Digging Into Linked Data: Perspectives From The Long Tail

By: Paromita Biswas and Andrea Leonard

Abstract

The success of the semantic web depends on widespread participation by cultural heritage institutions and other organizations in making connections between open, structured datasets. Large university libraries are beginning to make such connections. It is time for mid-size and smaller libraries to take the leap and establish themselves as playing a part in this web of data. In particular, digital collections of many of these libraries represent significant regional or local history collections; metadata of these collections exposed as linked data can bring visibility for these unique resources. But do these libraries have the resources to create semantic data? What kinds of resources and technical support do these libraries need? How much and what kind of training do their staff need for linked data projects? This presentation focuses on a collaborative linked data project between two mid-sized academic libraries--Western Carolina University and Appalachian State University. The libraries are members of the Western North Carolina Library Network and share a common catalog. Both libraries have significant special collections on Appalachian culture and history. Their project aims to expose a slice of their digital collections on Appalachia as linked data and build connections to related datasets on the web thereby exploring the possibilities of the semantic web. The project also serves as the testing bed for future such collaborative work, possibly on a larger scale. The presentation will highlight the successes and challenges faced by the presenters as they delved into this project. For example, what resources and training did they need? How successful were they in manipulating digital collections metadata in OpenRefine; navigating the intricacies of various data models such as those from Europeana and DPLA; sorting through the multitude of controlled vocabularies that are available as linked data on the web and selecting the best possible options? How difficult or easy was it to figure out linked data jargon, such as dereferenceable URIs and RDF skeletons? What kind of technical support was needed for setting up triples stores and querying linked data via SPARQL endpoints? The presenters hope this presentation will be a useful learning experience for those who are thinking of venturing into creating access for their special collections using linked data tools particularly for those from mid-size to small libraries.

Paromita Biswas (WCU) and Andrea Leonard (ASU). 2018. "Digging into linked data: perspectives from the long tail"; Presentation at the 2018 Texas Conference on Digital Libraries. NC Docks permission to re-print granted by authors.
Digging into Linked Data: Perspectives from the Long Tail

Paromita Biswas, Western Carolina University (WCU)
Andrea Leonard, Appalachian State University (ASU)
Texas Conference on Digital Libraries 2018
Hunter Library, WCU

Belk Library & Information Commons, ASU
Choosing the dataset: WCU

Oldest and most extensive collection on Appalachian culture and heritage

Number of items for project: 25
Choosing the dataset: ASU

Number of items for project: 20

Quilts found in:
- University Library Art Collection
- ASU Historical Photos
- Kirby and Eller Family Letters
Learning Curve: Concepts

Data Models/Mapping
Europeana Data Model: ProvidedCHO (core class) and its properties (EDM provides the option to map to Dublin Core terms)
Resource Description Framework (RDF)
Data in triple statements

Literals and URIs
5 star data
Dereferenceable URIs/Reuse URIs

RDF Serialization: Turtle-Terse RDF Triple Language
Easily readable and could be manually written

Mapping

<table>
<thead>
<tr>
<th>Metadata elements</th>
<th>EDM Predicate</th>
<th>Type of object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>dc: title</td>
<td>Literal</td>
</tr>
<tr>
<td>Creator</td>
<td>dc: creator</td>
<td>URI/Literal</td>
</tr>
<tr>
<td>Type</td>
<td>edm: type</td>
<td>URI</td>
</tr>
<tr>
<td>Medium of Original</td>
<td>dcterms: medium</td>
<td>URI</td>
</tr>
<tr>
<td>Date of Original</td>
<td>dc: date</td>
<td>Literal</td>
</tr>
<tr>
<td>Dimensions</td>
<td>dcterms: extent</td>
<td>Literal</td>
</tr>
<tr>
<td>Description</td>
<td>dc: description</td>
<td>Literal</td>
</tr>
<tr>
<td>Subject - Topic</td>
<td>dc: subject</td>
<td>URI</td>
</tr>
<tr>
<td>Subject - Craft</td>
<td>dc: subject</td>
<td>URI</td>
</tr>
<tr>
<td>Subject - Group</td>
<td>dc: subject</td>
<td>URI</td>
</tr>
<tr>
<td>Craft category</td>
<td>dc: subject</td>
<td>URI</td>
</tr>
<tr>
<td>Location</td>
<td>dcterms: spatial</td>
<td>URI</td>
</tr>
<tr>
<td>Source Institution</td>
<td>dc: publisher / dc: provenance</td>
<td>URI/Literal</td>
</tr>
<tr>
<td>Collection</td>
<td>dcterms: isPartOf</td>
<td>Literal</td>
</tr>
<tr>
<td>Inventory Number</td>
<td>dc: id</td>
<td>Literal</td>
</tr>
<tr>
<td>Copyright Information</td>
<td>dc: rights</td>
<td>Literal</td>
</tr>
<tr>
<td>Digital Publisher</td>
<td>dc: publisher</td>
<td>URI</td>
</tr>
<tr>
<td>Object URL</td>
<td>edm: isShownAt</td>
<td>URL</td>
</tr>
</tbody>
</table>

See:
*Heath and Bizer, How to Publish Linked Data on the Web*
Learning Curve: Tools

Getting the data ready; uploading to OpenRefine
Adding the RDF extension and reconciliation services
Reconciling controlled vocabularies with established URIs
(LCNAF/LCSH/VIAF/Getty/DBPedia/Wikidata)
See:
*Verborgh and De Wilde’s Using OpenRefine (2013)
*Hooland and Verborgh, Linked Data for Libraries, Archives and Museums (2014)

Edit RDF Skeleton (RDF Schema Alignment)
creating the URI for the ProvidedCHO:
<Namespace><class of thing described>/<local unique ID>
Inventory no./Omeka no.

Triple stores and SPARQL
Apache Jena Fuseki/practice querying
Apache Marmotta/visualization
Querying RDF/link data with SPARQL
*Certificate in XML and RDF-Based Systems (Library Juice Academy)
Tools--RDF Skeleton

RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://omeka.library.appstate.edu/rdf

Available Prefixes:
- foaf
- dcterms
- rdfs
- rdf
- dc
- add rdf:type
- add property
- add rdfs:label
- add rdfs:seeAlso

RDF Skeleton

Add another root node

OK Cancel

Base URI: http://www.wcu.edu/library:edit

Available Prefixes:
- foaf
- dcterms
- rdfs
- rdf
- dc
- add rdf:type
- add property
- add rdfs:label
- add rdfs:seeAlso

RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://www.wcu.edu/library:edit

Available Prefixes:
- foaf
- dcterms
- rdfs
- rdf
- dc
- add rdf:type
- add property
- add rdfs:label
- add rdfs:seeAlso

RDF Skeleton

Add another root node

OK Cancel
The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://omeka.library.appstate.edu/edit

This is a sample turtle representation of (up-to) the first 10 rows:

```turtle
@prefix dcterms: <http://purl.org/dc/dcmitype> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix vocab: <http://vocab.getty.edu/aat/> .
@prefix vocab:getty: <http://vocab.getty.edu/aat/> .
@prefix vocab:mechanism: <http://vocab.getty.edu/aat/> .
@prefix vocab:roles: <http://vocab.getty.edu/aat/> .
@prefix vocab:subject: <http://vocab.getty.edu/aat/> .

<<http://vocals.library.appstate.edu/ProvidedCH/930803> a vocab:ProvidedCH ;
  vocab:title "Quilt top: Brick pattern" ;
  vocab:creator "Buchanan, Bettie Hughes" ;
  vocab:subject "1900-1949" ;
  vocab:medium "Fabric, cotton" ;
  vocab:dimensions "63 cm x 73 cm" ;
  vocab:isPartOf "University Library Art Collection" ;
  vocab:isPartOf "2000.0414.04" ;".

<<http://vocals.library.appstate.edu/ProvidedCH/930802> a vocab:ProvidedCH ;
  vocab:title "Quilt top: Brick pattern" ;
  vocab:creator "Buchanan, Bettie Hughes" ;
  vocab:subject "1900-1949" ;
  vocab:medium "Fabric, cotton" ;
  vocab:dimensions "63 cm x 73 cm" ;
  vocab:isPartOf "University Library Art Collection" ;
  vocab:isPartOf "2000.0414.04" ;".

<<http://vocals.library.appstate.edu/ProvidedCH/830130> a vocab:ProvidedCH ;
  vocab:title "Quilt top: Brick pattern" ;
  vocab:creator "Buchanan, Bettie Hughes" ;
  vocab:subject "1900-1949" ;
  vocab:medium "Fabric, cotton" ;
  vocab:dimensions "63 cm x 73 cm" ;
  vocab:isPartOf "University Library Art Collection" ;
  vocab:isPartOf "2000.0414.04" ;".

<<http://vocals.library.appstate.edu/ProvidedCH/830129> a vocab:ProvidedCH ;
  vocab:title "Quilt top: Brick pattern" ;
  vocab:creator "Buchanan, Bettie Hughes" ;
  vocab:subject "1900-1949" ;
  vocab:medium "Fabric, cotton" ;
  vocab:dimensions "63 cm x 73 cm" ;
  vocab:isPartOf "University Library Art Collection" ;
  vocab:isPartOf "2000.0414.04" ;".

<<http://vocals.library.appstate.edu/ProvidedCH/830128> a vocab:ProvidedCH ;
  vocab:title "Quilt top: Brick pattern" ;
  vocab:creator "Buchanan, Bettie Hughes" ;
  vocab:subject "1900-1949" ;
  vocab:medium "Fabric, cotton" ;
  vocab:dimensions "63 cm x 73 cm" ;
  vocab:isPartOf "University Library Art Collection" ;
  vocab:isPartOf "2000.0414.04" ;".
```
Querying the data

SELECT
    ?w WHERE {
      wr:write {
        wd:dtl:ttlo; wr:dtl:titl;
        skd:descrip Description;
        adj:emisitn "Erl";
        ?w FILTER regex(Description, "bird", "i")
      }
    }

Table

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

Search: [Show $10 entries]

Citation

Appalachian State University, Art Department, Quilting Competition, 1970s, Photo 4

Description

This image shows Assistant Professor Judy Lumpre of the Art Department standing in front of a quilt, part of a quilting competition by the Art Department at Appalachian State University in 1970s. The quilt in the background includes images of birds and butterflies. This Art Department was one of the first departments of Appalachian State, and it is located in the College of Fine and Applied Arts. It includes concentrations in clay, drawing, fibers, metal, jewelry, painting, photography, printmaking and sculpture.

Subject

Art

Collection

Appalachian State University Historical Photos

URL: http://omeka.library.appalach.edu/items/animal/39694

Quilt: Bird Cage Design

URL: http://www.sos.ca.gov/digitalcollections/04450/04450.html

Description

"A Patchwork Quilt Identified on the verso (back) of the photograph as "Bird Cage Design". The quilt may have been sold through the Allentown Candy Shop in the early 1900s. Product literature indicates that customers could order a custom-made double bed quilt for $75.00. The photograph was originally available as an example of a quilt design available through the shop. It is shown as one of the quilts which were sold through Allentown during this time period. The quilter and subject of the photograph are unknown."
Current Project Status

Apache Marmotta

An Open Platform for Linked Data

The goal of Apache Marmotta is to provide an open implementation of a Linked Data Platform that can be used, extended, and deployed easily by organizations who want to publish Linked Data or build custom applications on Linked Data.

You can find more information about the project and the supported features on [http://marmotta.apache.org](http://marmotta.apache.org).

Links to common features

You can find all installed features on the module list on the right side. To get a quick access to common functionalities, we listed some links:

- **Import your data**: RDF and non-RDF formats are supported.
- **SPARQL your data**: full SPARQL 1.1 support including querying and updates.
- **Configure your database**: Marmotta comes with h2 embedded; configure your own database to handle bigger data.
- **Control your data**: the dataview gives an overview on the current data in the system. (Attention: there might be problems with the visualisation if big data sets.)
Data Views generated by Apache Marmotta

- **Graph of the “graphs”**
- **Types of classes**

**“Contexts”** ("graphs" in SPARQL):
The triple containers with their own URI

---

### Graphical representation of data:

- **Label**
- **Context**
- **Size**
- **Download**

<table>
<thead>
<tr>
<th>Label</th>
<th>Context</th>
<th>Size</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td><a href="http://linkeddata.library.appstate.edu:3030/marmotta/context/default">http://linkeddata.library.appstate.edu:3030/marmotta/context/default</a></td>
<td>10751</td>
<td>rdf+xml, turtle, ld+json, delete</td>
</tr>
<tr>
<td>W3C Linked Data Platform (LDP)</td>
<td><a href="http://www.w3.org/ns/ldp#">http://www.w3.org/ns/ldp#</a></td>
<td>8</td>
<td>rdf+xml, turtle, ld+json, delete</td>
</tr>
<tr>
<td>cache</td>
<td><a href="http://linkeddata.library.appstate.edu:3030/marmotta/context/cache">http://linkeddata.library.appstate.edu:3030/marmotta/context/cache</a></td>
<td>18</td>
<td>rdf+xml, turtle, ld+json, delete</td>
</tr>
</tbody>
</table>
Visualize SPARQL Query Results

This page allows you to enter custom SPARQL queries and visualize their result. Evaluate query visualizations here and then build your own custom Marmotta + Query.

```
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX usps: <http://www.w3.org/2000/10/swamp/pim/usps#>

SELECT DISTINCT ?lat ?long ?cityName
WHERE
{
    ?object geo:lat ?lat .
    ?object usps:cityName ?cityName .
}
```

Width: 800  Height: 400  Chart Type: Text

47.760176, -122.190953, Bothell
48.765438, -122.509527, Bellingham
47.971892, -118.971025, Coulee Dam
47.758933, -122.190670, Bothell
48.145159, -122.468436, Camano Island
48.765443, -122.511210, Bellingham
45.592697, -122.401628, Camas
46.270212, -122.892241, Castle Rock
47.626660, -122.665226, Bremerton

APACHE
MARMOTTA

login
Hope versus reality: Will we be able to do this with Apache Marmotta...?

Manual visualization via Darrin J. Ward's Mapping Tool

https://www.darrinward.com/lat-long/

![Map with markers indicating locations with coordinates and comments]
Challenges Overall

Time gaps & work loads
Learning curves & training
Metadata cleanup
Moving landscapes
Institutional growth angst
Challenges with systems
Benefits

- Theoretical to practical
- Open source technologies
- Collaboration and connections
- Groundwork for future projects
Next Step(s)

PHASE TWO
Thank you!

Tom Bennett, Operations & Systems Analyst: Apache Marmotta setup
Scott Goldstein, Web Librarian: giving SPARQL queries for visualization a try