

CARLI



Consortium of Academic and
Research Libraries in Illinois

“New” Developments Related to I-Share Cataloging Workflows

ICAT Forum, April 9, 2008
Casey Sutherland, CARLI Office

Three main topics to discuss

- (1) Changes to extracts from local databases for Universal Catalog loads
 - Main effect is on the Suppress/Replace Routine
- (2) Analysis of 2007 UC rebuild log files
 - “Second pass discards” is AGT!
 - Some problematic data in 035 \$a needs editing
- (3) Changes at OCLC regarding backloading
 - Volunteers to test backloading original cataloging?



(1) Changes to extracts for Universal Catalog loads

- As of November 13, 2007, CARLI began extracting data on an hourly basis from the local databases, that is fed into the Universal Catalog (UC).
- The ultimate goal is to allow more workflow flexibility in the “Suppress/Replace Routine.”
 - “Replace” means both bib overlay/replaces and edits to any 035 \$a data
 - Reminders about this routine’s details to follow in next section.
- We needed to find a balance between system performance concerns with predictability of outcomes for cataloging staff.



(1) Changes to extracts for Universal Catalog loads (cont.)

- The data is extracted from the local databases hourly, beginning at x:45.
 - X = the variable hour of the day.
 - The extract job completes in 3-10 minutes (across all databases).
 - The exact time the extract will start and complete in a particular DB will vary, by day/hour.
 - The exception is the early morning hours, due to circ batch job processing and the server bounce that happens daily.
 - No extracts done at 1:45 a.m., 2:45 a.m., or 3:45 a.m.
 - The 4:45 a.m. extract will include any transactions done after the 12:45 a.m. extract has completed.
- The data is still loaded into the UC once per day, beginning at 9 p.m.

(1) Changes to extracts for Universal Catalog loads (cont.)

- What does this mean for my workflow?
 - Reminder: the suppress sends a delete transaction to the UC; the replace sends an update transaction to the UC.
 - Multiple delete and update transactions on the same bib must be sent to the UC in separate files, so the right thing happens during the UC load.
 - Until Nov. 2007, with only one extract per day, the delete and update transactions had to take place on different days to be placed in separate files.
- Although you won't be able to see the results in the UC of the new extract routine until the next day, you can now do a suppress and a replace during the same work day!
 - This applies to both bibs with "Suppress from OPAC" on the System tab or "Suppress from UC" via 049 \$u nouc.



(1) Changes to extracts for Universal Catalog loads (cont.)

- If you do the suppress between x:00 and x:44 each hour, you can do the replace as early as the top of the next hour.
 - The extracted data is loaded in chronological order (per source DB) once the loads actually begin at 9 p.m.
- If you do the suppress between x:45 and x:59 each hour, you need to wait until the top of the following hour to do the replace.
- The clock that matters is the server's clock.
 - The bib history tab will show you what time the server records each transaction.

(1) Changes to extracts for Universal Catalog loads (cont.)

- Example 1:
 - Bib suppressed at 8:15 a.m. (per the History tab)
 - Bib can be replaced as early as 9:00 a.m. (same day)
- Example 2:
 - Bib suppressed at 9:44 a.m. (per the History tab)
 - Bib can be replaced as early as 10:00 a.m. (same day)
- Example 3:
 - Bib suppressed at 9:50 a.m. (per the History tab)
 - Bib can be replaced as early as 11:00 a.m. (same day)



(1) Changes to extracts for Universal Catalog loads (cont.)

- You won't actually see the results of these transactions until Day 2, because the loads will still occur once per day.
- As long as you watch the server "clock" (per the extracts), the load order will control the outcome.
 - When we've had problems with daily UC loads, it has been during the load process, not usually with the extracts.
 - The revised UC load scripts will honor the extract order, so the results should be predictable, even if there are load problems.
 - Hopefully, no more postings to techserv-ig telling libraries to wait another day to do the replace, if there are UC load problems!



(1) Changes to extracts for Universal Catalog loads (cont.)

- If you don't want to adopt the new routine, you can still do a suppress on "Day 1" and a replace on "Day 2."
 - For example, CARLI staff macro jobs that affect 035 \$a data will likely stick with the Day 1/Day 2 routine.
- Some ICAT team members mentioned they might do suppresses in the morning and replaces in the afternoon.
 - This should be fine as long as the suppress happens before 11:45 a.m. and the replace after 12:00 p.m.

(1) Changes to extracts for Universal Catalog loads (cont.)



- Questions about this topic???



(2) Analysis of 2007 Universal Catalog Rebuild log files

- CARLI usually rebuilds the Universal Catalog each summer.
- A UC rebuild means all of the bibs in the UC are deleted, and then reloaded via full extracts from all of the local databases, with “fresh” data.
 - This is the most efficient way to incorporate “new libraries” into the UCdb.
 - This is the best way to overcome human errors in the suppress/replace routine.

(2) Analysis of 2007 Universal Catalog Rebuild log files (cont.)



- After the summer 2006 UC rebuild, CARLI Office staff determined a new processing routine that would reduce the overall number of discards from the UC.
- Background info:
 - UC's duplicate detection matches on 035a data and the combination of LCCN and ISBN/ISSN in the first load.
 - Discards happen when the incoming bib matches more than one existing UC bib.
 - The loader program doesn't know which is the "correct" existing bib, so the incoming bib is not added/existing bib not replaced, and the library's holdings are not set on the new/existing bib.
 - This is A Bad Thing.

(2) Analysis of 2007 Universal Catalog Rebuild log files (cont.)



- When discards happen in the UC, the incoming bib (in MARC format) is copied to a file.
- The “new” routine is to re-load the discard file into the UC a second time, using a duplicate detection profile that matches only on the bib’s 035 \$a (usually the OCLC control number).
 - Many discards from the first load contain a perfectly good OCLC number.
 - We call this routine the “second pass discard” loads.

(2) Analysis of 2007 Universal Catalog Rebuild log files (cont.)



- During the 2007 UC rebuild, only 807 total bibs were discarded after the second pass was performed, as part of the initial rebuild!
 - The total number of bibs fed into the UC from all 71 databases during the 2007 rebuild was 24,240,517.
 - The total number of “first pass discards” during the 2007 rebuild was 48,662.
 - The “second pass” reduced this total to 807.
 - This “second pass discards” routine is A Good Thing!

(2) Analysis of 2007 Universal Catalog Rebuild log files (cont.)



- For daily loads since the 2007 UC rebuild, the number of discards after the “second pass” are still significantly less than the total number of discards.
- Stats below are from July-Oct 2007 and Mar 2008
(Nov 2007-Feb 2008 stats are unavailable for this analysis)
 - The total number of ongoing updates and deletes fed into the UC from all 71 databases was 1,166,070.
 - The total number of “first pass discards” (both updates and deletes) was 10,981.
 - The total number of discards remaining after the “second pass” (both updates and deletes) was 3,702.
 - This “second pass discards” routine is A Good Thing!

(2) Analysis of 2007 Universal Catalog Rebuild log files (cont.)



- What kinds of records are still being discarded after the “second pass”?
- CARLI staff analysis shows four main patterns (see Excel handout)
 - Different patterns in initial rebuild loads vs. ongoing daily loads
 - (1) Suppress/replace routine not performed properly
 - (2) False matches on “unprefixed” 035a data and/or 035a (XXXdb)nnnnnnn
 - (3) Incoming bib record contains more than one OCLC number in multiple 035s
 - (4) Incoming or existing UC bib has “second 035”



(2) Discard scenario #1: Suppress/replace routine problems

- Basic suppress/replace routine (required when adding, changing, or deleting the 035 \$a field on a currently unsuppressed bib):
 - When the bib is not currently suppressed (in either the bib's System tab or via the 049 \$u nouc option) ...
 - On Hour1/Day 1: suppress the bib (either method) as is, without making ANY changes to the 035 \$a data.
 - This action sends a delete transaction to the UC
 - On Hour2/Day 2: update/replace the bib to add/edit the 035 \$a data, and remove the suppression.
 - This action sends an update to the UC, with the new data
- This routine does not apply to any bib edits other than 035 \$a changes.

(2) Discard scenario #1: Suppress/replace routine problems (cont.)



- If the bib is currently suppressed (either method), no need to use the two-hour/day process, since the bib is not represented in the UC to begin with.
 - Edit the 035 \$a as needed on Hour1/Day 1
- More details on this routine found the revised version of “Safe Bibliographic Record Replacement Routines” available at:
<http://www.carli.illinois.edu/mem-prod/I-Share/cat/safebibrep.pdf>
- More details on using the “Suppress from UC” (049 \$u nouc) option available at:
http://www.carli.illinois.edu/mem-prod/I-Share/cat/UC_suppr_049u.html

(2) Discard scenario #2: “False matches” on 035 \$a data

- False matches on unprefixed 035 \$a data or 035 \$a (XXXdb)nnnnnn
 - When copying a bib from the UC into the local database, before saving the record, delete 035 \$a fields that look like these:
 - **035 \$a 4362554** (i.e., a single string of digits in \$a, with no prefix; this is usually the Voyager bib ID from the UC)
 - **035 \$a (XXXdb)2646153** (where the XXX is an I-Share library three letter code, followed by the bib ID from the local DB)
 - Do not delete the OCLC number, which looks like this:
 - **035 \$a (OCoLC)ocmNNNNNNNN (eight digits)**
 - **035 \$a (OCoLC)ocnNNNNNNNNNN (nine digits)**
 - **035 \$a ocmNNNNNNNNNN** (in some older Voyager databases)

(2) Discard scenario #2: “False matches” on 035 \$a data (cont.)



- New Access query on Shared SQL page to help libraries find records with 035 \$a (XXXdb)nnnnn in the local database.

<http://www.carli.illinois.edu/mem-prod/I-Share/secure/sql.html>

- New Macro on Shared Macro page that will delete the 035 \$a (XXXdb)nnnn in the local database, while leaving “good” 035 \$a data intact.

<http://www.carli.illinois.edu/I-Share/secure/macros/>

- CARLI staff will continue working with local DBs on macro projects to delete this and other undesirable 035 \$a data.
 - Since 2005, CARLI macros have deleted over 540,000 instances of undesirable 035 \$a data from the local DBs.

(2) Discard scenario #3: Bibs with multiple OCLC numbers

- Single bib record with more than one (different) OCLC number
 - Unclear what workflow is creating this scenario, but examples are found in multiple I-Share databases.
 - These records almost always produce UC discards!
 - Access query on Shared SQL page to help libraries find these records in the local database, for subsequent manual correction.

<http://www.carli.illinois.edu/mem-prod/I-Share/secure/sql.html>



(2) Discard scenario #4: Bibs with “second 035” data

- This began in November 2006 when OCLC changed how they output OCLC control numbers.
 - See the URL below for more details on the cause:
<http://www.carli.illinois.edu/news/16/65.html>
- CARLI scripts delete the second 035 for routine, automated bulk import loads.
- Library staff must set the Export preference on each PC that uses Connexion to delete the 035 field from bibs exported from OCLC and then imported manually via the Voyager cataloging client.

<http://www.carli.illinois.edu/mem-prod/I-Share/cat/oclcnmbrs.pdf>

(2) Analysis of 2007 Universal Catalog Rebuild log files (cont.)



- Questions on this topic???

(3) Changes at OCLC regarding backloading

- When CARLI worked with OCLC to implement backloading for the 2007 new I-Share libraries, it was discovered that OCLC’s policy on adding original bibs via backloading had been “relaxed.”
 - Previous OCLC policy on evaluation of “conditional adds” was very restrictive and not practical in our environment.
 - This resulted in the I-Share requirement that original cataloging be done in OCLC, not in Voyager.

(3) Changes at OCLC regarding backloading (cont.)

- New OCLC policy is more “relaxed” but still somewhat unclear:
 - A library’s data must still undergo an evaluation process.
 - Not sure how many records are needed for this evaluation.
 - Not sure how long the evaluation period will last.
 - OCLC batch services staff do these evaluations “as time allows.”
 - Still no guarantee that OCLC will approve a library’s records for “unconditional adds.”
- Good I-Share practice would mean staff would need to go back into Voyager bib and add the OCLC number, after it has been added to WorldCat.



(3) Changes at OCLC regarding backloading (cont.)

- But OCLC is willing to work with us on getting past the evaluation step to (hopefully) creating a standard, shorter evaluation process that would apply to all I-Share libraries.
 - OCLC recognizes that our data is more similar than it is different, due to our shared Voyager system.
- We need some volunteer libraries to do some original cataloging in Voyager, to be evaluated by OCLC.
 - If interested, contact the CARLI Office (support@carli.illinois.edu)
- If you are not a volunteer now, continue to do your original cataloging on OCLC until further notice!

(3) Changes at OCLC regarding backloading



- Questions about this topic???



For More Information

- CARLI website:
<http://www.carli.illinois.edu>
- Selected Cataloging Documents on CARLI/I-Share page:
<http://www.carli.illinois.edu/mem-prod/I-Share/cat.html>
- CARLI Office:
support@carli.illinois.edu

Thank You For Your Attention!