Making Library Collections Discoverable on the Web

Ted Fons
Executive Director, Data Services & WorldCat Quality
Some Trends in Libraries Today...

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 and staff budgets.

Trend 2: Search for Distinctiveness

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

Trend 3: Be Found on the Web

**Explain:** As users search everywhere, libraries want their collections to be found on the web.
So, what is OCLC going to do?
✓ Model things of interest to the web.

Focus first on data elements for web discovery

✓ Make those things available in structures familiar to the web.

Start with Schema.org
Manage entities, not records

✓ Improve library workflows.

Improve discovery

Reinvent cataloging
The library knowledge graph

A graph of relationships
The library knowledge graph
A graph of relationships
What will be better when we have that graph thingy?

I’m gonna make it better
What will be better?

Improved coherence of the data

*Explain:* Clustering related things (manifestations, items, formats, etc.)

Improved library applications

*Explain:* Improved discovery. Make it more web-like.
Reduce redundancy of data by managing entities in large aggregations.

Be Found on the Web

*Explain:* Using the data models and structures familiar to the web.
The library knowledge graph

Lots of things....if we do it right.
The library knowledge graph
Discovery works...
The year is 1327. Franciscans in a wealthy Italian abbey are suspected of heresy, and Brother William of Baskerville arrives to investigate. His delicate mission is suddenly overshadowed by seven bizarre deaths that take place in seven days and nights of apocalyptic terror.
Cataloging will be different...

- Managing the quality of Works
- Improving clusters
- Managing the quality of Persons
- Links to works, Other IDs

Consistent with RDA
So what progress have we made?
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

http://www.oclc.org/data
The Data Strategy: WorldCat Entities

**Work and Person Creation Process Flow**

There are three components to the pipeline for creating Work and Person entities. The harvest component extracts the data from the different sources. The map component identifies the objects and combines the triples through name recognition and authority linkages. The reduce component pulls together the entity descriptions and writes them out to HBase.
New “Common Ground” white paper

- Co-authored by the Library of Congress and OCLC Research
- This executive summary compares and contrasts the compatible linked data initiatives at both institutions
- A more detailed technical analysis will be released later this year

Download a copy at: oc.lc/CommonGround
“The alignment...is still accurate and is perhaps even more defensible than in 2013 because the primary BIBFRAME concepts are now more consistent with the corresponding concepts defined in the OCLC/Schema model.”

p. 8
Are you skeptical?

WELCOME
NATIONAL SOCIETY
OF SKEPTICS

I DON'T BELIEVE
WE'VE MET...

I DON'T BELIEVE
YOU DON'T BELIEVE
WE'VE MET...

Photo credit: http://www.quickmeme.com/img/4c
Can we measure impact?
Entities and library workflows

Discovery

The Name of the Rose

Summary:
The year is 1327. Franciscans in a wealthy Italian abbey are suspected of heresy, and Brother William of Baskerville arrives to investigate. His delicate mission is suddenly overshadowed by seven bizarre deaths that take place in seven days and nights of apocalyptic terror.

Works by

- Outliers: the story of success
- Blink: the power of thinking
- The tipping point: how little things can make a big difference

Works about

- TED Talks Malcolm Gladwell: What We Can Learn from
- Interview with Malcolm Gladwell, January 15, 2009
- The Ohio State University

Borrowing Options
- eBooks
- Printed Books
- Audio Books

Subjects
- Monastic libraries -- Italy – Fiction
- Semiotics – Fiction

en.wikipedia.org
Ongoing data modeling, research, experiments, and engagement

Measure impact of change, review actions, improve plans, repeat

- Release Person
- Release Works
- Release other Entities + add Articles

- Import and Export record data from WorldCat: MARC, UNIMARC, ONIX, BIBFRAME, etc.

- Begin Cataloging Entities
- Enhance Discovery
- Improve Other Workflows

| - | 2015 | 2016 | + |

- Build the Library Knowledge Graph
  - Connecting local systems, OPACs and the web via identifiers
Register

Add your holdings & metadata to WorldCat

Follow

Schema – Bib Extend
http://www.w3.org/community/schemabibex/

OCLC – Data
http://www.oclc.org/data

Ted Fons
Executive Director, Data Services & WorldCat Quality
fonst@oclc.org
@tedfons

Links and Entities