

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

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

LD4 Wikidata Affinity Group Call

2022-09-06

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

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

Topic: Dominic Byrd-McDevitt will talk about the Digital Public Library's Wikimedia program, an effort to provide national leadership around access and discoverability of digital collections by leveraging Wikipedia and its sister projects. 2 years ago, DPLA launched a digital asset pipeline to enable participating institutions in the DPLA network to share their collections with Wikimedia Commons. DPLA is continuing to innovate by taking advantage of Wikidata entities and Structured Data on Commons to continually synchronize data and improve discoverability. We'll discuss issues around large datasets, aggregation, reconciliation, and other challenges DPLA has faced.

Recording

[video2022-09-06.mp4](#)

Connection Information

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*

- Eric Willey, Illinois State University, emwille
- Diane Shaw, Smithsonian Libraries & Archives
- Teresa Hebron, Mountain West Digital Library/University of Utah
- Amanda Xu, National Agricultural Library
- Elena Sillitti, Casalini Libri, user: Sillitti E.
- Christa Strickler, University of Notre Dame
- Mohammed Awal Alhassan, Alhassan Mohammed Awal, Dagbani Wikimedians User Group
- Rhonda Super (UCLA retiree) Semantic Web LibGuide:
<https://guides.library.ucla.edu/semantic-web>
- Marianne Swierenga, WMU Libraries
- Stella Wisdom, British Library

- Prema Desai, National Agricultural Library
- Sherman Clarke, Alfred U and Avery Index
- Mary Aycock, Texas State University
- Cory Lampert, University of Nevada, Las Vegas
- Sarah Kasten, University of Notre Dame
- Steven W. Holloway, James Madison U.
- Xiaoli Ma, University of Florida
- Doreva Belfiore, Philadelphia WikiSalon (user:dorevabelfiore)
- Will Kent, Wiki Education, user:Will (Wiki Ed)
- Steve Baskauf, Vanderbilt Heard Libraries
- Beth Guay University of Maryland Libraries
- Amy Watson, Smithsonian Libraries & Archives
- Jamie Flood, User:JamieF, U.S. National Agricultural Library (NAL)
- Bob Kosovsky, New York Public Library | [user:kosboot](#)
- Mairelys Lemus-Rojas, Brown University
- Jessica C. Williams, Tiffin-Seneca Public Library, user:wispofacloud
- Kyla Jemison, University of Toronto
- Sarah Adams (New York Public Library + Semantic Lab @ Pratt)
- Sarah Osborne Bender , National Gallery of Art (DC)
- Jeannette Ho, Texas A&M University Libraries
- Soo Bae, National Gallery of Art (DC)
- Israel Yanez, California State University, Sacramento, user: iyanez916
- Emma Clarkson, New York Public Library
- Goodness Ignatius/ User:Olugold (Igbo Wikimedians User Group)
- Christine Fernsebner Eslao, Harvard | User:Fernsebner
- Meg Mering, University of Nebraska-Lincoln
- Susan Radovsky (Harvard Library) | User: Sradovsk
- Stephen Mantz, University of Pennsylvania, user: Liveoak957
- Pilar Ramirez Amurrio ([dokusare.com](#))
- Melanie Polutta (Library of Congress)
- JJ Dearborn (Smithsonian Libraries and Archives, Biodiversity Heritage Library)
- Karly Wildenhaus, New York Public Library
- Nora Zimmerman (Lafayette College)
- **Monika Soler Correa** (University of North Carolina at Charlotte)
- Richard J. Urban, (OCLC, Research Library Partnership) | User:musebrarian
- Huda Khan, Cornell
- Jere Odell, IUPUI
- Hollis Wittman, Virginia Tech

Call notes

(please feel free to edit/update notes for updates + corrections!)

Dominic Byrd-McDevitt will talk about the Digital Public Library's Wikimedia program, an effort to provide national leadership around access and discoverability of digital collections by leveraging Wikipedia and its sister projects.

- Intro
 - <https://dp.la/news/wikimedia-project-update>
 - Big Picture = all cultural institutions have onramp to leverage Wikimedia w/ tech support
 - Goals/Agenda for today = DPLA's Digital Asset pipeline; Wikidata & Structured Data on Commons; DPLA Network Participation
- Background @ US National Archives
 - Dominic Byrd-McDevitt was first Wikimedian-in-residence at US National Archives
 - First institution to do large scale bulk uploads to Wikimedia Commons, worked on first API
 - Held training Wikipedian Bootcamps
- DPLA's Wikimedia Program [<https://dp.la/news/wikimedia-project-update>]
 - similar efforts as US National Archives, but make it so that each institution doesn't have to "reinvent the wheel" to upload/engage w/ Wikimedia
 - DPLA is largest aggregator of cultural heritage in US, generally modeled after Europeana (Europeana is government agency whereas DPLA is not)
 - Goal = allow DPLA institution to contribute images to Wikimedia Commons and link back
 - pipeline = metadata (dpla aggregation) + media files (contributing institution) → Wikimedia Commons → Wikipedia
 - DPLA has one top level category Media contributed by the DPLA
https://commons.wikimedia.org/wiki/Category:Media_contributed_by_the_Digital_Public_Library_of_America
 - Outcomes: 3million + files to Wikimedia Commons (largest single contribution, 5% of all contributions); 115 million page views on Wikipedia; uploads from 200+ institutions across 10 hubs
- Structured Data on Commons (SDC) Work
 - worked w/ community to map DPLA data
 - example item on slides:
[https://commons.wikimedia.org/wiki/File:%22Don%27t_let_this_surprise_you%22_-_Bour_Quality_Coffees_Royal_Garden_Teas._-_DPLA_-_82455468412cc2cc82e4568f3dec05cc_\(page_1\).jpg](https://commons.wikimedia.org/wiki/File:%22Don%27t_let_this_surprise_you%22_-_Bour_Quality_Coffees_Royal_Garden_Teas._-_DPLA_-_82455468412cc2cc82e4568f3dec05cc_(page_1).jpg)
 - reference URL; publisher; retrieved
 - every file uploaded has at least 10 different statements, with many statements containing multiple values (e.g., subjects)
 - Outcomes: 40 million statements added; 200 million triples
 - Why Structured Data on Commons (SDC)?
 - data sync for metadata updates and iterations
 - discoverability, reliability, standardization

- allows for making comparisons and identify data differences when different harvests occur
- Image Citation / automatic capture generation
 - https://en.wikipedia.org/wiki/Charles_Robinson_Rockwood
 - image metadata citation shown in info box
 - https://en.wikipedia.org/wiki/Charles_Robinson_Rockwood#cite_note-1 = “Unknown Photographer, “Charles Robinson Rockwood”. Photographic Print from The History of Imperial County California by Finnis C. Farr p. 96.5, 1918. Google Books, [Charles Robinson Rockwood](#).”
 - structured data + caption can be generated by only providing the file name
 - Image citation -- see more:
 - https://en.wikipedia.org/wiki/Wikipedia:Image_citation/Sample

```

Live example: [[w:Charles\_Robinson\_Rockwood]]
[[w:Wikipedia:Image\_citation/Sample]]
{{com:embed_dpla}}
[[com:User:Dominic/SDC\_citation\_tests]]

```

- View It! (tool)
 - Check out the Meta page: https://meta.wikimedia.org/wiki/View_it!_Tool for installation instructions and to sign up for updates and beta testing
 - could help surface or provide a different way for users to find images
- DepictAssist
 - <https://public.paws.wmcloud.org/User:Dominic/depicts>
 - note: don't forget to log into Quick Statements before you explore
 - automated “depict” suggestions using other data present
 - “depicts” is not a common in metadata standards
 - idea is that this could be a way for the staff to engage with + enhance image descriptions once the collections have been updated
- Funding timeline
 - 2020 Sloan Foundation grant: launched digital asset pipeline; partner meetings/trainings/webinars; changes to data model
 - 2021 Wikimedia Foundation grant
 - Current Wikimedia Foundation grant
 - to improve image descriptions (Depict tool) and discoverability
 - key elements:
 - enrich DPLA data w/ more entities + reconcile w /Wikidata
 - Develop suggestion tool for “depicts” statements
 - work on image citation/caption gadget
 - share DPLA's work w/ global peer institutions
 - Challenges
 - no funding for setting up + maintaining a Wikimedia “program”

- How to make the program sustainable?
 - How to provide better support for partners?
- New DPLA Network Initiatives
 - Project page <https://dp.la/news/wikimedia-project-update>
 - Office hours = First Wednesdays of the month, 2-3pm EST
 - Wikimedia Working Group !

Wikimedia working group

- Rationale: Wikimedia Commons upload as a tech service belongs to DPLA Tech Team, but Wikimedia program will belong to the DPLA network
- Want to build and empower the community of practitioners to provide peer support and collaborate across institutions
- Opportunity to invite outside Wikipedia experts

- Indiana University–Purdue University Indianapolis (IUPUI)
 - largest contributor of Indiana DPLA hub
 - interest in creating a hub-level Wikimedia Project (program?)
- Get involved
 - contribute to DPLA
 - improve your institutions provided data
 - participate in Commons upload
 - become an early adopter/tester of View it! and DepictAssist
 - list of current DPLA hubs if your institution wants to contact their possible hub: <https://pro.dp.la/hubs/our-hubs>
 - contact Dominc → dominic@dp.la

Q&A + Discussion

- Is there any discussions/training for how source partners can merge that data back to their local records?
 - I'd also like to hear a response to Richard Urban's question on round-tripping of data between partner & WMC. Thanks!
 - A: haven't yet encountered the demand; would be interested in hearing from institutions interested in that; can be difficult when Wikimedia community advises metadata updates/corrections to provide that back to the institution
- Does beta testing for the View it tool only apply to participants in DPLA or anybody who has files uploaded to Commons?
 - Anyone can try View it! regardless of affiliation/participation
 - Tool has a lot of broad utility outside of DPLA or cultural heritage; can be helpful for reading any article in Wikipedia; side benefit is that it can surface images from digital collections
- I'd love to see a demo of DepictAssist, and/or hear more about workflows for maintaining/syncing structured metadata on Commons
 - primarily intended for participants in DPLA digital asset pipeline
 - tool runs a query against wikidata that matches items for that subject

- tool builds a batch for you; loads into quick statements for Wikimedia Commons
 - (if View it is installed) clicking on the “view” tab of a wikipedia article would show all images depicting the Wikidata item associated with the Wikipedia article
- At what point does the item being depicted become notable enough to be its out Wikidata item (rather than just an image on Wikimedia Commons) + ramifications for the “depicts” statement
 - question of data + community; not current consensus; doesn’t necessarily make sense in that it could obscure the thing you’re looking at
 - “depicts of depicts” problem = wikimedia commons search does not traverse wikidata info
 - anything coming from a catalog record could be notable enough for Wikidata; but Wikidata likely doesn’t want a bunch of items for nuanced images that are better suited to just live on
 - people have the strongest opinions re: paintings
- Have contributing institutions reported on impact on their collections' discoverability, e.g., statistics on inquiries from users/public?
 - info mainly recorded through tool baglama2
 - <https://glamtools.toolforge.org/baglama2/>
 - example: 47k page views via media contributed by Toledo-Lucas County Public Library → <https://glamtools.toolforge.org/baglama2/#gid=430&month=202208>
 - wouldn’t have been able to capture that usage info outside of this system
- Follow up Q: Just to be clear about this "depicts of depicts" -- would this mean that for an image of the painting Mona Lisa, the Wikimedia Commons (WC) depicts statement is "depicts --> Mona Lisa" and then on the Wikidata item for Mona Lisa you have “depicts —> woman”?
 - yes, that's the issue; might make sense for the data but doesn’t serve the purpose of Wikimedia commons search; some are purists with the data in that they want WC to just say “depicts mona lisa” but that doesn’t help Commons search
- The BHL-tech team is interested in hearing more about the pywikibot you are running to upload images from DPLA https://commons.wikimedia.org/wiki/User:DPLA_bot, is there a github repo we can look at?
 - DPLA can perform uploads for you
 - <https://github.com/dpla/ingestion3#wikimedia>
- What is “depth” in BaGLAMa 2? (<https://glamtools.toolforge.org/baglama2/>)
 - when you’re setting up a category, it’s a way to say “how far down the category tree” you’d want to include
 - i.e., depth of 0 would be just the category itself, depth of 1 would be the category and any subcategories, etc.

Announcements

- Next [Wikidata Working Hour](#): **Monday, September 12, 2022 at 10:00am PT / 1:00pm ET / 17:00 UTC / 7:00pm CEST**
See more information about our current Working Hour series [here!](#)
- Next [Wikidata Affinity Group Call](#): Tuesday, September 20
 - Michael Jones on using machine learning, IIF, and Wikidata with historical newspaper collections
- LD4 Wikibase Working Hour: Discussion on Wikibase and the PCC
 - **When:** Mon. 26 September 2022, 1PM Eastern ([time zone converter](#))
 - **Registration:** Please fill in [this form](#) to register
 - For the September 2022 Working Hour, we are partnering with our colleagues at the [Program for Cooperative Cataloging](#) (PCC) [Standing Committee on Applications](#) (SCA) for a facilitated discussion about use cases for a possible Wikibase instance. As background, the SCA just issued a [Wikibase Exploration Survey Report](#) recommending that the PCC further explore using Wikibase. The goal of this discussion is to get input from PCC participants about what use cases the PCC should focus on if it embarks on developing a Wikibase instance. All are welcome to attend and participate in the discussion! Please submit your comments, questions, and discussion prompts ahead of the discussion on the [meeting notes document](#).
- Wikidata's 10th birthday in October: [Call for community events](#)
- The 3rd Wikidata Workshop, Workshop for the scientific Wikidata community
 - 24 October 2022, <https://wikidataworkshop.github.io/2022/>
- Rhonda Super's Semantic Web LibGuide, <https://guides.library.ucla.edu/semantic-web>, will be migrating to the UC E-Scholarship Repository. Rhonda has retired, so the LibGuide will no longer be updated. A static final version will be available in the repository later this month. If any institution or person would like to continue the LibGuide or work with Rhonda to do so, please contact Rhonda at rasuper1@g.ucla.edu.
- Any other announcements?

Help us organise the LD4-Wikidata affinity group!

- Our monthly coord meetings + Slack channel ([#wikidata-coordination](#)) are open for participation
 - Next meeting: **Monday, September 12 at 1pm PT/4pm ET** (45min)
- Anyone can sign up for a rotational role and propose group call/working hour topics, either during each organising meeting or anytime through the coord Slack
- Runsheets are 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 2022-08-01](#)
 - [Week before 2021-08-08](#)
- 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>
- Rhonda Super's UCLA Semantic Web LibGuide is being retired. A final iteration will be deposited in the UC E-scholarship repository. If anyone is interested in taking this guide over or providing a platform to have it continue, please contact Rhonda Super at rasuper@earthlink.net