pinboard.in tag

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

@jaclark #lodlib
http://linkedjazz.org/
Overview

● What is Linked Open Data (LOD)?
● LOD in Practice
● Build a sample RDF
● LOD in Libraries
● Questions
What is Linked Open Data?

- Practice of publishing structured, machine actionable data with an open license for sharing and reuse.
What is Linked Data?

“a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried.”

http://en.wikipedia.org/wiki/Linked_data
What is Open Data?

“A piece of data or content is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and/or share-alike.”

http://opendefinition.org/
LOD in Practice

- Design Expectations for LOD
- Expressing and Encoding LOD
- Principles of Publishing LOD
Design Expectations for LOD

1. Use URIs as names for things
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*, SPARQL)
4. Include links to other URIs so that they can discover more things.

- Tim Berners Lee

http://www.w3.org/DesignIssues/LinkedData.html
Expressing and Encoding LOD

Resource Description Framework (RDF)

A structured data format for expressing relationships (descriptions) between things (resources).

http://www.xml.com/pub/a/2001/01/24/rdf.html
Expressing and Encoding LOD

The data model of RDF is based on the idea of a triple.

A triple is a simple statement with three distinct parts:
1. a subject
2. a predicate
3. an object
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Author</td>
<td>Publisher</td>
<td>OCLC #</td>
<td>ISBN</td>
<td>Notes</td>
</tr>
<tr>
<td>2 This house of sky: landscapes of a Western mind</td>
<td>Ivan Doig</td>
<td>New York: Harcourt Brace Jovanovich, 1992.</td>
<td>25629631</td>
<td>0151900558</td>
<td></td>
</tr>
<tr>
<td>3 To skin a cat: stories</td>
<td>Thomas McGuane</td>
<td>New York: Dutton/S. Lawrence, 1987.</td>
<td>15549047</td>
<td>0394755219</td>
<td></td>
</tr>
<tr>
<td>6 The horse whisperer</td>
<td>Nicholas Evans</td>
<td>New York: Delacorte Press, 1995.</td>
<td>35722048</td>
<td>0440222656</td>
<td></td>
</tr>
</tbody>
</table>
#5 “has ISBN” 0671776975
#5 “has author” Norman Maclean
(subject) (predicate) (object)
1. Use URIs as names for things
http://www.worldcat.org/oclc/1733412 (subject)

http://purl.org/dc/terms/creator (predicate)

http://dbpedia.org/page/Norman_Maclean (object)
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc ="http://purl.org/dc/elements/1.1/">
    <rdf:Description rdf:about="http://www.worldcat.org/oclc/1733412">
        <dc:creator rdf:resource="http://dbpedia.org/page/Norman_Maclean"/>
        <dc:identifier>ISBN 0671776975</dc:identifier>
    </rdf:Description>
</rdf:RDF>
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/
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/*
5 Star Linked Open Data

- make your stuff available on the Web (whatever format) under an open license
- make it available as structured data (e.g., Excel instead of image scan of a table)
- use non-proprietary formats (e.g., CSV instead of Excel)
- use URIs to denote things, so that people can point at your stuff
- link your data to other data to provide context

http://5stardata.info/
Build a sample RDF

1. Go to Maya Angelou’s Wikipedia page

2. Come up with 2 statements that can expressed as triples
Build a sample RDF

What are some triples you identified?

Final step:
Place a triple into RDF and test in W3C RDF Validation Service
A Graph Database

- A series of RDF triples that can be queried
- Expressions provide links between triples
- Non-hierarchical, arbitrary object relations
Bengie is a dog.  
Bonnie is a cat.  
Bengie and Bonnie are friends.  

Using these three simple statements, let's turn this into a data graph:

http://www.linkeddatatools.com/introducing-rdf
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
Introducing Graph Search

Photos I like

https://www.facebook.com/about/graphsearch
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 - Use Cases

- Web Scale Cataloging
  - Library of Congress BIBFRAME Initiative
  - http://www.loc.gov/bibframe/

- Annotated Texts and Books

- Recommendation Systems
  - http://linkedjazz.org/network/
Introducing Semantic Markup

- HTML5 semantic tags and RDFa that helps classify page types and 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
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
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

AM (After Microdata)
<dl>
<dt>Title</dt>
<dd>A River Runs Through It and Other Stories</dd>
<dt>Author</dt>
<dd>Norman Maclean</dd>
<dt>Publication date</dt>
<dd>October 1, 2001</dd>
<dt>ID</dt>
<dd>0226500667</dd>
</dl>
<dl itemscope itemtype="http://schema.org/Book">
  <dt>Title</dt>
  <dd itemprop="title">A River Runs Through It and Other Stories</dd>
  <dt>Author</dt>
  <dd itemprop="author">Norman Maclean</dd>
  <dt>Publication date</dt>
  <dd itemprop="pubdate">October 1, 2001</dd>
  <dt>ID</dt>
  <dd itemprop="isbn">0226500667</dd>
</dl>
Resources


Koster, Lukas (2011) Brief Introduction to Linked Data (open access) https://docs.google.com/document/d/1W6UOCLgxTyM0BiPfd5hs58dh4k6CUdLW354AjjtnJfk/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.
Are you too busy to improve?

No thanks!

We are too busy

Håkan Forss @hakanforss http://hakanforss.wordpress.com

This illustration is inspired by and in part derived from the work by Scott Simmerman, “The Square Wheels Guy” http://www.performancemanagementcompany.com/
Questions?

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