including identifiers to external data sources)⁷² and their publications.
- Integrate existing local name authorities or vocabularies that are important to your institution through the Mix'n'match tool.
- Batch-upload data. There is a well-documented process for identifying and uploading batches of data to Wikidata.⁷³ Data sets that might be a good fit for Wikidata include: data about people or other institutional entities that are relevant to your collections; extended information about the institution itself (especially

faculty, buildings in existing data sets); or data about geographical entities from the local region.

Wikidata in the Landscape of Scholarly Communication

Academic institutions and their libraries are interested in collecting and sharing bibliographic information related to the scholarship produced by their faculty and researchers. This not only facilitates visibility and recognition of their work, but also strengthens the institution's reputation. There are a number of resources and tools available to faculty, researchers, and their institutions for sharing their scholarly activities. One way this is accomplished is by displaying faculty profiles in institutional websites, which can contain both biographical and bibliographical data, or using academic network sites⁷⁴ to expose the data. For instance, managing identities using ORCID;⁷⁵ tracking citations in Google Scholar Citations; networking with other faculty and researchers with similar interests in ResearchGate, Academia.edu, and LinkedIn; or using VIVO,⁷⁶ a cross-institutional application used to track the productivity of faculty and researchers. Limitations to these approaches include the need for individual faculty or researchers to keep the data up-to-date (both biographical and bibliographical), and the fact that the data shared on these sites are not always structured or licensed in a way that would allow for them to be easily reused.

Academic institutions have demonstrated the appetite and readiness for both research library communities and the larger scholarly community to access, share, and discover scholarly materials without restrictions. The promise of linked data in libraries is in part associated with the importance of making research more visible for communities to use and reuse for the creation of new scholarship. Open bibliographic metadata initiatives in the Wikimedia arena include the Initiative for Open Citations⁷⁷ and WikiCite.⁷⁸

As the number of tools and platforms in open scholarship proliferates (see “101 Innovations in Scholarly Communication”),⁷⁹ there has been increased attention to ownership, portability and interoperability, and lock-in. Groups supporting “community-owned” or “academy-owned” scholarly infrastructure are investing in open source publishing platforms, preprints, and scholarly annotation tools. Wikidata is community-owned. It stores structured data that has open, reusable licenses, which makes the Wikidata knowledge base part of the open “connective tissue” that can help transcend particular tools and power discovery and integration of scholarship.

Scholarly Communication Exemplars

Wikidata is being used as a means of documenting and surfacing researchers, publications and research data in a number of ways. It provides an opportunity for sharing faculty scholarship on an open and accessible platform. The following are examples of current initiatives, tools, and use cases related to Wikidata:

- **WikiCite:** “a Wikimedia initiative to develop a database of open citations and linked bibliographic data to serve free knowledge.”⁸⁰ The WikiCite initiative has brought together a community of Wikidata contributors, open access advocates, and library metadata professionals, to model source materials (including journal articles and books) that can be used as structured data citations for Wikimedia projects and other platforms. The goal is to create the “Sum of All Citations.” At the time of this paper’s writing, the WikiCite data set now includes nearly 7 billion triples (subject-predicate-object) and over 150 million citation relationships among the works collected.⁸¹ To date, there have been three WikiCite conferences, bringing together Wikimedia contributors and community members, linked data professionals, and librarians to discuss and develop the future of bibliographic data on Wikidata. There are planned future events as well as a discussion mailing list that any interested people can join.⁸²

- **Scholia:** a web service that makes live SPARQL queries (SPARQL Protocol and RDF Query Language) to Wikidata, which allows for live browsing of the relationships between works, authors, institutions, and other metadata about the works and their creators.⁸³ In addition to rendering scholarly profiles, Scholia can be used as a bibliographic reference management tool. The bibliographic visualizations generated by the tool allow users to explore publications and their connections with other works. Scholia has served an important role in providing a better understanding of what is possible to achieve using Wikidata data.
- **Faculty profiles at IUPUI:** The University Library at Indiana University–Purdue University Indianapolis ran a pilot project in 2017⁸⁴ to test the viability of Wikidata as a repository for data related to campus faculty members and the scholarship they produce. The core faculty from the IU Lilly Family School of Philanthropy was selected as a use case. Items were created for the selected faculty, their co-authors (regardless of their institutional affiliation), and some of their publications. Connections between works and creators were established. Cited works (works listed in the references) were also added to Wikidata. These contributions facilitated the use of Scholia to generate the scholarly profiles. IUPUI, in continuing its support and involvement with open knowledge projects such as Wikidata, has continued creating entries for the campus faculty across multiple disciplines, with a focus on women faculty.

How Can an Institution Get Started Adding Scholarly Data to Wikidata?

- Institutions can systematically add data for faculty members and their publications to Wikidata using the Source MetaData⁸⁵ tool. This tool's performance is dependent, in part, on works having digital object identifiers (DOIs), and those DOIs appearing in publicly viewable ORCID records.

- Data can also be added ad hoc after being exported from Zotero⁸⁶ into the QuickStatements tool.⁸⁷
- Institutions can consider running Wikidata edit-a-thons⁸⁸ (for staff and/or as public events), especially during events like Open Access Week, Open Data Day, Mozilla Global Sprint, or Ada Lovelace Day, to add faculty and researcher profiles and publications to Wikidata.

Wikibase and Infrastructure for Linked Open Data

The software that drives Wikidata, Wikibase,⁸⁹ can offer value to the broader library community independent of Wikidata. The relatively lightweight technology stack and easy-to-use human-readable interface make Wikibase a viable piece of infrastructure for developing a linked data store. Wikibase is particularly useful in cases where data and data models are highly specialized or there are considerations that require greater control over the data. Wikibase has a growing community of users in the GLAM and research sectors.

There are a number of benefits to using Wikibase, as Matt Miller of the Library of Congress describes:

- Statement level provenance (what Wikibase calls references)
- Revision tracking and history
- A nice user interface for manual editing and curation
- An API to do bulk data work
- SPARQL endpoint⁹⁰

Wikibase provides something that a number of other platforms in the linked data environment have not: a dynamic, machine- and human-contributable and readable environment that supports diversity of language and data structure from the beginning. The growing number of Wikibase implementations (examples below) suggests opportunities for scholarly and GLAM reuse of the software as a generic data

store. Investment in community infrastructure means less lock-in to proprietary platforms and systems.

A Wikibase implementation may be a sandbox or drafting environment that allows for ingestion and reconciliation of data sets before they are merged into Wikidata. Or a Wikibase implementation may be the end point for managing a particular set of data or project.

Because the Wikibase software has only been readily available for independent deployment, many of the current examples of its deployment are experimental. There is a growing community of Wikibase adopters interested in running such experiments, identifying technical barriers to adoption, and facilitating increased feedback to the development of the software beyond its Wikidata application.

Wikibase Exemplars

- **OCLC Linked Data Wikibase Prototype:**⁹¹ Since 2017, OCLC has been collaborating with university research libraries to pilot the use of Wikibase as a sandbox environment for creating bibliographic and other metadata, and to explore its use in metadata reconciliation. The project has recently concluded, and its results will be shared in forthcoming conference presentations and a report.
- **Enslaved.org:** A Mellon-funded project from the Matrix Digital Humanities Center at Michigan State University, Enslaved.org uses Wikibase as a central platform for integrating discrete scholarly databases of slavery data. The goal of the project is to create a central platform to facilitate better research and storytelling about individual slave narratives as well as create tools for scholars to analyze and reflect on those sources.
- **Rhizome**, an arts organization based in New York City, has used Wikibase to document its collection of born-digital art and to practice digital preservation. Wikibase was chosen because it offers a flexible and customizable structure for both

modelling data and adding properties and qualifiers. Further, Rhizome needed highly specialized fields that may not have been appropriate for Wikidata.⁹²

- **Structured data on Wikimedia Commons:** The Wikimedia Foundation, with the support of the Alfred P. Sloan Foundation, has invested in the integration of Wikibase into the backend metadata storage for its Wikimedia Commons platform. Commons, which contains 50 million files, including millions that come directly from GLAM institutions, has long stored metadata as free text.

How Can an Institution Get Started Implementing Wikibase?

- Explore how and why institutions and projects have decided to go with separate Wikibase installations.
- Provide server space and staff time from institutional developers in order to launch and support a deployment of a local Wikibase.
- Pilot linked data projects within the Wikibase environment before building or deploying custom platforms.

Diversity, Equity, and Inclusion

Most collections in the library and broader GLAM community build on a number of assumptions about what constitutes valuable content. Though there is a growing attempt to increase representation and collection of underrepresented communities and knowledges in these collections, many of the descriptive practices that underpin collection catalogs and authorities don't provide the full flexibility to break out of colonial and patriarchal assumptions.

Wikidata offers an opportunity to document and increase the visibility of collections and research materials about underrepresented populations. In the last half decade, the Wikimedia movement has invested energy in populating biographies of notable individuals, and in turn, increasing the diversity of people represented in that

content through focused drives, such as those implemented by the Art+Feminism (focused on women in the arts), Women in Red (focused on missing women in English Wikipedia), WikiMujeres (a group focused on representation of women in Spanish Wikipedia), and AfroCrowd (outreach to people of African descent) initiatives. These community-led Wikimedia initiatives focus on diversity of representation of people included in the projects and connected to other data sources. By populating Wikidata with content about these individuals, and then supporting collaborative initiatives coming out of the Wikimedia community, libraries have an opportunity to engage networks of activists in fully documenting and describing these communities.

There are still considerable gaps in Wikidata content, but the Wikimedia movement's commitment to knowledge equity as part of the 2030 Wikimedia movement direction, and recent trends in organizing within the Wikimedia movement, suggest that a growing global community of practice is forming around using Wikidata to document marginalized knowledge.⁹³ Library and other institutional metadata is an important foundation for building out that growing data set to address these issues.

Several institutions are already developing projects that leverage Wikidata to better connect resources from external sites with content within the Wikimedia ecosystem to create an emergent and dynamically growing body of knowledge. For example, the Smithsonian's American Women's History Initiative⁹⁴ will be building on existing crowdsourcing and research efforts to identify women in their collections through open data practices that rely on Wikidata. These efforts will help the Smithsonian better understand what is already documented in their metadata by providing the crowdsourced context and will make it easier to surface the identified resources. Practically, this community-centered approach to enriching data should open up a number of opportunities for dynamically enriched explorations of women in the Smithsonian's collection.

Another example of an emergent applications of Wikidata is a recent project from a team at Yale University called Science Stories.⁹⁵ The new platform pulls data from Yale University Library, Wikidata, and other sources to create a dynamic interpretative environment for understanding Yale alumni who are women in the sciences. In this setting, Wikidata acts as both a source of data and a bridge between different authoritative sources for enabling the auto-generation of these women’s story pages.⁹⁶ These examples are emergent, but promising models of how Wikidata could offer a pathway for institutions and communities to surface underrepresented groups and the knowledge associated with them from collections.

Due to the multilingual nature of Wikidata, its data is being used to generate content for underrepresented language editions of Wikipedia. This work is being accomplished by the use of the “article placeholder,”⁹⁷ which dynamically generates information for the subject being searched in a particular Wikipedia edition based on statements available in Wikidata. For instance, the article placeholder has been adopted by the Welsh community of editors in an effort to increase access to knowledge in Welsh Wikipedia.⁹⁸ Other applications include Wikidata-powered infoboxes⁹⁹ and Listeria.¹⁰⁰

Beyond leveraging Wikidata as a hub among projects trying to better understand existing metadata, the platform allows for more deliberate efforts to create and accurately represent these communities. This has been best demonstrated in the work being led by Stacy Allison-Cassin at York.¹⁰¹ The open-ended data model for working with content allowed her team to work with the Indigenous communities represented in materials and records of the York University archives to create authorities that accurately represent the individuals and cultural practices in those ecosystems. Similarly, the Black Lunch Table project¹⁰² connects underrepresented artists in Wikipedia, and through Wikidata the project has run campaigns to collect images and research about these artists. The flexibility of Wikidata and Wikipedia allows

for the project to identify and expand description of these artists while respecting their self-determined identities.¹⁰³

How Can an Institution Get Started Improving the Diversity, Equity, and Inclusion of Wikidata?

To address gaps in Wikidata coverage, institutions can:

- Add descriptions to existing collections with biases because of how and when they were described, to include more diverse identities from sets integrated in Wikidata or other related sets.
- Create authorities, using Wikidata, for missing identities within existing collections, to increase the reliability of Wikimedia projects.
- Use Wikidata as a mechanism for connecting data sets focused on underrepresented groups or marginalized collections.
- Engage scholars working in underrepresented knowledge areas, such as non-Western, colonized, or other marginalized knowledge, to help extend existing sets of knowledge in Wikidata.

Community Outreach and Training

The Wikimedia community has a long history of collaborating with and training librarians to participate in Wikimedia projects.¹⁰⁴ Much of the collaboration has centered around holding community-facing Wikipedia events in libraries supporting outreach, digital literacy, or other skill-based work—and the library community is beginning to see Wikidata and Wikibase skills being shared in workshop and edit-a-thon settings similar to those used with Wikipedia. For example, the 2017 Canadian Music Edit-a-thon series included Wikidata integration.

There are also themed communities in Wikidata that can catalyze public contribution, and there is value in bringing actors beyond the “library metadata” workers into thinking and caring about metadata.

Community Outreach Exemplars

Sum of All Paintings is a volunteer-organized effort to represent “every notable painting” in the world on Wikidata.¹⁰⁵ This project has become a focal point for data modeling, not only for paintings and other works of art, but related topics like artistic methods, creators, and institutions holding the works. Sum of All Paintings and the larger cultural heritage community on Wikidata have been a driving force behind ingestion of data from GLAM institutions around the world.¹⁰⁶ This community has allowed for the creation of several prototype tools for browsing heritage collections, such as Crotos.¹⁰⁷

The Wikimedia community has been running the **Wiki Loves Monuments** campaign since 2011 to document listed monuments and heritage sites.¹⁰⁸ Since its inception, Wiki Loves Monuments has generated millions of freely licensed, high-quality images of heritage sites, and created a Guinness World Record¹⁰⁹ for the largest photography competition in the world. In the process of running the contest, the community collected and reconciled heritage registries from hundreds of jurisdictions, creating the biggest monuments database in the world. In the last few years, Wikimedia Sweden, with the support of a major Swedish grant and in collaboration with UNESCO (United Nations Educational, Scientific and Cultural Organization), converted much of that database into Wikidata, and imported additional heritage registries into Wikidata as well.¹¹⁰

How Can an Institution Get Started Doing Outreach on Wikidata?

- Devote staff time to Wikidata by writing participation into job descriptions, and giving release time and professional development time to participate in Wikimedia events. Examples of positions explicitly incorporating Wikidata include a coordinator of Wikipedia initiatives at Brigham Young University Library, in the library’s department of Marketing, Design, and Communications, and the digital initiatives metadata librarian

at IUPUI, who has participation, training, and advocacy for Wikimedia and open knowledge as secondary responsibilities.

- Look into a Wikipedian/Wikimedian in Residence. More than 150 libraries, museums, archives, and other institutions have hosted Wikimedians in Residence since 2010. “The Wikipedian in Residence is not simply an in-house editor: the role is fundamentally about enabling the host organisation and its members to continue a productive relationship with the encyclopedia and its community after the Residency is finished.”¹¹¹
- Convene different sectors of the library—technical services, scholarly communication, information literacy and public services, etc.—to look at opportunities in Wikidata to advance the library’s strategic goals.

Conclusion

It is unlikely at this point that Wikidata can replace systems needed to manage collection inventory, transactions, and other practical aspects of managing linked data within library systems. However, it is vitally important for the library community to invest in linked data solutions that benefit the larger ecosystem of scholarly metadata creators. Linked data should connect and make more interoperable the information created by different disciplines and fields, helping knowledge become less balkanized by the scholarly communities that create it. This requires a global community of collaborators, across many disciplines, and globally interoperable technology.

While libraries have generally seen the benefit of making their data openly available as linked data and enabling semantic data within their systems, the ability to implement projects or create linkable data has been out of reach for many institutions and organizations. Some of this is due to lack of resources, but some is due to a lack of accessible tools and techniques as well as a lack of consensus on application use beyond first adopters. Many cataloging systems do not generate linked data, and few make data available as linked open data. Research libraries, by

getting involved in the Wikidata community as users and contributors, can help address these barriers. Using infrastructure such as Wikibase and Wikidata allows integration into a variety of systems and tools that can be used to maximize interoperability and minimize hands-on work, and potentially allows for efficiencies across the research library sector.

This white paper advances a number of recommendations for research libraries to contribute to Wikidata, and the ecosystem of linked data that is growing around Wikidata in the form of other Wikibases. There are still a number of open questions about the the long-term needs for maintaining library engagement with open platforms, like Wikidata, but there is momentum in that direction. Before the library community can develop a shared strategy for working in this space of collaborative linked data, it is important to both run pilot projects and develop understanding and skill among library staff. Through such work, the library community can develop an understanding of the circumstances under which Wikimedia engagement is a competitive way to invest institutional resources.

Endnotes

1. “About ARL,” Association of Research Libraries, accessed April 3, 2019, <http://www.arl.org/about>; “Association of Research Libraries (Q4810036),” *Wikidata*, last edited March 10, 2019, <https://www.wikidata.org/wiki/Q4810036>.
2. Heather Joseph, “Securing Community-Controlled Infrastructure: SPARC’s Plan of Action,” *College & Research Libraries News* 79, no. 8 (2018), <https://doi.org/10.5860/crln.79.8.426>.
3. Metadata 2020, accessed April 10, 2019, <http://www.metadata2020.org/>.
4. National Academies of Sciences, Engineering, and Medicine, *Open Science by Design: Realizing a Vision for 21st Century Research* (Washington, DC: The National Academies Press, 2018), <https://doi.org/10.17226/25116>.
5. Stephan Bartholmei, Rachel Franks, James Heilman, Mylee Joseph, Vicki McDonald, Anna Raunik, Mia Ridge, and Mark Robertson, *Opportunities for Academic and Research Libraries and Wikipedia*, IFLA and The Wikipedia Library, endorsed by IFLA Governing Board in December 2016, <https://www.ifla.org/files/assets/hq/topics/info-society/iflawikipediaopportunitiesforacademicandresearchlibraries.pdf>.
6. Wikidata website, accessed April 15, 2019, https://www.wikidata.org/wiki/Wikidata:Main_Page.
7. Stacy Allison-Cassin, “Research Libraries and Wikimedia: A Shared Commitment to Diversity, Open Knowledge, and Community Participation,” *Wikimedia Blog*, October 4, 2017, <https://blog.wikimedia.org/2017/10/04/libraries-wikipedia-york-university-project/>.
8. “Strategy: Wikimedia Movement: 2017,” *Wikimedia Meta-Wiki*, accessed April 3, 2019, <https://meta.wikimedia.org/wiki/Strategy/>

[Wikimedia_movement/2017/Direction#Our_strategic_direction:_Service_and_Equity.](#)

9. Wikibase website, accessed April 15, 2019, <http://wikiba.se/>.
10. “Wikidata: WikiProject Open Access,” *Wikidata*, last edited August 18, 2017, https://www.wikidata.org/wiki/Wikidata:WikiProject_Open_Access.
11. “Wikidata: Wikiproject Cultural Heritage,” *Wikidata*, last edited April 3, 2019, https://www.wikidata.org/wiki/Wikidata:WikiProject_Cultural_heritage; Beat Estermann, “How Wikidata Is Solving Its Chicken-or-Egg-Problem in the Field of Cultural Heritage,” *SocietyByte*, November 2018, <https://www.societybyte.swiss/2018/11/07/how-wikidata-is-solving-its-chicken-or-egg-problem-in-the-field-of-cultural-heritage/>.
12. Beat Estermann and members of the eCH Specialized Group on Open Government Data, *Linked Open Data*, eCH, March 13, 2018, <https://www.ech.ch/dokument/ae48e8b-2307-44fb-b1d3-ad61ed4432bc>.
13. “Archives at,” *Wikidata*, last edited March 15, 2019, <https://www.wikidata.org/wiki/Property:P485>.
14. Mairelys Lemus-Rojas and Lydia Pintscher, “Wikidata and Libraries: Facilitating Open Knowledge,” in *Leveraging Wikipedia: Connecting Communities of Knowledge*, ed. Merrilee Proffitt (Chicago: ALA Editions, 2018), 143–158, preprint, https://scholarworks.iupui.edu/bitstream/handle/1805/16690/Lemus-Rojas_Pintscher_Wikidata_2017-07-03.pdf; to provide just one example of how this type of contribution can be utilized, the following URL is a query to Wikidata that returns a list of people who were born on the current day (in any year) and who have an archival collection at Yale University: <http://bit.ly/archives-at-yale-birthday>.
15. “Wikidata: Tools,” *Wikidata*, last edited January 18, 2019, <https://www.wikidata.org/wiki/Wikidata:Tools>.

16. “Wikimedia and Libraries User Group,” *Wikimedia Meta-Wiki*, last edited March 8, 2019, https://meta.wikimedia.org/wiki/Wikimedia_and_Libraries_User_Group.
17. See “Wikidata: Licensing,” *Wikidata*, last edited February 28, 2018, <https://www.wikidata.org/wiki/Wikidata:Licensing>, and “Wikidata: Open Data Publishing,” *Wikidata*, last edited August 24, 2018, https://www.wikidata.org/wiki/Wikidata:Open_data_publishing.
18. SNAC (Social Networks and Archival Context) website, accessed April 15, 2019, <https://snaccooperative.org/>.
19. VIAF (Virtual International Authority File), accessed April 15, 2019, <https://viaf.org/>.
20. “Surfacing Knowledge, Building Relationships: Indigenous Communities, ARL and Canadian Libraries,” York University, accessed April 10, 2019, <https://scbrwiki.library.yorku.ca/>.
21. *Honouring the Truth, Reconciling for the Future* (Winnipeg: Truth and Reconciliation Commission of Canada, 2015), http://nctr.ca/assets/reports/Final%20Reports/Executive_Summary_English_Web.pdf.
22. Bartholmei et al., *Opportunities for Academic and Research Libraries and Wikipedia*.
23. Edit-a-thons are in-person events to which an organizer invites members of the community to fill gaps in Wikimedia projects by writing Wikipedia articles, describing content on Wikidata, transcribing WikiSource, or other editing activities. For guidance on running these events, see “Running Editathons and Other Editing Events,” *Wikipedia Programs & Events Dashboard*, accessed April 10, 2019, <https://outreachdashboard.wmflabs.org/training/editathons>.
24. “Wikipedian in Residence,” *Wikimedia Outreach*, last edited April 8, 2019, https://outreach.wikimedia.org/wiki/Wikipedian_in_Residence.

25. See for example “Jazz Wikipedia Edit-a-Thon,” Kansas City Public Library (held at American Jazz Museum, Kansas City, MO, March 5, 2018), accessed April 10, 2019, <https://www.kclibrary.org/library-locations/offsite-location/adults-event/jazz-wikipedia-edit-thon>; or the very successful campaign for Art+Feminism, accessed April 10, 2019, <http://www.artandfeminism.org/>.
26. “Health Information on Wikipedia,” *Wikipedia*, last edited March 13, 2019, https://en.wikipedia.org/wiki/Health_information_on_Wikipedia.
27. Michael A. Keller, Jerry Persons, Hugh Glaser, and Mimi Calter, *Report of the Stanford Linked Data Workshop, 27 June–1 July 2011* (Washington, DC: Council on Library and Information Resources, October 2011), <https://www.clir.org/pubs/reports/pub152/stanford-linked-data-workshop/>
28. Stacy Allison-Cassin and Dan Scott, “Wikidata: A Platform for Your Library’s Linked Open Data,” *Code4Lib Journal* 40 (2018), <https://journal.code4lib.org/articles/13424>.
29. Karen Smith-Yoshimura, “Analysis of 2018 International Linked Data Survey for Implementers,” *Code4Lib Journal* 42 (2018), <https://journal.code4lib.org/articles/13867>.
30. Heather Joseph, “Securing Community-Controlled Infrastructure: SPARC’s Plan of Action.”
31. Metadata 2020, <http://www.metadata2020.org/>.
32. National Academies of Sciences, Engineering, and Medicine, *Open Science by Design*.
33. For context on this practice, see “FAIR Data,” *Wikipedia*, last edited December 19, 2018, https://en.wikipedia.org/wiki/FAIR_data.
34. Karen Smith-Yoshimura, “Analysis of International Linked Data Survey for Implementers,” *D-Lib Magazine* 22, no. 7/8 (July/August 2016), <https://doi.org/10.1045/july2016-smith-yoshimura>.

35. Karen Smith-Yoshimura, “The Rise of Wikidata as a Linked Data Source,” *Hanging Together: The OCLC Research Blog*, August 6, 2018, <http://hangingtogether.org/?p=6775>.
36. Alissa Hafele and Michelle Futornick, “Background and Rationale,” Linked Data for Production (LD4P) website, last modified February 6, 2017, <https://wiki.duraspace.org/display/LD4P/Background+and+Rationale>.
37. The Wikimedia Foundation has since 2003 supported Wikipedia and its global volunteer community through fundraising and infrastructure support, including legal and technical support. Because of the commitment to open licensing and open software, the projects and their content can be exported or replicated by non-Wikimedia Foundation organizations for any reason at any time. Moreover, the Wikimedia Foundation as of 2018 is developing an endowment to indefinitely keep the projects online regardless of current fundraising ability (see the Wikimedia Endowment website, accessed April 10, 2019, <https://wikimediaendowment.org/>).
38. Several blog posts and articles outline the current and potential use of Wikidata in libraries. These include: Alex Stinson, “Wikidata in Collections: Building a Universal Language for Connecting GLAM Catalogs,” *Down the Rabbit Hole: Facts, Stories, and People from across the Wikimedia Movement*, December 13, 2017, <https://medium.com/freely-sharing-the-sum-of-all-knowledge/wikidata-in-collections-building-a-universal-language-for-connecting-glam-catalogs-59b14aa3214c>; “Why Data Partners Should Link Their Vocabulary to Wikidata: A New Case Study,” *Europeana Pro*, August 7, 2017, <https://pro.europeana.eu/post/why-data-partners-should-link-their-vocabulary-to-wikidata-a-new-case-study>; and Daniel Mietchen, Gregor Hagedorn, Egon Willighagen, Mariano Rico, Asunción Gómez-Pérez, Eduard Aibar, Karima Rafes, Cécile Germain, Alastair Dunning, Lydia Pintscher, and Daniel Kinzler, “Enabling Open Science: Wikidata for Research (Wiki4R),” *Research*

Ideas and Outcomes 1: e7573 (December 22, 2015), <https://doi.org/10.3897/rio.1.e7573>.

39. “List of Wikipedias,” Wikipedia, last edited March 29, 2019, https://en.wikipedia.org/wiki/List_of_Wikipedias.
40. *Wikimedia Commons*, last edited March 18, 2019, <http://commons.wikimedia.org>.
41. *Wikisource*, last edited January 4, 2019, <https://wikisource.org>.
42. In Mediawiki, the core software underneath Wikidata, namespaces distinguish different types of data in a project, and create distinct storage spaces for that content.
43. For more information about items and how to create them, see “Help: Items,” *Wikidata*, last edited April 27, 2018, <https://www.wikidata.org/wiki/Help:Items>.
44. The authors recognize this description is a simplification of the potential issues related to the use of statements that can be changed and edited over time. This issue is discussed in more detail in Allison-Cassin and Scott, “Wikidata: A Platform for Your Library’s Linked Open Data,” and requires further discussion and investigation of the impacts and solutions in research libraries.
45. A complete list and tools to browse them can be accessed at “Wikidata: List of Properties,” *Wikidata*, last edited February 3, 2019, https://www.wikidata.org/wiki/Wikidata:List_of_properties; the current count of properties can be obtained from “Wikidata Propbrowse,” *Hay’s Tools*, accessed April 10, 2019, <https://tools.wmflabs.org/hay/propbrowse/>.
46. “Library of Congress Authority ID (P244),” *Wikidata*, last edited April 5, 2019, <https://www.wikidata.org/wiki/Property:P244>.
47. “QuickStatements,” Wikimedia Toolforge, accessed April 15, 2019, <https://tools.wmflabs.org/quickstatements/#/>.
48. OpenRefine website, accessed April 15, 2019, <http://openrefine.org/>.

49. “Linked Data Wikibase Prototype,” OCLC Research, accessed April 10, 2019, <https://www.oclc.org/research/themes/data-science/linkedata/linked-data-prototype.html>.
50. For example, The Andrew W. Mellon Foundation–funded digital humanities project at Michigan State University, “Enslaved: Peoples of the Historic Slave Trade,” accessed April 10, 2019, <http://enslaved.org/>.
51. Wikimedia Deutschland, “Wikibase-docker,” GitHub, <https://github.com/wmde/wikibase-docker>.
52. DBpedia website, accessed April 15, 2019, <https://wiki.dbpedia.org/>.
53. Alessandro Piscopo, “Wikidata: A New Paradigm of Human-bot Collaboration?,” preprint, submitted October 1, 2018, <https://arxiv.org/pdf/1810.00931.pdf>.
54. Note the “Programme,” SWIB18: Semantic Web in Libraries (held in Bonn, Germany, November 26–28, 2018), accessed April 10, 2019, <http://swib.org/swib18/programme.html>.
55. “Help: Sources,” *Wikidata*, last edited January 16, 2019, <https://www.wikidata.org/wiki/Help:Sources>.
56. For example, “Collective Biographies of Women ID (P4539),” *Wikidata*, last edited November 5, 2018, <https://www.wikidata.org/wiki/Property:P4539>.
57. Joachim Neubert and Klaus Tochtermann, “Linked Library Data: Offering a Backbone for the Semantic Web,” in *Knowledge Technology*, ed. Dickson Lukose, Abdul Rahim Ahmad, and Azizah Suliman (Berlin, Heidelberg: Springer, 2012), 37–45, https://doi.org/10.1007/978-3-642-32826-8_4.
58. Joachim Neubert, “Linking Knowledge Organization Systems via Wikidata,” in *2018 Proceedings of the International Conference on Dublin Core and Metadata Applications* (Silver Spring, MD: Dublin

Core Metadata Initiative, 2018), 3, <http://dcpapers.dublincore.org/pubs/article/view/3975>.

59. Allison-Cassin and Scott, “Wikidata: A Platform for Your Library’s Linked Open Data.”
60. University of Wisconsin–Madison Libraries, “Bibcard,” GitHub, accessed April 10, 2019, <https://github.com/UW-Madison-Library/bibcard>.
61. Dan Scott, “Enriching Catalogue Pages in Evergreen with Wikidata,” *Coffee|Code: Dan Scott’s Blog*, August 12, 2017, <https://coffeecode.net/enriching-catalogue-pages-in-evergreen-with-wikidata.html>.
62. See Gloria Gonzalez (@InformaticMonad), “Here’s a quick thread of updates on my talk at #wikicite!...,” Twitter, November 29, 2018, <https://twitter.com/InformaticMonad/status/1068310528710213633>.
63. “SNAC Ark ID (P3430),” *Wikidata*, last edited March 26, 2019, <https://www.wikidata.org/wiki/Property:P3430>.
64. “Get Your Vocabularies in Wikidata...so Europeana and Others Can Get Them,” *Europeana Pro*, September 1, 2017, <https://pro.europeana.eu/page/get-your-vocabularies-in-wikidata>.
65. Rebecca Snyder and DucPhong “Ducky” Nguyen, “Smithsonian Open Data Pilot,” *Smithsonian Collaboration Wiki*, June 25, 2018, <https://confluence.si.edu/display/LODPP/Smithsonian+Open+Data+Pilot>.
66. “Wikidata for Digital Preservation: Home,” Yale University Library, last updated October 6, 2017, <https://guides.library.yale.edu/WikidataDigitalPreservation>.
67. “Wikidata: Planning a Wikidata Workshop,” *Wikidata*, last edited March 13, 2019, https://www.wikidata.org/wiki/Wikidata:Planning_a_Wikidata_workshop; “Works in Progress Webinar: Introduction to Wikidata for Librarians,” OCLC Research

- (June 12, 2018), accessed April 11, 2019, <https://www.oclc.org/research/events/2018/06-12.html>.
68. “The Distributed Game,” Wikimedia Toolforge, accessed April 11, 2019, <https://tools.wmflabs.org/wikidata-game/distributed/>.
 69. “Mix’n’match,” Wikimedia Toolforge, accessed April 11, 2019, <https://tools.wmflabs.org/mix-n-match/#/>.
 70. At time of writing fewer than 17,000 such statements exist: “Property Talk: P485,” *Wikidata*, last edited February 12, 2019, https://www.wikidata.org/wiki/Property_talk:P485.
 71. To search for entities without descriptions: “Entities without Description,” *Wikidata*, accessed April 11, 2019, <https://www.wikidata.org/wiki/Special:EntitiesWithoutDescription/en>.
 72. Useful identifiers include, but are not limited to, author/academic schemes like ORCID, VIAF, ISNI, Scopus; social media profiles; entries in public databases for extracurricular activities such as sports or IMDb. See lists at “Wikidata: Property Navboxes,” *Wikidata*, last edited March 15, 2019, https://www.wikidata.org/wiki/Wikidata:Property_navboxes.
 73. “Wikidata: Data Import Guide,” *Wikidata*, last edited February 15, 2019, https://www.wikidata.org/wiki/Wikidata:Data_Import_Guide.
 74. Justin Shanks and Kenning Arlitsch, “Making Sense of Researcher Services,” *Journal of Library Administration* 56, no. 3 (2016): 295–316, <https://doi.org/10.1080/01930826.2016.1146534>.
 75. ORCID website, accessed April 9, 2019, <https://orcid.org/>; “Wikidata: ORCID,” *Wikidata*, last edited September 19, 2017, <https://www.wikidata.org/wiki/Wikidata:ORCID>.
 76. Jim Blake and Mike Conlon, “VIVO,” *DuraSpace Wiki*, <https://wiki.duraspace.org/display/VIVO>.

77. Initiative for Open Citations (I4OC) website, accessed April 11, 2019, <https://i4oc.org/>. Note that both the Wikimedia Foundation and ARL are signatories to this effort.
78. WikiCite website, accessed April 11, 2019, <http://wikicite.org>.
79. Bianca Kramer and Jeroen Bosman, “101 Innovations in Scholarly Communication: The Changing Research Workflow,” (poster, FORCE2015, Oxford, UK, January 12–13, 2015), <https://doi.org/10.6084/m9.figshare.1286826.v1>.
80. “WikiCite,” *Wikimedia Meta-Wiki*, last edited November 16, 2018, <https://meta.wikimedia.org/wiki/WikiCite>.
81. See statistics at the foot of “Scholia,” Wikimedia Toolforge, accessed April 12, 2019, <https://tools.wmflabs.org/scholia/>.
82. Wikicite-discuss Google Group, accessed April 12, 2019, <https://groups.google.com/a/wikimedia.org/forum/#!/forum/wikicite-discuss>.
83. “Scholia,” <https://tools.wmflabs.org/scholia/>; “Scholia (Q45340488): Recently Published Works on the Topic,” Wikimedia Toolforge, accessed April 12, 2019, <https://tools.wmflabs.org/scholia/topic/Q45340488>.
84. For information on the project, see “IUPUI University Library Commitment to Open Knowledge,” IUPUI Center for Digital Scholarship, updated February 20, 2019, <https://www.ulib.iupui.edu/digitalscholarship/openknowledge>; Mairelys Lemus-Rojas and Jere D. Odell, “Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project Using Wikidata and Scholia,” *Journal of Librarianship and Scholarly Communication* 6, no. 1 (2018), <https://doi.org/10.7710/2162-3309.2272>.
85. “Source MetaData,” Wikimedia Toolforge, accessed April 12, 2019, <https://tools.wmflabs.org/sourcemd/>.

86. See “Wikidata: Zotero,” *Wikidata*, last edited July 29, 2018, <https://www.wikidata.org/wiki/Wikidata:Zotero>.
87. “Help: QuickStatements,” *Wikidata*, last edited April 6, 2019, <https://www.wikidata.org/wiki/Help:QuickStatements>.
88. Edit-a-thons may be known by other names, including workshops and other kinds of event titles.
89. Wikibase website, accessed April 12, 2019, <http://wikiba.se/>.
90. Matt Miller, “Wikibase for Research Infrastructure — Part 1,” *Medium*, March 9, 2018, <https://medium.com/@thisismattmiller/wikibase-for-research-infrastructure-part-1-d3f640dfad34>.
91. See “Linked Data Wikibase Prototype,” OCLC Research; *Using Wikibase as a Platform for Library Linked Data Management and Discovery* (Dublin, OH: OCLC, October 16, 2018), webinar video, 89 minutes, <https://www.oclc.org/en/events/2018/devconnect-online-workshops/wikibase-as-platform-for-library-linked-data-management-and-discovery.html>.
92. Sandra Fauconnier, “Many Faces of Wikibase: Rhizome’s Archive of Born-Digital Art and Digital Preservation,” Wikimedia Foundation, September 6, 2018, <https://wikimediafoundation.org/2018/09/06/rhizome-wikibase/>.
93. “Strategy: Wikimedia Movement: 2017,” *Wikimedia Meta-Wiki*, accessed April 12, 2019, https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017.
94. American Women’s History Initiative website, accessed April 12, 2019, <https://womenshistory.si.edu/>.
95. Science Stories website, accessed April 12, 2019, <http://www.sciencestories.io/>.
96. “User: YULdigitalpreservation/ScienceStories,” *Wikidata*, last edited March 8, 2019, <https://www.wikidata.org/wiki/User:YULdigitalpreservation/ScienceStories>.

97. “Wikidata: Article Placeholder,” *Wikidata*, last edited November 13, 2017, https://www.wikidata.org/wiki/Wikidata:Article_placeholder.
98. “Tag Archives: Article Placeholders,” National Library of Wales, accessed April 12, 2019, <https://blog.library.wales/?tag=article-placeholders>.
99. “Learning Patterns/Using Wikidata on Wikipedia Infoboxes,” *Wikimedia Meta-Wiki*, last edited September 28, 2017, https://meta.wikimedia.org/wiki/Learning_patterns/Using_Wikidata_on_Wikipedia_infoboxes.
100. “User: ListeriaBot,” *Wikipedia*, last edited March 29, 2019, <https://en.wikipedia.org/wiki/User:ListeriaBot>.
101. Allison-Cassin, “Research Libraries and Wikimedia.”
102. The Black Lunch Table website, accessed April 12, 2019, <http://theblacklunchtable.com/>.
103. “Wikipedia: Meetup/Black Lunch Table/Juneteenth Photo Challenge 2018,” *Wikipedia*, last edited June 22, 2018, https://en.wikipedia.org/wiki/Wikipedia:Meetup/Black_Lunch_Table/Juneteenth_Photo_Challenge_2018.
104. See discussion in the papers released by IFLA, “Presenting the IFLA Wikipedia Opportunities Papers,” news release, January 17, 2017, <https://www.ifla.org/node/11131>.
105. “Wikidata: WikiProject Sum of All Paintings,” *Wikidata*, last edited February 13, 2019, https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings.
106. See the list of institutions imported en masse by the project at “Wikidata: WikiProject Sum of All Paintings/Collection,” *Wikidata*, last edited April 2, 2018, https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings/Collection.
107. Crotos website, last updated April 5, 2019, <http://zone47.com/crotos/>.

108. Wiki Loves Monuments website, accessed April 12, 2019, <https://www.wikilovesmonuments.org/>.
109. “Largest Photography Competition,” *Guinness World Records*, accessed April 12, 2019, <http://www.guinnessworldrecords.com/world-records/largest-photography-competition/>.
110. “Connected Open Heritage,” *Wikimedia Meta-Wiki*, last edited December 13, 2016, https://meta.wikimedia.org/wiki/Connected_Open_Heritage.
111. “Wikipedian in Residence,” *Wikimedia Outreach*, last edited April 8, 2019, https://outreach.wikimedia.org/wiki/Wikipedian_in_Residence#List_of_Wikipedians_in_Residence.

ASSOCIATION OF RESEARCH LIBRARIES



Association of Research Libraries

21 Dupont Circle, NW
Suite 800
Washington, DC 20036
T 202.296.2296
F 202.872.0884
ARL.org
pubs@arl.org

Wikimedia Foundation

Alex Stinson
Senior Program Strategist
Twitter: [@glamwiki](https://twitter.com/glamwiki)/[@sadads](https://twitter.com/sadads)
astinson@wikimedia.org

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

