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

Christa Strickler, Wheaton College (IL)

ABSTRACT Wikidata, a community-curated knowledge base related to Wikipedia, affects our access to information, wielding more power than many realize. Seeing an opportunity for improving access to knowledge and promoting their collections, libraries, archives, and other cultural heritage institutions have been experimenting with Wikidata in various ways. One burgeoning area of activity is in Wikidata's scholarly citation data, but that participation has largely concentrated in the sciences, leaving a gap in its theological and religious studies coverage. This presentation demonstrates how this gap matters to theological libraries and shows how you can contribute to efforts to fill it, even in small ways. Delving into Wikidata can be intimidating, but it doesn't have to be. As long as you have basic computer skills, you can find a way to participate.
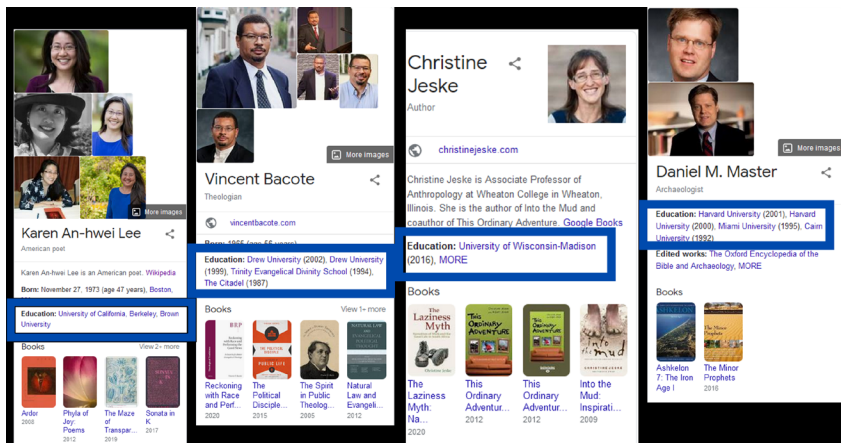
Imagine a world where librarians could connect people to scholarship on religion and theology, not just in our buildings and database silos, but also on the open web. Imagine a world where we could promote the scholarship produced by our institutions, even if our faculty members don't have ORCID IDs or know anything about scholarly identity management.

Such a world is possible.

Enter Wikidata, the somewhat obscure structured data counterpart to the well-known Wikipedia. The general public is mostly unfamiliar with Wikidata, but it plays a powerful yet mostly invisible role in the information they consume. As one example, Wikidata feeds into the data used by digital assistants like Siri and Alexa. An article from the magazine *Wired* tells of how companies like Google and Amazon scrape data from Wikidata and combine it with other sources to inform their digital assistants, though those companies often decline to tell how they use the data (Gross 2019).

Wikidata also influences the knowledge panels of search engines like Google and Bing. These knowledge panels highlight certain kinds of information, giving greater prominence to some search results over others. Pictured here are four Google knowledge panels featuring the provost and three faculty members at Wheaton College.

IMAGE 1. Four Google knowledge panels for Wheaton College administration and faculty: Karen An-hwei Lee (https://g.co/kgs/6T2nGx), Vincent Bacote (https://g.co/kgs/MCGibN), Christine Jeske (https://g.co/kgs/TE7UKQ), and Daniel M. Master (https://g.co/kgs/5qwKPP)

The highlighted areas show their educational background, which appeared in the knowledge panels after I and my cataloging assistant did a project to create Wikidata entries for all of Wheaton's faculty.

And it's not just on the open web that Wikidata has influence. Wikidata is also starting to be used in knowledge panels in library catalogs to enhance discovery. The Linked Data for Production group, including librarians from Cornell, Stanford, and the University of Iowa, has been working on a prototype using sources such as Wikidata to provide context around entities found in a library catalog. When they present this prototype at conferences, they use the example of someone searching for *Lincoln in the Bardo,* a novel by George Saunders. If you click for more information about the author while looking at the bibliographic record, you can see a knowledge card powered in part by Wikidata. This knowledge card displays contextual information about Saunders, such as a photograph and a list of authors who influenced his writing. In the testing phase for this

prototype, users reported that these kinds of enhancements made the catalog search results stickier (Usong and Khan 2020).

Imagine what it would be like for scholars in religion and theology to have "sticky" search results just like authors of popular novels. But as things stand now, they won't. At least, not many of them.

Wikidata has gaps, and a big one is in religion and theology. As a crowdsourced platform, Wikidata reflects the biases of its contributors, who are mostly male and from the Northern Hemisphere (Sengul-Jones 2021), and when it comes to data about scholarship, Wikidata skews towards the hard sciences. To visualize this gap, we will explore Scholia (https://scholia.toolforge.org/), a service that creates visual scholarly profiles for topics, people, organizations, and more using information from Wikidata. Unlike commercial services like Scopus and Web of Science, Scholia is freely available on the web. Unlike Google Scholar, Scholia has a transparent data source that can be freely used in other applications.

Let's begin by exploring some highlights of the Scholia profile for the topic of COVID-19 (https://scholia.toolforge.org/topic/Q84263196). Thanks to the work of volunteers, as of today, Scholia shows data for nearly 70,000 information sources (mostly scholarly articles) published in 2020 alone. The profile also displays the researchers who publish the most on the subject, a co-author graph, a graph of related topics, a list of journals with the most articles on the topic, and a list of the most commonly cited researchers of the topic. All of this comes from data added by volunteers in Wikidata.

Religion and theology, however, do not fare so well. The topic profile for Religion and Politics (https://scholia.toolforge.org/topic/Q16546879) shows 14 published works in total. The profile for the *Journal of Jewish Studies* (https://scholia.toolforge.org/venue/Q15750751) shows eight published works in total. This is a huge contrast to the over 70,000 articles listed on the topic of COVID-19. Granted, this specific comparison is not entirely fair, as COVID-19 is a hot topic these days, but there are far more than 14 works published on the topic of religion and politics than Scholia shows, and the *Journal of Jewish Studies* has published more than eight articles since its inception. The point is that religion and theology are not a priority for Wikidata editors right now. Try exploring this gap by using Scholia to look up your institution, or faculty from your institution, or a topic of interest, and see what you find.

Because of Wikidata's unseen influence in the data world, data gaps in Wikidata mean gaps in the information promoted in commonly used information sources like Google, Siri, and Wikipedia, all without many people knowing the gap is there. This is where we come in, as collectors and connectors of religion and theology. We can help fill these gaps.

You might wonder how we can do this. I think most of us can attest to being busy. Many of us are understaffed and overworked. Some of us are from smaller libraries and wear many hats. The good news is that even a small contribution to Wikidata can make a difference, and starting out is simpler than it looks.

The best place to start if you have little or no experience with Wikidata is to manually create an item for a faculty member at your institution. If you don't already have a Wikidata or Wikipedia account, you'll need to create one at the **Create account** link at the top of the main Wikidata page (https://www.wikidata.org).

I will show you how I created an item for Elizabeth Bloch-Smith, an adjunct faculty member at Princeton Theological Seminary, using only the information found on her faculty web page (https://www.ptsem.edu/people/elizabeth-bloch-smith).

1) Do a quick search to make sure the faculty member of interest is not already in Wikidata.

   IMAGE 2. Close-up of the search box on the Wikidata home page: https://www.wikidata.org
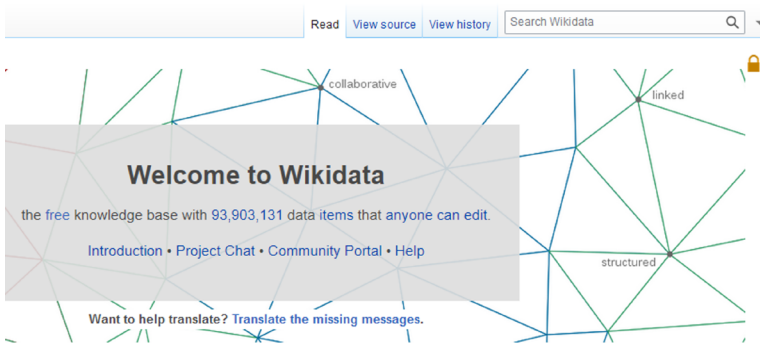
2) Once you've verified that you're not creating a duplicate item, you can create a new item.

IMAGE 3. Close-up of **Create a New Item** on the Wikidata home page: https://www.wikidata.org
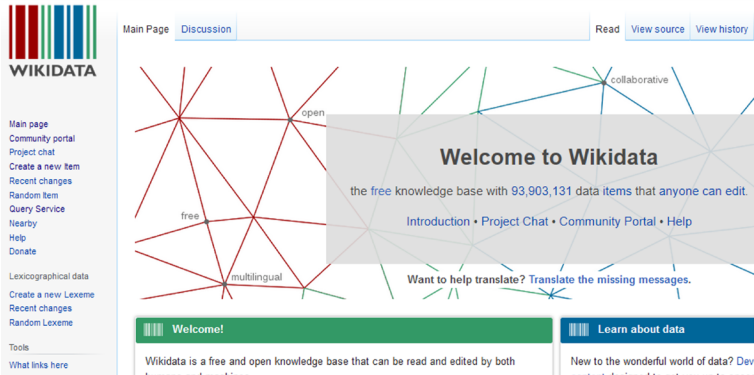
3) Add a label and a description using the instructions above the **Create a New Item** box. Keep in mind the warning about privacy for living people. Only input information for which you have evidence. This is why I'm only including information found on Dr. Bloch-Smith's web page.

IMAGE 4. Close-up of the **Create a New Item** page (https://www.wikidata.org/wiki/Special:NewItem)

Next, you will create a series of statements to make the item meaningful and useful. Each statement consists of a property and a value. You will need to tell Wikidata which property you want to use and which value you want to pair with that property. There are thousands of choices for property and value, so if you don't know which one to use, this part can be intimidating. For a person, you

start with the property "instance of" and the value "human." When you start to type those into the form, a dropdown menu will appear, and you can choose the correct option.

IMAGE 5. Creating a statement in the item for Elizabeth Bloch-Smith (https://www. wikidata.org/wiki/Q107196635)
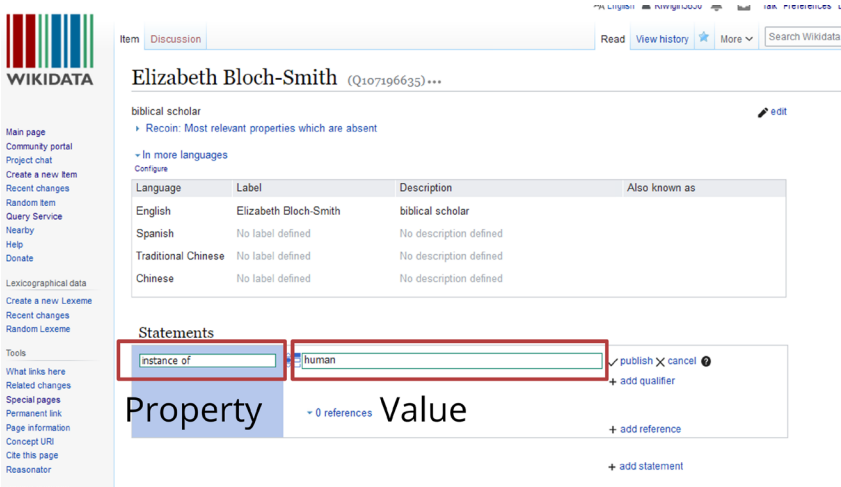
There are many properties I could choose next, like gender, occupation, or field of work, but those can get complicated, so we'll add an employer statement next. Once you have **employer** chosen as a property, you start typing the name of the employing organization in the **value** field, and if that organization has an item in Wikidata, you should be able to choose the correct option from the dropdown menu.

IMAGE 6. Creating an employer statement
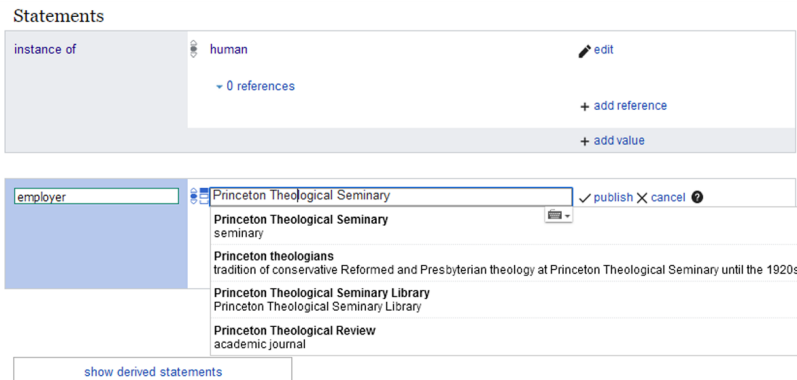
For nearly every statement that you make, it's best practice to provide evidence. There are different kinds of evidence you can supply, but the simplest type is a URL. Begin by selecting **add reference** within the statement you just created.

IMAGE 7. Adding a reference
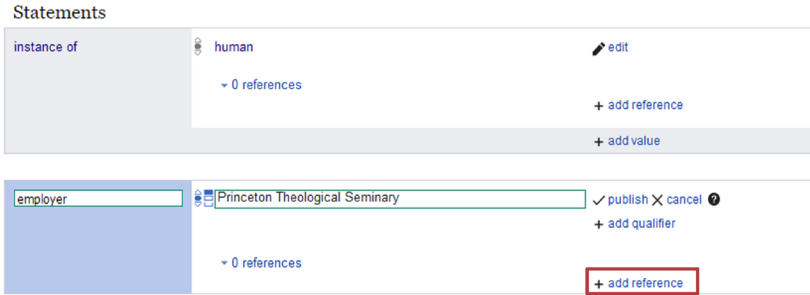
References are also property-value pairs. If you're using a URL, use the property **reference URL** and paste the URL into the **value** field. Once that's saved, you now have an item for Elizabeth Bloch-Smith that links her to her employer, Princeton Theological Seminary.

IMAGE 8. Adding a reference URL
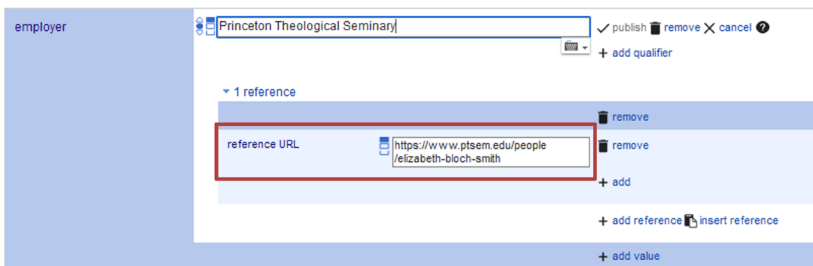
If possible, I also try to include an identifier of some sort, usually a Library of Congress authority ID. This connects the Wikidata item to other data sets, which helps bots come through later and add more data based on those data sets. There are many other properties we could add, and I'll give you a link to a web page with a list of potential options.

IMAGE 9. Adding a Library of Congress authority ID
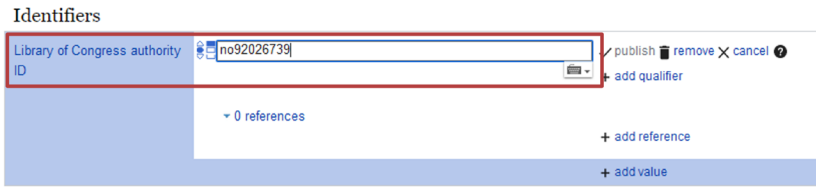
Now that I've created an item for Elizabeth Bloch-Smith, I can find her in Scholia. Right now, her profile is rather boring because she's not linked to any publications.

IMAGE 10. Scholia profile for Elizabeth Bloch-Smith (https://scholia.toolforge.org/author/Q107196635)
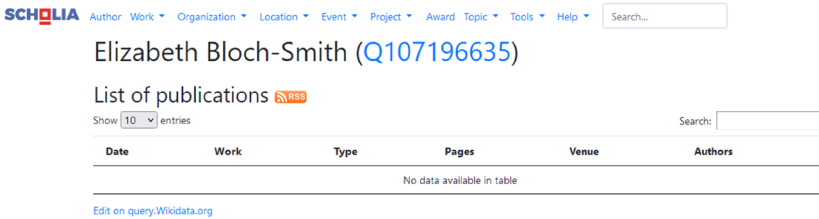
You can create items for her individual publications, but there are also batch methods to make the process quicker. Before I created her item in Wikidata, Elizabeth Bloch-Smith had five publications listed in Wikidata, but none of them were linked to her because she didn't have an item. I ran an author linking tool provided by Scholia, and in less than a minute I had those five publications linked to Dr. Bloch-Smith, and none of those publications were entered manually. Even though she does not have an ORCID ID, she has the beginnings of an academic profile.

IMAGE 11. Scholia profile for Elizabeth Bloch-Smith after running author linking tool (https://scholia.toolforge.org/author/Q107196635)

Also, because I linked Dr. Bloch-Smith to Princeton in Wikidata, she and her publications are now linked to Princeton's organization page in Scholia. All of this took me less than five minutes to create.

IMAGE 12. Scholia profile for Princeton Theological Seminary (https://scholia. toolforge.org/organization/Q909696)

Of course, what you get out of it depends on what you put into it. It takes time to build a more robust institutional profile like we've been doing at Wheaton College, with lists of Wheaton faculty and recent publications, along with a graph of Wheaton citations. But it can be done, just a few minutes at a time.

IMAGE 13. Scholia profile for Wheaton College (https://scholia.toolforge.org/organization/Q747179)

IMAGE 14. Scholia publication list for Wheaton College (https://scholia.toolforge.org/organization/Q747179)

IMAGE 15. Scholia citation graph for Wheaton College (https://scholia.toolforge.org/organization/Q747179)
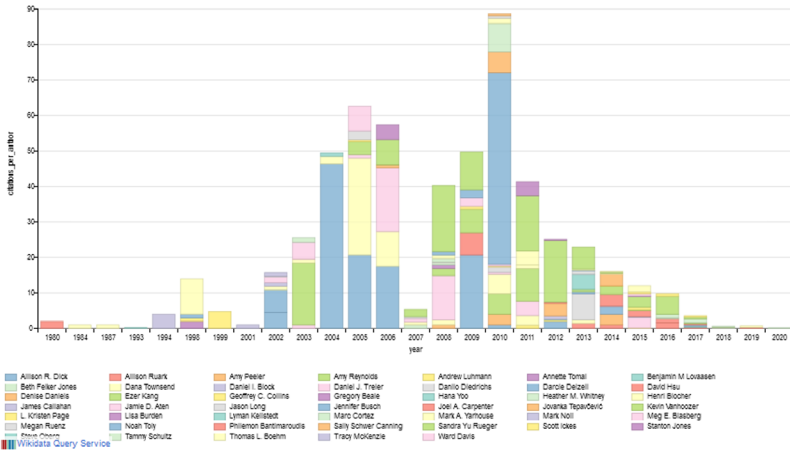
The beauty of Wikidata is that even contributing one item with a few statements can be helpful, because others can build on that one item. In May 2021, I added an item for Paul Kahle's article in *The Journal of Biblical Literature*, "Greek Bible Manuscripts Used by Origen," and I created an English-language label and description for it (https://www.wikidata.org/wiki/Q106835275). In June, a bot enhanced my description by translating it into several languages, adding value that I didn't have the skills or time to do myself.

If you're feeling ambitious and have some tech skills such as Python, there are other things you can do. Steve Baskauf of Vanderbilt University has built a Python tool called VanderBot, which scrapes the web profiles of Vanderbilt faculty and creates Wikidata items for them in a batch. He has made his documentation freely available (http://baskauf.blogspot.com/2021/03/writing-your-own-data-to-wikidata-using.html) and created YouTube videos to help you get started (https://www.youtube.com/c/SteveBaskauf/videos).

For those like me who are ambitious but don't have such tech genius, another option is to import citations into the citation management tool named Zotero (https://www.zotero.org/), export them into a data tool called OpenRefine (https://openrefine.org/), then reconcile the data to Wikidata and upload it.

IMAGE 16. Uploading data to Wikidata using OpenRefine

This may sound intimidating. Even if you're familiar with Zotero, I'm guessing that many of you don't use OpenRefine, even if you know what it is. It can be hard to figure out, and the available documentation isn't intuitive for those like me without coding skills, so I've spent hours figuring out how to make it work. The good news is that I've learned how to upload hundreds of citations in less than an hour, saving me the time of entering each citation manually. With the workflow I've created for uploading scholarly articles, I've made 18,500 edits to Wikidata this year alone, and that's while still wearing all my other hats of cataloging, supervising, information literacy instruction, collection management, etc. When I have some spare time, I upload article metadata or create researcher metadata to fill that religion and theology Wikidata gap.

But I can't do it alone, so I turn to you for help. From what I can see, very few Wikidata volunteers are adding citation data for religion and theology scholarship. I have seen some work coming out of Vanderbilt Divinity School, but that's the only concentrated effort I've encountered outside of my own. Additionally, my area of expertise is Christian theology, so help from others with expertise in other religions and disciplines is most welcome. To that end, I propose starting a community of practice for Atla members interested in using Wikidata for religion and theology citation data or even other projects of interest to our field of work. Whether you know how to code or you're only good at data entry, you can contribute. Even if you only have 10 minutes to contribute every so often, you can help.

Even if you only ever create one researcher profile, someone can come along and attach research articles to that profile.

To start, I've created a Wikidata page with a data model that describes recommended properties and values for items for scholars (https://bit.ly/3wlnRON).

IMAGE 17. Atla 2021 Mind the Wikidata Gap web page (https://bit.ly/3wlnRON)

Hopefully the data model I've provided there will help demystify the decision-making process for creating detailed items. I've also linked to data models for scholarly articles and journals, but I can customize some specifically for this page if there is interest.

Another option to explore is to create written documentation or short tutorial videos for workflows. One of the challenges about learning to use Wikidata and OpenRefine is that the available documentation is often highly technical and difficult to figure out, and the training videos are often an hour or more long, making it difficult to find the one piece of information you need.

If a Wikidata Religion & Theology Community of Practice is something that interests you, even if you're not ready to commit yet, please mail me at christa.strickler@wheaton.edu. I'm happy to answer any questions you might have about Wikidata.

Imagine a world where librarians could connect people to scholarship on religion and theology outside of our silos. Such a world is possible, if we, as collectors and connectors in religion and theology, collaborate to make it happen.

WORKS CITED

Gross, Tyler. 2019. "Inside the Alexa-Friendly World of Wikidata." *Wired*, February 10, 2019. https://www.wired.com/story/inside-the-alexa-friendly-world-of-wikidata/.

Sengul-Jones, Monika. 2021. "The Promise of Wikidata: How Journalists Can Use the Crowdsourced Open Knowledge Base as a Data Source." DataJournalism.com. February 10, 2021. https://datajournalism.com/read/longreads/the-promise-of-wikidata.

Usong, Astrid, and Huda Khan. 2020. "Libraries and Linked Data: A Tale of Two Technologies." Presented at the LD4 Conference. http://bit.ly/ld4pdiscovery.