4-8-2015

Linking to Our Future: Cataloging & Metadata in Transition

Sherry L. Vellucci
University of New Hampshire - Main Campus, sherryvellucci@unh.edu

Follow this and additional works at: https://scholars.unh.edu/library_pub
Part of the Cataloging and Metadata Commons, and the Scholarly Communication Commons

Recommended Citation
https://scholars.unh.edu/library_pub/77

This Presentation is brought to you for free and open access by the University Library at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in University Library Scholarship by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
Linking to Our Future: Cataloging & Metadata in Transition
Linking to Our Future
Cataloging & Metadata in Transition

April 8, 2015
11:00-12:30 pm,
Rm. 352

Sherry Vellucci
Today's Presentation

**Part 1:** Everything you need to know about the Web in (almost) 3 slides

**Part 2:** Linked data on the Web and in Libraries (almost)
Using the Web to create defined (typed) links between data from different sources that weren’t previously linked.

A method for sharing data—within and outside a library environment in a non-library-centered exchange format

Depends on everyone using standards that enable data exchange

Identifying relationships between entities is essential to the success of a linked data environment
Some Key Terms

- **Semantics**: The study of the meanings of words and phrases in language; the meanings of words and phrases in a particular context.

- **Relationship**: The way in which two or more people or things are connected.

- **Link**: An identifier attached to an element in a system to indicate or permit connection with other similarly identified elements, especially a hyperlink in a computer file.
More Key Terms

**Data:** The primary building block of information; factual statements or figures presented in a form that can be understood, interpreted, and communicated by a human being or processed by a computer; when contextualized by its use, data becomes information.

**Entity:** Independent, separate, or self-contained existence

**Record:** A collection of related items of information (as in a database) treated as a unit
Evolution of the Web
or,
How Did We Get Here?

Everything you need to know about the Web in 5 slides!

Internet  World-Wide-Web  Semantic Web
“Vague but exciting”
The World-Wide-Web

0 HTTP developed in 1991
0 Web of Documents
   0 Pages (HTML documents with hyperlinks)
0 Web of Discovery
   0 Search engines index and infer relevance
   0 Implicit relationships between documents
   0 Lack of semantics (not “typed” links)
0 All are designed for humans
“The semantic web is a vision of information that is understandable by computers, so computers can perform more of the tedious work involved in finding, combining, and acting upon information on the web.” – Wikipedia
The Semantic Web of Linked Data

- Linked Data is *not* about searching for specific documents or visiting particular websites, it’s about **identifying and connecting things**
  - Making connections between related data (entities) from different sources
  - Computer programs understand connections (for humans & machines)
  - Data is **sharable, extensible, and reusable**
  - Entity-based Architecture

- Relationships are Key!
Linked Data, or,
How I Learned to Stop Worrying and
love Cyber-Vagueness
What Do Computers Like?

**Unique Identifiers**
(they confuse easily, & need specific directions & information)

**Consistent, Predictable, Non-human-language Data**
(See Unique IDs; they are also verbally challenged)

**Controlled Vocabularies**
(See previous likes. They need the same language to play with other computers)

**To Understand Relationships**
(They are really into genealogy & don’t like blind dates)

... and the answer is ...

URI*s*!!!
Four Rules of Linked Data

1. Use URIs as names for things

id.loc.gov/authorities/names/n87929612

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

http://id.loc.gov/authorities/names/n87929612

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

4. Include links to other URIs so that they can discover more things

Tim Berners-Lee
Building the Web of Data

How will this work in a library?
Library Research: It’s About Relationships, Stupid!
## The Library Perspective

Deconstruct MARC records into individual entities (bits of data)

Create RDF statements that link entities by identifying relationships

### Linked Data Basics

<table>
<thead>
<tr>
<th>Records</th>
<th>Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link entities</td>
<td>Statements</td>
</tr>
<tr>
<td>Resource Description Framework (RDF)</td>
<td></td>
</tr>
<tr>
<td>Relationships defined in “triples”</td>
<td></td>
</tr>
<tr>
<td>Entity A → Has relationship to → Entity B</td>
<td></td>
</tr>
<tr>
<td>Record becomes a <strong>graph</strong> formed by triples</td>
<td></td>
</tr>
<tr>
<td>Entities represented by URIs (or verbal values)</td>
<td></td>
</tr>
<tr>
<td>Catalog (closed) → Web (open)</td>
<td></td>
</tr>
<tr>
<td>Not limited to libraries or bibliographic records</td>
<td></td>
</tr>
</tbody>
</table>

4/08/2015
RDF is a Data Model that makes statements about [Web] resources.

Literal/String = Value expressed as natural language.

Statements have 3 parts and are called triples:

- subject — predicate — object

What we're describing (entity 1)

Verb describes relationship between 2 entities

Value of what we're describing (Entity 2)

More Terminology & a RDF Graph
James C. Collins is author of Good to Great.

Subject: James C. Collins
Predicate: is author of
Object: Good to Great

sameAs:
- [http://id.loc.gov/authorities/names/n91123848]
- [http://purl.org/dc/elements/1.1/creator]
- [http://www.worldcat.org/oclc/46835556]
Library Data Stored as Entities

Library Knowledge Graph

Graph of Relationships

Adapted from: Richard Wallis. Linked Data: From Library Entities to the Web of Data.
Where are our users?

“When my community searches the Web for something we have [in the library], we better show up as an option.” — Chuck Gibson, Director & CEO Worthington Public Library

Adapted from: Richard Wallis. Linked Data: From Library Entities to the Web of Data.
**Not in Google**

**Why?**

**Google does not understand:**

MARC, ISBD, RDA, Z39.50, etc.
The real problem is that we don’t expose our collections very well on the web!

- Locked in library catalog silos that are not searchable on the Web, or
- Available from aggregator’s database—No direct access from Web search
Unlock Library Data

- Transforms library data so that it is of the web
- Becomes easily searchable
- Libraries can integrate outside data into what they already have
- Anyone can reuse library data to create their own applications

Fact: BnF converted to LD
- 80% + users now come from search engines
Connect different parts of the library’s system

- Different parts of the library structure (ILS, ERMS, Archive Finding Aids, LibGuides, etc.) would be able to share data more easily, allowing searches to easily jump from one area to another.
- Link to outside information relevant to a search.

https://vimeo.com/54674757
Jane Austen was an English novelist whose works of romantic fiction, set among the landed gentry, earned her a place as one of the most widely read writers in English literature. She was born on December 16, 1775, at Steventon, Hampshire, England.

Her novels, including "Pride and Prejudice," "Sense and Sensibility," and "Emma," are celebrated for their social commentary of 19th-century English society. Austen's letters and journals, which often reflect her thoughts on literature and society, are also noteworthy.

Austen's works are known for their strong female characters, whose relationships and personal growth are central to the narratives. Her novels often explore themes of marriage, social status, and personal fulfillment.

Quotes:

"I do not want people to be very agreeable, as it saves me the trouble of liking them a great deal."

― Jane Austen, from "Jane Austen's Letters"
Core Questions

- Is linked data and the Semantic Web achievable?
- Will it bring substantial benefits to libraries?
- Is linked data something we can ignore?
So, How Do We Get There?
- Being developed by Zepheira and the Library of Congress
- The foundation for future bibliographic information
- Eventual replacement for MARC
- A data model based on identifying information entities with URIs
What is schema.org?

Developed by:

Provides a hierarchical structure to identify specific types of:

Things
- Creative Work (Book, Movie, MusicRecording, Recipe, TVSeries ...)
- Person (alive, dead, undead, fictional)

Schema.org: Another Solution
# Person

A person (alive, dead, undead, or fictional).

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>additionalName</td>
<td>Text</td>
<td>An additional name for a Person, can be used for a middle name.</td>
</tr>
<tr>
<td>address</td>
<td>PostalAddress</td>
<td>Physical address of the item.</td>
</tr>
<tr>
<td>affiliation</td>
<td>Organization</td>
<td>An organization that this person is affiliated with. For example, a school/university, a club, or a team.</td>
</tr>
<tr>
<td>alumniOf</td>
<td>EducationalOrganization</td>
<td>An educational organizations that the person is an alumni of. Inverse property: alumni.</td>
</tr>
<tr>
<td>award</td>
<td>Text</td>
<td>An award won by this person or for this creative work. Supersedes awards.</td>
</tr>
<tr>
<td>birthDate</td>
<td>Date</td>
<td>Date of birth.</td>
</tr>
<tr>
<td>birthPlace</td>
<td>Place</td>
<td>The place where the person was born.</td>
</tr>
<tr>
<td>brand</td>
<td>Organization or Brand</td>
<td>The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.</td>
</tr>
<tr>
<td>children</td>
<td>Person</td>
<td>A child of the person.</td>
</tr>
<tr>
<td>colleague</td>
<td>Person</td>
<td>A colleague of the person. Supersedes colleagues.</td>
</tr>
<tr>
<td>contactPoint</td>
<td>ContactPoint</td>
<td>A contact point for a person or organization. Supersedes contactPoints.</td>
</tr>
<tr>
<td>deathDate</td>
<td>Date</td>
<td>Date of death.</td>
</tr>
<tr>
<td>deathPlace</td>
<td>Place</td>
<td>The place where the person died.</td>
</tr>
<tr>
<td>duns</td>
<td>Text</td>
<td>The Dun &amp; Bradstreet DUNS number for identifying an organization or business person.</td>
</tr>
<tr>
<td>email</td>
<td>Text</td>
<td>Email address.</td>
</tr>
<tr>
<td>familyName</td>
<td>Text</td>
<td>Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the name property.</td>
</tr>
<tr>
<td>faxNumber</td>
<td>Text</td>
<td>The fax number. Sherry Vellucci</td>
</tr>
</tbody>
</table>
Good to great: why some companies make the leap—and others don’t

Author: James C Collins
Database: WorldCat
Summary: Answers to commonly asked questions about how good-to-great principles can help societal organizations make the leap to greatness, using interviews with over 100 societal leaders.
Rating: 0 with reviews - Be the first.
Subjects: Leadership, Strategic planning, Organizational change.
View all subjects

You are connected to the UNH Durham network
Request item through ILL  Find Full Text at UNH  Find Full Text at UNH  Request item through ILL  Request item through ILL  Request item through ILL  Request item through ILL  Navigator RTM
Search the catalog at your library OCLC FirstSearch
Schema Creator for Book


**Name**
Good to Great

**URL**

**Description**
Book by Jim Collins

**Author**
James C Collins

**Data published**
2001

**Publisher**
HarperBusiness

**Book edition**
first

**ISBN**

---

**Good to Great**
Book by Jim Collins
Written by: James C Collins
Published by: HarperBusiness
Date published: undefined/undefined/2001
Edition: first
Available in Hardcover

**Code**

```html
<div itemscope itemtype="http://schema.org/Book">
    <strong>Good to Great</strong></div>
  </a>
  <div itemprop="description">Book by Jim Collins</div>
  <div itemprop="author" itemscope itemtype="http://schema.org/Person">
    Written by: <span itemprop="name">James C Collins</span>
  </div>
  <div itemprop="publisher" itemscope itemtype="http://schema.org/Thing">
    Published by: <span itemprop="name">HarperBusiness</span>
  </div>
  <div itemprop="bookEdition">first</div>
  <div itemprop="datePublished">2001</div>
</div>
```
Libraries should use

\texttt{schema.org}

to Describe & Link to the Web

BUT

Continue to apply other vocabularies & standards
Are we there yet?

Gartner Technology Hype-Cycle
Adapted by Eric Miller
Phases of Linked Data / BIBFRAME Adoption

**The Chasm**

- **Technology Enthusiasts**
- **Visionaries**
- **The Early Market**
- **Experimenters**
- **Early Implementers**
- **Data Publishers & Connectors**
- **Mainstream Workflow**
- **Conservatives**
- **Back Office Systems**
- **Skeptics**

---

<table>
<thead>
<tr>
<th>Early Market</th>
<th>Mainstream Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Clarify Space</td>
<td>- &quot;Final&quot; Standards &amp; Best Practices</td>
</tr>
<tr>
<td>- Determine the Need</td>
<td>- New businesses and models</td>
</tr>
<tr>
<td>- Define a Foundation</td>
<td>- &quot;There's Linked Data in there!?&quot;</td>
</tr>
<tr>
<td>- Draft Specifications</td>
<td>- &quot;Use other's data&quot;</td>
</tr>
<tr>
<td>- Test the Assumptions</td>
<td>- &quot;Participate - Publish, Share, Connect&quot;</td>
</tr>
<tr>
<td>- Evaluate Data, Processes, &amp; Gaps</td>
<td>- &quot;Begin to work at scale&quot;</td>
</tr>
</tbody>
</table>

---

*Sherry Vellucci*
Anything Else We Need?

- An ILS vendor that lets us control our own HTML!
  - Can’t take advantage of Linked Data without that!
- How do we get there?
  - Bother current vendor
  - Write microdata into RFPs
  - Collaboration

Adapted from Dorothea Salo; Libraries Storming Linked Data!
Be part of the web
not just on it
linked data
Open APIs
RDF
discover
enables
HTML
open
 LODLAM
Linked Open Data in Libraries Archives and Museums
VIAF
Virtual International Authority File
It Takes Community Buy-in
"If all of this sounds otherworldly and vague, it is because there is no specific vision of where these changes will lead us. The crystal ball is unfortunately shortsighted... The few things that are certain, however, point to the Web, and its eventual successors, as the place to be. For libraries, this means yet another evolutionary step in the library of our catalog: from metadata to metaDATA."
Libraries are moving from Cataloging to Catalinking!
Thank you!
Questions? Comments?

Sherry Vellucci
sherry.vellucci@unh.edu