Urge to Merge

Adventures in WorldCat Duplicate Resolution

Casey Mullin
Head of Cataloging and Metadata Services
WWU Libraries
February 25, 2020
Overview

• Setting the scene
• The OCLC Member Merge Project
• Adventures in merging (scores and sound recordings)
• Future possibilities
A visit to the score backlog...
What will WorldCat deliver??

<table>
<thead>
<tr>
<th>Record</th>
<th>Main Entry</th>
<th>Title</th>
<th>Publisher</th>
<th>Date</th>
<th>L</th>
<th>H</th>
<th>Holdings</th>
<th>Cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto A, major, for flute or violin (or violoncello) or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Bach, Carl Philipp Emanuel, 1714-1788.</td>
<td>Concerto for flute or violoncello or harpsichord, strings and continuo, A major</td>
<td>E. Eulenburg</td>
<td>1947</td>
<td>1 mg</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A better way??

1) Pick the first decent-looking matching record, and walk away slowly

2) Spend 5-10 minutes sorting through all the results and send a merge request to OCLC

3) ???
The OCLC Member Merge Project

• Began in 2013 with 4 member libraries (the “first cohort”)
  • University of California, San Diego
  • University of Chicago
  • University of Pennsylvania
  • University of Washington

• 4 additional members added in 2017 (the “second cohort”)
  • Brigham Young University
  • University of California, Berkeley
  • University of Maryland
  • Western Washington University
The OCLC Member Merge Project

• 10 additional members added in 2018 (the “third cohort”)
  • Butler University
  • College of DuPage
  • Harvard University
  • Indiana University
  • National Library of Medicine
  • Oregon State University
  • University of California, Irvine
  • University of Colorado, Boulder
  • University of Delaware
  • University of Oregon
The OCLC Member Merge Project

• 24 additional members added in 2019-2020 (the “fourth cohort”)

**Part 1**
- Alaska Resources Library and Information Services
- Boise State University
- Center for Research Libraries
- Hebraica Merge Team
- Library and Archives Canada
- Morgan Library & Museum
- New Mexico State Library
- New York University
- Penn State University
- State Law Library of Montana
- State Library of Oregon
- University of Akron
- University of Alberta
- University of Kentucky
- University of Wisconsin–Madison

**Part 2**
- Washington Research Library Consortium (acting as a funnel)
  - American University
  - George Mason University
  - George Mason University Law Library
  - George Washington University
  - George Washington University Law
  - Georgetown University
- Loyola Marymount University
- Sacramento State University
- Saint Louis University Law Library
The OCLC Member Merge Project

• Member merge libraries are productive!
• Statistics for 1st half of Fiscal Year 19/20

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained records</td>
<td>1630</td>
<td>1590</td>
<td>1403</td>
<td>3041</td>
<td>1755</td>
<td>1888</td>
<td>11307</td>
</tr>
<tr>
<td>Deleted records</td>
<td>2713</td>
<td>2468</td>
<td>2136</td>
<td>1802</td>
<td>2624</td>
<td>3327</td>
<td>15070</td>
</tr>
<tr>
<td>Total records processed</td>
<td>4343</td>
<td>4058</td>
<td>3539</td>
<td>4843</td>
<td>4379</td>
<td>5215</td>
<td>26377</td>
</tr>
</tbody>
</table>

Source: OCLC Expert Cataloging Community Sharing Session, ALA Midwinter, Philadelphia, PA, January 24, 2020
The OCLC Member Merge Project

Impact on OCLC operations

• Member merge vs. OCLC staff statistics

<table>
<thead>
<tr>
<th>Member Merge Project</th>
<th>July 2019</th>
<th>August 2019</th>
<th>September 2019</th>
<th>October 2019</th>
<th>November 2019</th>
<th>December 2019</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained records</td>
<td>1630</td>
<td>1590</td>
<td>1403</td>
<td>3041</td>
<td>1755</td>
<td>1888</td>
<td>11307</td>
</tr>
<tr>
<td>Deleted records</td>
<td>2713</td>
<td>2468</td>
<td>2136</td>
<td>1802</td>
<td>2624</td>
<td>3327</td>
<td>15070</td>
</tr>
<tr>
<td>Total records</td>
<td>4343</td>
<td>4058</td>
<td>3539</td>
<td>4843</td>
<td>4379</td>
<td>5215</td>
<td>26377</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OCLC Staff</th>
<th>July 2019</th>
<th>August 2019</th>
<th>September 2019</th>
<th>October 2019</th>
<th>November 2019</th>
<th>December 2019</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained records</td>
<td>2988</td>
<td>53757*</td>
<td>17732*</td>
<td>1382</td>
<td>1459</td>
<td>620</td>
<td>77938</td>
</tr>
<tr>
<td>Comparison of MMP and OCLC (as %)</td>
<td>55%</td>
<td>3%</td>
<td>8%</td>
<td>220%</td>
<td>120%</td>
<td>305%</td>
<td>15%</td>
</tr>
</tbody>
</table>

*Special project undertaken during August-September 2019
Sources: Ibid; Laura Ramsey, OCLC Metadata Quality
The OCLC Member Merge Project

Impact on OCLC operations

• OCLC Metadata Quality merge queue: turnaround time of up to one year
  • Book format merge requests in queue*: 19,500
  • Score format merge requests in queue*: 984
  • Sound recording format merge requests in queue*: 315

*Figures as of February 14, 2020; some requests include multiple merge sets
Source: Laura Ramsey, OCLC Metadata Quality
The OCLC Member Merge Project

Process for joining

• Library must be a PCC member (NACO participation is sufficient)
• Library contacts OCLC Metadata Quality to express interest (AskQC@oclc.org)
• If successful, library invited to join next training cohort
• Cohort members attend training webinars and undertake a common exercise set
The OCLC Member Merge Project

Process for joining (continued)

• Each cohort member is assigned an OCLC reviewer, and member contact person begins sending merge sets to their reviewer
  • All cohort members *typically* start with books/e-books format
  • Record sets are identified by member staff (in the course of their day-to-day cataloging) or provided by the OCLC reviewer (taken from the OCLC merge request queue)
  • Member staff evaluate record sets and recommend whether to merge any of the records, which records to merge, and which record is to be retained
  • After review, member is free to perform the merge themselves in OCLC Connexion (using special “Master mode” authorization level and “secret” merge menu)
    • OCLC staff must facilitate these Connexion “upgrades”
The OCLC Member Merge Project

Process for joining (continued)

• After ~100 merge sets are submitted and reviewed, OCLC reviewer grants member library independence on the books/e-books format

• Member may request training on additional formats (rinse and repeat the aforementioned process)

• Member maintains contact with reviewer for ongoing support/questions (similar to the NACO-Music Project)
  • Bad merges (accidentally) made by member library can only be reversed by OCLC staff; good to report these immediately!
  • Can always escalate a difficult merge to OCLC staff to let them decide
Community Center

- Documentation: general merging principles
  - When a record has internal inconsistencies, especially between coded and textual data, tend to favor textual data or a preponderance of evidence, when appropriate.
  - Self-reported duplicate records may be merged if the record reported has only the reporting library’s holdings attached.
  - Apply looser criteria if the duplicate record is a vendor record (coded Encoding Level 3) or is batch-loaded (Encoding Level M) and has not been edited by another library.
    - CM’s experience: a preponderance of records merged qualify for these looser criteria.
The OCLC Member Merge Project

Community Center

Documentation

• Field by field merge guidelines (by format)
  • Like adaptations of BFAS Chapter 4 “When to input a new record”
  • Guidelines are to be consulted in conjunction with BFAS Chapter 4
  • Excerpt from score guidelines to the right

300 (Physical Description)

Subfield $a$ (Extent)

Number of pages may vary by 3 (when the total number of pages exceeds 10) and still be considered a match if all else matches on both records.

25 pages may match 27 pages

Variations in paging due to different methods used to record paging may be ignored, if all else matches on the records.

1 score (various pagings) may match 17, 27, 3 pages
1 score (1 volume) may match 1 score (421 pages)

If one record has a number of pages and the other has a number of volumes greater than one they are not duplicates.

3 scores (251 pages) does not match 1 score (251 pages)
1 score (3 volumes) does not match 1 score (3 volumes in 1)

If one record has parts and the other does not they are generally not considered a match.

1 score (50 pages) does not match 1 score (51 pages) and 3 parts
1 score (50, 25 pages) does not match 1 score (51 pages) and 1 part

See also FMus and 348.
1) Which fields automatically transfer in the merge transaction?
   • Instances of a field beyond the first are usually not automatically transferred (e.g., 028)
   • Rules are complex and not completely spelled out in the documentation
   • Fields not automatically transferred may be manually copied and pasted from records to be merged, per cataloger judgment

2) Which record should be chosen as the retained record?
## The OCLC Member Merge Project

### Community Center

#### Merge matrix

<table>
<thead>
<tr>
<th></th>
<th>New Zealand National Bibliography (042 = nznb or toknb)</th>
<th>National Library Records (040/c = DLC, GWDNB, NLC, NLM, and ELvl = blank or 1; or 040/c = any symbol and 042 = lac or pcc)</th>
<th>LC/LAC CIP Records (040/c = DLC or NLC and ELvl = 8)</th>
<th>Member (anything not in the previous outlined categories)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Zealand National Bibliography (042 = nznb or toknb)</td>
<td>*Ask OCLC</td>
<td>Retain National Library record</td>
<td>Retain National Library record</td>
<td>Retain National Library record</td>
</tr>
<tr>
<td>National Library Records (040/c = DLC, GWDNB, NLC, NLM, and ELvl = blank or 1; or 040/c = any symbol and 042 = lac or pcc)</td>
<td>Retain New Zealand National Bibliography</td>
<td>Retain National Library record</td>
<td>Retain National Library record</td>
<td>Retain National Library record</td>
</tr>
<tr>
<td>LC/LAC CIP Records (040/c = DLC or NLC and ELvl = 8)</td>
<td>Retain New Zealand National Bibliography</td>
<td>Retain National Library record</td>
<td>*Ask OCLC</td>
<td>Retain LC/LAC encoding level 8</td>
</tr>
<tr>
<td>Member (anything not in the previous outlined categories)</td>
<td>Retain New Zealand National Bibliography</td>
<td>Retain National Library record</td>
<td>Retain LC/LAC encoding level 8</td>
<td><strong>Retain most complete</strong></td>
</tr>
</tbody>
</table>

*Ask OCLC in all other cases*
The OCLC Member Merge Project

Community Center

Other features

• Contact information for OCLC Metadata Quality staff and other member merge libraries

• Discussion forum
  • Questions and answers
  • Can subscribe to forum topics and receive emails instead of logging in
  • Would LOVE to see more music folks posting to the forum!
The OCLC Member Merge Project

WWU’s Merging Journey

- Early 2017: cohort training
- December 2017: independent in books/e-books
- March 2018: independent in scores
- October 2018: began reviewing books/e-books for Oregon State University
- February 2019: independent in visual materials
- September 2019: independent in cartographic materials
- October 2019: began reviewing books/e-books for State Library of Oregon
- January 2020: independent in sound recordings
The OCLC Member Merge Project

WWU’s Merging Journey

- Cumulative statistics, October 2017-January 2020
  - 2946 merge sets (retained records)
  - Average of 105 sets per month
- One of only 4 libraries independent in scores
  - One additional library joining program soon
- One of only 3 libraries independent in sound recordings
  - One additional library joining program soon
- WWU first member library to...
  - Attain independence in 5 formats!
  - Serve as reviewer for another library
The OCLC Member Merge Project

Seeing is believing
Adventures in Merging

General approach
“Targeted” searching vs. “dragnet” searching

• Targeted
  1. Command line query: mn:ee6360
  2. Triangulate record with the highest number of holdings that matches item
  3. Download “best” record, edit if necessary/inclined
  4. Go about your day

• Dragnet
     “Minimum viable precision”
  2. Pin all records that are at least a near match to the item in hand
  3. Disqualify legitimate* variant editions and merge the rest
  4. Pat self on back and download merged record into local catalog

*Remembering that Level M and Level 3 records warrant looser criteria
Adventures in Merging
Adventures in Merging

- All records have language of cataloging = English (good)
- All records but two have a precise publication date of 1967 (matches item in hand)
- Some records have Wq number(s), which do not appear on the item in hand, in the 245; will probably disqualify these
  - If Wq. numbers appear in square brackets, do not disqualify
- Record with 168 holdings is marked as DLC, and the title matches item in hand
  - This record will serve as the benchmark against which all other records will be compared
- Next step: reviewing records against the benchmark (in full display)
Adventures in Merging

Culling the Herd

• An initial scan of the 10 records found...
  • 2 records with significantly different titles; disqualified
  • 2 records for online versions; disqualified
    • Could filter these out from the outset in the command line search
  • 6 records remain as candidates for merging
    • 1 is Level I, DLC (benchmark record)
    • 4 are Level M
    • 1 is Level 7

• Next, tile windows vertically to see three full records at a time
  • Benchmark record stays on left; other records are minimized/expanded as needed

• Reverse engineer DDR
  • “Why didn’t DDR merge this record and is that a legitimate reason to keep it separate?”
Adventures in Merging

Concertos, Violoncello and string orchestra, W. 172, a major : #b concerto, a major for flute or violoncello or harpsichord (piano) and string orchestra / #c by Carl Philipp Emanuel Bach.
Adventures in Merging

Closing the Deal

• Connexion can merge up to 10 records in one transaction
• Identify any fields that will need to be manually copied over
  • In this case, happily none
• With all records open, begin the bib linking and merging process
• Merged record numbers will be added to the 019 field of the retained record
Adventures in Merging

Other Examples

1) no:1038072608 | no:1033587217 | no:1031379901 | no:1048200577

- Title: Rogers & Hammerstein’s Oklahoma! (or just Oklahoma!)
- Match point across all records: 024 (UPC)
- Imprint varies:
  - New York, NY: ℃b Decca Broadway/Verve/UMe, ℃c [2018]
  - London: ℃b Decca Broadway, ℃c [2018]
  - New York: ℃b Decca Broadway/Verve/UME, ℃c [2018]
  - [United States]: ℃b Verve, ℃c 2018.
- Two records have a matching edition statement (75th anniversary.) but their 505s don’t match
- Decision: do not merge any records
Adventures in Merging

Other Examples

2) no:41636395 | no:51939334

- 245s
  - My hero: ℓb from the Chocolate soldier: song / ℓc words by Stanislaus Stange; music by Oscar Straus.
  - My hero: ℓb from the chocolate soldier / ℓc music by Oscar Straus; words by Stanislaus Stange.
- Plate number matches (028)
- Imprint, pagination and dimensions match (260/264, 300)
- Series matches (490)
- BUT, edition statements do not match
  - Operatic, In D (d to g). ← My item in hand
  - Edition for high voice (d to g) in D.
- And... both records are encoding level I
- Decision: do not merge records
Adventures in Merging

Common Themes
(“Reverse-engineering DDR”)

• Order of elements in 245
• 245 subfield placement
• Source of title (title page vs. cover vs. caption)
• Choice of publisher/distributor
• Different supplied dates of publication
• Dates of multivolume sets (closed vs. open)
• Description of accompanying material (300 $e vs. 500)
• Wrong workform for the format
• And so much more...
Future Possibilities

• M-level records project at WWU
  • Thousands of original records were batchloaded into OCLC at the time of retrospective conversion (mid 1990s) and remain M-level to this day
  • Many of these can be matched and merged to full-level records without item in hand
  • Those that require examining item in hand and still cannot be merged will be upgraded to full-level (authenticated as BIBCO when feasible)
  • Ultimate goal: no records with 040 $a/$c XFF and ELvl M

• Orbis Cascade Alliance merging hub idea
  • 4 of our 37 members are merge libraries (3 are active)

• MOUG mergers affinity group??
Learn More

• Merge documentation is not password-protected
  • https://help.oclc.org/WorldCat/Metadata_Quality/Member_Merge

• Interested in learning more, or to become a member merge library?
  • Email AskQC@oclc.org

• Remember to “catalog defensively”
  • https://oc.lc/cataloging-defensively

• Questions for me: casey.mullin@wwu.edu

• Questions for Jay: weitzj@oclc.org

Thank you!!