Implementing Linked Data & Discovery: Liberating Library Data & Empowering Users

Jason A. Clark
Associate Professor
Head of Library Informatics & Computing
Montana State University (MSU) Library

ALA Annual Conference
ProQuest Exhibitor Session
June 27, 2015
pinboard.in tag

pinboard.in/u:jasonclark/t:lod/
twitter as channel (#hashtag)

@jaclark #lod
LOD in Libraries

- UNLV Linked Data Project
  http://www.library.unlv.edu/linked-data

- Worldcat
  http://www.oclc.org/data.en.html

- Europeana Linked Open Data
  http://labs.europeana.eu/api/linked-open-data/introduction/
LOD in Libraries

- **iiiiff**
  
  http://iiif.io/

- **DPLA**
  
  http://dp.la/info/developers/codex/responses/field-reference/

- **Vivo**
  
  http://vivoweb.org
Welcome to VIVO

VIVO is a research-focused discovery tool that enables collaboration among researchers across all disciplines.

Browse or search information on people, departments, courses, grants, and publications.

Search VIVO

limit search →  Search
Linked Data for Libraries: From Experimentation to Practice at Scale

Session Type: Presentation

Session Description
Libraries have been moving to embrace linked data for years now, slowly and fitfully it sometimes seems. Now the landscape is firming up as major players have begun committing to well-formed initiatives that move beyond experimentation to practice at scale. This panel presents a view of four diverse, larger-scale, linked data efforts where the “rubber is meeting the road.” Each of the panelists will briefly present an overview of their efforts. A moderated discussion will follow, giving the panelists and audience a chance to compare and contrast approaches, and the larger implications for libraries of this potentially disruptive innovation.
Welcome to the project wiki-space for the Linked Data for Libraries (LD4L) project. The project, supported by Cornell University Library, the Harvard Library Innovation Lab, and the Stanford University Library, received nearly $1 million two-year grant from the Andrew W. Mellon Foundation.

The goal of the project is to create a Scholarly Resource Semantic Information Store (SRSIS) that librarians and other domain experts and scholars add to information resources when they select, and use those resources, together with the social value evident from patterns of usage.

Our intent is to do so using existing ontologies and Open Source technology.

Project Pages

- Project Proposal:
  - Why Linked Data?
  - Previous Library-related Linked Data work at Cornell, Harvard, and Stanford
  - Rationale for the project
  - Project Description
  - Expected Outcomes
  - Intellectual Property and Sustainability

- Project Timeline
- Planned LD4L Workshop (tentatively set for January 2015)
- Project Team
Overview

• Why Linked Open Data (LOD)?
• Truths about LOD
• LOD in Practice at MSU
• Questions
Defining Terms and Truths
What is Linked Open Data?

- Practice of publishing structured, machine actionable data with an open license for sharing and reuse.
Truths about Linked Data
SELECT * WHERE {?s a ?o} limit 100
Limited Interfaces
Eureka SPARQL is starting to make sense!

Ugh SPARQL you make my brain hurt.

Working on loading data into a triplestore and then running SPARQL on it at #accessYYC hackfest
Learning curves are expensive, per Nannette.

#rwlod
Learning Curves
A Research Team Effort

Scott Young: Digital Initiatives Librarian at Montana State University @hei_scott

Jason Clark: Head, Library Informatics & Computing at Montana State University @jaclark

Patrick O’Brien: Semantic Web Research Director at Montana State University Library

Kenning Arlitsch: Dean of the Library at Montana State University @kenning_msu

Doralyn Rossmann: Head, Collection Development at Montana State University Library @doralyn
Need Staffing + Interest
Why LOD?

• Libraries as publishers in Web of Data
• Reuse of our data = Significance
• Can improve findability in search engines (search engine optimization)
• Levels of description in LOD enable entirely new applications
Why LOD at MSU?

- Semantic Publishing
- Machine Definitions
- Discovery
Principles of Publishing LOD

- Use URIs for description
- Make structured data available on WWW
- Allow for machine-actionability and sharing through common ontologies

* machine-actionable is a term used by Karen Coyle to mean readable and interpretable by computers or software agents. See her Library Technology Report on *Understanding the Semantic Web* at http://www.metapress.com/content/g212v1783607/*
LOD Ontologies & Vocabularies

URIs link to:

1. Reputable machine-readable endpoints
   - E.g., Worldcat, VIAF, or DbPedia

2. Open Ontologies & Vocabularies
   - E.g., Dublin Core Terms, Library of Congress Subject Headings

* See Linked Open Vocabularies (LOV) for a list of possible ontologies and vocabularies.
  http://lov.okfn.org/dataset/lov/
5 Star Linked Open Data

1. make your stuff available on the Web (whatever format) under an open license

2. make it available as structured data (e.g., Excel instead of image scan of a table)

3. use non-proprietary formats (e.g., CSV instead of Excel)

4. use URLs to denote things, so that people can point at your stuff

5. link your data to other data to provide context

http://5stardata.info/
LOD at Montana State University (MSU)
Case Studies

1. Book as Linked Data Platform
2. Linked People
3. Web Scale Cataloging
Semantic Publishing
Book as Linked Data Platform
The Swabian

He was slow in learning how to talk. "My parents were so worried," he later recalled, "that they consulted a doctor." Even after he had begun using words, sometime after the age of 2, he developed a quirk that prompted the family maid to dub him "der Depperte," the dopey one, and others in his family to label him as "almost backwards." Whenever he had something to say, he would try it out on himself, whispering it softly until it sounded good enough to pronounce aloud.

"Every sentence he uttered," his worshipful younger sister recalled, "no matter how routine, he repeated to himself softly, moving his lips." It was all very worrying, she said. "He had such difficulty with language that those around him feared he would never learn."  

His slow development was combined with a cheeky rebelliousness toward authority, which led
Web Book
The Research

Institute of Museum and Library Services
Sparks Ignition Grant
Book 1
demo: arc.lib.montana.edu/book/home-cooking-history-409
code: github.com/jasonclark/bib-template

Book 2
demo: arc.lib.montana.edu/book/opsis
code: github.com/msulibrary/bib-template-fiction

* Additional prototypes in development for textbook and an academic journal
Every Book its [machine] Reader

Apologies to Ranganathan
Gingerbread

Ingredients:

- 1/2 c. soft shortening
- 2 1/4 c. flour
- 2 T. sugar
- 1 tsp. soda
- 1 egg
- 1/2 tsp. salt
- 1 c. dark New Orleans Molasses
- 1 tsp. ginger
- 1 c. boiling water
- 1 tsp. cinnamon

Directions:

- Mix shortening, sugar and egg.
- Blend in molasses and cup of boiling water, cooled first.
- And flour, soda, salt, ginger and cinnamon.
- Beat until smooth.
- Pour into well-greased 9 inch square pan.
- Bake 45 minutes to 50 minutes at 350 degrees.
- Serve with whipped cream and applesauce.

Young Lyle Bergh grew up in a family that placed emphasis on the importance of food and eating well. The family's traditional recipe for gingerbread was a cornerstone of their cooking.

In 1940s northeast Montana's food options were limited. For inhabitants of Plentywood, Montana, and his family, it meant that home cooking, not eating out, was the norm. Meals like steak, potato salad, and coffee. Lyle would come home and have farm life will have breakfast cooked by his mother. Bee, cooked.

When the Bergh family headed to Seattle, Washington, it meant family had to be reunited, and there would be treats to be enjoyed. Going to Seattle meant visiting relatives and having meals together.

Tradition and Travel

by Jonathan Bergh

Every winter the Bergh family celebrates the season with a special tradition. It is the time of the year when the Bergh family...
Check out Opsi (in a browser)
arc.lib.montana.edu/book/opsis/

6:54 PM - 9 Feb 2015
Structured Data

Data Model + Machine Readability
<xml version="1.0"?>
<items>
  <item>
    <id>1</id>
    <name>Angel Food Cake</name>
    <creator>Laura Cameron</creator>
    <description>Angel food cake recipe written by Laura Cameron for Montana State University’s upper division history course.</description>
    <image>http://upload.wikimedia.org/wikipedia/commons/thumb/6/64/AngelFoodCake.jpg/200px-AngelFoodCake.jpg</image>
    <cookingMethod>Baking</cookingMethod>
    <ingredients>18 egg whites, 2 t. cream of tartar, 1 pinch salt, 1 1/2 cups white sugar, 1 cup cake flour, 1/2 cup confectioners sugar</ingredients>
    <recipeInstructions>&lt;li&gt;Preheat oven to 350 degrees&lt;/li&gt;
    &lt;li&gt;Sift cake flour and confectioners sugar together 5 times. Set aside.&lt;/li&gt;
    &lt;li&gt;In a large bowl whip egg whites with salt until foamy. Add cream of tartar and continue to beat until soft peaks form.&lt;/li&gt;
    &lt;li&gt;Quickly fold in flour mixture.&lt;/li&gt;
    &lt;li&gt;Pour into a 10 inch tube pan.&lt;/li&gt;
    &lt;li&gt;Bake for 45 minutes.&lt;/li&gt;&lt;/recipeInstructions&gt;
    <cookTime>PT45M</cookTime>
    <prepTime>PT15M</prepTime>
    <totalTime>PT60M</totalTime>
    <recipeCategory>dessert</recipeCategory>
    <recipeCuisine>American</recipeCuisine>
    <recipeYield>8 to 10 servings</recipeYield>
    <sameAs>http://dbpedia.org/page/Angel_food_cake</sameAs>
    <keywords>cream of tartar, confectioners sugar, vanilla extract, egg whites, cup cake flour, cake flour</keywords>
    <publisher>Montana State University Library</publisher>
    <dateCreated>2012-12-01T18:34Z</dateCreated>
    <dateModified>2012-12-01T18:34Z</dateModified>
    <datePublished>2012-12-01T18:34Z</datePublished>
    <inLanguage>en</inLanguage>
  </item>
</items>
Schema.org

Controlled Vocabulary
What is Schema.org?

This site provides a collection of schemas that webmasters can use to markup HTML pages in ways recognized by major search providers, and that can also be used for structured data interoperability (e.g. in JSON). Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right Web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

We invite you to get started!

View our blog at blog.schema.org.
What if?

Books were machine indexed for semantic discovery

Journal articles were semantically tagged at the page level
Gingerbread

Ingredients:
- 1/2 c. soft shortening
- 2 1/4 c. flour
- 2 T. sugar
- 1 tsp. soda
- 1 egg
- 1/2 tsp. salt
- 1 c. dark New Orleans Molasses
- 1 tsp. ginger
- 1 c. boiling water
- 1 tsp. cinnamon

Instructions:
Mix shortening, sugar and egg. Blend in molasses and cup of boiling water, cooled first. Add flour, soda, salt, ginger and cinnamon. Beat until smooth. Pour into well-greased 9 inch square...
Gingerbread - Jonathan Bergh

Table of Contents > Recipe and Essay Gingerbread. Ingredients: 1/2 c. soft shortening; 2 1/4 c. flour; 2 T. sugar; 1 tsp. soda; 1 egg; 1/2 tsp. salt; 1 c. dark New ...
Behavioral Metadata
Machine Definitions

Linked People

www.lib.montana.edu/people
Linked Data is People

Using Linked Data to Reshape the Library Staff Directory
"@context": "http://schema.org",
"@type": "Person",
"name": "Jan Zauha",
"jobTitle": "Outreach Librarian [Professor]",
"workLocation": "Learning and Research Services",
"email": "jzauha@montana.edu",
"image": "http://shelf.lib.montana.edu/people/meta/img/photos/janzauha-thumb.jpg",
"telephone": "406-994-6554",
"address": {
  "@type": "PostalAddress",
  "addressLocality": "Bozeman",
  "addressRegion": "MT",
  "postalCode": "59717-3320",
  "streetAddress": "P.O. Box 173320"
},
"colleague": [
  "shelf.lib.montana.edu/people/2",
  "shelf.lib.montana.edu/people/4",
  "shelf.lib.montana.edu/people/17",
  "shelf.lib.montana.edu/people/28",
  "shelf.lib.montana.edu/people/33",
  "shelf.lib.montana.edu/people/40",
  "shelf.lib.montana.edu/people/79",
  "shelf.lib.montana.edu/people/47",
  "shelf.lib.montana.edu/people/82",
  "shelf.lib.montana.edu/people/88",
  "shelf.lib.montana.edu/people/101",
  "shelf.lib.montana.edu/people/135",
  "shelf.lib.montana.edu/people/138"
],
"sameAs": [
  "http://scholar.google.com/citations?user=htx7S9oAAAAJ&hl=en",
  ""
],
"worksFor": "Montana State University (MSU)"
Jan Zauha

Title: Outreach Librarian [Professor]
Department: Learning and Research Services, Montana State University
Phone: 406-994-6554
Email: jzauha@montana.edu
Calendar: Schedule Appointment

Room: 117C
Vita
Get QR Embed Code

Back to Staff Home page
How are people and data understood?
Create distributed, linked, flexible data systems.

Add context and dimension to data.

Demonstrate relationships among data based on standardized and reusable vocabulary

Networked data is smarter data.
Applying the concept of linked data for our local community

Finding a real-world use-case for graph modeling using open source tools

What if our people could be discovered and understood?
Where are your people?

Where are your concepts?
2 headings found for *jason a clark*

<table>
<thead>
<tr>
<th>Heading</th>
<th>Type</th>
<th>Sample Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Clark, Jason A.</td>
<td>Personal</td>
<td>Building mobile library applications</td>
</tr>
<tr>
<td>2  Clark, Jason A., 1975-</td>
<td>Personal</td>
<td>Florence</td>
</tr>
</tbody>
</table>
Jason A. Clark

Personal Information

Biography

Associate Professor, Montana State University

Works


Country: United States

Keywords: machine learning - information architecture

Websites:
http://www.jasonclark.info
https://github.com/jasonclark
An Ontology
Jan Zauha

**Title:** Outreach Librarian [Professor]

**Department:** Learning and Research Services, Montana State University

**Phone:** 406-994-6554

**Email:** jzauha@montana.edu

**Room:** 117C

**Vita**

**Get QR Embed Code**

<http://chart.apis.google.com/chart>
Jan Zauha is defined/understood as an entity
Discovery
Web Scale Cataloging
SEO + Semantic Understanding
Designing Your Collections for Web Scale Search

James Willard Schultz

http://arc.lib.montana.edu/schultz-0010/
Introducing Semantic Markup

HTML5 semantic tags and RDFa that help classify page types and define the types of content on the page

“If Google understands the content on your pages, we can create rich snippets—detailed information intended to help users with specific queries.” https://support.google.com/webmasters/answer/99170?hl=en&ref_topic=1088472
Item: Women posing on college locomotive engine

Title: Women posing on college locomotive engine
Creator: unknown
Date: unknown
Description: Women posing on a college locomotive engine.
Notes:
Physical Description: Photo print - Black and White
Subjects: Locomotive
Keywords: locomotive
Photograph ID: parc-000432

Before RDFa
Item: Women posing on college locomotive engine

Title: Women posing on college locomotive engine
Creator: unknown
Date: unknown
Description: Women posing on a college locomotive engine.
Notes:
Physical Description: Photo print - Black and White
Subjects: Locomotive
Keywords: locomotive
Photograph ID: parc-000432
External Enumerations

"We define here some specific integration points through which selected externally maintained vocabulary can be published as part of schema.org markup"

www.w3.org/wiki/WebSchemas/ExternalEnumerations
additionalTypes, specialty

https://schema.org/additionalType, http://schema.org/specialty
Improved Metrics? Research in progress...
Images for michelle gollehon

Michelle Gollehon | Facebook
https://www.facebook.com/michelle.gollehon.3
Michelle Gollehon is on Facebook. Join Facebook to connect with Michelle Gollehon and others you may know. Facebook gives people the power to share and ...

Michelle Gollehon | LinkedIn
www.linkedin.com/pub/michelle-gollehon/2b/a16/568
View Michelle Gollehon's professional profile on LinkedIn. LinkedIn is the world's largest business network, helping professionals like Michelle Gollehon ...

Michelle Gollehon - Montana State University MSU Library
www.lib.montana.edu/people/about.php?id=97
Library Technician - Montana State University
People in the library - individual results. photo of Michelle Gollehon. Michelle Gollehon. Title: Library Technician. Department: Collection Development, Montana ...
Sara Mannheimer grew up in Gothenburg, Sweden, and was educated in the United States, the Netherlands, and the Czech Republic. She now lives in ...

Sara Mannheimer is a Swedish novelist. She hails from Gothenburg and was educated in the US, Holland, and the Czech Republic. Her debut novel Reglerna was nominated for the August Prize and won the debut writers' prize from Borås Tidning newspaper. Wikipedia

Born: 1967, Lund, Sweden
Awards: European Union Prize for Literature

People also search for
Lada Žigo
Gunstein Bakke
Jana Beňová
Adda Djørup
Laurence Plazenet
New Search and Browse
Staff Directory & Departments

Browse by Liaison Area
I want help with... + full list

Browse by Department
Administration
Collection Development
Learning and Research Services
Library Informatics and Computing
Resource Description and Metadata Services
Special Collections and Archives

Browse by Expertise Topic
I want help with... + full list

Related pages
Browse All Staff
Browse All Faculty and Professional(s)
Browse liaisons by department
Browse liaisons by name
Contact Us

Browse by staff last name
ABC | DEF | GHI | JKL | MNO | PQR | STU | VWX | YZ | Other
Jan Zauha

Title: Outreach Librarian [Professor]
Department: Learning and Research Services, Montana State University
Phone: 406-994-6554
Email: jzauha@montana.edu
Room: 117C

Liaison Role(s):
- History
- Philosophy
- American studies
- Writing composition
- English

Expertise and Skill(s):
- Information Literacy
- Professional Development
- Reading
- Performance studies

More Information:
- Graph Visualization
- Vita
- Get QR Embed Code

Back to Staff Home page
Linked Data Tools
RDF Search Engines
Unified search
Enter a query, e.g., name:(barack obama) birthdate:(04 08 1961)

jason clark

Ontology-based search [Show]

Found ~1 thousand results in 121 ms.

Showing results 1 to 10 of ~1 thousand

1. jason clarke
http://www.examiner.com/topic/jason-clarke

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td>jason clarke</td>
</tr>
<tr>
<td></td>
<td>jason clarke</td>
</tr>
<tr>
<td></td>
<td>jason clarke</td>
</tr>
</tbody>
</table>
Semantic Classification Tools
getSEMDantic [beta]
jason a. clark

Feed me a URL
Get Analysis

Key: Demo App What is this? View Code @jaclark Credit

github.com/jasonclark/getsemantic
Linked Data Creation Tools
Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; extending it with web services; and linking it to databases like Freebase.

Please note that since October 2nd, 2012, Google is not actively supporting this project, which has now been rebranded to OpenRefine. Project development, documentation and promotion is now fully supported by volunteers. Find out more about the history of OpenRefine and how you can help the community.

Using OpenRefine - The Book

Using OpenRefine, by Ruben Verborgh and Max De Wilde, offers a great introduction to OpenRefine. Organized by recipes with hands on examples, the book covers the following topics:

1. Import data in various formats
2. Explore datasets in a matter of seconds
3. Apply basic and advanced cell transformations
4. Deal with cells that contain multiple values
Crosswalking Tools
RDF Translator is a multi-format conversion tool for structured markup. It provides translations between data formats ranging from RDF/XML to RDFa or Microdata. The service allows for conversions triggered either by URI or by direct text input. Furthermore, it comes with a straightforward REST API for developers.

REST API

This on-line service provides an easily accessible API which allows for a couple of access methods:

1. Request raw code snippet served using the proper media type for the target data format:

   http://rdf-translator.appspot.com/convert/<source>/<target>/<uri>

Examples:

- URI, source data format, and target data format are given
- Input format is detected automatically
Your Website is Your API
Jan Zauha is defined/understood as an entity
The Linked Data Audience
Humans + Bots
Why LOD?

- Libraries as publishers in Web of Data
- **Reuse of our data = Significance**
- Can improve findability in search engines (search engine optimization)
- Levels of description in LOD enable entirely new applications
Resources


Koster, Lukas (2011) Brief Introduction to Linked Data (open access) https://docs.google.com/document/d/1W6UOCLgxFDyM0BIPfd5hs58dh4k6CUdLW354AjjtnJfk/edit

*See also LODLAM - Linked Open Data in Libraries, Archives & Museums at http://lodlam.net/*
Acknowledgements

I would like to thank Alison Hitchens for her excellent work in explaining and teaching these linked data concepts and definitions.

http://www.accessola2.com/olita/insideolita/wordpress/?p=60029

She was also gracious enough to distribute her work under the Attribution-Noncommercial-Share Alike 3.0 Unported license. And I am doing the same.
Questions?

twitter.com/jaclark
github.com/jasonclark
www.lib.montana.edu/~jason/talks.php