

OCLC Deduping Project Reports

When the deduping project was completed by OCLC, the duplicates which were identified were immediately resolved in MOBIUS by the matching of OCLC record #s.

Within the cluster catalogs, catalogers will need to manually resolve the duplicates. A cluster can choose to do nothing as far as MCO is concerned. The duplicates have been deduped in the central catalog. The cluster catalog is no worse off than before, since the duplicates were there before. This project has just identified them for us and made it possible for us to dedupe them.

It is important to realize that this project accomplished more than just locating duplicates. The project also added OCLC holdings to all Bridges records that matched OCLC records, even when there was no duplicate record. These do not appear in our duplicates lists, since they are not duplicated in Bridges. These can still be identified by the 003, but we would need to run our own lists to find them.

Jim demonstrated the method he recommends for resolving duplicates using MilCat. MilCat needs to be set for Multiple Windows (SETTINGS – WINDOW – MULTIWINDOWS MODE) for this to work.

1. Retrieve both records and tile them vertically.
2. Compare the records to ensure that they truly are duplicates. Follow the Bridges policies found in the document Standards for Merging Bibliographic Records on Bridges, posted on the Committee's webpage.
3. If the records are truly duplicates, decide which record to retain. Generally, it is easiest to keep the original OCLC record. However, if the record from the dedup project is clearly a more complete record, keep it instead.
4. If you are keeping the record with the 003 field of OCLC DEDUP (this is the record to which the OCLC # was added during the deduping project) retain the 003 in the record. This will allow us or MCO to create a review file of these records in the future, if some need should arise. Edit the 003 to read:

OCLC DEDUP b

(The space and lowercase b will identify Bridges ones uniquely)

5. Move any fields which should be retained from the record to be deleted to the record being retained. Follow the Bridges policies found in the document, Tags To Be Preserved During Merge or Overlay in Bridges, posted on the Committee's webpage.

6. Transfer all attached records from the record being deleted to the record to be retained.
7. On the record to be deleted, click on EDIT – TRANSFER ATTACHED. You will be given the opportunity to select the bib # of the record to be retained. (Usually this record is already displayed as one of your tiled records. Then you just need to accept the bib # that appears in the box). The advantage of MilCat over GuiCat is that in MilCat you can see and move order & checkin records from other accounting units.

Jim warned us to beware of mismatches, where the records are not really duplicates. A random sample done on MERLIN's database showed that approximately 6% of the records given OCLC #s were given the wrong OCLC #. So far, these appear to be instances of different editions of the same title. None, so far, have been mismatches involving different titles. The cataloger who has been working for MCO doing deduping for the various clusters has been assigned to work on the MERLIN dedup lists so that mismatches can be rectified. Since MERLIN records comprise a larger percentage of the master records than any other cluster, this will help rectify mismatches in the Union Catalog more quickly.

The question was raised as to whether to retain a bib if it had no attached records. Several of the Bridges libraries indicated they wanted these records retained, so the decision was made to retain them. NOTE: This policy refers to bib records that have never had any attached records. They are a result of the initial load into III. This does not refer to bib records where the attached items, etc. have been transferred to another bib record. These can be deleted when you are finished deduping.

Jim recommended opening the dedup list online via Word or another program which would allow you to delete an entry from the list as you finish it. This also allows you to copy and paste information from the entry if you need to email another holder on a record to check their copy of the title.

If you find a mismatch when comparing two records:

1. Go to OCLC and find the right OCLC # to insert in the 001 field of the bib record in Bridges.
2. Go to the "wrong" OCLC # record in OCLC and delete your OCLC holding symbol, which was added as part of the dedup project.

The following points came up during the discussion and questions/answers period:

- Only those libraries on the record which has the 003 OCLC DEDUP receive the report of duplication on their Dedup report.

- When items, orders, and/or checkins are moved from the record being deleted to the record being retained, the retained record will not show in the scope of the moved holders until the next day, unless they manually add themselves by editing the bib location field.
- If a record has a 912 tag containing DDR, it means the record was identified as a duplicate via the final matching algorithm used by OCLC. This algorithm is not as good as the others used by OCLC and so those records should be checked very carefully to see if they really are dups.
- We should ignore the “Headings Used For the First Time” sequence of records that came with the OCLC project duplicate reports. MCO failed to turn off the generation of this report when running the duplicate reports for Bridges.
- It is important for ILL folks to report mismatches which they find as they respond to OCLC ILL requests.
- Jim would like to be notified if we find any mismatches where the titles are different.