

Linking the Art in the Christian Tradition Database to Wikidata

Charlotte Y. Lew, Digital Scholarship and Special Projects Coordinator, Vanderbilt Heard Libraries

Steven J. Baskauf, Data Science and Data Curation Specialist, Vanderbilt Heard Libraries

ABSTRACT The Art in the Christian Tradition image collection and database was developed to support the Vanderbilt Divinity School's Revised Common Lectionary website, but it has become an important source of images in its own right. To improve discoverability, we started a project to create Wikidata items for all works in the collection, with the goal of linking those items to as many images in Wikimedia Commons as possible. We describe several challenges we faced in the early stages of the project. We provide an introduction to creating an item and refer to some useful tools for learning how to edit Wikidata and to perform bulk uploads. We end by describing how features of Wikidata can make artworks more discoverable and how we hope to improve the quality of our image metadata in the future.

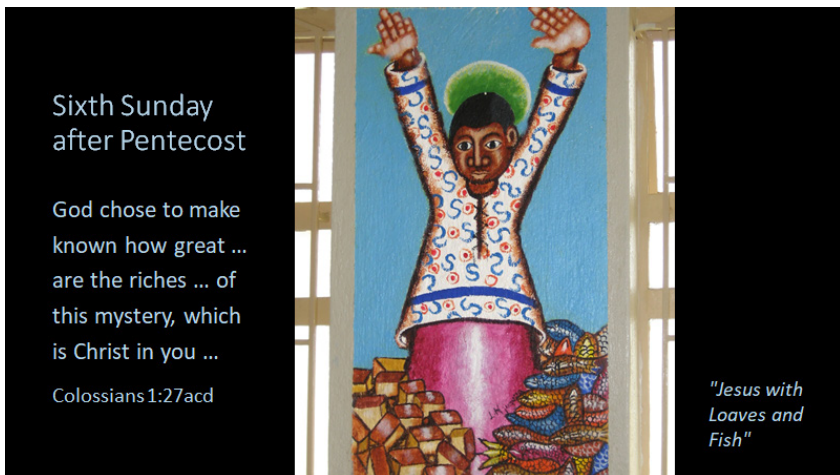


IMAGE 1 "Black Jesus." An ACT image and its associated RCL scripture. Image by klndonnely CC BY from <https://www.flickr.com/photos/genvessel/537739978/> via <https://diglib.library.vanderbilt.edu/act-imagelink.pl?RC=59089> [LinkingArtChristianImage1.png](#)

THE INTERWOVEN DESIGN OF ACT AND RCL

The scripture above was taken from the Vanderbilt Revised Common Lectionary website, abbreviated as RCL. The artwork matched with the scripture was retrieved from the Art in the Christian Tradition database, abbreviated as ACT. It manifests how the matched ACT art image works with the biblical scripture in the RCL.

The RCL website was established by Anne Richardson, the Special Projects Librarian of Vanderbilt Divinity Library, in 1995. She envisioned embedding religious art in weekly biblical readings when the project was first launched. The combination of RCL and ACT offers many tools for worldwide ministry and personal reflection. RCL and ACT pride themselves on celebrating diversity in ethnicity, practice, and social location through these interwoven aspects of Christian spirituality. RCL with ACT collection is the second-most visited website at Vanderbilt University. Moreover, it served more than a million global users in 2021.

ACT was developed for congregational and educational use, to enable leaders of worship and teachers to easily find visual art resources related to the scripture readings for each Sunday in the church calendar. ACT offers more than seven thousand images taken from street art, mosaics, frescoes, manuscripts, sculptures, architecture, and paintings. To optimize the search results, the image data are structured to be searchable by keyword, scripture reference, iconographic content, personal name, time period, and geographical location. Most of the images are freely available for use with an attribution requirement.

WHY LINK THE ACT DATABASE WITH WIKIDATA?

What is Wikidata? Why did we use this interface to feature the ACT image database? Based on the Wikidata homepage (<https://www.wikidata.org/>), “Wikidata is a free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, ... and to anyone in the world.” In other words, Wikidata is a central hub of linked data for all the Wikimedia projects. It is a multilingual platform for presenting and sharing information globally. Its impacts on research, metadata production, and collection visibility are immense. As is well known, open-access repositories in STEM disciplines are much more robust than those in the humanities. Therefore, increasing access to religion and theology collections on Wikidata is vital to researchers, instructors, students, and professionals in the theological schools.

PROGRESS REPORT

Linking the ACT image database with Wikidata is one of several Vanderbilt Wikidata projects. Before the ACT/Wikidata project was initiated, Steve Baskauf had contributed to transposing data of 6,800 images from the Vanderbilt Art Gallery to Wikidata. In the spring of 2021, Anne, Steve, and I (Charlotte Lew) started the project that aspired to elevate the reputation and increase the accessibility of the ACT database. The project site is here: [https://www.wikidata.org/wiki/Wikidata:WikiProject_Art_in_the_Christian_Tradition_\(ACT\)](https://www.wikidata.org/wiki/Wikidata:WikiProject_Art_in_the_Christian_Tradition_(ACT)). We began by setting the project's goal and generating four phases of an implementation plan. We use GitHub to record our progress and to create tickets for the encountered issues. Seeing a list of problems crossed out when solutions are delivered is fulfilling.

In phase one, we focused on images that exist in Wikimedia Commons and Wikidata, but that also have records in the ACT. Python scripts were utilized to retrieve around 3,300 images that fell into this category. We petitioned the Wikidata Foundation for a new ID for the ACT image collection and were assigned P9092. The newly appointed ID was used to link the existing Wikidata items to their corresponding records in the ACT database.

In phase two, we aimed to create new Wikidata items for the images existing in the ACT and Commons, which don't have Wikidata items yet. Despite some unsolved issues, we decided to value progress over perfection. As a result, phase two was completed in May 2022. Combining phases one and two, more than 5,000 ACT image records have been automatically imported to Wikidata. This couldn't have been done without Steve's programming talents. However, manually cleaning up and enhancing the data in advance is critical to making mass transposition possible. The tasks that require human intervention before automation materializes are illustrated below.

TASKS DEMANDING HUMAN INTERVENTION

Preventing Duplication

Inevitably, multiple images of the same work exist in Wikimedia Commons, and multiple records describing the same image exist in Wikidata. How to discover and disentangle the duplication can

be a challenge. We discretely analyzed the complicated duplication and learned that multiple Wikidata items had the same label, but their item types differed. For example, “The Good Samaritan” can be an artwork painted by Rembrandt or a biblical parable. We also noticed that some Commons images came with various versions, such as black and white or cropped. Correctly adding the ACT ID to a Wikidata item cannot be done without taking great discernment among duplicates.

Reconciling Inconsistent Data

After Python scripts scraped data from ACT and Commons, we faced the challenge of inconsistencies in the data, such as the inception date or creator. The creation date of an artwork fluctuates depending on the source of the data. The precise date is preferable, but a broader range of years can be acceptable. It is common for an image uploader or photographer to mistakenly use their name as an artist’s name, especially when they don’t know who created the art. Numerous artworks in the ACT database were created by unknown artists. Although several world-renowned museums and galleries provide reliable resources for verifying these inquiries, it takes time to excavate the accurate information.

WIKIDATA PROPERTY	VALUE	ACT PROPERTY	COMMONS PROPERTY
INSTANCE OF	P31	object/function	object type
IMAGE	P18	image	image
TITLE	P1476	title	title
CREATOR	P170	artist	artist
INCEPTION	P571	date	date
COUNTRY OF ORIGIN	P495	country of origin	country of origin
COUNTRY	P17	country	country
LOCATION	P276	current location, city/town	current location
COLLECTION	P357	collection	collection
GENRE	P242	genre	genre
MAIN SUBJECT	P84	subject	subject
MADE FROM MATERIAL	P306	material	medium
HEIGHT	P2048	height	dimensions
WIDTH	P2049	width	dimensions

TABLE 1 Properties used with ACT artworks in Wikidata

Determining Appropriate Properties




Identifying key properties to associate with item types is essential. The data models of ACT and Commons are different. We agreed to prioritize the ACT properties which better fit our needs. The table above indicates the list of potential properties to feature when creating a new Wikidata item. We know pairing scriptures with the artwork fulfills the goal of ACT design. Therefore, we are determined to find the suitable properties to include scriptures. After exploring several options, we discovered the properties of “based on” and “section, verse, paragraph, or clause” for the scriptures.

CREATING A NEW WIKIDATA ITEM

Wikidata, a well-structured database, offers an intuitive workflow for the users to create or modify data. The workflow is arranged in a linear movement, from labeling items with descriptions to building statements and then providing identifiers. Each addition shows a drop-down of options for users to pick as the following property to fill.

Art in the Christian Tradition



Download
Thumbnail image: 
Download
Medium image: 
Download Large
image: 

Title: Christ Walks on Water and Rescues Peter
Date: 20th century
Building: [St. Peter's Church](#)
Object/Function: Mosaic
City/Town: [Hamburg](#)
Country: [Germany](#)

Scripture: [John 6:1-21](#)
[Luke 9:51-62](#)
Person as [Jesus Christ \(Biblical figure\)](#)
Subject: [Peter, the Apostle \(Biblical figure\)](#)

IMAGE 2 Example from an ACT webpage. <https://diglib.library.vanderbilt.edu/act-imagelink.pl?RC=56964> – [LinkingArtChristian_Image2.png](#)



Summary [\[edit \]](#)

Description	Deutsch: St. Peter (Hamburg-Groß Borstel), Tympanon über dem Eingang English: St. Peter (Hamburg-Groß Borstel), Tympanon
Date	13 April 2012

IMAGE 3 Example from a Wikimedia Commons page. Image by Dirtsc CC BY-SA https://commons.wikimedia.org/wiki/File:Spgb_tympanon.jpg

The image used for demonstration is a mosaic titled “Christ walks on water and rescues Peter.” Comparing the data displayed in the ACT and Commons, we find the title, item type, date, location, and scriptures from the ACT. But the artist’s information is missing. The data included in Commons are relatively less helpful. “April 13, 2012” is inappropriate information because the date should be when the art piece was created rather than when the image was uploaded. The listed author, “Dirtsc,” is not the artist’s name but the username of the image uploader. After comparison, it is more effective and accurate to create a new item mainly based on the data we retrieve from the ACT database.

PROCEDURE FOR CREATING A NEW WIKIDATA ITEM FOR THE ACT IMAGE

See Wikidata item of “Christ Walks on Water and Rescues Peter” (<https://www.wikidata.org/wiki/Q111842119>) for details.

- Log into Wikidata if an account is already set up. Without a personal account, the editing won’t be attributed to the editor.
- To avoid duplication, enter the artwork’s title in the search box. The indication of “no match was found” provides the green light to create a new item.
- Click on “create a new item” under the Wikidata icon.

- “en” in the language field means the title is in English. Copy the title from the ACT database and paste it into the label field. Add item type and artist’s name in the description field. Because the artist’s name is unavailable, “mosaic by the artist unknown” was keyed in.
- After the item is created, move on to build statements. The first required property, “instance of,” is to specify the item type. The artwork is a mosaic. For the image property, insert the Commons image URL. This action makes the item linkable to the Commons. Make sure to assign a language to go with the title accordingly. If the information is available, add other statements, such as inception and location. The “based on” property and “section, verse, paragraph, or clause,” qualifiers are employed to indicate books and verses for biblical scripture. “Creator” is an important property to include, but the needed information is missing. The creator can be added when the data is uncovered. To provide references to all the statements, the ACT image page URL is inserted as the reference URL and specifies the retrieved date afterward.
- After adding statements, the Art in the Christian Tradition ID P9092 is used as an identifier. After the record number is entered, click the record number can link up the item with the ACT database.

When a statement is added, the “item not found” warning appears when the entry for a property or value is not an existing item in Wikidata. To ensure data are linkable, a new item must be created. However, don’t sweat over the incompleteness of item statements. Including the basic properties can be an easy start for beginners. Don’t forget that Wikidata values the effort of global collaboration to enrich its dataset. Items that seem presently incomplete can be enhanced by other editors later.

TRAINING AND COLLABORATION

A good tidings to share with whoever is interested in showcasing one’s institutional publications or special collections on Wikidata: To lower the barriers so more people can join the Wikidata editing force, the Vanderbilt Wikidata project team created a training tutorial called *Learn Wikidata* (<https://www.learnwikidata.net/>). This peer-led course provides training that ranges from the basic skills of creating a user account and editing items to more advanced techniques such as add-

ing gadgets to make editing more efficient. The tutorial comprises 20 learning topics individually featured in brief animation videos. The program is available in English, Mandarin Chinese, and Spanish. It also comes with closed captions to facilitate learning. To address the needs of librarians with various levels of experience with Wikimedia projects, Learn Wikidata offers a self-paced environment for users to customize their own learning pathways.

Wikidata provides a free platform for seminary libraries to publicize their institutional scholarship without having to invest in building an expensive and complicated scholarly communications infrastructure. To develop a community of practice around Wikidata, more theological librarians should be encouraged and empowered to participate in these critically important projects. The Wikidata religion and theology community needs more partnerships to foster global scholarly communications.

SPEEDING UP THE EDITING PROCESS WITH TOOLS

Although the Wikidata web interface is very easy to use, it is slow and somewhat tedious if you have a lot of data that you want to upload. Fortunately, there are several freely available tools that you can use to speed up the process. Two of them are well known: QuickStatements and OpenRefine. A less well-known tool, VanderBot, is a Python script that Steve wrote to upload CSV data to Wikidata.

QuickStatements (<https://www.wikidata.org/wiki/Help:QuickStatements>) is often used together with other tools that can output data into its format. For example, the commonly used reference manager Zotero can export data about publications in QuickStatements format so that it can then be uploaded en masse to Wikidata.

OpenRefine (<https://www.wikidata.org/wiki/Wikidata:Tools/OpenRefine>) is a more powerful tool for bulk upload. There is a greater learning curve to using it than QuickStatements, but for existing OpenRefine users it is a good option, since Wikidata support is built in. It is necessary to set up a schema in order for OpenRefine to map your data in such a way that the Wikidata API (application programming interface) can understand it. However, once set up, OpenRefine will help you reconcile your data against Wikidata items that already exist, preventing duplication. Christa Strickler described how she used OpenRefine in a project to add theological publications to Wikidata (Strickler 2021, 311-12).

For the ACT project, we chose to use the VanderBot script (<https://github.com/HeardLibrary/linked-data/blob/master/vanderbot/README.md>) to perform the data upload. It is somewhat less user-friendly than the other tools because it requires installing Python and running the script from the command line. However, since its data source is a simple CSV spreadsheet, once it is set up it is relatively easy to enter the data by hand, or to generate the CSV using another script. It also records detailed information identifying each statement and reference that is created, making it possible to change or delete the statement at a future time (Baskauf and Baskauf 2021). In our case, we used a series of Python scripts to combine data from the ACT database and Wikimedia Commons, clean and disambiguate, and generate appropriate references. The output of this series of scripts could then be fed directly into VanderBot to do the upload. Because the data were in the form of a spreadsheet prior to uploading, we could use a regular spreadsheet editor to proofread and to simply copy and paste values if we wanted to make any final changes.

As we mentioned previously, one challenge is to avoid creating duplicates of items that already exist in Wikidata. One of the Python scripts we used during processing searched for possible duplicates of a work by using the Wikidata Query Service to download the labels of all works made by the artist of that work, and then perform fuzzy string matching to identify any potential existing works that might be a duplicate. The final decision was made by a human who reviewed the matches to decide if the work should be excluded from the upload. Code used in the project is available in GitHub at <https://github.com/HeardLibrary/linked-data/tree/master/act>.

For most users, OpenRefine seems to have the best balance of adaptability with moderate difficulty. However, if you are comfortable working with code, Python scripts provide a powerful and flexible way to increase the efficiency of preparing and uploading data. We were able to create over 1,250 new artwork items for ACT works using this workflow.

FEATURES OF WIKIDATA THAT IMPROVE DISCOVERABILITY

A major motivation behind creating Wikidata items for ACT artworks was to improve their discoverability. There are three features that improve discoverability that are built in to Wikidata: aliases, multi-lingual capabilities, and querying.

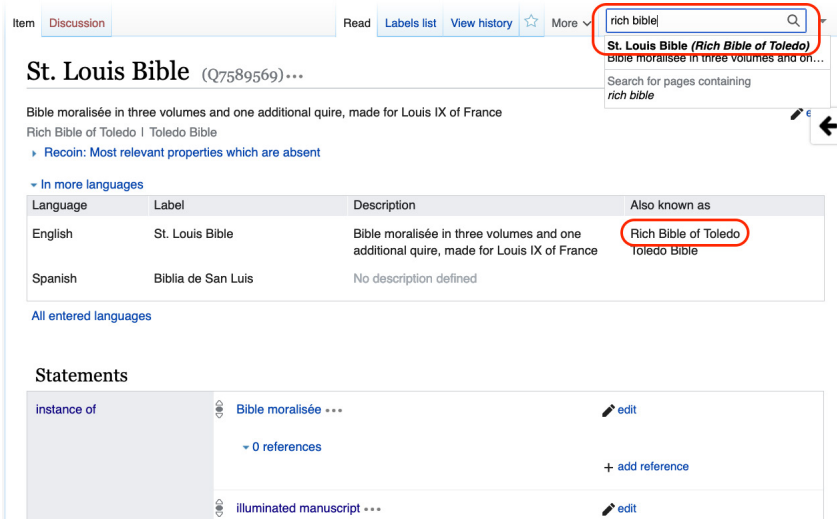


IMAGE 4 Searching Wikidata using an alias. <https://www.wikidata.org/wiki/Q7589569>.

When an item is created in Wikidata, it can be given a label and a description. However, it is also possible to list other possible labels that people may use for that item. These alternative labels are called “aliases.” For example, the St. Louis Bible (<https://www.wikidata.org/wiki/Q7589569>) has also been given the aliases “Rich Bible of Toledo” and “Toledo Bible”. Entering any of these terms in the search box allows the item to be found.

Tobias and the Angel (Q1635039)...

painting by the workshop of Andrea del Verrocchio

↳ [Recoin: Most relevant properties which are absent](#)

↳ In more languages

Language	Label	Description	Also known as
English	Tobias and the Angel	painting by the workshop of Andrea del Verrocchio	
Spanish	Tobías y el ángel (Verrocchio)	cuadro de Andrea del Verrocchio	
Egyptian Arabic	توبياس والملك	No description defined	
Asturian	No label defined	pintura de anónimu	
Breton	No label defined	taolenn dizanv	
Catalan	Tobies i l'àngel	quadre d'anònim	
Czech	Tobiáš a anděl	No description defined	
Danish	Tobias og englen	maleri fra Andrea del Verrocchios værksted	
German	Tobias und der Engel	Gemälde aus der Werkstatt von Andrea del Verrocchio	
British English	No label defined	painting by anonymous	
Esperanto	No label defined	pentraĵo de anonima aŭtoro	
French	Tobie et l'Ange	peinture de l'atelier de Andrea del Verrocchio	Tobias et l'Ange
Galician	No label defined	pintura de anónimo	
Hungarian	No label defined	anonymus festménye	
Ido	No label defined	pikturo da anonymous	
Italian	Tobiolo e l'angelo	pittura della bottega di Andrea del Verrocchio	
Japanese	トビアスと天使	No description defined	
Malayalam	തോബിയായുടെ ആന്റണി എഞ്ചൽ (വെറോച്ചിയോ)	No description defined	

IMAGE 5 Multilingual labels for a Wikidata artwork item. <https://www.wikidata.org/wiki/Q1635039>.

One of the key features of Wikidata is that it was designed to have multilingual capabilities right out of the box—multilingual labels are a core part of the model that underlies Wikidata. In the example above, the painting “Tobias and the Angel” (<https://www.wikidata.org/wiki/Q1635039>) is relatively famous and therefore has been labeled and described in many languages. The ability to create labels, descriptions, and aliases in many languages allows users around the world to find the works and to learn about them in their own languages.

This list is periodically updated by a bot. Manual changes to the list will be removed on the next update!

underlying query [Manually update list](#)

WDQS | [PetScan](#) | [TABernacle](#) | [Find images](#) | [Recent changes](#) | Query: `SELECT ?item WHERE { ?item wdt:P9092 ?id. ?item wdt:P31 wd:Q3305213. ?item wdt:P571 ?date. filter(?date < "1600-01-01T00:00:00Z"^^xsd:dateTime) }`





label	description	instance of	image	creator	inception
Mérode Altarpiece	15th-century painting by the workshop of Robert Campin	triplych reredos altarpiece painting		No/unknown value Robert Campin Workshop of Robert Campin Master of Flémalle	1430 1425
Tobias and the Angel	painting by the workshop of Andrea del Verocchio	painting		No/unknown value	140s
The Agony in the Garden	painting by El Greco or his studio	painting		No/unknown value El Greco	150s
Small Round Pietà	painting by an anonymous artist	painting		No/unknown value	14th century 1390

IMAGE 6 Art in the Christian Tradition Paintings before 1600 webpage. Example of a page built from a SPARQL query using the tool Listeria. [https://www.wikidata.org/wiki/Wikidata:WikiProject_Art_in_the_Christian_Tradition_\(ACT\)/Paintings_before_1600](https://www.wikidata.org/wiki/Wikidata:WikiProject_Art_in_the_Christian_Tradition_(ACT)/Paintings_before_1600).

The most powerful tool that is built into Wikidata is the ability to query its data. Because Wikidata is at its core structured and machine-readable, it is very well suited for querying. Some web pages, such as this list of ACT pre-1600 paintings ([https://www.wikidata.org/wiki/Wikidata:WikiProject_Art_in_the_Christian_Tradition_\(ACT\)/Paintings_before_1600](https://www.wikidata.org/wiki/Wikidata:WikiProject_Art_in_the_Christian_Tradition_(ACT)/Paintings_before_1600)) are actually generated based on a query of data in Wikidata. This type of page is generated by a tool called Listeria (<https://www.wikidata.org/wiki/Wikidata:Listeria>) and it is periodically updated automatically by a bot.

The underlying technology is a query language called SPARQL (<https://www.w3.org/TR/sparql11-query/>) that you can use to discover information in Wikidata. The Wikidata Query Service interface (<https://query.wikidata.org/>) can be used to construct and test queries. It also comes with several built-in visualizations that are useful with artworks, such as an image gallery and a timeline. These “canned” visualizations can be embedded in web pages, where they are updated live each time the page is loaded.

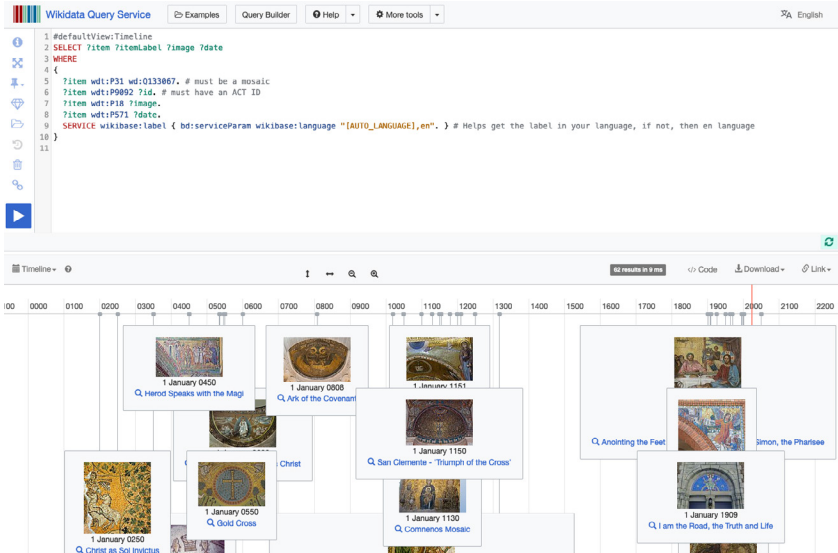


IMAGE 7 Timeline of mosaics in the Art in the Christian Tradition database. Example of a visualization built using a SPARQL query. Link to conduct the query: <https://wiki/SEHm>.

It is possible to perform the SPARQL queries “under the hood” as part of custom Javascript code, then to use the retrieved data to build custom web pages that display the data. A good example of this is the University of Edinburgh’s website that seeks to humanize the victims of witchcraft trials (<https://witches.is.ed.ac.uk/>). In addition to using queries to generate visualizations, some visualizations in that website provide links to the Wikidata pages about the accused, allowing for more detailed exploration of their lives.

FUTURE ENHANCEMENTS AND VALUE ADDED

We have additional metadata available from the ACT database that we have not yet added to the Wikidata items for the artworks. This includes the material medium of the artwork, scriptural references associated with the artwork, the subject, and the icon class. Adding values for these metadata properties can improve search capabilities and make it possible to create custom search pages that use these values to direct users to appropriate artworks through SPARQL queries incorporated in the page Javascript.

Since Wikidata is an open knowledge graph, we can build on the previous work of others, and others can improve the metadata about

items we have created. This is potentially a big benefit if your project is being carried out by a small group with limited resources.

REFERENCES

- Baskauf, Steven J. and Jessica K. Baskauf. 2021. "Using the W3C Generating RDF from Tabular Data on the Web Recommendation to Manage Small Wikidata Datasets." *Semantic Web Journal*. <https://doi.org/10.3233/SW-210443>
- Strickler, Christa. 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: Seventy-fifth Annual Conference of Atla*: 301-14. <https://doi.org/10.31046/proceedings.2021.2978>