

[LD4-Wikidata Affinity Group web page](#)
[Google Drive Folder](#)
[WikiProject Practical Wikidata for Librarians](#)

[Slack channel: <https://ld4.slack.com>]

LD4 Wikidata Affinity Group

2023-05-02

9am PT / 12pm ET / 16:00 UTC / 6pm CEST ([Time zone converter](#))

Facilitator: Eric Willey, Illinois State University, ld4-wikidata-coord@googlegroups.com

Speakers: Fudie Zhao, University of Oxford

Topic: Wikidata in Digital Humanities projects

Recording:

https://drive.google.com/file/d/1bKzMO_GUrU92F19Er4HjBuY46PSxhdq/view?usp=share_link

Presentation materials

A systematic review of Wikidata in Digital Humanities projects

<https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqac083/6964525>

Attendance

(if this is your first meeting, please write your name, Wiki* username, institution or affiliation, and interest in Wikidata if you'd care to share)

- Ann Myatt James, George Washington University
- Eric Willey, Illinois State University
- Pilar Ramirez Amurrio (dokusare.com)
- Brandon Katzir, Oklahoma State University
- Alan Munshower, Virginia Tech University Libraries
- Mary Aycock (Texas State University)
- Bob Kosovsky (New York Public Library) | [user:kosboot](#)
- Sara\$h Kasten (University of Notre Dame)
- Crystal Yragui (University of Washington)
- Mairelys Lemus-Rojas (Brown University)
- Karen Li-Lun Hwang
- Jeannette Ho (Texas A&M University)
- Tyler Thorsted (Brigham Young University) [User:Thorsted](#)

- Cliff Landis (AUC Woodurff Library, user:clifflandis)
- Lisa Shiota, University of Maryland (Lshiota)
- Sasha Frizzell, Binghamton University
- Soo Bae, National Gallery of Art
- Kerri Pfister, Frick Art Reference Library
- Elizabeth Peterson, University of Oregon
- Gerry Walden, Harvard University
- Mingyan Li, University of Illinois Chicago
- Tricia Lampron, University of California, Irvine
- Jack OMalley, Frick Art Reference Library
- Megan Sallabedra (Getty Research Institute)
- Karly Wildenhaus (New York Public Library)
- Huda Khan (Stanford - joined late)
- Hilary Thorsen (Stanford)
- Patricia Thurston (Yale)
- Jessika Davis (Semantic Lab @ Pratt)
- Giovanna Fontenelle (Culture and Heritage, WMF)
- Natasha Zaleski (New Jersey State Library)
- Iván E Calimano (Retired)
- Alexandra Provo (NYU)
- **Monika Soler Correa** (University of North Carolina at Charlotte)

Agenda

- Welcome + announcements

Notes:

Presentation: A systematic review of Wikidata in Digital Humanities projects by Fudie ZHAO (DPhil Candidate in East Asian history and Digital Humanities)

- In 2018 worked at SOAS Library IT Service and introduced to library systems (metadata, etc.). Did a project on how to enrich authority datasets to improve discovery service
- Also completed a master's thesis on Wikidata
- Focused on multilingual datasets from Chinese, Korean and Japanese sources, working on integrating Wikidata into a researcher's workflow in order to address the issue of scattered research. Hoped to publish resources she's collected as linked open data.
- Did face technological barriers from lacking a computer science or information science background, but Wikidata is intuitive and provides opportunities for collaboration.
- She tried to research what people have done with Wikidata and DH, discovered there wasn't a systematic review of Wikidata in Digital Humanities projects, so she authored one: <https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqac083/6964525>
- Wikidata relationship with Digital Humanities community: DH can pose challenges to research as it is quite cross-disciplinary. It seems that 2019 is the moment that Wikidata attracted increased attention.

- Related Works and research gap:
 - Widely adopted & systematically reviewed in IS/CS and the library domain.
 - However, there lacked a review regarding Wikidata and DH.
 - Paper helped explain/justify research using Wikidata to potential collaborators who might not be familiar with it.
- Asked 4 research questions:
 - 1. How is Wikidata described in the current DH literature?
 - Divided into 3 concepts: Technology Stack, Platform and Content.
 - People have diverse understanding of Wikidata: most commonly addressed was “content provider”. Since, DH is multi-disciplinary, people from different backgrounds bring different vocabulary (“authority file” used in libraries for instance).
 - 2. To what end is Wikidata being experimented within the DH domain?
 - Used a taxonomy developed in the European Union to describe digital humanities practices
 - At least 4 commonly seen:
 - Annotation and enrichment
 - Researchers often have own text corpus, linguistic dictionaries.
 - Metadata curation
 - Integrate authority datasets (choose ID to obtain URI)
 - Improvement of metadata quality (often seen in libraries)
 - Sharing their unique metadata for public use
 - Using Wikidata info to enhance local records
 - Visualize and analyze metadata
 - Use visualization tools provided by Wikidata to analyze distribution and other aspects of metadata
 - Modelling
 - Knowledge modelling in the context of the Semantic Web, reference for the creation of data models
 - Name entity recognition (NER) and related tasks and miscellaneous (pedagogy)
 - Occurs when projects are using natural language processing.
 - Used in pedagogy in at least one classroom to teach archaeology students about use of Linked Open Data. Wikidata is quite intuitive and doesn’t require a lot of background knowledge to understand.
 - In data aggregation projects, it provides one of the resources to link with domain-specific data
 - A source of data itself for research: different than annotation.
 - 3. What is the potential of embracing Wikidata in data-related activities?

- Data consumption (45 out of 50 projects consumed data from Wikidata)
 - Although many concerns expressed about data quality due to lack of control, this data shows that the majority actually consume it anyway.
 - Data publication and exchange
 - Wikidata as a platform, technology stack, and a repository
 - Data integration (linking hub for domain resources)
 - Data production (low-tech approach for publishing linked open data)
 - Publish data on Wikidata to consume data
- 4. What are the challenges and possible solutions regarding Wikidata's data quality?
 - Technical challenges, such as identifier mismatches, data model incompatibility
 - Concern about quality due to open model
 - Contemporary and notable more prevalent, less so for historical and local data
 - Projects using Wikidata prioritized data coverage over data accuracy.
 - Can be difficult to detect challenges as they aren't always addressed as such.
- Adopted a methodology for systematic review(Kitchenham's guidelines for Systematic Review (2004)) that's commonly used in software engineering and medical science. Audience can actually track-back inclusion/exclusion criteria.
 - First decisions need to be made on which databases and keywords used for searching and language used (focused on English first)
 - Inclusion and exclusion criteria
 - Data collection and analysis
- What can we learn from the projects?
 - Evaluate its quality against other available domain sources. Why do projects use it despite concerns about data quality?
 - They contextualize their perception of Wikidata's quality by comparing or combining it with other data sources available. Best if:
 - Entities in question are well-known, popular or infamous (notability)
 - Newly emerging or niche fields lacking other domain-specific sources
 - Cross-domain tasks because Wikidata is generic source
 - Different projects may perceive Wikidata's quality
 - Often used in combination with other sources, such as DBPedia, Wikipedia, VIAF, GeoNames. Projects often have local datasets available for them for specific names/entities and use Wikidata to cover national/international names.

- Form a community of practice
 - More users that use it and experts contribute to it, more likely that domain will have high-quality datasets.
 - For each field, DH needs their community of practice that are suitable for their purposes.
- Design a workflow that orchestrate technical and labor resources from the projects and Wikidata
- Normally a technician can identify methods to automate some steps. But for projects with less support, they have access to tools developed by Wikidata (Mix'n'match) and other open source tools (such as OpenRefine which helps with identifier matching).
- Conclusion:
 - Wikidata is more than a knowledge base: it can also be considered a content provider, platform, technology stack.
 - Mainly used for annotation research materials, curating metadata
 - Used more for consumption than publication.
 - Projects should take into consideration:
 - available domain sources
 - community practices
 - workflow design to balance resources
- Question to ask audience:
 - How can we better collaborate? Feedback from GLAMs are always welcome!
- Illinois State University: Is looking to hire a digital humanities librarian. Scholarly Communications librarian there did a survey and discovered that technology support is something their university community wanted the library to be more active in, so that's one area the library can work to assist them with.
- At Oxford, several institutes exist related to digital humanities. Now have a digital humanities degree. Libraries are quite active in organizing events related to DH.
- Q: Curious to know more about pedagogical use of Wikidata.
 - A: Not a common use in terms of what projects report, so that's why she noted the one archaeological one. Teachers had the students curate entities on Wikidata in order to generate and visualize linked data sets.
 - She has also created a 3 week-mini course.
 - Some problems in some properties (e.g., "citizenship" in the historical context). But it is a great way to teach students with little background in data to understand how it all works.
 - Comments about other DH projects:
 - There have been a number of presentations in the Affinity Group which were DH projects (like the Scottish witches <https://wikimedia.org.uk/2018/03/data-on-the-history-of-scottish-witch-trials-added-to-wikidata/>) - seems like Wikidata is more a tool to enable relationships not available in other ways

- For Wikidata pedagogy you might also be interested in Stacy Allison-Cassin and Cora Coady, Adam Cavanaugh, and Devin Frede's talk on Wikidata in the classroom. Link to agenda: https://docs.google.com/document/d/13TI0ox1wh4T9dXagcidt9Es_pR8am29n9DamGggxczF0/
- Q: From your activities, have you proposed some sort of model/properties that could be better used? Asian materials, particularly rare historical sources, are unique in their needs.
 - A: This is what we're often facing as historians. Can you see what projects have already exported/used Wikidata in the area? Each researcher has own analytical framework/historical sources. One place in which an unified model can be developed is when researchers curate their metadata (such as a bibliography). As for primary sources and analytical framework, found it is difficult to establish a general model for everyone in the community to use.

Announcements

- Next [Wikidata Working Hour](#):
 - Friday, May 12, 2023 at 10:00am PT / 1:00pm ET / 17:00 UTC / 7:00pm CET
 -
- Next [Wikidata Affinity Group Call](#)
 - Tuesday, May 16 at 9am PT / 12pm ET / 16:00 UTC / 6pm CEST
 -
- Next [Wikibase Working Hour](#)
 - No sessions are planned as of May 2023. We will probably do a session at the 2023 LD4 Conference on Linked Data in July, though--stay tuned for details!
-
- SAVE THE DATE! LD4 2023 Conference July 10-14, 2023
 - <https://sites.google.com/stanford.edu/ld4-community-site/conferences>
-

Help us organise the LD4-Wikidata affinity group!

We're moving to make it easier for more of us to participate in organising LD4-Wikidata affinity group programming! Here's what you can expect in coming weeks:

- Starting **this month**, we're opening up coordination meetings + channel on Slack
 - Coordination meetings take place once a month!
 - Current schedule: Monday at 1pm PT/4pm ET ([Time zone converter](#))
 - Next call -- save the date: **May 8th at 1pm PST/4pm EST** (Zoom link in channel)
 - Join LD4 Slack channel [#wikidata-coordination](#) to join the conversation and get further updates!
- Anyone can sign up for various rotational roles and propose group call/working hour topics, either during each organising meeting or in the days following through the coord Slack
- Runsheets will be available for each role, and role hand-off will be a team effort. Depending on capacity/comfort level, you are welcome to take on a smaller role then grow into bigger roles.
- Here are roles that we've identified as gaps! These are some starting points, but we invite any and all sorts of contribution based on your skillset:
 - note-taking for a call
 - documentation -- help refine and build on key pages, and continue to develop/nudge others to add content
 - help communicate/promote programming
 - present during a group call on some topic
 - lead a working hour on some topic
 - help facilitate calls

Weekly summary

Listed in [Wikidata:Status updates](#) and posted to the [mailing list](#) / [on-wiki](#) to your talk page

- Recent summaries
 - [Week before 2023-04-17](#)
 - [Week before 2023-04-11](#)
- Help write [the next summary](#)!

More editing communities

- [Live Wikidata Editing](#), a Wikipedia Weekly Network live streaming show hosted by Jan Ainali and Albin Larsson. Wide variety of topics covered, archive of YouTube videos going back to March 2020.
- Telegram Wikidata channels:
 - Join Wikidata channel: <https://t.me/joinchat/leCRo0j5Uag1qR4Tk8Ftsg>
 - Join Wikidata Help channel: <https://t.me/wikidatachat>
 - Join Wikidata Ex Música channel: <https://t.me/exmusica>

Join Wikidata Lexicographical data: <https://t.me/joinchat/TwVx898JKJTAajm4>

