Linked Data and Open Data Licensing

George Buzash
Attorney, OCLC Legal Department

Karen Calhoun
VP Metadata, OCLC

Gordon Dunsire, Member, W3C Library
Linked Data Incubator Group (LLD XG)
Outline

1. A completely nontechnical introduction to linked data — Karen
2. Issues in making library metadata available as open linked data — Gordon
3. Thoughts about open data licensing and WorldCat Rights and Responsibilities for the OCLC Cooperative — Karen
4. Discussion
A COMPLETELY NONTECHNICAL INTRODUCTION TO LINKED DATA
You will be hearing more about research data management and linked data later today.

Simon Bains
Edinburgh University Library

Vic Elliott
The Australian National University
Tim Berners-Lee on linked data: February 2009

“Unlocking the potential of the Web”
“Raw data now!”
What would “unleashing the potential of the Web” with linked data look like?
The Open Linked Data Cloud, 22 Sept 2010

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch
http://lod-cloud.net/
Show me writers for the *New Yorker* born in Columbus, Ohio
Other artists born in Columbus > George Bellows > Collections in Columbus Museum of Art

Polo at Lakewood

www.columbusmuseum.org/collection/detail.php?id=46
Theoretically, you could then link to other information, for example, what libraries “know” about George Bellows and his work ...
Linked Personal Names

*** This page is a placeholder. Its main purpose is to serve links to the RDF representation and not yet to be visually pleasing or useful to a human. ***

Person: Strindberg, August, 1849-1912

Other names
- Strindberg, Johan August, 1849-1912
- Häved, Ulf, 1849-1912
- Strindberg, August Johan, 1849-1912
- Estrindberg, August, 1849-1912
- Strindberg, August, 1849-1912

DBpedia
- http://dbpedia.org/resource/August_Strindberg

Wikipedia
- https://sv.wikipedia.org/wiki/August_Strindberg

VIAF
- http://viaf.org/viaf/54154627

Depiction
- http://libris.kb.se/images/August_Strindberg.jpg

misc.
- text.rdf+n3
- application/rdf+xml
- OpenLink RDF browser

LIBRIS

VIAF

OCLC
The world's libraries. Connected.
GLOBAL COUNCIL
What is the problem we are trying to solve?

• The methods used to organize Web content are *inefficient* and *fail to optimize* the intellectual effort invested in them.

• Collections of interest to the communities that libraries serve are *underutilized* because they are *not visible enough* on the Web.

• The methods used by libraries to organize collections (authorities, vocabularies, classification—in all their language forms) are *underutilized* because they are bound to library collections (but not the larger Web).

Is linked data the solution?
Charter of the W3C Library Linked Data Incubator Group (LLD XG)*

http://www.w3.org/2005/Incubator/lld/charter

To help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web — focusing on Linked Data—in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future.

Source: Presentation by Antoine Isaac, 21 July 2010.
www.slideshare.net/antoineisaac/w3c-library-linked-data-incubator-group

*OCLC is among 10 W3C initiating members of the XG
ISSUES IN MAKING LIBRARY METADATA AVAILABLE AS LINKED DATA
Will a linked data approach be adopted by enough libraries to justify investment in developing linked data systems?

- Focus shifts from whole records to “atomic” metadata
  - Improves share-ability; reduces duplication
- Requires data reformat and system reengineering
  - Vendors will do this (only) if there is market demand
- We have done this before (card to MARC; UKMARC to MARC21, etc.)
  - And we replace/renew systems every 5-10 years
Who will build linked data services and applications from linked data sets?

- Not just vendors and libraries with development infrastructure
  - Broader cultural heritage/memory communities
  - Other communities who find our data useful
  - Individuals and organizations with an “open” approach

- Innovative applications
  - Beyond current (low) expectation and imagination
  - Fewer barriers = more action
If many libraries release their catalogs as linked data, who will deduplicate the data, and how?

- Duplication identified by professionals, trusted amateurs (baby-boom retirees), machines, end-users
- Data and agents from other communities will help
- Identify and tag as “same as”
  - Once only (by a human) for any pair of records
  - Thence on the fly (by machine) for atomic data
- Nothing should be deleted
Who owns legacy catalog records? What will encourage libraries to release their records as open linked data?

- Ownership is complex
  - Create, aggregate, copy, amend, re-aggregate, …
  - May be impossible to determine legally
  - Clearer for unique resources (special collections)
- Little or no extrinsic value in most records
  - Who wants to buy, other than libraries?
- Libraries very cautious about legal issues
  - Waive IPR and a great barrier is removed
SOME THOUGHTS ON OPEN DATA LICENSING
Some members beginning to express interest in releasing their catalogs as open data

Dear OCLC,

What are OCLC’s views on open data licensing?

Yours truly,

A member library
WorldCat Record Use Policy: A new approach

- New policy created by members, reviewed by the community, and implemented 1 August 2010: [www.oclc.org/worldcat/recorduse/](http://www.oclc.org/worldcat/recorduse/)
- Scope = copies of WorldCat records—not the WorldCat database itself
- A code of good practice for members of a cooperative based on shared values, trust and reciprocity in understanding rights and responsibilities—instead of data ownership, detailed provisions or restrictions
- Outlines rights to transfer data to individuals, consortia and public agencies, other libraries and scholarly institutions, (members or nonmembers), and third parties
  - Machine-to-machine access to a member‘s records okay (linked data, Z39.50, SRU/SRW, OAI, etc.); see FAQ #3. [www.oclc.org/worldcat/recorduse/policy/questions/](http://www.oclc.org/worldcat/recorduse/policy/questions/)
Members asked to exercise their rights in the context of OCLC community norms laid out in WCRR

- A balancing act: enable BOTH broad sharing AND sustainability of WorldCat

- “While, on behalf of its members, OCLC claims copyright rights in WorldCat as a compilation, it does not claim copyright ownership of individual records.”

- Reliance on self-governing behaviors

- Members asked to ensure awareness of WCRR to organizations to which they make their data available.
Open Data Licenses we considered: “Conformant” Licenses

<table>
<thead>
<tr>
<th>License</th>
<th>Domain</th>
<th>By</th>
<th>SA</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Data Commons Public Domain Dedication and Licence (PDDL)</td>
<td>Data</td>
<td>N</td>
<td>N</td>
<td>Dedicate to the Public Domain (all rights waived)</td>
</tr>
<tr>
<td>Open Data Commons Attribution License</td>
<td>Data</td>
<td>Y</td>
<td>N</td>
<td>Attribution for data(bases)</td>
</tr>
<tr>
<td>Open Data Commons Open Database License (ODbL)</td>
<td>Data</td>
<td>Y</td>
<td>Y</td>
<td>Attribution-ShareAlike for data (bases)</td>
</tr>
<tr>
<td>Creative Commons CCZero</td>
<td>Content, Data</td>
<td>N</td>
<td>N</td>
<td>Dedicate to the Public Domain (all rights waived)</td>
</tr>
</tbody>
</table>

http://www.opendefinition.org/licenses/
You are free:

To Share: To copy, distribute and use the database.

To Create: To produce works from the database.

To Adapt: To modify, transform and build upon the database.

Provided:

You attribute in the manner specified in the license.

http://www.opendatacommons.org/licenses/by/summary/
ODC Licensing FAQ: Distinguishing between a database and its contents

WHY DO YOU DISTINGUISH BETWEEN THE “DATABASE” AND ITS “CONTENTS”? 

The simplest answer is because they may have separate rights. For example, consider a database of photographs. Here there are the rights in the database and quite separate individual copyrights in the photographs. Or consider the example of Freebase which contains textual material and images from Wikipedia as well as user contributed material. While Freebase controls the database the individual items of contents need to have their own separate license.

Of course much of the time the Licensor of the database is also in the position to license the

Our preliminary thoughts on Open Data Licensing

- We are considering recommending ODC-BY
- Distinguishes between the database and its contents (or portions of contents)
- [Member DATABASE NAME] would be the name of the member’s or group’s catalog, and the member or group = the licensor
- License notice wording in accordance with instructions in ODC licenses
- Still under investigation—your input invited and welcome
Discussion