2021-03-26 Cornell LD4P3 Meeting notes

Date: 26 Mar 2021

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

Regrets: Lynette, Jason

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 (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)
      - 2021-03-26 Tim discussed usability testing, note syndetics as well as D&A work. Not expecting to be able to do this until the summer or early Fall
- **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
    - 2021-02-19 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
      - 2021-03-05 Tim resolved a number of issues. Next week will return to work on this and deal with influence-for and influenced-by presentation
      - 2021-03-19 Tim still working on this
  - Considering KPAOW zero (streamlined knowledge panels). Have begun discussing what should go in a streamlined version.
    - Huda Khan Scheduled meeting with D&A reps April 9th, 9-10 (Moved the LD4P3 meeting to this time)
  - Usability testing for DASH
    - Getting ready for usability tests in April
    - Need to finalize usability tasks for both authors and subjects
  - User reps D&A meeting: Need to re-follow up
    - Good to have knowledge panel lite mockups or examples ready to show
      - Started some mockup brainstorming here: https://jamboard.google.com/d/1PXLAu0XTzaQhe9vw6OcAdsqJItFJ4U-FStgm9lA1CHc/viewer?f=0
    - Also need to show entity page examples
      - SVDE Works
        - 2021-03-19 Steven has looked at SVDE works and is thinking about how to find works via ISBN to then find other Instances. Is looking via SPARQL on Dave's ingested data
        - 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
      - eResources
        - Discussion with small group regarding call number classifications on e-resources, where there is/isn't sufficient metadata compared to equivalent/related physical objects. Possibility of gathering some useful examples around needs/wishes around e-resource discovery.
        - 2021-03-26 - Other CUL staff met a couple of weeks ago. Examples where print and e-resource versions have differing metadata quality. Can we connect the resources better and/or copy metadata from one to the other? Has connection to our work to identify different versions. Nothing for LD4P3 to do until they have come up with appropriate examples

Linked-Data Authority Support (WP2)

- Qa Sinopia Collaboration – Support and evolve QA+cache instance for use with QA
  - 2021-03-19:
    - No meetings this week for QA/Sinopia this week.
    - See Steven's comments on ShareVDE data. Dave wants to look at making a direct connection to ShareVDE GraphQL API and translate it on-the-fly to something that QA can work with. There are going to be some complexities with how to structure queries and extended context based on the variability of data shapes.
- Search API Best Practices for Authoritative Data working group
  - 2021-03-19:
    - Created the charter for the next working group describing the expected outputs for change management. I plan to announce the charter on Monday and begin reaching out to individuals to get folks on board.
  - 2021-03-26 - Lynette will speak about results of first group at Discovery Affinity group next week
- Cache Containerization Plan - Develop a sustainable solution that others can deploy
  - 2021-03-19
    - Lynette began documenting the deployment to AWS adding in an overview, background knowledge, and architecture sections. Next is to start looking at how to use the templates.

- Greg is going to follow up with Dave about how to work with him on his containerization efforts
    - Expect to work on demo screencast after refining documentation
  - 2021-03-26
    - Greg and Lynette have been working together on documentation and this has been very useful in uncovering missing things. Greg found that he needs to make sure AWS actions can be performed by a less powerful user
    - Greg has discussion containerization of cache infrastructure. Greg has some tasks to help with

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_lfXUUj8=/
- 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)?
- 2021-03-19 Jason/Steven/Simeon created separate diagrams and Jason is working on a combined picture

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 ldpaths 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.
  - ACTION - Steven Folsom to write up state of current analysis and store a snapshot of the spreadsheet on the LD4P3 wiki
- OCLC Linked Data / Entities Advisory Group
  - 2021-03-26 No updates, some emails
- 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
- 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.
- Authorities in FOLIO
  - Hope to include URIs as part of Cornell FOLIO migration, possible LD4P work
  - 2021-03-19 In LTS there is a task group that has a proposal for authority management in FOLIO (absent new features). Being reviewed with request for scripting work to create reports etc.. Includes insertion of URIs into MARC

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
- LD4 Conference 2021 - proposals due April 30
  - Discovery - suggestion of discussion form
  - Steven thinking something around shapes and compatibility and round trips
  - Lynette/Greg/Dave - containerization, should have documented product by then
- code4lib - Expecting to attend: Huda, Steven, Lynette, Greg
  - Great opening keynote about capturing of indigenous information connected to mapping, discussion of what information is shared or not. https://terrastories.io/
  - Steven notes everyone dealing with questions of sovereignty and a agency in building collections
  - Steven's Discogs poster at code4lib went well
  - Steven notes session on technologies to fix problematic terminology in upstream data sources and/or as an overlay/replacement strategy. There are at least 3 or 4 Blacklight implementation to replace "Illegal Alien" in discovery systems
- Lynette doing a QA presentation at Samvera partner call in June

Next Meeting(s), anyone out?:

- 2021-04-02 ...