Challenges for Linked Data

Table of Contents

- Overview
- Challenges
  - Reconciliation Services (Things to Things)
  - Entity Resolution (Strings to Things)
  - Lexicalization (Things to Strings)
  - Caching of External Fragments / Entities
  - External Entity as Subject in Fedora
  - Controlled Vocabulary Management
  - Use with SEO
  - SPARQL Queries with Fedora
  - Hydra Stack over a Triple Store
  - Interoperability
  - LD Path
  - Linked Data Publishing
  - Role of Solr

Overview

This page includes a list of challenges related to working with triples and triples stores that represent non-trivial levels of work to address each challenge.

What you can do?

- Add a new challenge to the list.
- Add your organization under Organizations Working on This if you are actively working on this or are interested in working on this. Optionally, add a contact for your organization.
- Update Potential Solutions sections for any challenge you are working on to describe what approach you are taking.

Challenges

Reconciliation Services (Things to Things)

Description

Given a URI, how do you find other URIs that represent the same thing? This is primarily used to go from a local URI to an authoritative URI, or vice versa. Also addresses issues of how are the various URIs are represented in the local triple store?

Examples:

- 2 URIs that are the same thing (owl:sameAs)
- local URI mapped to authority URI (owl:sameAs)

Potential Solutions

Entity Resolution (Strings to Things)
**Description**

Given a string (e.g. "Mark Twain"), how do you find URIs (local, external, or authoritative) that represent this thing? This is complicated by the fact that various string labels (e.g. "Mark Twain", "Twain, Mark") may represent the same thing.

**Examples:**

- 2 labels (e.g. "Mark Twain", "Twain, Mark") that should resolve to the same URI (e.g. http://viaf.org/viaf/50566653)
- I have a string (e.g. "Mark Twain"), give me a URI (e.g. http://viaf.org/viaf/50566653)

**Potential Solutions**

**Organizations Working on This**

---

**Lexicalization (Things to Strings)**

**Description**

Given an URI, how do you get a user friendly label (e.g. "Mark Twain")? This includes other challenges related to labels. ([video presentation](#))

**Examples:**

- I have a URI (e.g. http://viaf.org/viaf/50566653), give me a label (e.g. "Mark Twain")
- I have a URI (e.g. http://viaf.org/viaf/50566653), give me a label in a specific format (e.g. "Twain, Mark" instead of "Mark Twain")
- I have a URI, give me the label in another language.

**Potential Solutions**

Use a linked data fragments server or sidecar triplestore to cache the triples describing the external URI

**Organizations Working on This**

Amherst College will use a caching layer (either an existing LD Fragments server or a simple database) along with a resolver to extract a label (e.g. in order, look for skos:prefLabel, mads:authoritativeLabel, rdfs:label)

---

**Caching of External Fragments / Entities**

**Description**

For example, a label is retrieved from an external vocabulary source (e.g. Library of Congress) and stored locally for performance and to insure availability in the case where the external source's server is down. What process is used for caching values and refreshing cached values?

**Potential Solutions**

LD Cache and Linked Data Fragments was discussed at Hydra Connect 2015 as a potential solution.

**Organizations Working on This**

In the context of "Lexicalization" (above), Amherst College will be working on this. Certainly, external values will be cached; cache invalidation will be an out-of-band process that periodically checks for updates (monthly? annually?).

---

**External Entity as Subject in Fedora**

**Description**

Fedora requires that an entity be a resource within Fedora to be able to make statements about that entity with the entity as subject.

**Potential Solutions**

- Create a resource in Fedora and use owl:sameAs to record the external URI for the entity. Not ideal. Would prefer use the external URI for the entity without having to create a resource within Fedora.

**Organizations Working on This**
### Controlled Vocabulary Management

**Description**

Have a controlled vocabulary with URIs identifying terms.

**Potential Solutions**

- [Questioning Authority](#) gem has been used for controlled vocabularies.
- See Oregon Digital's [ControlledVocabularyManager](#)

**Organizations Working on This**

- Oregon Digital: [https://github.com/OregonDigital/ControlledVocabularyManager](https://github.com/OregonDigital/ControlledVocabularyManager)

### Use with SEO

**Description**

How to leverage linked data for search engine optimization?

**Potential Solutions**

- Looking for standard reason recipe and crosswalks to [schema.org](https://schema.org) from various vocabularies.
- Include linked data on a web page
- Make data available for machine-to-machine queries

**Organizations Working on This**

### SPARQL Queries with Fedora

**Description**

How to use SPARQL queries to get information stored in Fedora?

**Potential Solutions**

- Use [camel messaging](#) already supported by Fedora to sync Fedora with a triple store and perform SPARQL queries over that triple store.

**Organizations Working on This**

Amherst College will be using the camel-based synchronization mechanism to replicate fedora metadata to a searchable index.

### Hydra Stack over a Triple Store

**Description**

Define requirements and work required to be able to have the Hydra Stack (e.g. Blacklight, Solr) backed by a triple store instead of Fedora.

**Potential Solutions**

- Tom Johnson and Trey Terrell have a path but time / development is needed
- Active Triples 1.0 needs to be released and some Productive Refactoring of ActiveFedora

**Organizations Working on This**

### Interoperability
Description
This was in the notes from Hydra Connect, but there wasn't a description. If others know what is meant by interoperability, please fill in a description here.

Potential Solutions

Organizations Working on This

**LD Path**

Description
**LDPath** is similar to XPath, but for RDF. Supports translations from RDF to formats such as JSON using a very compact notation.

Potential Solutions
With the fcrepo-transform extension module, one can apply LDPath programs to any repository resource as described in the Fedora documentation.

Organizations Working on This
Amherst College will likely be using this, especially as part of a solr indexing pipeline.

**Linked Data Publishing**

Description
Following best practices for publishing linked data.
- does Hydra only manage linked data?
- or also serve as a means for publishing linked data?

Potential Solutions

Organizations Working on This

**Role of Solr**

Description
Role of Solr when working with triple stores.

Potential Solutions

Organizations Working on This
Cornell - See active_triples-solrizer for an example of having solr based on triples in a triple store.