**ABSTRACT** WikiProject Christian Hymns began as one individual’s involvement in a pilot project to test Wikidata as an identity management platform, but it has continued on as a standalone project to improve the coverage of global Christian hymnody in Wikidata and connect the resulting data to library resources. The complexity of hymnody creates challenges in the Wikidata environment, with tunes ranging from African American spirituals and medieval chant to contemporary Christian worship melodies and adaptations of folk music. The addition of texts from a variety of sources and languages increases the complexity. This paper describes the creation of the project, discusses the difficulties encountered in developing a data model, shows visualizations that explore various aspects of the data, and demonstrates connections with library resources while considering ongoing challenges.

**BACKGROUND**

Wikidata, a community-curated structured data repository related to Wikipedia, offers an opportunity for cultural heritage institutions to experiment with new ways of providing access to their resources. Freely editable, Wikidata allows anyone with an Internet connection to create and edit metadata for nearly anything in the world, from chemical formulas to buildings to Cookie Monster. Thanks to its public domain Creative Commons license, Wikidata’s metadata can be reused in all sorts of applications, including Wikipedia infoboxes, search engine knowledge panels, and artificial intelligence applications such as Siri and Alexa. This free and open model has attracted participation from galleries, libraries, archives, and museums look-
ing for new ways to enhance their collection metadata and share it outside of traditional library data silos.

As part of this wave of participation, the Library of Congress Program for Cooperative Cataloging (PCC) launched a pilot program during the COVID-19 pandemic in the fall of 2020 to test Wikidata as an identity management platform that might make it easier for catalog and metadata librarians to create name authority records (https://www.wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot). I was working from home as a catalog librarian at Wheaton College at the time, and I needed a project to help fill the work-from-home hours, so I signed up for the pilot program. Participants were allowed to create their own projects, so I decided to work on Christian hymns for the following reasons. First, I have an educational background in music and Christian theology, with performance experience in sacred choral music, so I had personal expertise in the knowledge domain. Second, my family owns a personal collection of hymnals and related reference sources that I could refer to without relying on access to library-owned resources while the building was closed for the pandemic. Third, there had been little activity in Christian hymns in Wikidata up to that point, so there was a large knowledge gap in that subject area. Finally, the Wheaton College Archives & Special Collections owns an extensive hymnal collection comprising more than 2,000 Protestant hymnals from the 1790s to the 1980s (https://archives.wheaton.edu/repositories/5/resources/774), and I hoped to connect my Wikidata work with the library's hymnal collection.

For my project scope, I chose one hymnal, *Glory to God: The Presbyterian Hymnal (Glory to God 2013)*. I had a copy of it at home, so it was easily accessible. My main sources for information were the hymnal itself; the hymnal's companion volume (Daw 2016), which had background information on the hymns and hymn writers in the hymnal; the *Canterbury Dictionary of Hymnology* (https://hymnology.hymnsam.co.uk), an online subscription resource covering a variety of topics in global Christian hymnody; and Hymnary.org (https://hymnary.org), a free online index of hymns and hymnals. Hymnary.org was especially important because each tune, text, and person that it indexes has an identifier that is a valid property in Wikidata, so I wanted to use those identifiers to link the data in Hymnary.org to Wikidata. Additionally, Hymnary.org's data is downloadable as a .csv file, making it easier to set up batch uploads of data to Wikidata.
With all of this in place, I was ready to start. I sat down with the hymnal, intending to add or edit each hymn manually. However, it quickly became apparent that this was going to be far more complicated than I initially imagined. This hymnal has 852 hymns, for starters. But then there were all of the composers, hymn writers, translators, and other factors that I hadn’t considered. I had failed to consider the data model.

**DATA MODELING**

Wikidata has a variety of data models that can help one figure out how to format data for Wikidata items in various subject areas. For example, there are data models for books ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Books](https://www.wikidata.org/wiki/Wikidata:WikiProject_Books)), music ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Music](https://www.wikidata.org/wiki/Wikidata:WikiProject_Music)), and so on. If Wikidata editors consistently follow these models, then the data is easier to use because it follows expected patterns.

When I started, there were few items related to hymns in Wikidata, so there was no data model to draw from. The data that was already in Wikidata had been entered inconsistently, so I could not easily infer a data model from existing data. Having no idea what I was getting into, I began creating items without a data model. Image 1 shows the data model in my head. There would be one Wikidata item for each hymn, and one for each person involved with the creation of the hymn. Pretty simple, right? But then there was reality. From the moment I began looking at the first hymn, “Holy, Holy, Holy,” with English text and a Korean translation, I started to realize that the model in my head didn’t fit reality. And it went downhill from there. What you will see next are some of the challenges that I faced in relation to data modeling.

![IMAGE 1](image-url) Original mental construct of a data model for hymns

A fundamental challenge I encountered is that one hymn does not equal one Wikidata item -- at least, not exactly. One way of looking at a hymn is to see it as one entity, basically like a song, as you can see in Image 2 on the left. That was how I originally thought of
it. However, from an academic and liturgical viewpoint, hymns are more like two entities combined: a text and a tune, and you can mix and match tune and text pairings if they have the same hymn metre, which is the number of syllables for the lines in each stanza. This second version is how Hymnary.org models it, and it uses identifiers for both text and tune. Once I had a better sense of my data, this was also how I wanted to model it because that's how scholars would likely want to use it.

**IMAGE 2** The model on the left shows as a hymn as one entity, like a song; the model on the right shows a hymn as two entities: a text and a tune.

Hymnary.org then combines tune and text into a hymn instance, with its own identifier, for each time a particular combination of text and tune is published in a hymnal (see image 3). If the same combination of a text and tune appears in 100 hymnals, then Hymnary.org will record 100 instances for it.

**IMAGE 3** Hymnary.org's use of a hymn instance for each time a combination of text and tune is published in a hymnal.

I thought about doing this but decided against it. If I were to implement it, there would be at least 1,391 instance items titled “Amazing Grace” in Wikidata, which would make it difficult for people wanting to find other Wikidata items with that title, such as the album *Amazing Grace* by Aretha Franklin. I also chose a hymn instance, but I’m not implementing it the same way as in Hymnary.org, with one instance for each time it shows up in a hymnal. Instead, Image 4 shows...
how I use it for a unique combination of text and tune, regardless of how many times that combination shows up in different hymnals.

rddefined: Instance of a unique combination of text and tune, regardless of the hymnal in which it appears

**Image 4** My use of a hymn instance for a unique combination of text and tune, regardless of how many hymnals use the combination.

Even though I have now figured out a model that incorporates the concepts of hymn instance, tune, and text, putting it into practice can get complicated, as the following example shows. Image 5 shows the Wikidata item for “Jesu, Meine Freude,” the hymn text written by Johann Franck in 1653.

**Image 5** Wikidata item for “Jesu, Meine Freude.” [https://www.wikidata.org/wiki/Q118445051](https://www.wikidata.org/wiki/Q118445051)
There is also a Wikidata item for the hymn text “Jesus, Priceless Treasure,” which is Catherine Winkworth’s 1863 translation of “Jesu Meine Freude” into English (https://www.wikidata.org/wiki/Q118445068). The German text and the English translation are typically set to the hymn tune titled JESU, MEINE FREUDE, composed by Johann Crüger in 1653, and this hymn tune has its own Wikidata item (https://www.wikidata.org/wiki/Q109747915). Then there is the version of the tune JESU, MEINE FREUDE that was harmonized by Johann Sebastian Bach in his motet Jesu, meine Freude (https://www.wikidata.org/wiki/Q118445111). I keep this separate from the original tune because some hymnals include the Bach harmonization and some do not, and I consider Bach significant enough to make the distinction. For the text and tune alone we already have four Wikidata items. Now we combine them to get more. There is a Wikidata item for the hymn instance of “Jesu, meine Freude,” which is the combination of the original German text with the original unharmonized tune (https://www.wikidata.org/wiki/Q17050641). Another item exists for “Jesu, meine Freude,” the hymn instance that combines the original German text with the harmonized tune (https://www.wikidata.org/wiki/Q118445130). Yet another hymn instance item describes “Jesus, Priceless Treasure,” which combines the English translation of the text with the harmonized version of the tune (https://www.wikidata.org/wiki/Q118445101). At some point, I would like to figure out how best to connect the German text and harmonized tune with the item for Bach’s motet (https://www.wikidata.org/wiki/Q892166), but that has some unresolved complications. This comes to eight items involved with one hymn, not counting the items related to the composer, the hymn writer, the translator, and the harmonizer.

Another challenge was to ensure that the data models that I created would dovetail into the models already in Wikidata. My musical content overlapped with music data models already in existence, so I needed to make sure that I didn’t contradict what was already in place. Image 6 shows the Wikidata item for the hymn tune “All Things Bright and Beautiful,” by William Henry Monk. When I first started creating items for hymn tunes, I would create a Wikidata statement claiming that the item was an instance of a hymn tune.
After creating several items in this manner, I discovered that Wikidata’s WikiProject Music ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Music](https://www.wikidata.org/wiki/Wikidata:WikiProject_Music)) preferred all musical works to be an instance of a musical work/composition to help with standardized data querying across all musical works in Wikidata. The type of musical work should then be indicated by using musical form and genre statements. Image 7 shows how I adapted my data model to reflect the WikiProject Music model and fixed the data in the items I had already created. Hymn tunes became instances of musical work/composition, with the form of hymn tune.
Similarly, hymn instances such as “To God Be the Glory” by Fanny Crosby and William Howard Doane were once described as an instance of a Christian hymn (see image 8).
Image 8  Wikidata item of “To God Be the Glory,” before it was edited to reflect the new model: https://www.wikidata.org/wiki/Q2855428

But now the recommended model is to describe them as an instance of a musical work/composition, in the form of a song, with the genre of Christian hymn (see Image 9). This makes them consistent with the WikiProject Music model. There is still data clean-up needed, however, because other Wikidata editors have created hymn items that do not conform to this.
Yet another challenge is determining what to do when to fit into existing models in Wikidata when they are insufficient or incorrect. When I first started working on this project, the Wikidata item for the concept of “Christian hymn” conflated the idea of a hymn text with the idea of a hymn that combines a text and a tune. The concept of Christian hymn was defined as a type of Christian song, but the item had other properties that indicated that a Christian hymn could also be defined as a hymn text, as you can see on the right side of Image 10, where it says a Christian hymn is also known as a hymn text.
In some ways, this definition is accurate. A hymn text alone could be called a hymn, but for purposes of describing hymn texts separately from Christian hymns that combine a text and tune, this was a problem. Also problematic was that the item allowed the use of a Hymnary.org text ID, tune ID, and instance ID to be used to describe a full hymn (see Image 11). This does not match Hymnary.org’s data model at all. The text ID only refers to a text, the tune ID only refers to a tune, and Hymnary.org’s concept of a hymn instance was different from anything in Wikidata.

Before I could start adding hymn texts to Wikidata, I had to sort out this mess. I could have just fixed it all without telling anyone, but I wanted to be a good Wikidata citizen, so I asked the community. Every Wikidata item comes with a discussion page, so I described my conundrum on the “Christian hymn” item discussion page (https://www.wikidata.org/wiki/Talk:Q856713) to see if anyone had any problems with my proposed solution. Nobody did, so I began fixing the problem. I created an item for a hymn text (see Image 12), then moved
over all of the properties from “Christian hymn” that were only appropriate for the hymn text alone, such as the Hymnary text ID.

Another challenge I’ve come across is the modelling of hymn tunes not originally written as hymn tunes. Some songs that appear in hymnals, particularly contemporary ones, are not technically hymns at all, such as national anthems, African American spirituals, gospel songs, traditional folk music, Taizé chant, and contemporary Christian pop music. The growing diversity in our hymnals is a beautiful thing to see, but with diversity comes complexity, which creates challenges in data modeling. My current solution is to describe the tune both as a hymn tune and as a more generic melody, adding extra de-
tails so that others looking for the tune can find it whether they’re looking for it as a hymn tune or in another form. For example, Bí-NÍU (see Image 13) is an adaptation of a Bunun rice-pounding song from Taiwan. For this tune, I’ve still claimed that the musical form is a hymn tune, but I have also added melody as a form. To reflect the cultural origins of this tune, I have added other statements to indicate that the original work comes from the Bunun people of Taiwan. It is not a perfect solution, so I may change my data model in the future if a better option presents itself.

**IMAGE 13** Wikidata item for the hymn tune Bí-NÍU: [https://www.wikidata.org/wiki/Q114273586](https://www.wikidata.org/wiki/Q114273586)
Now that I have worked out some of the modeling challenges, I have created a simplified visual of my data model (see Image 14), to show how the different pieces fit together.

![Simplified diagram of data model for the project](image14)

Each box represents a category of item, and the rows in each box indicate some of the properties that can be used to describe that kind of item. The hymn instance is at the center of the model, and this hymn instance connects to hymnal works and editions, hymn texts and translations, hymn tunes and adaptations, and all of the people involved in creating and editing them. You can see the comprehensive version of this in table format on the project page (https://www.wikidata.org/wiki/Wikidata:WikiProject_Christian_Hymns#Models), which contains information on describing composers and hymn writers, hymn instances, hymn texts, and hymn tunes.

**DATA VISUALIZATIONS**

A benefit of using Wikidata for this project is the ability to use pre-made data visualizations that open new avenues for exploring the data. These data visualizations are created using the SPARQL query language in the Wikidata Query Service (https://query.wikidata.org/). I'm still learning SPARQL and don’t know how to do anything complicated, but here are some samples of what I’ve done so far.
The most comprehensive visualization is a graph that explores the connections between all of the items in the project (https://w.wiki/4DWJ). It’s rather unwieldy since the project has become so large, but I love browsing it to see the connections that I might not notice otherwise. In Image 15, you can see Martin Luther connected to several tunes and texts for which he is responsible, along with connections to other Protestant Reformation figures (such as Philipp Melanchthon, Martin Bucer, and John Calvin), which shows relationships between major Reformation figures and the hymnody developed during the time period. You can also see Luther’s influence by Augustine, whose writings have also made appearances in Christian hymn texts.

**IMAGE 15** Close-up of graph visualization of the project, focused on Martin Luther and his connections in hymnody: https://w.wiki/4DWJ}
The online timeline creation tool Histropedia (http://histropedia.com) can use Wikidata queries to generate timelines. The timeline shown in Image 16 includes all people in the project who have a recorded birth date.

You can zoom in and out of a time period to see more people. This one focuses on the Protestant Reformation era, so you can see people such as Martin Luther, Thomas Tallis, John Calvin, and Theodore Beza. You can further focus the timeline by clicking on the facets on the left to narrow by a particular religion. This is only as good as the data it draws from, however, and Wikidata is weak in coverage of religion, so that needs some cleaning up. An additional quirk that I haven’t been able to figure out is that when Wikidata records two potential birth dates for a person, they appear twice in the timeline, as Image 16 shows with Martin Luther and Martin Bucer. I’m hoping to fix that at some point. You can also create timeline facets for other characteristics such as geographic location.

Data visualizations have also helped me identify gaps in my content, so I’ve changed my project scope as a result. This map data visualization from 2021 (see Image 17) shows that the bulk of my hymn writers and composers at that time had been born in the United Kingdom, Western Europe, and the East Coast of the United States.
Also in 2021, almost 200 hymn writers in the project were English speakers, 60 were German speakers, and a few other languages trailed behind, mostly Western European (see Image 18). Data like this inspired me to diversify my work, so I have expanded my scope to include a more global variety of hymns.
While this new scope has come from a variety of sources, I have particularly focused on *Sound the Bamboo* (2006) and its hymnal companion (Loh 2013). The hymnal contains hymns from a variety of Asian locations, while the hymnal companion contains background information about the hymns and their composers. Because of the Asian emphasis, the current map visualization (see Image 19) shows more hymn writers from Japan, Korea, China, India, Indonesia, and the Philippines, but I’ve also seen expansion in Brazil, Panama, and parts of Eastern Europe, the Middle East, and Africa.

![Image 19](https://w.wiki/4JKk)

English language speakers still dominate at 326, but I’ve seen inroads in Japanese, Korean, Hindi, Mundari, Burmese, and others. The updated bar graph with all of the new languages is too large to show in a legible way, but you can see an up-to-date version at [https://w.wiki/4KiH](https://w.wiki/4KiH). In order to increase the project’s global diversity in the future, I’ve begun compiling a list of hymnals to focus on later that include hymns from Asia, Africa, Latin America, and the Caribbean ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Christian_Hymns#Other_Sources_for_Focus](https://www.wikidata.org/wiki/Wikidata:WikiProject_Christian_Hymns#Other_Sources_for_Focus)).

**CONNECTIONS TO LIBRARY MATERIALS**

Because I’ve been spending so much time working on the hymn tunes, hymn texts, and composers and writers involved in Christian
hymnody, I haven’t been able to make many connections to library materials yet, but that’s still a major goal of the project. I need a critical mass of hymn data in Wikidata first, and then I can create batch processes to connect that data to library resources. Before I can accomplish that, a major task is to sort out the hymnals. Up to this point, I’ve been working at the level of texts, tunes, and people. None of these things are library materials. But hymnals and hymn reference sources are. The reference sources aren’t so much of a problem, but the hymnals are trickier for two main reasons. The first problem is that there isn’t a good model in Wikidata for music scores. Wikidata’s WikiProject Music has a model for musical compositions at the work level ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Music#Composition_properties](https://www.wikidata.org/wiki/Wikidata:WikiProject_Music#Composition_properties)), but nothing for published manifestations that I’ve found, so I’ll need to come up with something on my own. I will likely make them somewhat like books, with separate items for works and editions as is recommended in WikiProject Books ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Books](https://www.wikidata.org/wiki/Wikidata:WikiProject_Books)).

The reason I would want to do this is that some hymnals come in multiple editions. One example is the Genevan Psalter, which was important in the Protestant Reformation and is still influential in contemporary hymnody. It came in eight editions ranging from 1539 to 1562. These eight editions vary in content, so it would be important to differentiate them in Wikidata. However, I haven’t yet figured out how to indicate that while the editions are in volumes like a book, their content is mostly musical rather than textual.

The other problem is creating an effective batch process to add hymnal items to Wikidata. I could enter them manually, but that would take a lot of time. Hymnary.org lists 6,523 hymnals in their database, and they don’t index every hymnal in existence. I would like to create a batch process to save time. One option is to use the data from Hymnary.org. There I can do a browse search for all of the hymnals in their collection ([https://hymnary.org/search?qu=%20in%3Ahymnals](https://hymnary.org/search?qu=%20in%3Ahymnals)), download the data in .csv format, and upload the data to Wikidata in a batch using the data tool OpenRefine ([https://openrefine.org/](https://openrefine.org/)). However, for this approach to work for connecting the Wikidata items to library materials, the data needs to include OCLC identifiers, and Hymnary does not include that data. When I’ve manually looked up hymnal titles from Hymnary in OCLC, the data often doesn’t match, so I have to sort out that problem.
However, this does not mean that library connections aren’t already happening. A good example is the hymn tune ANTIOCH, composed by Lowell Mason. Some may be more familiar with it as the tune for the Christmas carol “Joy to the World.” Image 20 shows the Wikidata item for ANTIOCH displayed in Reasonator (https://reasonator.toolforge.org/?&q=60463566), a data visualization tool related to Wikidata.

The top section highlighted in Image 20 shows reference sources that describe the tune. Somebody wanting to know more about ANTIOCH could look it up in Reasonator and find library-owned reference sources that would tell them more about that tune, if the data is recorded in Wikidata. The section shown in Image 21 lists some hymn books in which the tune is published.
The section highlighted in Image 22 shows that this tune can be found in the Yale University Library Digital Collections.

If you look that up in the Wikidata item for ANTIOCH, you will be directed to the appropriate page in the Yale University Library Digital Collections, which has a digital scan of the tune from a copy of Lowell Mason's *Occasional Psalm and Hymn Tunes* (1837): [http://digital.library.yale.edu/digital/collection/1027_2/id/28](http://digital.library.yale.edu/digital/collection/1027_2/id/28). The “External sources” section shown in Image 23 links to other identifiers, two of which direct to library data.
These identifiers are important for connecting this Wikidata work to discovery of library resources. As part of the Linked Data for Production project, Cornell University has created a prototype in their discovery layer, Blacklight, that uses data from Wikidata to enhance library catalog records for classical music (Clarkson et al. 2023). Their work is reliant on the linking of these library identifiers with Wikidata identifiers to pull the data together. This tells me that the hymn data I’m creating in Wikidata could be usable in similar ways.

CONCLUSION

Though the project started as a small part of the PCC Wikidata Pilot, it has continued past the pilot to become its own project: WikiProject Christian Hymns (https://www.wikidata.org/wiki/Wikidata:WikiProject_Christian_Hymns). It currently includes 1,364 hymn writers, composers, and translators; 1,901 hymn tunes; 895 hymn texts; and 570 hymn instances. Even more valuable than this is the complex web of relationships built between these items, a growing knowledge graph that has the potential to connect researchers of Christian hymnody with library resources in new ways. There is much work to be done before this can be truly valuable for those who are interested in global Christian hymnody. Among other things, much more data needs to be added and quality assessment processes need to be developed, but a solid foundation has been laid for future work.

REFERENCES

Clarkson, Emma, Steven Folsom, Huda Khan, and Astrid Usong. 2023. “Linked Data via Vivaldi: A Musical Journey from Wikidata to the Library Catalog.” Presented at the LD4


