

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

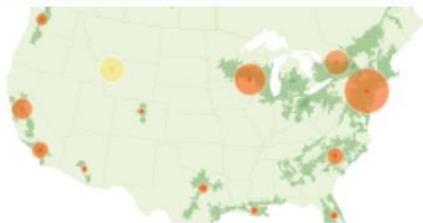
Senior Library Consultant, OCLC EMEA



Titia van der Werf

Senior Program Officer, OCLC Research

Research Areas



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 Centers

Whether in a college campus, 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.

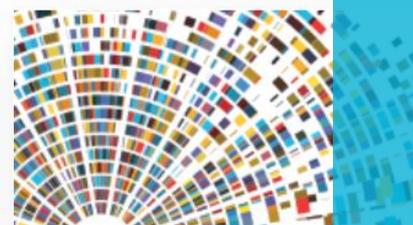
OCLC Research . oclc.org/research . hangingtogether.org



Research Collections & Support

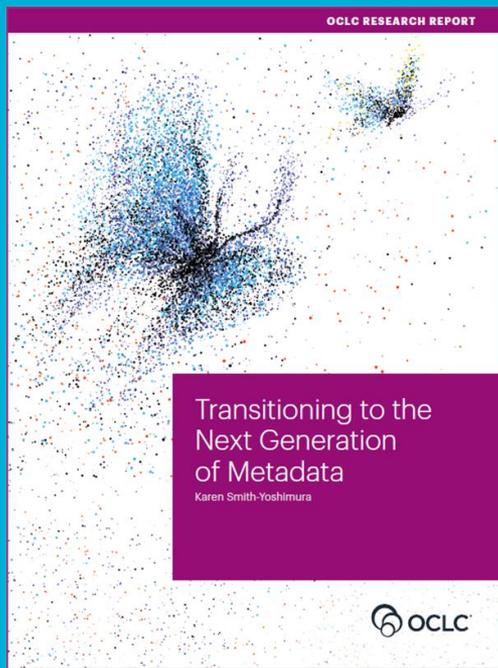


User Research



Library as Enterprise

INTRODUCTION



Cultivating understanding

OCLC
Research
Library
Partners
Metadata
Managers
Focus
Group
Discussions



OCLC RESEARCH REPORT

Transitioning to the
Next Generation
of Metadata

Karen Smith-Rothemund

OCLC

Cultivating understanding

OCLC
Research
Round Table
Discussions
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.

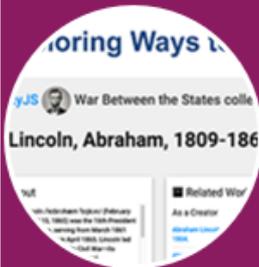


**CLOSING PLENARY
WEBINAR**
Tuesday 13 April 2021,
15:00 (CET)

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)

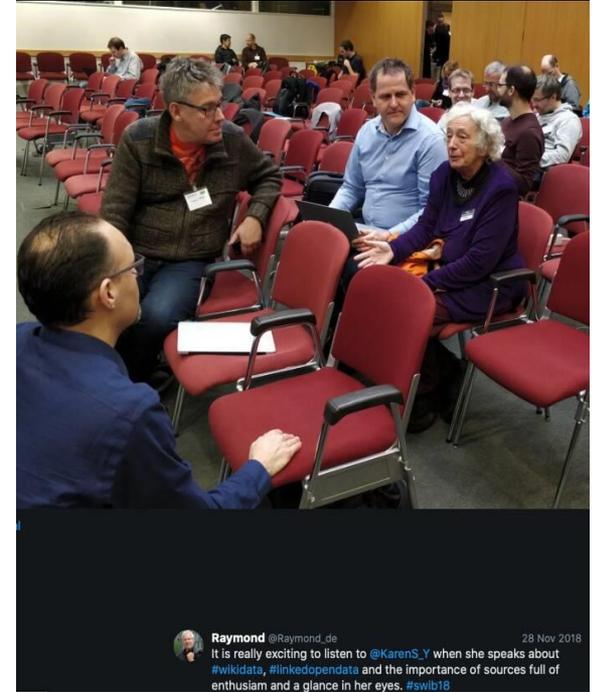
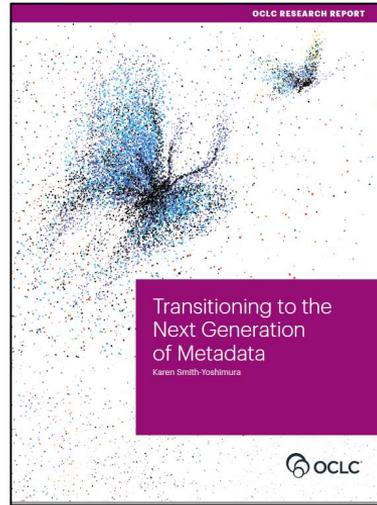


**OCLC RESEARCH LIBRARY
PARTNERS METADATA
MANAGERS FOCUS GROUP
DISCUSSIONS**



Karen Smith-Yoshimura

OCLC Research Senior Program
Officer (retired Nov 30, 2020)



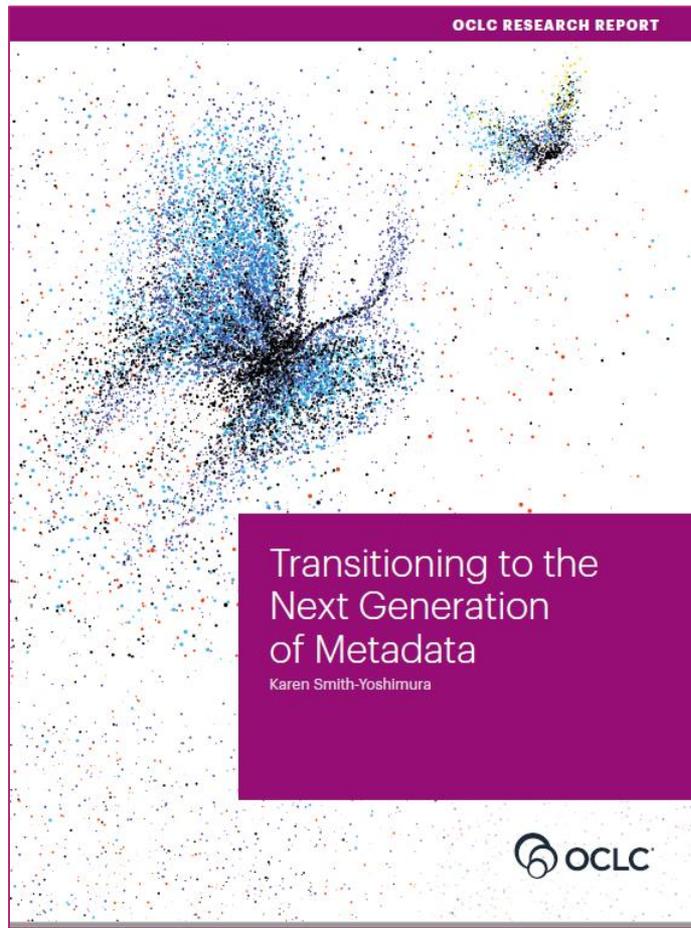
Context

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

Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata*. Dublin, OH: OCLC Research.
<https://doi.org/10.25333/raqd-b343>



Transition to Linked Data & Identifiers

Changing Resource Description Workflows

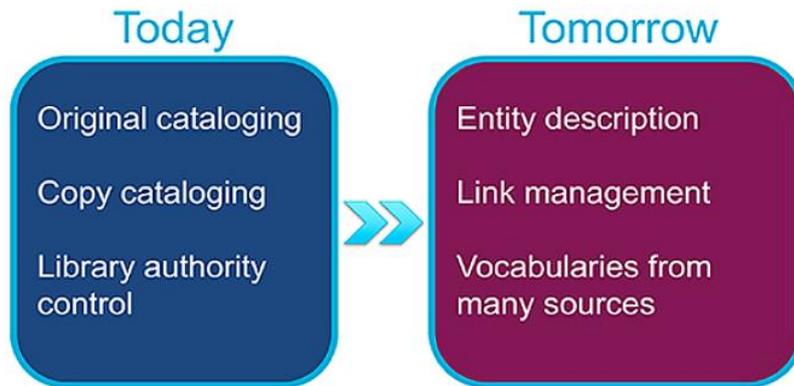


FIGURE 1. [“Changing Resource Description Workflows”](#) by OCLC Research¹⁵

“Persistent identifiers were viewed as crucial to transitioning from current metadata to future applications.”

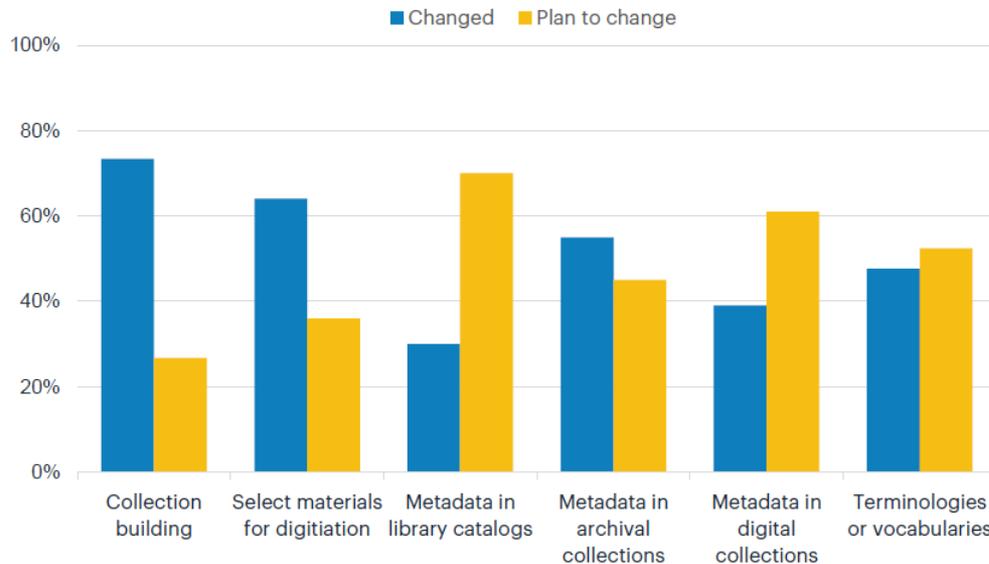
One Wikidata Identifier Links to Other Identifiers and Labels in Different Languages



FIGURE 4. One Wikidata identifier links to other identifiers and labels in different languages

“Identity management poses a change in focus...to describing entities ...and the relationships among them.”

What Areas Have You Changed or Plan to Change Due to Your Institutions EDI Goals and Principals?

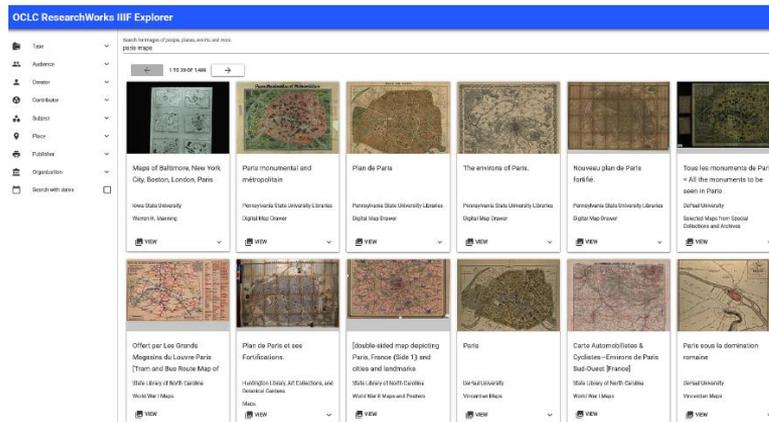


“Addressing language issues is important as libraries seek to develop relationships and build trust with marginalized communities.”

Describing “Inside-Out” and “Facilitated” Collections

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

The OCLC ResearchWorks IIIF Explorer Retrieves Images about “Paris Maps” across CONTENTdm Collections



<https://researchworks.oclc.org/iiif-explorer/>

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

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”

Distribution of 465 Indigenous Language Codes in the Australian National Bibliographic Database



New applications

Plus:

- Metrics
- Consultancy
- Semantic indexing

UK Hatchette’s “River of Authors” Generated from the British Library’s Catalog Metadata



Bibliometrics

Preparing for Future Staffing Requirements

A *culture shift* is needed: from pride in production alone to valuing opportunities to learn, explore, and try new approaches to metadata work.

Conclusion

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**



GRANTS DATABASE /

OCLC, Inc.

Entity Reconciliation for Linked Open Data

to support the development of an infrastructure to reconcile entities, such as names, for linked open data

Location: Dublin, OH, United States

Amount: \$2,436,000

Date: Dec. 9, 2019

Length: 24 months

Program Area: Scholarly Communications

Area of focus: Access and Library Services

Reference number:1907-06977

Advisory group members



UNIVERSITY OF MINNESOTA



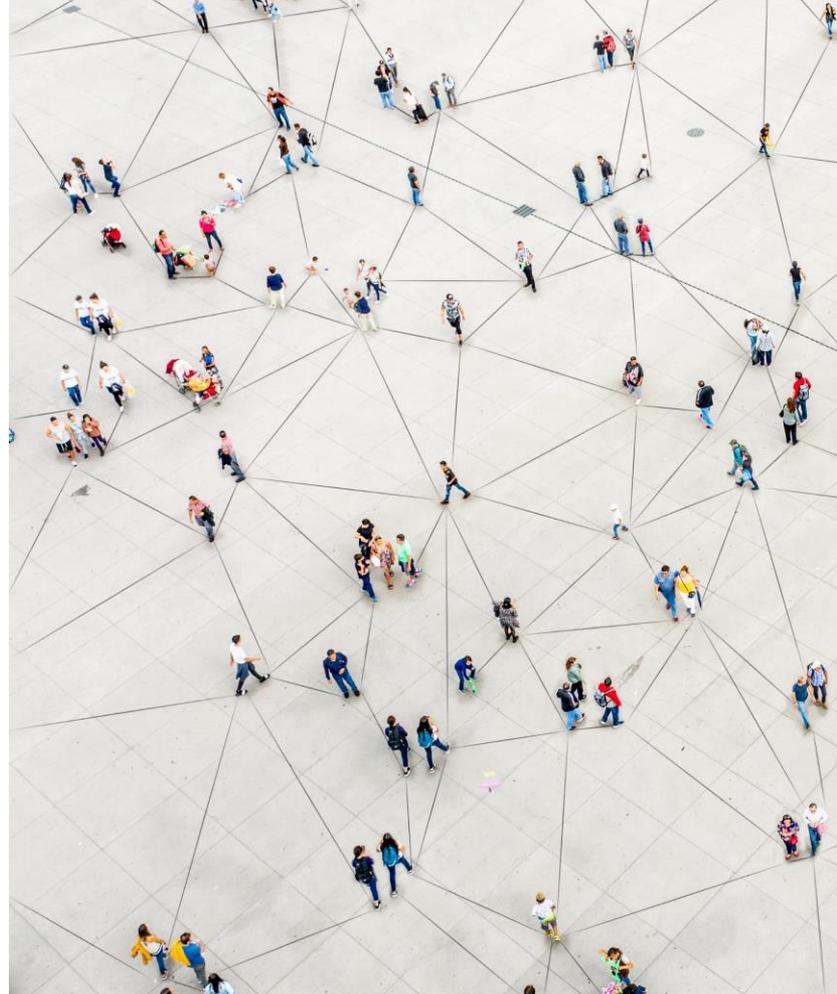
Yale



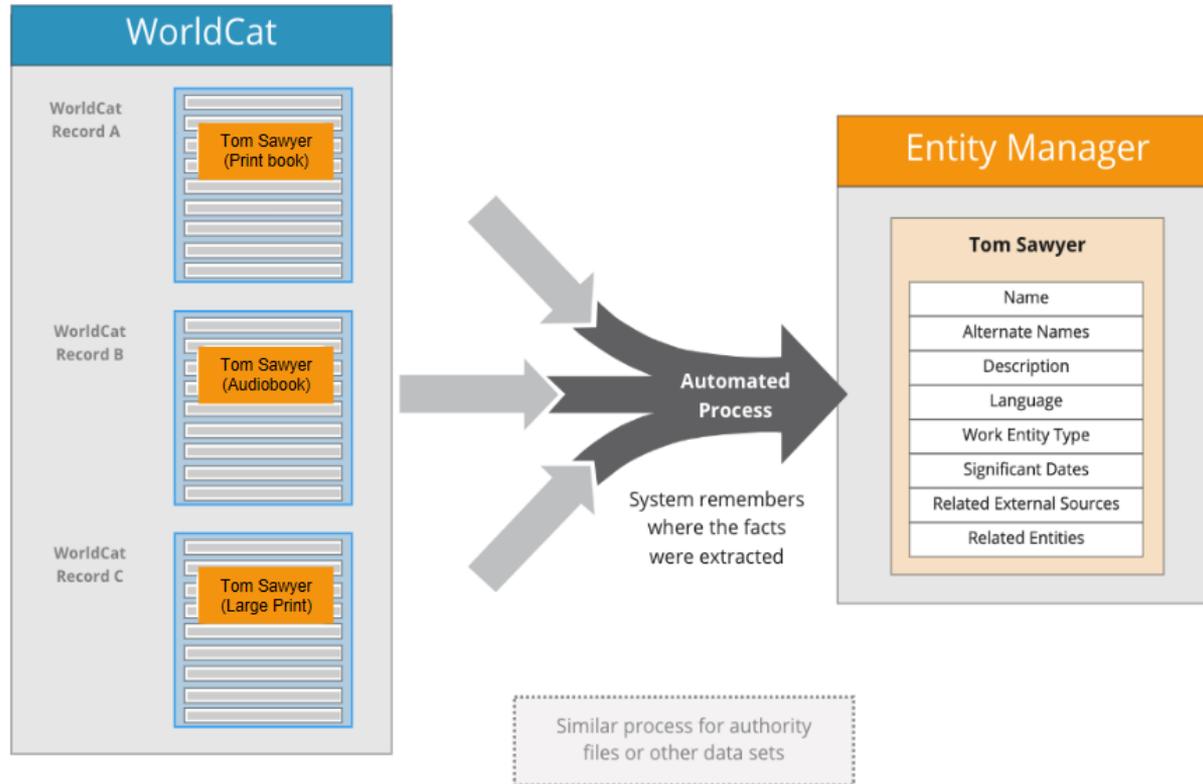
National Library Board
Singapore

What are we building?

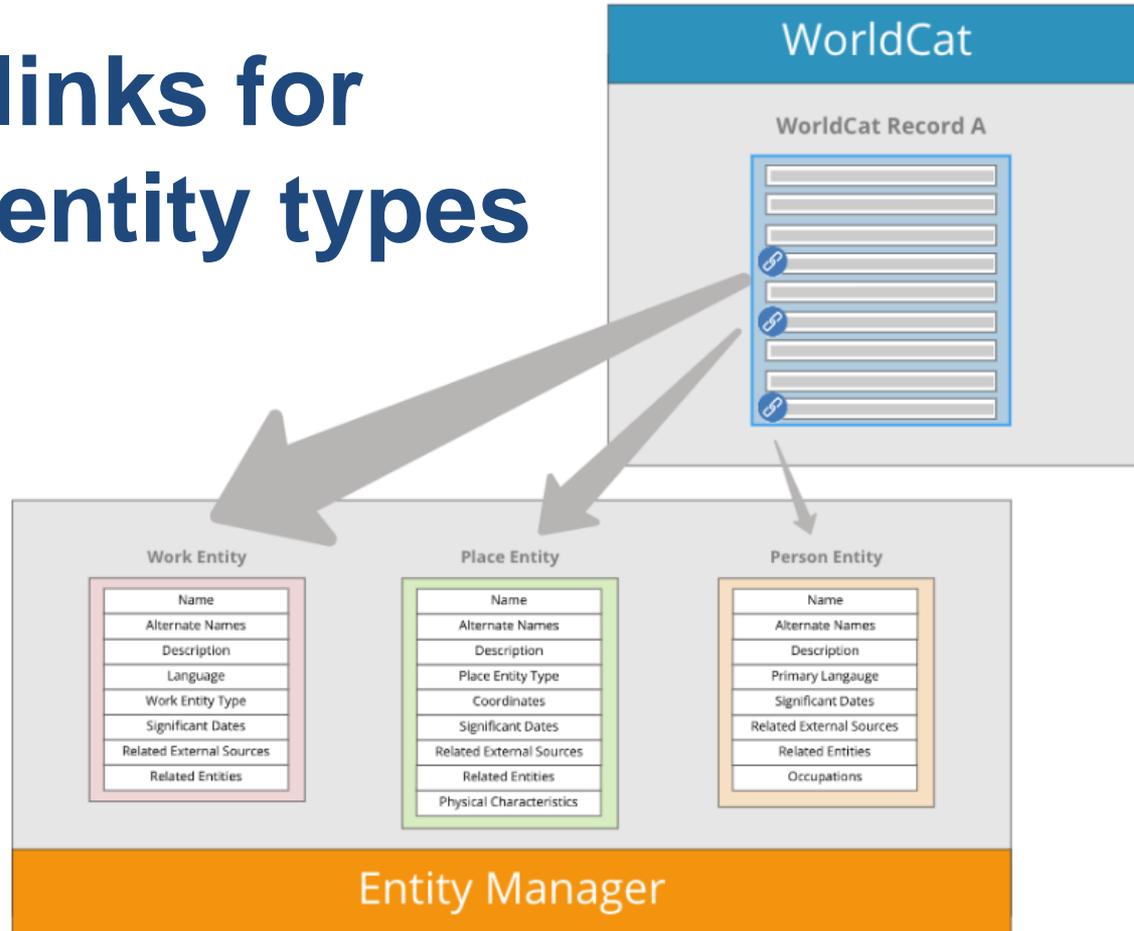
- An **Entity Ecosystem** that facilitates:
 - Creation and editing of new entities
 - Connecting entities to the Web
- A **community of users** who can:
 - Create/Curate data in the ecosystem
 - Imagine/propose workflow uses
 - Communicate easily with each other and with OCLC to iteratively improve the prototype
- Provide services to:
 - **Reconcile data**
 - **Explore the data**



How entities are built from WorldCat data



Multiple links for multiple entity types



Timeline of activities

2020

January -
June

- Wikibase: Millions of entities
- Simple search & read



July -
December

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



2021

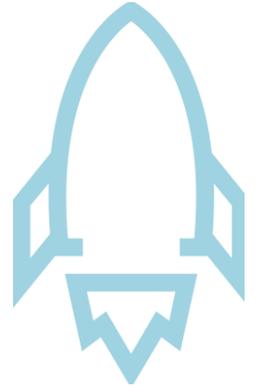
January -
June

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



July -
December

- 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:
<https://www.oclc.org/en/worldcat/oclc-and-linked-data.html>

**Because
what is
known must
be shared.®**

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

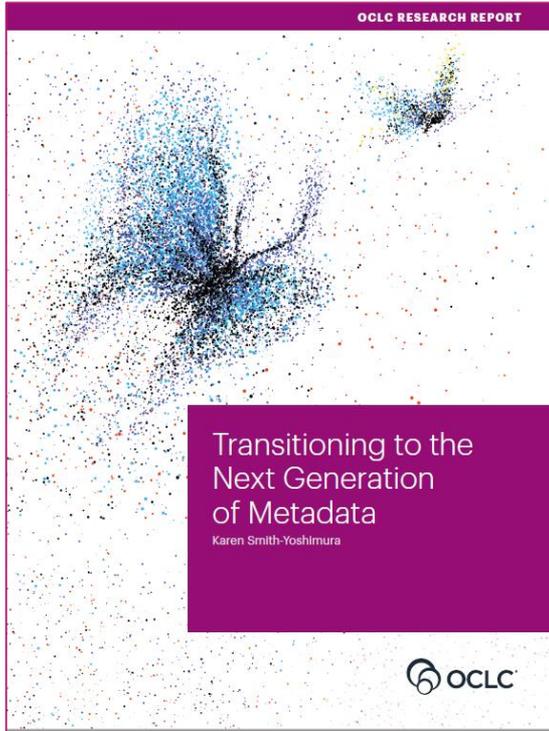
**During the first two weeks of
March 2021.**



CLOSING PLENARY WEBINAR

**Tuesday 13 April 2021,
15:00 (CET)**

Inspiration



oclc.org/nextgen-metadata-report

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](#)

Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata*. Dublin, OH: OCLC Research.

<https://doi.org/10.25333/rqgd-b343>

Main discussion question

*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](https://www.hangingtogether.org)

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](#)

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 new 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 Discussion Series on the Next Generation of Metadata: the recordings, slides and blog posts: <https://www.oclc.org/go/en/events/next-generation-of-metadata/Register.html>

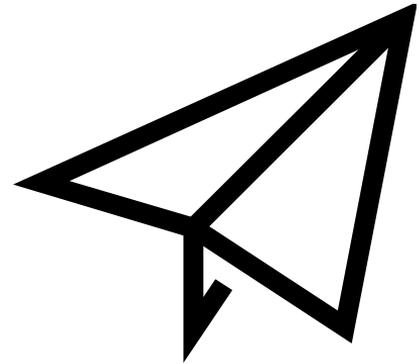
OCLC Research Blog "Hanging Together": <https://hangingtogether.org>

**Because
what is
known must
be shared.®**

**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.®**