Appalachian STATE UNIVERSITY
BOONE, NORTH CAROLINA

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 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

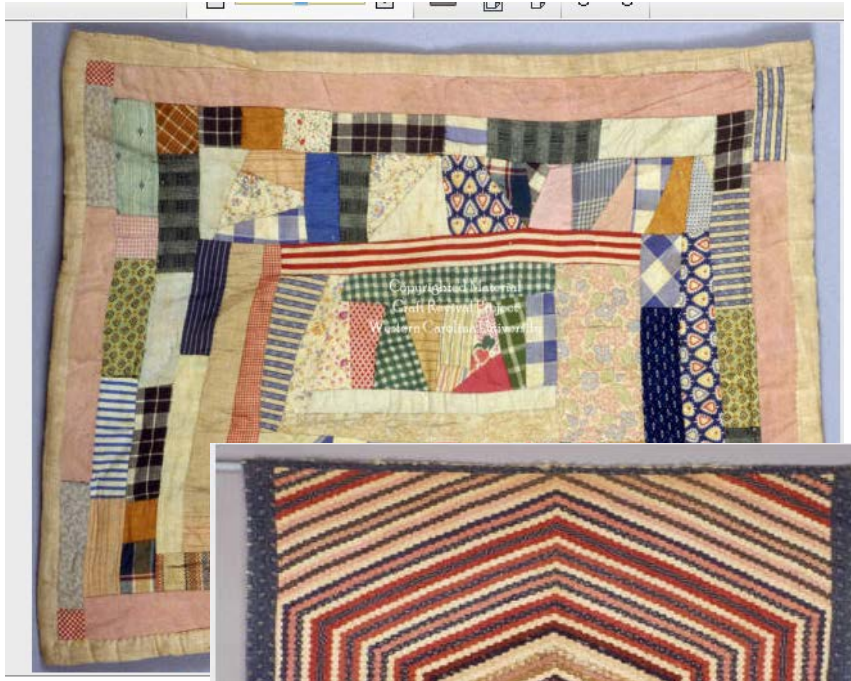
Belk Library & Information Commons, ASU

Choosing the dataset: WCU

Telling the story of an exciting regional movement
Creating handicrafts and preserving traditions from the 1890s to 1940s

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

See:
*Silvia B. Southwick, A Guide for Transforming Digital Collections Metadata into Linked Data, *Journal of Library Metadata* (2015)
*Heath and Bizer, *How to Publish Linked Data on the Web*
*Sauermann et al. *Cool URIs for the Semantic Web* (2006)

Mapping

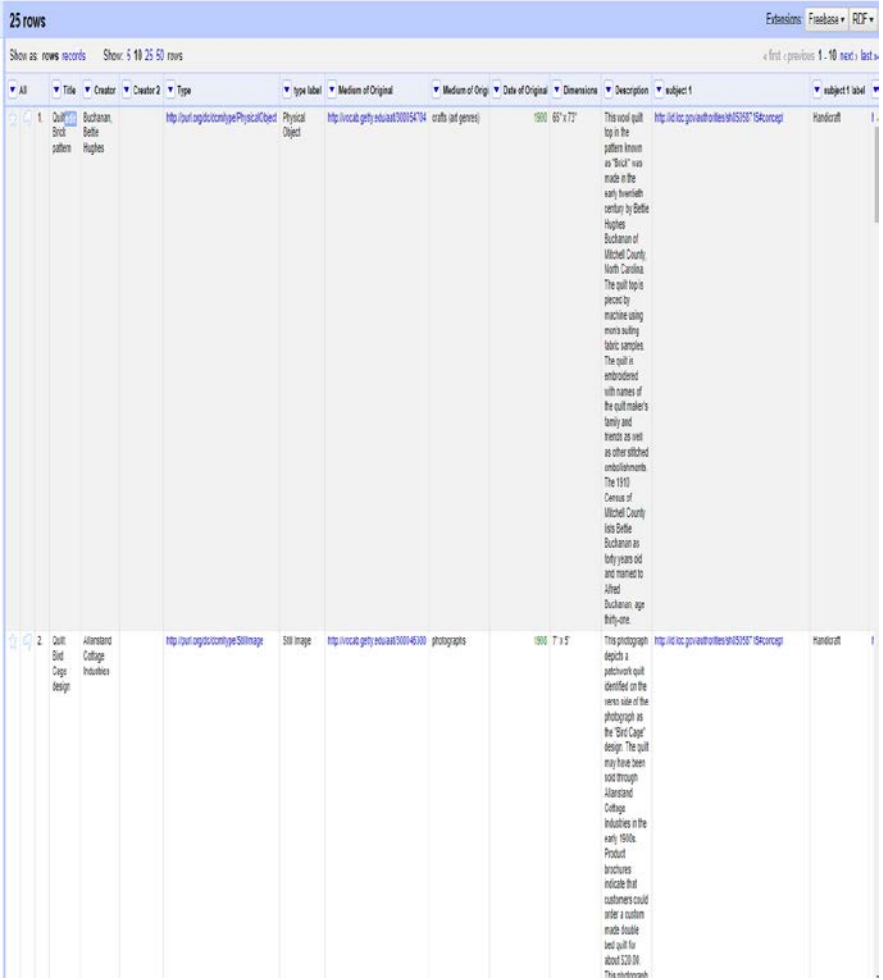
Metadata elements	EDM Predicate	Type of object
Title	dc: title	Literal
Creator	dc:creator	URI/Literal
Type	edm:type	URI
Medium of Original	dcterms:medium	URI
Date of Original	dc:date	Literal
Dimensions	dcterms:extent	Literal
Description	dc:description	Literal
Subject - Topic	dc:subject	URI
Subject -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

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/linked 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
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

RDF Skeleton | RDF Preview

Available Prefixes: rdf owl xsd dcterms rdfs dbpedia foaf edm dc add prefix manage prefixes

- Omeka Number URI
  - edm:ProvidedCHO
  - add rdf:type
  - dc:title → Title cell
  - dcterms:isPartOf → Collection cell
  - dc:description → Description-Text cell
  - dc:subject → Subject 1 URI (add rdf:type)
  - dc:subject → Subject 1 URI (add rdf:type) → rdfs:label → Subject 1 label cel (add property)
  - dc:subject → Subject 2 cell
  - dc:subject → Subject 2 URI (add rdf:type) → rdfs:label → Subject 2 label cel (add property)

Add another root node

OK Cancel

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

RDF Skeleton | RDF Preview

Available Prefixes: dbo dbp rdf owl xsd dcterms rdfs dbpedia foaf edm dc add prefix manage prefixes

- Inventory Number URI
  - edm:ProvidedCHO
  - add rdf:type
  - dc:title → Title cell
  - dc:creator → Creator cell
  - dc:creator → Creator 2 cell
  - dc:type → Type URI (add rdf:type)
  - dc:type → Type URI (add rdf:type) → rdfs:label → type label cell (add property)
  - dcterms:medium → Medium of Original URI (add rdf:type)
  - dcterms:medium → Medium of Original URI → rdfs:label → Medium of Original Lab

Add another root node Save

OK Cancel

RDF Preview (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://www.wcu.edu/library edit

RDF Skeleton | RDF Preview

This is a sample Turtle representation of (up-to) the *first 10* rows

```
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix dbp: <http://dbpedia.org/property/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix edm: <http://www.europeana.eu/schemas/edm/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .


<http://www.wcu.edu/ProvidedCHO/MHC_80_23_002> a edm:ProvidedCHO ;
        dc:title "Quilt top: Brick pattern" ;
        dc:creator "Buchanan, Bettie Hughes" ;
        dc:type <http://purl.org/dc/dcmitype/PhysicalObject> .

<http://purl.org/dc/dcmitype/PhysicalObject> rdfs:label "Physical Object" .

<http://www.wcu.edu/ProvidedCHO/MHC_80_23_002> dcterms:medium <http://vocab.getty.edu/aat/300054704> .

<http://vocab.getty.edu/aat/300054704> rdfs:label "crafts (art genres)" .

<http://www.wcu.edu/ProvidedCHO/MHC_80_23_002> dc:date "1900"^^xsd:date ;
        dcterms:extent "65\" x 73\"" ;
        dc:description "This wool quilt top in the pattern known as \"Brick\" was made in the early twentieth century by Bettie Hughes Buchar
        dc:subject <http://id.loc.gov/authorities/sh85058715#concept> .
```

OK Cancel

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 edit

RDF Skeleton | RDF Preview

This is a sample Turtle representation of (up-to) the *first 10* rows

```
<http://omeka.library.appstate.eduProvidedCHO/366> a edm:ProvidedCHO ;
        dc:title "Nine Patch" ;
        dcterms:isPartOf "University Library Art Collection" ;
        dc:description "Jerra Unglesbee has been quilting for 20 years having taught herself to quilt when she moved to Watauga County.  She
        dc:subject <http://id.loc.gov/authorities/sh85058715#concept> , "http://id.loc.gov/authorities/subjects/sh85109863#concept" , <http:/
        dc:creator "Unglesbee, Jerra " ;
        dc:date "2009"^^xsd:date ;
        dc:rights "Copyright for the images in the University Library Art Collection site is held by Appalachian State University. The images
        dcterms:medium <http://vocab.getty.edu/aat/300014067> ;
        dc:type <http://purl.org/dc/dcmitype/PhysicalObject> ;
        dcterms:extent "55\"x75\"" ;
        edm:isShownAt <http://omeka.library.appstate.edu/items/show/366> .

<http://omeka.library.appstate.eduProvidedCHO/309> a edm:ProvidedCHO ;
        dc:title "Blue Trees" ;
        dcterms:isPartOf "University Library Art Collection" ;
        dc:description "Marni Goldshlag made her first quilt in 1975, but didn't become seriously interested in the quilt as an art form unti
        dc:subject <http://id.loc.gov/authorities/sh85058715#concept> , "http://id.loc.gov/authorities/subjects/sh85109863#concept" , <http:/
        dc:creator "Goldshlag, Marni " ;
        dc:date "2006"^^xsd:date ;
        dc:rights "Copyright for the images in the University Library Art Collection site is held by Appalachian State University. The images
        dcterms:medium <http://vocab.getty.edu/aat/300386847> .

<http://vocab.getty.edu/aat/300386847> rdfs:label "fabric art (visual works)" .

<http://omeka.library.appstate.eduProvidedCHO/309> dc:type <http://purl.org/dc/dcmitype/PhysicalObject> ;
        dcterms:extent "34x27" ;
        edm:isShownAt <http://omeka.library.appstate.edu/items/show/309> .
```

OK Cancel

Fuseki

Apache Jena Fuseki | dataset | manage datasets | help | Server status:

Dataset: /cmsdata

query | upload files | edit | info

Upload files

Load data into the default graph of the currently selected dataset, or the given named graph. You may upload any RDF format, such as Turtle, RDF/XML or TRiG.

Destination graph name Leave blank for default graph

Files to upload + select files... | upload all

ASU_LATEST_metadata_TCDL-xlsx-xls.ttl *39.4kb*

Result: **success**. 304 triples

WCU-Data_TCDL-presentation_1-xlsx-xls.ttl *49.2kb*

Result: **success**. 502 triples

Apache Jena Fuseki | dataset | manage datasets | help | Server status:

Dataset: /cmsdata

query | upload files | edit | info

Available graphs | list current graphs

default graph (801 triples)

graph: default

```
97      dc:subject        "http://id.loc.gov/authorities/sh85109863#concept" ;
98      dc:subject        <http://id.loc.gov/authorities/sh85058715#concept> , <http://i
99      dc:title          "Jacob's Ladder/Underground Railroad" ;
100     dc:type           <http://purl.org/dc/dcmitype/PhysicalObject> ;
101     dcterms:extent    "59\"x76\"" ;
102     dcterms:isPartOf  "University Library Art Collection" ;
103     dcterms:medium    <http://vocab.getty.edu/aat/300014067> ;
104     edm:isShownAt     <http://omeka.library.appstate.edu/items/show/367> .
105
106 <http://omeka.library.appstate.eduProvidedCHO/365.0>
107     a                 edm:ProvidedCHO ;
108     dc:creator        "Deerfield Methodist Church" ;
109     dc:date           "2009.0"^^xsd:date ;
110     dc:description    "The women of Deerfield United Methodist Church in Boone, North
111     dc:rights         "Copyright for the images in the University Library Art Collect
112     dc:subject        "http://id.loc.gov/authorities/sh85109863#concept" ;
113     dc:subject        <http://id.loc.gov/authorities/sh85058715#concept> , <http://i
114     dc:title          "Grandmother's Fan" ;
115     dc:type           <http://purl.org/dc/dcmitype/PhysicalObject> ;
116     dcterms:extent    "57\"x75\"" ;
117
```

discard changes | save

Querying the data

```
SELECT *

WHERE {
  ?s dc:title ?title;
 dc:description ?description;
     edm:isShownAt ?url .
   FILTER regex(?description, "bird", "i")
}
```

QUERY RESULTS

Table | Raw Response

Showing 1 to 2 of 2 entries

Search: Show 50 entries

	s	title	description	url
1	<http://omeka.library.appstate.eduProvidedCHO/9694>	"Appalachian State University, Art Department, Quilting Competition, 1970s, Photo 4"	"This image shows Assistant Professor Judy Humphrey of the Art Department standing in front of a quilt, part of a quilting competition by the Art Department at Appalachian State University (1967-current) during the 1970s. The quilt in the background includes images of birds and butterflies. The Art Department was one of the first departments at Appalachian State and is located in the College of Fine and Applied Arts. It includes concentrations in clay, drawing, fibers, metals/jewelry, painting, photography, printmaking and sculpture."	<http://omeka.library.appstate.edu/items/show/9694>
2	<http://www.wcu.edu/ProvidedCHO/SHCG_FLG3_7_4>	"Quilt: Bird Cage design"	"This photograph depicts a patchwork quilt identified on the verso side of the photograph as the "Bird Cage" design. The quilt may have been sold through Allanstand Cottage Industries in the early 1900s. Product brochures indicate that customers could order a custom made double bed quilt for about $20.00. This photograph was probably used as an example of a quilt design available through the shop. Not much is know about the quilts which were sold through the Allanstand shop, but this appears to be a pattern available from the Mountain Cabin Quilters of Jackson County who sold quilts through Allanstand during this time period. The quilt maker and photographer are unknown."	<http://wcudigitalcollection.contentdm.oclc.org/cdm/singleitem/collection/p4008coll2/id/1211/rec/16>

Digital Collections
Appalachian State University

Home | Collections | About | Exhibits

Files

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

Description

This image shows Assistant Professor Judy Humphrey of the Art Department standing in front of a quilt, part of a quilting competition by the Art Department at Appalachian State University (1967-current) during the 1970s. The quilt in the background includes images of birds and butterflies. The Art Department was one of the first departments at Appalachian State and is located in the College of Fine and Applied Arts. It includes concentrations in clay, drawing, fibers, metals/jewelry, painting, photography, printmaking and sculpture.

Collection

Appalachian State University Historical Photos

Subject

Art

Citation

"Appalachian State Uni
Department, Quilting C
Photo 4," *Appalachian*
Digital Collections, acce
2018,
http://omeka.library.app
ow/9694.

Request a copy

Western Carolina University — Hunter Library Digital Collections

Digital Initiatives Home | Browse Collection | Hunter Library

Search | Advanced Search

Home » Craft Revival » Quilt: Bird Cage design

Add tags | Comment

Quilt: Bird Cage design

View Description

Download

Description

Title: Quilt: Bird Cage design

Graphs/Relationships

Graph of the data model

https://www.w3.org/RDF/Validator/

Current Project Status

login

APACHE MARMOTTA ™

Query and Update
- SPARQL
- LDPath
- LDP

Generic
- Storage Backend: KiWi
- Core Services
- Security
- Users

Others
- Linked Data Caching
- Reasoner
- Versioning

About | Configuration | Logging | Tasks | Import | Export | Data Views | Context Manager | Prefix Manager | System | Webservice

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.

Links to common features

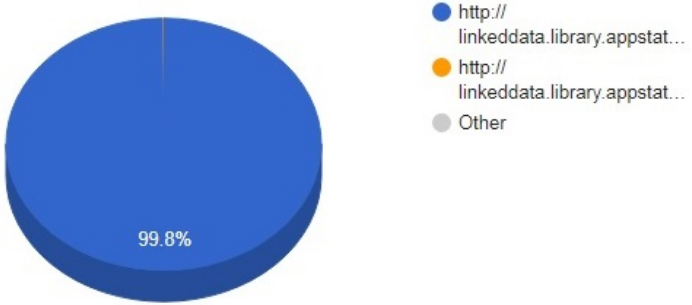
You can find all installed features on the module list on the right side. To get a quick access to common functionalities, we listed some links:

- 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

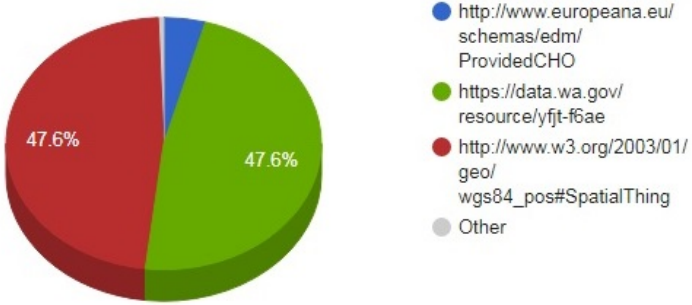
Graphs

All Graphs in the system. Click to select!



Label	Context	Size	Download	
default	http://linkeddata.library.appstate.edu:8080/marmotta/context/default	10751 triples	rdf+xml turtle ld+json	delete
W3C Linked Data Platform (LDP)	http://www.w3.org/ns/ldp#	8 triples	rdf+xml turtle ld+json	delete
cache	http://linkeddata.library.appstate.edu:8080/marmotta/context/cache	18 triples	rdf+xml turtle ld+json	delete

All classes in the System. Click to select!



Graphical representation of data:

- Graph of the “graphs”
- Types of classes

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

Visualize SPARQL Query Results

This page allows you to enter custom SPARQL queries and visualize their resu
evaluate query visualisations 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/swap/pim/usps#>

SELECT DISTINCT ?lat ?long ?cityName

WHERE

  {

    ?object geo:lat ?lat .

    ?object geo:long ?long.

  ?object usps:cityName ?cityName.}
```

Width: 800 Height: 400 Chart Type: Text

Query

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

login

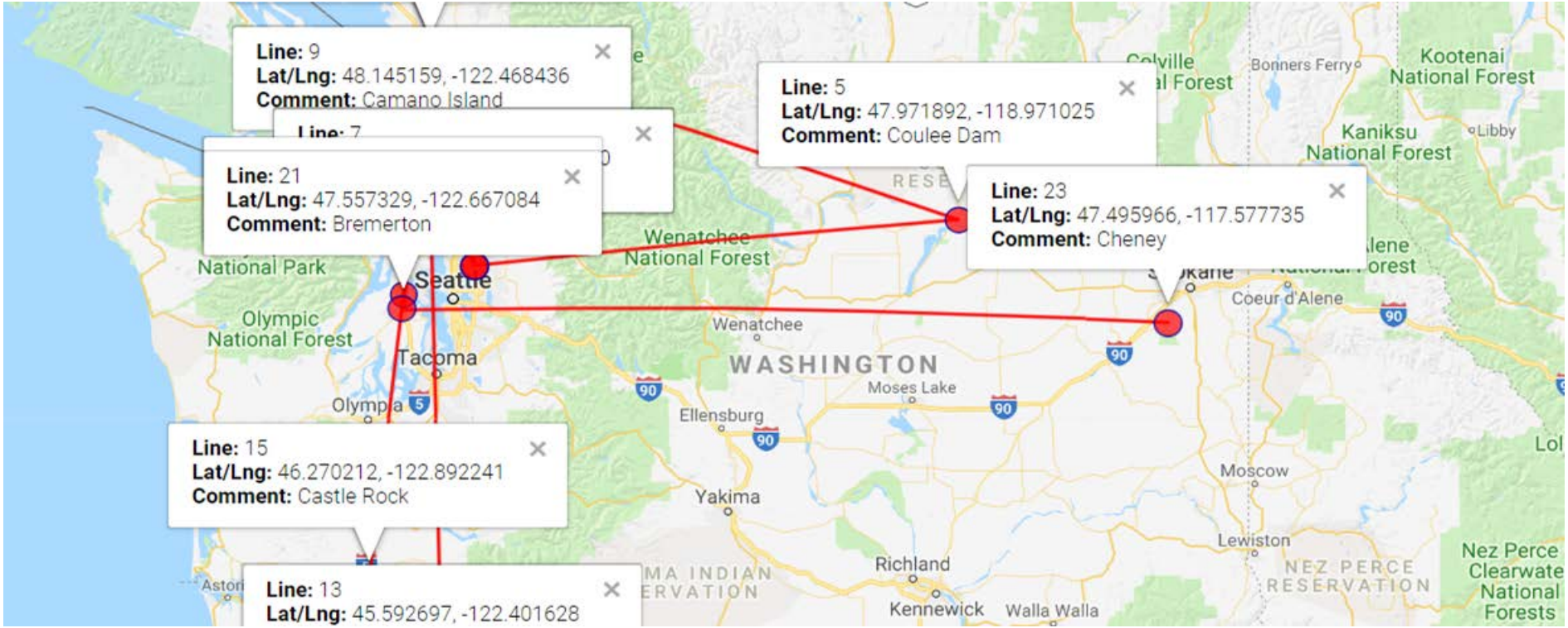
APACHE MARMOTTA™

Hope versus reality:

Will we be able do with this with Apache Marmotta…?

Manual visualization via Darrin J. Ward's Mapping Tool

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



Challenges Overall

Time gaps & work loads

Learning curves & training

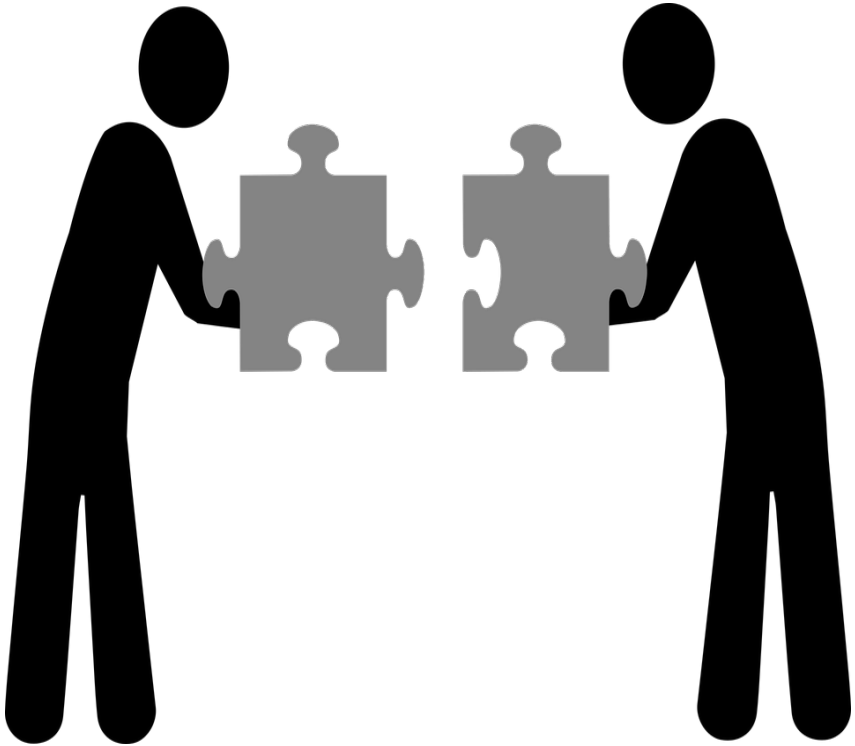
Metadata cleanup

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