

The Intersection of Open Metadata and SALALM: Adding SALALM Presidents and Their Scholarship to Wikidata

Mairelys Lemus-Rojas (IUPUI University Library)

Timothy A. Thompson (Yale University Library)

Introduction

Cultural heritage institutions are rooted in acquiring, preserving, and providing access to information. Sharing data openly beyond the proprietary content management systems used by many of our institutions is of interest to the GLAM (Galleries, Libraries, Archives and Museums) community. GLAM institutions have been increasingly embracing the idea of contributing information to open knowledge projects such as Wikipedia, the online encyclopedia; Wikimedia Commons, media repository for all Wikimedia projects; and Wikidata, a multilingual knowledge base.

Although Wikidata is a fairly new project in the Wikimedia Foundation ecosystem, it has already been embraced by many cultural heritage institutions. These institutions are not only contributing and bringing awareness of the project by way of hosting events in an effort to train library and museum staff, but they are also starting to become more conscientious of the value of sharing their digital collections more openly. Institutions such as the Rijksmuseum, New York Public Library, Metropolitan Museum of Art, National Gallery of Denmark, among others, have taken a step further and adopted an open access policy for sharing digital images and are actively contributing their resources. Their efforts concentrate on sharing digital images under an open license, which allows free use, benefiting open knowledge projects such as Wikimedia Commons (Maher and Tallon 2018). Other institutions have been interested in providing a presence for

underrepresented subjects in projects like Wikipedia and Wikidata. For instance, a pilot project conducted at the University of Miami Libraries in 2014 focused on creating Wikipedia articles for a set of the library's Cuban Theater collections comprised of Cuban and Cuban American playwrights, costume designers, actors, directors and theater companies (Lemus-Rojas and Thompson 2018). More recently, York University in Canada led a national effort called "Music in Canada @ 150" to provide a more balanced representation in Wikipedia and Wikidata for Canadian music (Allison-Cassin and Scott 2018).

Despite the fact that the Wikidata community of editors is steadily making contributions to the project with the goal of building a more comprehensive data set, there is much work in terms of content contribution and data curation that still needs to happen in order to achieve this goal. The SALALM (Seminar on the Acquisition of Latin American Library Materials) community is well positioned to make a significant impact in this regard by ensuring that works related to Latin American studies and their creators are represented in Wikidata. Contributing data to the Wikidata knowledge base enhances discovery for these creators and their scholarship and facilitates the use of data by external tools and projects. In this chapter, we share our experience organizing a Wikidata workshop as part of the SALALM 63 conference taking place in Ciudad de México from July 1-4, 2018. The focus of this workshop was on providing participants with an overview of the Wikidata knowledge base and sharing examples of services and tools developed by the community to query, contribute, and visualize data in Wikidata. The ultimate goal of this event was for participants to create a presence in Wikidata for SALALM presidents and their scholarship.

Wikidata

The Wikidata knowledge base offers a free, community-owned solution to sharing data more openly. Wikidata was developed in 2012 to serve as the centralized repository to be used by all Wikimedia projects, but has since proven to serve a purpose beyond its original intent. It acts as a hub connecting data to external data sources. The structured linked data stored in the knowledge base can be read by both humans and machines, which has inspired many to develop tools to aid with content contribution as well as to display the data in ways that provide users the option to interact with it, offering a more unique experience. Wikidata is multilingual, meaning that data in hundreds of languages is stored in one central location, making it appealing to smaller language communities. Anyone can contribute to Wikidata. The community has over 18,000 active users who are contributing and curating data in the knowledge base (“Wikidata:Statistics” 2018).

SALALM Wikidata Workshop

The Wikidata workshop we facilitated was planned with the idea of raising awareness of Wikidata within the SALALM community. We wanted to make sure to share some background information about Wikidata and explain why it was important to get involved with the project. We specified in the proposal that no previous experience using or contributing to Wikidata was necessary to participate in the workshop since we intended to provide them with the tools they needed to start contributing. The session took place at El Colegio de México on July 1, 2018, the first day of the conference. It was hands-on, focusing on teaching participants how to create entries in Wikidata for those who have served as presidents of SALALM and include at least one of their publications.

Planning the Workshop

As part of the preparation for the workshop, Wikidata items were created for the last two SALALM presidents (2016-2017 and 2017-2018)—an exercise that gave us a better idea of what we needed to prepare for the event. Because information in Wikidata is stored as structured linked data, we needed to make sure that SALALM, as an organization, was present in Wikidata to be able to connect it with the items we were going to create for the presidents. That way, we could make the claim that a particular individual served as SALALM president during a certain period of time.

We put together a spreadsheet in Google Sheets with the names of all SALALM presidents since 1972. In order to gather the information needed about them, we searched for identifiers in VIAF (Virtual International Authority File), Library of Congress Authorities, ISNI (International Standard Name Identifier), Research Gate, Scopus, Google Scholar, ORCID, Twitter, and the Researcher site. We also gathered links to LinkedIn and personal profiles sites hosted by their affiliated institution. Additionally, we included links to either an article or a book chapter written by them. The idea behind doing this work in anticipation of the workshop was to allow participants to focus on getting familiar with Wikidata and how to contribute content to it, rather than spending time searching for information to add.

Another component was the creation of a property guide, in English, Spanish, and Portuguese (Brazilian), for creating items about individuals and works (see Table 1). While doing this work, we identified a few Wikidata properties needing labels in Portuguese and addressed them before the workshop. Creating the property guide in SALALM's official languages provided participants with the option to contribute in their preferred language. In addition, we provided examples of how to use properties such as affiliation, member of, and

position held (marked with an asterisk in Table 1) when creating entries for SALALM

presidents.

English	Spanish	Portuguese (Brazilian)
instance of	instancia de	instância de
sex or gender	sexo o género	sexo ou gênero
given name	nombre de pila	primeiro nome
family name	apellido	sobrenome
country of citizenship	país de nacionalidad	país de cidadania
date of birth	fecha de nacimiento	data de nascimento
place of birth	lugar de nacimiento	local de nascimento
native language	lengua materna	língua materna
languages spoken, written or signed	lenguas habladas, escritas o signadas	línguas faladas
occupation	ocupación	ocupação
employer	empleador	empregador
educated at	estudió en	alma mater
affiliation*	afiliación*	afiliação*
member of*	miembro de*	membro de*
position held*	cargo ocupado*	cargo ocupado*
LinkedIn personal profile URL	perfil LinkedIn	perfil LinkedIn
VIAF ID	identificador VIAF	identificador VIAF
Library of Congress authority ID	identificador autoridades de la Biblioteca del Congreso de EEUU	identificador LCCN
ISNI	ISNI	ISNI
ResearcherID	identificador ResearcherID	identificador ResearcherID

ResearchGate person ID	identificador ResearchGate	identificador ResearchGate
Scopus Author ID	identificador Scopus de autor	identificador de autor Scopus
Google Scholar author ID	Google Scholar author ID	identificador Google Acadêmico
ORCID iD	ORCID	ORCID
Twitter username	nombre de usuario en Twitter	nome de usuário no Twitter

Table 1. List of the properties used for the individual language guide (English, Spanish or Portuguese (Brazilian)) to distribute as needed.

Moreover, we created a registration form using Google Forms to gather information about the workshop participants. In the form, we asked for their names, email, affiliated institution, and Wiki username. For those who did not already have a registered account in a Wiki project, we included a link for them to create one. An e-mail including an abstract for the Wikidata Workshop and the link to the registration form was sent out to the LALA-L List. Once we received confirmation from the participants, we created an event page on the Wikimedia dashboard and included the usernames shared with us so that we could track event contributions.

Data Gathering Analysis

The list of SALALM presidents from 1972 to 2018 comprised 46 names. Based on the confirmed workshop participants, and in case others were to join last minute, we selected 10 names for inclusion to Wikidata. Searching for what we identified as most commonly found identifiers for researchers, we found that representation varied on these sites (see Table 2). For instance, eight out of the ten selected names had entries in VIAF, and seven had a Library of Congress authority record and a LinkedIn personal profile. Six had an entry in ISNI, and only four were present in Scopus. Profiles in Google Scholar and ResearchGate were only found for

two individuals, while the ORCID and Researcher site only had one profile. Lastly, only three twitter handles were identified. We speculate that the unbalanced representation of data on the various sites searched could be the result of not being able to match the profile with the individual due to lack of information.

Number of SALALM Presidents	46
Group selected for Workshop	10
VIAF ID	8
Library of Congress authority ID	7
ISNI	6
ResearcherID	1
ResearchGate person ID	2
Scopus Author ID	4
Google Scholar author ID	2
ORCID iD	1
Twitter username	3
LinkedIn personal profile URL	7

Table 2. SALALM presidents data gathering results

Demographics and Outcomes

Workshop participants came from institutions such as Brigham Young University, Harvard University, St. Cloud State University, University of Pittsburgh and the Universidad del Claustro de Sor Juana. The gender gap among Wiki contributors is a serious concern. We were encouraged to see that four of our six participants were female. The total edit count of the contributions made during the workshop was 187 edits. Although this is not a significant number, it certainly is a step in the right direction, and we hope to continue bringing more SALALM members and their scholarship to Wikidata. Contributions, including the pre-work we did in anticipation of the event, resulted in nine new Wikidata entries for SALALM presidents and the enhancement of an existing one (see Figure 1). In addition, seven items for publications produced by seven SALALM presidents were also created and properly linked to the item for the creator (see Figure 2).

Daisy V. Domínguez librarian Q47800459	Molly Molloy librarian Q55318630	Mark L. Grover American historian Q6768449 Wikipedia article	Suzanne M. Schadl curator Q55155159
Paula Covington Librarian Q55318628	Angela Carreño librarian Q55318617	Dan Hazen librarian Q55318643	Paloma Celis-Carbajal librarian Q55318878
Martha Mantilla librarian Q55318619	Nerea A. Llamas librarian Q55339851		

Figure 1. SALALM presidents search result using Wikidata VizQuery¹

¹ Wikidata VizQuery: a tool that allows users to search and retrieve data from Wikidata <http://tools.wmflabs.org/hay/vizquery/>

Publication date	Work	Researchers
2017-09-13	Many Languages, Many Workflows: Mapping and Analyzing Technical Services Processes for East Asian and International Studies Materials	Nerea A. Llamas
2014-09-30	American Indians in feature films: beyond the big screen	Daisy V. Domínguez
2011-04-22	Reference as Outreach: Meeting Users Where They Are	Suzanne M. Schadl
2007-07-25	The Cooperative Conundrum in the Digital Age	Dan Hazen
2005-12-07	The Internet in Latin America	Molly Molloy
2000-02-01	The Politics of Linking Educational Research, Policy, and Practice: The Case of Improving Educational Quality in Ghana, Guatemala and Mali	Martha Mantilla
1999-01-01	Large scale collection assessment	Mark L. Grover

Figure 2. List of works by SALALM presidents generated by Scholia

Using External Tools to Visualize the Data

There are a variety of external tools created by Wiki contributors to help visualize Wikidata data. Two of these tools that are appropriate for the type of data we were contributing to Wikidata were Scholia and Reasonator.

Scholia, a web-based service tool, makes live SPARQL queries to Wikidata and uses the results to generate scholarly profiles. It can be accessed at <https://tools.wmflabs.org/scholia>. Scholia displays all the data that Wikidata holds in relation to the person, in this case, and it also brings in other relations such as coauthors. The scholarly profile generated by Scholia provides all the information in relation to the individual, the scholarship produced, and their connection with other researchers. Using this tool to browse Wikidata data facilitates the searching of authors and their scholarship. It is also possible to generate a profile for the organization, allowing users to see the interactions among scholars. Since only seven works were added as part of the workshop, and no entries were added for coauthors, it is not possible at this point to analyze the scholarship produced by the SALALM presidents. As we continue adding more works, we will be able to make connections that will facilitate the rendering of coauthorship

graphs. At a minimum, we can now see a list generated by Scholia, with seven works by individuals affiliated to SALAM where none were available before we started this initiative (see Figure 2).

Reasonator is a tool that tries to make sense of Wikidata data through simple reasoning (“Reasonator” 2018). Figure 3 shows a snippet of the profile of one SALALM president where we can see employment history, languages spoken, written or signed, affiliation, LinkedIn personal profile URL, position held (where the service to SALALM was included), membership information, as well as unique identifiers for Google Scholar, Scopus, Twitter and LibraryThing.

Daisy V. Domínguez (Q47800459)

Daisy Domínguez
librarian

Daisy V. Domínguez is a librarian.

She studied at City College of New York until 2010, Long Island University until 2005, and New York University until 1997. She is/was president from 2016 until 2017 for Seminar on the Acquisition of Latin American Library Materials. She is/was a member of Seminar on the Acquisition of Latin American Library Materials. She worked for City College of New York.

Other properties	
employer	City College of New York senior college of the City University of New York (CUNY) in New York City
languages spoken, written or signed	English West Germanic language originating in England with linguistic roots in French, German and Vulgar Latin
affiliation	Seminar on the Acquisition of Latin American Library Materials
LinkedIn personal profile URL	https://www.linkedin.com/in/daisy-v-domínguez-a379b412/
position held	president non-political leader of an organization, company, community, club, trade union, university or other group of : Seminar on the Acquisition of Latin American Library Materials start time : 2016 end time : 2017
member of	Seminar on the Acquisition of Latin American Library Materials
educated at	City College of New York senior college of the City University of New York (CUNY) in New York City end time : 2010 academic degree : Master of Arts academic major : history
	Long Island University university in Brooklyn, New York, United States end time : 2005 academic major : library science
	New York University private research university in New York, NY, United States end time : 1997 academic degree : Bachelor of Arts academic major : Latin American studies

Free images Google search

External sources	
Google Scholar author	QDrpYq8AAAAJ
Scopus Author	39961270500
Twitter username	daisilla

Social media

LibraryThing	daisilla
--------------	----------

Wikimedia projects

Concept cloud

Figure 3. Example of SALALM president profile generated by Reasonator

The completeness of the profile depends on the information that is stored in Wikidata for a particular individual. In this case, there is also information on educational institutions attended, attributed work(s), a timeline, and related media. The full profile can be accessed at:

<https://tools.wmflabs.org/reasonator/?&q=47800459>.

Conclusion

Facilitating a Wikidata Workshop as part of the SALALM conference served as an opportunity to bring awareness of the project and to share with the community the importance of contributing to it. We hope to have inspired participants to use their subject expertise and continue contributing content to the knowledge base after the event. Adding entries for SALALM presidents and their works in Wikidata will ensure a more balanced representation of the scholarship produced by the SALALM community. In general, working with open platforms such as Wikidata helps to break down data silos and to open up our data to a wider community of users.

References

Allison-Cassin, Stacy, and Dan Scott. "Wikidata: a platform for your library's linked open data."

The Code4Lib Journal, 2018. <https://journal.code4lib.org/articles/13424>

Katherine Maher and Loic Tallon, "Wikimedia and the Met: A Shared Digital Vision,"

The MET (blog), April 19, 2018. <https://www.metmuseum.org/blogs/now-at-the-met/2018/wikimedia-and-the-met-digital-vision>

Lemus-Rojas, Mairelys, and Timothy A. Thompson. "Bringing archival collections to Wikipedia with the Remixing Archival Metadata Project (RAMP) editor." In *Leveraging Wikipedia: Connecting Communities of Knowledge*, edited by Merilee Proffitt, 143-158. Chicago, IL: ALA Editions, 2018

Preprint of an article that will be published in the Papers of the Sixty-Third Annual Meeting of the Seminar on the Acquisition of Latin American Library Materials (SALALM)

"Reasonator." *GeoHack - Ujjain*. Accessed August 01, 2018.

<https://tools.wmflabs.org/reasonator/>

“Wikidata:Statistics.” *Wikidata, 2018*. Accessed August 25, 2018.

<https://www.wikidata.org/wiki/Wikidata:Statistics>