

# 2021-06-11 Cornell LD4P3 Meeting notes

Date: 11 Jun 2021

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

Regrets: Simeon

## Misc.

- For 6/25: SWIB and Euro BF proposals due very shortly

## Discovery (WP3)

- <https://github.com/LD4P/discovery/projects/2> for issues etc.
- Draft of a discovery plan: [https://docs.google.com/document/d/1zKYW7FQVVNvyd0XjW0qWznX9PC3jbmOE6Kz\\_yygPjs/edit?usp=sharing](https://docs.google.com/document/d/1zKYW7FQVVNvyd0XjW0qWznX9PC3jbmOE6Kz_yygPjs/edit?usp=sharing)
- **Research: how to go from knowledge graph to an index**
  - [Research decision points](#), [Use cases](#)
- DASH! (Displaying Authorities Seamlessly Here)
  - Dashboard design meeting kickoff [notes](#)
  - User reps D&A meeting: Expect next follow-up in August ([Slides](#): from user reps meeting 2021-04-09 and result was "not no")
  - [https://docs.google.com/document/d/1PgQi3xobsPhr9DUHU\\_YGeimL1OjNiiTdkjNWb36r3Gg/edit](https://docs.google.com/document/d/1PgQi3xobsPhr9DUHU_YGeimL1OjNiiTdkjNWb36r3Gg/edit)
  - Usability testing and followup for DASH: [Usability results](#)
    - 2021-05-21
      - [Usability results](#), a few little things to finish up
      - [GitHub issues](#)
    - 2021-06-04
      - Working on a few DASH! bugs, want to be able to turn some features on/off (e.g. influences) for user reps to consider options
      - How long to continue working on DASH! ? Tim and Huda to discuss what could be done by mid-July, discuss next week
        - 2021-06-04 Discuss next week. A motivation for DASH! work is to have more to show to user reps in August. For bigger items perhaps show "paper" designs for feedback
        - 2021-06-11 Tim and Huda had a meeting to discuss priorities and approach. Main concerns were showing user reps advancement in features and . Tim working on mockups on redesign; will not implement – but will present to user reps to decide which options. Not prioritizing anything functional at-present. Will make sure pages work well when there is not enough data on the page. Will put prototype in position that we feel more comfortable when people play with it. Concerned that if show same thing at user reps, will be unproductive... but question whether anyone will remember. With more robust prototype, hope will yield more of a decision. Tim does not have much time to work on this (2.5-3 weeks) – aiming for user rep meeting in August.
  - Video for DASH!, theme?
    - Sonic? Roadrunner?
- BANG! (Bibliographic Aspects Newly GUI'd)
  - [Jamboard link](#)
  - 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
      - 2021-04-23 Waiting on indexing of PCC data, have learnt more about the basis for the old OCLC concordance file
      - 2021-05-07 Steven didn't have much luck getting data from SVDE, learning GraphQL endpoint but also problems with timeouts there (HTTP 503)
      - 2021-06-11: At impasse. new modeling is represented in GraphQL data but fuller data are in RDF. Need to talk to SVDE when have QA/Sinopia conversation. Asked for test data but unsure when we'll have it all. Could consider doing this via Stanford Institutional data - though not ideal. **ACTION**: Steven will ping Anna to inquire on existing thread
  - What is the space of Work ids that we might use and their affordances?
    - OCLC Work ids, SVDE Opus (Work), LC Hubs (more than Hubs), what else?
    - Connections to instances, how to query, number
    - 2021-05-07 ACTION - Huda to start analysis
  - Other SVDE entities
    - 2021-05-07 ACTION - Huda will reach out to Jim Hahn about entities other than Works represented in SVDE - DONE
    - Summarized here: [Jamboard link](#) - U Penn Enriched Marc: Work Ids in 996 Field. 1.2 million with OCLC Work IDs in > 1 description. ~3.9 million with OCLC Work IDs in only one record.
  - 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
  - 2021-06-04
    - To plan BANG! we need to think about what can be done with the available data. Perhaps take some concrete examples to consider what LC and SVDE data might give us, no longer sure what we could do with current OCLC works data (hope that entity work will provide new data later)
    - What about providing users with better access using alternative labels etc. that might better match their expectations, including different languages via VIAF connections. Much of our catalog data around languages is very bad because we use roman transliterations based on LC rules that are not well sync'd with actual practices in other locales.
    - Other possible datasets? Wikidata information is quite sparse (see jamboard). We get Syndetics ToC data for the catalog now, are there other structured data sources for ToC? Perhaps also look at wikicite – could suggest articles even if we don't generally have article level data. ACTION - Huda to ask Jesse whether there are any open structured datasets for ToC, even if much smaller.
  - 2021-06-11
    - Huda asked Jesse about open structured datasets.
    - Huda reached out to Filip Jakobsen from Samhaeng; asked whether anything we can learn about use cases around people wanting to search across institutions to see what works exist (in ReSHARE capacity); Filip made two points: people do not benefit from looking at separate pages for Works and Instances (e.g.: conceptual distinction is not useful for users); users do not want multiple pages per institutions that has that work. If 35 instances that are same across institutions, they don't care for them to be separated. Context here is ILL – and wonder whether that would be true in local library's catalog. Filip had diagram that showed mapping b/t hubs and opi (opuses). ACTION: look at what works are and how would we map concrete examples... can you walk thru end-to-end representation of information for a few concrete examples.
- DAG Calls
  - 2021-05-28 Had high level topics overview discussion. Interesting comments and philosophical discussions about the benefits of linked data, demonstrations/examples that are useful to cite, BIBFRAME value proposition for discovery. Going forward plan to look more at user research and then work through high level areas
  - 2021-06-04 In next meeting will look at topics other than BIBFRAME and then talk about user research

## Linked-Data Authority Support (WP2)

- [Qa Sinopia Collaboration](#) – Support and evolve QA+cache instance for use with Sinopia
  - 2021-06-04
    - Continued discussions on Sinopia-QA/cache-ShareVDE. We focused mostly on the flows of data. I am moving towards identifying user stories that describe what needs to happen for each path for the flow of data. I've started capturing the user stories and questions about them in [User Stories for Sinoipia-QA/cache-ShareVDE](#).
  - 2021-06-11
    - In [User Stories for Sinoipia-QA/cache-ShareVDE](#) - 4 tabs: list of models being supported for this grant period, how to create new data, editing data, and summary of what is work remaining to meet workflows identified and agreed upon. Ideal: API with RDF. Adequate option: SVDE provides downloads and we cache those in Dave's system. Undesired: provide API that is not RDF and we're expected to connect to that. For dereferencing, similar options – ideal is they dereference and we can do via QA or Sinopia can connect directly. Either they provide dereference point or we do. What has to happen at the three levels we support is all up to what SVDE chooses
- [Best Practices for Authoritative Data working group \(focus on Change Management\)](#)
  - 2021-06-04
    - Next meeting is this coming Monday. We will continue discussing [type of changes](#). Added a couple of working documents to explore [Existing Change Management Approaches](#) (including general discussions about what should be supported), [Authority Data in MARC](#), and [Technology, Standards, and Specifications](#). I reorganized the types
  - 2021-06-11
    - Continue to talk about [types of change](#) and what data goes with that. Key piece: challenges associated with doing this approach at all; related to the fact that it is linked data. If something more complex like BF where you have to traverse the graph, cannot just say what the new label is (for instance). Three audiences: full cache audience who wants everything. application audience who has only cached small subset of data who want minimal info with limited interaction to update cache. Humans who want focused interactions with change management.
- [Cache Containerization Plan](#) - Develop a sustainable solution that others can deploy
  - Consider moving live QA instance from EBS to container version? Need to consider update mechanisms CI/CD. Agree that this is a good direction and Greg/Lynette will discuss
  - 2021-06-04
    - Greg focused on FOLIO
  - 2021-06-22
    - Lynette received from Dave what she needs to create a container. did not immediately work so needs to dig into it

## 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](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\\_lfXUJ8=](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
- 2021-04-23 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.
  - 2021-04-23 Write-up complete: [SVDE PCC/Sinopia PCC Template Analysis, March 2021](#)
  - DONE – remove for next week's notes**
- 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
  - 2021-04-23 Still nothing back from OCLC, need to reach out again
- 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
  - 2021-04-23 Waiting on <https://github.com/cul-it/blacklight-cornell> to change, not yet planned
- Authorities in FOLIO
  - Hope to include URIs as part of Cornell FOLIO migration, possible LD4P work
  - 2021-04-23 Likely going ahead in August
  - 2021-06-11: Devs in FOLIO are working on MARC authority storage and basic features for maintaining authorities. mock-ups provided and have asked for feedback and test cases (positive and negative)

## 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
  - 2031-05-07 Meeting scheduled for 5/13 with group. [Document to capture discussion/references](#).
    - Research paper format: "**Word count**: Research articles should be between 6,000 and 9,000 words (including notes but not references)". Peer reviewed. Have emailed KULA folks just to confirm type of article.
- [LD4 Conference 2021](#) - response on proposals will be announced April 30; conference is July 12-23
  - 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)
    - 5/7: Received word that affinity groups will be getting reserved time they can use as desired, so may go ahead with BOF format (will need to discuss with co-chair/DAG members)
  - Lynette/Greg/Dave - containerization, should have documented product by then
    - 4/9: Submitted abstract for presentation
    - 5/21: Accepted
  - Lynette – possibly something about the working group, perhaps updated version of code4lib
    - 4/9: Submitted as a lightning talk
    - 5/21: Accepted
  - [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
  - Call is out <https://www.casalini.it/bfwe2021/>, proposals due June 30
  - PCC Workflow Simplified view?
  - FOLIO integration?
  - MUST DISCUSS 6/25**
- SWIB virtual again this year, call for proposals out, due June 28
  - MUST DISCUSS 6/25**
- TODAY:** Lynette doing a QA presentation at Samvera partner call in June

## Next Meeting(s), anyone out?:

- 2021-06-18: Juneteenth
- 2021-06-25: