Unveiling Equity Challenges in Training Needs: Leveraging Wikidata for Linked Data Resource Description in GLAM communities

Greta Heng
San Diego State University

Michael.C.Herrick
San Francisco State University

Abstract

This research will document the issues preventing the broad majority of GLAM metadata workers in the United States from participating in Wikidata projects, addressing the unequal representation issue. These findings could significantly impact Wikimedia's Knowledge as a Service and Knowledge Equity aims. The researchers are already conducting a small training-based case study on a broad cross section of metadata skill sets in the California State University system. Wikimedia Foundation Research funding will greatly broaden the data pool by allowing the researchers to diversify the metadata creation publics studied and generalize the research findings.

Introduction

Background:

- To date, most Wikidata contributors and training participants come from large GLAM institutions. Training perspectives and needs outside elite institutions are being overlooked.
- While there are many free Wikidata trainings and conferences available, there is a research gap evaluating them and examining the training needs for Wikidata editors, especially for those from public and small GLAM institutions.
- Research has generally focused on equity issues in Wikipedia communities, not Wikidata. The Wikidata-focused research has targeted populations actively moving metadata techniques from theory to practice. Those examining Wikidata equity issues are more focused on the impact on the cultural heritage sector and open education than the equal representation from all types of GLAMs. This study aims to unearth issues among a broad population of future Wikidata practitioners in GLAM institutions

Goals:

- Document the training needs of GLAM metadata workers in attaining higher levels of Wikidata knowledge
- Identify the obstacles for under-represented communities to participate in Wikidata training
- Contribute to increasing production competencies to benefit Wikimedia deliverables most effectively

Benefits:

- Suggest solutions to enable greater uptake, by the general library community, of Wikidata in its workflows and applications, especially for underrepresented institutions
- Fill the gap of current Wikidata training strategies to encourage contributions from GLAM staff with diverse backgrounds as they learn the Wikidata environment.

Research questions:

- Is there an uneven representation of libraries actively participating in Wikidata within the United States? What are the barriers to the equitable adoption of Wikidata?
- What are the key inhibiting factors, from a learning perspective, to broad adoption of Wikidata for identifier minting by regular GLAM metadata staff, and how can these barriers be relieved?
- What are the learning needs of metadata practitioners with various levels of Wikidata knowledge and metadata skill sets?

Date:

- A CSU-only training-focused case study is beginning January 2024.
- With Wikimedia Foundation funding, the expanded case studies will occur between October 2024 and March 2025 with research outputs in Spring 2025.

Related work

Libraries across the United States actively engage in Wikidata editing, benefiting from various training series, groups, and Wikimedian events such as the PCC Wikidata pilot program and LD4 Wikidata affinity group. While studies have reflected on the challenges and advantages of Wikidata (Hahn & Ockerbloom, 2021; Provo et al., 2021; Zhang, Biswas, & Dagher, 2023), little attention has been given to analyzing the training provided to participants, particularly those with different backgrounds in Wikidata and different metadata skill set to begin with.

Knowledge equity is focused on addressing the systemic biases and discriminations by promoting equal footing to marginalized communities and their knowledge traditions (Allison-Cassin & Tharani, 2020). While some studies (Deng et al., 2023; Sigalov, Cohen, & Nachmias, 2023) look into the equity issues in Wikidata, their focus is on Wikidata's impact on cultural heritage information and open education. Our study fills a large gap existing within the total context of the LD4 Wikidata Affinity Group's current activity. Examining the recent meeting notes of this group reveals a predominant representation from large academic libraries and research institutions. Additionally, the group focuses on innovating and documenting Wikidata techniques of use to GLAMs and not on understanding the barriers to greater community uptake and learning.

Methods

Research base:

- California State University (11 campuses, in progress)
- Broadened to other GLAM metadata workers in the United States, dependent on funding

Research tools:

- Based on the initial findings of CSU case studies, provide a broader Wikidata resource description and visualization synchronous learning workshop.

- Conduct “in training” observations, interviews and surveys among this larger scale training group to evaluate the training strategies

Expected output

To share our research within the Wikimedian and GLAM communities, we will:

- Build a main training Wikiproject page
- Partner with Wikipedians, Wikimedians, and organizations
- Present the research findings at local and international Wiki conferences
- Publish an article in a peer-reviewed journal in library science
- Presenting this study at library conferences.

Risks

- Not enough non-CSU training participants
- Not able to line up a suitable trainer for the main workshop where data will be collected
- Inadequate identification and contextualization of training benchmarks

Community impact plan

This research aligns with Wikipedia's 2030 Strategies of Knowledge as a Service and Knowledge Equity. It emphasizes training and collaboration for the broadest possible GLAM metadata creation communities. By tailoring systematic training resources for diverse levels of expertise, it encourages broader participation in Wikidata editing. Additionally, it will empower underfunded institutions and LD novices to engage with Wikidata thus promoting knowledge equity and expanding contributions to the LD and Wikidata communities.

Evaluation

- Number of participants in main training workshop and responses received for surveys
- The richness of Wikiproject cites about Wikidata training and resources
- Feedback from presentations and publications

Budget

~30-40K

- Interviews with research institutions: ~$2k [Research assistant (summer 2024)]
- Conferences: ~4k [Principal Investigators (spring/summer 2025)]
- Compensation for participants’ travel expenses for synchronous training: ~20k [Fall/Winter 2024-2025)]
- Synchronous training workshop organization: ~4k

Prior contributions

- Greta Heng’s main research interest is in LD. She has worked on several Wikidata projects and published an article investigating issues in Chinese cultural heritage resource descriptions on Wikidata.

- Michael Herrick has three decades of experience training library and archives staff in metadata creation working in a broad range of cultural and linguistic contexts. He is the chair of the CSU Linked Data Task Force and is currently pursuing several research topics related to metadata literacy.

Neither PI has previously received WF funding

References

1. Allison-Cassin, S., & Tharani, K. (2020). *Critical issues in Knowledge Equity [online session]*. WikiCite Discussion Series https://www.ifla.org/node/93341.
2. Deng, S., Heng, G., Xu, A., Zhu, L., & Li, X. (2022). Enhance the Discovery and Interoperability of Culturally Rich Information: the Chinese Women Poets WikiProject.
3. Hahn, J., & Ockerbloom, M. (2021).Penn Libraries Linked Data Vision. https://upenn.app.box.com/s/b9ghloozr07rnuj5s4ci313go1k91h9t.
4. LD4 Wikidata Affinity Group Call Meeting Agenda (February, 2023). https://docs.google.com/document/d/1p8Av_xgNgIm_m7eFoIOtRlFzpby-8O7C_X528JH94_M/edit#
5. Provo, A.,Roylance, A., Katzoff, B., Chou, C., Stevens, G., Burak, G., Kim, J., Allgood, J., Lin, N., & Razizadeh, Y.. *Everyone can Edit: Cross-collaboration at NYU on the Program for Cooperative Cataloging Wikidata Pilot*. 2021 LD4 Conference on Linked Data in Libraries. Virtual.
6. Sigalov, S. E., Cohen, A., & Nachmias, R. (2023). Open Educational Resources for Literacies, Diversity, Equity and Inclusion: The Case of Integrating Wikidata into Higher Education. In Open Educational Resources in Higher Education: A Global Perspective (pp. 279-306). Singapore: Springer Nature Singapore.
7. Zhang, E., Biswas, P., & Dagher, I. (2023). Reflections on the PCC Wikidata Pilot at UCLA Library: Undertaking the PCC Learning Objectives. *Cataloging & Classification Quarterly*, 1-38. https://doi.org/10.1080/01639374.2023.2269416