

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

[Google Drive Folder](#)

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

[LD4P Wikidata WikiProject](#)

[Mailing List on Google Groups](#)

==

LD4 Wikidata Affinity Group

2020-06-16

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

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

Recording

- https://stanford.zoom.us/rec/share/eAtNuzb_HNLcK_97GzcBJ95MN2-T6a8hHRI-PYOxUzBli57Y2FVCCHaTmzXYHnq?startTime=1592321728000

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)

- Diane Vazine-Goetz (OCLC)
- John Riemer (UCLA)
- Elizabeth Roke (Emory)
- Michelle Durocher (Harvard)
- Susan Radovsky (Harvard) Username: Sradovsk
- Peter Rolla, Loyola Marymount University, interested in using Wikidata for authority control
- Christa Strickler, Kiwigirl3850 (Wikidata user name), Wheaton College (Illinois), interested in Wikidata for authority control and promoting faculty scholarly work
- JShieh (Smithsonian Lib)
- Mairelys Lemus-Rojas (IUPUI)
- Allison Bailund (SDSU)
- Graham Jevon, EAP GJ, British Library (Endangered Archives Programme), interested in wikidata and archival metadata

- John Bolan (University of Toronto, interested in using Wikidata for our Indigenous Perspectives Collection) wikidata username: johnbolan
- Christine Fernsebner Eslao (Harvard)
- Lena Denis (Harvard Map Collection)
- Marc McGee (Harvard Library)
- Karen Smith-Yoshimura (OCLC)
- Michael Stewart (University of Delaware), Wiki* username: MLeonStewart, interests include ethics (also involved in LD4 ethics group) and wikidata's possible role in authority control
- Sarah Kasten (Univ. Notre Dame)
- John Mark Ockerbloom (Penn)
- Jim Hahn (Penn)
- Rhonda Super (UCLA) - Semantic Web LibGuide: <https://guides.library.ucla.edu/semantic-web>
- Patricia Dragon (East Carolina University)
- Diane Shaw, Smithsonian Libraries
- Gila Prebor (Bar Ilan University, Israel)
- Karen Hwang, Asian American Arts Centre/Semantic Lab
- Arcadia Falcone (Stanford)
- Jennifer Page, Montgomery College, interested in using Wikidata for authority control
- Neil M. Frau-Cortes, [@dhneil](#), (University of Maryland Libraries).
- Bob Kosovsky / user: kosboot / Wikimedia NYC
- Tom Christie, Muhlenberg College, interested in using Wikidata for authority control
- Iman Dagher (UCLA)
- Joy Panigabutra-Roberts (University of Tennessee Libraries)
- Sharon Reidt (University of Connecticut)
- Chew Chiat Naun (Harvard)
- Mary Aycock (@maryacat, Texas State University, interested in using Wikidata for authority control)
- Paul Burley (Northwestern)
- Jens Ohlig (Wikimedia Deutschland, Wikidata Development team)
- Barbara Tysinger (Wikidata:Hennap; University of North Carolina at Chapel Hill, Special Collections Cataloger, NACO)
- Isabel Quintana (Harvard University)
- Adam Zukowski (Towson University)
- Matt Miller (Library of Congress)
- Eric Willey (Illinois State University)
- Robert J. Rohrbacher (Stanford; RJRohrbacher)
- Ben Riesenberg (U. of Washington).
- Emily Baldoni (Georgetown)
- Jeff Young (OCLC)
- Daniel Lovins (Yale)
- Tim Kiser (Michigan State University)

- Larisa Waslh (UofChicago)
- Stella Wisdom (British Library)
- Jim Soe Nyun (UCSD)
- Beth P. Camden (Penn)
- Irene Wu (Royal Ontario Museum Library & Archives)
- Anna Slawek (University of Toronto Libraries)
- Wanda K. Gunther (UNC-Chapel Hill; interested in authority control, PCC/NACO)
- Matthew Haugen (Columbia University; user:Matthewhaugen; interested in PCC/NACO, rare materials applications, ethics)
- Meg Wacha (CUNY, Wikimedia NYC)
- Sarah Hovde (UMD Libraries, @sehovde)
- Denise Soufi (UNC-Chapel Hill)
- Edric Mesmer (University at Buffalo)
- Kathleen Boyd (SDSU)
- Michelle Echols (WVU Libraries); No experience using Wikidata; Looking to learn more
- Nicole Smeltekop (Michigan State University)
- Greta de Groat (Stanford, user name Silentdivas)
- Joanna Dyla (Stanford)
- Robert Talbott (UC Berkeley)
- Ryan Mendenhall (TrMendenhall, Columbia University)
- Kate Henke (Wheaton College, MA)
- Yawen Kang (Wheaton College, MA)
- Karly Wildenhaus (Frick Art Reference Library)
- Stephen Hearn (University of Minnesota)
- Elisa Sze (University of Toronto Libraries)
- Kerri Pfister (Frick Art Reference Library)
- Dean Seeman (University of Victoria)
- Robert Fernandez (Prince George's Community College, Wikimedia District of Columbia)
- Suzanne Pilsk (Smithsonian)
- Honor Moody (Harvard)
- Sarah Ann Adams (NYPL Metadata)
- Kurt Blythe, Kcblythe, UNC-Chapel Hill
- Amy Watson (Smithsonian Libraries)
- Alexandra Provo, Aap1890 (NYU)
- Steve Baskauf (Vanderbilt Heard Libraries)
- Anna-Sophia Zingarelli-Sweet (George Eastman Museum Menschel Library)
- Paromita Biswas (UCLA)
- David Bigwood (Lunar and Planetary Institute)
- Jill Crane (St. Mary's University (San Antonio, Tex.))
- Huda Khan (Cornell)
- Alex Jung (University of Toronto Libraries)
- Bree Midavaine (Philadelphia Museum of Art)
- Tricia Lampron (UIUC)

- Dan Belich (Harvard)
- Asako Shiba (Washington State Library)
- Kay Teel (Stanford University Libraries)
- Melanie Wacker (Columbia University)
- May Chan (University of Toronto Libraries)
- Nomsa Mathabela (University of Swaziland Libraries)

Agenda

- **Announcements**
 - Heritage Connector: Public webinar on Wikidata and cultural heritage collections, 14.00–16.00 BST, Friday 19 June 2020: [Info](#); [Registration](#)
 - [Library Data as Linked Open Data](#), Jason Evans, National Wikimedian of National Library of Wales
 - Next Wikidata Working Hour on Monday, June 22, 2020 at 8am PDT / 11am EDT / 15:00 UTC / 5pm CEST ([Time zone converter](#))
 - Zoom link for calls will be changing:
<https://stanford.zoom.us/j/91072562402?pwd=ckRaQUpyZ1cxM2N2bEJ6TjJnamx1dz09>
 - Password: 698682
 - [Call for co-facilitators](#) --open until June 19th
- **The Program for Cooperative Cataloging Wikidata Pilot**
[Slides](#)
 - PCC existed for 40 years
 - Collaborative effort in cooperative metadata
 - 700+ institutions
 - NACO
 - Authority control in PCC
 - Identity management using URIs for differentiation as opposed to text strings
 - Can have associated elements collocating on the fly to distinguish similarly named entities
 - PCC Strategy to accelerate movement towards identity management and creation of URIs
 - More identifiers and more entities we encounter in our daily work receiving identifiers
 - Collaborate with others who have similar goals
 - Pilot sponsored by PCC, but open to anyone
 - Attractions to Wikidata
 - Large size
 - LCNAF contains 11 million
 - ISNI 10 million
 - WorldCat Identities 62 million
 - Wikidata over 50 million
 - Scope--all types of entities
 - Bigger pool of collaborators

- Prospect of greater coverage of identifiers
- Existing gadgets and tools that can be used
- Learning Objectives
 - Interested in collaboration with non-library partners
 - Past pilots have included a pilot project with ISNIs
 - Current pilot URIs in Marc pilot--hybrid/transitional phase of library community moving to linked data
 - Many participants have a need to create new URIs for entities in their work
 - Production rate
 - Reality--always have more content/collections to make available and limited staff time
 - What would we learn
 - How would other tools we use fit in
 - Requirement for documenting sources for data elements
 - Proportion of needed entities already in Wikidata or need to be created
 - UI ease of use for data entry and workflow integration with other tools
 - How to relate one entity to another
 - How are duplicate entities surfaced and managed
 - Ease of finding and using documented best practices
 - Use of automation to use human intellect and staff time for things machines can't do well
 - Batch searching, loading
 - Quality assurance (QA tools)
 - How to use to minimize what needs to be done manually
 - Training sources
 - Which training materials best to get started and for ongoing development
 - Learning about the community to work with and contribute
- Anticipated Timeline
 - Collect input from community
 - June 23 review suggestions and complete planning work
 - Last week of June: Issue survey for those volunteering to participate
 - July 6: Initial deadline for collecting feedback
- Discussion
 - Jens Ohlig: Works for Wikimedia Germany and close to Wikidata Community--pondering options to support and connect with members of community; will discuss with others
 - Paul: How do you plan on working with cultural differences between PCC and Wikidata communities? PCC--Global North, resource-rich institutions, dominated by "big actors" and Wikidata community more diverse in geographic representation, language communities, institution types, and a decentralized "membership" Are PCC members ready to work in Wikidata environment?

- Michelle: Some members of task group have revealed some of these differences; thinks ready; have been discussing it for a long time; Excited to work with the community; some have already been participating and educating themselves about Wikidata and the community; will have a lot to learn, but that's part of excitement of launching pilot; will likely be some missteps, but hope can learn together
- Neil M Frau-Cortes: It should be noted that some of us are both active in Wikidata and also PCC participants.
- Larisa: Will PCC community have its own set of standards? Quality wise, selection of data elements, etc
 - Beth Camden--PCC Policy Committee member and Chair of the Standing Committee on Training; great thing about doing pilot is haven't set standards yet; group that works in the pilot will be shaping what might be our best practices, policies eventually; approach for training of using what is out there already is a great place to start rather than reinvent the wheel; down the road we might need more PCC specific
- Bob Kosovsky: One big difference between the cataloging culture and Wikidata culture: Cataloging tends to do it as best as possible the first time; Wikidata participants depend on each other to correct/emend. (Making it as correct as possible the first time is not uppermost among Wikidata editors
 - Michelle: Excellent point and one library community has been working towards as well; focusing on what do I need right at the moment, what am I trying to do here and having iterative and evolving
- Joy Panigabutra-Roberts: I am both a PCC member and started to contribute to Wikidata as well. I benefit from both communities for creating the linked data that can be used by the world's users.
- Eric--Thank you for this presentation! My anecdotal experience is that Wikidata/Pedia tends to be stronger on more popular and modern authors and their works. Libraries might be able to contribute through items from their special collections/local/rare materials items and authors, which might have more support from admin as increased discoverability. Just a thought.
 - Michelle: One of very large attractions for library metadata practitioners contributing to Wikidata is the scope and impact of the work and how they might have a broader impact themselves; not thinking just of our library catalogs, but of sharing our work as broadly as possible and blending it with the work of others
- Bob Kosovsky: Once PCC decides that the project is a success, will there be a report, and what will happen next?
 - Michelle: Going lightweight on reports--trying to operationalize learning--move it right into everyday experiences at work; hopes will be a combination of the two; definitely want to share learning of pilot and find best way of doing that

- Elizabeth: Have issues of scalability come up? I remember hearing that Wikimedia did not want OCLC to load all their identities into Wikidata.
 - Karen S-Y: Yes, might have; OCLC investigating scalability of Wikidata/Wikibase--still under investigation; discussing with Wikimedia Deutschland; progress on federating wikibase instances might take care of scalability
 - Paul Frank: doesn't think there's an issue because following model used in URIs and MARC pilot where attendees or participants curate data at individual level; might use machine manipulation down the road, but not to start out
 - Paul Priebe: Could you describe the relationship between the URIs in MARC pilot and the Wikidata pilot?
 - Michelle: Thought overlapping pilots would help by giving participants in URIs in MARC pilot platform through which to create new entities should they need them; thought would be mutually beneficial to each other to run in parallel
 - Naun: URIs in MARC is part of transition plan where we work with existing MARC infrastructure and interest of getting URIs into MARC is a way of bridging gap between what we're doing now and what we may want to do in the future--there's a need to relate what data we already have to other sources describing the same entities; could also see a move over time to rely on more external sources and less on authorities
 - Karen Coyle: Has anyone analyzed the differences between library "rules" relating to name forms vs Wikidata? Will this be a bone of contention between the two groups? Can this become a discussion between the groups?
 - Paul: Wonderful opportunity to get away from the boundary approach that we take to our identities and so much of a reliance on a labeled form of name where we can be using an identifier instead; sees a lot of ways for improving cataloging practices based on different models that are in place
 - Sarah Ann Adams: Trying to reconcile lists of indirect order names to Wikidata difficult and wonders if component of this project could be how we can batch update wikidata entities with aliases in indirect format to help with reconciliation with LC Authorities
 - John Riemer: Virtual International Authority File and possible connectivity between Wikidata and other contributions could be a way of reconciling different forms of names
 - Diane: We do include Wikidata entities in the clusters when they match another authority; I published some statistics on that and will have to dig them out; regularly harvest with the Wikidata and match it to the other sources, so that might be way to do that reconciliation

- Christine Fernsebner Eslao: It might also suggest adding direct order names, for the purpose of reconciliation, to our own data, or to build inversion into the recon service. Might suggest entering data for “given name” and “family name” properties might be good practice to support inversion.
 - John Mark Ockerbloom: I do wonder a bit about maintaining consistency importing inverted-name forms directly. Someone actually created a property that imports (inverted library-style) author names from my Online Books catalog w/o asking me, and I worry about those falling out of date.
- May Chan: University of Toronto Libraries has scaffolded engagement approach to share that might dovetail to the PCC pilot. I am happy to share about this to see if there are overlaps, which will inform our approach; University of Toronto libraries is new to contributing to NACO and Wikidata, so trying to take advantage of being new to both to learn things and correlate them; hoping in cooperation with archives to bring Wikidata training to our tri-campus environment; would apply training to describe local collections and apply alternative subject headings for indigenous language and indigenous resources because didn't want to use what was in LC subject file; idea of creating a NACO authority and Wikidata entity at same time; haven't worked out workflow yet, but still ironing out process with more traditional NACO workflow; Alex Jung has been helping with delivering some pre-training--very basic training to get people familiar; shares a lot of the points around goals that Michelle and John identified in slides; very short term goal is introduce staff to concrete way of creating and publishing linked data; very pragmatic approach--see which staff are interested in being part of the communities; Wikidata requires different skills sets than technical ones we're used to; Alex has been delivering training to Archives and then librarians in general and this summer one for catalogers; providing environment and opportunity for those who want to learn; welcome any recommendations on how to engage and train staff; wonders if approach might be similar to PCC pilot
 - Michelle: thinks there's a lot of commonality in the themes you discussed and a lot interest in learning by doing and diving right in
 - Alex: On my end looks like customized training delivery to different library/archives functions
 - Meg Wacha--City University of New York and Wikimedia NYC; been in library and Wikimedia communities for last 10 to 11 years; treatment of person data in Wikidata--operating with a policy originally developed for Wikipedia, which was describing certain people in very particular circumstances and as move to Linked Open Data there are a number of concerns around privacy and provenance of statements

that we create, so I was wondering what conversations, if any, has the PCC had regarding person data and how it may be treated during the pilot

- Michelle: PCC has had discussions to acknowledge those issues; doesn't think we have answers
- Paul Frank: came up at Operations Committee meeting; strong feeling in community that in our current cataloging standard, our data has opened up with the possibility of including a lot more information than might be necessary for bibliographic purposes and there are the beginnings of a move to pull back slightly; a lot was started from the GDPR developments in Europe with amount of personal data being put on web; we've seen an uptick in number of living authors asking for personal information be removed from their authority record and we abide by those requests; strong acknowledgement in the community that we are putting way too much information in our authority record; as we go into pilot this is something can take a look at outside of PCC practice/policy and implement much quicker because we don't go through bureaucratic hoops of setting a new policy; move away from including so much personal information for living people
- Alex Provo: Just want to plug for another affinity group hosted in the LD4 slack: ethics in linked data and folks are welcome to join the ethics group calls!
 - Christine Fernsebner Eslao:
<https://wiki.lyrasis.org/display/LD4P2/Ethics+in+Linked+Data+Affinity+Group>
- Karen S-Y: From both Project Passage and the recent OCLC Research Library Partners Archives and Special Collections Linked Data Review Group the PROVENANCE of statements in Wikidata becomes even more critical
- Meg Wacha: Thanks so much for that input! I think one of my concerns re: Wikidata is that very few know it exists and even fewer know how to request that the information is removed (if we provide a way to do so outside of the ticketing system)
- Christine Fernsebner Eslao: It's also a problem that very few people know how to go about asking LC to remove or change data.
- Chiat Naun Chew: Maybe also worth mentioning another aspect of the discussions about person identities, which is the workflows (where, when, whom by) by which those descriptions are created. It's generally accepted that traditional library authority procedures don't scale, even for library catalogues.

- Jim Hahn: One useful model I see for leveraging networks of expertise is to organize internal organization teams with expertise in Wikidata who can help support quality control and feedback. I was impressed with the DCN process for leveraging professional networks in data curation: <https://datacurationnetwork.org/resources/workflows/>
- Alex Provo: I saw WikiProjects mentioned on the training sources slide; I'm still learning about them/getting involved, but I think that might be a good way to work together in the pilot.
- Edward Hugh Summers: I apologize if this has already been covered: will the PCC collaboration be considering the possibility of running its own Wikibase installation that can interact with others like the one at wikidata.org?
 - Michelle: Talked about whether wanted pilot to be on Wikidata or dedicated PCC Wikibase; strong belief we would learn most by diving into the Wikidata community, learning from the community at large, and contributing broadly; there is interest in understanding how federated wikibases could have a role in the PCC in the future for dedicated uses, but not a specific goal of the pilot; want to have open and inclusive foray in to Wikidata
 - Jackie Shieh: I think it'd be a greater challenge for cataloging metadata professionals to accept <statements> that might not be 100% accurate.
- Joy Panigabutra-Roberts: Updating data in Wikidata is quicker; your data got published instantly. Wikidata also accommodate UTF-8 scripts - non-Roman scripts like Thai that is not yet available in LCNAF. Example with the scripts: <https://www.wikidata.org/wiki/Q62393357>
- Sarah Hovde: re: PCC/cataloging culture and Wikidata culture, many catalogers are accustomed to working in shared cataloging environments, where we can expand on/improve existing records as more information becomes available or as an institution prioritizes - there may be more commonalities of practice than we expect
- **Outstanding questions/comments**
 - Karen Coyle: What use will be made of the identities in wikidata?
 - Improving/augmenting data that is already there. Creating new entities. Any number of uses of the data might occur, some of which we cannot anticipate.
 - Jeff Young: the advantage of having a new source of identifiers is important. Stephen asked how use could be demonstrated before systems have been developed to make better use of Wikidata and identifiers. Jeff noted that linked data functionality may develop organically within the context of HTML as Google has been encouraging.

- Diane: VIAF has value as a query-able source for enriching the information about entities.
- Lori Robare: structured data enables new kinds of visualization and organization of metadata.
- John Riemer: the opportunity to increase the representation of under-represented groups and people.
- Christine Fernsebner Eslao: Has there been any coordination with the IFLA wikidata working group? <https://www.ifla.org/node/92837>
- Joy Panigabutra Roberts: We need to get our library platform vendors at the table. Our work can go unused by the platform and our effort may not be seen to have the ROI - impact on our local users.
 - Chiat Naun Chew: engagement with them is certainly important. But I suspect part of the attraction of wikidata/wikibase is also to reduce our dependence on existing systems that can take literally decades to change
 - Meg Wacha: This is an opportunity to move library infrastructure back to the control of libraries
 - P Dragon: What I took from Joy's comment is the important point that we need to be able to demonstrate the impact this work can have on our public before we are able to invest a lot of time into it.
 - Joy Panigabutra-Roberts: My problem is how to convince my superior to support our engagement now if I can't prove the ROI now. We have a small team and justification of additional work and learning is my challenge now.
 - Christa Strickler: Hard to convince colleagues and administration who don't understand metadata/linked data work.
 - Alex Jung: Value demonstration--This is an issue we also come back to in Wikimedian-in-Residence conversations, and from the survey Hilary and I put together recently, it was a strong "want" across a number of responses.
- Bob Kosovsky: Names are actually easy to do in both NACO and Wikidata. Is there any plan to deal with subjects as part of the pilot?
 - Michelle suggested that the pilot might go in unexpected directions if participants want to pursue them.
 - Naun wondered whether the question had some particular approach in mind for subjects
 - Hilary is following up with Bob for additional information.
- Neil M Frau-Cortes: I would like to see pcc bibliographic records to become linkable entities that we could use in order to make references in Wikidata entities. It would be so helpful to be able to quickly cite all those nice resources in print
 - Jeff Young: OCLC Number is a Wikidata property

- Lori Robare: her group has been using them to reference records in WorldCat.org. They have not been limiting their use to PCC records.
 - LCCNs are also available as identifiers and DOIs.
 - Paul Priebe: Any ideas regarding staff training while we're in a telecommuting environment?
 - Lori: training for Wikidata could work well in a remote-work environment. She found a lot of documentation and support for her work.
 - Hilary: some training resources: https://www.wikidata.org/wiki/Wikidata:WikiProject_Linked_Data_for_Production/Practical_Wikidata_for_Librarians#Training_Resources
 - John: Do others have experiences to share with regard to training in a virtual environment?
 - Alex Jung: Perhaps we could build out a Slack channel (or something else that works well) around PCC convo?
 - Michelle: Slack has been drawing a lot of people in and could be a good venue
 - Hilary: Slack channel for the Wikidata Affinity group has seen some use and could be attractive, especially to smaller groups.
 - Naun: wondered how many catalogers are using the Slack environment.
- **Next call Tuesday, June 30, 2020 at 9am PDT / 12pm EDT / 16:00 UTC / 6pm**

June/July LD4 Wikidata Affinity Group Calendar

- **Wikidata Working Hour**
 - Monday, June 22, 2020 at 8am PDT / 11am EDT / 15:00 UTC / 5pm CEST ([Time zone converter](#))
- **Affinity Group Call**
 - Tuesday, June 30, 2020 at 9am PDT / 12pm EDT / 16:00 UTC / 6pm CEST ([Time zone converter](#))
- **Wikidata Working Hour**
 - Friday, July 10, 2020 at 11:30am PDT / 2:30pm EDT / 18:30 UTC / 8:30pm CEST ([Time zone converter](#))
- **Affinity Group Call**
 - Tuesday, July 14, 2020 at 9am PDT / 12pm EDT / 16:00 UTC / 6pm CEST ([Time zone converter](#))