

2021-04-02 Cornell LD4P3 Meeting notes

Date: 02 Apr 2021

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

Regrets: none!

Discovery (WP3)

- <https://github.com/LD4P/discovery/projects/2> for issues etc.
- Draft of a discovery plan: https://docs.google.com/document/d/1zKYW7FQVNVydxjW0qWznX9PC3jbmOE6Kz_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
 - 2021-04-02 Tim has implemented a D3 based graphical representation for Influenced by and influence for. Currently not limit to the number of related entities shown (e.g. many many influence-for for William Faulkner). When no examples shown the header is not shown, where there is no image (or not a png or jpg) then a circle is shown. Discussion:
 - Agreement that we should put a limit (perhaps 12) on the number of influenced-by or influence-for entities, ideally with a count of total and link to show all
 - Debate about best way to show all for long lists
 - Suggestion that we need to make the data source plain because the data is debatable/subjective and often incomplete
 - Would be good to test which related entities are in catalog and add links to the appropriate info pages
 - Ideally we will not show the Influence tab when there is no data for influenced-by or influence-for entities
 - Considering KPAOW zero (streamlined knowledge panels). Have begun discussing what should go in a streamlined version.
 - [Huda Khan](#) Scheduled meeting with D&A reps April 9th, 9-10 (Moved the LD4P3 meeting to this time)
 - Neither Jason nor Simeon are available for the meeting
 - Usability testing for DASH
 - Getting ready for usability tests in April
 - Need to finalize usability tasks for both authors and subjects
 - 2021-04-02 - Four people responded to request and have scheduled tests for next week (Apr 7 & 9), will review entity pages
 - User reps D&A meeting: Need to re-follow up
 - [Slides](#):
 - Good to have knowledge panel lite mockups or examples ready to show
 - [Go to link](#) Started some mockup brainstorming here: <https://jamboard.google.com/d/1PXLau0XTzaQhe9vw6OcAdsQJltFJ4U-FStgm9IA1CHc/viewer?f=0>
 - Also need to show entity page examples
 - SVDE Works
 - 2021-03-19 Steven has looked at SVDE works and is thinking about how to find works via ISBN to then find other Instances. Is looking via SPARQL on Dave's ingested data
 - 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
 - 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

Linked-Data Authority Support (WP2)

- [Qa Sinopia Collaboration](#) – Support and evolve QA+cache instance for use with QA
 - 2021-04-02:
 - No meetings this week for QA/Sinopia this week.
 - Dave and Lynette met to talk about remaining issues. Dave made progress on some a Lynette looking into a label issue on the QA side. Open questions about MeSH that Dave is looking into
 - Plan to do before and after comparison to discuss before moving into production
 - Lynette will give an update on this work package for the partner meeting; have suggested that Sinopia team will give update on UI side
- [Search API Best Practices for Authoritative Data working group](#)
 - 2021-04-02
 - Interest in group heavily skewed toward cataloger/curator side, hoping to get more from authority providers and developers
 - Will also reach out to AGROVOC because their structure is strong and easy to work with, they handle languages very well. Results lack ranking however
 - Expect to solidify membership next week and then schedule meetings
- [Cache Containerization Plan](#) - Develop a sustainable solution that others can deploy
 - 2021-03-26
 - Greg and Lynette have been working together on [documenting the deployment to AWS](#) and this has been very useful in uncovering missing things. Greg found that he needs to make sure AWS actions can be performed by a less powerful user
 - Greg has discussion containerization of cache infrastructure. Greg has some tasks to help with
 - 2021-04-02
 - Lynette notes mimemagic issue with Rails that derailed (pun intended) much
 - Greg had to work around issues with env file that connected container to wrong DB, not yet sure of scope of the problem created
 - Continuing work on documentation and permissions
 - Have given Dave and empty container as a starting point for him to start work when ready, don't expect this yet
 - What will it take to get the indexing process containerized? For LCNAP this will take significant resources, incremental updates will be important. Should perhaps start small to get the process going for something small (e.g. RDA registry). Would be good to start working on this.

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

- C.f. Stanford functional requirements document: https://docs.google.com/document/d/18H6zYGwKuKg3SZqm9Q_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_lfXUJ8=
- 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
- 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
 - 2021-03-19 In LTS there is a task group that has a proposal for authority management in FOLIO (absent new features). Being reviewed with request for scripting work to create reports etc.. Includes insertion of URIs into MARC
 - 2021-04-02 Jason meeting with Debra later today

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
 - Lynette – possibly something about the working group, perhaps updated version of code4lib
 - [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?:

- 2021-04-09 - Jason and Simeon not available, Huda will lead meeting to discuss with D&A reps