CHAPTER 17

STRUCTURING BIBLIOGRAPHIC REFERENCES: TAKING THE JOURNAL *ANAIS DO MUSEU PAULISTA* TO WIKIDATA

Éder Porto Ferreira Alves,[1] *Paul R. Burley,*[2]
and João Alexandre Peschanski[3]

[1] Grupo de Usuários Wiki Movimento Brasil, [2] Northwestern University, [3] Faculdade Cásper Líbero

Abstract

This chapter provides a step-by-step process for large-scale contributions of articles from scholarly publications to Wikidata, a collaborative data store project of the Wikimedia Foundation. Tools and processes in Wikidata, Zotero, and Google Sheets in particular are described; they relate both to the Wikidata platform and standard spreadsheet programs. The case of the Brazilian journal Anais do Museu Paulista is used to illustrate the process that can then be replicated with other publications and in other contexts.

DOI: https://doi.org/10.3998/mpub.11778416.ch17.en

Correspondence concerning this article should be addressed to João Alexandre Peschanski, Av. Paulista, 900, Bela Vista, São Paulo, SP, 01310–100 Brazil. Email: japeschanski@casperlibero.edu.br

Introduction

Wikimedia-supported knowledge projects have seen robust acceptance in the Global South, notably in Brazil. Wikimedia projects are attractive in Global South communities for decentering the use of the English language and the low cost of use and access. In contrast, institutional repository and data platforms are prohibitively expensive, either in cost or maintenance, for Global South galleries, libraries, archives, and museums (GLAMs). These tools additionally require technology out of the reach of Global South institutions. Resources for "community-owned infrastructure, and robust metadata to facilitate open scholarship practices" (ARL Task Force on Wikimedia and Linked Open Data, 2019) have grown in both size and depth in Brazil and other Global South communities. In contrast, the high cost of computer hardware, software, and Internet connectivity in Global South GLAMs, as well as a lack of technology specialists, is a constraint unlikely to change at present or in the near future. Internet connectivity in Brazil is limited to 74 percent in museums and 66 percent in libraries (Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação, 2018). Some of these infrastructure and technological obstacles are addressed by demonstrating the use of Wikidata to index items of scholarly articles in the Brazilian context.

Sources on Wikimedia Projects
Referencing reliable sources is an essential component of a Wikipedia article (Orlowitz, 2018), yet the quality of referencing has varied across Wikipedia instances in different languages (Lewoniewski et al., 2020). Some progress has been made in citation inclusion, but in general the

addition of references to a Wikipedia article remains an excessively technical endeavor. An instance of a reference is limited to its article and cannot be shared among articles; more importantly, it is difficult to move a reference between different-language Wikipedias, even with the support of translation tools.

Wikidata is a free, collaborative knowledge base (Vrandečić & Krötzsch, 2014), and it can be used to overcome existing inefficiencies in the Wikipedia referencing model. The creation of items for scholarly articles and other periodical literature for a Wikimedia project is a useful contribution in its own right. A single, simple recent improvement to the article citation creation process in Wikipedia is the ability to create a reference based on a Wikidata identifier (QID), which can then be reused independently across different-language Wikipedias. A more advanced utility, not yet fully developed, would be a system in which article citations generated from Wikidata items are automatically updated or improved when the corresponding Wikidata item is enhanced.

Figure 1 highlights the imbalance of items for scholarly articles from North America and Europe in Wikidata in comparison to those elsewhere. This chapter specifically aims to motivate editors from the Global
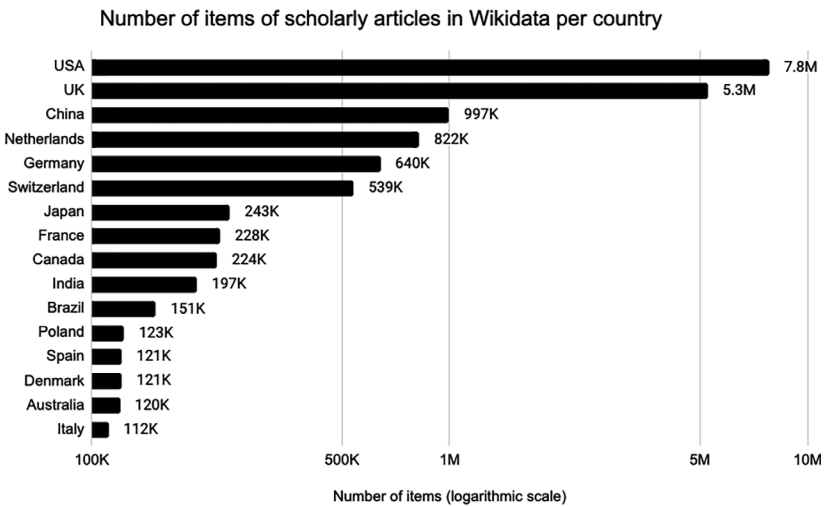
Figure 1 Number of items for scholarly articles in Wikidata by country of publication. Retrieved on September 26, 2020. Data available at https://doi.org/10.5281/zenodo.4051056.

South and underrepresented communities to engage in the large-scale contribution of items for scholarly articles and related academic literature to Wikidata. For these editors, tools and processes that are commonly used in resource-richer communities are often not helpful, as they depend on data that are already well-structured and easy to automatically feed to Wikidata. Commercial journal platforms and common metadata structures are well established in the Global North; in contrast, the situation in Brazil and elsewhere is ad-hoc. The process we present in this chapter is more easily replicable in Global South contexts.

Context

The edit-a-thon emerged among Wikimedia knowledge communities to increase content and depth of a subject area of common importance to its members or provide instruction on Wikimedia tools or practices. As the portmanteau of "edit" and "marathon" suggests, they are "in-person or virtual events where Wikimedia community members write or enhance Wikipedia articles, upload or edit metadata for images in Wikimedia Commons, add or enhance structured data in Wikidata, or other Wikimedia-related knowledge project activities" (ARL Task Force on Wikimedia and Linked Open Data, 2019). Edit-a-thons at Stanford University, Indiana University-Purdue, University of Indianapolis, and Laurentian University are described in the literature (Allison-Cassin & Scott, 2018; Keller et al., 2011; Lemus-Rojas & Pintscher, 2018). Descriptions of edit-a-thons outside the Global North in scholarly literature, however, are lacking.

The activities of Grupo de Usuários Wiki Movimento Brasil (English: Wiki Movement Brazil User Group), a national-level Wikimedia umbrella organization in Brazil, range from group editing projects to instruction on advanced tools. The user group has organized two distinct activity modalities: edit-a-thons, known as *maratonas de edição* or *editatonas* in Brazilian Portuguese, and Wikidata Labs. In contrast to the edit-a-thon, the Wikidata Lab emerged in 2017 as periodic events to share resources and capacities for the integration of Wikidata into other Wikimedia projects, especially Wikipedia. This series of events was awarded the 2019 WikidataCon Award, in the Category Outreach. Wikidata Labs evolved into connecting Wikimedians to GLAMs and set up a space for working directly on their collections.

The monthly events of the group, sustained activities, and tight community of practitioners of Wiki Movimento Brasil led to a large-scale project focused on content related to the Ipiranga Museum, commonly known as the Museu Paulista. The Museu Paulista opened in the late nineteenth century in a monumental building by the Italian architect Tommaso Gaudenzio Bezzi. The structure commemorated the independence of Brazil and its early collections emphasized the natural history of the country. The long directorship of the historian Afonso d'Escragnolle Taunay reoriented its collections to emphasize the independence movements and establishment of the federal republic of Brazil, the history of the state of São Paulo, and historical and cultural objects of the early twentieth century. The museum was integrated into the University of São Paulo in 1963. Bezzi's museum structure is now a federally protected monument in its own right. Activities of the museum include the maintenance and exhibition of its physical collection and support of research related to its activities. The museum closed in 2013 due to financial problems and is expected to reopen in 2022.

Wiki Movimento Brasil and Museu Paulista partnered on April 4, 2020, to organize Wikidata Lab XXI: Structuring Bibliographic References. The first part of the workshop was a webinar on how to automate the creation of Wikidata items of scholarly articles. The second part of the workshop was a collective effort among attendees to index articles from the primary journal of the museum, *Anais do Museu Paulista* (English: "Annals of the Paulista Museum"), into Wikidata. Twenty-six editors participated in the project, and their work resulted in creating 876 items: 511 for journal articles and 365 for authors. Ultimately, approximately 31,000 statements were added to Wikidata.

The *Anais* is a scholarly journal published by the Museu Paulista since 1922. It serves not only to disseminate scholarship on the museum itself, but it is also one of the primary history and museology journals in Brazil. The *Anais* dates to the early period of Taunay's directorship of the museum and reflects his focus on the formation of the Brazilian nation. Taunay's influence is mirrored in the journal's early subtitle, "Organ of the History of Brazil Section, and Especially of São Paulo, of the Paulista Museum" (Bittencourt, 2012). It draws heavily

on contributions from academics associated with the museum and the University of São Paulo. Journal articles are, with few exceptions, by Brazilian scholars in Brazilian Portuguese. The editorial emphasis on the arts and history of Brazil continues to the present in the publication, which has a stated editorial objective to "discuss . . . themes related to material culture as a mediator of social practices, as well as innovative approaches to historical and museological processes." A new series of the journal was launched in 1993 (see figure 2).
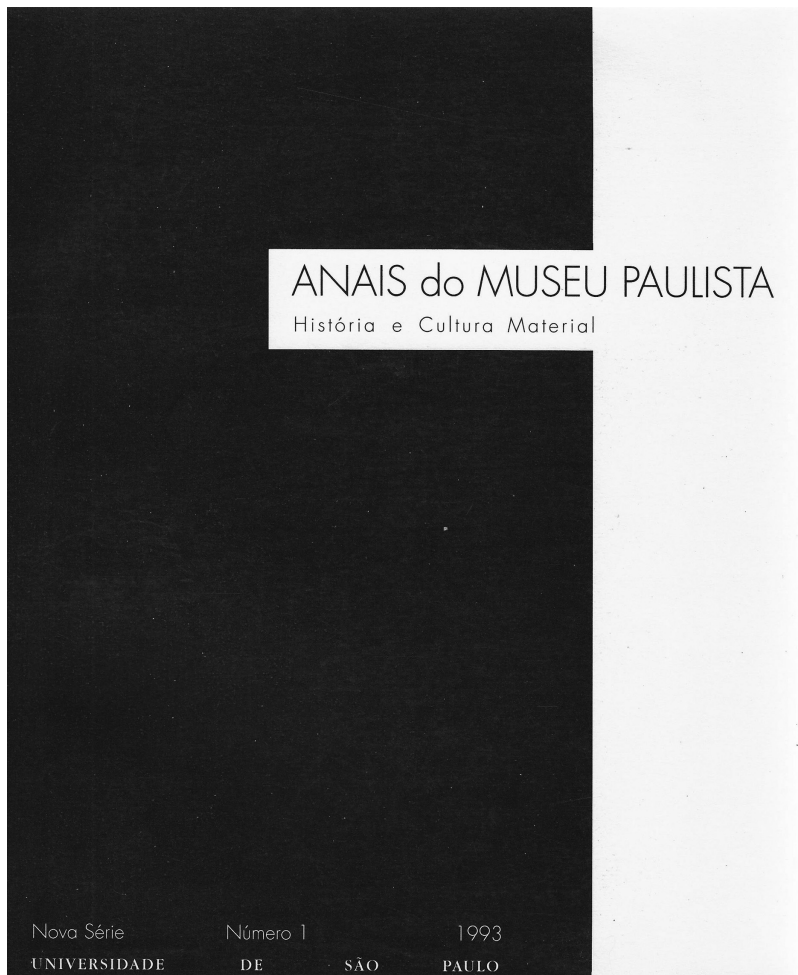
Figure 2 Cover of the first issue of the New Series of *Anais do Museu Paulista*, 1993.

Importing Scholarly Articles into Wikidata via Zotero: Step-by-Step Process

The eight steps below provide a detailed description a process to ingest scholarly articles to Wikidata using Zotero, a bibliographic citation management software. The process is the same used to ingest a body of articles of the *Anais do Museu Paulista* into Wikidata.

Step 1: Download Zotero

In this step, we use Zotero, an open-source software to manage bibliographic data. Zotero Desktop is the desktop version of the tool and Zotero Connector is an extension for browsers to save online references to Zotero. The combination of the two tools aims to make bibliographic references fully integrated, interoperable, and synchronized. These tools are available to download at www.zotero.org/download/. Tutorials for how to work with Zotero are available at www.zotero.org/support/.

Step 2: Import a Set of Articles into Zotero

Begin by creating a list of identifiers and use the "Add Item by Identifier" button on Zotero Desktop to import them all at once (see top of figure 3). As of 2020, ISBNs, DOIs, PMIDs, and arXIv IDs are the only supported identifiers. Alternatively, add identifiers manually using the Zotero Connector extension in the browser by clicking on the "Save to Zotero" button at the top right corner of the article web page (see bottom of figure 3).

The challenge in this step is to decide on which alternative is less time-consuming or skill dependent: to produce a list of identifiers in a text file or spreadsheet or to add them one by one. If an article does not have any of the four identifiers, rely on Zotero Connector to add the article to the library.

Step 3: Create an Account on Wikimedia Projects

Users are not required to create an account or login in order to edit Wikipedia, Wikidata, or Wikimedia Commons. However, an account and
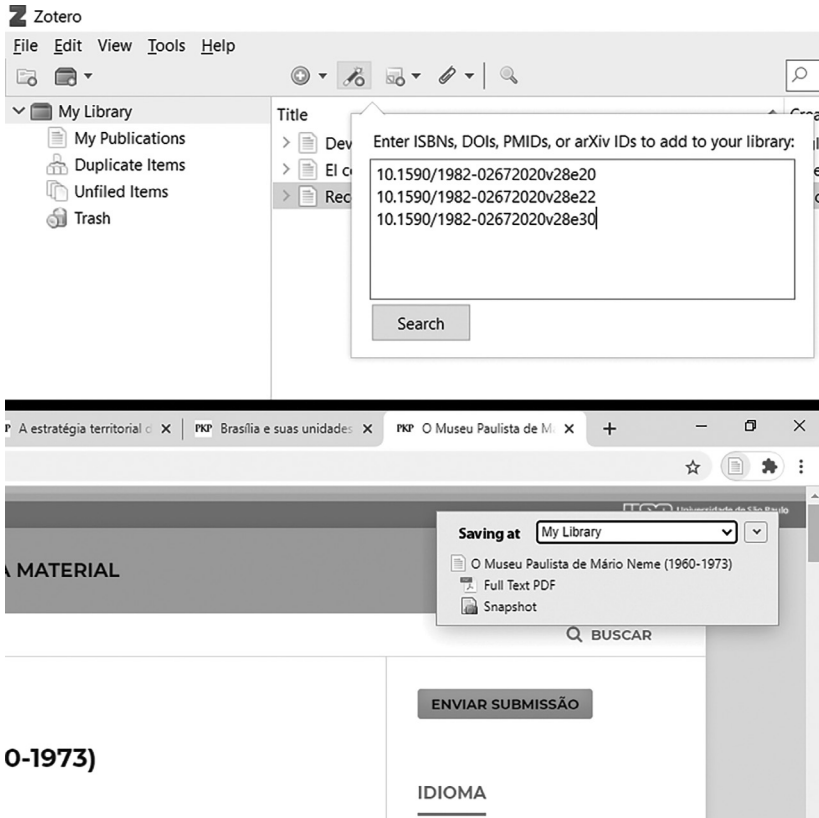
Figure 3 From top to bottom: screenshots of Zotero Desktop and Zotero Connector, showing usage of the "Add Item(s) by Identifier" and "Save to Zotero" buttons, respectively. Retrieved on September, 4, 2020.

username are required, for any functionality beyond basic edits. Additionally, Wikidata user access levels only allow experienced users to batch import datasets (Wikidata contributors, 2020). To access this functionality:

- Create an account by following the instructions at www.wikidata. org/wiki/Special:CreateAccount, and
- Make fifty valid edits or more and wait at least four days. This will establish you as an autoconfirmed user, and it will also allow you to use more advanced tools on Wikidata.

*Step 4: Download, Install, and Set Up the QuickStatements Translator*

Zotero and Wikidata are integrated with the use of a translator, a script that converts metadata stored in one format to another. The translator, which is called *Wikidata QuickStatements.js* and is part of a Zotero extension called *zotkat*, will generate a set of text commands to be imported into QuickStatements, an online tool program that can read and execute commands to create or edit a Wikidata item (Wikidata contributors, 2021). For this step:

- Download Wikidata QuickStatements.js file at https://github. com/UB-Mannheim/zotkat, and
- Paste the Wikidata QuickStatements.js file into the "translators" folder (see figure 4) of your Zotero installation.
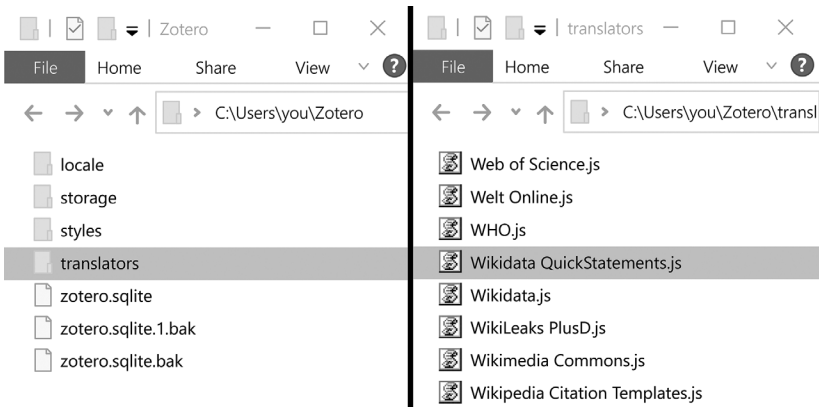
Figure 4 From left to right, screenshots of Zotero folder and translators folder. Retrieved September 9, 2020.

After this, restart Zotero and open "Program Preferences" under the Edit menu. To complete this step, highlight the "Export" tab, and select "Wikidata QuickStatements" under the "Default Output Format" dropped down menu, as indicated in figure 5.
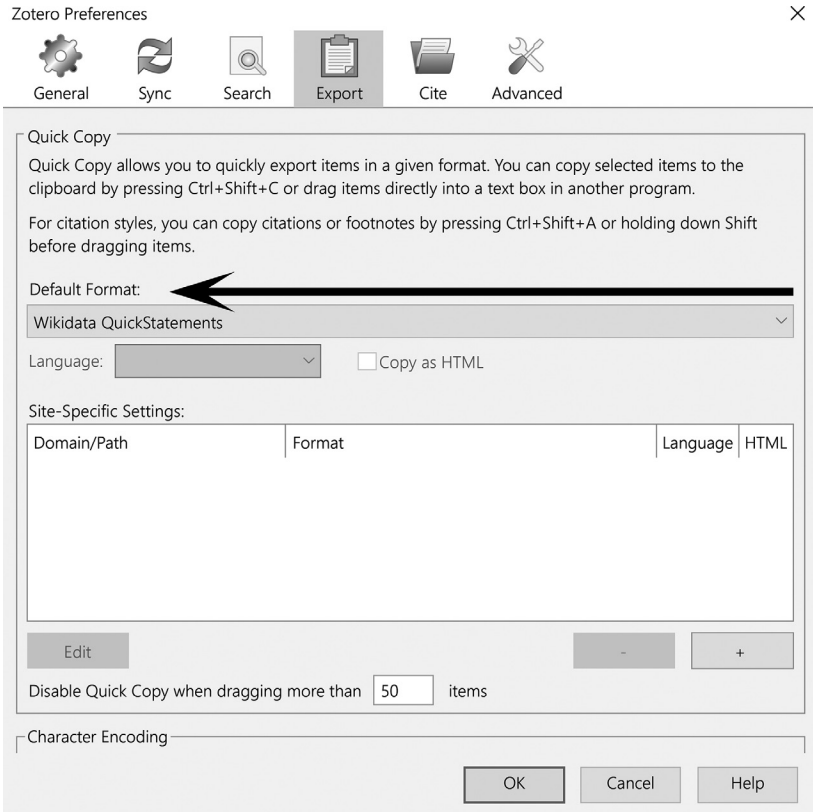
Figure 5 Screenshots of Zotero Preferences window, showing the Wikidata QuickStatements format chosen under "Default Format." Retrieved September 9, 2020.

Step 5: Check for Duplicates in Wikidata

It is important to not create duplicate items in Wikidata, so you will first need to check for your journal's articles in Wikidata before uploading new items. There are several strategies to check for duplicates. One strategy is to query Wikidata for items for scholarly articles from the journal on which you are working using Wikidata Query Service, a user-friendly interface to build and run data queries on Wikidata, providing an overview of the statements of a set

of items. An example of a query of scholarly articles from *Anais do Museu Paulista* can be found at https://w.wiki/bYF. To run a similar query for another journal, switch the Brazilian journal Wikidata QID (Q50426299) in row 11, as shown in figure 6, to the Wikidata item QID of your target journals.


```
   Wikidata Query Service    ⮞ Examples   ❓ Help  ▾    ⚙ More tools  ▾   🈂 English
 1  SELECT DISTINCT
 2  ?Article ?ArticleLabel
 3  (GROUP_CONCAT(?AuthorLabel;SEPARATOR="; ") AS ?Authors)
 4  ?DOI ?ISBN10 ?ISBN13 ?PMID ?arXIv
 5  WHERE {
 6     SERVICE wikibase:label {
 7        bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,pt-br,pt,es,fr,de,ja".
 8        ?Article rdfs:label ?ArticleLabel.
 9        ?Author rdfs:label ?AuthorLabel.}
10
11     ?Article wdt:P1433 wd:Q50426299.
12     OPTIONAL {?Article wdt:P50 ?Author.}
13     OPTIONAL {?Article wdt:P356 ?DOI.}
14     OPTIONAL {?Article wdt:P212 ?ISBN13.}
15     OPTIONAL {?Article wdt:P957 ?ISBN10.}
16     OPTIONAL {?Article wdt:P698 ?PMID.}
17     OPTIONAL {?Article wdt:P818 ?arXIv.}
18  } GROUP BY ?Article ?ArticleLabel ?DOI ?ISBN10 ?ISBN13 ?PMID ?arXIv
```

Figure 6 Query commands to generate a list of identifiers of the articles from *Anais do Museu Paulista*. Retrieved September 11, 2020.

Once you have performed a query, the data can be downloaded in several formats and compared to the list in Zotero. To do this:

- Download the query result in. csv format and open it into a spreadsheet software,
- Import your list of identifiers into another sheet of the same spreadsheet,
- Write a MATCH function to compare and match the identifiers from your list with the identifiers from the query result; for an example of a match function on Google Sheets, see https://support.google.com/docs/answer/3093378, and
- Exclude the articles with a match from your Zotero library.

Step 6: Upload to Wikidata via QuickStatements
To upload references to Wikidata, use the QuickStatements tool, mentioned before and available at https://quickstatements.toolforge.org/ along with Zotero. Begin in Zotero by:

- Selecting all the references you have imported and want to upload to Wikidata, and
- Clicking on "Edit" then "Copy as Wikidata QuickStatements" to copy the commands to the clipboard.

Then in the QuickStatements tool:

- Click on "Log in" at the right top corner to log in to your account,
- Click on "New batch" and paste the commands into the text field, and
- Click on "Import V1 commands." Look over the first articles to be uploaded, per Wikidata community guidelines, before clicking on "Run."

It might take some time for QuickStatements to upload the references to Wikidata, depending on the number of commands and the Quick-Statements server itself. Help on how to use QuickStatements is available at www.wikidata.org/wiki/Help:QuickStatements. Beware of the following possible mistakes and pitfalls: if the website of origin of your reference is not well structured, Zotero might import the items with duplicated information; Zotkat translator is not fully developed yet and some fields are not translated into QuickStatements commands, for example, the license; some properties present in Zotero do not yet exist in Wikidata and therefore are not imported.

Step 7: Check Completeness of Item Properties Using Wikidata Query Service
To check the completeness of data imported into Wikidata as part of a project, build a dashboard for Wikidata properties. An example of a dashboard for the *Anais do Museu Paulista* was generated using the

Wikidata Query Service at https://tinyurl.com/articles-by-journal-qid. In order to replicate the dashboard for a different journal using the query from the *Anais do Museu Paulista*, change the journal QID at row 39 of the query.

The zotkat tool does not import the "journal of publication" statement to Wikidata. The items created, therefore, do not initially have this declaration on Wikidata. In order to monitor them, build a query using their associated unique identifiers. An example query for a subset of *Anais do Museu Paulista* articles is available at https://tinyurl.com/articles-by-identifiers. To replicate the process for a different set of articles, substitute their identifiers at row 42 of the query.

Step 8: Add the "Journal of Publication" Statement using PetScan
To add the "journal of publication" statement to the items created, use PetScan (https://petscan.wmflabs.org), a tool that lists items based on a query code and can add statements to them in Wikidata. To do this:

- Add the identifiers of the articles created at line 42 of the query at https://tinyurl.com/articles-by-identifiers,
- Copy the query code and paste it into the SPARQL field at the "Other Sources" tab in PetScan (see the top half of figure 7),
- Click "Do it!" and wait until the results show up,
- At the beginning of the results list (see the bottom half of figure 7), include the journal of publication statement that needs large-scale editing at the top right corner. An example for including the journal of publication for scholarly articles of Anais do Museu Paulista is P1433 (property for "published in") and Q50426299 (item for "Anais do Museu Paulista"), and
- Click "Start QS" to open a Quick Statements window with the commands and follow the instructions as indicated before in step 6.

Figure 7 From top to bottom, screenshots of the Pet Scan tool, highlighting the SPARQL and Command list fields, and the results list. Retrieved from https://tinyurl.com/petscan-anaismp26092020.

Data Visualization

Bibliographic visualization is an important outcome of an initiative to create items on scholarly articles in Wikidata (see figure 8). Scholia is a web service to create scholarly profiles and their associated visualizations based on bibliographic information in Wikidata (https://scholia.toolforge.org/) (Nielsen et al., 2017). Profiles may be built for individuals, organizations, works, locations, events, awards, and research topics. In the case of academic journals, it displays lists of publications and research topics as well as information on authors and citations. Displays are built
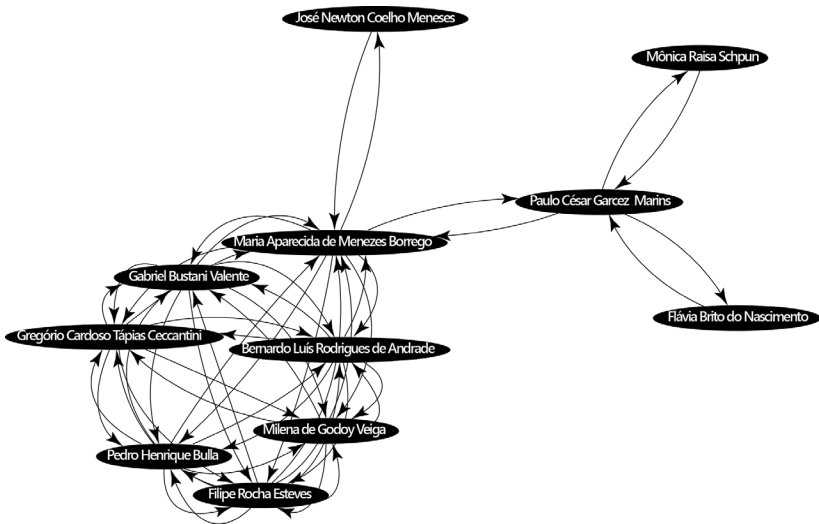
Figure 8 The coauthor graph—here, a fragment from Anais do Museu Paulista—is an output of the Scholia tool for publications. Available at https://w.wiki/XEw.

with the Wikidata Query Service; accordingly, information is updated in Scholia as items in Wikidata are added and enhanced. Scholia provides examples and a menu on the top to browse possible profiles. The profile of *Anais do Museu Paulista* may be found at https://scholia.toolforge.org/venue/Q50426299. Other bibliographic elements may be explored by changing the QID ("Q50426299") at the end of the previous URL.

Conclusion

The process laid out in this chapter demonstrates a strategy for how to democratize the contribution of scholarly literature to Wikidata and to facilitate the diversity of cultural and regional origin of the literature that is contributed to the project. We relied on Zotero and a set of Wikidata tools to provide a large-scale body of article indexing from the *Anais do Museu Paulista*, a scholarly journal of the Museu do Ipiranga. Our case study and related guidelines seek to motivate and illustrate the inclusion of bibliographic literature from the Global South without an advanced skill set, an information-heavy technology,

or a well-structured dataset. The combination of tools and processes presented above forms a catalyst to promote knowledge equity in Wikidata, and in a wider context, bibliographic discovery and access.

References

Allison-Cassin, S., & Scott, D. (2018). Wikidata: A platform for your library's linked open data. *The Code4lib Journal*, *40*. https://journal.code4lib.org/articles/13424.

Association of Research Libraries Task Force on Wikimedia and Linked Open Data. (2019). ARL white paper on Wikidata: Opportunities and recommendations (p. 59) [White paper]. *Association of Research Libraries*. www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf.

Bittencourt, V. L. N. (2012). Revista do Museu Paulista e(m) capas: Identidade e representação institucional em texto e imagem [Revista do Museu Paulista in its covers: Identity and institutional representation through text and images]. *Anais do Museu Paulista*, *20*(2), 149–54. http://dx.doi.org/10.1590/S0101-47142012000200006.

Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação. (2018). TIC Cultura: Pesquisa sobre o Uso das Tecnologias de Informação e Comunicação nos Equipamentos Culturais Brasileiros [TIC Cultura: Research on the use of information and communication technologies in Brazilian Cultural Institutions] (p. 4) [Report] *Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação*. www.cetic.br/media/analises/lancamento-pesquisa-tic-cultura-2018.pdf.

Keller, M. A., Persons, J., Glaser, H., & Calter, M. (2011). Report of the Stanford Linked Data Workshop, 27 June-1 July 2011. *Council on Library and Information Resources*. https://www.clir.org/pubs/reports/pub152/stanford-linked-data-workshop/

Lemus-Rojas, M., & Pintscher, L. (2018). Wikidata and libraries: Facilitating open knowledge. In Proffitt, M. (Ed.), *Leveraging Wikipedia: Connecting communities of knowledge* (pp. 143–58). ALA Editions.

Lewoniewski, W., Węcel, K., & Abramowicz, W. (2020). Modeling popularity and reliability of sources in multilingual Wikipedia. *Information*, *11*(5), 263.

Meneses, U. T. B. de. (Ed). (1993). Anais do Museu Paulista: História e Cultura Material, (1993). *1*(1). https://www.revistas.usp.br/anaismp/issue/view/380

Nielsen, F. Å., Mietchen, D., & Willighagen, E. (2017, May 28). Scholia and scientometrics with Wikidata. *Joint Proceedings of the 1st International Workshop on Scientometrics and 1st International Workshop on Enabling Decentralised Scholarly Communication. 1st International Workshop on Scientometrics and 1st International Workshop on Enabling Decentralised Scholarly Communication*, Portorož, Slovenia. https://doi.org/10.5281/ ZENODO.1036595.

Orlowitz, J. (2018). The Wikipedia Library: The largest encyclopedia needs a digital library and we are building it. In Proffitt, M. (Ed.), *Leveraging Wikipedia: Connecting communities of knowledge* (pp. 1–25). ALA Editions.

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, *57*(10), 78–85. https://doi.org/ 10.1145/2629489.

Wikidata contributors. (2020, April 27). Wikidata:User access levels. In *Wikidata*. Retrieved January 20, 2021, from www.wikidata.org/w/index. php?title=Wikidata:User_access_levels&oldid=1167897989.

Wikidata contributors. (2021, January 13). Help:QuickStatements. In *Wikidata*. Retrieved January 20, 2021, from www.wikidata.org/w/index. php?title=Help:QuickStatements&oldid=1340541986.