

Wikidata

Brenna Wheeler

10/15/2019

Introduction

According to its own Wikidata entry, Wikidata began in 2012 as a “free knowledge database project hosted by Wikimedia and edited by volunteers” (“Wikidata” 2019). As with many projects under the Wikimedia Foundation auspices, Wikidata relies heavily upon these volunteers for a wide variety of tasks, including generating content, repairing vandalism, and organizing the entire Wikidata volunteer community. The fulfillment of these tasks by volunteers reveals a successful adaptation of Peer Production methods from the Free and Open Source Software community, specifically Commons-Based Peer Production (CBPP). For this paper, I will be examining Wikidata’s adaptation of Peer Production using the three characteristics of Peer Production established by Benkler (2016):

- “Harnessing diverse motivations”
- “Decentralization of conception and execution”
- “Separation of governance and management from property and contract”

The adaptation of Peer Production has created a large database of structured linked data, which anyone can access and use. Finally, I will be looking at Wikidata’s relationship with the famous Wikipedia, and how the two Peer Production platforms inform one another.

Harnessing Diverse Motivations

Wikidata harnesses a diverse set of motivations by allowing anyone to easily contribute to the project. As soon as a user is introduced to the homepage, they are immediately provided information about how to contribute under the “Get Involved” section, which links to additional information about Wikidata, tutorials for adding data, a guide to using data, and the community portal.

To get users familiar with creating and editing entries, Wikidata has created two tutorials: “Items” and “Statements”. These tutorials walk a user step-by-step through the process of adding data by allowing the user to edit the page for “Earth” in a sandbox environment. Each step has additional details about rules and best practice for properly logging data. By completing these tutorials, users can learn how to contribute information that is consistent and properly linked to the data currently available.

Once a user has completed the tutorials, Wikidata provides a series of WikiProjects to get new users involved in areas of the database in need of the most help. The first project linked on the Wikidata Introduction page is a project called “Labels and Descriptions”, which provides a large list of Identifiers that are missing labels and descriptions in different languages. Additional WikiProjects focus on specific topics, such as LGBT-related content or items related to Ancient Greece, to assist users with more specialized knowledge about these subjects to find and develop data in those areas.

For users with advanced coding skills, Wikidata also welcomes the contribution of bots to save community members time and effort. However, since bots can make a lot of changes in a very short amount of time, there are policies in place to limit the work they can do. All bots need to have a limit on the number of changes made per minute, and their edits need to be logged with a special bot flag, which can only be obtained from a bureaucrat user and the approval from the Wikidata community (“Wikidata:Bots” 2019).

All the methods of contribution discussed so far is mostly for individuals, but Wikidata also allows organizations to donate their large quantities of data to the database. To make a “Data Donation”, the organization must reach out to the Wikidata community to determine which data to add, and they must work with

the community to import the chosen data (“Wikidata:Data Donation” 2019). It is interesting to note that most organizations listed on Wikidata’s “Data Donation” page manage cultural heritage collections, such as UNESCO, the British Museum, and the National Libraries of France and Australia. This will be discussed further below, when I compare Wikidata with its sibling site, Wikipedia.

Decentralization of Conception and Execution

To discuss ideas and actions, the Wikidata Community has several options. The main project chat is hosted on Wikidata as an individual page, with a layout similar to the layout of a Wikipedia article. A user can post a new “discussion” by adding a section to the page, and responses are indented under the first post. After 7 days, all discussions are added to the Archive Index, which acts as an open repository for past problem solving and discussions (“Wikidata:Project Chat”, 2019).

In addition to the main project chat, Wikidata provides similar pages for special topics. Each governance role has its own “Noticeboard”, which allow other users to post a message for them to address and respond to. Additional pages for requests of queries, deletions, comments, bots, and permissions are also listed and routinely checked by users responsible for those areas. When certain actions need community approval, they are posted for discussion and voted for on the relevant page.

By hosting these “chats” on Wikidata, the site allows for full transparency of these discussions. Once a new post is added, all members of the Wikidata community can see it without needing to log in or verify credentials. Anyone can post a discussion, and anyone can respond to the problem. However, the users who can *fix* the problem may be limited to roles with the correct permissions, such as deleting problematic pages is limited to Administrators and changing permissions is limited to Bureaucrats.

I will also note here that Wikidata does have an option for direct communication. The community has a live chat, hosted on two different websites: Telegram and IRC. Unfortunately, the author had technical issues while trying to access these platforms, so the nature of these chat forums are unknown.

Separation of Governance from Property and Contract

According to Benkler (2016, pg. 2), this characteristic contains two important parts, each of which will be addressed below:

- Inputs and outputs are open commons or common property
- Governance utilizes combinations of participatory, meritocratic, and charismatic systems rather than proprietary or contractual models

Open Commons

This section can be addressed by looking at the copyright policies associated with Wikidata and the larger Wikimedia Foundation. According to “Wikidata:Introduction” (2018), all data in Wikidata is published under the *Creative Commons Public Domain Dedication 1.0*, so content on Wikidata can be used, copied, and distributed without paying fees or requesting permission. WikiMedia Terms and Conditions also require that “all submitted content be licensed so that it is freely reusable by anyone who cares to access it” (“Terms of Use” 2019). By establishing an open policy for submitted and presented content, Wikidata and the larger WikiMedia Foundation fulfill the first bullet listed above in ensuring that inputs and outputs are open commons.

Governance

To manage the community and the project, Wikidata has established five governance roles: administrators, bureaucrats, checkusers, oversight, and property creators. Each role handles a different aspect of managing the Wikidata community, from deleting unwanted pages to adding and taking away permissions from users.

This creates a decentralized leadership where the power is divided among different groups, rather than held in one area. The different roles are as follows:

- **Administrators:** oversee deleting pages, blocking users, monitoring for vandalism, monitoring for role abuse, and returning vandalized pages to their pre-vandalized forms; requires a week of discussion and a community vote with at least eight supporting votes and over 75% of total votes in support (“Wikidata:Administrators” 2019).
- **Bureaucrats:** can change permissions and rights for other users; requires a permissions request and at least 15 supporting votes in a community vote (with 80% of total votes in favor) after a week of discussion (“Wikidata:Bureaucrats, 2017).
- **CheckUsers:** can examine IP address information of users on Wikidata and block users in case of vandalism or “sock puppetry”, where one user creates multiple accounts for improper use; requires a permissions request and at least 25 supporting votes (with 80% support) after two weeks of discussion (“Wikidata:CheckUser”, 2019).
- **Oversight:** also called “suppression”; can hide and restore revisions, usernames in histories, and portions of log entries in a page’s editing history in cases of hiding non-public personal information, hiding potentially libelous information, removing copyright violations, and removing blatant attack names in editing histories; requires a permissions request and at least 25 support votes (with 80% support) after two weeks of discussion; rights is given by a bureaucrat formally requesting access from the Wikimedia Foundation (“Wikidata:Oversight” 2019).
- **Property Creators:** can create properties; requires a nomination from themselves or someone else; user needs to be a trusted member of the community and have shown a satisfactory understanding of how Wikidata works; an administrator can add the flag (“Wikidata:Property Creators”, 2019).

As listed above, four of these roles require a period of discussion and a voting process by the community members. The only exception is the Property Creators, whose role was created when the Wikidata community that decided only a select number of people should have technical ability to create Properties (apart from the administrators), but the new properties still require community consensus. By publicly posting their interest, discussing their merits, and voting on new members of each role, the Wikidata community fulfills Benkler’s “participatory, meritocratic, and charismatic” requirements mentioned earlier (Benkler 2016, pg. 2). However, Wikidata policies also limit the power of several roles. For example, the Bureaucrats policy explicitly state that their ability to grant permissions “should not be used for political control; to apply pressure on editors; or as a threat against another editor in a dispute. There must be a valid reason to check a user” (“Wikidata:CheckUser”, 2019). By dispersing the power and voting with the consensus of the community on the user’s merit, the governance is very similar to a Meritocracy model (Gardler and Hanganu, 2013).

Wikidata vs Wikipedia

Wikidata has a strong relationship with Wikipedia. According to its Introduction page (2018), Wikidata supports a structured, multilingual data that can easily be reused by other Wikimedia projects. In a casual discussion about the subject, Liaison Librarian Gina Bastone, who organizes the Wikipedia Edit-a-thons hosted at the Perry-Castañeda Library, explained to me that in her experience, Wikipedia is often used to fill the content of Wikidata.

Such a close relationship could have an impact on the data bias within Wikidata. Already, there have been several studies done on the male-dominated demographics of Wikipedia (Glott, Schmidt, and Ghosh, 2010; Antin, Yee, Cheshire, and Nov, 2011). Another study by Konieczny and Klein (2018), shows that these demographics is one of the major factors involved in the lack of women and non-binary representation in Wikipedia biographies. Even if the gender bias for Wikidata is unknown (or at least, the author was unable to locate relevant data), it is still incredibly possible that the gender bias of Wikipedia is impacting the content of Wikidata.

The largest difference between Wikidata and Wikipedia is the structure of information, which in turn shapes the communities interested contributing to these databases. Wikidata allows a highly structured and efficient method of presenting data, which is making it more popular among academics studying information science, Natural Language Processing, and medical industries (Mora-Cantalops, Sánchez-Alonso, & García-Barriocanal, 2019). Wikidata’s structure has also caught the attention of Galleries, Libraries, Archives, and Museums (GLAMs).

Earlier, I mentioned that many the data organizations listed on Wikidata’s “Data Donations” page were cultural heritage GLAMs. Since 2010, GLAMs have been involved in contributing to WikiProjects, with Wikidata recently becoming the favored database (Stinson, Fauconnier, and Wyatt, 2018). For this paper, I interviewed with Paloma Graciani Picardo, the Metadata Librarian at the Harry Ransom Center, to discuss the GLAMs’ interest in Wikidata. According to her, “Libraries and archives see an opportunity to put their collections out there. They can use the ‘archived at’ [element] or, whenever you put something into Wikidata, you can add a source”. She also described that some librarians see Wikidata as a potential central hub of identifiers, where catalog records could point to for additional information, other relevant collections, and identifiers assigned by other organizations, such as the Library of Congress or WorldCat. A similar project had been started at the Harry Ransom Center, where she and a graduate student were reconciling local names in the HRC Metadata Taxonomies with data from Wikidata. However, there are still ongoing discussions about whether all the data in a collection is appropriate for Wikidata, or if Wikidata should focus in on major, widely-known concepts (Graciani Picardo, 2019; Stinson, Fauconnier, and Wyatt, 2018). Although Wikidata has a higher concentration of cultural heritage involvement, the involvement usually comes from wealthy GLAMs with highly educated staff from Europe and North America, so this continues to “perpetuating the dominance of colonialism-tainted collections and professional practices in retelling history”, as phrased by Stinson, Fauconnier, and Wyatt (2018). Cultural heritage institutions in Ghana, Indonesia, and Argentina are working to upload their collections, which would help balance out the bias, but there are still many challenges that need to be addressed (Stinson, Fauconnier, and Wyatt, 2018).

Conclusion

Wikidata relies heavily upon the successful implementation of peer production policies in order to create a large database of structured data. It harnesses diverse motivations by allowing contributions to be made easily by anyone. It decentralizes conception and execution of ideas by allowing members multiple and specialized areas for topics to be addressed, discussed, and voted upon. Finally, it also separates government from property by making all data and discussions be open to the public, and allowing transparent, open voting in almost all changes within the community. These policies have allowed Wikidata to grow into an more efficient database for supporting other Wikimedia sites and GLAMs.

References

- Antin, J., Yee, R., Cheshire, C., & Nov, O. (2011). Gender Differences in Wikipedia Editing. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (pp. 11–14). New York, NY, USA: ACM.
- Bastone, G. (2019). Personal Communication.
- Benkler, Y. (2016) “Peer Production and Cooperation” in J. M. Bauer & M. Latzer (eds.), *Handbook on the Economics of the Internet*. Northampton, MA: Edward Elgar Publishing, Inc.
- Gardler, R. and Hanganu, G. (2013). “Governance Models” on *OSSWatch*. Retrieved from <http://oss-watch.ac.uk/resources/governancemodels>
- Glott, R., Schmidt, P., & Ghosh, R. (2010). Wikipedia survey–overview of results. *United Nations University: Collaborative Creativity Group*, 1158-1178. Retrieved from http://www.ris.org/uploadi/editor/1305050082Wikipedia_Overview_15March2010-FINAL.pdf
- Graciani Picardo, P. (2019). Personal Interview.

- Konieczny, P., & Klein, M. (2018). Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media & Society*, 20(12), 4608–4633. <https://doi.org/10.1177/1461444818779080>
- Mora-Cantalops, M., Sánchez-Alonso, S., & García-Barriocanal, E. (2019). “A systematic literature review on wikidata”. *Data Technologies and Applications*, 53(3), 250-268. doi:10.1108/DTA-12-2018-0110
- Stinson, A. D., Fauconnier, S., & Wyatt, L. (2018). Stepping beyond libraries: The changing orientation in global GLAM-wiki. *JLIS.it, Italian Journal of Library and Information Science*, 9(3), 16-34. doi:10.4403/jlis.it-12480
- “Terms of Use” (2019) on *WikiMedia* Foundation. Retrieved from https://foundation.wikimedia.org/wiki/Terms_of_Use/en#3._Content_We_Host
- “Wikidata” (2019) on *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Q2013>
- “Wikidata:Administrators” (2019). On *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:Administrators>
- “Wikidata:Bots” (2019). On *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:Bots>
- “Wikidata:Bureaucrats” (2017). On *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:Bureaucrats>
- “Wikidata:CheckUser” (2019) on *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:CheckUser>
- “Wikidata:Data Donation” (2019) on *Wikidata*. Retrieved from https://www.wikidata.org/wiki/Wikidata:Data_donation
- “Wikidata:Introduction” (2018) on *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:Introduction>
- “Wikidata:Oversight” (2019). On *Wikidata*. Retrieved from <https://www.wikidata.org/wiki/Wikidata:Oversight>
- “Wikidata:Project Chat” (2019) on *Wikidata*. Retrieved from https://www.wikidata.org/wiki/Wikidata:Project_chat
- “Wikidata:Property Creators” (2019). On *Wikidata*. Retrieved from https://www.wikidata.org/wiki/Wikidata:Property_creators