2020-09-04 Cornell LD4P3 Meeting notes

Date: 04 Sep 2020

Attendees: Huda, Lynette, Jason, Greg, Simeon

Regrets: Tim, Steven

Actions from 2020-08-21 Cornell LD4P3 Meeting notes

- Simeon Warner to coordinate work on final report with Michelle
  - Due Aug 31, not yet sure about coordination around partner report
  - 2020-09-04 - Still to do
- Huda Khan to reach out to Tracey (cc. Jason) to review/refine discogs plan before taking to Q&A (after September 2nd)
  - 2020-09-04 - Seems OK to go ahead with this now

Agenda

- Discovery (WP3)
  - Preliminary draft of a discovery plan- intended to get feedback: https://docs.google.com/document/d/1zKYW7FQVVNvyd0XjjW0qWznX9PC3jbmOE6Kz_yygPjs/edit?usp=sharing
    - Which to start with?
      - Discogs - everyone likes this, are we OK showing the data. D&A Reps + Tracey would be a good group to think about this.
      - Autosuggest - need to clarify added value of linked data over what could be done with current indexes, and address updates
      - Open Syllabus - can we rely on that project? are there other (linked) data sources
      - Knowledge Panel -
    - How much does bang-for-buck influence our decision? E.g. discogs would be great for music but that is a relatively small number of items
  - Parallel strands:
    - strand 1: production piece (i.e.: picking features we've already worked on to then push those into production): discogs, autosuggest. Feature branches in the code to work on features in isolation. Still need to discuss discogs with Tracey (post- 9/2). Preliminary pass with discogs features using latest production (from a few days ago) successful (tried on dev vm and will try on ld4p3 demo using continuous integration later)
      - Checkpoints:
        - Discussion with Tracey re. use case and benefit
        - LTS engagement re. metadata
        - Production requirements and functionality
        - Discussion with D&A User Reps and dev team
    - strand 2: research: how to go from knowledge graph to an index - what decisions are needed. What are the data sources for each (e.g.: how many Cornell faculty in wikidata)? Present: reviewing data sources and questions. Should have more worked out in a week or two. Main areas of concern: browsing and dashboard... and anything we can do to help patrons navigate our collections and how we can highlight an entity... and what does that mean for the index? How do we capture the relevant bits of the graph for an index? Is there a repo for this? Not yet... but can use discovery repo (https://github.com/ld4p/discovery) we already have to capture any queries or related work.
      - Design requirements next
    - Can we get Blacklight fork to not hit the production catalog every time we do a pull request? Not clear way to specify that all pull requests should go to the fork rather than the main branch from which you forked. Huda is investigating with input from Code4Lib inquiry
      - SF - to add data analysis queries in https://github.com/LD4P/discovery.
  - LD4P3 demo blacklight site
    - Greg working on full CI/deploy but with a manually initiation for now : 2020-08-28: DONE
    - Huda looking into avoid PRs on the LD4P3 fork going against the CUL-IT repo
    - 2020-09-04 - Greg and Huda working on getting deployment from a branch, still issues to resolve
- Linked-Data Authority Support (WP2) - A key element of this work package is a sustainable solution that others can deploy. Questions of budget for deployment. Need to get all code into LD4P repository. What would a good end-product look like both for our maintenance and for others to use
  - Qa Sinopia Collaboration
    - 2020-09-04 - primarily discussed ShareVDE. Touched base with Tiziana with questions from the meeting.
      - API
        - will support search for works and instances (can search CKB or institutional data)
        - will support search of items (items will not be clustered in the CKB)
        - is in the planning stages and they hope to complete the plan in the next 2 weeks
        - goal is to release the API in June 2021
      - CKB
        - will include Instance entities (reconciled master instances)
      - Hubs
        - do not support hubs in the way defined by LOC
        - Opus is their equivalent abstraction (not sure how closely they are aligned)
      - Meeting with ShareVDE technical team scheduled for Tues, Sept 14
  - Cache Containerization Plan agreed with a prioritization of containers
    - 2020-09-04 - Tried switching to Postgres instead of MySQL. Still having issues. Tried to get Hyrax docker running as a sanity check. It also had problems. Tom said that Docker on Mac has been problematic for folks. He works on Ubuntu. Tried to install VirtualBox VM to be able to run Ubuntu, but the install failed.

- [Search API Best Practices for Authoritative Data working group](#) is still working through use cases, this seems very important to take sufficient time
    - 2020-09-04
        - Working on developer user stories. Not as clear cut as expected. Some confusion around UI user stories and API user stories. Working on reorganizing to those two sets.
        - Need to get contact information from Steven to share with PCC. Looking at creating a survey that will allow for prioritization by PCC and the working group.
- LD4P2 wrap up - 2020-09-04 - DONE
- 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?
    - 2020-08-21 Jason drafted an introduction for LTS staff and no capacity right now in LTS, postpone until folks are back on campus. Question of how we much we need to seed discussion with a first pass. Discuss approach again 2020-09-04 or 11
- Other Topics
    - OCLC Linked Data / Entities Advisory Group
        - Recap of usability test results, API testing (get, search), recap of shapes (gap between wikibase use and ontologies we understanding in the community)
    - PCC Sinopia Profiles Working group
        - Going through large spreadsheet comparing Sinopia profiles with BSR and CSR, close to having a comparison, will be part of report to POCO
            - 2020-08-28: done comparing BSR with existing profiles. Now trying to clean-up the sames/diffs so that the spreadsheet is shareable. Prelim report due soon that might confirm direction or send on new course
    - PCC Task Group on Non-RDA Entities
        - Meeting today to decide on which models to take, e.g. RDA/RDF values vs. abbreviated list of codes
        - Question of going to LC about using/hosting a small vocab of entity types or use an external vocab
        - Write-up will go to POCO
        - 2020-08-28: lobbying for shallow and focused set of entity types that would allow to say (at least in MARC) that we're working with something that RDA does not currently permit and not try to enact a deep hierarchy. Examples: RDA does not give direction on events description (we do this now... Olympics produces brochures... but are also Events with a host, agents, etc.). MARC allows events in both agent and subject... RDA does not have distinction/recognition.
            - More granular descriptions might happen in places like Wikidata; the Task Group will be attending a PCC Wikidata

- Default branch name
    - Instructions for creating a [New Repo with Main Branch](#)
    - Instructions for renaming an existing branch TBD. Will be bringing back info from the Samvera community working group on branch renaming.
    - 2020-08-21: Lynette has done testing with tool that does the transformation, process is one-time destructive process but does keep settings and update PRs. Might still be good to wait for github that has said it plans to have tools by end of year ([Impact Analysis](#), [Analysis results of promising tool](#)) We are evaluating multiple tools and approaches for renaming. The analysis will give you a sense of what is being tested.
    - 2020-08-28: Nothing new. Kate Lynch will present our findings at Samvera Connect
    - 2020-09-04: Oct 1 - all new branches will be main unless you set the default in organization
- Upcoming meetings
    - Virtual Blacklight Summit in early October. (Main site [link](#) and [document](#) to add lightning talk/breakout session ideas) - Huda is on unofficial committee. Expecting various follow-ons from earlier meeting this year. (Discussions still in progress and, at some point, draft program will be shared with the community).
        - Topics list sent to Blacklight list and other channels... people vote.
    - European BIBFRAME Summit; registration now open. Occurring 9/22-9/23 (9-11:30am Eastern each day): http://www.casalini.it/bfwe2020/ (no cost)
    - DCMI 2020 Virtual: https://www.dublincore.org/conferences/2020/programme/

Next Meeting(s), anyone out?:

- ...