



ISSN 2162-3309 | JLSC is published by the Pacific University Libraries | <http://jlsc-pub.org>

Volume 6, General Issue (2018)

Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia

Mairelys Lemus-Rojas & Jere D. Odell

Lemus-Rojas , M. & Odell, J. (2018). Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia. *Journal of Librarianship and Scholarly Communication*, 6(General Issue), eP2272. <https://doi.org/10.7710/2162-3309.2272>

This article underwent semi-anonymous peer review in accordance with JLSC's peer review policy.



© 2018 Lemus-Rojas & Odell. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

PRACTICE

Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia

Mairelys Lemus-Rojas

Digital Initiatives Metadata Librarian, Indiana University-Purdue University Indianapolis

Jere D. Odell

Scholarly Communication Librarian, Indiana University-Purdue University Indianapolis

INTRODUCTION Wikidata, a knowledge base for structured linked data, provides an open platform for curating scholarly communication data. Because all elements in a Wikidata entry are linked to defining elements and metadata, other web systems can harvest and display the data in meaningful ways. Thus, Wikidata has the capacity to serve as the data source for faculty profiles. Scholia is an example of how third-party tools can leverage the power of Wikidata to provide faculty profiles and bibliographic, data-driven visualizations. **DESCRIPTION OF PROGRAM** In this article, we share our methods for contributing to Wikidata and displaying the data with Scholia. We deployed these methods as part of a pilot project in which we contributed data about a small but unique school on the Indiana University-Purdue University Indianapolis (IUPUI) campus, the IU Lilly Family School of Philanthropy. **NEXT STEPS** Following the completion of our pilot project, we aim to find additional methods for contributing large data collections to Wikidata. Specifically, we seek to contribute scholarly communication data that the library already maintains in other systems. We are also facilitating Wikidata edit-a-thons to increase the library's familiarity with the knowledge base and our capacity to contribute to the site.

Received: 08/03/2018 Accepted: 10/18/2018

Correspondence: Mairelys Lemus-Rojas, Indiana University-Purdue University Indianapolis, 755 West Michigan Street, Indianapolis, IN 46202, mlemusro@iupui.edu



© 2018 Lemus-Rojas & Odell. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

Universities, schools, departments, research labs, and other academic units have a strong interest in collecting and displaying bibliographic information about the work created by their scholars. With access to a departmental website or a university content management system, many academic organizations meet this need by listing publications on a web page or by providing links to the curriculum vitae files of affiliated scholars. Although widespread, this approach to collecting and sharing bibliographic information is cumbersome for both its creators and its readers. These web-based lists are often out of date, prone to errors and irregularities, and not structured in ways that facilitate reuse. There are many partial solutions to this problem, but most suffer from limitations that slow adoption (relying on individuals to populate their own profiles) or exclude users (using expensive, proprietary systems limited to subscription data). Wikidata, a knowledge base for structured linked data, provides an open data alternative—one that is machine readable, free to edit and use, and maintained in the commons. Unlike systems that depend on authors to populate their own data profiles, Wikidata was designed for crowdsourced data entry. When data is gathered from publicly available sources (for example, web pages, online catalogs, publications, and bibliographic databases), it can be contributed to Wikidata with references to their sources. Because all elements in a Wikidata entry are linked to defining elements and metadata, other web systems can harvest and display the data in meaningful ways. Thus, Wikidata has the capacity to serve as both “a linking hub” at the center of disparate data systems and as a data source for faculty profiles for Research Information Management Systems (RIMS), Current Research Information Systems (CRIS), and other data-driven faculty profile systems (Neubert, 2017). Scholia is an example of how third-party tools can leverage the power of Wikidata to provide faculty profiles and bibliographic data-driven visualizations.

In this article we describe our efforts to use Scholia, an open source tool for displaying faculty profiles and authorial networks sourced from Wikidata, to build a faculty profile collection for a unique school on the Indiana University-Purdue University Indianapolis (IUPUI) campus. We share the methods for contributing bibliographic information to Wikidata and the current features available to users of Scholia. We also reflect on the challenges, limitations, and rewards for libraries that seek to use these tools to curate and share information about their institutions.

LITERATURE REVIEW

Faculty Profile Systems

Numerous web-based tools for recording and displaying data about scholarly productivity

are available to authors, including, for example, Open Researcher and Contributor ID (ORCID), Google Scholar Profiles, and even LinkedIn. In addition to sites like these that help authors create a scholarly bio page or something resembling a web-based curriculum vitae, many other systems are available for recording faculty productivity for annual reviews, grants, and other institutional uses. An existing Wikipedia table currently lists more than 50 systems—grouping these sites together under the label of “Research Networking” tools (Comparison of research networking tools, n.d.). Similarly, Heller (2015) provides a table of attributes for seven systems that make available data-driven scholarly profile pages. Both of these lists group tools that were developed to meet diverging needs—some tools were developed for individuals to track citations and other uses of their works (e.g., Google Scholar and ImpactStory), while others were developed for individuals to network with scholars with shared interests (e.g., ResearchGate and Academia.edu), to manage identities (ORCID), or to collect and assess scholarly productivity at the institutional level (e.g., VIVO, Symplectic Elements, and Digital Measures).

The literature on these tools tends to focus on two user groups, researchers and institutions. Shanks and Arlitsch (2016) review tools for their utility to individual researchers and group them in three categories: identification, networking, and citation management. In contrast, Givens (2016); Givens, Macklin, and Mangiafico (2017); and Bryant, Namachchivaya, and Speer (2017) focus on Research Information Management Systems (RIMS), also known as Current Research Information Systems (CRIS). RIMS (such as VIVO, Converis, Symplectic Elements, Elsevier’s PURE, and Activity Insight by Digital Measures) do have features that will be useful to individuals. PURE, for example, displays author profiles that include altmetrics and limited coauthorship networks. Symplectic Elements provides similar features and also permits individuals to edit their profiles. Nonetheless, most RIMS were not designed and have not been deployed to meet the online identity management needs of individual authors; rather, they were developed to collect and manage data about academic units—grant-funded teams, departments, schools, and universities. Because universities and other institutions subscribe to and deploy them, RIMS are more likely to include data about the grant-seeking activities of affiliated faculty members and other data that will not be shared beyond the offices of university administrators.

In contrast to faculty profile systems, researcher networking sites, and RIMS, Wikidata collects data about any topic—the scope of its content and its contributors is limited only by the community of users that defines terms and supplies and curates entries. Wikidata was not designed to be used to self-publish a biographical web page or for enterprise-only data tracking. However, it does include data about scholarly works, academic authors, and their institutions. Academic libraries are beginning to explore ways to leverage Wiki-

data's "relatively user-friendly interfaces" to create linked open data about their collections (Allison-Cassin & Scott, 2018). In addition, Wikidata includes communities of users and developers working to establish the site as a foundation for open science. These communities include groups that seek to increase the site's ability to serve as a virtual research environment, such as Wikidata: WikiProject Wikidata for research, a group that began with the intention to acquire funding from the European Commission's Horizon 2020 program (Mietchen et al., 2015), and Wikidata: WikiProject Chemistry, a group that defines the properties and curates data related to chemistry (Wikidata: WikiProject Chemistry, n.d.). Similarly, and with a specific focus on bibliographic information, a community of users is involved in Wikidata: WikiProject Source MetaData, which serves as the hub for groups working on the capacity for Wikidata to serve as a repository for bibliographic and citation data (Wikidata: WikiProject Source MetaData, n.d.). Notable efforts of the Source MetaData project include WikiCite and a key tool for the service that we describe, Scholia.

OVERVIEW OF THE PROJECT SYSTEMS

Wikidata

Wikidata is a free knowledge base where structured linked data are stored. It was developed in 2012 with the purpose of serving as the hub for all projects under the umbrella of the Wikimedia Foundation, a nonprofit organization. Wikidata (accessible at <https://www.wikidata.org/>) can be edited and read by both humans and machines. The data in Wikidata are created and maintained by volunteers from around the world with diverse subject expertise, and they are published under the Creative Commons Public Domain Dedication 1.0, which means that they are free to use, modify, and distribute (Wikidata: Data access – Wikidata, 2017). All the edits made in Wikidata can be accessed at any time and are recorded in their respective namespace (storage space for different types of data, e.g., *main namespace* used only for items, *property namespace* used for properties, *user namespace* used for user pages, etc.). This allows for changes to be reverted whenever needed. The number of new Wikidata items increases daily. At the end of May 2017, 26 million content pages had been added to the site (Lemus-Rojas & Pintscher, 2018); a year later that number had increased to more than 48 million (Statistics – Wikidata, n.d.).

Wikidata follows the *anyone can edit* model used by Wikipedia as a way of fostering engagement with the project. The Wikidata knowledge base is multilingual in itself, which means that data in hundreds of languages is stored in one central location, making it language independent. The multilingual nature of Wikidata makes it possible for editors

to contribute to the same entry regardless of their language expertise or preference. Wikidata contains entities that are either properties or items. They both have their individual namespaces and Wikidata pages containing labels, descriptions, and aliases that can be entered in different languages (Wikibase/DataModel/Primer, 2017). Both properties and items are assigned unique identifiers that start with the letter “P” for property or “Q” for item, followed by numbers.

Properties must be proposed by editors and undergo a discussion among editors who have the option to support or oppose their creation. If the community reaches a favorable consensus, a Wikidata property creator or administrator is able to create the property. Once available in the knowledge base, the property can be used by any editor. This community-controlled process explains why there are only about 5,000 properties in Wikidata.

Unlike properties, items can be created by anyone. Items represent entities (including people, topics, works, organizations, objects, and concepts) and contain a series of statements describing them. Each statement consists of one property and at least one value (known as a property-value pair), one or more qualifiers, and one or more references. For instance, as seen in Figure 1, the item we created for Una Osili, professor of economics and philanthropic studies and one of the faculty members chosen for the pilot project, contains the property *employer* (P108), and the item *Northwestern University* (Q309350) as its value. *Northwestern University* is an item in itself, and in this case is being used as the value for the property *employer*. In order to be able to use items as values in a statement, they first need to be created. More specific details can be added to values present in the statements in the form of *qualifiers*. Qualifiers allow contributors to provide more accurate information about the claim being made and are also recorded using a property-value pair. In the example of Una Osili’s entry, one can see that she held a position as teaching assistant at Northwestern University from 1996 to 1998—this additional information given in the form of qualifiers specifies Osili’s term of employment and position held at the university. As positions are added to Osili’s entry, a timeline of her work history begins to take shape. For these positions, like all claims in Wikidata, contributors are encouraged to add supporting references (see Figure 1).

Information in Wikidata is created, curated, and maintained by a community of about 18,000 active contributors (Statistics - Wikidata, n.d.). In doing this work, a variety of tools and web applications have been created by community members to facilitate content contribution and to display the data in more effective and meaningful ways. One of the tools, Scholia, was created to display bibliographic data.

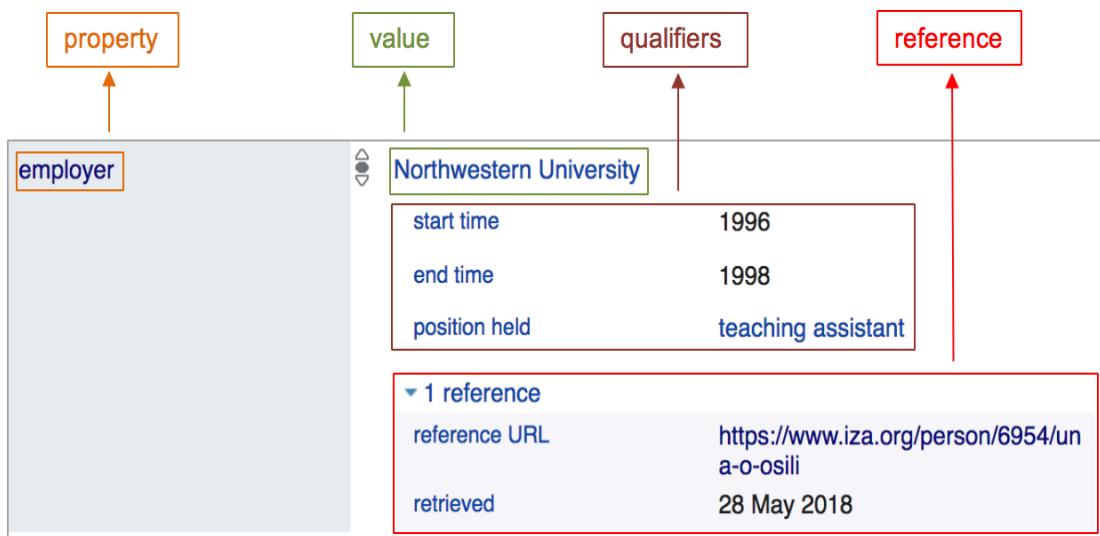


Figure 1. Una Osili's example of employer statement in Wikidata. Complete entry accessible at <https://www.wikidata.org/entity/Q32979516>

Scholia

The development of Scholia was inspired by the work of WikiCite, “an initiative (and a series of events) aiming to build a bibliographic database in Wikidata to serve free knowledge” (WikiCite, n.d.). Scholia is an open source web service hosted by the Wikimedia Toolforge and can be accessed at <https://tools.wmflabs.org/scholia/> (Nielsen, Mietchen, & Willighagen, 2017). Still in active development, its source code is available on GitHub, a web-based code-repository hosting service (Nielsen, 2018). With Scholia, it is possible to generate scholarly profiles using the structured data that reside in Wikidata. The tool works by making live SPARQL calls (SPARQL Protocol and RDF Query Language)—a language used for querying databases, thus providing the most up-to-date information in its display.

In Scholia, a user can perform a search by using what the tool identifies as “aspects” (Nielsen et al., 2017). The available aspects are *Author*, *Work*, *Organization*, *Venue*, *Series*, *Publisher*, *Sponsor*, and *Topic*. Examples of all of these aspects generated by Scholia are provided on its home page. Scholia has an autocomplete search feature, which means that as soon as a user starts typing into the search box, a list of suggested matches is displayed along with their description, when available. In this way, entries with the same name can be easily identified. For instance, when searching for Una Osili in Scholia, after her item was created in Wikidata, one can see that her name displays under the search bar as *Una Osili - professor of economics and philanthropic studies* (see Figure 2). Osili’s scholarly profile displays a list of publications, the number

of publications and pages produced per year, the journals where the works are published, a co-author graph, topics, associated images, education and employment timelines, locations, and citation statistics (including most-cited work, citations by year, and a list of citing authors).

Search

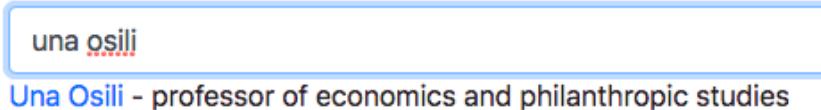


Figure 2. Una Osili search result in Scholia

Scholia provides users with a different way of exploring and experiencing the data contained in Wikidata. The tool exposes a bibliographic subject's relationships (authors, publications, pages, citations, etc.) by providing a graphical interface for interacting with the aggregated data.

DESCRIPTION OF PROGRAM

IU Lilly Family School of Philanthropy Pilot Project

To explore the potential of Wikidata, and more specifically of Scholia as a scholarly profile generator, we decided to conduct a pilot project in the summer of 2017. The goal of the pilot project was to provide a presence in Wikidata for faculty members at IUPUI, their publications, and their coauthors. Aligning with IUPUI University Library's commitment to contributing to open knowledge and open access projects, we sought to explore the possibility of offering the creation of scholarly profiles as a library service. As an added benefit, in doing this work we are also contributing to the corpus of free, structured linked data scholarly records in Wikidata.

We started by identifying potential subjects for the work and decided to select the IU Lilly Family School of Philanthropy for the pilot project. Located on the IUPUI campus and established in 2012, the School of Philanthropy is the first school of its kind—dedicated entirely to research and education in the field of philanthropy (Lilly Family School of Philanthropy, n.d.). We selected all 18 faculty members who were listed as core faculty on the school's website at the time, and subsequently added a newly hired faculty member for a total of 19. We started by checking Wikidata to make sure that none of the faculty members or the school were already in the knowledge base. We created an entry for the school and proceeded to cre-

ate a spreadsheet in Google Sheets to record identifiers pertaining to the faculty. Including one or more identifier on the author's Wikidata entry facilitates connections with other data sources where web users can find additional information about the author. We searched for identifiers in ORCID, Scopus, ResearcherID, Google Scholar, ISNI (International Standard Name Identifier), Twitter, and VIAF (Virtual International Authority File). From VIAF, we used the VIAF ID and the LCCN (Library of Congress Control Number) as well as any birthdate information present on the authority record. We also searched LinkedIn for faculty members' personal profiles.

After we completed our initial data gathering process, we created entries for the faculty in Wikidata, following a list we put together of the most commonly used (but not required) existing properties for authors, educators, and researchers (see Table 1).

During the content creation process, we found some additional information about the faculty and included other existing properties that were not part of our initial reference list (see Table 2).

After collecting identifiers and creating initial entries for the faculty members, we began collecting additional biographical and bibliographic information. We gathered this information from each faculty member's profile on the school's website. In most cases, the school's website included brief biographies and links to curriculum vitae. Using these publicly available sources, we were able to gather information about faculty members' professional careers (e.g., publications, employment history, and education). In cases where we did not find sufficient publication information, we turned to Scopus and Google Scholar. Using Zotero, we built collections of publications for each faculty author and exported the collections as CSV files. Once the publication information was gathered, we searched Wikidata to make sure they were not already available, and proceeded to create entries in the knowledge base. Initially, we created entries for the journal articles manually, following another list of properties as a reference (see Table 3). We found ourselves checking Scholia after making a contribution in Wikidata to see how the profiles were coming together. This helped us get a better understanding of what was possible to achieve with Scholia once the data were properly linked in Wikidata (see Table 3).

Although we were excited about contributing to Wikidata and were closely watching the growth of the Scholia profiles, we also came to the realization that reaching our desired goal of adding all selected faculty members, their coauthors, and at least one article linked to the faculty member in Wikidata was going to be more time consuming than expected. For that reason, and in an effort to make the workflow more efficient, we decided to try the Source MetaData tool (accessible at <https://tools.wmflabs.org/sourcemd/>), which searches

Property Label	Property ID
instance of	P31
given name	P735
family name	P734
sex or gender	P21
country of citizenship	P27
date of birth	P569
place of birth	P19
languages spoken, written, or signed	P1412
occupation	P106
field of work	P101
employer	P108
award received	P166
educated at	P69
affiliation	P1416
LinkedIn personal profile URL	P2035
ORCID iD	P496
Scopus Author ID	P1153
Google Scholar author ID	P1960
ISNI	P213
Twitter username	P2002
VIAF ID	P214
Library of Congress authority ID	P244
ResearcherID	P1053

Table 1. Primary list of Wikidata properties for individuals (in no particular order)

Property Label	Property ID
spouse	P26
doctoral student	P185
doctoral advisor	P184
work location	P937
official website	P856
NLA (Australia) ID	P409
BIBSYS ID	P1015
SUDOC authorities ID	P269
BNE ID	P950
GND ID	P227
NUKAT (WarsawU) authorities ID	P1207

Table 2. Secondary list of Wikidata properties for individuals (in no particular order)

Property Label	Property ID
instance of	P31
title	P1476
author*	P50
author name string	P2093
main subject*	P921
publication date	P577
cites*	P2860
full work available at*	P953
published in*	P1433
volume	P478
issue	P433
page(s)	P304
number of pages*	P1104
language of work or name*	P407
DOI	P356
arXiv ID	P818

Table 3. List of Wikidata properties for articles

* Properties marked with an asterisk (*) were added manually (the Source MetaData tool did not add them at the time)

for bibliographic metadata based on PMID (PubMed identifier for scholarly works), PM-CID (PubMed Central unique identifier), or DOI (digital object identifier). When any of these identifiers are entered, the tool pulls in metadata describing the item, in this case a journal article. It then presents a link to QuickStatements (accessible at https://tools.wmflabs.org/wikidata-todo/quick_statements.php), which is another tool that executes automatic edits in Wikidata. Using this set of tools helped us move faster, but one limitation was that we could only work on articles that had identifiers, which in our case were mostly those with DOIs. After we created a couple of entries, we realized that we still needed to manually add some properties in order to provide more complete scholarly profiles. We added the following properties on the resulting article entries created using the Source MetaData and QuickStatements tools: *published in* (P1433), *number of pages* (P1104), *language of work or name* (P407), and *author* (P50). The tools automatically assign authors' names to the *author name string* (P2093) property (see Figure 3), which is used for unspecified names. What this means is that unless we created an entry for all coauthors, their names were not going to be listed on the profiles generated in Scholia for the faculty members selected. For instance, Una Osili had 14 coauthors that were not part of the selected group of faculty for the pilot project. As a result, we had to create items for all those authors, who were mostly from other institutions, to be able to get an accurate coauthors graph in Scholia. We approached this new challenge in two ways: by creating entries manually and by using the Resolve authors tool (accessible at https://tools.wmflabs.org/sourcemd/new_resolve_authors.php). The Resolve authors tool allows users to link documents (articles, books, etc.) to authors. If the author is not found in Wikidata, then the tool offers the option to create them. After creating the items for all coauthors, we added the property *author* in the entry for the publication to include their names (see Figure 4). Another property we manually added to some of the publications was *cites* (P2860). This property was created during the first WikiCite event in 2016 with the purpose of establishing a connection between creative works.

author name string	Chin Hua Kong ...
	series ordinal 3

Figure 3. Example of the author name string property automatically added to publications in Wikidata



Figure 4. Example of the author property added manually to publications in Wikidata

After we finished creating the items for our faculty, their coauthors, and their publications in Wikidata, we were able to interact with all the data in Scholia. For example, by searching for Una Osili in Scholia, we can see a nearly complete scholarly profile, accessible at <https://tools.wmflabs.org/scholia/author/Q32979516>. This includes her full name followed by a unique identifier in parentheses (assigned by Wikidata at the top of her page), and a link to her ORCID profile. When an item in Wikidata is linked to its corresponding article in the English Wikipedia, then Scholia displays a brief description of the subject as found on the online encyclopedia with a link to its page (Nielsen et al., 2017). But this is not applicable in this case since there is no article about Una Osili in Wikipedia. A list of publications including date, title of the work, type of work, number of pages, journal where the article was published, and authors is presented in a table (see Figure 5). It then shows a color-coded bar chart for the number of publications Osili produced per year and her level of involvement (first author, last author, or solo author). Another bar chart displays the number of pages she has written color-coded by article (See Figure 5).

Scholia also displays information related to the journals where the author's work was published. Here too, the bubble chart is color-coded by journal and the size of the bubbles corresponds with the number of articles published in each journal. The same information is displayed in a table form. A coauthors graph shows Osili's level of collaboration with other scholars in the field, internally and externally (see Figure 6). Another color-coded bubble chart is used to generate topics, and a list of topics for Osili's authored works is shown in table form. Associated images are displayed in addition to timelines for her education and employment history. The next section is an academic tree that displays the connection between doctoral advisors and their students. In this case, we did not have any sourced

author

Una Osili (Q32979516)

 <https://orcid.org/0000-0001-7447-8857>

List of publications

Show entries

Search:

Date	Work	Type	Pages	Venue	Authors
2017-05-26	Philanthro-metrics: Mining multi-million-dollar gifts	scholarly article		PLOS ONE	Robert P. Light, Jacqueline Ackerman, Chin Hua Kong, Katy Börner, Una Osili
2017-04-01	Beyond Grantmaking: An Investigation of Program-Related Investments by U.S. Foundations	scholarly article	25	Nonprofit and Voluntary Sector Quarterly	Heng Qu, Una Osili
2014-07-31	Examining the Association of Religious Context with Giving to Non-Profit Organizations	scholarly article	15	European Sociological Review	Pamala Wiepkink, Una Osili, René Bekkers
2014-04-01	The Global Dynamics of Gender and Philanthropy in Membership Associations: A Study of Charitable Giving by Lions Clubs International Members	scholarly article	21	Nonprofit and Voluntary Sector Quarterly	Una Osili, Debra Mesch, Amir Daniel Hayat, Xiaonan Kou
2014-03-01	Crises and confidence: Systemic banking crises and depositor behavior	scholarly article	15	Journal of Financial Economics	Anna Paulson, Una Osili
2010-06-01	Giving to family versus giving to the community within and across generations	scholarly article	25	Journal of Population Economics	Una Osili, Partha Deb, Cagla Okten
2009-04-01	Do Immigrants and Their Children Free Ride More Than Natives?	scholarly article	7	The American Economic Review	Una Osili, Jia Xie

Figure 5. Una Osili's list of publications in Scholia

information to add to Wikidata, which is why the academic tree section is blank. Scholia also displays a map of all locations associated with the faculty member, a table with her most-cited works, a bar chart with a count of citations per year, and another table listing citing authors. A user is able to change the display of any of these sections at any time, as well as access the query that was used to generate the data. In some of the sections, Scholia even includes a link to preconfigured queries in the Wikidata Query Service that aids the user in figuring out other pieces of information that could be missing. For instance, the first link provided on the coauthors graph section, shown in Figure 6, can be used to search for coauthors whose items have not been created yet, but whose names appear listed in the item for the publication in Wikidata under the *author name string* property (see Figure 3).

Since Scholia ignores the values present in this field when rendering the coauthors graph and publications table, items need to be created for the coauthors and their names entered in the *author* field for the publication (see Figure 4). When a contributor knows what items need to be created based on the results of the query, they can use the links provided in Scholia to access external tools to add content to Wikidata, which in this case is the Resolve authors tool.

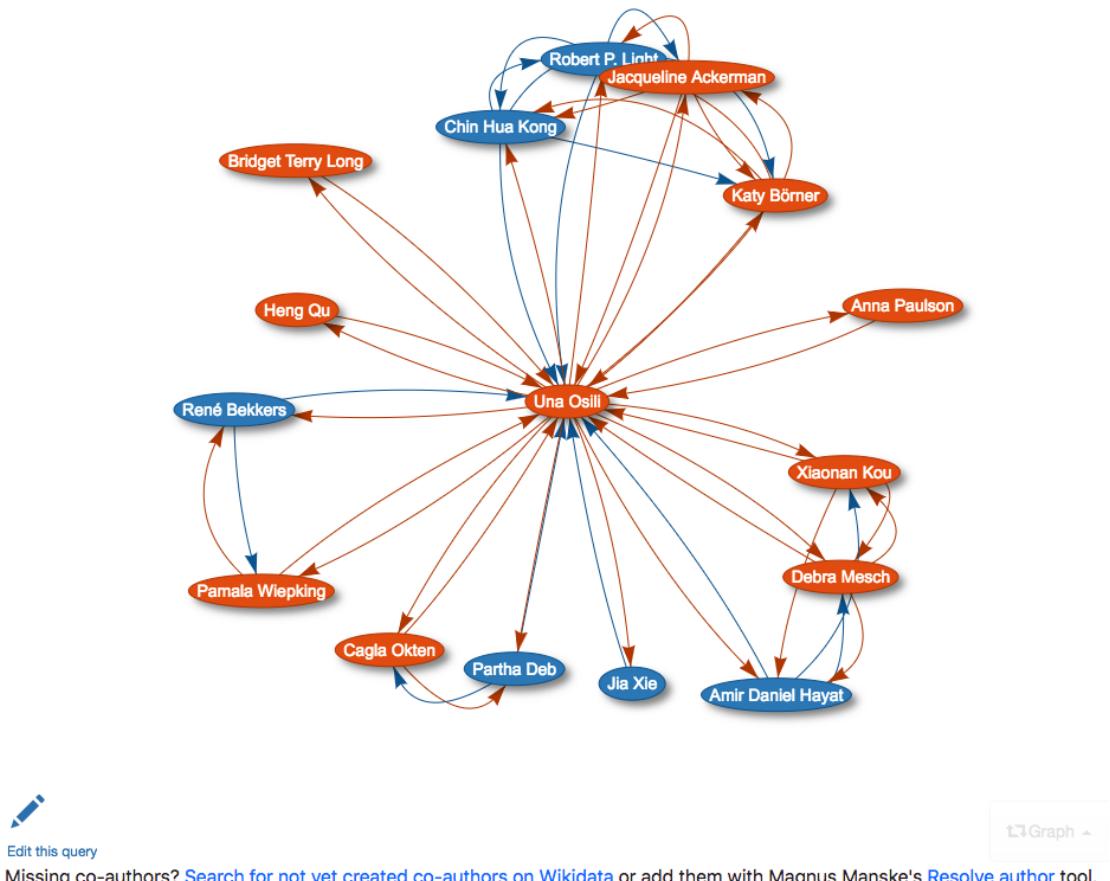


Figure 6. Una Osili's coauthors graph in Scholia

While interacting with the faculty profiles in Scholia, one has the ability to explore the usage of works authored by these faculty members. For instance, if one clicks on the title of the article “Crises and Confidence: Systemic Banking Crises and Depositor Behavior” (see the fifth entry in Figure 5), Scholia will display a page with information specific to the work. A title with a unique identifier generated by Wikidata, the DOI for the work, a list of

authors and their ORCIDs, topics, citations to the work (number of citations, publication date, citing work), cited works (number of citations, publication date, cited work), citation graphs, and citations per year (citations from others) are all elements found in Scholia for works. The citation graph for this work (see Figure 7) displays the names of the first authors for the works and their connection to our faculty member Una Osili. The directional arrows indicate a work citing another work. For instance, the directional arrow pointing to Osili from David-Jan Jansen indicates that he cited Osili's article. Additional information can be found by clicking on any of the nodes, which allows users to better understand the citation graph. By clicking on Jansen, one learns that his article was written in English and published in the *Journal of Financial Services Research*, and that it cites Osili's article. This illustrates how powerful Scholia can be in exploring the scholarly literature.

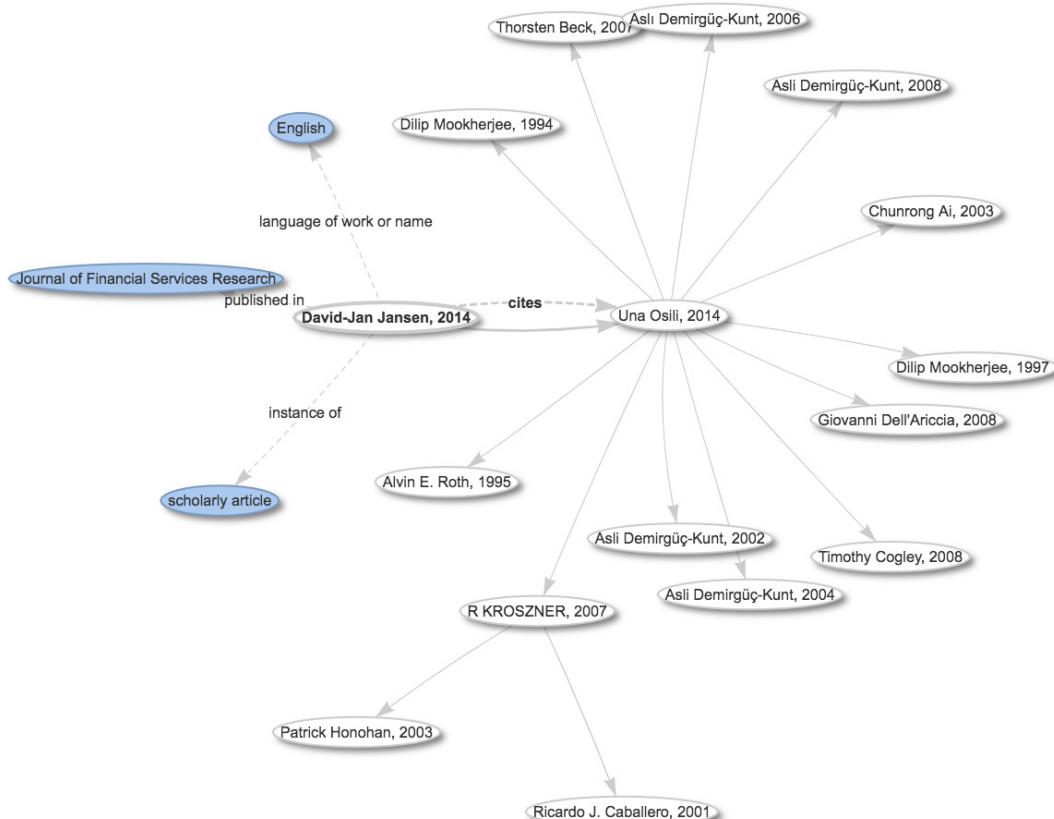


Figure 7. Citation graph in Scholia for the article Crises and confidence: Systemic banking crises and depositor behavior by Una Osili and Anna Paulson

OUTCOMES

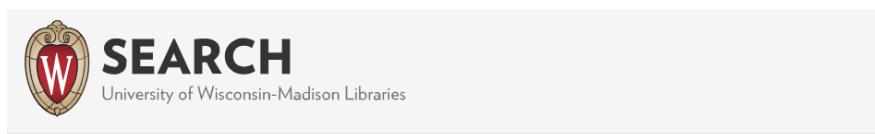
The IU Lilly Family School of Philanthropy pilot project served as an opportunity for us to explore the potential of Wikidata and Scholia to create and generate scholarly profiles for the faculty. We started with 18 selected faculty and recently included a newly hired faculty member, for a total of 19. In addition to creating items in Wikidata for the faculty, we created 58 items to represent coauthors. Since the initial goal was to be able to generate scholarly profiles for the faculty, we also created 110 publication items. Of these publications, we selected three and created a total of 39 items for works cited. Part of that work is represented in the citation graph generated by Scholia as shown in Figure 7. We also looked for works that cited the three publications we selected and created entries for them (David-Jan Jansen's article is one example). In this way, we started to slowly interconnect works—a process that resulted in a nicely rendered citation graph in Scholia. When we concluded the pilot, we estimated a combined edit count of 8,000, but we have since made additional contributions to the project.

One of the things that makes Wikidata so powerful is that it has the capacity to store identifiers that point to external data sources, allowing data connectivity. At the same time, and despite the fact that Wikidata is still growing and far from complete, its data are being used to enrich existing external data sources. For instance, a project called BibCard, developed by the University of Wisconsin–Madison Libraries, uses linked data to generate library data knowledge cards. This project pulls in linked data from external data sources, among which is Wikidata, to enhance the library catalog (Meyer, 2018). Searching for Una Osili on their search page provides a list of publications by the faculty member. When clicking on the title for the work, a bibliographic record displays and at the bottom of the page there is a section that can be expanded to check data about the authors from other sources (see Figure 8). Here one may see how the data from the Wikidata item we created for Osili is being displayed in the library catalog, as well as the details about the author currently available to users. Traditionally, library catalogs do not display information about authors beyond their names. This is an excellent example of how community-driven contributions to Wikidata's open-linked data platform results in reuse by other systems—in this case, Wikidata enhances the representation of authoritative data in library catalogs (see Figure 8).

NEXT STEPS

Integrating with the Mission

The work of building database-driven, faculty profiles using an open source project (Wikidata) and a web service-based tool (Scholia) aligns with the mission and values of both our



SEARCH
University of Wisconsin-Madison Libraries

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](#)

Harvard University

Northwestern University

Identities on The Web

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

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

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

Figure 8. Una Osili's information in the University of Wisconsin–Madison Libraries catalog

library and our campus. Our library seeks to transform “information to new and more accessible formats” while also connecting a “wider community of learners” with resources (IUPUI University Library, 2015). Likewise, IUPUI’s emphasis on “civic engagement,” “external partnerships,” and “cultural contributions to the well-being of the citizens of Indianapolis, the state of Indiana, and beyond” encouraged us in our exploration of this potential service to our constituents on and off campus (IUPUI, 2018). As an open-linked data system, Wikidata serves as a platform to contribute and curate data about our institution in

the commons. The data that we have contributed throughout this pilot project can be freely accessed on Scholia and other sites by stakeholders in our communities (students, philanthropy scholars, policymakers, citizens, and many more). Similarly, the Wikidata entries that we created will continue to be enhanced by other contributors and will be linked to an ever-growing network of related entries. As an open-data alternative, Scholia and Wikidata can also meet the needs of librarians working for smaller institutions. At the same time, Scholia and Wikidata provide a welcome antidote to a marketplace that seeks to capture data about university authors and their works to build expensive, proprietary information tools.

As of 2013, more than half of all annually published journal articles were owned by five large publishers. In some fields, such as psychology, this consolidation has topped 70% of all articles published (Larivière, Haustein, & Mongeon, 2015). These publishers now have near monopoly control of the scholarly publishing marketplace—one that drains library budgets and limits access to knowledge. One of these companies, Elsevier, recently rebranded itself as an “information analytics business” and has made a series of investments to purchase an entire pipeline of scholarly data platforms (Krauss, 2017). As universities make purchasing decisions about research information systems and academic analytics tools, Wikidata and Scholia have the capacity to provide a subscription-free, community-owned, open-source alternative. IUPUI is a public university campus, largely supported by student tuition, state and federal taxes, grants, and philanthropic gifts. With this in mind, our pilot project demonstrates the ability to give a portion of the data back to the communities that funded it to begin with.

Addressing the Challenges

Although we think that Wikidata and the tools and web services that it enables, like Scholia, have the potential to be integrated with existing library data curation activities, our pilot project revealed a number of challenges for those who seek to use this approach to building a collection of open faculty profile data. These challenges may be grouped under two topics: institutional buy-in and labor.

Institutional Support and Issues of Trust

In our experience, faculty authors are unaware of Wikidata and unsure of the value of their bibliographic presence on the site. On the one hand, it is possible that academic authors may be suspicious of a site that *anyone can edit*—while at the same time, librarians and other information professionals may be wary of participating in open community platform that may expose contributors to discriminatory behavior and harassment.

Our pilot project did not aim to build trust in Wikimedia projects, but we do believe it contributes to a better understanding of how data can be contributed and curated by the commons in a responsible and respectful manner. The Wikimedia Foundation Terms of Use encourage users to “support a civil environment” and to “not harass other users” (Wikimedia Foundation, n.d.). They also ask users to disclose and avoid potential conflicts of interests. The Terms of Use introduce two aspects that will be of interest to library-supported Wikidata projects like the one we describe. First, perhaps reflecting the fact that data is less likely to result in disputes about nuance, or the fact that conflicting data points can coexist in the knowledge base, or perhaps in an indication of the site’s age, Wikidata, unlike Wikipedia, has not been the subject of high-profile news stories about discrimination and harassment (Paling, 2015; Koren, 2018). In addition, the conflict of interest policy in the Terms of Use discourages users from contributing data about themselves. This introduces both a strength and a limitation for using Wikidata as a RIMS tool. On the one hand, Wikidata for scholarly profile data can avoid the lag that other systems may face while waiting for faculty authors to login and add or approve data. On the other hand, Wikidata depends on third parties for data entry (these include librarians, scholars, other individuals, and robots (or “bots”) that batch-ingest data from PubMed, CrossRef, and other scholarly sites). While this reduces the sense of ownership a faculty member may have for data about their work, it also reflects the reality of many RIMS implementations. For example, three years into an implementation of VIVO, Duke University reported that only 35% of faculty had logged in—at the same time, 80% of faculty had VIVO profiles with data contributed by third parties (Givens et al., 2017, p. 242).

As we contribute data for authors in the School of Philanthropy, we are building a collection for demonstrating the potential of Wikidata and Scholia. Many customers of vendor-supplied data systems expect a demonstration of the product before making an investment. By showing how Scholia and Wikidata work together, our colleagues in the library and on our campus will be able to make more informed decisions about whether or not to support or participate in the effort.

Library Labor

Given the usefulness of a strong demonstration project with a large and accurate collection of data about the selected faculty unit, we are pleased that labor costs of contributing data are declining. New tools for batch processing Wikidata contributions have been introduced, and the tools that we used for the pilot project continue to be updated and enhanced by their developers. In addition to working on batch processing methods using existing tools from Wikimedia Toolforge, we are also exploring offline approaches (for example, cleaning and preparing metadata with scripting languages, including R and Python). As these tools

reduce the labor costs of contributing large data sets to the site and as the number of contributors increases, the goal of building complete bibliographic data collections for faculty groups and institutions will be more readily accomplished.

In addition to exploring batch processing methods for contributing to Wikidata, we are also hosting Wikidata edit-a-thons and informal sessions for library staff. In these sessions we aim to increase familiarity with Wikidata and to build the capacity of the library to contribute data. To date, these sessions have focused on topics of interest to our colleagues and to our communities, such as adding data about women authors from IUPUI. As we build interest in Wikidata among our colleagues in the library, the capacity for the organization to contribute complete faculty data collections to the site will grow.

Although contributing bibliographic data to Wikidata is not labor-free, the effort aligns with other metadata projects and services supported by libraries. In our experience, the effort involved in creating a Wikidata entry for an author and adding an entry for one of their publications is less onerous than the traditional library task of original cataloging. Furthermore, at IUPUI, the library already curates and enriches bibliographic metadata for faculty works—notably, for the campus-wide open access policy and the open access publishing fund. While these projects do not represent a complete data set of all publications authored by faculty at IUPUI (monographs, for example, are not included in the OA policy), they do include more than 3,000 publications per year, and will accumulate as a representative sample over time. Given that the library already manages these data sets offline and in other web-based systems, much of the labor of preparing the data for Wikidata is already complete. Deploying another researcher information system would require similar work from the library, from another organization on the campus, or as an (outsourced) investment in a proprietary data collection. At our institution, the library is well-positioned to do this work using open source systems, and our pilot project with Wikidata and Scholia will help to demonstrate this capacity.

CONCLUSION

This pilot project aligned with our library's efforts to build and support an open scholarly communication culture at IUPUI. Wikidata, despite being relatively new to the Wikimedia ecosystem, has already shown steady growth in the scholarly data it collects and makes available for anyone to use. At the same time, new tools for displaying data (such as Scholia, and others used for contributing and curating Wikidata entries) are reducing barriers for those that see a value in maintaining scholarly communication data in the commons. The methods that we have described will help libraries establish a similar program for a small school or department. Successfully completing a project like ours will show a library's commitment

to the value of a community-driven, open, linked data environment. As the labor costs of contributing decline, Wikidata is becoming a powerful platform for sharing and discovering institutional bibliographies. All libraries with an interest in linked open data or open scholarly communication systems will benefit from learning more about and contributing to Wikidata.

REFERENCES

- Allison-Cassin, S., & Scott, D. (2018, May 4). Wikidata: A platform for your library's linked open data. *Code4Lib Journal*, 40. Retrieved from <https://journal.code4lib.org/articles/13424>
- Bryant, R., Namachchivaya, B., & Speer, J. (2017). *Supporting Research Information Management in the research university: Partnerships, challenges, and possibilities*. <https://doi.org/10.5703/1288284316488>
- Comparison of research networking tools and research profiling systems. (n.d.). In Wikipedia. Retrieved February 27, 2018, from https://en.wikipedia.org/wiki/Comparison_of_research_networking_tools_and_research_profiling_systems
- Givens, M. (2016, March 15). Keeping up with . . . Research Information Management Systems. Retrieved from http://www.ala.org/acrl/publications/keeping_up_with/rims
- Givens, M., Macklin, L. A., & Mangiafico, P. (2017). Faculty profile systems: New services and roles for libraries. *Portal: Libraries and the Academy*, 17(2), 235–255. <https://doi.org/10.1353/pla.2017.0014>
- Heller, L. (2015, July 16). What will the scholarly profile page of the future look like? Provision of metadata is enabling experimentation. [Blog post]. LSE Impact Blog. Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/16/scholarly-profile-of-the-future/>
- IUPUI. (2018). Vision & mission: About. Retrieved from <https://www.iupui.edu/about/vision-mission.html>
- IUPUI University Library. (2015). Mission & vision statements. Retrieved from <https://ulib.iupui.edu/about/mission>
- Koren, M. (2018, October 2). One Wikipedia page is a metaphor for the Nobel Prize's record with women. Atlantic. Retrieved from <https://www.theatlantic.com/science/archive/2018/10/nobel-prize-physics-donna-strickland-gerard-mourou-arthur-ashkin/571909/>
- Krauss, A. (2017, September 20). Publishers increasingly in control of scholarly infrastructure and this is why we should care. [Blog post]. The Knowledge G.A.P. Retrieved from <http://knowledgegap.org/index.php/sub-projects/rent-seeking-and-financialization-of-the-academic-publishing-industry/preliminary-findings/>

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>

Lemus-Rojas, M., Pintscher, L. (2018). Wikidata and libraries: Facilitating open knowledge. In M. Proffitt (Ed.), *Leveraging Wikipedia: Connecting communities of knowledge* (pp. 143–158). Chicago: ALA Editions.

Lilly Family School of Philanthropy. (n.d.) Founding of the school. Retrieved from <https://philanthropy.iupui.edu/about/founding.html>

Meyer, S. (2018). Enhancing library discovery with linked open data. Retrieved from OSF Code4Lib 2018 talks: <https://osf.io/h2cav/>

Mietchen, D., Hagedorn, G., Willighagen, E., Rico, M., Gómez-Pérez, A., Aibar, E., . . . Kinzler, D. (2015). Enabling open science: Wikidata for Research (Wiki4R). *Research Ideas and Outcomes*, 1, e7573. <https://doi.org/10.3897/rio.1.e7573>

Neubert, J. (2017, September). Wikidata as a linking hub for knowledge organization systems? Integrating an authority mapping into Wikidata and learning lessons for KOS mappings. Proceedings of the 17th European Networked Knowledge Organization Systems Workshop. Thessaloniki, Greece. Retrieved from <http://ceur-ws.org/Vol-1937/paper2.pdf>

Nielsen, F. Å. (2018). Scholia: Wikidata scholarly profile. Retrieved from <https://github.com/fnielsen/scholia>

Nielsen, F.Å., Mietchen, D., & Willighagen, E. (2017). Scholia and scientometrics with Wikidata. https://doi.org/10.1007/978-3-319-70407-4_36

Paling, E. (2015, October 21). Wikipedia's hostility to women. *Atlantic*. Retrieved from <https://www.theatlantic.com/technology/archive/2015/10/how-wikipedia-is-hostile-to-women/411619/>

Shanks, J., & Arlitsch, K. (2016). Making sense of researcher services. *Journal of Library Administration*, 56(3), 295–316. <https://doi.org/10.1080/01930826.2016.1146534>

Statistics - Wikidata. (n.d.). In Wikidata. Retrieved May 15, 2018, from <https://www.wikidata.org/wiki/Special:Statistics>

Wikibase/DataModel/Primer - MediaWiki. (2017). In MediaWiki. Retrieved from <https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>

WikiCite. (n.d.). Retrieved April 23, 2018, from <https://meta.wikimedia.org/wiki/WikiCite>

Wikidata: Data access - Wikidata. (2017). In Wikidata. Retrieved from https://www.wikidata.org/wiki/Wikidata:Data_access

Wikidata: WikiProject Chemistry - Wikidata. (n.d.). In Wikidata. Retrieved April 23, 2018, from https://www.wikidata.org/wiki/Wikidata:WikiProject_Chemistry

Wikidata: WikiProject Source MetaData. (n.d.). Retrieved April 23, 2018, from https://www.wikidata.org/wiki/WikiProject_Source_MetaData

Wikimedia Foundation. Terms of use - Wikimedia Foundation. (n.d.). Retrieved October 5, 2018, from https://foundation.wikimedia.org/wiki/Terms_of_Use/en