2019-03-08 Cornell LD4P2 Meeting notes

Date: 08 Mar 2019

Attendees: Jason, Steven, Lynette, Huda, Simeon

Regrets: Tim

Agenda & Notes

Review actions from 2019-03-01 Cornell LD4P2 Meeting notes

- Steven Folsom keep track of cataloger perspective on work to support LC Names is in QA with context
  - 2019-02-22 - Tools call this week was Steven, Lynette and Dave. Went through the spreadsheet and David Eichmannis going to go through the top 3 issues (https://github.com/LD4P/qa_server/issues/53 https://github.com/LD4P/qa_server/issues/50 https://github.com/LD4P/qa_server/issues/47 – no updates on these since Tuesday)
  - 2019-03-01 – Was discussed again on Tuesday, Dave has been busy with SHARE-VDE so no progress on these issues. Will continue to emphasize these as the top three priorities
  - 2019-03-08 - David Eichmannis focusing on SHARE-VDE which is needed for Sinopia and is thus the project priority, progress this week. E. Lynette Rayle will emphasize our priorities again for QA (https://github.com/LD4P/qa_server/issues/53 https://github.com/LD4P/qa_server/issues/50 https://github.com/LD4P/qa_server/issues/47) when Dave has time, discuss on Tuesday
- Steven Folsom to reach out to Nancy to see what data we get from this about needs, will then reach out to cohort to check results and see if anything is missing
  - Update: Received a confirmation from Nancy that they are happy to share. The survey closes 2019-03-08.
- E. Lynette Rayleto add comment on plan for Wikidata https://github.com/LD4P/qa_server/issues/15
  - DONE, Christine Fernsebner Eslao has been doing investigation of wikidata queries, E. Lynette Raylewill discuss status
- E. Lynette Rayle LDPath gem is doing load for graph every time it tries to access something (SLOW)... may need to fork LDPath gem to make more efficient since there will not need to reload graph every time (for our use case)
  - Still issue with LDPath gem, working Chris Beer to resolve
  - 2019-03-08 – no reply from Chris, will follow-up with issues in LDPath repo
- E. Lynette Rayle investigate & document algorithm for selection of languages with label and/or no label. In cases like AgroVoc where languages are handled well then we just want to document the (good) algorithm; but in other cases where language information is wrong we need to be able to turn that off (need more config for this)
  - 2019-02-22 – no update
  - 2019-03-01 – no update, will be part of QA documentation update that is ongoing
  - 2019-03-08 – very nearly done... will be in QA release, hopefully today. Config will have wildcard for "all-languages" which will then not select data for just a single language
- Huda Khan & Steven Folsom to go through spreadsheet https://docs.google.com/spreadsheets/d/1gklgTQUNd9nr3xTl5kdFYK51SRG6TeMRH678N1FtNQ0/edit?usp=sharing assessing difficulty and "linked data wow" factor (BfB)
  - 2019-02-22 Discussion of whether we are doing anything that will impact our source (MARC) data or whether this work is just additions at the discovery layer (that may come from other sources such as SHARE-VDE clusters) – agreement that we are interested in experiments that will merge any useful data at the indexing stage. If we wanted to move any of this to production then we'd need to have conversations about linking in MARC but that could be done later / if / when. Huda Khanand Steven Folsomwill meet next week to come up with proposals with an assessment of tractability and wow factor.
  - 2019-03-01 Met yesterday and assessed "wow" in spreadsheet, suggested that three: relationships, knowledge cards and semantic navigation have the best potential for "wow". Huda Khanand Steven Folsom to come up with wireframe proposals to present to D&A group on March 19.
- Huda Khan is working on whether usability work on catalogs has been done internally as well as externally to CUL. Code4Lib journal and another has some details
  - 2019-01-18: Have been looking at usability work and literature. Found some older material on Blacklight tests (when facets new), not so much after that. Some log analysis on SearchWorks. Also looking at linked data browser tools. Interesting dissertation from a few years ago on LD exploratory search. See discovery folder, literature exploration NOTES DOCUMENT, initial discussions with Tim regarding knowledge panels etc. , and also Stanford NOTES FROM LAST SUMMER
  - 2019-02-15: Astrid is doing more user sessions in the next week or so. Huda will then review. Also discovery discussion group had a meeting. For reference, knowledge panel slides with screenshots/examples plus questions
  - 2019-02-22 Huda Khan will schedule a DOG (discovery on the ground) meeting with goal of working toward plans/proposals for implementation (DIG) (flow DAG DOG DIG DUG)
  - 2019-03-01 DOG meeting on Monday with Cornell and Stanford folks
  - 2019-03-08 Astrid has completed user sessions, Huda has completed literature review. Will work together to experiment with designs (building on work from Chalet meeting). Results so far suggest specialized research users have very little interest in knowledge panels, but it seems that undergrad users may have interest. Plan UX testing of ideas with undergraduates.
- Jason Kovariwill discuss in LTS whether there a URIs from Share-VDE data that are useful and might be added to catalog
  - 2019-02-22 ON HOLD until we have the real (with clusters) Cornell data back from Casalini (expected early March). Then have discussion in LTS – people will need to see / explore data before forming opinions on utility
  - E. Lynette Rayle 's first look at indexed data (see https://github.com/LD4P/qa_server/issues/16#issuecomment-468697789) doesn't make any sense... waiting for feedback from Dave
  - 2019-03-08 The SHARE-VDE response from services.ld4l.org is BROKEN, invalid ntriples (first line properly formed, rest missing < > and quotes??). Lynette will contact Dave about this.
- Simeon Warner re: Enhanced Discovery work, will talk with Jessie about Stanford plans and then talk to budget & finance
  - 2019-02-22 Still exploring remote work vs contract work, will need budget change approval
- Issues: Infrastructure Authority Requests https://waffle.io/LD4P/ld4p2-cornell

Status updates and planning

- Wikimedian-in-residence Cornell visit

- - Steven is leading organization, have suggested a draft agenda to review with Hilary
- Cohort engagement
- Discovery developer search
- Authority Lookups for Sinopia (Production QA instance: https://github.com/ld4p/qa_server/issues/)
  - Lynette has worked with Greg to get pingdom monitoring up for the QA server. Lynette will get notification if down.
- Prep for Cataloging Sinatra and other 45's (Discogs data, https://github.com/ld4p/qa_server/issues?q=is%3Aissue+is%3Aopen+label%3ADiscogs)
  - Tim – The tests are done and I've moved issue 36 to review/pending. I'll close all 3 related issues once the pull request is into QA and accepted. Before doing the PR, I ran Rubocop on all the Discogs code, and as a result had to do some refactoring. (It was news to me that methods can only be 14 lines long and modules 100.) This work is just about done, then I'll do the PR.
- Sinatra Profile Work
- Enhanced Discovery (see also https://wiki.duraspace.org/x/sJI7Bg)
  - "DOG": Discovery on the Ground. A few meetings in with an evolving set of participants.
    - Objective: A better understanding of the problem space. What questions and areas of development hold promise for prototypes and for more near-term concrete implementations we can roll out at our institutions, and how can we contribute this work back to the technological communities (e.g. back to the Blacklight core code)?
    - Have done some brainstorming and gathering of information/context. Astrid should be ready to report on user interviews by next week. Meeting planned for Monday so may have a better sense of concrete ideas to test with mockups and prototypes (that would be good to try at some point).
    - May be good to begin involving Cornell D&A/Usability group "on the ground" participants, based on availability/interest
- Travel and meetings (see LD4P2 Cornell Meeting Attendances)
  - US2TS
    - Huda notes interesting intersection between ontologies and machine learning, good to know what has been done practically
    - What might we learn about discovery investigations?
    - Panel is about domain knowledge graphs – excitement about our work was primarily about how to make connections between datasets. Good to get feedback on use of QA (screen shots: Tim Worrall want screenshot supporting the story "I searched for a work in discogs, results show extended context to allow selection of correct record, application returns BIBFRAME RDF derived from Discogs JSON, ready to be included as a new work in cataloging tool. Note no URI to reuse from Discogs, have to create one but retain local Discogs identifier in data.", QA/BFE - slide 33 and on in SWIB slides https://docs.google.com/presentation/d/1V64pncM5DnJJfY3-9SwT3iCzlKF9iLa3CQfxscK1_dI/edit#slide=id.g4826bbbf7e_0_39)
  - LD4 conference and pre-meeting with Harvard folks
    - Expect to send: Jason, Steven, Huda, Lynette, Simeon
    - Simeon has reserved minivan for Wed 8 through Sun 12 May
    - Everyone work on agenda document.
  - LD4P June 2019 Meeting: Agenda formation: https://wiki.duraspace.org/x/p4B7Bg
- Next meetings: