2021-02-12 Cornell LD4P3 Meeting notes

Date: 12 Feb 2021

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

Regrets: Jason

Last meeting: 2021-02-05 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/1zKYW7FQVVNvyd0XjjW0qWznX9PC3jbmOE6Kz_yygPjs/edit?usp=sharing
- **Strand 1: production piece**
  - Production requirements and functionality – Production decision points
  - **Discogs** data use - in production since January 2021
    - **ACTION**: Tim will speak to Melissa about tracking for the Discogs link.
      - 2021-02-12 Had discussion with Melissa and now Tim will create a issue for the March D&A sprint. Followup will be to track that this gets done
    - **ACTION**: Tim Worrall will raise usability testing for D&A queue (don't carry forward in notes as now outside of LD4P)
      - 2021-02-05 This won't happen until the next sprint, starting second week of March
    - **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
    - Steven notes LD discussion where he was asked about how external data is acknowledged and shared with the group how we do it on D&A. Miel Vander Sande:

      > *Anyone who has examples of websites that somehow cite / display the source of external #metadata? for example: "this description is from #wikipedia / #wikidata " #semweb #digitalarchiving #archiving #linkeddata #provenance*
      >
      > *— Miel Vander Sande (@Miel_vds) February 11, 2021*

- **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
      - 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.
          - Done. IRB application submitted.
  - 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
- Tim working on the 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

Linked-Data Authority Support (WP2)

- Qa Sinopia Collaboration – Support and evolve QA+cache instance for use with QA
  - 2021-02-12:
    - Dave, Steven, and I met to review the status of issues related to the indexing approach. 4 were closed (returning too many results, newline issue, all tests passing for geonames and ligatus). 2 more will likely be closed today (accuracy tests for locgenres and locnames). 4 others are under active exploration by Dave (getty looking for subject to end with -place and -agent, some subauths not returning any results, mesh hierarchy and description issues). 5 other issues awaiting exploration. Some may be fixed by the current set of issues.
    - Vivian at Stanford sent email to Anna at ShareVDE to restart discussions about moving data round trip from Sinopia to ShareVDE, form Stanford's MARC converted records to ShareVDE, from ShareVDE to Sinopia, and direct access to ShareVDE through QA.
    - Michelle announced the release of ShareVDE's PCC data. They are provided as downloaded data, not direct access to ShareVDE. Dave is reviewing for import to the cache.
      - When we do have a round-trip, where is the source of truth? How will we track this? How does QA deal with this and what portions of the graph should it handle? Lynette has been looking at GraphQL (which maybe isn't directly usable) for information about how to define a graph shape/selection in a flexible way
- Cache Containerization Plan - Develop a sustainable solution that others can deploy
  - 2021-02-05 Greg says "I've finished expanding the CloudFormation template to include a container service, which will make the container much more useful to adopters. There are still several steps which are not automatic, involving networking and security groups, and making those automatic will require the template to create the cluster of machines that the whole thing runs on. That will be super nice, but it will be a pretty big expansion of the template."
  - 2021-02-12 Not much progress, next step is to add cluster to template
- Search API Best Practices for Authoritative Data working group
  - ACTIONSteven Folsom will send document to PCC
    - Pending announcement mentioned below
  - ACTION E. Lynette Rayle will send announcement to Samvera and put a note on #general, also link from main LD4P page.
    - 2021-02-12:
      - The summary document is almost complete. There is one question about the wording around broader and narrower terms and related user stories. I am adding in an introduction that describes the working group and defines the 3 perspectives for catalogers, developers, and providers.
      - Began writing the announcement that Steven and I can use for the various communities – any feedback due by Monday 2/15

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/1TVPBFak7DkfjBptKl-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
  - 2021-02-12 Attempts to test have been frustrated by lack of documentation of testing desires. Some users have been providing feedback about expected metadata details
- 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 - WAIT until we can use github tools January 2021
  - Lynette is signed up to be a beta tester for github. Target for public release is end January
  - Github have provided the change button, Lynette has provided feedback as a beta tester. There is an issue that after the change to `main` someone can repush the `master` branch, has suggest a deny list of branch names. Links in comment will get forwarded to `main` unless there is a new branch with the old name
  - For forked repos there will be a message
  - What will it take to make the change to `main` for LD4P repos?
    - Lynette has documented HERE
    - Lynette will rename for qa_server, work with Greg on qa_server container; document on LD4P wiki
    - For our blacklight fork we should wait until the D&A team change
  - 2021-02-12 E. Lynette Rayle still to reach out to fork owners
    - Think that forks aren't too problematic for LD4P repos because we don't have folks contributing back
    - Have done rename for CUL-IT qa_server and will for LD4P qa_server
    - ACTION: (DONE) Lynette will review set of LD4P repos with Cornell primary ownership so that we can change these to use `main` branch, later propose that other LD4P repo owners do it
      - Created Renaming of LD4P Repositories page to identify Cornell repos, provide instructions, and track progress.

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?:

- 2021-02-19 Steven out, Lynette might be out