Data Designed for Discovery

Roy Tennant
Senior Program Officer, OCLC Research
A few introductory remarks

• This is the *Research* view of linked data
• We (OCLC) have experiments and prototypes, but no products or production services (yet)
• We (OCLC Research) have been working with linked data for as long as anyone in the library world
• Our (OCLC Research) playground is the entirety of WorldCat (380 million records) and a parallel computing cluster
• Stay tuned for more information on production services
WHY LINKED DATA?
What we have to work with
What we have to work with

- A collection of text strings...
- Taken from the piece itself...
- Sometimes “enhanced” with inferred parentheticals (e.g., [1975] )...
- Or additional statements not on the piece (e.g., subject headings)
- Punctuation, which may or may not be present, is used (*inconsistently*) for structure
- *Mostly uncontrolled and only loosely connected to anything else*
- *Designed for description rather than discovery*
THE PROBLEM
Actually, A Number of Problems

- Identification Problems (illustrated next):
  - The Title Problem
  - The Names Problem

- Quality Problems (illustrated next):
  - The Legacy Problem (strings are not controlled terms; often, they cannot be turned into them)

- Linkage Problems:
  - The Web Problem (records aren’t enough, you need links)
  - The Language Problem (showing the right translation for a given user)
<table>
<thead>
<tr>
<th>Num</th>
<th>Mark</th>
<th>TITLES (1-12 of 18)</th>
<th>Year</th>
<th>Entries Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Hamlet Prince Of Denmark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy As It Is Acted At The Theatres Royal In Drury</td>
<td>1776?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lane And Covent Garden By William Shakespear : Shakespeare, William,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy As It Is Acted At The Theatres Royal In Drury</td>
<td>1779</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lane And Covent Garden Written By Shakespeare, William,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy As It Is Now Acted At The Theatres Royal In Drury</td>
<td>1734</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lane And Covent Garden Written By Willia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy As It Is Now Acted By His Majestys Servants</td>
<td>1750</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Shakespeare, William,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy By Mr Wm Shakespear : Shakespeare, William,</td>
<td>1720</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy By Shakespeare As Performed At The Theatre</td>
<td>1777</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Royal Covent Garden Regulated From The Prompt B : Shakespeare, William,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy By Shakespeare As Performed At The Theatres</td>
<td>1800?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Royal Regulated From The Prompt Book By Permis : Shakespeare, William,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy By W Shakespeare : Shakespeare, William,</td>
<td>1770</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Hamlet Prince Of Denmark A Tragedy By William Shakespear Collated With The Best</td>
<td>1750</td>
<td></td>
</tr>
</tbody>
</table>
John Rock was an American obstetrician and gynecologist. He is best known for the major role he played in the development of the first hormonal contraceptive, colloquially called "the pill".

Born: March 24, 1890, Marlborough, MA
Died: December 4, 1984
Education: Harvard University, Harvard Medical School

People also search for:
- Gregory Goodwin Pincus
- Min Chueh Chang
- Katharine McCormick
- Carl Djerassi
- Margaret Sanger
## 245: Title Statement

You may request a report on a particular date.

### 1 Jan 2015

<table>
<thead>
<tr>
<th>Subfield</th>
<th>Number of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>91</td>
</tr>
<tr>
<td>i</td>
<td>1</td>
</tr>
<tr>
<td>j</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
</tr>
<tr>
<td>l</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
</tr>
<tr>
<td>q</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>188</td>
</tr>
<tr>
<td>t</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
</tr>
<tr>
<td>v</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td></td>
</tr>
</tbody>
</table>

### 1 Jan 2014

<table>
<thead>
<tr>
<th>Subfield</th>
<th>Number of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>91</td>
</tr>
<tr>
<td>i</td>
<td>1</td>
</tr>
<tr>
<td>j</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
</tr>
<tr>
<td>l</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
</tr>
<tr>
<td>q</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>188</td>
</tr>
<tr>
<td>t</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td></td>
</tr>
<tr>
<td>v</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td></td>
</tr>
</tbody>
</table>

### Overall number of Holdings: 2,211
THE SOLUTION
First, define ALL "Entities" Linked Data
Quick Definitions

**entity**

/ˈɛntɪti/

*noun*

a thing with distinct and independent existence.

**relationship**

/rɪˈleɪʃ(ə)nʃɪp/

*noun*

the way in which two or more people or things are connected
also known as "triples" about "entities"...establish relationships with...
...with actionable links from authoritative data hubs

Wikidata and VIAF

http://experiment.worldcat.org/entity/work/data/369081611
WorldCat Works

http://id.loc.gov/authorities/subjects/sh85101653.html
Library of Congress Subject Headings
The tipping point: how little things can make a big difference

An introduction to the Tipping Point theory explains how minor changes in ideas and products can increase their popularity and how small adjustments in an individual's immediate environment can alter group behavior.
<table>
<thead>
<tr>
<th>workExample</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://schema.org/workExample">http://schema.org/workExample</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/697286558">http://www.worldcat.org/oclc/697286558</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/150382268">http://www.worldcat.org/oclc/150382268</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/614497669">http://www.worldcat.org/oclc/614497669</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/301835491">http://www.worldcat.org/oclc/301835491</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/42463626">http://www.worldcat.org/oclc/42463626</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/633715391">http://www.worldcat.org/oclc/633715391</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/87987548">http://www.worldcat.org/oclc/87987548</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/438886367">http://www.worldcat.org/oclc/438886367</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/220075830">http://www.worldcat.org/oclc/220075830</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/22014811">http://www.worldcat.org/oclc/22014811</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/101931910">http://www.worldcat.org/oclc/101931910</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/470858303">http://www.worldcat.org/oclc/470858303</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/18632187">http://www.worldcat.org/oclc/18632187</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/902424325">http://www.worldcat.org/oclc/902424325</a></td>
</tr>
<tr>
<td><a href="http://www.worldcat.org/oclc/635195984">http://www.worldcat.org/oclc/635195984</a></td>
</tr>
</tbody>
</table>
OCLC releases WorldCat Works as linked data

DUBLIN, Ohio, 28 April 2014—OCLC has made 197 million bibliographic work descriptions—WorldCat Works—available as linked data, a format native to the Web that will improve discovery of library collections through a variety of popular sites and Web services.

Release of this data marks another step toward providing interconnected linked data views of WorldCat. By making this linked data available, library collections can be exposed to the wider Web community, integrating these collections and making them more easily discoverable through websites and services that library users visit daily, such as Google, Wikipedia and social networks.

"Bibliographic data stored in traditional record formats has reached its limits of efficiency and utility," said Richard Wallis, OCLC Technology Evangelist. "New technologies, influenced by the Web, now enable us to move toward managing WorldCat data as entities—such as 'Works,' 'People,' 'Places' and more—as part of the global Web of data."

OCLC has created authoritative work descriptions for bibliographic resources found in WorldCat, including the OverDrive collection. These works will be made available through the WorldCat Linked Data Service.
Zen and the art of motorcycle maintenance: an inquiry into values,

Author: Robert M. Pirsig
Database: WorldCat
Summary: Acclaimed as one of the most exciting books in the history of American letters, this modern epic became an instant bestseller upon publication in 1974, transforming a generation and continuing to inspire millions. A narration of a summer motorcycle trip undertaken by a father and his son, the book becomes a personal and philosophical odyssey into fundamental questions of how to live. The narrator's relationship with.

Rating: ★★★★★ based on 3 rating(s)
Read more...

More like this: User lists, Similar items

Find a copy in the library

Enter your location: 28372
Find libraries

Displaying libraries 1-6 out of 3370 for all 145 editions

<< First Prev 1 2 3 Next >>

Show libraries holding just this edition or narrow results by format
Acclaimed as one of the most exciting books in the history of American letters, this modern epic became an instant bestseller, transforming a generation and continuing to inspire millions. A narration of a summer motorcycle trip undertaken by a father and his son, the story grew into a philosophical odyssey into fundamental questions of how to live. The narrator's relationship with his son leads to a powerful self-reckoning that is both an austerely beautiful process for reconciling science, religion, and humanism. Resonant with the confusions of existence, Zen and the Art of Motorcycle Maintenance is an inspiring and touching and transcendent book of life.
OCLC’s linked data resources

WorldCat Works: 5 billion RDF triples

WorldCat Catalog: 15 billion triples

FAST: 23 million triples

ISNI: 10-50 million triples

VIAF: 2 billion triples
VIAF aggregates identifiers

Chomsky, Noam, 1928-....
Chomsky, Noam.
Chomsky, Noam (Noam Avram), 1928-
Noam Chomsky American linguist, philosopher
VIAF ID: 89803084 (Personal)
Permalink: http://viaf.org/viaf/89803084
ISNI: 0000_0003_8149_3996

Noam Chomsky
Noem חומסקי
نوم شومسكي
Noem Nom
Chemsky, Nom
Chomsky, Noom
Chemsky, Noom
Chomsky, Nom
Chomsky, Nome
Chomsky, Nom
Chomsky, Nom
Wikidata disseminates identifiers

Jerry Brown (Q152451)

VIAF identifier: 1320107

Authority control:
- WorldCat
- VIAF: 1320107
- LCCN: n80131844
- ISNI: 0000 0001 1487 0443
- GND: 119129352578
- ULAN: 500274877
OCLC’S 2015 INTERNATIONAL LINKED DATA SURVEY
SOURCE: KAREN SMITH-YOSHIMURA
2015 responding institutions by type

71 institutions total

- Academic library: 31%
- National library: 20%
- Network: 14%
- Government: 10%
- Scholarly: 8%
- Public Library: 7%
- Museum: 6%
- Other: 4%
- Other: 4%

Total: 71 institutions
What is published as linked data

- Authority files
- Bibliographic data
- Digital collections
- Descriptive metadata
- Encoded archival descriptions
- Geographic data
- Ontologies/vocabularies
- Data about museum objects
- Datasets
- Other
<table>
<thead>
<tr>
<th>2015 linked data sources most consumed</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIAF (Virtual International Authority File)</td>
<td>41</td>
</tr>
<tr>
<td>DBpedia</td>
<td>36</td>
</tr>
<tr>
<td>GeoNames</td>
<td>35</td>
</tr>
<tr>
<td>id.loc.gov</td>
<td>35</td>
</tr>
<tr>
<td>Resources we convert to linked data ourselves</td>
<td>17</td>
</tr>
<tr>
<td>Getty's AAT</td>
<td>16</td>
</tr>
<tr>
<td>FAST (Faceted Application of Subject Terminology)</td>
<td>15</td>
</tr>
<tr>
<td>WorldCat.org</td>
<td>15</td>
</tr>
<tr>
<td>data.bnf.fr</td>
<td>12</td>
</tr>
<tr>
<td>Deutsche National Bib Linked Data Service</td>
<td>12</td>
</tr>
</tbody>
</table>
SOLVING PROBLEMS & MOVING TOWARD A LINKED DATA FUTURE
Improving the Discovery Experience

Mockup
Eiseley, Loren C. 1907-1977

Loren Eiseley (September 3, 1907 – July 9, 1977) was an American anthropologist, educator, philosopher, and natural science writer, who taught and published books from the 1950s through the 1970s. During this period he received more than 36 honorary degrees and was a fellow of many distinguished professional societies. At his death, he was Benjamin Franklin Professor of Anthropology and History of Science at the University of Pennsylvania. He was noted as a “scholar and writer of imagination and grace,” which gained him a reputation and record of accomplishment far beyond the campus where he taught for 30 years. Publishers Weekly referred to him as “the modern Thoreau.” The broad scope of his many writings considered such diverse topics as the mind of Sir Francis

http://wikipedia.org/wiki/Loren_Eiseley
<table>
<thead>
<tr>
<th>Person</th>
<th>Organization</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>Event</td>
<td>Work</td>
</tr>
</tbody>
</table>
Exploring Ways to Use Linked Data

EntityJS  
War Between the States collection

Lincoln, Abraham, 1809-1865.

About
Abraham Lincoln (/æbrəˈhæm lɪnˈkən/) (February 12, 1809 – April 15, 1865) was the 16th President of the United States, serving from March 1861 until his assassination in April 1865. Lincoln led the United States through its Civil War—its bloodiest war and its greatest moral, constitutional and political crisis. In so doing he preserved the Union, abolished slavery, strengthened the federal government, and modernized the economy. Reared in a poor family on the western frontier, Lincoln was a self-educated lawyer in Illinois, a Whig Party leader, state legislator during the 1830s, and a one-term member of the Congress during the 1840s. He promoted rapid modernization of the economy through banks, canals, railroads and tariffs to encourage the building of factories; he opposed the war with Mexico in 1846. After a series of highly publicized debates in 1858 during which he opposed the expansion of slavery, Lincoln lost the

Related Works
As a Creator
Abraham Lincoln: miscellaneous papers, 1840-1864.
Abraham Lincoln campaign leaflet 1864.
Abraham Lincoln collection, 1824-1865.
Abraham Lincoln collection, 1847-1864.
Abraham Lincoln correspondence.
Abraham Lincoln Document, ca. early 1860s.
Abraham Lincoln note and photograph, 1863 August 20,.
Abraham Lincoln papers (at Chicago History Museum), 1800-[ongoing].
Abraham Lincoln papers, 1774-1948.

Related Media

Related People

Related Organizations

Related Places

Source: Wikimedia Commons
### Allegiant Editions

<table>
<thead>
<tr>
<th>Cover</th>
<th>Edition Details</th>
<th>Date</th>
<th>Language</th>
<th>Format</th>
<th>Audience</th>
</tr>
</thead>
</table>
| ![Allegiant Cover](image1) | **Allegiant**  
Veronica Roth.  
| ![Allegiant Cover](image2) | **Allegiant**  
Veronica Roth.  
| ![Allegiant Cover](image3) | **Allegiant**  
Veronica Roth.  
ISBN: 1410467848, 9781410467846. | 2014 | English  | Large print book | Adolescent |
| ![Allegiant Cover](image4) | **Allegiant**  
ISBN: 9780007569502, 0007538022, 0007569505,  
0007534949, 9780007538027, 9780062297335,  
9780007534944, 0062287338. | 2013 | English  | Book | General |
Offering the right translation
Offering the right translation
Bringing Authority Control to the Web
Prototyping New Services

- **Person Lookup Service** – An experimental service for looking up OCLC Person Entities

- **Scenario:**
  - A library wants to disambiguate a name
  - It sends the name text string to our API
  - We check all of our aggregated authority files and send back the best match(es)
  - Each response comes with one or more URIs (e.g., to LCNAF, Wikidata, ISNI, etc.)
  - The library inserts this data into their record, *turning a text string into an actionable link on the web*
In Summary: Why Linked Data?

- A better user experience
- Greater Web visibility
- Develop better models of resources not well served by current standards
- Replicate existing library functions more cheaply and efficiently
- Improve data integration
- Improve internal data management
EASING THE TRANSITION
Collaborating on BIBFRAME

• Working with the Library of Congress and others to finalize the BIBFRAME standard
• Beginning to explore what working with it at scale will mean
Working With the Web

• Modeling bibliographic data using Schema.org
• Collaborating on expanding the Schema.org with additional bibliographic elements at bib.schema.org
• Syndicating WorldCat data to search engines using Schema.org markup
Learning About Changing Workflows
COPY CATALOGING FOR NON-RARE BOOKS USING OCLC CONNEXION CLIENT

The following workflow represents the current work process in place at the UC Davis library when

ORIGINAL NON-RARE BOOK CATALOGING USING OCLC CONNEXION CLIENT

The following workflow represents the current work process in place at the UC Davis library when
Making MARC “Linked Data Ready”

Least machine-processable

If you must use free text:
- Use established conventions
- Use standardized terms

Algorithmically recoverable

- Use the most specific fields appropriate for a descriptive task
- Minimize the use of 500 fields
- Obey field semantics
- Avoid redundancy

Most machine-processable

- Use uniform titles
- Use added entries with role codes (7xx and $4)
- Use 041 for translations, including intermediate translations
- Use indicators to refine the meaning
Creating Standards for URIs

The Charge
How should URIs be added to MARC records to ease the transition to Linked Data?

Participants
• British Library, German National Library, Library of Congress, National Library of Medicine, OCLC.
• University libraries at Cornell, Columbia, George Washington, Harvard, Ohio State, Stanford, University of Washington
Summary Remarks

• We are in a major transition that will take YEARS to navigate
• We don’t know yet exactly what the future holds…
• ...but we know that it will be more linked and machine readable (actionable) than ever before
• And that’s a Good Thing
For More Information
Thank you!

Roy Tennant

@rtennant
tennantr@oclc.org

Together we make breakthroughs possible.