The uses of Wikidata for galleries, libraries, archives and museums and its place in the digital humanities

Stacey Cook

As open source, collaborative and crowd-sourced media become increasingly popular, it is essential that the community of galleries, libraries, archives and museums (GLAMs) take a central role within these digitally based projects. This article looks specifically at Wikimedia's Wikidata project and provides an overview of Wikidata's various uses for GLAMs. It gives a brief outline of where Wikidata fits into broader digital humanities scholarship and how it can contribute to digital humanities projects, whether within academia, or through public engagement. Finally, it provides examples of some potential digital humanities projects that GLAMs can explore to foster further engagement with digital scholarship and the larger Wikimedia community, as well as finding increasingly creative, informative, innovative and meaningful ways to disseminate and present the reliable data and source information which they hold.

Introduction

With the growing popularity of open source, collaborative and crowd-sourced media, it is essential that the community of Galleries, Libraries, Archives and Museums (GLAMs) are central to these digitally based projects. GLAMs are uniquely qualified to provide reliable data and source information for a large portion of what is found on Wikimedia sites, such as census data (for the exact population of a city), artwork (for the dates and locations where pieces of art were created and the materials used to make them) and historical events (for contextualization through authentic images or journal entries). By having this information, the community of users on Wikimedia can incorporate it into their cultural and digital projects, making this information more accessible to the public, and able to be used in new and creative ways by GLAMs.

It is important to have an understanding of the growing scholarship of the digital humanities, and what is often referred to as the largest digital humanities project of all time, Wikipedia.[1] However, as things shift and are constantly updated in the digital world, Wikipedia is slowly becoming outdated while a newer project – Wikidata – moves into the spotlight. Wikidata offers a more connected experience for the user and a more accessible source for

Stacey Cook is an MA candidate in Anthropology and the Digital Humanities at Carleton University, Canada. She had the opportunity to work under the Wikipedian-in-Residence at Library and Archives Canada and helped to host its first Wikipedia Edit-A-Thon. She continues to be interested in the open source movement within academia.

1 FLORES Pepe, 'Is Wikipedia the largest-ever digital humanities project? Exploring an emerging relationship,' last modified 17 Aug. 2016, at https://blog.wikimedia.org/2016/08/17/wikipedia-largest-digital-humanities-project.

https://doi.org/10.3828/comma.2017.2.12

data that aims to capture the sum of all knowledge, make it available in every language, and be a completely accessible database for everyone (human or not) to scrape data from. This development offers new opportunities for GLAMs to contribute and use information found in Wikidata, adding depth to our own cultural projects and digital scholarship.

This article was originally written as part of a digital humanities student practicum at Libraries and Archives Canada (LAC) in Ottawa, Ontario. As a result, LAC is central to the examples discussed. I hope that other archivists can use its insights to better understand how they, too, can pursue digital humanities projects and use Wikidata to incorporate the Wikimedia community into their own institutions.

What is Wikidata?

Wikidata is an ongoing project to create and store linked, dynamic data, data which is structured in a way that makes it easily understood by computers, and which, through the Wikidata database, can be accessed, and used, by anyone.[2] It works in a similar way to Wikimedia Commons, a depository for images, videos and soundclips where many GLAMs already upload media that can be used to edit other Wikimedia projects. Wikidata is the repository for this linked data, aiming to keep it centrally stored so that each Wikipedia page or Wikimedia project no longer has to be maintained separately. It aims to be:[3]

- free
- ·collaborative
- multilingual;

and to act as

- a secondary database, which
- collects structured data, and
- supports Wikipedia and other Wikimedia projects

How does Wikidata work?

Central to an understanding of Wikidata is its collection of data into a defined structure. This structure is invaluable as it allows data to be processed in many languages and by humans, bots and other computer programs. Users enter information using pre-defined fields: "Items", "Labels", "Properties", "Values", "Descriptions" and "Aliases." Let's use Ottawa as an example of how this works in practice:[4]

- **Items** represent all the things in human knowledge (e.g., people, concepts, objects and ideas). Our Item is Ottawa.
- Each item has a **Label**, which will be the Item's title. It should be the most common name of the Item and the shortest version, so in our example the Label would be "Ottawa" rather than "The city of Ottawa" or "The capital city of Canada."

2 Wikidata project page, 'Introduction' at https://www.wikidata.org/wiki/Wikidata:Introduction.
3 VRANDECIC Denny, 'The rise of Wikidata' in *IEEE Intelligent Systems*, 28, no. 4 (July 2013), p. 90.
4 And see Wikidata, 'Ottawa (Q1930)' at https://www.wikidata.org/wiki/Q1930

- Each label has a unique identifier to distinguish it from other items with the same Label, typically a **Q** followed by a few numbers: Ottawa is Q1930.
- Each item has a **Description** to distinguish the item from others, so a description for Ottawa could be "capital city of Canada."
- **Aliases** are used to identify any other names for the item, and can link these names to the item. One Alias for Ottawa is Bytown.
- **Properties** and **Values** work together to create a Statement, and this is where data becomes central. Properties of Ottawa could be its location, population, mayor, or even website. The corresponding Values of these properties would be Ontario, 934,243, Jim Watson, ottawa.ca. Together, these Properties and Values make Statements about Ottawa: it is located in Ontario, its population is 934,243 people, its mayor is Jim Watson, and its website is ottawa.ca.

How can GLAMs use Wikidata?

Many GLAMs are taking steps to incorporate Wikipedia into their work, from uploading media files to the Wikimedia Commons, to hosting edit-a-thons for the community to come together and work with their materials, or even hiring Wikipedians-in-residence, resident experts who can help coach institutions on how best to work with the Wikimedia community to make their collections available online. Wikidata is an excellent tool to further these initiatives and is increasingly being applied due to the ease with which collections and information about them can be made more widely available and used by a global community.

Like Wikimedia Commons, Wikidata has a portal to upload data. This could be as simple as importing census figures; genealogical data such as births, deaths and military service records; or updating the "Library and Archives Canada" Item page[5] with data about its specific number of holdings. External databases and websites can be linked on an item page; the LAC page links to its Twitter account, and could similarly add a Facebook page, Wikipedia page, or even access to podcasts or blogs. At the time of writing, the LAC item page had no links to any other Wikimedia projects other than the Wikimedia Commons page, so this could be one opportunity for growth. We could even consider a community Wikimedia-themed scavenger hunt, where users find instances of LAC data among Wikimedia Projects and link these to the Item page, creating more accessibility and traffic to the Commons and through this, to the physical holdings.

One key aspect of Wikidata is its ability to link Properties and Items. Specifically, each Item has its own page which can be linked to other relevant Wikimedia projects. Wikimedia Projects are vast and diverse, but anyone can start a Project, or contribute to those already existing. One particularly extensive Project is "Wikidata: Wikiproject sum of all paintings."[6] The Project aims to get an item for all "notable" paintings that are on display in museums or were created by a "notable" artist (and Wikidata Notability criteria are available to aid selection).[7] If GLAMs wanted help with particular Wikimedia Projects – translations, for example, or adding descriptions to images – those Wikimedia Projects could then be linked

5 Wikidata item page, 'Libraries and Archives Canada (Q913250)' at https://www.wikidata.org/wiki/Q913250

6 Wikidata project page, 'Wikidata:WikiProject sum of all paintings' at https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings

7 Wikidata project page, 'Notability' at https://www.wikidata.org/wiki/Wikidata:Notability

to their institution's item page. By utilizing these links, institutions can draw more attention to various projects and help users find information that is of value for their own Projects; GLAMs could consider running link-a-thons – similar to edit-a-thons – structured around linking data and connecting it to Wikimedia Projects.

Through Wikidata, GLAMs can also contribute to user and community engagement, enhancing accessibility and engagement with these holdings; in particular it can help provide sources relevant to what users are interacting with. It is easy to track what is being used from dumps to the Wikidata and this data provides insight about which topics might be most productive to upload content about. Everything in Wikidata has a specific query code, which can be searched for using the Wikidata Query Service (WDQS). From a research standpoint, this service is valuable for understanding how people use data, where in the world the data is being used, and who it is connecting to. Query search results always reflect what has been added to, deleted from, and edited within Wikidata, and what information has been uploaded by users and other GLAM institutions around the world. This provides insight into what contributors consider important to upload, what information is available and how information changes over time. By better understanding the flow of data and digital materials, we can, for example, better assess what is essential to digitize and what will contribute to more global digital projects.

Critique on cultural interpretivism

The gathering of data always involves a form of cultural interpretation. We know the limits of the Wiki community: that it is a completely community-run project, reliant on volunteer editors for its content. Its pages are also predominantly English-language, a further acknowledged limitation. But we must also be continually critical of Wikidata and the very data points we input from the perspective of their cultural significance. The Items, Labels and Values that Wikidata relies on to create its structured data are all culturally significant because they form institutional boundaries around each data point. This means that data has to fit into a certain Item category which has already been established by Wikidata. Creating new Items is a process that must be approved by Wikidata, and new Items must fit certain criteria.[8] By categorizing data in this way, we institutionalize it into how Wikidata believes data should be presented and disseminated. The query results themselves, also produced by Wikidata, are similarly bound to a certain method of understanding how data should be presented and worked with. While this can be helpful for certain Projects, those working in culturally sensitive data need to remain aware of how contributed data might be presented.

For example, LAC has a mission to incorporate more indigenous languages into Wikimedia, allowing these languages, as well as the information they provide, to be more accessible to the Wikimedia-using world. However, the "colonizing effect" of such contributions needs to be considered, given that these indigenous languages are being presented through a Western understanding of access to knowledge. While making more indigenous languages and knowledge available is an excellent way to contribute to ongoing reconciliation, and to give more visibility to disappearing languages, it is vital, at the same time to consider how that traditional knowledge has been passed down and disseminated, for

8 Wikidata special page, 'Create new Item' at https://www.wikidata.org/wiki/Special:NewItem

example through stories and oral histories. While Wikidata can make a great contribution by acting as a repository for these histories, we are reminded that in doing so, those histories pass through a Western data system and are categorized in accordance with a Western-centric way of thinking. Thus GLAMS should remain both careful and critical of the way some knowledge is presented on Wikipedia, recognizing its limitations as a tool. But by being creative, flexible and considerate, by working closely with archivists and cultural experts, and consulting community members as to how they see Wikimedia best protecting their data and communities, we can develop best practices for using these online tools to disseminate culturally significant data.

Wikidata and the digital humanities

As a discipline, digital humanities aims to make the social sciences and humanities more accessible, innovative and diverse. Methodologically, it aims to broaden the scope of traditional scholarly work through, for example, open source culture, a movement that promotes free and accessible tools and scholarship beyond the paywall of academic journals: data visualization, mapping, crowd-sourcing, maker culture, video games, massive open online courses (MOOCs) and more.[9] In doing so, digital humanities aims to change the ways in which scholarship is carried out in college and universities, and at the same time it allows other institutions to explore new avenues in how their data and holdings are interpreted.

In a similar thread, Wikipedia and the Wiki-movement are at the centre of what the digital humanities aim to be, and they are often considered the largest digital humanities project to date:[10] it fully encompasses open source culture and crowd-sourcing, and these allow other projects to be shaped around its metadata which can be used both directly by humans and by machine-to-machine transmission. Digital humanities scholars are always on the quest for more databases where they can access information and Wikidata could be a primary source; not only is all of its material is accessible but, significantly, it is all properly sourced so users can see where all of this material is coming from.

Many digital humanities projects have objectives similar to those of Wikidata: translating sources, digitizing archival records and books, scraping databases for information, collecting and listing material in various digital forms that can then be searched. SPARQL (pronounced "sparkle") is a digital humanities project behind-the-scenes of Wikidata which collects and models data, picking up patterns and reading descriptions, then presenting it as information. One key aspect of Wikidata, and one which provides a particularly interesting bridge between it and the digital humanities, is SPARQL's ability to create visualizations from its various data queries. It can create charts, maps and other visualizations, limiting the ongoing need to develop new tools for topic modelling. Further, by being able to run a search query from such a large database information becomes more readily available and accessible through that searchability.

While Wikidata is a huge database, its drawback to those working in the digital humanities is the lack of specificity, and of detailed information, in the material included. Many

9 See University of California, Los Angeles, 'The Digital Humanities Manifesto 2.0,' 2009, at http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/

10 FLORES, 'Is Wikipedia the largest-ever digital humanities project?'

scholars are working in depth in a very specific area and while Wikidata can help link them to relevant data, there is not as much material as most scholars would like, and expect, to use. Unfortunately, there appears to be a gap between the community of Wikimedians and digital humanities scholars; the latter are not using Wikimedia for many of their projects, while Wikimedians are focusing less on digital humanities projects and more on projects centred on gathering data and editing pages. There is a disconnect: while digital humanities scholars use many of the same tools that Wikidata uses, they do not also contribute their own findings to Wikidata.

There are a number of reasons for this. Wikimedia, and specifically Wikipedia, still has a bad reputation in scholarly circles as not being a reliable source. Students are discouraged from using Wikipedia as anything other than a starting point for research, and are also discouraged from citing Wikipedia as a source. However, the view that Wikimedia is not of use within scholarly contexts is becoming increasingly outdated. Recent research (2017) from the Massachusetts Institute of Technology,[11] while noting that Wikipedia's subject coverage matched more than 90% of undergraduate subjects studied at top-tier research universities, indicated that, more significantly, Wikipedia has the potential to shape science—and does—through its use of accessible and open data. Yet, while open data has far-reaching implications for academic research – it allows a wider spread of information, ease of citation, more peer review and collaboration – many scholars are not so ready to accept this open source policy. In her book *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*, Kathleen Fitzpatrick explains that traditional academic publishing is based on values of "individuality, originality, completeness and ownership," a set of principles which are in direct conflict with Wikipedia, as something more structured, co-authored, constantly being re-written and owned by everyone who chooses to use it. Accepting open source means changing the norms of academic publishing, and academic publication is a key determinant for universities in allocating funding, or appointing to tenured positions. Fitzpatrick successfully published the manuscript of *Planned Obsolescence* online for comments and review, and an example that can give us more hope that open source publishing could be the way of the future in academia.

What types of digital humanities projects can LAC pursue?

Wikidata is lacking in one major respect: while we can use data and filter it, there is a lack of analysis of how the data is used and this type of information is needed in order to create meaningful projects, ones which will be relevant to people, and answer their questions. This is where LAC comes in. How can we get people to access LAC's holdings? How can we better educate the public about history? What are the best practices to engage with people in meaningful ways? How can we make LAC's holdings public and accessible while maintaining its integrity as an institution? These questions, and the answers to them, must guide our use of data and inform future projects. LAC's use of the digital humanities is not limited to Wikidata: the possibilities are endless with its holdings. By making data digital, and accessible, we can make interactive and multi-layered digital maps of our cities, create historically accurate video games, dive into virtual reality, or even work on online courses

11 THOMPSON Neil and HANLEY Douglas, 'Science is shaped by Wikipedia: evidence from a randomized control trial' in *Social Science Research Network*, MIT Sloan Research Paper No. 5238-17 Sept. 2017.

for those interested in LAC's content. The following suggestions for possible projects at Library and Archives Canada, whether aimed at scholarship, as outreach to communities of researchers and scholars, or just for fun, could be mirrored within other GLAMs:

- **Digitizing content**: Many digital humanities projects are aimed at simply digitizing content, often around a theme – the works of Charles Dickens or artwork by Monet – and making it accessible online. While this is something LAC continues to work on, partnering with universities could be an excellent way to do it more swiftly. Possibilities include digitizing political speeches, or all of the artwork in the National Gallery, or of other famous Canadian artists like the Group of Seven.
- **Massive Open Online Courses (MOOCs)**: LAC could create a MOOC that is available to all. Its topic could be as simple as "how to do historical research" with lessons in Canadian history, or take a specific focus, whether on marginal groups, through indigenous histories, or on important women in Canada, etc. Adding digital humanities content and information about Wikimedia to the MOOC would get more people involved in community edit-a-thons, and scholarly research and partnerships.
- **Gamifying content:** Games can be an excellent way for people to get excited about learning history, or even about how to do archival research. Wikipedia has a space adventure-themed game that teaches new users how to edit, and there is a lot of potential for GLAMs to create smaller games which also teach some history, or techniques such as running search queries in a particular database. For instance, you might take part in a scenario in which, as someone with little knowledge about your past, you make inquiries about your family history at an archives As the genealogical query unfolds, you discover that you are related to someone famous or historically significant. "Digital Zombies" is an example of such a game. Developed at the University of California-Riverside its pedagogic aim is to teach students how "to navigate online and physical research spaces" but it does this through a game in which the player must avoid becoming a digital zombie, someone who loves the Web too much, that "it's rotting their brains."[12]
- **Virtual Reality (VR)**: This is a growing field with a lot of potential for public exhibits and GLAMs; its use can push the boundaries of "traditional" exhibit, and allow people to experience history in a new way. Imagine being immersed in a battle, or being able to see the building of certain landmarks through digitized historic photos used to create a VR experience. Already quite common are apps which link GPS data with archival data in real time as you walk around. These apps have a tourism as well as educational purpose, functioning for example as a walking tour providing both directions through the city and relevant images, news articles, and videos to your phone at specific landmarks. Tours could be created for different time periods, or created around themes – historic battles, architecture, taverns, etc.

Conclusion

The Wikimedia community is both wonderful and challenging. It requires ongoing learning, continuous research, innovation and community engagement. While still relatively new to

12 'Digital Zombies' at http://digitalzombies.ucr.edu/

the Wikimedia circles, Wikidata offers an innovative way to collect and disseminate information that remains true to the core tenets of digital humanities: it is free and open, collaborative and structured in such a way that is aimed at accessibility. There are many ways GLAMs can use Wikidata, from uploading their own data and databases online through the Wikidata portal, to linking social media and online content to Item pages, or focusing on community events such as edit-a-thons for linking data and entering material into Wikidata.

At the same time, as noted above, it is important to continue to look at how we approach data, and ensure that we are aware of the systemic privileging of Western knowledge systems, and the possibility that this may not be the most effective way of presenting all data. Wikidata, and Wikimedia in general, are the centre of what many in the digital humanities stand for: being open, collaborative and innovative. Here we have an opportunity for growth, and a change in some scholarly reticence about use of Wikidata. By using the Commons to show how many sources are available, and having more sources and data dumps directly from GLAMs and other recognized institutions, Wikidata can become a place full of trusted primary materials as well as a place where scholars find information linking them to trusted archival holdings. Lastly, GLAMs are not tied to Wikimedia in their pursuit of digital humanities projects, although this is a great place to start. By incorporating aspects of the digital humanities it is possible to be even more creative, informative, innovative and meaningful in the way we disseminate and present information.

Finally, here are some additional projects and resources to light the way for your own project:

- Bodleian Digital Libraries, "A Bodleian Libraries Blog." Martin Poulter, the Wikipedian in Residence at the Bodleian Libraries in Oxford. Poulter's blog discusses digitizing books, creating research from Wikimedia, as well as considering how the university and other public institutions can use Wikimedia. See https://blogs.bodleian.ox.ac.uk/digital/author/poulterm/
- "Why Digital Humanities + Wikimedia" is Wikimedia's own page of suggestions for better outreach with the digital humanities. See https://outreach.wikimedia.org/wiki/Why_Digital_Humanities_%2B_Wikimedia
- DH Commons Projects is a database of digital humanities projects. Some have links to access the project, others are just a description, but it is a great source for what DH people are doing and the diversity of projects in many disciplines. See https://dhcommons.org/projects[13]
- "London (Q84)" is an example of a "picture-perfect" Wikidata page. Described as "Wikidata – in pretty!" it shows the potential for all Wikidata Items, including maps, links to external sites, links to Wikimedia Commons photos etc. See https://tools.wmflabs.org/reasonator/?q=Q84.

13 Inaccessible, Oct. 2018.