

# 2020-02-14 Cornell LD4P2 Meeting notes

Date: 14 Feb 2020

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

Regrets: John

## Agenda & Notes

### Review actions from 2020-01-31 Cornell LD4P2 Meeting notes

- **Huda Khan** to discuss with Astrid and David possible collaboration with U Chicago over usability (and maybe others in DOG team)
  - Notes: [https://docs.google.com/document/d/1\\_V7JfAKqSn63-G0zupLZigt8o7R3v\\_WXC7hWlt8w744/edit](https://docs.google.com/document/d/1_V7JfAKqSn63-G0zupLZigt8o7R3v_WXC7hWlt8w744/edit)
  - 2020-01-10 Chicago report will be discussed in DAG meeting on Jan 21 (expect to happen as scheduled). Follow up with David and Astrid to look at what we could apply from what they learned and what additional user studies they could help us with
  - DAG presentation did occur and it was very interesting ([link to meeting](#)) – in depth interviews to understand research needs with open-ended questions, need more thought to understand how we might apply lessons from this
  - 2020-02-14 Chicago uses vufind which already has a notion of similar items, questions of what types of semantic similarity might be interesting. Also discussion of author browse and articles; presentation of introductory materials for adjacent topics – they are going to discuss internally what to investigate further. [Notes from U Chicago](#)
- **E. Lynette Rayle** QA performance
  - Dave made some changes to address issues with 50x responses, still possible issues under high load (possibly something on the Sinopia side but waiting to look at IP address to logging to identify whether same source is hitting us with same request)
  - Ongoing discussion about the need build more complex indexes to deal with slowness of complex queries
  - 2020-01-31 - Dave, Steven, Lynette had a meeting this week. Created a list ([HERE](#)) . Dave is still optimistic about perf improvements in SPARQL with change from CONSTRUCT to SELECT, but not sure when Dave will be able to try this. Also looking at indexing approaches with smaller sources than SVDE, will try LOC which is expected to take 3 days. Will also work to cache context when needed and not to request it when it isn't needed. Steven noted this on #authorities channel. Three categories of approaches: 1) amount of extended context, 2) efficiency of queries, 3) scalability of requests. Longer term there are questions of lookup vs. autocomplete modes
  - 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.
- 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
- **John Skiles Skinner** to continue discussion with Hathi trust about an API or access to their index
  - 2020-01-31 There is investigation but not sure whether it will result in something we can use
  - 2020-02-14 Some more discussions with Hathi and suggestions was to use current search with debug facility that includes things like facet values in machine readable form (requires either 1) a user account for testing, or 2) to use IP access for our dev machine but there is some issue of fixed external IP for our dev VMs)
- **Huda Khan Tim Worrall John Skiles Skinner** to finish up [lessons learned](#) from BAM!
  - 2020-02-14 Will be done by next week...
- See items under travel/conferences

### Status updates and planning

- Discovery presentation on 3/3: what is agenda? who is speaking? who is announcing to CUL?
  - Goal is to disseminate our week and then get some feedback about what might be promising
  - We need to have a way to take good notes – designate a note taker
  - View from staff involved with virtual libraries would be interesting, also D&A user reps
  - Need to be careful about managing expectations for stage this work is at and what might happen or not going forward
  - ACTION - Jason to write up, circulate and then send out on Tuesday
- Cataloging Sinatra and other 45's (Discogs data, [https://github.com/ld4p/qa\\_server/issues?q=is%3Aissue+is%3Aopen+label%3ADiscogs](https://github.com/ld4p/qa_server/issues?q=is%3Aissue+is%3Aopen+label%3ADiscogs))
  - Lookups for place not usable, relies on work from Dave to fix: [https://github.com/LD4P/qa\\_server/issues/248](https://github.com/LD4P/qa_server/issues/248) & [https://github.com/LD4P/qa\\_server/issues/240](https://github.com/LD4P/qa_server/issues/240)
  - Work is being done but places are not being recorded
  - Steven: Revised the Cornell Sinopia documentation (<https://docs.google.com/document/d/1vatjuDOAy5Qzi-Jj-JmH9zytyjUd6c3X4DzMWvybL4/edit#heading=h.3tc5xzitbah8>) to reflect the Discogs lookup and UI changes
  - 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 this above limitation. If we want to later use the data we'll have to create URIs for the Instances. Also some concerns about how readily our BF could be transformed into MARC if that is what we wanted (because it is XSLT the very particular form of the BF RDF/XML is critical)
- 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)
  - Open meeting March 3, 2-3:30pm in Mann 102 and should Zoom it too
  - How will we decide what to take forward from KAPOW!, BAM! and SMASH!? (or as Tim put it, "what happens in late February?")
  - SMASH!
    - Demo from Huda
    - Tim on Agra at the moment, just work on write up still to do
    - Next step is to devise and implement user tests for all three experiments in this phase
- Authority Lookups for Sinopia (Lookup infrastructure: [https://github.com/LD4P/qa\\_server/projects/2](https://github.com/LD4P/qa_server/projects/2), Authority requests: [https://github.com/LD4P/qa\\_server/projects/1](https://github.com/LD4P/qa_server/projects/1))
  - Lynette has worked on some issues with the monitoring page, hope to make a release sometime today together with new CERL support from Dave

- Additional extended context work is pending revisions from Dave but less important than the performance issues
- LODLAM discussions
  - Getty cache LC data for robustness, something we have thought of as a motivation for caching behind QA
  - Kevin had a session on notifications in the context of library data and the need for APIs to support our workflows, ties to FOLIO and authority maintenance ([Lynette's summary notes from this session](#))
- Travel and meetings (see [LD4P2 Cornell Meeting Attendances](#))
  - [Knowledge Graph Conference](#) (Columbia University), May 6-7, workshops 5/4-5/5
    - Workshops due 1/15 (not doing this), proposals due 2/28, talks 20mins
    - [Huda Khan](#) will lead effort to consider proposal
      - [Preliminary brainstorming](#)
  - LD4P2 cohort and partner meeting, LC, 20 & 21 April
    - Hotels are expensive those days
  - LD4 Conference at College Station, TX (TAMU) - May 13/14, 2020
    - Proposals went in 2020-01-31 – Discovery, QA/Sinopia, BOF for best practices, Intro to LD workshop
    - Expect hear back in next 2 weeks, registration will open then (first come first served)
  - rdfs:seeAlso [Conferences Related to Linked Data in Libraries](#)
- Next meetings: