

2021-04-23 Cornell LD4P3 Meeting notes

Date: 23 Apr 2021

Attendees: Lynette, Tim, Huda, Steven, Simeon

Regrets: Greg, Jason

Partner Meeting Reflections

- Like presentations on template alignment and data sources from Nancy and Steven
- Understood that templates are specific about whether we expect a bnode or named resource. Tools need to know this. Both SHACL and ShEx support these distinctions
 - In Sinopia bnode might indicated that one doesn't want a separate form
- Understood that we don't have partner agreement on some decisions
 - What is the role of the different SVDE datasets?
 - Shapes of data: Sinopia/SVDE/LC – all differ
 - Expectation of what ecosystem supports in short- and long-term. Who edits what/where? Connections to data in existing library systems
 - What is role of PCC data pool (and how does SVDE implementation support that role)?
 - What does it mean to have local "changes" that overlays a shared resource? How is that described (e.g. a deletion)?
 - Do we intend to keep MARC and BIBFRAME in parallel for the same items? If so, are those sync'd?
- How to get agreement?
 - First document the approaches/options in more detail. Use pictures
 - Steven's dataflow questions: [SVDE PCC/Sinopia PCC Template Analysis, March 2021](#)
 - What does "local" mean?
 - Learn from the way we authorities to talk about reuse workflows – Works and Instances will be used through linking the same way we have for a long time with authorities; we make edits in the primary copy of these data (e.g. in LC NAF for name authorities, in SVDE for the CKB, etc.)
 - Need to better describe caching needs in our linked/reused view (c.f. working group on change management to support this efficiently)

Discovery (WP3)

- <https://github.com/LD4P/discovery/projects/2> for issues etc.
- Draft of a discovery plan: https://docs.google.com/document/d/1zKYW7FQVVNvyd0XjvW0qWznX9PC3jbmOE6Kz_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_YGeimL1OjNiiTdklNWb36r3Gg/edit
 - Usability testing for DASH
 - 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
 - 2021-04-23 - Five people helped with usability tests. Setting up final results document. High level:
 - Uncertainty of info button vs author link (or lack of understanding if same/separate). Tim has created issue in D&A
 - Good understanding of author page in general, opportunities for some clarification in labeling etc., influences tab well understood
 - On subject page the timeline scroll/zoom fast creating interaction problems, broader/narrower mixed understanding (difficult when subject is an event), "show all subject" checkbox not found easily, timeline/map connection was understood
 - User reps D&A meeting: Expect next follow-up in August
 - [Slides](#): from 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
 - 2021-04-23 How much more work to do before August? Candidates:
 - Follow up on results from usability tests
 - Work on how smoothly the prototype works
 - Think about more consistent look between pages - Tim thinking of making some mockups of redesign
 - Video for DASH!, theme?
- BANG! (Bibliographic Aspects Newly GUI'd)
 - 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 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
- 2021-04-23 Waiting on indexing of PCC data, have learnt more about the basis for the old OCLC concordance file
- 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
 - 2021-04-23 Might be able to use LC publisher ids in BANG!, Steven will look at whether there is a dump available

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
 - 2021-04-23
 - Open questions from partner meeting
 - Lynette plans to review open issues by testing against
 - Lynette would like to do comparison of Sinopia's current UI for lookups with suggestion from best practices working group
- [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.
 - 2021-04-23
 - First meeting May 10, same time slot as old group. Will look at use cases first (keeping cache in sync, keeping cached labels in sync in application, URIs cached in application)
- [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_ifXUUj8=/
- 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
- 2021-04-23 Working in Miro board

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.
 - o 2021-04-23 Write-up complete: [SVDE PCC/Sinopia PCC Template Analysis, March 2021](#)
- OCLC Linked Data / Entities Advisory Group
 - o 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
 - o 2021-04-23 Still nothing back from OCLC, need to reach out again
- PCC Task Group on Non-RDA Entities
 - o 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
 - o 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](#)
 - o Created [Renaming of LD4P Repositories](#) page to identify Cornell repos, provide instructions, and track progress
 - o 2021-04-23 Waiting on <https://github.com/cul-it/blacklight-cornell> to change, not yet planned
- Authorities in FOLIO
 - o Hope to include URIs as part of Cornell FOLIO migration, possible LD4P work
 - o 2021-04-23 Likely going ahead in August

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"
 - o Folder [Link, CFP + Brainstorming](#)
 - o 2021-03-26 Abstract accepted, paper due June 25
- [LD4 Conference 2021](#) - response on proposals will be announced April 30; conference is July 12-23
 - o Discovery - suggestion of discussion form
 - Huda submitted status of discovery update, BOF for DAG, and discussion of opportunities for effects and use of linked data (60min discussion)
 - o Lynette/Greg/Dave - containerization, should have documented product by then
 - 4/9: Submitted abstract for presentation
 - o Lynette – possibly something about the working group, perhaps updated version of code4lib
 - 4/9: Submitted as a lightning talk
 - o [Document for brainstorming](#) (in case anyone wants to use it)
- BIBFRAME in Europe workshop - September 21-23 15:00–18:00 CEST = 9am-12noon EST
 - o Call expected early May, by late June
- Lynette doing a QA presentation at Samvera partner call in June

Next Meeting(s), anyone out?:

- 2021-04-30