Practical Work in Linked Data using Digital Collections:
Unleashing the Expressivity of Data

CNI Project Briefing
April 1, 2014

Silvia B. Southwick
Digital Collections Metadata Librarian
UNLV Libraries
silvia.southwick@unlv.edu

Cory K. Lampert
Head, Digital Collections
UNLV Libraries
cory.lampert@unlv.edu
Agenda

• Motivation
• Environment
• UNLV Linked Data project
• Technologies used for transforming metadata into linked data
• Visualizations of linked data (demos)
• Next steps and Q & A
What is Linked Data?

- Linked Data refers to a set of best practices for publishing and interlinking data on the Web.

- Linked data (Web of Data) is an expansion of the Web we know (Web of documents).
- Data needs to be machine-readable.
- Example: [http://5stardata.info/](http://5stardata.info/)
How we Started

• Conferences and “buzz”
• Curiosity and professional development
• Exploration and pilot project
• Compelling results; sharing impact of what we’ve learned
• Assessment
• Much more to do...
Current Practice

• Information encapsulated in records
• Records contained in collections
• Very few links are created within and/or across collections
• Links have to be manually created
• Existing links do not specify the nature of the relationships among records

This structure hides potential context (links) within and across collections
What we can do with linked data

• Free metadata from silos
• Expose rich relationships
• Leverage powerful, seamless, interlinking of data from multiple sources
• Discover and query data in new ways
• More precise searching
• More opportunities to repurpose data
Making the Case for Linked Data in Academic Library Digital Collections

Problem:

– Rich metadata is being lost when adopting a standard that is designed for interoperability (Dublin Core)

– Rationale for adopting linked data is being disseminated, but there is very little practical implementation to serve as reference; no “recipe” or uniform solution

– Evolving beyond records takes resources and requires embracing an exciting but uncertain future
Graphical Representation: Part of a Record

- Entertainer
- Frank Sinatra: depicts, has creator
- Las Vegas News Bureau
- Theatrical producer
- Jack Entratter: depicts, has genre
- Photographic print
- has profession
Examples of records

Showgirls

Menus

Dreaming the Skyline
UNLV Linked Data Project

Goals:

• Study the feasibility of developing a common process that would allow the conversion of our collection records into linked data preserving their original expressivity and richness

• Publish data from our collections in the Linked Open Data Cloud to improve discoverability and connections across our collections and with data from other related data sets on the Web
PROJECT IMPLEMENTATION
Actions

Phase 1
- Clean data
- Export data

Phase 2
- Import data
- Prepare data
- Reconcile
- Generate triples
- Export RDF

Phase 3
- Import data
- Publish

Technologies

CONTENTdm

Open Refine

Mulgara / Virtuoso
Phase 1

• Clean data
• Export data
Clean / Export Data

Technology: CONTENTdm

• Increase consistency across collections:
  – metadata element labels
  – use of well-known CVs
  – share local CVs
  – etc.

• Export data as spreadsheet
Phase 2

- Import data
- Prepare data
- Reconcile
- Generate triples
- Export RDF
OpenRefine

• Open source

• It is a server – can communicate with other datasets via http

• Open Refine and its RDF extension should be installed

Screenshots to show some of the functions we have used
Import

Google Refine - Mozilla Firefox

Create Project
Open Project
Import Project

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.

Locate one or more files on your computer to upload:

This Computer
Web Addresses (URLs)
Clipboard
Google Data

Get data from

Browse... No files selected

Next »
<table>
<thead>
<tr>
<th>Source</th>
<th>Original Collection</th>
<th>Date</th>
<th>Site Name</th>
<th>Graphic Elements</th>
<th>Collection Subject</th>
<th>DC Type</th>
<th>Genre (TGM)</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-19</td>
<td>Las Vegas Show Costume Designs Collection;</td>
<td>1945; 1946; 1947; 1948; 1949; 1950; 1951; 1952; 1953; 1954; 1955</td>
<td>Facet, Text filter</td>
<td>Still Image</td>
<td>Costume design drawings</td>
<td>eng</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Edit cells**: Transform, Common transforms, Fill down
- **Sort**: Blank down
- **View**: Split multi-valued cells, Join multi-valued cells, Cluster and edit
- **Reconcile**:
Facet view for Graphic Elements after splitting

<table>
<thead>
<tr>
<th>Graphic Elements (TGM)</th>
<th>change</th>
</tr>
</thead>
<tbody>
<tr>
<td>112 choices</td>
<td></td>
</tr>
<tr>
<td>Sort by: name count</td>
<td></td>
</tr>
</tbody>
</table>

- Advertising 1
- Aerialists 1
- Airplanes 1
- Animals 2
- Aprons 2
- Audiences 34
- Automobiles 1
- Ball gowns 10
- Ballet 2
- Banjos 1
- Bathing suits 7
- Beads 39
- Beards 1
- Belts (Clothing) 13
- Berets 1
- Bikinis (Bathing suits) 7
- Billboards 1
- Blazers 3
- Bloomers 1
- Blouses 2
- Bonnets 1
- Boots 18
- Boudoirs 1
<table>
<thead>
<tr>
<th>ID</th>
<th>Digital ID</th>
<th>Title</th>
<th>Individual Creator</th>
<th>Group Creator</th>
<th>Description</th>
<th>Costume Details</th>
<th>Name of Shoe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sho0000119</td>
<td>Costume design drawing, yellow calypso costume, circa 1945-55</td>
<td></td>
<td></td>
<td></td>
<td>Pencil sketch on tracing paper of female dancer in samba costume, with notations of colors and fabrics.</td>
<td>Samba</td>
</tr>
<tr>
<td>2</td>
<td>sho0000114</td>
<td>Costume design drawing, samba costume pencil sketch, circa 1945-55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reconciliation
### Specifying Reconciliation service

#### Add SPARQL-based reconciliation service

- **Name:** A human readable name

**Endpoint details**
- **Endpoint URL:**
- **Graph URI:** Leave empty to use the default graph
- **Type:**
  - Generic SPARQL (poor performance)
  - This determines the syntax that will be used for search

**Label properties**

Select properties that are used to label resources in the endpoint. These properties will be used to match resources:

- rdfs:label
- skos:prefLabel
- dcterms:title
- dc:title
- foaf:name
- Other...

[OK | Cancel]
### Activating Reconciliation

<table>
<thead>
<tr>
<th>Site name URI</th>
<th>Graphic Element</th>
<th>Graphic URI</th>
<th>Collection Subj</th>
<th>DC Type</th>
<th>Genre (TGM)</th>
<th>Genre URI</th>
<th>Lan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facet</td>
<td></td>
<td><a href="http://id.loc.gov/vocabulary/graphicMaterials/m002610">http://id.loc.gov/vocabulary/graphicMaterials/m002610</a></td>
<td>Still image</td>
<td>Costume design drawings</td>
<td>Choose new match</td>
<td><a href="http://id.loc.gov/vocabulary/graphicMaterials/tgm002607">http://id.loc.gov/vocabulary/graphicMaterials/tgm002607</a></td>
<td>eng</td>
</tr>
</tbody>
</table>

**Steps to Start Reconciliation:**

1. Click on the 'Reconcile' drop-down menu.
2. Select 'Start reconciling...'

Additional options include:
- Facets
- QA facets
- Actions
- Copy reconciliation data...
- Discover related RDF datasets.
<table>
<thead>
<tr>
<th>Site name URL</th>
<th>Graphic Elements</th>
<th>Graphic URI</th>
<th>Collection Subj</th>
<th>DC Type</th>
<th>Genre (TGM)</th>
<th>Genre URI</th>
<th>Language</th>
<th>Is Part Of</th>
<th>Rights</th>
</tr>
</thead>
</table>

This material is protected by copyright. Personal, non-commercial use of this material without prior permission is permitted. For further information please contact Collections at http://digital.library.unlv.edu/contact.
Europeana Data Model

Based on the EDM documentation at http://pro.europeana.eu/edm-documentation

Legend:
* not implemented by UNLV -- Aggregation class is under consideration
gray background -- not yet implemented by Europeana
blue font -- properties pertaining to the edm vocabulary
Creating a 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://digiloc7.library.unlv.edu:8860/ edit

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

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edm: <http://www.europaeana.eu/schemas/edm/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix dcterms: <http://purl.org/dc/terms/> .

<http://digiloc7.library.unlv.edu:8850/ProvidedCNO/SHO000119> a edm:ProvidedCNO ;
  dc:identifier "112233445566" ;
  dc:description "Sketch of female dancer in yellow calypso costume with skirt with long train, ruffled short sleeves, .
  dc:800106 20070606162239 ;
  edm:source "Digital Library Collection II, Presentation" ;
  dc:language "en" ;
  dcterms:identifier "en" ;
  edm:hasType <http://id.loc.gov/vocabulary/graphicMaterials/tgm002610> ;
  skos:Concept .

  dc:description "Still Image" ;

<http://id.loc.gov/vocabulary/graphicMaterials/tgm002607> a skos:Concept .
```

---

---
### Exporting RDF files

#### Table: RDF Export Options

<table>
<thead>
<tr>
<th>Export Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF as RDF/XML</td>
<td>Converts the data into RDF/XML format.</td>
</tr>
<tr>
<td>RDF as Turtle</td>
<td>Converts the data into Turtle format.</td>
</tr>
</tbody>
</table>

#### Example Data:

<table>
<thead>
<tr>
<th>Site name URI</th>
<th>Graphic URI</th>
<th>Collection Subj</th>
<th>DC Type</th>
<th>Genre (TGM)</th>
<th>Genre URI</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costumes</td>
<td><a href="http://id.loc.gov/vocabulary/graphicMaterials/gm002610">http://id.loc.gov/vocabulary/graphicMaterials/gm002610</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eng</td>
</tr>
<tr>
<td>Dancers</td>
<td><a href="http://id.loc.gov/vocabulary/graphicMaterials/gm002607">http://id.loc.gov/vocabulary/graphicMaterials/gm002607</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>eng</td>
</tr>
</tbody>
</table>
Phase 3

- Import data
- Publish
- Query
Mulgara Triple Store: Import
<table>
<thead>
<tr>
<th>URIs</th>
<th>URIs</th>
<th>URIs</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://dbpedia.org/resource/Sun_City_South_Africa">http://dbpedia.org/resource/Sun_City_South_Africa</a></td>
<td><a href="http://w3.org/1999/02/22-rdf-syntax-ns#type">http://w3.org/1999/02/22-rdf-syntax-ns#type</a></td>
<td><a href="http://www.europeana.eu/schemas/edm/Place">http://www.europeana.eu/schemas/edm/Place</a></td>
</tr>
</tbody>
</table>

**URIs from:**
- UNLV
- Europeana
- Dbpedia
- Geonames
- LOC
Visualization Open Source Tools

- OpenLink Virtuoso Pivot Viewer
- RelFinder
- Gephi
  - Linked Jazz
OpenLink Pivot Viewer

• Good for displaying images
• Selection of images through SPARQL Queries
• Allows refinements using facets
• Allows creating dynamic “collections”
SPARQL Query

```
describe ?thing
where{
?thing a edm:ProvidedCHO.
?thing edm:hasType ?tmuri.
?tmuri skos:prefLabel "Costume design drawings".
?collection dc:title "Showgirls".
optional {?thing foaf:depiction ?image }
}
```
Example of Pivot Viewer

• Link to the demo:
  C:\Users\Silvia\Desktop\videosCNI\PivotViewer.mp4
Examples of RelFinder

• Good to show relationships:
  – Among people
  – Among “things”

• Show type of relationships

Demos:

– African American Experience in Las Vegas (Oral History):
  C:\Users\Silvia\Desktop\videosCNI\AAE_relationships.mp4

– Cross collections people relationship:
  C:\Users\Silvia\Desktop\videosCNI\Frank_relationships.mp4
Gephi

• Good for showing relationships

• More sophisticated interface design

• We are working toward implementing Gephi for our digital collections
Gephi

Linked Jazz demo:

C:\Users\Silvia\Desktop\videosCNI\LinkedJazz.mp4
Next steps for the UNLV project

• Transform all digital collections into linked data (parallel structure)
• Evaluate alternative interfaces
• Increase linkage with other datasets
• Publish our collections metadata as Linked Open Data
• Design and assess user friendly interfaces to access and display our data and related data from other datasets
• Produce a cost benefit analysis to inform future plans for the development of digital collections
Our Experience

- Project led, implemented and managed by two busy faculty librarians
- No model to follow; our model was experimentation and research
- With interest and motivation, Linked Open Data is a feasible goal
Thank You!

Questions?