23 September 2020

# (Id)entity Management:
# What is the shape of the future?

**John Chapman**

Senior Product Manager, Metadata Strategy and Operations

OCLC

# Warning: Spoiler Alert

- Some problems have been addressed
- Thorny problems remain
- Many tools and workflows will be required
- Many things will only reveal themselves along the way

# Starting with the goal in mind

- Can we design our systems to be less brittle?
- Can we "swap out" components instead of replacing the whole thing?
- What are the correct "components" of distributed systems?
- What's going to be centralized, and what will be done locally?

# Thorny Problems affecting answers

- Archives and Special Collections
- ”Entity Alignment”, mixing & matching
- Representing and facilitating local communities & collections

# Archives and Special Collections

- Descriptive data models for special collections in linked data environments
- Challenges around multilinguality
- Ethical issues, community engagement
- Expressing relationships and change over time
- The long tail of authorities & identifiers in special collections
- Sustainability

Archives and Special Collections Linked Data: Navigating between Notes and Nodes

OCLC Research Archives and Special Collections Linked Data Review Group

# Entity Alignment

- Given an explosion of "mixing and matching," how can we align?

- Can we bridge cultural and temporal semantic differences?

- What tools will be most efficient/useful/durable over time?

Entity
Management
Infrastructure
(2020-21)

OCLC

# Local knowledge, local communities

- Indigenous peoples
- Local history
- Local semantics
- Naming
- Collecting
- Who does the work?

- Aggregating without homogenizing

**Works in Progress Webinar: Decolonizing Descriptions: Finding, Naming and Changing the Relationship between Indigenous People, Libraries and Archives**

This webinar examines how two different organizations—the Association for Manitoba Archives and the University of Alberta Libraries—began the processes of examining subject headings and classification schemes as they relate to Indigenous people.

**Works in Progress Webinar: Dynamics of change—continuing the conversation on library naming, finding, and relationship-building with Indigenous peoples**

Learn about how the University of Toronto Libraries is reckoning with both cultural change and the development of culturally appropriate metadata on a system-wide scale.

**Works in Progress Webinar: Operationalizing the CARE Principles for Indigenous Data Governance**

What are the CARE Principles? Learn how they work to address the historical legacy of data inequities that impact Indigenous Peoples. Presenters will also discuss how the CARE Principles promote a broader understanding of the FAIR Principles when applied to Indigenous data and collections.
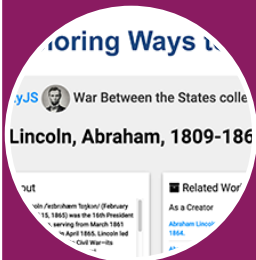
# What's OCLC's vision?

- New infrastructure
- A variety of tools and data sources, enabling new workflows
- Ongoing standards, rules, and best-practice conversations
- Targeted research

# Why OCLC?

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

# Why OCLC?



CONTENTdm
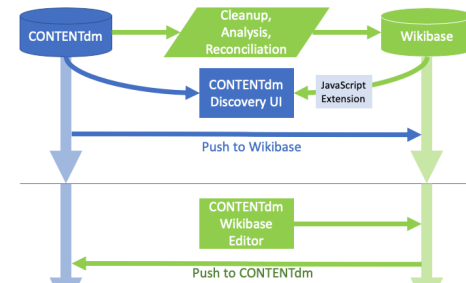Linked Data
Pilot
(2019-20)
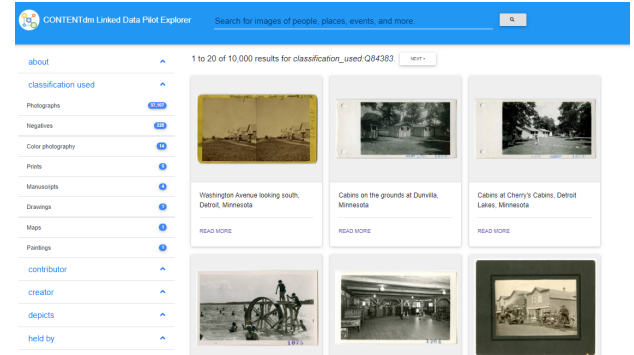
Entity
Management
Infrastructure
(2020-21)

OCLC

# CONTENTdm Linked Data Pilot

CONTENTdm Linked Data Pilot (2019-20)

## Project goals

- Develop scalable methods and approaches for richer, state-of-the-art machine representations of entities and relationships.

- Make visible connections that were formerly invisible.

- Prototype an application for library staff to:

  - convert existing record-based metadata into linked data

  - manage and publish the resulting entities and relationships

# Entity Management Infrastructure

- Two-year, $2.436M Mellon Foundation grant, matched by OCLC
- Production infrastructure for Creative Work and Person entities
- Support for multiple descriptive and encoding standards
- Use of persistent identifiers
- **Most importantly: a collaboration with the library community**

Entity
Management
Infrastructure
(2020-21)

OCLC

# Our goals

- Address infrastructure needs identified by libraries
  - Expand on non-MARC metadata management
  - Link library data to non-library data… and shared data to local data
  - Provide ID creation services to help "at the point of need"
  - Stand behind entity URIs
- Operate at a large scale – and be sustainable
- Complement other efforts
- Deliver products and services December 2021

# A semantic continuum

**Shared Entity Management Infrastructure (2020-21)**

**CONTENTdm Linked Data Pilot (2019-20)**

| | |
|---|---|
| Shared entities… | …localized, heterogeneous, de-centralized collections |
| Machine-matching, highly automated reconciliation… | … tools for hand-matching, semi-automated reconciliation |
| Widely applicable context: Persons & creative works | Granular context: About, Depicts, Annotations, Notes |
| Combines bib and auth work | Combines object and context description |

Custom applications and interfaces needed

**This is the new Knowledge Work**

OCLC

# Thank you!

**John Chapman**

Senior Product Manager,
Metadata Strategy & Operations

**chapmanj@oclc.org**

**https://blog.oclc.org/next/author/john-chapman/**

Because
what is
known must
be shared.®

OCLC®