10-9-2013

Not Just for Geeks: A practical approach to linked data for digital collections managers

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

Cory K. Lampert  
*University of Nevada, Las Vegas, cory.lampert@unlv.edu*

Follow this and additional works at: [http://digitalscholarship.unlv.edu/libfacpresentation](http://digitalscholarship.unlv.edu/libfacpresentation)  
Part of the [Cataloging and Metadata Commons](http://digitalscholarship.unlv.edu/libfacpresentation)

**Repository Citation**
Available at: [http://digitalscholarship.unlv.edu/libfacpresentation/117](http://digitalscholarship.unlv.edu/libfacpresentation/117)

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.
Not Just For Data Geeks!

A Practical Approach to Linked Data for Digital Library Managers

Cory Lampert and Silvia Southwick
Salt Lake City
October 9, 2013
presenters

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

- Cory Lampert
  - Head of the Digital Collections Department
  - cory.lampert@unlv.edu
Today’s Agenda

- Welcome
- Morning
  - Linked Data Basic Concepts
  - Creating Triples and applying EDM activity
  - UNLV Linked Data Project
- Lunch
- Afternoon
  - Phases of data transformation demo and activities
  - Open Refine – data clean-up
  - Mulgara, SPARQL
- Discussion and Wrap-Up
Linked data: Why all the fuss?

- 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 do we mean by: “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).
What we do now produces:

- 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 – DATA IS TRAPPED!
Where linked data can take us:

- Free data from silos
- Expose relationships
- Powerful, seamless, interlinking of our data
- Users interact/query data in new ways
- Search results can be presented in more sophisticated ways
How?

- Our records can be deconstructed and assigned identifiers; creating data that can be used in Web architecture (HTTP, URIs)

- Data can be expressed in triples; statements that are machine-readable when transformed into Resource Description Framework (RDF)

- Linked data can be queried using SPARQL - SPARQL Protocol and RDF Query Language -- to retrieve and manipulate data stored in RDF.
Concept: Graph

- A graph is a collection of objects (represented by "nodes") any of which may be connected by links between them.
- Graphs are human readable.
- Graphs can represent a metadata record showing what is known about the item; relationships.
- Triples are the simplest form of a graph.
Concept: Triples

A triple is a statement, consisting of three parts:
- (a "subject" and an "object")
- and a relationship between them (a verb, or "predicate").

The subject-predicate-object triple forms the smallest possible RDF graph (although most RDF graphs consist of many such statements).
A Uniform Resource Identifier (URI) is simply a recognized standard for identifiers.

- URIs can be used to uniquely identify virtually anything
- URIs play a key role in enabling Linked Data because they represent “things” (subjects) in a machine-readable form
- URIs are used in HTTP web architecture
Principles of Linked Data

1. Use URIs as names for things (people, organizations, artifacts, abstract concepts, etc.)

2. Use HTTP URIs so that people can look up those names

3. When someone looks up a URI, provide useful information, using the standards (RDF) to create statements

4. Beyond describing the item, include links to other URIs so that people can discover other related items
Where do we start?

We already have the information we need in our metadata records to create triples. We just need to think of it differently:

- **Subjects** – **Objects** - **Predicates**
- Each metadata field may produce one or several statements
- One metadata record can produce many, many, triples
Expressing metadata as triples

What are possible triples for this “thing”?

- `<this thing> <created by> <Las Vegas News Bureau>`
- `<this thing> <is a> <photographic print>`
- `<this thing> <depicts> <Frank Sinatra>`
- `<this thing> <depicts> <Jack Entratter>`

---

- `<Frank Sinatra> <knows> <Jack Entratter>`
- `<Jack Entratter> <knows> <Frank Sinatra>`

---

- `<Frank Sinatra> <is an> <entertainer>`
- `<Jack Entratter> <is a> <theatrical producer>`

---
Expressing records as triples: graph example

- Triples are expressed as:
  subject – predicate – object

- Examples:
  Frank Sinatra -- is an – entertainer
  Frank Sinatra – knows – Jack Entratter
Triples and RDF

Once we have triples we need to:

- Assign URIs to each subject
- URIs definitely are used for subjects, and might also represent objects.
- URIs are essential for constructing RDF statements

These steps take the human readable graph and make it machine readable!
Examples of records

- Showgirls
- Menus
- Dreaming the Skyline
Graphical representation of the photo triples
Adding triples from the other records

What are the URIs for subjects, predicates and objects?
Triples: Text/Graph → RDF

Source: Introduction to RDF at http://www.linkeddatatools.com/introducing-rdf

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:feature="http://www.linkeddatatools.com/clothing-features#">
    <rdf:Description rdf:about="http://www.linkeddatatools.com/clothes#t-shirt">
        <feature:color rdf:resource="http://www.linkeddatatools.com/colors#white"/>
    </rdf:Description>
</rdf:RDF>
```
ACTIVITY: Brainstorming Triples

- Look at the metadata record you brought and think about subject-object-predicates
- Start listing some possible triples in text on handout
- When you have several, try to graph one of the triples
- Work individually or in groups and discuss
Getting From Triples to the next step

Once we understood triples we needed to answer some questions:

• Which triples to create? (literal, outgoing links, incoming links, triples that describe related resources, triples that link to descriptions, triples that indicate provenance of the data, etc.)
• Which vocabularies that will be adopted for predicates?
• How to get URIs for the cv terms used as value of the digital collection fields
• How to specify URIs for new “things”
A Little Help From EDM

- Data model is a way to help us to bring some order to the chaos of all these triples.
- Europeana Data Model gives us a framework to help organize, structure, and define how we create triples and express them in RDF.
- Adopting a current model is preferable to creating your own (interoperability).
Vocabularies

In addition to the data model we explored how data could be reconciled with existing linked datasets/vocabularies.

- Thesaurus of Graphic Materials and LoC
- DCMI Type Vocabulary
- Friend of a Friend Vocabulary (FOAF)
- Geonames
- Creative Commons Rights Expression vocabulary
- Schema.org

Many more at: http://lov.okfn.org/dataset/lov/
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
grey background -- not yet implemented by Europeana
blue font -- properties pertaining to the edm vocabulary
ACTIVITY: Exploring EDM

http://pro.europeana.eu/c/document_library/get_file?uuid=99ce6a74-8e55-4321-917a-65bdff1fe5bc&groupId=51031

- Refer back to triples (the ones you wrote in text/English)
- Translate the verbs (properties) by browsing through the EDM core and contextual classes using handout
- Discuss and work in groups
Is this work worth it?

- We add value by creating rich metadata records at our institutions.
- When these records are harvested as Dublin Core they lose some of that context.
- When harvested metadata records are automatically transformed into linked data (OCLC) they lose even more.
- You get “linked data” at a cost.
How can we create rich linked data?

Create a complementary data structure that would allow dynamic interlinking among data

How?
- Export records from the collections
- Deconstruct these records by extracting data from them
- Apply vocabularies
- Adopt a common model to express data
- Publish data in a data space (Linked Data Cloud) where links among data are created automatically
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
How we started

• Created a study group in the Library (members from various areas of the library)
• Watched webinars on the topic and have discussions after the webinars
• Created an internal wiki with linked data resources
• Participated in linked data interest groups
• Follow the literature on this topic
Phases of the project

- Literature Review
- Evaluating Technologies
  - Research existing technologies and best practices
  - Develop small experiments with technologies
  - Make decisions of which technologies to adopt, adapt or develop
- Data preparation
  - Select and prepare records from digital collections to participate in the project
- Run process to generate data from the original records
- Publish on the Linked Data Cloud
- Assess results
The Linking Open Data Cloud

Project Perspective: Sara

- Why we needed Sara’s help
- How she accelerated our learning
- What she has learned so far
- Her thoughts on linked data beyond digital collections
Project Demo
Discussion

- Questions?
- What type of linked data explorations are happening in your organizations?
- Challenges?