

2020-03-20 Cornell LD4P2 Meeting notes

Date: 20 Mar 2020

Attendees: Tim, Steven, Jason, Huda, Lynette, John, Simeon

Regrets:

Agenda & Notes

Review actions from [2020-03-06 Cornell LD4P2 Meeting notes](#)

- [E. Lynette Rayle](#) QA performance
 - 2020-02-14 Dave has made progress this week, is moving to have all data in index tailored to search in order to avoid SPARQL queries at search time, results are in a blob of RDF. Working on CERL first, expect to get this our soon. Will then try MeSH and OCLC FAST, then LC.
 - 2020-02-21 CERL was deployed with the new index strategy but no before and after to compare. However, this is small so we need to wait for LC or such to get a sense of possible improvement
 - 2020-02-25 New authorities have been brought online all the way through to Sinopia. These include CERL (searching person, corporate, imprint, or all of these) and Ligatus. Additionally, MeSH has been updated to include extended context and support for searching by subject or publication type.
 - 2020-03-06: Dave in process of converting everything over. Unsure of status for any authority, including MeSH. When LC done, we'll know whether this has impact since there are considerable usage data for LC. Dave and Lynette each working on other projects at the moment.
 - 2020-03-20: Dave is tied up with dept. COVID-19 planning. Log analysis shows two classes of error: 1) 502 is Dave's cache not responding, 2) 500 error which depends on query, Java null pointer error from Dave's system, and fewer of a third 3) timeout error
- [Simeon Warner](#) to ask Adam Smith to investigate cost and any issues with setting up a D&A Beta system to allow broader testing of some discovery ideas from this work.
 - 2020-03-20: STILL NOT DONE
- [John Skiles Skinner](#) to continue discussion with Hathi trust about an API or access to their index
 - 2020-02-28 HathiTrust have allowed institutional accounts to add a query parameter to get XML output, may also provide IP based access for prototypes. Have already made demo with a mock-up of access
 - 2020-03-06: Huda sent them IP Address... and then confirmed that it was indeed ours. Follow-up needed. [John Skiles Skinner](#) will do that before next meeting.
 - 2020-03-20 - Have IP set up for dev VM but only works from there only. More important update is that XML content will be turned off for a while unless we have particular need. Huda thinks that even if we use this for WHAM we could rely on canned responses for a while
- [Huda Khan](#) to copy [lessons learned](#) from BAM! into the main wiki and check scripts into github
 - BAM! lessons learned in [wiki](#) (DONE)
 - Scripts making way into GitHub (in progress)
 - 2020-03-06: NOT DONE. URLs need replacement; Huda will replace with comment + dummy URL. Will be completed 3/13
 - DONE – links: <https://github.com/LD4P/discovery/tree/master/scripts/bam> (also SMASH! scripts, for reference: <https://github.com/LD4P/discovery/tree/master/scripts/smash>)
- [E. Lynette Rayle](#) to ask Tiziana about SHARE-VDE APIs for real-time up-to-date search and for possible engagement in linked data best practices for authoritative data working group
 - 2020-03-06: NOT DONE. Will email by 2020-03-13
 - Will instead email a larger set of people (individually) about proposed best practices working group, will delay a bit until "new normal" has set in

Status updates and planning

- Enhanced Discovery (see also <https://wiki.duraspace.org/x/SJl7Bg> and <https://github.com/LD4P/discovery/projects/1>)
 - SMASH! (dev to run through 7 Feb, then user testing, video and write-up) – dev complete, video done, Hitchcock homage and cameos still under consideration, [lessons learned document](#) (DONE) in process and also [annif](#) use summary
 - Remains:
 - edit demo video for SMASH!
 - create Hitchcock video for talent show...
 - **Next 4 months: WHAM (H may or may not be capitalized) <--- KEY TOPIC**
 - https://docs.google.com/document/d/1u88qVOhp92C6Y1N9dNpNtjV_XOnusFB0nb2aJ-Q10FM/edit
 - **Consensus from discussion that top priority should be work on the autosuggest, with possibility of additional work on discogs or syllabus. Goal for the next week is to develop plan for autosuggest work and to understand any possible issues**
- What from the grant will be affected by COVID-19?
 - raised during March all-hands meeting as topic for which Michelle would like an update when feasible
 - Expect QA & discovery work to continue as planned
 - Expect cataloging work as part of cohort to be slowed
 - We note that any usability work will have to be remote, there might be opportunity to have some LTS staff help with usability
- Discovery presentation 3/3 debrief
 - positive feedback from many; high engagement from attendees
 - open syllabus data had positive review
 - timeline: visuals! there was at least one person who really liked this
 - knowledge panel – critique was wrt: info overload but not that this was not worth-while
 - auto-suggest and no-search-result both well-received
 - discogs metadata was well-received - method of bringing in trusted data. there are use cases where we may wish to index discogs data for search

- recording is in Drive. notes will be there. is it alright to send out follow-up email thanking people for attending with a link to the video? Questions raised about privacy, value for viewers and whether this should be public v. CUL-only. Notes summarizing can go on wiki.
 - DECISION: put video in LD4P-Internal. Can share internally for those who request.
 - Follow-up: summary of what we think we've learned. Goal is to prioritize work based on strongest feedback. Wait until next Friday to share broadly, assuming we've made decisions at that point.
 - this affords us 3.5 months to work on moving 1-3 items toward production... but not making it production-ready. includes analyzing existing infrastructure and consider whether formal usability testing is possible/advisable (using usability working group)
 - we are not looking at new work... this is to take current work forward
- Cataloging Sinatra and other 45's (Discogs data, https://github.com/ld4p/qa_server/issues?q=is%3Aissue+is%3Aopen+label%3ADiscogs)
 - Lookups for place not usable and hence places are not being recorded, relies on work from Dave to fix: https://github.com/LD4P/qa_server/issues/248 & https://github.com/LD4P/qa_server/issues/240
 - Have a currently insurmountable issue with nested profiles. When create Work profile with nested Instance profile there isn't a URI for the Instance (it just gets hung from a bnode). Without a URI the title of the Instance doesn't get indexed. The Sinopia team are unable to fix this in the near term.
 - Cataloging work continues with the above limitations
 - 2020-02-28 Steven update – I did a bunch of PCC profile and LOC policy related writing/correspondence; met with Huda, Tim, and John to discuss the Discovery Event (happy to help facilitate/mediate/rove on the day of the event); worked with Sinopia team to understand title search and display bugs that have been affecting Sinatra work (Jeremy has created https://github.com/LD4P/sinopia_editor/issues/2090 which looks at part of the problem); I still need to clean up the QA/Sinopia priority list to reflect the work completed by Lynette and Dave.
 - 2020-03-06: nothing new to report. 2090 issue above remain open - was not about imbedded templates but b/c label lacked lang tag... was not being indexed. Catalogers are providing feedback on UI concerns
 - 2020-03-20 not expecting progress in the near future
- Authority Lookups for Sinopia (Lookup infrastructure: https://github.com/LD4P/qa_server/projects/2, Authority requests: https://github.com/LD4P/qa_server/projects/1)
 - MeSH had typo at the Sinopia-level: FIXED; prs merged
 - FAST: EventName entity is now MeetingName. Until fix is in, QA Server is down. Cached data remain correct... until update, that'll work. Different configs already in place for direct access v. access to cache.
 - ISNI: requested feedback; Steven sent email summarizing good example of data; Lynette used that to discuss how QA would interact with those data. Challenges: no primary label (closest is ISNI #). Alt labels: huge list with no language tagging but clearly in various scripts and languages. All are equal so presents challenges. In ISNI UI, show person's name - clearly the Eng name is somewhere... they really need lang tag OR pref label. This is a problem beyond ISNI
 - Dave is working on indexing diacritics and various grammatical characters.
 - Need answer whether attempts to improve performance by indexing will indeed improve performance. LC will be the true test.
 - SHARE-VDE: issue arose re: a query where a subject was VIAF; one search brought in all of VIAF. Problem with SVDE data. Steven contacted them to ask about the data... and Dave is filtering this out on his end as a temp patch. Unclear whether they addressed issue and we have not seen it yet due to not having an updated data dump
- Travel and meetings (see [LD4P2 Cornell Meeting Attendances](#))
 - LD4P2 cohort and partner meeting, LC, 20 & 21 April
 - in-person is CANCELLED; will be virtual
 - Current discussions that Michelle has organized are reviewing status of cohort and whether some remote meeting is needed
 - [Knowledge Graph Conference](#) (Columbia University), May 6-7, workshops 5/4-5/5
 - Now remote and Huda confirmed, not quite sure of format
 - [LD4 Conference 2020](#) at College Station, TX (TAMU) - May 13/14, 2020
 - Hold on planning for a few weeks
 - [DCMI](#) (9/14-9/17, Ottawa): CFP due 4/13, not cancelled
 - Lincs: Linked Infrastructure for Networked Cultural Scholarship (Guelph, 5/7-5/9). Due March 16th.
 - check in on 3/13 as to whether anyone is submitting
 - [rdfs:seeAlso Conferences Related to Linked Data in Libraries](#)
- Next meetings:
 - 2020-03-20: