

2020-04-24 Cornell LD4P2 Meeting notes

Date: 24 Apr 2020

Attendees:

Regrets:

Agenda & Notes

Review actions from [2020-04-17 Cornell LD4P2 Meeting notes](#)

- [E. Lynette Rayle](#) QA performance
 - 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
 - 2020-04-03 On hold because Dave is otherwise tied up
 - 2020-04-10 Lynette has done an analysis of current performance [Performance Analysis for QA Lookups](#) – The Getty authorities are pretty much all sub-second which is not too bad, however averaging over all authorities we are around 2s. We find that for ARGOVOC and GEONAMES direct access (without extended context) but when going against the cache we find typically 3-4s with the QA portion (graph load and normalization) approaching a second – Lynette is not yet sure why Geonames has such a large normalization hit for example, why is the graph load time long for Argovoc?
 - 2020-04-17 Did an analysis sorting response time by complexity and size but it didn't show a clear picture, will try a little more analysis next week. When Ligatus and CERL were added - why is Ligatus fast and CERL slow? Dave expects to have more time soon
 - 2020-04-24 Lynette did an analysis of performance to try to understand whether speed is clearly related to data size or complexity of extended context. Result are that there isn't clear correlation. Tried to parallelize parts of QA and in some places saw slowdown, one place found speed improvement where the complexity is high. However, in the complex cases the times are often still rather long (0.5–2s) but not markedly longer than somewhat less complex queries. Still the worst cases are because of the retrieve time from Dave's cache, he is looking at why CERL is slow when we might not expect it to be. Unfortunately no clear path to improving everything from the QA side: Lynette will try to understand why the OCLCFast graph load is so slow.
- [E. Lynette Rayle](#) to set up best practices working group around linked data APIs for authorities documenting on [Charter 1 - Best Practices for Authoritative Data Working Group](#), with target start in early May
 - 2020-04-17 Heard from Christine @ Harvard, Rob @ Getty, Kirk @ OCLC, likely ISNI, nothing from wikidata, have also asked Samvera community – have enough to move forward with a meaningful working group
 - 2020-04-24 Yes from wikidata, will start early to mid-May

Status updates and planning

- Enhanced Discovery - **WHAM!** (see also <https://wiki.duraspace.org/x/sJl7Bg> and <https://github.com/LD4P/discovery/projects/1>)
 - See: [Organizing doc](#) and [Pseudonym thoughts](#)
 - See updates in [organizing doc](#)
 - John thinks that any code changes we will make for this work will be to the Blacklight gem rather than to the Cornell code. May be able to make a gem that is applied to the Blacklight gem.
 - Plan to review use cases and scenarios with Jenn & Frances, maybe also catalogers involved in authority work
- 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.
 - 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)
 - 2020-03-27 A couple of updates being put into production now: ISNI now available, Wikidata wasn't working properly because of a lack of response header which has now been added so that it plays nicely with Sinopia, also fix with CERL. Doing some infrastructure work to use Redis for caching and avoid a memory leak issue, working with Greg to understand issues
 - 2020-04-03 Recently done: cleaned up and updated UBER Issue for Authorities ([#253](#)), UBER Issue for Indexing ([#239](#)), OCLC FAST cache vs. direct ([Sinopia #2125](#)), Use Case: Authority Lookup UI option in place of auto-complete ([Sinopia #2123](#))
- Travel and meetings (see [LD4P2 Cornell Meeting Attendances](#))
 - LD4P2 remote partner meetings
 - Sinopia topics
 - Possible need for lookups and not just auto-complete to meet cataloger needs: https://github.com/LD4P/sinopia_editor/issues/2123
 - [Knowledge Graph Conference](#) (Columbia University), May 6-7, workshops 5/4-5/5
 - Now remote and Huda confirmed, sticking with original dates, May 7, 2:20pm in 20min slot
 - Blacklight summit happening May 7,8 virtual
 - [LD4 Conference 2020](#) at College Station, TX (TAMU) - May 13/14, 2020
 - Planning going ahead for meeting to work out virtual presentations. There was a survey of presenters to understand how they might want to present. Steven likes the idea of recorded sessions followed by office hours.
 - [DCMI](#) was 9/14-9/17, Ottawa - now postponed 'til Fall 2021
 - rdfs:seeAlso [Conferences Related to Linked Data in Libraries](#)
- Next meetings
 - ...