11-13-2013

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: https://digitalscholarship.unlv.edu/libfacpresentation

Part of the Cataloging and Metadata Commons

Repository Citation

Available at: https://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 Library Faculty Presentations 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
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

**Description**

Jack Entratter and Frank Sinatra watch rehearsals at the Sands Hotel.

**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

Menus

Dreaming the Skyline
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

Entertaine (TGM URI) —Frank Sinatra (LoC URI) —Las Vegas News Bureau (Local URI)

foaf:depicts

dc:creator

rdaGr2:professionOroccupation

Jack Entratter (Local URI) —Photographic print (TGM URI)

foaf:depicts

edm:hasType

rdaGr2:professionOroccupation
@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> 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

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

Technologies

- CONTENTdm
- Open Refine
- Mulgara / Virtuoso
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

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

<table>
<thead>
<tr>
<th>Get data from</th>
<th>Locate one or more files on your computer to upload:</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Computer</td>
<td>Browse... No files selected.</td>
</tr>
<tr>
<td>Web Addresses (URLs)</td>
<td></td>
</tr>
<tr>
<td>Clipboard</td>
<td></td>
</tr>
<tr>
<td>Google Data</td>
<td></td>
</tr>
</tbody>
</table>

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.
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 screen casts

<table>
<thead>
<tr>
<th>All</th>
<th>Digital ID</th>
<th>Title</th>
<th>Individual Creator</th>
<th>Group Creator</th>
<th>Description</th>
<th>Costume Detail</th>
<th>Name of Show</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sho000119</td>
<td>Costume design drawing, yellow calypso costume, circa 1945-55</td>
<td></td>
<td></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>
</tr>
<tr>
<td></td>
<td>sho000114</td>
<td>Costume design drawing, samba costume pencil sketch, 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></td>
</tr>
</tbody>
</table>
Facet
<table>
<thead>
<tr>
<th>Cluster</th>
<th>Show years</th>
<th>Site Name</th>
<th>Graphic Elements</th>
<th>Collection Subject</th>
<th>DC Type</th>
<th>Genre (TGM)</th>
<th>Language</th>
<th>Is Part Of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1945; 1946; 1947; 1948; 1950; 1951; 1952; 1953; 1954; 1955</td>
<td>Still Image</td>
<td>Costumes; Dancers; Turbans; Skirts;</td>
<td></td>
<td>Still Image</td>
<td>Costume design drawings</td>
<td>eng</td>
<td><a href="http://digital.library.univ.edu/7r7/dig.5">Link</a></td>
</tr>
</tbody>
</table>
### Split multi-value cells

<table>
<thead>
<tr>
<th>Source</th>
<th>Original Collect</th>
<th>Date</th>
<th>Site Name</th>
<th>Graphic Elem</th>
<th>Collection Subj</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>Text filter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Edit cells</td>
<td>Transform...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Edit column</td>
<td>Common transforms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Transpose</td>
<td>Fill down</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sort...</td>
<td>Blank down</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>View</td>
<td></td>
<td>Split multi-valued cells...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reconcile</td>
<td>Join multi-valued cells...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cluster and edit...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Original Collect</td>
<td>Date</td>
<td>Site Name</td>
<td>Graphic Elements</td>
<td>Collection Subj</td>
<td>DC Type</td>
<td>GE Type</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>------</td>
<td>-----------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>2000-19</td>
<td>Las Vegas Show Costume Designs Collection;</td>
<td>1945;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1946;</td>
<td></td>
<td>1947;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1948;</td>
<td></td>
<td>1949;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950;</td>
<td></td>
<td>1951;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1952;</td>
<td></td>
<td>1953;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1954;</td>
<td></td>
<td>1955;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Split multi-valued cells in column Graphic Elements (TGM)
Undo

1070 records
Facet view for Graphic Elements
Reconciliation

<table>
<thead>
<tr>
<th>Digital ID</th>
<th>Title</th>
<th>Individual Creator</th>
<th>Group Creator</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></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sho000114</td>
<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>
<td>Samba</td>
</tr>
</tbody>
</table>
Specifying Reconciliation service
<table>
<thead>
<tr>
<th>Site name URI</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>
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://digioc7.library.univ.edu/8860/edit

---

**Available Prefixes:**
- dc
- edm
- foaf
- owl
- rdf
- skos
- dcterm

---

**RDF Skeleton**

- (row index) URI
  - edm:ProvidedCHO
  - add rdf.type

- dc:creator
  - edm:Agent
  - add rdf.type

- dc:creator
  - Group creator URI
  - edm:Agent
  - add rdf.type

- dc:description
  - Description
coll

- edm:isRelatedTo
  - Name show URI
  - edm:Event
  - add rdf.type

- edm:hasRelation
  - Site name URI
  - edm:Place
  - add rdf.type

---

**RDF Preview**

Title
coll

Individual creator URI

Group creator URI

Description
coll

Name show URI

Site name URI

---

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://digloc7.library.unlv.edu:8890/edit

RDF Skeleton

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

```turtle
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edm: <http://www.europa.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 rdfs: <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:8890/ProvidedCHO/sho000119> a edm:ProvidedCHO;
  dc:title "Costume design drawing, yellow calypso costume, circa 1945-55";
  dc: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/ProvidedCHO/sho000119> a skos:Concept .

<http://digloc7.library.unlv.edu:8890/ProvidedCHO/sho000119> edm:hasType <http://id.loc.gov/vocabulary/graphicMaterials/tnm002607>
  dcterms:format "eps" ;
  dcterms:rights "This material may be protected by copyright. Personal, including educational and academic, use of this material
is permitted free of charge under the U.S. Copyright Act, Title 17, Section 107. For permission to use
outside the exemption, please contact: Center for Library Collections, http://digital.library.unlv.edu/ .

<http://id.loc.gov/vocabulary/graphicMaterials/tnm002607> a skos:Concept .
```

Exporting RDF files
**Actions**

- Prepare data
- Export data

**Technologies**

- CONTENTdm
- Open Refine
- Mulgara / Virtuoso

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

- Import data
- Publish
- Query
Mulgara Triple Store: Import
Mulgara Semantic Store

Graph URI: http://showgirls

Query Text: 

File: 

Submit Query Clear Query

Results: (1 query, 7.643 seconds)


A simple SPARQL query

```
Select *
Where {?s ?p ?o} limit 100
```
http://dbpedia.org/resource/Sun_City_South_Africa
http://dbpedia.org/page/El_Rancho_Hotel_and_Casino
http://library.unlv.edu:8890/Concept/Cest_Magnifique_(Lido_Paris_France)
http://library.unlv.edu:8890/Concept/Cest_Magnifique_(Lido_Paris_France)
http://library.unlv.edu:8890/Concept/Casino_de_Paris
http://library.unlv.edu:8890/Concept/Casino_de_Paris
http://library.unlv.edu:8890/Concept/Cocorico_(Lido_Paris_France)
http://library.unlv.edu:8890/Concept/Cocorico_(Lido_Paris_France)
http://library.unlv.edu:8890/Concept/Fan_Dance
http://library.unlv.edu:8890/Concept/Fan_Dance
http://library.unlv.edu:8890/Concept/Hallelujah_Hollywood!
http://library.unlv.edu:8890/Concept/Hallelujah_Hollywood!
http://library.unlv.edu:8890/Concept/Hello_America!
http://library.unlv.edu:8890/Concept/Hello_America!
http://library.unlv.edu:8890/Concept/Jubilee!
http://library.unlv.edu:8890/Concept/Jubilee!
http://library.unlv.edu:8890/Concept/Lido_(Paris_France)
http://library.unlv.edu:8890/Concept/Lido_(Paris_France)
http://library.unlv.edu:8890/Concept/Lido_de_Paris_(Las_Vegas_Nev.)
http://library.unlv.edu:8890/Concept/Lido_de_Paris_(Las_Vegas_Nev.)
http://library.unlv.edu:8890/Concept/Lido_de_Paris_(Las_Vegas_Nev.)
http://library.unlv.edu:8890/Concept/Lido_de_Paris_(Las_Vegas_Nev.)
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 skirt front, and hat with trim 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?