

2022-03-25 Cornell LD4P3 Meeting notes

Date: 25 Mar 2022

Attendees: Huda, Greg, Steven, Lynette, Simeon

Regrets: Jason

Discovery (WP3)

- <https://github.com/LD4P/discovery/projects/2> for issues etc.
- Draft of a discovery plan: https://docs.google.com/document/d/1zKYW7FQVNVyD0XjjW0qWznX9PC3jbmOE6Kz_yygPjs/edit?usp=sharing
- **Research: how to go from knowledge graph to an index**
 - [Research decision points](#), [Use cases](#)
- 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.
 - [References/bibliography list \(beginning\)](#)
 - BANG! preliminary design-ish/data questions [link](#)
 - 2022-03-18
 - BANG!: Finished LCCN analysis for Hubs . Used the same 10,000 sample method as before
 - For aggregations by hub (i.e. multiple LCCNs all under the same hub)
 - 497 sets with > 1 LCCN where hub > 1 work, covering a total of 1840 LCCNs. These match to our Cornell catalog: 38 sets of LCCNs comprised of a total of 284 LCCNs.
 - For relationships where two hubs are related via property
 - 277 sets where two hubs are related and each set has > 1 LCCN, comprised of a total of 674 LCCNs.
 - 224 sets related via hasTranslation, 53 related through relatedTo
 - Matches: 33 sets comprised of 123 total.
 - 26 sets related via hasTranslation, 7 sets related via relatedTo
 - We seem to be getting more useful results from LC Hubs than SVDE/PCC data
 - Steven thinking about what these number suggest for catalog services; also want to compare with what we have from the old OCLC concordance file
 - Huda asked Michelle about doing analysis on Sinopia data. API has metrics systems that provides info on what facilities are being use
 - We note possibilities for query against POD data in VuFind instance that will be set of the BD (or local instance were we to build one). We could filter data and build an ISBN index of modest size for example
 - 2022-03-25
 - Huda has looked at finding related works by ISBN in other IPLC libraries using POD data, thinking about how this could be tied to easier BD request functionality
 - Huda looking at tidying and writing up scripts from analysis work ...
 - Steven has asked Jeff Mixter what we can find from OCLC about versions of same thing across their very broad data
 - Huda hasn't looked at Sinopia API yet. Nancy has commented that Instance template is most used
 - PLAN: Wrap up analysis next week and Huda/Steven will give run-through at meeting, then look at literature and decide what to implement, do prototype in May, test in June
 - Inventory of [Work-like features in Discovery](#) - SF will add inventory of Requests features
- DAG Calls
 - 2022-03-25 Planning open office hour next week on Knowledge Panel white paper and group planning. On April 12 have plans to talk about challenges and opportunities of doing new discovery and LD work at their institutions, including how folks at institutions with fewer resources and do things, or what they need
- Document started re: Comments, Questions and Suggestions offered during the myriad of presentations provided. Huda will add link [here](#)

Linked-Data Authority Support (WP2)

- [Qa Sinopia Collaboration](#)
 - 2022-03-25
 - No meeting with Stanford this week. Much of the Stanford team had conflicts.
 - Met with Greg and Dave on Containerization. More on that below.
 - homosaurus - No change. Dave has in triple store. Steven defined context and validations. Lynette configured QA. Waiting on Dave to index and put up the search API endpoint.
 - Dave still needs to fix of total_number_found to make pagination work to get pagination working again in Sinopia. Once that is done, it will just feed through to Sinopia without any additional work.
 - Meeting with OCLC this afternoon to discuss collaboration and our usage.
 - National Library of Norway has requested some additional authorities, one has dataset, other has API
- [Best Practices for Authoritative Data working group \(focus on Change Management\)](#)
 - 2022-03-25
 - Met with Kevin at LOC to talk through the LOC activity stream and how Dave can use it to do an incremental update. There is a path forward for this. ([notes](#))
 - Scheduled to meet next Tues afternoon with Getty to look at how it can be used for incremental update.
 - Scheduled to meet next Thurs afternoon with Sinopia team to strategize how to update cached labels in the Sinopia app.
 - Working on updating the Consumer Processing section in the [EMM Change Document API](#) based on the results of these meetings.
 - There is still a question about date handling and what we want to recommend. Options are endTime, startTime, published, updated
- [Containerization](#)
 - 2022-03-25

- Worked on QaServerContainer support for customizing Qa and QaServer ([PR #56](#)). Used environment variables to set values in their respective initializers. Will need to copy translations from S3 which hold some app specific text overrides for the footer. Would be nice to have these as environment variables too, but translations can't use environment variables. Greg and I need to work through several things.
 - how to add the new env variables to the deploy
 - how to update the deploy when the image changes (the templates were used for the initial deploy)
 - how will -int, -stg, and -prod be supported? currently only -prod is supported.
 - Greg and Lynette will work together to see what it takes to move the container version into production
- Dave, Greg, and Lynette met [for status update and next steps for containerizing the cache](#). Talked through current workflow for QaServer. Greg will take first stab at setting up the cache search API endpoints in a Tomcat container for 2 authorities.
 - Greg has hosted indexes in S3 for Dave so he can point cache server at them. Dave hasn't had time to work on it. Now Greg trying to do this and has war files from Dave

Other Topics.

- Sinolio - Sinopia-FOLIO
 - 2021-12-17 - Work Cycle finished, sprint video out
- OCLC Linked Data / Entities Advisory Group
 - 2021-12-10 OCLC presented at bigheads meeting this week, in testing
- PCC
 - 2021-01-21 Definitions and non-RDA final report to POCO (hopefully) to be submitted next week
 - 2022-01-14 Nothing new to report.
- Authorities in FOLIO
 - 2022-03-25 Some transitions in team. Useful meeting with Jenn, Frances, Nick, and Darcy to decide what needs to be provided to build queue. Mockups look good and allow filtering on types of change (new, deleted, updated). Quite different indexing requirements for data maintenance vs discovery

Upcoming meetings/presentations

- KULA submission for special Issue "The Metadata Issue: Metadata as Knowledge" accepted, final proof text that Huda will accept straightforward changes and ask group for final feedback by April 1.

Next Meeting(s), anyone out?:

2022-04-01 - April Fool's -