Exposing Missing Links: From CONTENTdm digital collections to the Linked Open Data cloud

Silvia B. Southwick
University of Nevada, Las Vegas, silvia.southwick@unlv.edu

Follow this and additional works at: http://digitalscholarship.unlv.edu/libfacpresentation

Part of the Cataloging and Metadata Commons

Repository Citation

Available at: http://digitalscholarship.unlv.edu/libfacpresentation/110

This Presentation is brought to you for free and open access by the Library Faculty/Staff Scholarship & Research at Digital Scholarship@UNLV. It has been accepted for inclusion in Presentations (Libraries) by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
Exposing Missing Links

From CONTENTdm digital collections to the Linked Open Data cloud

Best Practices Exchange
November 13-15, 2013

Silvia B. Southwick
Digital Collections Metadata Librarian
Agenda

• Linked Data basic concepts
• UNLV Linked Data project
• Technologies
• Transforming metadata into linked data
• Next steps
Linked Data Overview

• My collections are already visible through Google; so who cares
• This is a topic for catalogers
• It’s too technical / complicated / boring

Actually ...
• Linked data is the future of the Web
• Data will no longer be in trapped in silos imposed by systems, collections, or records
• Exposed open data presents new opportunities for users
What is Linked Data?

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

  - Data needs to be machine-readable

  - Linked data (Web of Data) is an expansion of the Web we know (Web of documents)
Current Practice

• Data (or metadata) 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 links within and across collections
What we can do with linked data

• Free data from silos
• Expose relationships
• Powerful, seamless, interlinking of our data
• Users interact or query data in new ways
• Search results would be more precise
• Data can be easily repurposed
How can we create linked data?

• Our metadata records are deconstructed in triples (statements) that are machine-readable
• Triples are expressed as: **Subject – Predicate - Object**
  
  For example:  
  This book – has creator – Tom Heath  
  This book – has title – Linked Data: Evolving the…”
• **Subjects, predicates** and most **objects** should have unique identifiers (URIs) creating **data** that can be used in Web architecture (HTTP)
• These statements are expressed using the Resource Description Framework (RDF)
• Linked data can be queried using SPARQL
So, what?

• We already have the metadata!

• We need to transform them into triples

• Each metadata field may produce one or several statements

• One metadata record can produce many, many, triples
### Example of a metadata record

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digital ID</strong></td>
<td>sh0000077</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td>Photograph of Frank Sinatra and Jack Entratter at rehearsal, Las Vegas, 1954</td>
</tr>
<tr>
<td><strong>Group Creator</strong></td>
<td>Las Vegas News Bureau</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Jack Entratter and Frank Sinatra watch rehearsals at the Sands Hotel.</td>
</tr>
</tbody>
</table>

| **Name of Show** | Ziegfeld Follies |
| **Identified Individuals** | Sinatra, Frank, Entratter, Jack |
| **Source** | Image Number: 0287 0037 |
| **Original Collection** | Sands Hotel Collection |
| **Date** | 1954 |
| **Site Name** | Sands Hotel and Casino |
| **Graphic Elements (TGM)** | Theatrical producers & directors, Entertainers |
| **DC Type** | Still Image |
| **Genre (TGM)** | Pictures, Photographs, Photographic prints |
| **Language** | eng |
Expressing metadata as triples

• <this thing> <has creator> <Las Vegas News Bureau>
• <this thing> <has genre> <Photographic print>
• <this thing> <depicts> <Frank Sinatra>
• <this thing> <depicts> <Jack Entratter>

• <Frank Sinatra> <has profession> <entertainer>
• <Jack Entratter> <has profession> <theatrical producer>
Examples of records
Showgirls Collection
- published in
  - Photo of Sinatra and Enntratter
    - is a photographic print
      - created by Las Vegas News Bureau

Frank Sinatra
- born in Hoboken, NJ
- has nickname Ol' Blue Eyes
- is an entertainer
  - performed in Ziegfeld Follies
    - presented at Sands Hotel
      - located in Las Vegas

Jack Entratter
- is a theatrical producer
  - knows Frank Sinatra
    - has nickname Ol' Blue Eyes

Martin Stern
- is an architect
  - created Sands architectural drawings
    - published in SKY Collection
      - published in Menu
        - created by Sammy Davis Jr.
          - published in Menu Collection
Graphical Representation

Vocabularies

• Controlled Vocabularies for values (objects) containing term URI
• Vocabularies for predicates (e.g., Dublin Core, foaf, skos, etc.)

Subject URIs

• Re-use existent URIs
• Create URIs for unique “things” or for “things” that do not have yet URIs
How can I transform textual triples into machine-readable?

• We need a **data model**

• **Europeana Data Model** gives us a framework to help organize, structure, and define which predicates we are going to use

• Adopting an existing model is preferable to creating your own (interoperability)
Triples with URIs & EDM model predicates

Entertainer (TGM URI)

Frank Sinatra (LoC URI)

Las Vegas News Bureau (Local URI)

Theatrical producer (TGM URI)

Jack Entratter (Local URI)

Photographic print (TGM URI)

rdaGr2:professionOrOccupation

foaf:depicts

dc:creator

edm:hasType
Machine-readable triple

@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix edm: <http://www.europeana.eu/schemas/edm/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<http://digloc7.library.unlv.edu:8890/ProvidedCHO/sho000071> dc:creator

<http://digloc7.library.unlv.edu:8890/ProvidedCHO/sho000071> foaf:depicts
<http://id.loc.gov/authorities/names/n50026395> .

<http://digloc7.library.unlv.edu:8890/ProvidedCHO/sho000071> edm:hasType
http://id.loc.gov/vocabulary/graphicMaterials/tgm007779 .
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
Actions

- Prepare data
- Export data

Technologies

- CONTENTdm
- Open Refine
- Mulgara / Virtuoso

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

- Import data
- Publish
Prepare / Export Data

Technology: CONTENTdm

• Increase consistency across collections:
  – metadata element labels
  – use of CV, share local CVs
  – etc.

• Export data as spreadsheet

Create mapping between metadata elements and EDM model predicates
OpenRefine

• Open source

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

• Install Open Refine and its RDF extension

Screenshots to show some of the functions we have used
OpenRefine first screen
<table>
<thead>
<tr>
<th>Title</th>
<th>Individual Creator</th>
<th>Description</th>
<th>Costume Detail</th>
<th>Name of Show</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costume design drawing, yellow calypso costume, circa 1945-55</td>
<td></td>
<td>Sketch of female dancer in yellow calypso costume with skirt with long train, ruffled short sleeves, and floral headdress.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costume design drawing, samba costume pencil sketch, circa 1945-55</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></td>
</tr>
<tr>
<td>Title</td>
<td>Description</td>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costume design drawing, yellow calypso costume, circa 1945-55</td>
<td>Sketch of female dancer in yellow calypso costume with skirt with long train, ruffled short sleeves, and floral headdress.</td>
<td>Virgin islands</td>
<td></td>
<td></td>
</tr>
<tr>
<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>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Digital ID</td>
<td>Title</td>
<td>Individual Creator</td>
<td>Group Creator</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>39</td>
<td>sho0000124</td>
<td>Costume design, drawing, feathered turquois...</td>
<td>Mackie, Robert Gordon, 1940; Aghayan, Ray</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>sho0000128</td>
<td>Costume design, drawing, long black showgirl costume with split skirt, Las Vegas, 1974</td>
<td>Mackie, Robert Gordon, 1940; Aghayan, Ray</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Original Collect</td>
<td>Date</td>
<td>Site Name</td>
<td>Graphic Elem</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<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></td>
</tr>
<tr>
<td>2000-19 in vlyn</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></td>
</tr>
</tbody>
</table>

For split multi-value cells, use the 'Split multi-valued cells...' option.
<table>
<thead>
<tr>
<th>Source</th>
<th>Original Collect</th>
<th>Date</th>
<th>Site Name</th>
<th>Graphic Elements</th>
<th>DC Type</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></td>
<td>Costumes</td>
<td></td>
</tr>
<tr>
<td>Frasny, Daniel</td>
<td>Beardsley, Aubrey, 1872-1898; Bernard, Bruno, 1912-1987; de Otley, Murray, 1; English, Don 4</td>
<td>2000-19</td>
<td>Las Vegas Show Costume Designs Collection;</td>
<td>Costumes</td>
<td></td>
</tr>
<tr>
<td>15 choices</td>
<td>Sort by name count</td>
<td>Cluster</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Facet view for Graphic Elements

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

- Advertising 1
- Aerialists 1
- Airplanes 1
- Animals 2
- Aprons 2
- Audiences 34
- Automobiles 1
- Ball dresses 10
- Ballet 2
- Banjos 1
- Bathing suits 7
- Beads 39
- Beards 1
- Belts (Clothing) 13
- Berets 1
- Bikinis (Bathing suits) 7
- Billboards 1
- Blaziers 3
- Bloomers 1
- Blouses 2
- Bonnets 1
- Boots 18
- Boudoirs 1
### Reconciliation

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Description</th>
<th>Costume Detail</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>Pencil sketch on tracing paper of female dancer in samba costume, with notations of colors and fabrics.</td>
<td></td>
<td>Samba</td>
</tr>
<tr>
<td>sho000114</td>
<td>Costume design drawing, samba costume pencil sketch, circa 1945-55</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Specifying Reconciliation service
### Activating Reconciliation

#### 2540 rows

**Show as:** rows records  **Show:** 5 10 25 50 rows

<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>Genre URI</th>
<th>Lan</th>
</tr>
</thead>
</table>

- **Dancers**
  - Choose new match
  - Copy reconciliation data...

- **Turbans**
  - Choose new match
  - Discover related RDF datasets...
### Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts.

### Google refine - LD1 Showgirls

#### 2540 rows

<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>

### Dancers

<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>

### Skirts

<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>

### Costumes

<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 is allowed without restriction. This material is not for educational use. If you are interested in using this material, please contact the University of Nevada, Las Vegas for further information. The use of this material is restricted.
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://digilic7.library.univ.edu/8860/edit

Available Prefixes:
- dc
dc-title
- edm
edmProvidedCHO
- add rdf.type
- add prefix
- manage prefixes

- Title
- Individual creator URI
- Group creator URI
- Description
- Costume Details
- Name show URI
- Site name URI
- edm.happenedAt
- edm.Place
- add rdf.type
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:8890/edit

RDF Skeleton

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

```turtle
@prefix dct: <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 dc: <http://purl.org/dc/elements/1.1/> .

<http://digloc7.library.unlv.edu:8890/ProvidedCHM/eho000019> a edm:ProvidedCHO ;
dct:title "Costume design drawing, yellow calypso costume, circa 1945-55" ;
dct:description "Sketch of female dancer in yellow calypso costume with skirt with long train, ruffled short sleeves, and


dc:type "Still Image" .

<http://digloc7.library.unlv.edu:8890/ProvidedCHM/eho000019> skos:concept .

<http://digloc7.library.unlv.edu:8890/ProvidedCHM/eho000019> edm:hasType <http://id.loc.gov/vocabulary/graphicMaterials/tgm002607> ;
dd:language "eng" ;
dd:rights "This material may be protected by copyright. Personal, including educational and academic, use of this material is permitted provided that proper attribution is given to the University of Nevada, Las Vegas. For other uses you must obtain permission from the University." .
```

RDF Preview
Exporting RDF files
Actions

- Prepare data
- Export data

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

Technologies

- CONTENTdm

- Open Refine

- Mulgara / Virtuoso
Mulgara Triple Store: Import
Mulgara Semantic Store

Graph URI: http://showgirls

Query Text:

Submit Query Clear Query

File: Browse... Upload

Results: 1 query, 7.643 seconds
Query Executed:
load <http://mulgara.org/virtual/M:workshopLD1-Showgirls.ttl> into <http://showgirls>
A simple SPARQL query

Select *

Where {?s ?p ?o} limit 100
<table>
<thead>
<tr>
<th>URL</th>
<th>URL</th>
<th>URL</th>
<th>URL</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://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.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>
<td><a href="http://www.europeana.eu/schemas/edm/Place">http://www.europeana.eu/schemas/edm/Place</a></td>
</tr>
</tbody>
</table>
SPARQL: Querying Data

• Using Virtuoso triple store PivotViewer
Query

Dynamic Collection

SPARQL

Query Service Endpoint
http://digeol7.library.unlv.edu:8890/sparql

Default Data Set Name
(Graph IRI)

Query Text
```
describe ?thing
where{
?thing a edm:ProvidedCHO.
?thing edm:hasType ?tmuri .
?tmuri skos:prefLabel "Costume design drawings" .
optional {?thing foaf:depiction ?image }
}```

Resultset Options

Timeout
Costume design drawing, red ruffled Spanish flamenco costume with rimmed hat and flared cuffs, circa 1965-75

Red ruffled Spanish flamenco pantaloons costume with flared white cuffs, white ruffles on shirt front, and hat with brim and flowers on female dancer.
Next steps for the UNLV project

• Transform digital collections into linked data (parallel structure)
• Increase linkage with other datasets
• Design interfaces to access and display our data and related data from other datasets
• Evaluate alternative designs from user’s perspective
• Produce a cost benefit analysis to inform future plans for the development of digital collections
Thank You!

Questions?