

Adding humanities and social science publications to Wikidata: a contribution to the Open Science movement

Mazen El Makkouk*, Cláudia De Souza**, Carlos Suárez Balseiro***

* mazen.el@upr.edu

<https://orcid.org/0009-0007-1078-7352>

College of Communication and Information. University of Puerto Rico-Rio Piedras Campus,
Puerto Rico

** claudia.souza@upr.edu

<https://orcid.org/0000-0002-4168-9399>

College of Communication and Information. University of Puerto Rico-Rio Piedras Campus,
Puerto Rico

*** carlos.suarez5@upr.edu

<https://orcid.org/0000-0002-3908-9618>

College of Communication and Information. University of Puerto Rico-Rio Piedras Campus,
Puerto Rico

Abstract: This is a case study which describes and evaluates the process of creating Wikidata items for academic journals and articles published by the University of Puerto Rico in the humanities and social sciences, as part of an open science initiative. It suggests preliminary steps that can be taken to improve indexing and to register local geographic, disciplinary, and cultural contexts for humanities and social science journals on Wikidata. Finally, argues that these steps can promote a different way of evaluating scholarship in these fields, more appropriate to the humanities and social sciences but nevertheless promoting scholarly cooperation in a spirit consistent with the ethos of open science as perceived by the author.

Keywords: electronic information resources; editing; humanities; case studies

1. Introduction

In the opening paragraph of the call to papers for the 27th International Conference on Science, Technology and Innovation Indicators (STI 2023), the transition towards open science is presented as part of a larger “transition towards a healthier academic culture.” Probably the most jarring difference between open science and the established academic culture is the issue of evaluation, and the relationship between publication and author prestige. Perhaps this single fact makes open science appear at the head of a democratizing movement in the academy.

The wider picture is less clear. The possibility of collaboration through open science, as a democratizing trend, is certainly promising and alluring, even if the democratic nature of the conditions through which open science practices have become established are certainly contested (Mirowski 2018). The cultural momentum to collaborate may have grown out of economic

conditions, and these same conditions may have set the stage for certain types of collaboration and not others.

Whatever the reason or conditions of its rise and current momentum, it is certainly the case that the social good of open science is currently conceived very heavily through the lens of a certain kind of scientific collaboration. One downside to this is that the humanities are left out conceptually from open science discourse (such as the FAIR Data Principles - Findable, Accessible, Interoperable, and Reusable, or the Horizon 2020 policy statement), which creates limited and uncertain ground for thinking about the good of an open humanities and for their practical implementation.

This study begins from this uncertain ground, encountered when the author, as part of a graduation course requirement for completing a Master of Information (MIS) degree, was given the task of adding journals and articles published by the University of Puerto Rico to Wikidata. This task was given as part of a university initiative carried out by the Unit for Monitoring and Analysis of Scientific Research in Puerto Rico (UMAIPUR 2023), to improve the availability of academic research published in Puerto Rico and to enhance the visibility and complement the openness of scholarly exchange between Puerto Rico and its immediate contexts, especially the Caribbean (De Souza & Balseiro, 2022).

The online home for these journals, the University of Puerto Rico Academic Journals Portal (*Portal de Revistas Académicas de la Universidad de Puerto Rico*), was set up with the explicit intention of making research available to open science. However, its language at the same time betrays the uncertainty of how to conceive most of the journals of the site, which belong to the humanities and social sciences. While the opening sentence describing the website describes its broader context as “the processes of scientific communication,” within that it finds room to describe the contents of the journals as both “the results of research projects” and of “creative work” (*Portal*, 2023).

2. Purpose of the study

My induction into open science took place through taking, as a library science student, the course “What is Open Science?” provided by the Global Health Network (<https://globalhealthtrainingcentre.tghn.org/what-open-science/>). While I immediately perceived how the humanities were excluded from the discourse, I was nevertheless impelled to imagine the inclusion of the humanities by “translating” some of the foci of the course, and especially to think if those imagined possibilities could be turned into practical editing steps in Wikidata.

The following were my preliminary ideas:

1. What is the need for the humanities to be open? Is there anything in the humanities to match the urgency of open science fields? If that urgency could be found, how would it be communicated?
2. Do the humanities have data? What can count as data? And how can that data be turned into “knowledge”?
3. Perhaps it is the inverse that is true? Perhaps what the humanities have is knowledge, which needs to be turned into data?

4. Perhaps the open humanities could work on articulating central challenges of society, or break down large challenges into smaller ones, which could then be rallying points for research with results more discrete than is usual in the humanities, or which at least are more immediately shareable (a kind of sharing which doesn't typically happen in the humanist academy).

These questions are too broad and tentative to attempt to answer here, even if I do approach an answer in practical terms. I share them because they help explain my trajectory in the search for Wikidata editing options, and can be a useful starting point for further work and discussion.

3. Methodology

This is a case study in which I describe and evaluate the steps I was able to take in the Wikidata editing process, with a view to their replication by others as well as further reflection on their appropriateness and efficacy in the immediate context of Wikidata and the larger context of open science. In different contexts and areas, similar studies were carried out by Hitz-Gamper et al., 2019; Lemus-Rojas & Lee, 2019; Obregón Sierra, 2022; Obregón Sierra, 2021.

4. Preliminary results

4.1 Overview of creating journal items

There are a total of 29 journals available on the Portal. Of these, 17 are extant, of which only one, a journal of agriculture, *The Journal of Agriculture of the University of Puerto Rico*, could be conceived as reporting on the kind of universally replicable, data-centric science that open science typically engages in. The other journals were about various social sciences (library science, psychology, history, political science, administration, business, and education) and the humanities.

Between the months of January and April of 2023, I created Wikidata items for 13 thirteen journals, manually and from scratch, and augmented the record for 6 journals that already had items on Wikidata.

I used existing Wikidata items about journals as models, such as the item for the journal *Nature* (Q180445). I also referred to the Wikiproject Periodicals page (https://www.wikidata.org/wiki/Wikidata:WikiProject_Periodicals) for guidance on properties as well as a source for further exemplar items. To further explore available properties, I made use of the propbrowse tool (<https://hay.toolforge.org/propbrowse/>).

To learn about properties used to describe open access, I referred to items for open access journals.

4.2 Overview of creating article items

I experimented with both manual and automatic item creation. It was not possible to automate item creation through the Source MD (Source MetaData) tool, which helps to create articles from

DOIs, because most UPR articles don't have a DOI. The tool did not recognize the few that did have a DOI, namely those of the journal *Revista AnálisiS*, such as the following <https://doi.org/10.54114/revanlisis.v18i1.19489>.

As an alternative, I used Zotero to export metadata to the Quickstatements tool, as described by Bianchini (2021) and Allison-Cassin et al (2019).

To find which article items on Wikidata used a particular property, in order to also then see how the property was being used, I used a SPARQL query to find all instances of a scholarly article with a particular property, such as this query for all scholarly articles with a FAST ID property: <https://w.wiki/6W5J> (the result: one article).

4.3 Main subject at the journal level

In addition to labelling journals as “academic journal” and “scientific journal,” I also wanted journals to be labelled by academic discipline. Where that attribute didn't exist on Wikidata, I created an item for it. For example, I created items for “library science journal” and “philosophy journal.” In creating these items, it was sometimes necessary to create an item for the academic discipline itself, such as an item for “library science” as a discipline.

Furthermore, I added as many subject labels in the “main subject” field as I could find. I started with what was self-reported by the journal on its webpage, followed by what was reported in any indexing or cataloguing platform which listed the journal and contained a list of its subjects, such as Latindex and WorldCat. I also surveyed article titles from several issues of each journal to get an idea of content.

4.4 Main subject at the article level

As suggested by Thornton, the [main subject](#) property is of prime importance for the discovery of academic articles (2022). Often, I was able to choose items already available on Wikidata, but this was not always the case, and items needed to be created.

As recommended by Deng (2023), I bolstered the functionality of main subject items as linked data by adding a FAST ID to them where it was lacking.

I found that the FAST ID could only be added to the Wikidata item for the main subject, not to the article item itself. For example, in the item for the article “Descolonizando la Biblioteca (Q117084689), I listed as one of the main subjects the item [decolonization](#) (Q230533). I added the FAST ID to this last item. It is technically possible to add it to the article item itself, but the Wikidata template automatically turns it into an identifier for the article, which would not be accurate. OCLC allows FAST identifiers to be used as both name and subject authorities, but I found that Wikidata treats them as the former only.

4.5 Author disambiguation

Automatic adding of article items only contributes author names as strings to Wikidata, not author items. These items have to be added manually, therefore, for the most part, especially for authors who are not well-represented digitally, which is the case for many of the authors in this project. However, it is still useful to attempt some automatization in creating author items, as this can improve the representation of the author. Once an article is represented in Wikidata, entering its Q number in the Scholia tool will make it possible to choose the “improve venue” option, which links to an author disambiguator tool. This can help to find authors who already have a digital identifier, on a platform such as ResearchGate for example. Any digital foothold such as this, when found, could be used by the disambiguator tool to add (or add links to) other relevant data in the author item, such as other articles published by the author.

According to Donald (2020) adding a Library of Congress authority id (i.e. a name authority) to author items is an important first step, even when further information or the time to add it is not available, because it can establish a foothold that can help bots add information to the Wikidata author item in the future. I made limited progress with this, as only more established authors have a name authority in the LOC.

4.6 References to location

Reacting to Knöchelmann’s discussion of the open humanities, namely his contention that it may not be possible to have “a single space for an international discourse” (2019, p.8) as is the ideal in open science, I was impelled to think about what could be done on Wikidata to establish and nurture a local context for scholarship.

In addition to the standard property [country of origin](#), I found and used further properties, such as [location](#), [operating area](#), and [place of publication](#). Using the properties mentioned, I entered “Puerto Rico” as well as “Caribbean” as a geographic designation for the journals.

I also included “Puerto Rico” and “Caribbean” as a main subject for all journals, as this was self-reported by most journals.

4.7 Indexing

Where an item for a bibliographic index existed, I described a journal’s indexing using the [indexed in bibliographic review](#) property. In some cases, a bibliographic index existed but its item did not describe it as such. Fatcat was one such index. By going to the item for Fatcat and adding that it was an “instance of” “bibliographic index,” I was able to then create statements saying that the journals in my project were indexed by Fatcat. Fatcat was an attractive option because of its association with Internet Archive, a supporter of the Wikidata WikiCite initiative (James Hare, 2023), because its open indexing (without ranking) was more consistent with the open science ethos, and because it was free of charge.

On Fatcat, journals are designated as “venues.” Where a venue item did not exist for a UPR publication, I was able to add one. I did so by using my credentials as a Wikidata editor to establish an editing profile on Fatcat. One of the options for establishing a Fatcat venue item was through a QID. As I had already created journal items for my journals on Wikidata, I was able to

use that to establish a Fatcat venue ID on Fatcat where it did not exist. Once established, I added this ID to the Wikidata item for the journal, as a Fatcat ID.

5. Discussion

It was only at the later stages of my work, in the process of adding article items to Wikidata, and especially in the process of looking for Library of Congress and FAST subject authorities to add to the items designated as main subjects of the articles, that I realized that a contribution needed to be made, in relation to one of distinguishing aspects of the humanities as brought up by Knöchelmann: that the humanities are “recursive” (2019, p. 2). In the context of this project’s focus on the Caribbean, recursiveness can be usefully understood as the attempt to inhabit—to understand and to relate to—a particular place, and where the best answers are holistic, drawing on data that might be scattered across disciplines, collected and connected over the course of lifetimes. Examples of this kind of recursiveness are a library science article relating the history of library associations in Cuba,¹ or articles published in issues dedicated to one scholar or activist, such as Socorro Giron Torres in *Ceiba*² or Bernardo Vega in *Op. Cit.*³

In light of this, I explored and made preliminary editing attempts in two further directions, which I will discuss here.

5.1 Cooperation other than citation

[Cites work](#) is currently the most common property for drawing a picture of relationships between scholars and scholarship, followed by properties that can be used to describe academic relationships of student and teacher, such as [student](#), [student of](#), and [doctoral advisor](#).

The last three properties are useful to begin documenting a relationship between scholars in a local context, but this documentation needs to be more flexible and appropriate to the ways scholars in a local context can be inspired by each other, or inspired by key figures who are not necessarily scholars but who play key roles in local history. To this end, the use of the [influenced by](#) property is suggested. The use of this property would be appropriate for authors of tribute articles such as the ones mentioned above. Further use of this property would have to depend on self-reporting by authors, or on editors with a close knowledge of authors.

5.2 Wikidata editing possibilities hospitable to recursive knowledge

Tools such as Scholia can help match articles on the same topic, but how can authoritative topics be found for articles whose subject is local? For example, for the article on library associations in

¹ Enrique, L. E. P., Guzmán, M. F., & Rodríguez, R. A. M. (2022). Asociación Cubana de Bibliotecarios en Villa Clara: apuntes para su historia. *Acceso. Revista Puertorriqueña de Bibliotecología y Documentación*. <https://revistas.upr.edu/index.php/acceso/article/download/19888/17351>

² <https://revistas.upr.edu/index.php/ceiba/issue/view/2535>

³ <https://revistas.upr.edu/index.php/opcit>

Cuba, mentioned above, it was not possible to find a heading that narrowed the focus to Cuba. A topic item, such as “Library associations in the Caribbean,” might be created, especially if an editor knew of the existence of other articles on the topic. However, the approach of the humanities, as well as, to a certain extent, many of the social sciences, is to seek connections across the totality of a field, which means that an article might contain important information even if it not clearly or directly related to a stated or perceived main subject.

This author suggests that a web of Wikidata statements describing central influences, ideas, and events can provide a useful backdrop for describing author and article items at a higher level of granularity than that provided by FAST subject authorities. On Wikidata, items can be linked to other items relevant to a rich description of the subject matter, even if those items are outside the main subject field.

Figures and events central to local history referred to in an article can be described with the [significant event](#) and [significant person](#) properties. Items can be created for central ideas and classified by various levels of granularity, such as *theory* (Q17737), or, even more simply, *idea* (Q131841). Ideas with a cultural valence could be classified as a *tradition* (Q82821). At the discretion of the editor, these items could be used in the [main subject](#) field, but where the idea is not central, such ideas could be referenced with [references work, tradition or theory](#), a property currently used for creative works but which this author believes can be usefully adapted for articles in the humanities and social sciences.

It is hoped that in the future, this kind of granularity would provide a better ground for the discovery of information in one article that may be relevant to multiple disciplines, and which may facilitate collaboration across disciplines. One example from the present project is an article in the Journal of Agriculture whose main subject is the species of Coccinellidae (ladybugs) in Puerto Rico. A closer inspection of the article reveals a rich account of the cultural significances of this insect, but without any indication of the existence of this information in the abstract or keywords. Conceivably, linking this article to an item representing the idea that “Coccinellidae have healing powers,” could help locate the cultural information surrounding this insect, in a journal that is not ostensibly about culture. It could also conceivably make it possible for researchers from different disciplines to cooperate on documenting aspects of Puerto Rican and Caribbean culture.

6. Conclusions

Misgivings about the economic impetus behind it notwithstanding, Open Science does have the potential to foster a new democratizing ethos in the academy. This can include the humanities and social sciences.

As a first step towards this inclusion, this article details practical steps that can be taken to add humanities and social science journals to Wikidata, such as improving indexing and registering local geographic, disciplinary, and cultural contexts at higher levels of granularity.

The ideal level of granularity that the article proposes is a rich picture of connections among scholars, scholarship, and the world it refers to, rather than records of individual achievements registered through rankings and citation counts.

Open science practices

The journals discussed in this paper can be found in the *Portal de Revistas Académicas* of the University of Puerto Rico, at <https://revistas.upr.edu/>. The corresponding entries in Wikidata, of journal items and article items (work in progress), are openly available. The editing work of this author on Wikidata is being done under the username *Melmakko* (<https://www.wikidata.org/wiki/User:Melmakko>). The editing work of this author on FatCat is being done under the username *Melmakko-wikipedia* (<https://fatcat.wiki/editor/pl5hyat4yvgsrckdt3xeadh4wm>).

Author Contributions

Mazen El Makkouk

Conceptualized the research aims, conducted the investigation process, and wrote the original draft of this paper.

Cláudia De Souza

Principal Investigator of the project ‘Open Science Movement in the Caribbean Region’. Professor responsible for the assignment of tasks and control of the research schedule in the framework of the Course CINF6998. Involved in planning and supervised the findings of the work. Reviewed the intermediary draft of the article.

Carlos Suárez Balseiro

Co-investigator of the project ‘Open Science Movement in the Caribbean Region.’ Coordinator of the University of Puerto Rico Academic Journals Portal. Responsible for the original idea of the study in Wikidata. Contributed to the design and implementation, as well as contributed to journal and article sample selection. Helped supervise the project. Approved final draft of the article for publication.

Competing interests

The authors have no competing interests.

References

Allison-Cassin, S., Armstrong, A., Ayers, P., Cramer, T., Custer, M., Lemus-Rojas, M., ... & Stinson, A. (2019). ARL white paper on Wikidata: Opportunities and recommendations.

- Bianchini, C. (2021). Wikidata for J LIS. it: a new step forward mapping Italian library and information science journals. <https://doi.org/10.4403/jlis.it-12680>
- De Souza, Claudia & Suárez Balseiro, Carlos. (2022). La ciencia abierta en la Universidad de Puerto Rico. *Iberoamerican Congress of Open Science*. <https://doi.org/10.5281/zenodo.7268140>
- Deng, Sai. (2023) Linked Data, Wikidata and Their Implementations. *Faculty Scholarship and Creative Works*. 1171. <https://stars.library.ucf.edu/ucfscholar/1171>
- Donald, Margaret: Populating Wikidata with articles and authors: a how-to (2020). [video] *WikiCite 2020 - Open citations & linked bibliographic data*, Wikimedia Foundation (WMF). <https://doi.org/10.5446/51067>
- Hare, John. (Jan 31, 2023). WikiCite Meeting [virtual]. <https://etherpad.wikimedia.org/p/wikicite-monthly-2023-01>
- Hitz-Gamper, B. S., Neumann, O., & Stürmer, M. (2019). Balancing control, usability and visibility of linked open government data to create public value. *International Journal of Public Sector Management*, 32(5), 451–466. <https://doi.org/10.1108/IJPSM-02-2018-0062>
- Knöchelmann, M. (2019). Open Science in the Humanities, or: Open Humanities? *Publications*, 7(4), 65. <https://doi.org/10.3390/publications7040065>
- Lemus-Rojas, M., & Lee, Y. Y. (2019). Using Wikidata to Provide Visibility to Women in STEM. *International Conference on Dublin Core and Metadata Applications*, 126–131.
- Longley Arthur, P., & Hearn, L. (2021). Toward open research: A narrative review of the challenges and opportunities for open humanities. *Journal of Communication*, 71(5), 827-853.
- Mirowski, P. (2018). The future (s) of open science. *Social studies of science*, 48(2), 171-203.
- Strickler, C. M. (2021). Mind the Wikidata Gap: Why You Should Care About Theological Data Gaps in Wikipedia's Obscure Relative, and How You Can Do Something About It. *Atla Summary of Proceedings*, 301-314.
- Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *The Journal of Academic Librarianship*, 47(2), 102326.
- Thornton, Katherine. (2022) Wikidata Workshop, online. SWIB 2022 Conference.
- Obregón Sierra, Á. (2022). Inserción de metadatos de las bibliotecas españolas en Wikidata: Un modelo de datos abiertos enlazados. *Revista Española de Documentación Científica*, 45(3), <https://doi.org/10.3989/redc.2022.3.1870>
- Obregón Sierra, Ángel. (2021). Análisis bibliométrico con Wikidata: El caso Comunicar. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, 47.