Thursday 3 June 2021

**Transitioning to the Next Generation of Metadata**

*Days of Knowledge Organization 2021 Conference - OsloMet*

Titia van der Werf  
Senior Program Officer, OCLC Research

Ellen Hartman  
Senior Library Consultant, OCLC EMEA
Titia van der Werf
Senior Program Officer, OCLC Research
Collective Collection

Libraries are operating in an increasingly complex network of information supply and demand. OCLC Research provides evidence and models that help libraries optimize access to collections through coordinated resource sharing.

Data Science

To be discoverable by today's online users, traditional library data must be transformed. OCLC Research analyzes bibliographic data to derive new meaning, insights, and services for use by libraries and information seekers.

Libraries as Community

Whether on a college campus, in a rural town, or an urban center, libraries are the heart of communities. OCLC Research equips libraries with the knowledge, skills, and confidence to respond to ever-evolving community needs.
INTRODUCTION
Cultivating understanding
Cultivating understanding

OCLC Research Round Table Discussions on Next Generation Metadata
Experimenting and building

Publish linked data - FAST, VIAF, WorldCat (2009 - )

EntityJS Research Project (2013)

Person Entity Lookup Pilot (2014)

CONTENTdm Metadata Refinery (2015-16)

Project Passage (2017-18)

CONTENTdm Linked Data Pilot (2019-20)

Entity Management Infrastructure (2020-21)
Karen Smith-Yoshimura
OCLC Research Senior Program Officer (retired Nov 30, 2020)
Context

Format-specific metadata management based on curated text strings in bibliographic records understood only by library systems is nearing obsolescence, both conceptually and technically.

In short, the metadata could be better, there is not enough of it, and the metadata that does exist is not used widely outside the library domain.
• The Transition to Linked Data and Identifiers
• Describing “Inside-Out” and “Facilitated” Collections
• Evolution of “Metadata as a Service”
• Preparing for Future Staffing Requirements

https://doi.org/10.25333/rggd-b343
Persistent identifiers were viewed as crucial to transitioning from current metadata to future applications.
Identity management poses a change in focus...to describing entities...and the relationships among them.
"Addressing language issues is important as libraries seek to develop relationships and build trust with marginalized communities."
Describing “Inside-Out” and “Facilitated” Collections

- Archival collections
- Archived websites
- Audio and video collections
- Image collections
- Research data

“Metadata underlies all discovery regardless of format, now and in the future…”

https://researchworks.oclc.org/iiif-explorer/
Libraries’ expertise in metadata standards, identifiers, linked data, and data sharing systems as well as technical systems can be invaluable to the research life cycle.
Evolution of “Metadata as a Service”

New applications

Plus:
  • Metrics
  • Consultancy
  • Semantic indexing

Bibliometrics
A culture shift is needed: from pride in production alone to valuing opportunities to learn, explore, and try new approaches to metadata work.
The next generation of metadata will become even more focused on entities rather than record-based descriptions of an institution’s collections.
Ellen Hartman
Senior Library Consultant
TOWARDS A SHARED ENTITY MANAGEMENT INFRASTRUCTURE
OCLC’s involvement with linked data

- **Publish linked data - FAST, VIAF, WorldCat (2009 - )**
- **EntityJS Research Project (2013)**
- **Person Entity Lookup Pilot (2014)**
- **CONTENTdm Metadata Refinery (2015-16)**
- **Project Passage (2017-18)**
- **CONTENTdm Linked Data Pilot (2019-20)**
- **Entity Management Infrastructure (2020-21)**
Project SEMI

- Two-year, $2.436M grant, matched by OCLC
- Production infrastructure for Work and Person entities
- Support for multiple descriptive and encoding standards
- Use of persistent identifiers
- Most importantly: a collaboration with the library community
Advisory group members
What are we building?

• An **Entity Ecosystem** that facilitates:
  o Creation and editing of new entities
  o Connecting entities to the Web

• A **community of users** who can:
  o Create/Curate data in the ecosystem
  o Imagine/propose workflow uses
  o Communicate easily with each other and with OCLC to iteratively improve the prototype

• Provide services to:
  o Reconcile data
  o Explore the data
How entities are built from WorldCat data
Multiple links for multiple entity types
Timeline of activities

2020

- Wikibase: Millions of entities
- Simple search & read

2021

- Ramp up: 10s of millions of entities
- Creation and editing tools

- Add more data sources
- Advanced querying, new UI tools

- Scale and refine
- Prepare for release

Production release
How libraries can use data on January 1, 2021:

• Find information on OCLC entities on the free website
• Create, maintain and curate entity descriptions using a separate interface
• Work with and display entity data using different APIs
Questions?

Links

OCLC Research: https://oclc.org/research

Report: "Transitioning to the Next Generation of Metadata":
https://doi.org/10.25333/rqgd-b343

OCLC and Linked Data:
FINDINGS FROM THE ROUND TABLE DISCUSSIONS
The OCLC Research Series on Next Generation Metadata

OPENING PLENARY WEBINAR
Tuesday 23 February 2021, 15:00 (CET)

INTERACTIVE ROUND TABLE
During the first two weeks of March 2021.

CLOSING PLENARY WEBINAR
Tuesday 13 April 2021, 15:00 (CET)
This report served as inspiration and background reading for the OCLC Research Series on the same topic.

Also: Spanish translation of Karen’s report on the BNE website

How do we make the transition to the Next Generation of Metadata happen at the right scale and in a sustainable manner, building an interconnected ecosystem, not a garden of silos?
Participants and outputs

• **Broad interest**: representatives from 71 institutions (67% university libraries) - from 17 countries

• **Insightful reports**: 8 blog posts + translations on HangingTogether.org
Maps of next generation projects from the different language sessions
Predominant topics

1. Investing in authorities and identifiers
2. Interlinking data
3. Managing multiple scales
4. Professional development
Investing in authorities and identifiers

Context: Open Government Data

- strong focus on **transforming** and **publishing** library data
- is driving **collaboration** at **national** levels

*e.g.* [French National Bibliographic Transition Program](https://www.oclc.org)
Investing in authorities and identifiers

Aim: To achieve a critical mass of interoperable library data

- increasing the usefulness of authority data through reconciliation
- populating data sources and systems with library data
- embedding PIDs in the value-chain
  
  e.g. ISNI => VIAF => LC/NACO
Interlinking data

“When you refer to data that has already been defined by others, you relinquish control over that piece of information, and that can be a mental barrier against doing linked data the proper way. It is much safer to store and manage all the data in your own silo. But the moment you can let go of that, the world can become much richer than you could ever achieve on your own.”

[Dutch session]
Managing multiple scales

“There are multiple, co-existing ‘right scales’ for next generation metadata that need to be interconnected and managed as a whole.”

[French session]
Knowledge sharing/professional development

• Next generation metadata is often a “stretch assignment” for staff, on top of everything else
• “Old generation” systems are not ready yet
• Having to experiment with news tools (e.g. Wikidata) is not easily done. Learning curves are steep
• Regular training for staff in these areas is not always available
• Looking to OCLC for leadership in this area
How is OCLC helping libraries to prepare for the next generation of metadata?

1. Cultivating understanding
2. Experimenting
3. Building infrastructure
Questions?

Links

OCLC Research Blog "Hanging Together": https://hangingtogether.org
Do you want to receive the latest OCLC research reports, events, products and news?

Update your email preferences at: oc.lc/mail
Thank You!

Titia van der Werf
Senior program officer
Titia.vanderwerf@oclc.org

Ellen Hartman
Senior library consultant
Ellen.hartman@oclc.org

Because what is known must be shared.