Metadata Advisory Group: Setting Sail in a Changing Landscape

Dr. Axel Kaschte
Product Strategy Director, OCLC EMEA

Richard Wallis
Technology Evangelist, OCLC

Barbara Block
Head of Department for Library Services,
Gemeinsamer Bibliotheksverbund (GBV), Göttingen, Germany
Globalization

Dr. Axel Kaschte
Product Strategy Director, OCLC EMEA
Globalization

Global Library Network ↔ Global Information Network
Global Library Network
WorldCat Globalization

To extend the reach of WorldCat as the core metadata management tool

What: Country specific rules and interfaces and support more languages

Why: Support new collections & workflows (E,P,D) and transition of legacy systems

When: In progress
The Global Data Network

- Industry trend: metadata management becomes global
- Game changer: more data comes from providers than from libraries
- Game changer: more end users search the Web than the libraries
Globalization through Localization

<table>
<thead>
<tr>
<th></th>
<th>NL</th>
<th>GER</th>
<th>CAN</th>
<th>SPA</th>
<th>ITA</th>
<th>FRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorities</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Linking bib-bib bib-auth</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UNICODE Extensions</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UNIMARC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Globalization

Global Library Network ↔ Global Information Network
The Internet of Things & Library Metadata
The Internet of Things & Library Metadata

Linked Data & Library Metadata
Survey about linked data in libraries

OCLC Research Library Partners – Metadata Managers Focus Group

QUESTION SET 1

SHIFT TO LINKED DATA FOR PRODUCTION
Context Shift to Linked Data for Production

- **Linked data** may well be the next common infrastructure both for communicating library data and embedding it into the fabric of the semantic web...
- Much of a research library’s routine production is tied directly to its **ILS**...
- As libraries plan for this shift in infrastructure, what are the **key questions** that must be discussed and resolved?
Questions Shift to Linked Data for Production

a) Are you experimenting with or planning a shift to linked data for some workflows? If so, what workflows are under consideration?

b) What are the most pressing issues you feel must be considered?

c) What plans do you have for analyzing your workflows to understand the requirements for the next-generation data models?
39 responses from:

• Australian National University
• Brigham Young University, Utah
• British Library
• Brown
• Case Western Reserve
• Columbia
• Cornell
• George Washington University
Australian National University

a) Not yet

b) Efficient workflows – radical change needed. Interested in linking research data management and traditional research outputs

c) New workflow for digitization of theses

Answers
a) Experimented with BNB as LoD and looked at schema.org: there isn’t enough evidence yet to support a business case.

b) Infrastructure and tools

c) Metadata has to be capable of responding to changed functional requirements and technologies
a) We will be watching the LC and NLM BIBFRAME pilots. Meanwhile we are populating our existing data with identifiers that will make them “linked data ready”.

b) library data interaction with data outside traditional library boundaries

c) Authority work, e.g. using ISNI
New York University

a) currently developing a linked data user story for implementing authority control

b) basic end-user facing systems need to be the current focus of development

c) We are collecting requirements to improve our various discovery systems.
We would more than likely rely on our partners like the CDL, OCLC, or DPLA to provide our data in an aggregate linked data data set. As tools move towards linked data, so will we but we are really looking to our partners to do the heavy lifting for now.

UC Berkeley

Answers

a) Are you experimenting?
b) Most pressing issues?
c) What to analyze?
The aim of the OCLC Metadata Advisory Group (MAG) is to evolve global metadata management, in a way that it helps libraries and partners in all regions providing their services more efficiently and with better quality, by providing tactics & strategies as input to OCLC.
The Roles of the MAG

- Find consensus for discussed topics within the group
- Formulate unified recommendations for OCLC metadata services

What Topics?
MAGnificent changes

Barbara Block
Gemeinsamer Bibliotheksverbund (GBV), Göttingen, Germany
Six library networks for scientific libraries in Germany

The *Gemeinsamer Bibliotheksverbund (GBV)* is the common library network of seven German federal states and the Foundation of Prussian Cultural Heritage.
Gemeinsamer Bibliotheksverbund

450 members

- University, college and government sponsored research libraries within the territory of the participating federal states and of the Prussian Cultural Heritage Foundation

- Government libraries, public libraries and special libraries like TIB/UB Hannover and ZBW Kiel
Cataloguing in GBV: Basics

- CBS database

<table>
<thead>
<tr>
<th></th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>title records with holdings</td>
<td>39,053,000</td>
</tr>
<tr>
<td>holdings</td>
<td>97,042,000</td>
</tr>
<tr>
<td>GND</td>
<td>5,218,000</td>
</tr>
</tbody>
</table>

- Cataloguing format: Pica3, with title-title-linking, title-authority-linking

- Cataloguing tool: WinIBW

- Different cataloguing rules in the course of time, shift to RDA in 2015/2016
Cataloguing in GBV: Cataloguing format
Cataloguing in GBV: Electronic media

- The number of titles and holdings of electronic documents is increasing rapidly

<table>
<thead>
<tr>
<th>New titles and holdings</th>
<th>2012 Titles</th>
<th>2012 Holdings</th>
<th>2013 Titles</th>
<th>2013 Holdings</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-books</td>
<td>460.000</td>
<td>1,472.000</td>
<td>513.000</td>
<td>1,590.000</td>
</tr>
<tr>
<td>e-books</td>
<td>429.000</td>
<td>1,859.000</td>
<td>863.000</td>
<td>4,536.000</td>
</tr>
<tr>
<td>p-journals</td>
<td>21.000</td>
<td>74.000</td>
<td>18.000</td>
<td>55.000</td>
</tr>
<tr>
<td>e-journals</td>
<td>14.000</td>
<td>570.000</td>
<td>16.000</td>
<td>630.000</td>
</tr>
</tbody>
</table>

- 95% of e-title records are imported from publishers or vendors

- 90% of e-holdings are generated by offline programs (e-book packages, patron driven acquisition (PDA))
The aim of the OCLC Metadata Advisory Group (MAG) is to evolve global metadata management, in a way that it helps libraries and partners in all regions providing their services more efficiently and with better quality, by providing tactics & strategies as input to OCLC.
The future of cataloguing: Issues in GBV (1)

- The amount of manually catalogued data decreases
- The amount of data from publishers, vendors etc. increases
- The cataloguer becomes metadata manager
  - Needs new skills (and new tools)
  - Manages entities and identifiers

MAG – results of discussion

- Education of staff in the evolving metadata landscape so that the staff can create high quality linked metadata
- Use and reuse of global identifiers (instead of text strings) to describe efficiently resources with high quality
The Future of Cataloguing in GBV: Issues (2)

- The cataloguing platform is only a part of a complex information infrastructure
- Discovery needs high quality metadata
  - Use of global identifiers (e.g. ISNI, ORCID, VIAF)
  - Content clustering (FRBR, work entities)
  - Linked open data

MAG – results of discussion

- Use and reuse of global identifiers (instead of text strings) to describe efficiently resources with high quality
- Populate the metadata with links that go beyond the library catalogue to improve exposure
GBV: Search for new ILS

ILS to replace LBS3/LBS4
- From software to service
- Metadata workflow between cataloguing system and ILS
- Electronic resource management

MAG – results of discussion
- Understand metadata structures and systems to make reasonable decisions about future tools
GBV: Improvement of the visibility of a library

- In the funding organization
  - Collections
  - Services

MAG – results of discussion

- Tools that identify the distinctiveness in the collections and the contributors
- Make the library the central metadata hub for the institution
  - Scientific output
  - Connection to research information systems
GBV: Improvement of the visibility of a library

- In the international research community
  - Web presence
  - Linked data

MAG – results of discussion

- Offer the possibility to find connections and make connections between different works, peoples, projects, ... entities to create new ideas, recognize structures ...
MAG`s actions 1

- **Step 1:**
  Identify the global identifiers:
  Start from schema.org and identify the entities for which librarians should care and for which global identifiers should be provided

- **Step 2:**
  Describe the scope and role of the future cataloguer/metadata manager in an entity based world
MAG ’s actions 2

- **Step 1:**
  Initiate a study about researchers’ workflows and the systems they are using

- **Step 2:**
  Role of the library
  - Relevant entities (outside of common library entities)
  - Use cases
OCLC’s tasks

- Define business model for identity services
- New role model between individual and national libraries, regional and global service providers
  - Who is doing what?
  - What kind of central services can libraries rely on?

Additional questions

- Use of (national) authority files in WorldCat
- Use of authority files in WorldCat Local
- Merge of metadata from different sources, e.g. national bibliographies, regional catalogues or vendors in WorldCat
- Synchronization with WorldCat
Metadata Advisory Group: Making Library Collections Discoverable on the Web

Richard Wallis
Technology Evangelist, OCLC
Trend 1: Shift from collections to services

**Explain:** Libraries shift their focus from collections on the shelf to the services they offer. Pressure on collection budgets.

Trend 2: Search for Distinctiveness

**Explain:** As licensed content is available in commodity collections, academic libraries strive to declare their distinctiveness. Rise of OA.

Trend 3: Be Found on the Web

**Explain:** As users search on the web, libraries want their commodity and distinctive collections to be found on the web.
Landscape: Vision for the Ecosystem

"DBpediaLogo" by The DBpedia Team Licensed under Public domain via Wikimedia Commons
The Library Data Revolution

THE LIBRARY KNOWLEDGE GRAPH
Umberto Eco

Born: January 5, 1932, Alessandria, Italy
Spouse: Ranita Ramgo (m. 1962)
Movies: The Name of the Rose
Parents: Giulio Eco, Giovanna Eco
Awards: Strega Prize, Prix Médicis étranger

Get updates about Umberto Eco
Keep me updated

Books

- The Name of the Rose
  - 1980
- Foucault's Pendulum
  - 1988
- The Prague Cemetery
  - 2010
- The Island of the Day Before
  - 1954
- Baudozino
  - 2000

Subjects

- Semiotics
- Religions
- Thought and Thinking
- Art—Philosophy
- Foucault’s Pendulum
- Vodou

Quotes

“The real hero is always a hero by mistake; he dreams of being an honest coward like everybody else.”

Find Umberto Eco works at:

Libraries near me | Online Retailers
Library data stored as entities
The library knowledge graph

A graph of relationships
The library knowledge graph

Works
The library knowledge graph

Putting entities in library workflows
The Library Data Revolution

ENTITIES AND LIBRARY WORKFLOWS
The year is 1327. Franciscans in a wealthy Italian abbey are suspected of heresy, and Brother William of Baskerville ... mission is suddenly overshadowed by seven bizarre deaths that take place in seven days and nights of apocalyptic terror.

Monastic libraries --Italy –Fiction
Semiotics --Fiction
Entities and library workflows

Cataloging

Improve data quality
  • Link to authoritative sources

A new approach to cataloging
  • Point and click cataloging
  • Managing entities instead of managing records

Consistent with RDA
Entities and library workflows

Web exposure

Be found on the web
Connect your users to unique content

What the web requires for web exposure

• Aggregation
• Familiar structures
• A Network of Links
• Entity Identifiers
The Library Data Revolution

WHAT’S HAPPENING
15% of Web use Schema.org markup!
Schema Bib Extend Community Group

The mission of this group is to discuss and prepare proposal(s) for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing. The group will seek consensus around, and support for, proposal(s) to the W3C WebSchemas Group. This Community Group will not, itself, produce technical specifications.

Reports

No reports yet published. The Chair is responsible for publishing reports. More about publishing...

News

Add new post →

Recipes and Guidelines

Richard Wells

Recipes and Guidelines section is being added to the Group Wiki.

Check it out for an introduction to using Schema.org mark up in general, but with an emphasis on bibliographic data. Included are examples for typical bibliographic situations such as library holdings and citations.

Participants

Richard Wells
Chair

View all 80 Participants →

http://www.w3.org/community/schemabibex/
Model things of interest to the web.

Make those things available via structures familiar to the web.

Improve library workflows.

OCLC’s Approach to Data Modeling

http://www.BiblioGraph.net
The Library Data Revolution

ENTITIES AND WORLDCAT
The Work Entity

- **197+ million Work descriptions and URIs**
- Schema.org + BiblioGraph.net
- RDF Data formats
  - RDF/XML, Turtle, Triples, JSON-LD
- Links to WorldCat manifestations
- Links to Dewey, LCSH, LCNAF, VIAF, FAST
- Open Data license via Linked Data Explorer
  - 2015: Discovery API, Metadata API
- Released April 2014
The Person Entity

- **98+ million Person descriptions and URIs**
  - Person entities with authority: 20.2 million
  - Person entities without authority: 78.3 million
- **Schema.org + BiblioGraph.net**
- Harvested from WorldCat data and enriched from other hubs
- **RDF Data formats**
  - RDF/XML, Turtle, Triples, JSON-LD
- **Links to WorldCat Works. Added links from WC Works.**
- **Open Data license via Linked Data Explorer**
  - 2015: Linked Data Explorer, Discovery API
Gandhi, an autobiography : the story of my experiments with truth /
Gandhi, an autobiography: the story of my experiments with truth /
Mohandas Karamchand-Pensiero Gandhi

- type
- additionalName
- alternateName
- birthDate
- deathDate
- familyName
- givenName
- name
- sameAs

http://schema.org/sameAs

http://opac.kb.nl/PPN?PPN=068712030
http://id.worldcat.org/fast/220915
http://viaf.org/viaf/71391324
http://id.loc.gov/authorities/subjects/n79041926
http://id.worldcat.org/fast/33892
http://viaf.org/viaf/4230936
http://id.loc.gov/authorities/subjects/n87902362
http://viaf.org/viaf/27238439
OCLC Entity-Based Data Strategy

- VIAF, ISNI, FAST Publish Linked Data
- WorldCat.org Linked Data Release – using Schema.org
- Internal agreement on data strategy
- Evangelism
- Research & Design with Data Architecture Group
- Data mining of WorldCat resources

2012

2013

2014

- WorldCat Works Released

- Application Integration
  - WorldCat Discovery
  - Analytics
  - Discovery API
  - Cataloging

- More Entities Released
  - Person
  - Manifestation
  - Organization
  - Concept

- New Products
- Continuing Evangelism

- New Services
- Continuing Innovation
Our Task: Identifying the Problems To Solve

Trend 1: Shift from collections to services

As librarian, I want high quality data from around the world so that I can support my university’s international business school.

Trend 2: Search for Distinctiveness

As a researcher, I want to identify original sources so that I can produce original findings in my grant funded research project.

Trend 3: Be Found on the Web

As a library director, I want to show that my library has global relevance and a durable future so that I can prove value to my skeptical chancellor.