

# 2019-12-20 Cornell LD4P2 Meeting notes

Date: 20 Dec 2019

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

Regrets: Lynette

## Agenda & Notes

### Review actions from 2019-12-13 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)
  - Will revisit question in January; Chicago supposed to have report by now... regardless, understanding collaboration in January when better understand resources and what to test
  - 2019-12-20 Expecting update in January

### Issues/Blockers (for time since call at 10)

- Lynette: Connection time outs were happening due to increased requests (fixed); cache layer is returning a lot of 500s - Dave is working on this. No blocker - work itself is happening. Adding IP address to logging to identify whether same source is hitting us with same request.
- John: getting documentation from the Internet Archive. Needs to know whether there is an API to full text search the books. Recommend look into HathiTrust (Michelle Paolillo is CUL rep for Hathi and can point you to documentation).
- Tim: no blockers or issues
- Huda: looking at semantic search to define data requirements. no blockers

### Status updates and planning

- BAM!
  - We watched and it was awesome!
  - Plan to have an open meeting in mid-February to which we encourage D&A and other folks to attend – March 3, 2-3:30pm in Mann 102 and should Zoom it too
- Review of SMASH! brainstorming results from
  - Meeting last Friday: <https://docs.google.com/document/d/1ahhkmSGfKE8TCz781jIjB2mbvmFq85Wc8wnYKSC9RQ/edit>
  - Beginning to put together potential SMASH! examples: <https://docs.google.com/document/d/1u2NYavlpzjbqrmSvDpuFnmCXSljVW9tTfaU5B4N4zrw/edit>
  - John is looking at how we might do better with searches that currently yield zero results. Experience to date suggests that full-text approaches are the most promising. Have looked at archive.org, hathi and google books. So far, Google books is the only service with an available API. In discussion with Hathi and we might prefer that, or perhaps mix sources. Question of analysis of Google vs. Hathi results – certainly different but we don't have a good sense of how
  - Estimate is that less than 1% of search results that yield zero results, most are legitimate queries on obscure topics rather than simple typos
  - Blacklight has some "did you mean?" facility but it isn't good so we and every else turns it off
  - Tim is looking at autocomplete functionality against VIAF, VIAF via Dave, and FAST
  - Huda is looking at results from search of subject terms (including alternate labels) in Dave, and then considering broader and narrower terms in order to offer more subject categories to look at. Thinking of looking at subject headings obtained from results from catalog search
- Prep for 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))
  - STARTING with items not in Discogs but AWAITING more work in Sinopia to import data. Sinopia work cycle 2 (through December 6) will we hope include the ability to read in RDF back from Trellis. We hope that we can leverage this to import RDF from a lookup in Discogs or ShareVDE.
  - Work is going on. Questions about where to put data and roles, and how many abstractions of work are required (e.g. for performance and for composition). Lookups are too slow and too buggy
- Enhanced Discovery (see also <https://wiki.duraspace.org/x/sJl7Bg> and <https://github.com/LD4P/discovery/projects/1>)
  - BAM! (to run through end October)
    - For reference: [Initial draft of BAM plan](#)
    - [BAM Screenshots](#) (Tim's call number screenshots added.)
    - Tying up video next week, John has submitted his section (DONE)
    - Beginning draft for [lessons learned](#)
  - SMASH! (to run through end of January)
  - How will we decide what to take forward from KAPOW!, BAM! and SMASH!?
- 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))
- Travel and meetings (see [LD4P2 Cornell Meeting Attendances](#))
  - 5th International LODLAM SUMMIT at the The Getty Center in Los Angeles. February 3-4, 2020
    - Steven, Lynette and Simeon will attend
    - E. Lynette Rayle will put QA in for the "tool challenge": CFP due 1/8
  - [Knowledge Graph Conference](#) (Columbia University), May 6-7, workshops 5/4-5/5
    - Proposals due 1/15
    - Huda Khan will lead effort to consider proposal
  - LD4 Conference at College Station, TX (TAMU) - May 13/14, 2020
    - Jason is on PC
  - rdfs:seeAlso [Conferences Related to Linked Data in Libraries](#)
- Next meetings:

