# 2015-02-24 breakout: Services on linked data

# Services on linked data

LD4L Workshop Breakout Session, Tuesday, February 24

facilitator: Jon Corson-Rikert

### Risk of not knowing what to search for

- Providing discovery endpoints
  - "hardened" SPARQL endpoints may be less prone to down time e.g., Fuseki documentation states that "authentication and control of the number of concurrent requests can be added using an Apache server"
- · publishing starting points with examples and standard extracts may help
  - o emulate Social Explorer http://socialexplorer.com as a way to query the contents of a larger data source, in that case census data
  - the linked data fragments technology (http://linkeddatafragments.org) may facilitate hosting linked data without the server-side overhead and risk of a public SPARQL endpoint
- VIVO/Vitro 'rich export' augmenting standard linked data responses with standard queries
  - o e.g., get all a person's publications from a single request rather than client having to issue multiple requests

## Synchronizing harvested information

- · Risk of harvested or aggregated information going out of sync
  - Resource sync standard addressed the need to repeatedly synchronize and update
- Semantic Web crawling leveraging HTML web crawler experience
  - o what's attached
  - what has changed

### Desire to be able to query on different axes

• e.g., query OCLC Works by VIAF identifier to get a list of works by that author

#### **Reconciliation services**

- not necessarily centralized or monopolies
- would work best in an iterative mode, with curation and provenance to manage difference of opinion (or evidence)
  - o who's made that assertion differentiate librarians from crowdsourcing
  - o some way to express variable confidence levels
- incorporate feedback from users
- need protocols could leverage a common API for reconciliation building on the OpenRefine API specify as much metadata as you have, get ranked results back
- surface (publish) the results known servers, as with annotations select which servers to request responses or harvest data from
  - o notifications of new matches?
  - $^{\circ}$  ability to +1 or thumbs-up the connection to corroborate Reddit gets a lot of traction that way
  - o repeating assertions in multiple repositories
- · sameAs.org but with other expressions for and levels of confidence in the relationship

#### **Validation**

- RDF data shapes working group
- DCMI tutorial on RDF validation
- Measure the consistency of ontology use
- · Linked data needs mashup tools that test connections and illustrate bringing data together

# **Ontology extension mechanisms**

- Schema.org extensions being proposed and managed on GitHub
  - Bib Extend group and BiblioGraph

# Ability to push bookmarks

- Small graphs of data, consumable by others, to a platform similar to Mendeley but not limited to bibliographic material
- · A service where I can push the results of my search, organized by topic
- Add things to a collection I have
- · Similar to an annotation service

- You search, you refine it, you step back now only save as bookmarks at one level
- Nobody can use your web bookmarks nowHide the URIs behind a UI

#### **Additional ideas**

- · Semantic autotagging
- Nanopublications breaking academic articles into independent assertions with a mechanism to agree/disagree
- Side wikis a plugin for the Netscape browser where a wiki could be associated with any web page and display additional, user-entered content or commentary on any web page
- individual libraries will become the authorities for special collections items, people, events

  o queries to a central area would find a match
- cache the sameAs so don't have to re-query; everybody who consumes has the cross-links
   the sort of thing that OCLC might end up doing could be any type of object logical to start with works
   regular expressions to apply against EAD to suggest what is linked to; feed into a system to validate, then give pointers to the link
- a clustering algorithm to track the number of times a link between two entities is traversed, effectively shortening the distance between them
- a better page rank algorithm for linked data
- anybody a favorite semantic search engine (no too siloed)
- · visualizations have to be crafted individually