

2019-03-22 Cornell LD4P2 Meeting notes

Date: 22 Mar 2019

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

Regrets:

Agenda & Notes

Review actions from 2019-03-15 Cornell LD4P2 Meeting notes

- **Steven Folsom** keep track of cataloger perspective on work to support LC Names is in QA with context
 - 2019-03-15 - Dave is focused on SHARE-VDE as his top priority (https://github.com/LD4P/qa_server/issues/16), when that has been moved forward we need to emphasize our other priorities again (https://github.com/LD4P/qa_server/issues/53 https://github.com/LD4P/qa_server/issues/47)
 - 2019-03-22 - Per Tuesday meeting most SHARE-VDE issues resolved for now, Dave is hoping to have progress with LCNAF and LCGFT by next Tuesday
- **Steven Folsom** to reach out to Nancy to see what data we get from the "Profiles Wrangler" survey about QA needs, will then reach out to cohort to check results and see if anything is missing
 - 2019-03-08 - Received a confirmation from Nancy that they are happy to share. The survey closes 2019-03-08.
 - 2019-03-15 - Paloma mentioned before the survey closed that <https://www.rbms.info/vocabularies/> and <http://www.ligatus.org.uk/lob/> were named. Waiting for a complete list. The former, I (Steven) don't believe, is LOD; we'll have to confirm and come up with a plan.
 - List of datasets named in the Profile Wrangler Survey:
 - RBMS (No RDF yet, but the editorial group is working towards it)
 - Language of Bindings, Ligatus - lobt (no RDF)
 - Canadian Subject Headings (No RDF)
 - RVM - Répertoire de vedettes-matière (No RDF)
 - Wikidata (direct access via QA work)
 - ISNI (No RDF)
 - LCDGT (pending in QA)
 - OLAC video games (pending in QA)
 - [RightsStatements.org](https://rightsstatements.org)
 - Creative Commons licenses
 - Video Game Ontology
 - Steven will make sure there is a github issue for each vocabulary and try to find an owner for each. We will need to find out whether there are URLs for these terms – might not be able to support all
- **Jason Kovari** will check with Tim Thompson whether it is OK to use the same transformation of RBMS vocabs that was used in LD4L-Labs
- **E. Lynette Rayle** LDPATH gem has two problems: 1) bug in LDPATH parsing paths, 2) it is doing load for graph every time it tries to access something (SLOW)... may need to fork LDPATH gem to make more efficient since will not need to reload graph every time (for our use case)
 - 2019-03-15 - Tracking in [Issue #56](#), have not yet put example in LDPATH repo for Chris
 - 2019-03-22 - Still working on example for LDPATH issue. Will hope to speak with Chris at LDCX
- **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)
 - 2019-03-08 - 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
 - 2019-03-22 - [Documentation](#) has been started but still has some TODOs to fill in information. Created issue to track ([Issue #61](#))
- **Huda Khan** & **Steven Folsom** to go through spreadsheet <https://docs.google.com/spreadsheets/d/1gklgTQUNd9nr3xTI5kdFYK51SRG6TeMRH678N1FtNQ0/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 Khan** and **Steven Folsom** will 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 Khan** and **Steven Folsom** to come up with wireframe proposals to present to D&A group on March 19.
 - Based on D.O.G. and D.A.G. work, generated [set of ideas](#) (to elicit feedback). Based on bringing in relationships from external data (that hopefully go beyond what is already in the authorities). Areas to try: cross-references, relationships, other sources (e.g. databases, etc.).
- **Jason Kovari** will discuss in LTS whether there are URIs from Share-VDE data that are useful and might be added to catalog
 - 2019-02-22 – 2019-03-22 ON HOLD until we have the good (with clusters) Cornell data back from Casalini (not sure when expected). Then have discussion in LTS – people will need to see / explore data before forming opinions on utility
 - Output for QUADS in SVDE:

<<http://share-vde.org/sharevde/rdfBibframe/Provenance/CORNELL136709>> a <<http://www.w3.org/ns/prov#Entity>> ;

<<http://www.w3.org/2000/01/rdf-schema#label>> "CORNELL136709" ;

<<http://www.w3.org/ns/prov#wasAttributedTo>> <<http://share-vde.org/sharevde/rdfBibframe/Agent/CORNELL>> .

CORNELL136709 refers to <https://newcatalog.library.cornell.edu/catalog/136709>.

- [Simeon Warner](#) re: Enhanced Discovery work
 - 2019-03-15 Working on posting position again...
- Issues: [InfrastructureAuthority Requests https://waffle.io/LD4P/ld4p2-cornell](https://waffle.io/LD4P/ld4p2-cornell)

Status updates and planning

- Waffle.io is <https://waffle.io/closing-its-doors> May 16 – do we move all work areas to github projects or to adopt something else?
- Wikimedian-in-residence Cornell visit
 - What data in wikidata is of most interest for discovery and other projects?
 - How accurate/good is wikidata in various domains?
 - Stanford priority to get some knowledge card work done in time for May meeting – we are interested in working on this too
 - see google doc
- Cohort engagement
- Discovery developer search
- Authority Lookups for Sinopia (Production QA instance: https://github.com/ld4p/qa_server/issues/)
- 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 will update discogs code to use user token instead of keys that don't currently work in discogs
 - https://github.com/LD4P/qa_server/issues/49 - not expecting to be able to make progress until Sinopia update in early April
 - Tim working in getting up to speed with React
- Sinatra Profile Work
- Enhanced Discovery (see also <https://wiki.duraspace.org/x/SJl7Bg>)
 - Huda is now chairing the LD4 DAG with Tom Cramer
 - LD4 "DOG" continues for the meantime with discussions around findings, processes for investigation and implementation (i.e. "on the ground"). At appropriate juncture, good to bring in additional Cornell discovery/Stanford discovery implementation folks.
 - Simeon – I think we need a set of real user discovery stories to use as examples when we talk about discovery. As I hear, "my current example is Lincoln" it all sounds so arbitrary. I think we should be saying "Let's take the example of an undergraduate student searching for "lincoln" in order to find material for an American History 101 assignment on this and that (this example came from a user study we did in Cornell's Olin library...)" even if the actual terms and results are the same.
 - John Updike story is from the old D&A integration layer stories
 - Might get some stories from the Harvard and Chicago input in the DAG
 - Also input/user stories from [Astrid Usong](#) based on her user interviews
 - ACTION - [Jason Kovari](#) and [Huda Khan](#) will work on some more expansive stories that relate well with our collections
 - Suggestion that bringing together archives and special collections with the catalog would be interesting
 - Is there an opportunity for gamification around the catalog? Could have this as an opportunity for supervised reconciliation work, or correction of errors
 - Importance of transparency about data sources and how/who decides what sources are appropriate – under what circumstances is it appropriate to use data from Wikipedia/Wikidata?
 - ==> PLAN TO FOCUS ON IDEAS FOR A NEAR TERM WORKPLAN NEXT FRIDAY
- Travel and meetings (see [LD4P2 Cornell Meeting Attendances](#))
 - LD4 conference and pre-meeting partners
 - 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>
 - Michelle (2019-03-21, email): "June in-person partner meeting in DC: planning hasn't started yet, it would be good to have at least one person from each of Cornell/Harvard/Iowa SLIS/LC/PCC/Stanford involved in the planning. Let me know if you want to take part."
 - Partner meeting is day after cohort; can we wait to plan until after May meeting?
- Next meetings:
 - Lynette and Steven not available for 29 Mar meeting
 - Jason not available for 5 Apr meeting
 - Steven out April 19th.