Linked Open Data

Robin Murray  
Vice President, Global Product Management

Jim Michalko  
Vice President, Research Library Partnership
Session Overview

• Presentations 1:15 - 2:15
  - Linked Data within the OCLC Strategy Robin Murray
  - The proposed approach to open licensing Jim Michalko
  - Open questions.

• Table discussions 2:15 - 3:15
  - Set topics + anything else...

• BREAK 3:15 - 3:30

• Report back 3:30 - 4:15
Linked Data within the OCLC strategy

Robin Murray
Vice President, Global Product Management
Overview

- Linked Data Background & Value Proposition
- Linked Data within OCLC product strategy
  - WorldShare Platform
  - Data Exposure Services
  - Presenting a coherent web of library data
- Linked Data within the OCLC business model
  - Opportunity or threat?
- Open Data Licensing Proposal - Jim
Linked Data - Background

The World-Wide-Web
- Web of Documents / Pages
- Structured & linked by HTML
- Pages and Links Human Readable

Link Data
- Web of Data
  - about ‘things’
- Structured & linked by RDF (Resource Description Framework)
- Data and Links Machine Readable
  - Requires Additional Structure: Schema, Vocabularies, Identifiers
Linked Data - Value

- Enables machine processing of data for new applications
- Atomic data is more ‘linkable’ and more re-usable
  - Relates ‘data fragments’ to real things
  - Link data fragments about the same ‘thing’
- Links create value
  - Builds ‘super-records’
  - Links generate Search-Engine-Optimization
- Enables ‘collective innovation’ on data
- Makes Data Work Harder
A trivial example...

- Authority control in a web of data...
Value Proposition - Authority Control - Currently...

Current authority control copies in well-formatted string of characters to the destination record/entry. Value proposition - well-formatted data (for those who care...)

Rupert Murdoch
But... in a “Web of Data”...

With consistent & coherent data infrastructure headings have explicit navigational links that can be exposed and traversed.

(subset of links in WorldCat Identities)
Authority Control in a “Web of Data”...

Copy string of characters and create navigation links to “web of data”.

Value Proposition:
Well-formatted data
AND
Automatic data enrichment and navigation

Rupert Murdoch

Web of Data

Australia Biography

News Corporation

Geoff Pryor

Fox News

James Murdoch

Names

Outfoxed Rupert Murdoch’s war on journalism

Business the Rupert Murdoch Way...

The man who owns the news...

...
Murdoch, Rupert, 1931-....
VIAF ID: 85338033 (Personal)
Formal link: http://viaf.org/viaf/85338033

Preferred Forms

4x4's: Alternate Name Forms (3)

Selected Titles

About

Personal Information
Nationality:
GB - United Kingdom
AU - Australia

External Links
Wikipedia - Rupert Murdoch
WorldCat Identifiers

Record Views

<foaf:name>Nag Copaleen, Myles</foaf:name>
<foaf:name>Copaleen, Myles</foaf:name>
<rdaGr2:dateOfBirth>1911-10-10</rdaGr2:dateOfBirth>
<rdaGr2:dateOfDeath>1966-04-01</rdaGr2:dateOfDeath>
<owl:sameAs rdf:resource="http://dcpedia.org/resource/Brian_O'Nolan"/>
<owl:sameAs rdf:resource="http://libris.kb.se/resource/auth/196381"/>
<owl:sameAs rdf:resource="http://www.idref.fr/027051013/id"/>
<owl:sameAs rdf:resource="http://d-nb.info/gnd/119012871"/>
Linked Data within OCLC product strategy

- The WorldShare Platform
- Data Exposure Services within the WorldShare Platform
- Presenting a coherent web of library data
The core infrastructure underlying all OCLC applications.

Designed from the ground up for Web-Scale operations

Open to the community to enable “Collective Innovation” and Sharing
Major focus has been at the application layer.

For Data Exposure Services and Linked Data the focus is here.
Web Services

Business Logic Services

Data Exposure Services

Data

Dewey  ... Terminologies  Identities/VIAF  WorldCat

Categories of Web Services

Core Data Assets

WorldCat Identities /VIAF

Terminologies

Dewey

WorldCat

Categories of Web Services

Core Data Assets
Data Exposure Services

Provides publicly consumable interface for exposing data in a variety of formats

WorldCat API

- FY10: 2.5M / month
- FY11: 9.5M / month
- FY12: 15M / month
Linked data / RDF representation is just another data format to express records in. Make this an option for the service.

RDF provides a structured way to explicitly represent links between data objects. Library data has many implicit links. Making these explicit has obvious value...

Data Exposure Services

WorldCat API

Bib RDF

links

Data

Dewey... Terminologies Identities /VIAF WorldCat
Expose other data sets in the same way and provide links between each data set to produce a global web of linked data.
Linking User Interfaces / Pages...

Linked Data

Data Exposure Services

Dewey API
... API
Terminlogy API
Identities API
WorldCat API

Data

Dewey
... Terminologies
Identities /VIAF
WorldCat

Things

Concepts
... Names
Works
Manifestations

RDF representation points to worldcat.org representation of the record
Linking User Interfaces / Pages...

Linked Pages
- Dewey Pages
- ... Pages
- Terminology Pages
- Identity Pages
- WorldCat.org Bib records

Linked Data
- Dewey RDF
- ... RDF
- Terminology RDF
- Identity RDF
- Bib RDF

Data Exposure Services
- Dewey API
- ... API
- Terminology API
- Identities API
- WorldCat API

Data
- Dewey
- ... Terminologies
- Identities / VIAF
- WorldCat

Things
- Concepts
- ... Names
- Works
- Manifestations

RDF representation for all data sets point to a page representation. The page provide user navigation between data sets.
License Terms...

Linked Pages
- Dewey Pages
- ... Pages
- Terminology Pages
- Identity Pages
- WorldCat.org Bib records

Linked Data
- Dewey RDF
- ... RDF
- Terminology RDF
- Identity RDF
- Bib RDF

Data Exposure Services
- Dewey API
- ... API
- Terminology API
- Identities API
- WorldCat API

Data
- Dewey
- ... Terminologies
- Identities / VIAF
- WorldCat

Things
- Concepts
- ... Names
- Works
- Manifestations
Consistent, Coherent and Comprehensive data infrastructure, suitable for public consumption.
In existence today...

Requires
- Greater coherence & consistency
- Completing against all entities

Data Exposure Services
- Dewey API
  - ... API
  - Terminology API
  - Identities API
  - WorldCat API

Data
- Dewey
  - ... Terminologies
  - Identities /VIAF
  - WorldCat

Things
- Concepts
- ... Names
- Works
- Manifestations

Online License Terms

Requires
- • Greater coherence & consistency
- • Completing against all entities

OCLC - The world's libraries. Connected.
WorldShare Platform: Data Exposure Services - WORK IN PROGRESS

- Data Modeling across all entities
- Consistency & Coherence of
  - API Style / RDF representation style / URI’s / UI styling and linking / Syndication and search engine exposure / ...
- Licensing approach
- ➔ Data exposure services including Linked Data
  Will roll-out incrementally with continuing enhancements
Linked Data within the OCLC Business Model...

- Core Data Infrastructure Value Proposition

- Business Model: Access to the WorldShare Platform

- Linked Data - Threat or Opportunity?
Core Data Infrastructure Value Proposition...

- **Expose & Syndicate** - expose data services for public consumption through API’s linked data, syndication and applications.

- **Build Business Logic Services** - deliver business logic services based on underlying data.

- **Mine, Extend & Build Orthogonal Data Sets** - build derived data sets such as identities, terminologies...

- **Structure** - build additional structure to support discovery and navigation - build works, manifestations, merge holdings...

- **Enhance & Augment** - enhance sparse records by merging attributes from related records. Add attributes from augmented evaluative data sets such as reviews, ratings, etc.

- **Aggregate** - acting upstream in data supply chains, aggregate data across print, licensed and digital.

1. All this is required for Coherent & Comprehensive Linked Data Exposure.

2. There is a substantial cost of this activity - shared by the co-operative.

3. This picture is just another representation of the WorldShare Platform.
Business Model: Access to the WorldShare Platform...

- A “Platform view” of current services:
  - Current FirstSearch can be viewed as an “Institutional Data discovery and exposure license”
    - Linked data is just another exposure mechanism
  - Current Cataloging can be viewed as “Institutional data management, registration, supply and re-use license.”
    - Linked data is just another format
- The combination is a substantial section of the WorldShare Platform...
Business Model: Access to the WorldShare Platform...

- If you subscribe to Cataloging and FirstSearch then you have access to the WorldShare Platform:
  - Linked Data exposure
  - Data Services and Business Logic Services
  - Access to App Gallery and Apps.
  - Benefit from the collective innovation of the co-operative...
  - ...
Opportunity or Threat?

- IF: collectively we can re-envision cataloging as ‘registering nodes in a global web of data’.

- AND: we can position WorldCat as the trusted, global web of library data

- THEN: This dramatically increases the global value and utility of metadata management with WorldCat

- Significant Opportunity to ‘Grow the Denominator’
Linked Open Data

• Important to OCLC; shape its future services

• Lots of experimentation
  • Individual institutions e.g. University of Cambridge
  • Secondary aggregators e.g. Europeana, DPLA

• Asked for guidance
  • How release their catalog data consistent with WCRR?
  • N.B. Member data (particularly re-use and transfer of their WorldCat data to non-members)
Final phase of process

One year ago...
Introduction
Issues
Early thoughts
One year ago

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

ODC-BY = Open Data Commons - Attribution License

Inputs and interaction since then:
## Interactions and Implementations

### OCLC consultations:
- Management Team
- Board of Trustees
- (former) Record Use Policy Council
- Multiple member institutions

### Implementations:
- OhioLINK
- FAST
- VIAF
- University of Cambridge (COMET)
Review

An implementation step that balances

- WorldCat Rights and Responsibilities for the OCLC Cooperative (WCRR)
- Rights over WorldCat
- Expectations of confident re-use of released data by the broad community of web developers
WCRR - the big responsibilities

- abide by the WC R&R policy [3.B.1];
- ensure awareness of the WC R&R by those with access to WorldCat [3.B.1];
- give attribution to OCLC when making use of substantial amounts of WorldCat data [3.B.4];
- not engage in mass distribution of WorldCat data directly to non-members without OCLC's permission [3.B.6.3];

AND
“to make reasonable efforts to ensure that the subsequent re-use and transfer of their WorldCat data by non-members is consistent with this policy and OCLC's public purposes and supports the long-term viability and utility of WorldCat” [3.B.4].
OCLC Rights and WorldCat

- OCLC does not have contractual right to an OCLC member’s data
- No service agreement conditions imposed
- No intellectual property claim in individual bib records
- A copyright claim in the entire aggregation
  - Based on compilation and value-added investment

Currently

- Nothing legally enforceable
- WC R&R is code of good practice; community norms
- Issues called before Global Council
Expectations of re-use

- Availability that anticipates innovative re-use and fuller integration into the broader web environment
  - Including commercial sites
- Best practice provides data along with a license that clearly sets out the terms under which the data is being made available
  - Without a license data won’t be re-used
  - Can signal the norms of the community that created the data
Arriving at a recommendation

- Attribution (per WCRR - attribution)
- Awareness (per WCRR - community norms)
- Asserts compilation copyright (per WCRR - sustainability)
- Meets expectations (per web best practice - Open Data)

**YIELD**

Open Data Commons Attribution License
with reference to community norms i.e. WCRR
Why?

Prefer ODC-BY because

• Distinguishes between the database and its contents (or portions of content)
  • No Creative Commons license deals with database

No commercial restriction

• Search engines
• Commercial traffic back to library
• Non-commercial is not open
• It’s the emerging community practice
ODC-BY Plain Language Summary

This is a human-readable summary of the ODC-BY license. Please see the disclaimer below.

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

As long as you:

Attribute: You must attribute any public use of the database, or works produced from the database, in the manner specified in the license. For any use or redistribution of the database, or works produced from it, you must make clear to others the license of the database and keep intact any notices on the original database.

http://opendatacommons.org/licenses/by/summary/
Recommendation on Open Data Licensing for WorldCat data

• ODC-BY - Open Data Commons Attribution License
• Request attribution for member institution and OCLC
• Provide linkage to OCLC Community Norms
  • For now, the WCRR document
  • Desirable: plain language summary for non-OCLC members
Facilitates Linked Open Data work

- All the contemplated interactions in a LOD world are supported by this approach
- Will automatically be updated as Open Data Commons offers enhancements

Next steps

- Explicit Advice, Discussion and feedback from Global Council
- Board Approval
- Implementation
  - Documentation, toolkit, and FAQ on website
  - News announcement of recommendation
Some examples

- **COMET** (University of Cambridge)
- **OhioLINK** (OCLC and OhioLINK)
- **VIAF** (OCLC and contributors)
http://viaf.org/viaf/data/#norms

Use of VIAF URIs:

- To facilitate interoperability, OCLC strongly encourages the use of VIAF URI's in all appropriate circumstances. The canonical structure of a VIAF URI is http://viaf.org/viaf/[numerical value] (Example: http://viaf.org/viaf/49224511)

Compliance with Section 4.3 of the ODC Attribution License

- Adherence to ODC Attribution instructions for the correct assertion of attribution is encouraged. The preferred form of attribution for VIAF is: "This [title of report or article or dataset] contains information from VIAF (Virtual International Authority File) which is made available under the ODC Attribution License."
- Special cases: In circumstances where providing the full attribution statement above is not technically feasible, the use of canonical VIAF URIs is adequate to satisfy Section 4.3 of the ODC Attribution License.

Files Available

void dataDump http://viaf.org/viaf/data/viaf-20120219-links.txt.gz
Questions?
Table Discussions: 
- 3:15
Suggested Topics

• What Opportunities do you see for libraries and OCLC in Linked Data?

• How well does the approach support these opportunities?

• MOST IMPORTANTLY - The board is seeking explicit advice regarding the Open Data License recommendation from Global Council. In particular is this an implementation step consistent with the approved community norms? - either provide this or propose a mechanism to achieve this.

Appoint a reporter