The Match & Merge Story:
Improving Search Results, Reducing User Confusion

Tyler Hauck
Senior Metadata Librarian
The Research Role of Discovery Services

- Provide extensive resource coverage
- Expose the best resources
- Connect users to most relevant content for their needs
- Provide the most reliable access
Discovery benefits most from metadata that is...

- Highest quality
- Most extensive
- Most detailed
Ex Libris Central Discovery Index (CDI)

Unified content index
- Originates from a variety of sources
- Represents several different material types
- Contains approximately 3.6 billion individual records (“physical records”)

Content is processed by Ex Libris Content Operations

To be used by Ex Libris discovery services
- Summon
- Primo
Content Ingestion Process for Discovery: High-Level Overview

1. Identify New Content to add to the Indexes
2. Review & Evaluate Sample Content
3. Approve Sample Content
4. Sign Data-Sharing Agreement
5. Acquire/Deliver Full Content Set
6. Deduplication (Match & Merge)
7. Post-Ingestion Maintenance and Support
8. Indexing
9. Data Standardization, Mapping, Link Rules, and Enrichment by the Metadata Librarians
CDI: Surfacing Search Results

Delivered to discovery layers in real-time

Based on the user’s query, subscriptions, and customizations
What is Match & Merge?

System for improving search results retrieved from the CDI.

Three primary functions to achieve discovery goals
- Deduplication of citations
- Optimization of metadata
- Optimization of access
Match & Merge: How Does it Work?

Series of criteria governing which physical records can be combined, or “merged”, into a “logical record”, and under what conditions.
Match & Merge: Logical Records

- Logical records consist of two or more physical records, called “participant records”
- Represent a combination of all available metadata points
- The single, comprehensive record can be surfaced by the discovery layer

Differentiated from “grouped” or “best available” record approaches
Cancer: The RAS Renaissance
by Ledford, Heidi
Nature, ISSN 0028-0836, 04/2015, Volume 520, Issue 7547, pp. 278-280

DOI
Start/End Pages
Subject Headings
Abstract
And More
Match & Merge: Rules

**External identifiers**
- DOI
- PMID
- ISBN/EISBN
- ISSN/EISSN
- LCCN
- OCLC

**Internal identifier**
- Publication level
- Article level

**Other**
- Title only (Reference)
- URI only (Dissertation)
Match & Merge: Filters

Filters are the inverse of rules, dictating circumstances under which records cannot be merged

Mismatched metadata
- DOI
- PMID
- URI
- Language

Other
- Institutional Repositories
- Record exclusion flag
- “Bad Titles”
- Overmatch
A Word About Transitive Merge

Allows all records to merge into the same logical record, since both records A and C can merge with record B.
User-Uploaded Content

- Content uploaded by Primo users will not M&M
- Content uploaded by Summon users will M&M by default*
M&M Results – Display

Metadata surfaced for the user is sourced from all available records
Match & Merge Results: Access

**Summon**
Separate links are displayed for catalog records

**Primo**
API selects what it determines to be the most reliable link available, based on user’s rights and record metadata

Some resources in Primo always link via SFX/Alma
Match & Merge Challenges

Perfection is hard to achieve in any system

We encounter and work to resolve several issues

- Balance
- Granularity of indexing
- Transitive merge
- Metadata errors
- Exceptions
Addressing Match & Merge Challenges

- Periodically reevaluate and adjust the rules and filters
- Correct metadata errors
- Some problems cannot be fixed
Everyone Benefits From Match & Merge!

- Users get better search results
- Librarians can render services more effectively
- Content providers get their data represented in fair, unbiased manner, and have an equal opportunity to be linked to
- Helps Ex Libris comply with terms of Open Discovery Initiative (ODI)
Continuing Content Collaboration Initiatives

• Customer-focused
  • Content Working Group
  • Alma Community Zone Management Group
  • Ex Libris Idea Exchange

• Industry-level
  • NISO Committees: Discovery & Interchange Topic, NISO KBART Standing, and KBART Automation Committees
  • LOC Program for Cooperative Cataloging – CIP Partnership Program
  • Metadata 2020: Incentives for Improving Metadata Quality; Research Communications
  • Ex Libris Provider Zone
Q&A

cdi_info@exlibrisgroup.com