

2021-02-05 Cornell LD4P3 Meeting notes

Date: 05 Feb 2021

Attendees: Jason, Tim, Huda, Lynette, Simeon, Steven

Regrets: Greg

Last meeting: [2021-01-22 Cornell LD4P3 Meeting notes](#)

Agenda

1. OK to move meetings that conflict with FOLIO Fridays to 9-10am? – Will do it, try to highlight
2. Quick updates as below
3. What should the next thread for production work be?
4. [Jason Kovari](#) to lead discussion functional requirements efforts

Discovery (WP3)

- <https://github.com/LD4P/discovery/projects/2> for issues etc.
- Draft of a discovery plan: https://docs.google.com/document/d/1zKYW7FQVVNvyd0XjjW0qWznX9PC3jbmOE6Kz_yygPjs/edit?usp=sharing
- **Strand 1: production piece**
 - Production requirements and functionality – [Production decision points](#)
 - Implementation work
 - 2021-01-29: In Production - DONE
 - Assessment: outside of LD4P; consider engaging D&A group with usability testing and/or user feedback. For browse, melissa put in some tracking to monitor user interaction with browse. can we do something similar with Discogs (e.g.: are people clicking the highlight link and/or the link to get to discogs? What have we learned from virtual shelf browse tracking)
 - **ACTION ITEM:** Tim will speak to Melissa about tracking for the Discogs link.
 - 2021-02-05 Still need to follow up with Melissa DONE (2021-02-05)
 - **ACTION ITEM:** [Tim Worrall](#) will raise usability testing for D&A queue (don't carry forward in notes as now outside of LD4P)
 - 2021-02-05 This won't happen until the next sprint, starting second week of March
 - ACTION [Huda Khan](#) [Tim Worrall](#) [Steven Folsom](#) to develop a post for #general, #discovery and partners email about Discogs going live
- **Strand 2: research: how to go from knowledge graph to an index**
 - [Research decision points](#), [Use cases](#)
 - First goal: DASH! dashboard (full page for entity) that extends on the idea of an embedded knowledge panel, aim to have functional prototype for end of year
 - DASH! (Displaying Authorities Seamlessly Here)
 - Dashboard design meeting kickoff [notes](#) - will also try to understand what our data will support or connections to other data sources
 - https://docs.google.com/document/d/1PgQi3xobsPhr9DUHU_YGeimL1OjNiiTdklNWb36r3Gg/edit
 - 2021-01-29: Huda working to get scripts in place to populate index; bringing in period-O info; focused on locations with Wikidata URIs for consistency. Subject headings: script that takes-in components & breaks those out... and parses into timeline info. On Dave's fuseki, 34 distinct temporal terms with labels. Will finish today with actual index. Will break-down the loading to increase load speed
 - **ACTION ITEM:** adding Wikidata URIs for any subject headings and broader/narrower URIs to index (today) - DONE for dev VM index, to do next for LD4P3 Solr index (will be done today)
 - 2/5/21: Dev VM index populated with the content above using a script that will be reused for populating the LD4P3 discovery solr index.
 - Each LCSH now has a Wikidata URI where available, broader and narrower URIs where available, periodO extracted spatial URIs and time periods where available, a list of components for multi-part subject headings with URIs and labels for each component, separate lists for temporal components with start and end times mapped, and geographic components with LCSH URIs and equivalent Wikidata URIs extracted and added.
 - One error (for review perhaps later): two of the LCSH temporal component labels have special characters and are not being matched
 - Reached out to IRB to ask about testing: if we want to disseminate results as research data, need to do IRB protocol; has a follow-up. Waiting to hear back but will submit protocol if no word. Simeon's interpretation of reply is that we are crossing line into research and the approval will likely be positive. Depending on how we describe what we're doing it either falls under research OR improving a product... but we're essentially doing research to improve a product so yes to IRB review.
 - **ACTION ITEM:** IRB did respond and say they wish us to proceed with sending in an application. Huda will work on this and reach out with any questions if needed.
 - 2021-01-29: Next piece to go into production? discuss in Discovery on the Ground and then here
 - For reference: [Use Cases](#) and [production idea](#) discussion
 - Tim was looking to reuse some of the ajax calls Huda did for knowledge panel; reworking portions of HTML but focused on getting LoC URI and calling wikidata. making good progress. Considering layout and not duplicating what we already have in catalog. Working on authors/subjects. Wants another week to work on this
 - 2021-02-05
 - What would D&A user reps favor?
 - Concern that full KP linked from ⓘ button is too much
 - Is "KP-lite" on autosuggest a good route? We think users would find this valuable. Are there options that minimize index changes?

- What warrants a KP?
- What is the redundancy between KP work and DASH!? Does dashboard mean a fundamental change or is just an enhanced KP?
- We need to be aware of which options require significant indexing changes. There is already a sense that we want to add ids to the index
- What about the open syllabus project? This relied on the open syllabus API, not sure whether it is available in LD. Essentially a mapping from domainCSIP codes ISBN, very few wikidata connections
- What would be the smoothest next step for production?
- Which option would give us real linked data connections via URI?
- Steven notes that LTS authorities in FOLIO group is looking at the insertion of URIs into MARC records (resources willing)
- ACTION - [Huda Khan Tim Worrall](#) document options and implications as preparation for user reps presentation in order to get a steer on where to continue experimentation with a view to future implementation
 - Dashboard (perhaps for some entity types only)
 - Autosuggest with KP-lite
 - Regular facet with KP
 - Open syllabus related items
- ACTION - [Huda Khan](#) to line up meeting with D&A user reps

Linked-Data Authority Support (WP2)

- [Qa Sinopia Collaboration](#) – Support and evolve QA+cache instance for use with QA
 - 2021-01-29: no Stanford meeting. Still 5 issues; one for GeoNames that does not pass test though target URI is there. Remaining 4 are in Dave's court. Steven checking whether any other authorities have line breaks; Dave considering generalized approach to repair the line break in the triples. Exploring whether problem in Jena Fuseki or if it is an artifact of making this a blob in the indexing process.
 - 2021-02-05: Met with Stanford. The discussion focused on closing the loop to get ShareVDE data into Sinopia. Specifically, this is data derived from Stanford catalogers submitting RDF translated from their MARC data that was submitted to ShareVDE. The plan is for Stanford to take the lead to establish connections with ShareVDE to start up meetings focused on this topic. We documented questions in meeting notes.
- [Cache Containerization Plan](#) - Develop a sustainable solution that others can deploy
 - 2021-01-29: lots of documentation and templating work; one critical piece needing completion (getting service that takes all pieces together and keeps container running in service into a template). rest is good. nearing point where we can put in front of others and say 'give it a spin'. Likely finishing up the template of last critical service piece for next week
 - 2021-02-05 Greg says "
- [Search API Best Practices for Authoritative Data working group](#)
 - 2021-01-29: met and did informal voting in group for next topics; also has survey (posted in Slack) that has descriptions of each area, potential areas and video discussion in meeting to ensure sense of the potential topics. Will then send to broader group to get sense of what broader community wants as next steps. 4 areas:
 - 1. change management (source data to consumers), 2nd (tie) among group
 - 2. language processing (anything to do with language),
 - 3. moving forward from what has been done (user stories to specific recommendations),-- 1st among group
 - 4. Linked data specific approaches and what implications does it have for our data being in LD, what are the tools and what can be leveraged (producer side)? 2nd (tie) among group
 - 2021-01-29: plus delta done. plus saying everything well organized and a delta saying that everything was really well organized. steering of conversation based on what was brought to meeting ahead of time. Concerns expressed around whether time will be allocated to adopt recommendations
 - 2021-01-29: work done on supporting cataloging inclusion (categorized version of user stories); meeting one more time next Monday to work through rest.
 - 2021-01-29: got to a good place with good output!
 - 2021-02-05 Met again on Monday, some small refinements (advance search and filter are interconnected). Expecting to get review finished by today. [Steven Folsom](#) will send document to PCC, [E. Lynette Rayle](#) will send announcement to Samvera and put a note on #general, also link from main LD4P page.

Developing Cornell's functional requirements in order to move toward linked data

- C.f. Stanford functional requirements document: https://docs.google.com/document/d/18H6zYGwKuG3SZqm9Q_cxkZThcdmBjknE6HdtQ-RRzk/edit#heading=h.4fu64x8jzm6e
- What does success look like? And then how do we get there?
- Miro board (diagramming): https://miro.com/app/board/o9J_lfXUUj8/
- Notes space: <https://docs.google.com/document/d/1TVPBfak7DkfjBptkl-pCMWQnOaiWHB0XCHswiB3Fr9g/edit?usp=sharing>
- 2021-02-05 discussion
 - Purpose? Vision for mid-term (3-5 years) transition to support linked-data at Cornell. May include things we don't yet have or cannot yet do, but not long-term vision of post-MARC environment
 - Important to understand sources of truth (primary data) and where there is derivative data
 - Imagine landscape with items described in multiple formats including at least MARC, BF, DC (eCommons), JSTOR
 - Imagine all items indexed and discoverable via D&A
 - Functions of "Aggregated index, allowing pivoting & ETL"
 - Includes current functionality of Frances' indexing
 - Does it include any editing?
 - Is there interaction with CULAR?
 - Includes indexing associated with DCP
 - What interfaces or functionality do we expect for the connecting lines?
 - Do we need a diagram for now (or at least July 1, 2021 with Voyager gone)?

Other Topics

- OCLC Linked Data / Entities Advisory Group
 - Request for UI and API testing from Jan 25

- Lynette has Cornell key (a WSKEY) for testing
 - Call discussed seeding of data. Data for person includes VIAF and other sources; place includes geonames. Steven, Huda, Jason and Lynette signed up for user testing
 - 2021-01-29: User testing happening now
 - 2021-02-05 Have been lots more details of testing hopes etc.
- PCC - Sinopia collaboration
 - 2021-02-05 Charge to form a new group for documentation, mentoring etc is under reviews
- PCC Task Group on Non-RDA Entities
 - 2021-01-15 PCC reviewed proposal but no decisions made yet, looking at description wrt cataloger use, discussion will continue
- Default branch name - WAIT until we can use github tools January 2021
 - Lynette is signed up to be a beta tester for github. Target for public release is end January
 - Github have provided the change button, Lynette has provided feedback as a beta tester. There is an issue that after the change to `main` someone can repush the `master` branch, has suggest a deny list of branch names. Links in comment will get forwarded to `main` unless there is a new branch with the old name
 - For forked repos there will be a message
 - What will it take to make the change to `main` for LD4P repos?
 - Lynette has documented [HERE](#)
 - Lynette will rename for qa_server, work with Greg on qa_server container; document on LD4P wiki
 - For our blacklight fork we should wait until the D&A team change
 - 2021-01-29: no updates. working group for Samvera meeting next week... Samvera is discussing moving away from GH as a whole. QA server investigation – quite a few forks off QA server so not sure it is the best to start - Lynette will reach out to the fork owners and document the aspects. Tech side is not challenging, community side is.
 - 2021-02-05 [E. Lynette Rayle](#) still to reach out to fork owners

Upcoming meetings

- <https://kula.uvic.ca/index.php/kula/announcement/view/1> . Call for Proposals - Special Issue: "The Metadata Issue: Metadata as Knowledge". Due January 31, 2021 (abstract 300-500 words). Includes "The use of linked open data to facilitate the interaction between metadata and bodies of knowledge" and "Cultural heritage organization (libraries, archives, galleries, and museums) and academic projects that contribute to or leverage open knowledge platforms such as Wikidata"
 - Folder [Link](#), [CFP + Brainstorming](#)
 - 2021-02-05 Huda submitted abstract
- [code4lib](#) - virtual next year
 - Expecting to attend: Huda, Steven, Lynette
- Lynette doing a QA presentation at Samvera partner call in June
- Mann All-Staff : Huda (Dashboard), Tim (Discogs) and Lynette (QA) presenting next Wednesday at Mann All-Hands meeting (2/7)
 - Done, positive feedback
 - Chat comments about dashboard, questions about how that works for e-resources. Jason thinks that for journals we may have good data but the subjects will be higher-level unless the journal is very focused - but won't have article data that we can manipulate. For e-books there are options and we might consider at-scale enhancement perhaps with connection to physical book

Next Meeting(s), anyone out?:

- ...