Applied Linked Data Call 2015-07-23

Attendees:

- Steven Anderson (Boston Public Library)
- Trey Pendragon (Oregon State University)
- Arwen Hutt (UC San Diego)
- Corey Harper (New York University)

Linked Data Fragments Update:

- Comments requested for https://github.com/ActiveTriples/linked-data-fragments/pull/18
  - (Unit tests do work locally if you have marmotta up. Otherwise should Marmotta be mocked out? Or have Travis run an instance?)
  - Likely mock out the marmotta part for the tests.
- Another backend option should likely be included in the initial release.
  - In RDF.rb and there is a model called Repository that has a good interface. Can then do this in-memory using that without a configured cache store.
- Configurable routing is likely to be challenging
  - No one has really done this before so research and a stab at it will likely just need to be done.

Broader / Narrower SKOS concepts in an application attempt by Trey Terrell report back.

- Trouble with weighting...
  - In the past, they used the “all” field to search but then all text within it would all have the same weight. Need to really do better weighting for this to work.
  - Only narrower was useful and would need to be less weighted than the all field… doesn’t want to split them up to then have to add that to the solr config for each field.
    - Could maybe have three groups: high weight, medium weight, and low weight? Then use some character indicator with wildcard to determine which category the solr field belongs to.
    - Default solr field would be high weight perhaps to prevent having to redo the solr schema. Then new fields can use this for medium / low weight.
      - But… then sadly solr doesn’t support negative regular expressions.
      - But… what if you copy something twice anyway? How does that affect solr weighting?
  - Could perhaps add this type of support to the code that we are working on (even just alt labels would be great).

Question to Trey on the removal of type-ahead resolution in Oregon Digital?

- Currently if you, say, type into Geographic field, will pop-up with all of the results from Geonames. Can then try to pick the correct uri then. But...
  - Only as fast as search endpoints are.
  - Ran into cases where one field would search multiple vocabularies but their labels are the same… how do you know you are picking the right one? Only have label and uri.
    - Conversation with Oregon Digital users is that they never uses the type ahead… they just go and find the uris rather than relying on it.
  - So nobody uses type-ahead. Great for demos but actually doesn’t help!
    - Labels aren’t a good enough context.
  - New thinking now is:
    - Why can’t people just put free text in a field?
      - So just type stuff in. There will be a button or something like that will then give them a popup with more context.
      - Could also allow one to configure what vocabulary(ies) to use for the field too in this interface.
  - What is context?
    - Description
    - Broader and narrower terms
    - Type of what that element is
    - And most importantly: How many documents in the repository are using this term.
      - If it is zero, the user is likely picking the wrong one.
  - New term for this type of thing is: Metadata Enrichment Interface.
  - Some other notes about why this came up regarding mediated deposit in the repository:
    - People at OSU who aren’t librarians won’t understand how to pick the right terms. So allow them to type it in free terms… then perhaps someone else enriches it downstream using the above interface. It doesn’t have to be done by the entry user in this case.
    - Also don’t want to stop people from putting in keywords. If a match doesn’t exist in any linked data source, perhaps they could mint a local linked data entry for it.
  - Action item of making a repo on project hydra labs for this type of thing and adding the user stories from Oregon Digital there. This could be useful for other institutions.

Linked Data Fragments Standup:

- Will be on Friday, July 31st at 11:00 AM PST / 2:00 PM EST on the same Google Hangouts link.

Next Official Meeting:

- Next meeting will be August 6th at 9:00 AM PST / Noon EST.