

2021-04-16 Cornell LD4P3 Meeting notes

Date: 16 Apr 2021

Attendees: Lynette, Huda, Tim, Steven, Simeon

Regrets: Greg, Jason

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
- **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_YGeimL1OjNiiTdkinWb36r3Gg/edit
 - 2021-04-16
 - Tim working on changes to entity pages for authors, may incorporate some usability feedback
 - Future work: should review work based on usability tests. Perhaps make a few more changes (e.g. entity pages for subject and authors look more different than they need to (e.g. relations back to catalog)) the wrap this work until more discussion in August
 - Usability testing for DASH
 - 2021-04-02 - Four people responded to request and have scheduled tests for next week (Apr 7 & 9), will review entity pages
 - 2021-04-16 - Usability testing is underway, will finish next week. Useful input on what users see/understand and what they struggle with. Users don't know what a repository is, can guess what digital collections means. Users tend to think that the author name link will get more information, don't see/understand info button – could mean adding tool tips and/or making button more obvious or even flip button/link or one link for info+search
 - User reps D&A meeting: Need to re-follow up
 - [Slides](#):
 - Had user reps meeting 2021-04-09 and result was "not no"
 - Positive feedback on linking back to catalog results
 - Questions about links and display from wikidata, other sources
 - End was to discuss more in August with possible discussion of how to move into production
 - User reps have slides and wanted demo link
 - Could have lite KP might to start, possible later entity pages – or perhaps both together
- Next Demo that Needs New Acronym
 - Expect to include Works. Need to do something beyond what we already have live from the OCLC concordance data.
 - Full OCLC concordance us 343M rows, and gzipped the file is 3.3GB
 - SVDE Works
 - 2021-02-26 Have to develop SPARQL queries to pull out certain sorts of connected Work. Don't expect data to be very dense but do expect that we would get useful connections between print and electronic for example. We already have a link based on the OCLC concordance file from several years ago.
 - ACTION - [Steven Folsom](#) and [Huda Khan](#) to work on building an equivalent of the OCLC concordance file based on SVDE data and then do a comparison to see how they are similar and different
 - 2021-04-02 Steven and Huda met to think about putting together queries to extract a similar dataset. ([Document for recording queries](#)). Open questions about the counts – got 16k works from one view, got about 8k where limited to case with at least one instance. These numbers are much much lower than expected
 - 2021-04-16 Steven working with Dave on how to pull our SVDE data. Dave still working through some errors in ingest of SVDE data – this needs to be resolved before looking for concordance. Has asked Frances for 2015 concordance
- eResources
 - Discussion with small group regarding call number classifications on e-resources, where there is/isn't sufficient metadata compared to equivalent/related physical objects. Possibility of gathering some useful examples around needs/wishes around e-resource discovery.
 - 2021-03-26 - Other CUL staff met a couple of weeks ago. Examples where print and e-resource versions have differing metadata quality. Can we connect the resources better and/or copy metadata from one to the other? Has connection to our work to identify different versions. Nothing for LD4P3 to do until they have come up with appropriate examples
 - How do our new mockups support ideas of local overrides or correction to address DEI issues?
 - Question of process for reporting and dealing with errors
 - Use of wikidata means that individuals can make some changes there - direct citation in the data makes that plain and more easily actionable
 - Subject headings
 - Currently we don't show how many other headings a sub-heading appears in, this could be calculated
 - E.g. if someone is looking at "Architecture - Norway" would it be useful to see other headings that include Norway. Or would a use simply search for Norway?
 - Questions about whether some types of subheading (such as temporal or spatial) might be more usefully connect than others
 - Perhaps see if this comes up in the testing
 - 2021-04-16 - When we have better understanding of Work data, go back to group to discuss possibilities for e-resources
- Publisher authorities/ids
 - At Cornell we haven't tried to connect authorities with publishes
 - LC working on connecting to publisher identifiers - utility is things also published by a publisher
 - Also possible interest in series and awards

Linked-Data Authority Support (WP2)

- [Qa Sinopia Collaboration](#) – Support and evolve QA+cache instance for use with QA
 - 2021-04-16
 - Released to production the new indexing scheme. This release cleans up all tests with known failures marked as pending. Added in a few UI improvements to remove inactive authorities from the Monitor Status page, adjust the threshold for marking an authorities uptime as concerning, and on the Check Status page show a dogear for pending accuracy tests.
 - Need to think about how best we represent recent status for authorities https://lookup.ld4l.org/monitor_status. Lynette has been working on two thread: uptime fraction and performance/speed
 - Agree to change window to last month to give better indication or recent status
 - In the weekly meeting, we identified the ShareVDE/PCC data interactions as a topic for the larger group in the quarterly meeting next week. We also discussed potential topics around lookup enhancements. The following topics would be discussed in the weekly Qa/Sinopia meeting.
 - compare list of priorities identified in the working group with what's available now in Sinopia
 - review Astrid's mockups and compare with Sinopia current UI
 - discuss and prioritize additional UI improvements
 - discuss which authorities are supported and whether we should be looking at additional authorities
 - In ShareVDE meeting, it was highly focused on the needs of the Stanford team. Discussed question of how to transform data in and out of SVDE
- [Search API Best Practices for Authoritative Data working group](#)
 - 2021-04-16
 - Membership is complete with 14 members. Producers include LOC, OCLC, MeSH, ShareVDE who have an interest in the creation of change management documents. Dave is joining the group to represent the needs of downstream consumers of change management documents. There is a poll out to determine the date of the first meeting.
- [Cache Containerization Plan](#) - Develop a sustainable solution that others can deploy
 - 2021-04-09
 - Greg: Working on permissions issue. Need to rethink approach.
 - Need certain amount of permissions to create whole suite, including ability to create permissions. Somewhat circular.
 - Considering breaking permissions out into a separate piece.
 - Create these for yourself or give to admin to install for you
 - If using admin permissions, do not need this piece. Not recommended.
 - Goal: granular set of permissions recorded in document.

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

- C.f. Stanford functional requirements document: https://docs.google.com/document/d/18H6zYGwKuCg3SZqm9Q_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_lfXUJj8=
- Notes space: <https://docs.google.com/document/d/1TVPBFak7DkfjBptKI-pCMWQnOaiWHB0XCHswiB3Fr9g/edit?usp=sharing>
- 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
- 2021-04-02 Jason/Steven/Simeon discussed Jason's work on a merged CUL ecosystem diagram, plan to discuss on 2021-04-16

Other Topics

- PCC/Sinopia and SVDE shape analysis
 - 2021-03-19 Steven has been working through a spreadsheet of 400+ lines to compare the shape of SVDE data with the PCC/Sinopia profile. He is finding that there are many many differences which will severely limit how well Sinopia will be able to consume and edit SVDE data. For the purposes of QA/Sinopia cloning, Steven could come up with some lpaths but not sure whether the amount of data will be useful. Steven expects to be able to share the spreadsheet at the next Sinopia/SVDE meeting. Going forward we need to consider the role of versioning/documenting shape changes and validation at both scale and single descriptions. Justin's validation scripts: <https://github.com/LD4P/dctap>. Tom Baker's csv2shex: <https://github.com/tombaker/csv2shex>
 - 2021-03-26 Steven finished working through the spreadsheet comparing SVDE data with the PCC profile. Notes that he is looking only from the side of the PCC profile and would thus miss other things in SVDE data. Patterns around different types of work in SVDE data (e.g. Opus and other higher level works have very different shapes). Difficult pattern of double-reified relationships between works. Steven will let SVDE/QA folks know about completion of the work. Need to find a way toward alignment.
 - ACTION - [Steven Folsom](#) to write up state of current analysis and store a snapshot of the spreadsheet on the LD4P3 wiki
 - 2021-04-02 - Started...
- OCLC Linked Data / Entities Advisory Group
 - 2021-04-02 Michelle asked about connecting QA to the OCLC Entity Backbone as part of updates for partner meeting, Lynette has reached out about API
 - 2021-04-09 Lynette has not heard back yet from OCLC
- PCC Task Group on Non-RDA Entities
 - 2021-03-19 Group headed by standing committee on standards will formally propose a list of non-RDA entity types. Steven will join. Deliverables by June
 - 2021-04-02 Many participants involved in ILS/LSP migrations so work delayed until July
- Default branch name - Working through repositories in [Renaming of LD4P Repositories](#)
 - Created [Renaming of LD4P Repositories](#) page to identify Cornell repos, provide instructions, and track progress.
- Authorities in FOLIO
 - Hope to include URIs as part of Cornell FOLIO migration, possible LD4P work
 - 2021-04-09: CUL-IT in favor of proposal but resourcing will have to wait until after FOLIO. (URIs are cool.)

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](#)
 - 2021-03-26 Abstract accepted, paper due June 25
- **LD4 Conference 2021** - proposals due April 12 – brief descriptions, ~200 words, with structured questions too
 - Discovery - suggestion of discussion form
 - Steven thinking something around shapes and compatibility and round trips
 - Lynette/Greg/Dave - containerization, should have documented product by then
 - 4/9: Submitted abstract for presentation
 - Lynette – possibly something about the working group, perhaps updated version of code4lib
 - 4/9: Submitted as a lightning talk
 - [Document for brainstorming](#) (in case anyone wants to use it)
- Lynette doing a QA presentation at Samvera partner call in June

Next Meeting(s), anyone out?:

- 2021-04-23