

[LD4-Wikidata Affinity Group web page](#)

[Google Drive Folder](#)

[LD4-Wikidata Affinity Group Glossary](#) (info about the LD4P2 grant and related terminology)

LD4 Wikidata Affinity Group

2019-11-05

9am PST / 12pm EST / 17:00 UTC / 6pm CET ([Time zone converter](#))

Chair: Hilary Thorsen, Wikimedian in Residence, Linked Data for Projection project
(thorsenh@stanford.edu)

Recording:

https://stanford.zoom.us/recording/play/D1VuJKF4stRiEysoXj0Q8h40T_-bVcAOMonncNCI_Ouq0WpC-642R0aXLoJJEEeNE?startTime=1572972276000

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)

- Adam Zukowski, Towson University (curious observer)
- Jeff Young and Karen Smith-Yoshimura (OCLC)
- Lena Denis (Harvard Map Collection)
- Jeannette Ho (Texas A&M University)
- Ben Riesenbergl
 - [riesengrey](#)
 - University of Washington
 - Interested in (among other things) reconciling library linked-data entities with Wikidata items
- Barbara Bushman (NLM)
- Jordan Patterson (Memorial University of Newfoundland)
- Mike Bolam (Univ of Pittsburgh)
- Allen Kwan (Ontario Digital Service)
- Stephan Schindehette (OCLC)
- Huda Khan (Cornell)
- Rhonda Super(UCLA): Semantic Web LibGuide:
<https://guides.library.ucla.edu/semantic-web>
- JShieh (Smithsonian Lib)
- Jamie Flood
 - Jamie-NAL
 - National Agricultural Library, USDA
 - Sharing Library collections/items on WikiData using WikiData for article creation/editing (new tool/coding being used)

- Cliff Landis (AUC Woodruff Library, [@clifflandis](#))
- Patricia Dragon (East Carolina University)
- Alla Roylance (NYU)
- Joy Panigabutra-Roberts (U of Tennessee Libraries)
- Chew Chiat Naun (Harvard)
- Honor Moody (Harvard)
- John Hostage (Harvard)
- Will Kent (Wiki Education)
- Lihong Zhu (Washington State University)
- Lori Robare (University of Oregon)
- Kay Teel (Stanford University)
- Liam Wyatt / Wittylama (Europeana & Wikicite)

Agenda

- Announcements
 - WikidataCon 2019 [program](#) and [recordings](#)
 - Will recap highlights at November 19th call
 - [WikiConference North America](#) November 8-11 in Cambridge, MA, USA
 - Will recap highlights at November 19th call
- Martin Poulter, Wikimedian in Residence, University of Oxford on using Wikidata to describe the internal contents of a book, projects at the University of Oxford, and the Dictionary of National Biography (& other biographical dictionaries) in Wikidata ([slides](#))
- Next call 19 November 9am PST / 12pm EST / 17:00 UTC / 6pm CET ([Time zone converter](#))

Notes

- [Martin's Slides](#)
- Notability
 - Wikidata tries to capture top levels of FRBR
 - You wouldn't normally capture individual items in Wikidata unless they are special collections material--not mass-produced things
- Showcase query <https://w.wiki/BQG>
 - Shows us that google books has this book and --could look for later editions of political books that Google books has
- Mass-produced books not generally notable
- An individual biography in A Dictionary of the Booksellers and Printers - just two lines of text! - is notable enough for Wikidata
 - Anything that has a page on any of the Wikimedia projects can be represented in Wikidata
 - This book has a page in Wikisource, and so does each individual biography
- Wikisource

- Free library--bit like Gutenberg, but better
- Any freely available text is in scope
- It's a community who do volunteer to share knowledge and promote research
- Take scans of a book and fix OCR (optical character recognition)
 - Fixed by human users to make a verified electronic version of the text
 - Different than wikipedia--just judging what text is visible on page and capturing some of the layout and appearance
 - Transcription project going on now for 1860 Leaves of Grass by Walt Whitman
 - Yellow means page is fair representation of text
 - Yellow then becomes promoted to green once checked by 2 users
 - Blue pages have something complicated--an image, page not legible, etc
 - Big task of transcribing book is broken into small pieces that community can contribute to
 - Made available in ebook formats that people can download
- Wikidata and Wikisource
 - Book editions have Wikidata representations
 - Author profiles and topic profiles
 - Author--get birth and death years
 - Authority control links
 - Wikipedia article
 - All of this information drawn in from Wikidata
- Biggest achievement of english Wikisource (French is biggest) is out of copyright Dictionary of National Biography (nation is UK)
 - Has 30,000 biographies
 - Whole thing has been transcribed and individually bookmarkable on Wikisource and on Wikidata
 - Can be treated like a database
 - <https://w.wiki/BQf>
 - Can do queries to find people in this resource and another bibliographic source
- Lots of biographical dictionaries in progress
 - Long, categorised list here: https://en.wikisource.org/wiki/Wikisource:WikiProject_Biographical_dictionaries
 - Grove Dictionary (out of copyright version) a bit more than half-done
 - Who's Who of American Women is about a quarter done
 - Biographical dictionaries can be added
- Wanted to add a biographical dictionary
 - The Plomer Dictionary: a dictionary of booksellers and printers who were at work in England, Scotland, and Ireland from 1641-1667
 - Imported the scans to Wikisource

- Announced in staff letter at Oxford that they were doing this and Wikisource community contributed as well
 - Only needed to be able to recognize English to do this task
- Once have corrected text, need to mark up the structure of the book--pseudo xml syntax for describing the different chunks of the book
 - Colleague Kat Steiner went though and did it for the whole book
 - Created table of contents with biographies
 - Biographies could then be represented in Wikidata
 - Wanted to add the people from the book to Wikidata
 - Used OpenRefine
 - Used GoogleSheets to do it
 - Extensions enable you to do look-ups to any API
 - Could describe a series of steps and processing take link to biography in Wikisource to get the text of biography and do different tests on it
 - Get the name of the person
 - Get that they're a human being
 - If there's a 4 digit number, flourishing date
 - Bookseller or printer appear in first sentence, then they're the occupation of the person
 - If the place names appear, then those are the workplace of the person
 - Used the spreadsheet to extract some properties from the text
 - Fed into QuickStatements
 - Queries like people described in the book who worked in publishing in Oxford visualized as timeline
<https://w.wiki/BQN>
- Flora Graeca (The Sibthorp & Bauer Expedition)
 - Botanical book with paintings of plants, represented in Wikidata as hundreds of art works that are "published in" the book.
 - Gave different ways for content of book to be explored and presented
 - Each painting deserves a Wikidata item.
 - Wikidata has tree of life info and species are part of a genus, family, etc.
 - One could ask for all images in book that depict species in a given family. Wikidata knows which species are in which family- we don't have to supply this.
 - Can get common names from Wikidata
 - <http://glam-discovery.bodleian.ox.ac.uk/botany/>

- Got coordinates from Wikidata for Mediterranean and visualise on map
- Question from Jackie--CML encoding at Oxford?
 - Martin--a lot of efforts around this at Oxford. Wikisource is providing definitive versions of text--just gets text in form that's readable, not TEI, but helps to have correct version of text before marking up in TEI--can play a part in the process, but performs a different function. Wikisource likes to compare the text against images in terms of formatting like font size. Feminism portal on Wikisource--Martin was able to add some texts from Oxford. Moving text from Oxford to Wikisource strips out some mark-up, shown to broader audience. Wikidata can point people to versions of a text at various institutions
- Jackie-How is it possible to express relationship between versions?
 - Full text **online at**
 - Different identifiers like Project Gutenberg, Google Books
 - Edition or version of property
 - Properties attached to the editions
 - Digital surrogate would be attached to the edition
- Jackie-Entity reconciliation to avoid duplicates before importing
 - Recommends using OpenRefine for reconciliation, so you can do reconciliation before import
 - Martin manually merged duplicates after importing--not best practice, but can be fixed
- Martin has created showcase examples <https://w.wiki/BQG>
 - MacCauley edition
- Karen-Do you record copy-specific info for a particular edition?
 - Record on Commons will say where book has come from, so text can be traced to a specific university library
 - WikiSource isn't interested in transcribing marginalia, but gives a paper trail to that specific item
- Liam--mentioned that WikiSource isn't interested in marginalia. Is there a way to represent palimpsests in Wikidata?
 - Possible to add marginalia, but not core
 - Martin not aware of, but hasn't been in the scope of his work at this point; maybe need a dedicated WikiProject for that
- Does Wikisource have a way of dealing in transcription with text errors, non-standard spelling, etc. that could be problems for searching
 - You can sic tag--shows that you have corrected and you can mouse over and see the original
- Where do Wikisource discussions happen?
 - Talk pages for transcription projects and finished online book are two different layers; talk pages for transcription projects are a place to discuss

and whoever sets up the project can decide based on past work, etc. how best to transcribe the item

- Had a transcription public event for Mary Somerville to transcribe a paper and part of a book by her.

https://en.wikisource.org/wiki/Author:Mary_Fairfax_Somerville

- Jeff— query for palimpsest: <https://w.wiki/BWf>
- Martin--We harness immense pedantry for peaceful ends--library and Wikimedia communities
- Martin—"We harness the right kind of obsessive"