From Prototype to Production: Turning new ideas into useful services

Barbara Bushman, Assistant Head, Cataloging & Metadata Management Section, National Library of Medicine
Sara DeSmidt, Senior Product Analyst, OCLC Metadata Services

March 27, 2019
Agenda

• Review the current state of cataloging
• A future state example of knowledge work
  • Linked data pilot project
    • Background
    • Benefits
    • Case study
    • Project impacts
Current state
The library community has a data problem.

Our current standards do not result in data that is easily shared, managed, or discovered by users.
| 082 04 | 082.04 | S2 23 |
| 100 1# | 100 1 | S2 Adams, Douglas, d1952-2001, S2 author. |
| 245 14 | 245 14 | S2 The hitchhiker's guide to the galaxy : S2 a trilogy in five parts / S2c Douglas Adams. |
| 246 14 | 246 14 | S2 Hitchhiker's guide to the galaxy : S2b the complete trilogy of five |
| 264 #1 | 264 #1 | S2 London : S2b Pan Books, S2c 2017. |
| 264 #4 | 264 #4 | S2c 01992 |
| 300 ## | 300 ## | S2xii, 732 pages : S2c 20 cm. |
| 336 ## | 336 ## | S2a text S2b txt S2c 2 rdacontent |
| 337 ## | 337 ## | S2a unmediated S2b n S2c 2 rdamedia |
| 338 ## | 338 ## | S2a volume S2b nc S2c 2 rdacarrier |
| 500 ## | 500 ## | S2a "Don't panic"--Cover. |
| 505 0# | 505 0# | S2a The hitch hiker's guide to the galaxy -- The restaurant at the end of the universe -- |
| 520 ## | 520 ## | S2a Don't Panic: collected together in the Hitchhiker's Guide Trilogy are the five titles: an introduction from the author and bonus short story, Young Zaphod Plays It Safe. |
| 600 10 | 600 10 | S2a Dent, Arthur S2c (Fictitious character) S2v Fiction. |
| 600 10 | 600 10 | S2a Prefect, Ford S2v Fiction. |
| 600 17 | 600 17 | S2a Dent, Arthur S2c (Fictitious character) S2c 2 fast S2s 0 (OCoLC)fst00890366 |
| 600 17 | 600 17 | S2a Prefect, Ford. S2c 2 fast S2s 0 (OCoLC)fst01075077 |
| 650 #0 | 650 #0 | S2a Human-alien encounters S2v Fiction. |
| 650 #0 | 650 #0 | S2a Galaxies S2v Fiction. |
Linked Data basics

• In 2006 Tim Berners-Lee shared a set of common practices for exposing data on the Web in way that:
  – Allows many parties to participate in a Web of data
  – Fosters creating connections in a Web of data
  – Produces new knowledge and added value
From records to Linked data

082 04 $a 823.92 $2 23
245 10 $a The hitchhiker's guide to the galaxy : $b a trilogy in five parts / $c Douglas Adams.
246 10 $a The hitchhiker's guide to the galaxy : $b the complete trilogy of five
264 #1 $c 01903
300 ## $a xii, 732 pages : $c 20 cm.
336 ## $a text $b txt $c 2 rdataset
337 ## $a unmediated $b n $c 2 rdatafile
338 ## $a volume $b nc $c 2 rdcarrier
500 ## $a "Don’t panic"--Cover.
505 09 $a The hitchiker's guide to the galaxy -- The restaurant at the end of the universe --
520 ## $a Don't Panic: collected together in the Hitchhiker's Guide Trilogy are the five titles introduction from the author and bonus short story, Young Zaphod Plays It Safe.
600 10 $a Dent, Arthur Sc (Fictitious character) $v Fiction.
600 20 $a Prefect, Ford $v Fiction.
600 20 $a Prefect, Ford. $2 fast $0 (OCoLC)fas00090366
600 20 $a Prefect, Ford. $2 fast $0 (OCoLC)fas00075077
650 00 $a Human-alien encounters $v Fiction.
650 00 $a Galaxies $v Fiction.
Linked Data triples

- Subject
- Predicate
- Object

Subject: Andrew Pace’s blog is called Hectic Pace.

Why Linked Data?

Better outcomes
- A better user experience
- Greater web visibility

New functionality
- Develop better descriptions of library resources not well served by current standards

Cleaner data
- Replicate existing library functions more cheaply and efficiently
- Improve data integration
- Improve internal data management
- Greater web visibility
- A better user experience
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
Triples, triples, everywhere!!!

Other library-related linked data resources

  - Linked data representation of the MeSH biomedical vocabulary
- **Library of Congress** ([id.loc.gov](id.loc.gov))
  - Enables developers to interact with vocabularies found in data & standards as linked data.
  - 100,000+ requests/day; 100-500 million triples
- **Getty Art & Architecture Thesaurus**
  - A structured vocabulary for generic concepts related to art and architecture.
  - 100,000+ requests/day 100-500 million triples
- **Bibliothèque nationale de France** ([data.bnf.fr](data.bnf.fr))
  - Data produced by the Bibliothèque nationale de France more useful on the Web.
  - 10,000 – 50,000 requests/day; 100-500 million triples
- **Deutsche National Bibliothek**
  - Publishes authority and bibliographic data in RDF to make the data accessible to the semantic Web community.
  - 100-500 million triples
- **Linked Data for Libraries** ([https://www.ld4l.org/](https://www.ld4l.org/))
Project goals

*Work with our members* through a foundational shift in the collaborative work of libraries, communities of practice, and end-users—dramatically improving efficiency, embracing the inclusive, diverse, and earnest OCLC membership, and *empowering a new and trusted knowledge work enabled by the web*.

Andrew K. Pace  
Executive Director, Technical Research
About the Project

- December 2017-September 2018
- 16 OCLC Member libraries
- The objective was to evaluate a framework for reconciling, creating, and managing bibliographic and authority data as linked data entities and relationships
- To learn more, see: [http://oc.lc/ld-prototypes](http://oc.lc/ld-prototypes)
About the Project
Project expectations

- Set an end-date for experimentation
- Demand participation
- Communicate frequently
- Document everything
- Not every idea is a winner; fail fast and move on

Photo by rawpixel on Unsplash
Build – Observe – Adjust

PHASE 1
- Entity editor
- Reconciliation API
- SPARQL endpoint

PHASE 2
- Explorer UI
- OpenRefine API
- Wikimedia commons links

PHASE 3
- Retriever
- Serialization API
- Taxonomy navigation
- Property-based constraints
- Improved search experience
BENEFITS
Hands-on Participation

• Hands-on cataloging experience in linked data environment
• Practical workflows and integration with other components of existing metadata environments
Paradigm-shifting experience

• Cascading descriptions and relationships
• Conceptualizing properties & relationship types
• Working on libraryland’s edges
• Cataloging beyond text strings
CASE STUDY
Office Hours and Assignments

• Office Hours
  • Held weekly to bi-weekly, with partner participation and demonstration

• Selection of a Postcard
  • Opportunity to see how non-textual resources might be handled in a linked data environment
  • How to describe depictions of historical figures, places, events
Assignment: Postcards

• Cataloging in Passage meant:
  – Coming to terms with language versatility and browser confusion
  – Dealing with descriptive details at a different level
Assignment: Postcards

520 0 \a Postcard featuring a black and white image of Princess Maria Josefa of Saxony. She is seated with her hands folded in her lap. She is wearing a nurse’s uniform which consists of a long, high-necked dress and white cap. There is a Red Cross armband on her left arm. During WWI she nursed the wounded at the Augarten Palace in Vienna, which served as a hospital.

540 0 \a This item may be under copyright protection; contact the copyright owner for permission before re-use. \a DNLM

600 0 2 \a Maria Josefa, \a Princess of Saxony, \a 1867-1944.

650 1 2 \a Famous Persons \a https://id.nlm.nih.gov/mesh/D005197

650 1 2 \a Nurses \a https://id.nlm.nih.gov/mesh/D009726

650 2 2 \a World War I \a https://id.nlm.nih.gov/mesh/D0047828

651 2 \a Austria \a https://id.nlm.nih.gov/mesh/D001317

655 2 \a Postcard \a https://id.nlm.nih.gov/mesh/D056571

700 1 \a Zwirnerl, Michael, \a former owner. \a DNLM

710 2 \a Postkartenverlag Brüder Kohn, \a publisher.
Assignment: Postcards

Postcard of Maria Josepha

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Postcard of Maria Josepha</td>
<td>Postcard of Maria Josepha in her nursing uniform during World War I</td>
</tr>
<tr>
<td>German</td>
<td>Postkarte von Maria Josefa</td>
<td>Postkarte von Maria Josefa in ihrer Krankenpflegeuniform während des Ersten Weltkrieges</td>
</tr>
</tbody>
</table>

Also known as:
- Postcard of Archduchess Maria Josepha
- Postcard of Princess Maria Josepha
- Postkarte von Prinzessin Maria Josefa
- Postcard of Archduchess Maria Josepha
- Postcard of Princess Maria Josepha
- Postkarte von Prinzessin Maria Josefa

Fewer languages
Assignment: Postcards
Relationships in the Explorer

Princess Maria Josepha of Saxony

**About**

Princess Maria Josepha of Saxony (31 May 1867 – 28 May 1944) was the mother of Emperor Charles I of Austria and the fifth child of George of Saxony and Infanta Maria Anna of Portugal.

【From dbpedia】

Archduchess of Austria

Also known as: Maria Josefa Sachsen, Prinzessin, 1867-1944

【Public domain】

【Wikimedia Commons image】

**Depicted by**

Postcard of Maria Josepha
Postcard of Maria Josepha in her nursing uniform during World War I
PROJECT IMPACTS
Reflection

• Opportunity to collaborate on a project report
• Influence on workflows
Going forward

• Continue collaboration
  – Metadata Initiatives Advisory Group will offer feedback on OCLC production services
  – Ponder questions with the community
    • Transparency of and governance of data model
    • Location for descriptive information
Going forward

• NLM’s LD plans
  – MeSH RDF
  – LD4P2 participation
Thank you!

Barbara Bushman
Assistant Head, Cataloging & Metadata Management Section,
National Library of Medicine
bushmanb@mail.nlm.nih.gov

Sara DeSmidt
Senior Product Analyst
desmidts@oclc.org