

2021-02-19 Cornell LD4P3 Meeting notes


Date: 19 Feb 2021

Attendees: Greg, Lynette, Tim, Simeon, Huda

Regrets: Steven

Last meeting: [2021-02-12 Cornell LD4P3 Meeting notes](#)

Discovery (WP3)

- <https://github.com/LD4P/discovery/projects/2> for issues etc.
- Draft of a discovery plan: https://docs.google.com/document/d/1zKYW7FQVNVyD0XjW0qWznX9PC3jbmOE6Kz_yygPjs/edit?usp=sharing
- **Strand 1: production piece**
 - Production requirements and functionality – [Production decision points](#)
 - **Discogs** data use - in production since January 2021
 - **ACTION (after March D&A sprint)** - Tim to follow up on implementation and look at data from tracking use of the Discogs
 - **ACTION (during March D&A sprint):** [Tim Worrall](#) will raise usability testing for D&A queue (don't carry forward in notes as now outside of LD4P)
 - **ACTION** [Huda Khan](#) [Tim Worrall](#) [Steven Folsom](#) to develop a post for #general, #discovery and partners email about Discogs going live
 - 2021-02-12 Started text (as in copied Steven's content). In progress. Will send out draft text today. Also discussed in DOG meeting on Monday with useful feedback
 - DONE. Emails sent and documentation page set up at [Production: Discogs](#)
- **Strand 2: 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
 - 2021-01-29: Huda working to get scripts in place to populate index; bringing in period-O info; focused on locations with Wikidata URIs for consistency. Subject headings: script that takes-in components & breaks those out... and parses into timeline info. On Dave's fuseki, 34 distinct temporal terms with labels. Will finish today with actual index. Will break-down the loading to increase load speed
 - **ACTION ITEM:** adding Wikidata URIs for any subject headings and broader/narrower URIs to index (today) - DONE for dev VM index, to do next for LD4P3 Solr index (will be done **still** today). Can be marked done after.
 - 2021-02-12 Each LCSH now has a Wikidata URI where available, broader and narrower URIs where available, periodO extracted spatial URIs and time periods where available, a list of components for multi-part subject headings with URIs and labels for each component, separate lists for temporal components with start and end times mapped, and geographic components with LCSH URIs and equivalent Wikidata URIs extracted and added.
 - One error (for review perhaps later): two of the LCSH temporal component labels have special characters and are not being matched
 - 2021-02-19: Kicked off script, will mark when done but not done yet.
 - Reached out to IRB to ask about testing: if we want to disseminate results as research data, need to do IRB protocol; has a follow-up. Waiting to hear back but will submit protocol if no word. Simeon's interpretation of reply is that we are crossing line into research and the approval will likely be positive. Depending on how we describe what we're doing it either falls under research OR improving a product... but we're essentially doing research to improve a product so yes to IRB review.
 - **ACTION ITEM:** IRB did respond and say they wish us to proceed with sending in an application. Huda will work on this and reach out with any questions if needed.
 - 2021-02-19 DONE. IRB application submitted.
 - 2021-02-19 Tim has been working on entity page. Notes a number of issues with the Historopedia timeline such as items with same date being hidden, but performance is good.
 - 2021-02-05
 - What would D&A user reps favor?
 - Concern that full KP linked from  button is too much
 - Is "KP-lite" on autosuggest a good route? We think users would find this valuable. Are there options that minimize index changes?
 - What warrants a KP?
 - What is the redundancy between KP work and DASH!? Does dashboard mean a fundamental change or is just an enhanced KP?
 - We need to be aware of which options require significant indexing changes. There is already a sense that we want to add ids to the index
 - What about the open syllabus project? This relied on the open syllabus API, not sure whether it is available in LD. Essentially a mapping from domainCSIP codes ISBN, very few wikidata connections
 - What would be the smoothest next step for production?
 - Which option would give us real linked data connections via URI?
 - Steven notes that LTS authorities in FOLIO group is looking at the insertion of URIs into MARC records (resources willing)
 - **ACTION** - [Huda Khan](#) [Tim Worrall](#) document options and implications as preparation for user reps presentation in order to get a steer on where to continue experimentation with a view to future implementation
 - Dashboard (perhaps for some entity types only)
 - Autosuggest with KP-lite

- Regular facet with KP
- Open syllabus related items
- Brainstorming [notes](#)
- 2021-02-12 Agreement that streamlined KP is a good starting point, with possibility of later extension to a full dashboard. Autosuggest and open syllabus good alternative options.
- ACTION - [Huda Khan](#) to line up meeting with D&A user reps
 - 2021-02-12 Ready to set up meeting
 - 2021-02-19 Sense that we shouldn't burden the D&A user reps at this point
- Tim working on the author panel and timeline, will continue working on this next week. Also spent time on graphical representation of influenced-by and influence-for. Steven thinks that the lists of results may be ordered by some sense of strength, so perhaps top results are useful
- Tim on ESMIS next two weeks, Huda will work on tasks for user tests and perhaps pilot test with the team (pending IRB approval). Open questions about how to recruit students right now - perhaps talk to usability WG

Linked-Data Authority Support (WP2)

- [Qa Sinopia Collaboration](#) – Support and evolve QA+cache instance for use with QA
 - 2021-02-19:
 - No meeting. Dave is working on the ShareVDE PCC data. "Share-VDE release is in six parts, each with numerous nq (quad) files, each with numerous records. Translating Part1 to a single nt (triple) file results in a 2.7GB file, ready for loading into the triplestore. Parts 2-6 are currently in conversion." Checking with Dave to see what was in the quad position that is being lost in the conversion to n-triples.
 - We still have 5 issues that are under exploration. Dave says he is close on fixing the GETTY_TGN and ULAN issue where subject URIs should be the RWO version with ` -place` and ` -agent` appended, respectively.
 - Still waiting for a reply from ShareVDE to Vivian's request to restart meetings.
- [Search API Best Practices for Authoritative Data working group](#)
 - 2021-02-19:
 - Group is officially ended and documents tidied up
 - Announcement includes survey for feedback and survey for topics for next group. Will give 2 weeks to respond
 - ACTION [Steven Folsom](#) and [E. Lynette Rayle](#) will send out announcement on Monday. Plan is to send to PCC list, LD4P3 list, LD4 #general slack, Samvera Community list, Samvera #general slack, ShareVDE #aims_sg slack, several authorities (e.g. Getty, MeSH Bio-portal, etc.)
- [Cache Containerization Plan](#) - Develop a sustainable solution that others can deploy
 - 2021-02-19 Greg completed CloudFormation template that allows someone to spin up a QA service in AWS easily. About 500 lines of template code that brings this very close to being a turnkey solution (in services-ci branch). Greg notes pre-reqs for spinning this up: S3 bucket for configs etc. which could be added to another template.
 - When complete Lynette will test, then ask Dave to test, then ask Stanford folks. Greg will also create a demo screencast.
 - What about replacing the current QA setup with this new approach? Would need to check authority configuration and correct setup for load. Lynette notes need to copy over the DB to retain history
 - Next steps
 - start to look at containerize Dave's setup. Two steps: 1) code to serve from cache, 2) indexing process
 - think about instructions for a vanilla linux server setup

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
- What does success look like? And then how do we get there?
- Miro board (diagramming): https://miro.com/app/board/o9J_lfXUUj8=
- Notes space: <https://docs.google.com/document/d/1TVPBfak7DkfjBptKI-pCMWQnOaiWHB0XCHswiB3Fr9g/edit?usp=sharing>
- 2021-02-05 discussion
 - 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
 - Functions of "Aggregated index, allowing pivoting & ETL"
 - Includes current functionality of Frances' indexing
 - Does it include any editing?
 - Is there interaction with CULAR?
 - Includes indexing associated with DCP
 - What interfaces or functionality do we expect for the connecting lines?
 - Do we need a diagram for now (or at least July 1, 2021 with Voyager gone)?

Other Topics

- OCLC Linked Data / Entities Advisory Group
 - Request for UI and API testing from Jan 25
 - Lynette has Cornell key (a WSKEY) for testing
 - Call discussed seeding of data. Data for person includes VIAF and other sources; place includes geonames. Steven, Huda, Jason and Lynette signed up for user testing
 - 2021-02-19 Huda finished UI testing (Seymour Schwartz for the win). Involved assessment of amount of information presented. Lynette hasn't got response to query about access key, interesting in testing new search in API as well as CRUD facilities
- PCC - Sinopia collaboration
 - 2021-02-05 Charge to form a new group for documentation, mentoring etc is under reviews
- PCC Task Group on Non-RDA Entities
 - 2021-01-15 PCC reviewed proposal but no decisions made yet, looking at description wrt cataloger use, discussion will continue

- Default branch name - Working through repositories in [Renaming of LD4P Repositories](#)
 - 2021-02-19
 - Created [Renaming of LD4P Repositories](#) page to identify Cornell repos, provide instructions, and track progress.
 - ACTION - Huda to look at blacklight and discovery repos
 - ACTION - Steven/Jason to look at HipHip
 - Lynette notes that Stanford have already dealt with their repositories

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](#)
 - 2021-02-05 Huda submitted abstract
- [code4lib](#) - virtual next year
 - Expecting to attend: Huda, Steven, Lynette
- Lynette doing a QA presentation at Samvera partner call in June

Next Meeting(s), anyone out?:

- 2021-02-26 Tim out