Librarians' adventure into LODLAM

LODLAM Training Day
August 19, 2014
San Jose, CA

Silvia Southwick & Cory Lampert
UNLV Digital Collections
Agenda

• How the adventure started
• UNLV Linked Data project
• Technologies used for transforming metadata into linked data
• Visualizations of linked data (demos)
• Next steps and Q & A
Example of linked data visualization

- Pivot Viewer

C:\Users\Silvia\Desktop\videosCNI\PivotViewer.mp4
How we started (2012)

• UNLV Libraries study group
• Literature Review
• Survey of technologies
• Experimentation
• Design of a project
Why? Examples of current records
Making the Case for Linked Data

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

– Evolving beyond records takes resources and requires embracing an exciting but unknown future
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 Data Cloud to improve discoverability and connections with other related data sets on the Web
# Main Concepts / Technologies

**CONCEPTS**
- URIs
- Triples
- RDF
- HTTP Protocol
- Triplestores
- SPARQL

**Technologies**
- Digital Asset Management (CONTENTdm)
- OpenRefine
- Karma (hierarchical data structure)
- Mulgara / OpenLink Virtuoso
- SPARQL end points
Actions

- Prepare data
- Export data

Technologies

- CONTENTdm
- Open Refine
- Mulgara / Virtuoso

Import data
- Clean data
- Reconcile
- Generate triples
- Export RDF

Import data
- Publish
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
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 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.

Get data from
- This Computer
- Web Addresses (URLs)
- Clipboard
- Google Data

Locate one or more files on your computer to upload:
Browse... No files selected

Next »
### Facets

The image shows a screenshot of a database interface with a focus on facets and records. The interface includes various columns such as **Show Collect**, **Date**, **Site Name**, **Graphic Element**, **Collection Subject**, **DC Type**, **Genre (TGM)**, **Language**, and **Is Part Of**. Each record is represented with specific data entries:

- **Costumes**: Dancers, Turbans, Skirts
- **Still Image**: Costume designs
- **Genre (TGM)**: Drawings
- **Language**: eng

The facet list includes categories such as:

- Graphic Elements (TGM)
  - 884 choices
  - Sort by: name, count
  - Airplanes; Chorus girls; 1
  - Animals; Dogs; Pets; Women; 2
  - Aprons; Trousers; Men; Vests; Stripes; 1
  - Audiences; Beads; Capes (Clothing); Costumes; Dancers; Theatrical productions; 1
  - Audiences; Beads; Costumes; Hats; Theatrical productions; 1
  - Audiences; Beads; Costumes; Chaps; Coats; Costumes; Cowboy boots; 1

The interface also includes options for **Facet/Filter**, **Undo/Redo**, and **Show as: rows, records** with options to show 5, 10, 25, or 50 records. The page is part of a project with the URL `127.0.0.1:3333/project?project=1698768213599`.

The interface is part of the Google Refine tool, which is used for data cleaning and transformation tasks.
## Split multi-value cells

### Table

<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 in vlyn collection</td>
<td>Las Vegas Show Costume Designs Collection;</td>
<td>1945; 1946; 1947; 1948; 1949; 1950; 1951; 1952; 1953; 1954; 1955</td>
<td></td>
<td>Facet</td>
<td>Still Image</td>
<td>Costume design drawings</td>
<td>eng</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Text filter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Edit cells</td>
<td>Transform...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Edit column</td>
<td>Common transforms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transpose</td>
<td>Fill down</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sort...</td>
<td>Blank down</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>View</td>
<td>Split multi-valued cells...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Reconcile</td>
<td>Join multi-valued cells...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cluster and edit...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<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 in vlyn collection</td>
<td>Las Vegas Show Costume Designs Collection;</td>
<td>1945; 1946; 1947; 1948; 1949; 1950; 1951; 1952; 1953; 1954; 1955</td>
<td>Costumes; Dancers;</td>
<td>Still Image</td>
<td>Costume design drawings</td>
<td>eng</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Diagram

The diagram illustrates the process of splitting multi-value cells in a dataset. It shows the different steps and tools available for handling multi-valued data, including facet creation, text filtering, cell editing, column transformation, transpose, sort, view, reconcile, join, and cluster and edit functions.
Facet view for Graphic Elements after splitting
Reconciliation

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Costume Details</th>
<th>Name of Sho</th>
</tr>
</thead>
<tbody>
<tr>
<td>sho000119</td>
<td>Costume design drawing, yellow calypso costume, circa 1945-55</td>
<td>Based on SPARQL endpoint... Add reconciliation service</td>
<td></td>
</tr>
<tr>
<td>sho000114</td>
<td>Costume design drawing, samba costume pencil sketch, circa 1945-55</td>
<td>Pencil sketch on tracing paper of female dancer in samba costume, with notations of colors and fabrics.</td>
<td>Samba</td>
</tr>
</tbody>
</table>
## 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></td>
<td>Facet</td>
<td><a href="http://id.loc.gov/vocabulary">http://id.loc.gov/vocabulary</a></td>
<td>Still image</td>
<td>Costume design</td>
<td>drawings</td>
<td><a href="http://id.loc.gov/vocabulary/genre/tgm002507">http://id.loc.gov/vocabulary/genre/tgm002507</a></td>
<td>eng</td>
</tr>
<tr>
<td></td>
<td>Text filter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Edit cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Edit column</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sort...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>View</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reconcile</td>
<td>Start reconciling...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Dancers**

- Choose new match
- Copy reconciliation data...

**Turbans**

- Choose new match
- Discover related RDF datasets...
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
Mapping between Showgirls and EDM
Implementing Mapping (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

<table>
<thead>
<tr>
<th>Available Prefixes:</th>
<th>do rdf dc edm foaf add osw rdf skos dcterms +add prefix +manage prefixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(row index) URI</td>
<td></td>
</tr>
<tr>
<td>x edm:ProvidedCHO</td>
<td></td>
</tr>
<tr>
<td>add rdf.type</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RDF Skeleton</th>
<th>RDF Preview</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="RDF Diagram" /></td>
<td></td>
</tr>
</tbody>
</table>

- **dc:title** → **Title**
- **dc:creator** → **Individual creator** [URI URI]
- **dc:creator** → **Group creator** [URI URI]
- **dc:description** → **Description** [cell]
- **edm.isRelatedTo** → **Name show** [URI URI]
- **edm.happenedAt** → ** edm:Place** [cell]
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://digloc7.library.unlv.edu:8850/ edit

**RDF Skeleton**

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edm: <http://www.europeana.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://digloc7.library.unlv.edu:8850/ProvidedCHO/sho000119> a edm:ProvidedCHO ;
dct:title "Costume design drawing, yellow calvans costume, circa 1945-55" ;
dcm:description "Sketch of female dancer in yellow calvans costume with skirt with long train, ruffled short sleeves, and
```

**RDF Preview**

```
<http://id.loc.gov/vocabulary/graphicMaterials/tgm002607> a skos:Concept .
<http://id.loc.gov/vocabulary/graphicMaterials/tgm002609> a skos:Concept .
```
Exporting RDF files

<table>
<thead>
<tr>
<th>Site name URI</th>
<th>Graphic Exer</th>
<th>Graphic URI</th>
<th>Collection Subj</th>
<th>DC Type</th>
<th>DC URI</th>
<th>DC URI (TGM)</th>
<th>Genre (TGM)</th>
<th>Genre URI</th>
<th>eng</th>
<th>Description</th>
</tr>
</thead>
</table>
**Actions**

- Prepare data
- Export data

- Import data
- Clean data
- Reconcile
- Generate triples
- Export RDF

**Technologies**

- CONTENTdm
- Open Refine
- Mulgara / Virtuoso
Phase 3

- Import data
- Publish
- Query
Mulgara Triple Store: Import
Graph URI: http://showgirls

Query Text:

Submit Query Clear Query

File: Browse... Upload

Results: (1 query, 7.643 seconds)


Simple SPARQL query

```sparql
Select *
Where {?s ?p ?o} limit 100
```
Visualization Open Source Tools

• OpenLink Virtuoso Pivot Viewer

• RelFinder
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" .
optional {{?thing foaf:depiction ?image }}
}
```

Costume Design Drawings
Showgirls
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
Next steps for the UNLV project

• Transform all digital collections into linked data
• Increase linkage with other datasets
• Publish as Linked Open Data
• Design and assess user friendly interfaces
• Work on integrating data from diverse local systems
• 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?

UNLV’s Linked Open Data Blog: http://library.unlv.edu/linked-data

Cory Lampert
Head, Digital Collections
cory.lampert@unlv.edu

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