


13th of September, 2022

Linked data and international standards for cultural heritage, Brussels

Next generation metadata as a transformative change

Dr. Annette Dortmund, Senior Product Manager, OCLC

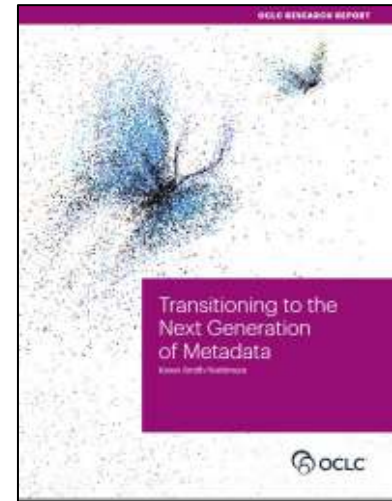
 @libsun

 <https://orcid.org/0000-0003-1588-9749>

“Transitioning to the Next Generation of Metadata”

Metadata work is undoubtedly now in an era of transformation and transition.

- The transition to **linked data and identifiers**
- Describing **“inside-out”** and **“facilitated”** collections
- Evolution of **“metadata as a service”**
- With effects on future **staffing requirements**



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

Why is metadata changing?

- Traditional library metadata was and is made by librarians conforming to rules that are mainly used and understood by librarians.
- It is record-centered, expensive to produce, and has historic size limitations.
- Metadata is limited in its coverage, notably not including articles within scholarly journals or other scholarly outputs.
- The infrastructure has been inadequate for managing corrections and enhancements, inducing an emphasis on perfection that has exacerbated the slowness of metadata creation.
- It is not scalable enough to meet requirements given rapid digital change.

In short

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.

How is the creation process changing?

- Metadata is no longer created by library staff alone.
- Publishers, authors, and other interested parties are equally involved in metadata creation.
- Metadata creation has also been pushed forward in the scholarly life cycle, with publishers creating metadata records earlier than in the traditional cataloging process.
- Metadata can now be enhanced or corrected by machines or by crowdsourcing.

How is the metadata itself changing?

- Machine-readable cataloging (MARC) was created to replicate the metadata traditionally found on library catalog cards.
- We are transitioning from MARC records to assemblages of well-coded and shareable, linkable components, with an emphasis on references
- We are eliminating anachronistic abbreviations not understood by machines.
- Instead of relying only on library vocabularies such as subject headings and coded lists, the developing assemblages can accommodate vocabularies created for specific domains, expanding the metadata's potential audiences.

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

Transitioning ... accelerated

The COVID-pandemic has accelerated

- the digital shift (from print to electronic resources)
- and the “move to open” (from closed/paywalled to open access)

Libraries are coping with an increasingly complex and messy **hybrid** environments

- Updating “old” metadata (MARC) with “new” data such as PIDs, to lay the foundation for “new” linked data and workflows
- Employing identifiers from different sectors
- Analog (special) collections still important for Humanities

Transitioning ... to pluralization

Strong push for Diversity, Equity and Inclusion (DEI)

- Need for ethical and responsive metadata
- Libraries are engaging their communities in descriptive practices and need more agile vocabularies (compared to e.g. LCSH)
- Linked data seen by some as technology that might help
- Streamlining and automating mainstream data workflows required to free up resources for pluralization

Transitioning ... but how?

Professional development of librarians is a challenge

- The new generation of librarians lacks the traditional cataloging know how
- The competencies and skills required to work with next general metadata are not yet clearly defined
- Much learning-by-doing

New standards and practices are a moving target

Transitioning ... workflows

Identity management poses a change in focus, from providing access points in resource descriptions to describing the *entities* in the resource (work, persons, corporate bodies, places, events) and establishing the relationships and links among them.


Transitioning ... quality criteria

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

Transitioning ... infrastructure

- How will “cooperative cataloguing” work in a next generation metadata environment?
- How will libraries share statements about resources?
- Should there be a centralized linked data store for libraries?
- What will “trustworthy provenance” look like in this context, and also in the context of DEI?
- Will there be a stronger focus on peer-to-peer sharing?
- Should libraries engage in general purpose infrastructures (such as Wikibase/Wikidata) or create their own, more specific environments?

Polarity of requirements

 Project Passage (2017-18)	 CONTENTdm Linked Data Pilot (2019-20)	Shared, homogeneous, and centralized entities accounting for the reality of localized, heterogeneous, de-centralized collections
		Machine-matching, highly automated reconciliation with tools for hand-matching, semi-automated reconciliation
		Well-accepted context: Persons and Works	Granular context: About, Depicts, Annotations, Notes
		Blurs the line between bibliographic and authority work	Blurs the line between object description and context description

A large centralized infrastructure is needed. Custom applications and interfaces are needed.

This will facilitate new Knowledge Work in libraries.



Why a library-specific ecosystem?

Wikipedia/Wikidata has the goal to represent selected knowledge aspects, but is not designed to support the process of knowledge creation.

Libraries have a central role in organizing knowledge and make it discoverable – as a support for the publication and research life cycle.

Libraries have developed specific policies, standards and quality criteria: authoritativeness, reliability and completeness of metadata.



Why a library-specific ecosystem?

Libraries deal with immense scale, high performance and long-term commitment.

Complementing general purpose ecosystems (e.g. Wikidata), not replacing them or made redundant by them.

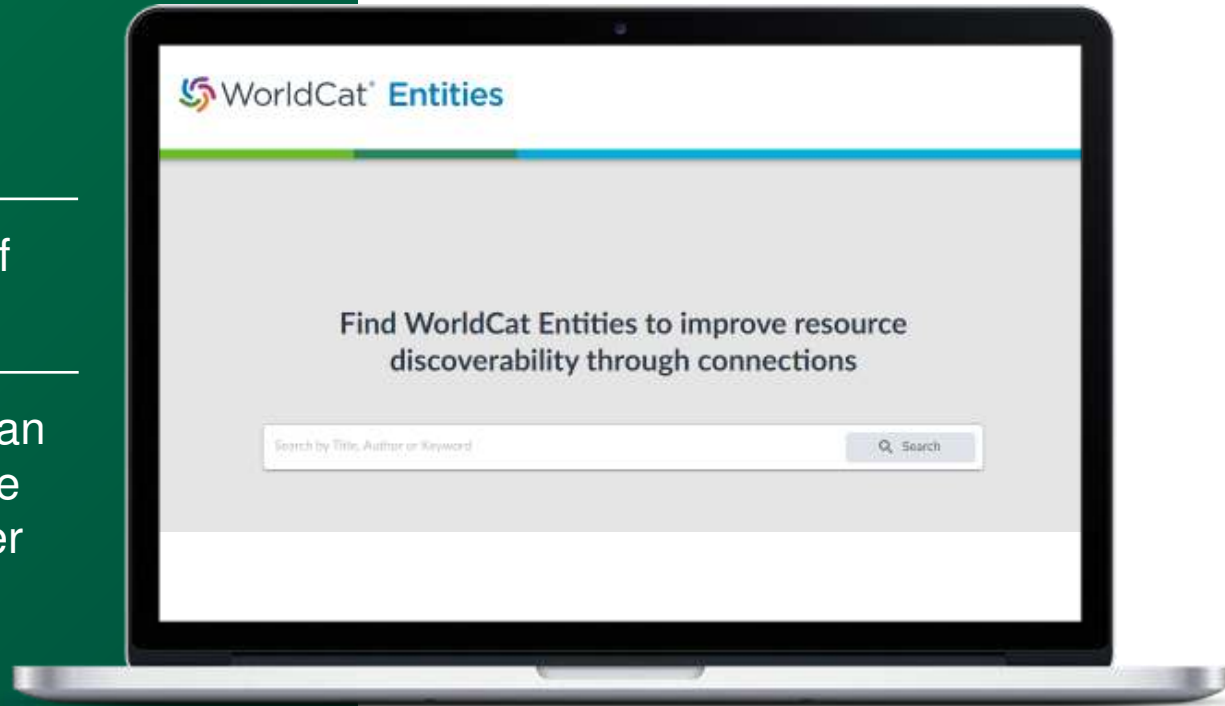
WorldCat Entities

More than 150M entities
searchable on publicly
available website

Entities contain descriptions of
creative works and persons

Library metadata specialists can
look up entity URIs and make
new connections that empower
discovery and research

entities.oclc.org

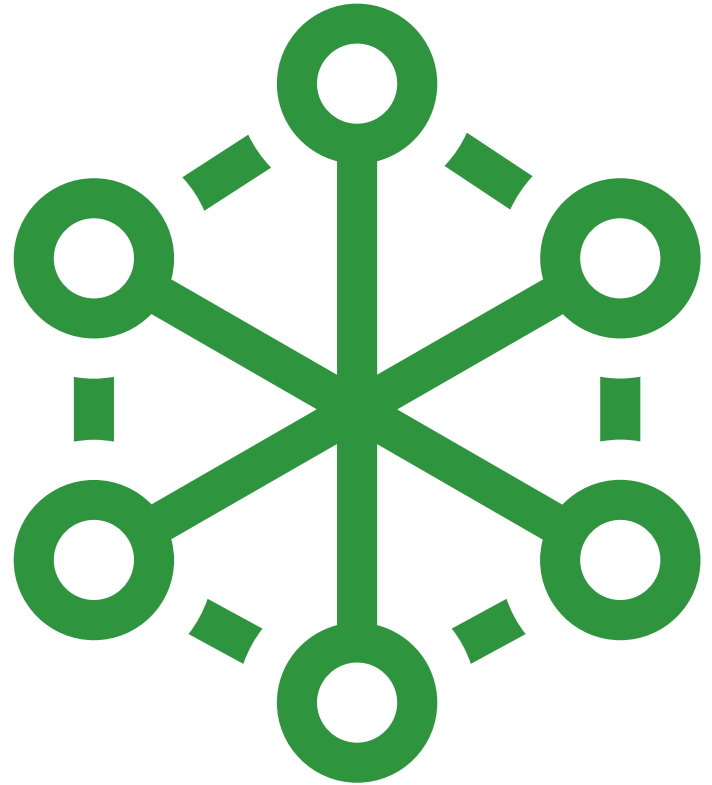


Linked Data: next steps

Commence development partner phase to refine **OCLC Meridian** (an entities management tool) and **APIs**

Collaborate with key libraries globally to explore how these tools can best be integrated into library workflows

oclc.org/linkeddata



OCLC's role in realizing the promise of Linked Data

We believe in the value of libraries, and the value of library cooperation.

We support library cooperation with a shared technical and human infrastructure has been our mission for 50 years.

We understand what is needed to work with linked data at scale and in the context of library workflows as a result of a decade of research and hands-on learning.

We create and continue to develop global, sustainable infrastructures.

We work cooperatively with libraries to meet the community's needs while ensuring solutions achieve critical mass.

Thank you!

Dr. Annette Dortmund

Senior Product Manager, OCLC

annette.dortmund@oclc.org