2022-02-18 Cornell LD4P3 Meeting notes

Date: 18 Feb 2022

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

Regrets: none

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
- **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.
  - 2022-01-28
    - Huda extending data source writeup and will spend a couple of days building a tool to characterize/explore data in DAVE
    - What past research is there on works and what is useful to users?
      - K Coyle and K Godby work from a few years ago suggested ways to find particular expressions (bf: works) useful, works (bf: hubs / opera) help users get there
      - Steven to check past D&A user studies - No institutional memory or documentation of tests for "Other forms of this work"
  - 2022-02-11
    - Huda has works on HTML/Javascript graph explorer code and showed demo. Steven notes how this will be useful to explore datasets. Where to take this work? Perhaps discuss with Dave based on the set of queries used, and to see whether this can run over his Lucene/ES index for more responsive.
    - Huda working to get counts of items under the same LC Hub. Steven and Huda will discuss the data further
    - For past research on works and usefulness to users, Huda will also revisit FRBR/BIBFRAME + user research references previously suggested for KULA paper
  - 2022-02-18
    - References/bibliography list (beginning)
      - Plan: Continue reviewing and looking at references
    - Huda added a "feeling lucky" page to fuseki UI which picks random statements to show so people can start exploring. Needs to send info to/show to Michelle (Stanford), and ask Jim Hahn (Penn) if interested.
      - Plan: Add click functionality on the graph to generate a side panel with details for that node/entity. May be easier to understand than spaghetti expansion
      - Steven mentioned classes/predicates summary to Dave
    - Hubs analysis: Huda changed approach after discussing with Steven what we're trying to explore with this analysis.
      - Main question for ShareVDE and Hubs aggregation: Can this data yield relationships between/groupings of works that yield related items in the catalog? We want to extract sets of ISBNs grouped under an aggregation or under a property between works, and then see if any of those sets yield at least two catalog matches (i.e. translate into relationships between items in the catalog)
      - Approach for ShareVDE: Evaluate how many opera there are with at least two works and at least two related ISBNs, and how many of these ISBN groups yield at least two catalog matches. (Only one catalog match means that, if we were on that item page, we would not have any related items to see using this data).
        - Steven notes that a link via Hub from an ISBN we hold to one we don't hold is a possible ILL use case
      - Now for Hubs: Get sample of hubs from LOC search. Changing start parameter to page through list to work around LOC side throttling. For each hub, see if ISBN set can be generated. For each set with > 1 ISBN, determine if there > 1 catalog matches.
        - Plan: Continue to do so. Current results: For 4000 hubs from LOC, 87 sets of ISBNs (with > 1 ISBN) where hub has > 1 work = 367 unique ISBNs => catalog matches where you have at least two catalog items for that ISBN set: 12 ISBN sets yield matches for total of 73 ISBNs
- DAG Calls
  - 2022-02-18
    - Had first crossover call with WAG group focusing on DASH! and usability testing. Some questions to follow up on, report of our page selecting one citizenship statement where wikipedia has multiple statements that might be a bug
    - Next week will talk more about this work and also about getting and using feedback from use reps etc.

Linked-Data Authority Support (WP2)

- Qa Sinopia Collaboration
  - 2022-02-18
    - No meeting with Stanford this week. Steven, Dave and Lynette met to discuss current priorities.
    - Current priorities include
      - fix of total_number_found to make pagination work to get pagination working again in Sinopia (Dave)
      - add homosaurus authority - Dave uploaded in cache, but needs to run indexing, Steven defined context defined, Lynette defined QA config and waiting on index to be able to test prior to deploy
      - create documentation of search queries for cache (Dave)
      - try using LOC or Getty activity streams to update the cache (Dave). This is proof of concept. It may prove insufficient as none of the feeds include patches. But makes for a good exploration.
      - Update pagination in QA linked data module to use json-api as output (Lynette)
- Best Practices for Authoritative Data working group (focus on Change Management)
  - 2022-02-18
    - Meeting on Monday to review EMM Change Document API and Notifications Example

- Updates include incorporating suggestions from last meeting, adding reference links at the start of each section, specifying property examples common across activity types in one place, adding more content across all sections. Sent both recommendations document and notifications examples to the group for review.
      - There are two other use cases which need examples, but wanted to get feedback on the notifications before working on those.
  - Containerization
    - 2022-02-18
      - QaServer - Evaluating status of docker deploy to determine how close it is to being ready to become production.
        - working as expected: has all authorities installed; spot tests showed all queries tested were successful; check status page is functioning; API Documentation displays; Fetching a single URI works
        - not working as expected:
          - monitor status page gets an error that looks like it might be database related
          - not sure how to connect to debug
          - once working, it needs a nightly job
          - most customizations are for the monitor status page, so unable to evaluate customization requirements for production
        - Greg and Lynette will work on these items, perhaps consider framework for maintenance jobs and maybe consider for nightly rebuild too (needs two requests at specific times)
      - Cache Search API - Planning meeting next week with Dave and Greg

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
    - 2022-02-18 Mary met with team and making progress with deletes, Frances is getting experience building out the index with hope to have and API Nick can work against by end February. Steven has diagram of vision

Upcoming meetings/presentations

- https://kula.uvic.ca/index.php/kula/announcement/view/1 . Submission for special Issue "The Metadata Issue: Metadata as Knowledge" accepted
- 2/24 giving presentation to PCC/Sinopia interest group
- 3/9 ALA Linked Data interface group
- 3/11 ALA Catalog form and function group

Next Meeting(s), anyone out?:

2022-02-25 ... moved to 2022-02-24 9am - Jason and Steven out

2022-03-05 - Simeon out, Jason will run