



Batch Deletion for Dummies: Removing BIB and MFHD Records from your Database with Bulkimport

Endeavor Users Group Meeting
Friday April 21, 2006
Chicago, Illinois



Welcome and Introductions

Paul Anderson
DeVry University
Panderson@devry.com

Laura Guy
Colorado School of Mines
lguy@mines.edu

What We Hope to Accomplish

We'll show you the server-side steps to delete large numbers of BIB and/or MHFD records:

- Efficient
- Fast
- Powerful
- You must have direct access to the server
- Don't need to install Oracle Client

Using Bulkimport You Can:

- Delete BIB records without MFHDs.
- Delete BIB and all attached MFHD records.
- Delete MFHDs from BIB records and leave the unattached BIB record in place.
- Delete one of several MFHDs from a BIB, leaving the other MFHDs and the BIB untouched.

Steps to a Successful Deletion Project

1. “Spec” out your project – goals and steps.
2. Do preliminary setup in System Administration.
3. Remove any attached ITEM records.
4. Use Marcexport to create a file of records to delete.
5. Use the Prebulk utility to create an interleaved file of bibliographic and MFHD records (if needed).
6. Use Bulkimport to delete the records using the appropriate import parameters.

Step 1: Specification

- Determine your goals: what do you want to do?
- Determine steps you will need to accomplish your goals.
- Review the documentation.
- “Measure twice cut once” principle.

Step 2: Preliminary Setup in System Administration

- Create a Bibliographic Duplication Detection Profile.
- Create a Bulkimport Rule for Batch Delete.

Duplicate Detection Profile

- Set Duplicate Handling to **Replace**.
- Field Definitions should be set to Match on **BBID (Bibliographic Record Number)**

Creating the Bulkimport Rule

- Be sure to use the Bibliographic Duplication Detection Profile you created.
- The Owning Library must match the database you are working in.
- The **Load Bib/Auth Only** checkbox on the Rules tab must be checked.

Step 3: Remove Item Records

- Use Access Reports and Pick & Scan, or some other method, to remove ITEM records by creating a list of barcodes you want to delete.
- Some manual cleanup may be necessary.

Step 4: Create a MARC file of BIB Records to delete

- Step 4a: Create a list of BIB Numbers for records you wish to delete.
 - Use Access reports to identify records.
 - Must use Bibliographic Record Numbers as identifiers
 - Transfer list of BIB Numbers in .txt format to the Voyager server
 - Keep track of record counts for later checks

Step 4: Create a MARC file of BIB Records to delete

Sample Access Query

```
SELECT MFHD_MASTER.MFHD_ID, LOCATION.LOCATION_CODE,  
BIB_MFHD.BIB_ID  
FROM BIB_MFHD RIGHT JOIN (LOCATION INNER JOIN  
MFHD_MASTER  
ON LOCATION.LOCATION_ID = MFHD_MASTER.LOCATION_ID) ON  
BIB_MFHD.MFHD_ID = MFHD_MASTER.MFHD_ID  
WHERE (((LOCATION.LOCATION_CODE)=[LOCATION CODE]));
```

Step 4: Create a MARC file of BIB Records to delete

- Step 4b: Use the Marcexport utility in conjunction with your file of BIB Numbers to export the bib records to a MARC file.
 - Remember to Review Your Log File!
 - Keep in mind character set issues (use defaults)
 - You export BIBs ONLY – if you're using Prebulk and even if you're deleting selected MFHDs later!

Step 4: Create a MARC file of BIB Records to delete: Export parameters.

- ❑ Export the file as a BIB records only file if you're going to use Prebulk to delete some but not all MFHDs.
- ❑ Specify the input filename of BIB Numbers to identify records to be exported.
- ❑ Specify the output filename of MARC records to be exported.
- ❑ `/m1/voyager/xxxdb/sbin/Pmarcexport -rB -mM -t<input file name> -o<output marc filename>`

Step 4: Create a MARC file of BIB Records to delete: Export parameters.

If you're deleting BIBs with all attached MFHDs (not using Prebulk):

- ❑ Specify the input filename of BIB Numbers to identify records to be exported.
- ❑ Specify the output filename of MARC records to be exported.
- ❑ `/m1/voyager/xxxdb/sbin/Pmarcexport -rG -mM -t<textfilename> -o<output marc filename>`

Step 5: Create an interleaved file of BIBs with MFHD's to delete.

- To delete some MFHD's, the BIBs in the MARC file must have a MFHD attached with the location of the record to delete.
- Use Prebulk to add MFHD's to BIBs in export file.
 - Set up prebulk configuration file.
 - Run the Prebulk utility on your MARC file.

Setting up the Prebulk configuration File

- Prebulk configuration files are usually stored in the /m1/voyager/xxxdb/local directory.
- Default file named Prebulk.cfg
- Archive multiple cfg files for different prebulk jobs. (Prebulkx.cfg)
- Prebulk.cfg files **MUST** be edited with vi editor on the server.
- Data elements on each line must be separated with a <TAB>

Setting up the Prebulk configuration File (continued)

- Prebulk.cfg file is divided into stanzas
- 5 stanzas required in prebulk.cfg file
 - [OVERRIDES]
 - [MFHDTAG]
 - [LOCATIONS]
 - [CALLTYPES]
 - [MAPPING]

Overrides Stanza

- The overrides stanza is required to tell prebulk to create an interleaved MFHD with each record processed.
- `CREATEMFHD=yes`

MFHDTAG Stanza

- The MFHDTAG stanza tells prebulk where to look in the MARC record for MFHD data.
- MFHDTAG stanza must be present even though no data are being pulled from the MARC record.
- Use XXX as the value.
- XXX forces prebulk to use location code specified in the [LOCATIONS] stanza.

[LOCATIONS] Stanza

- Specifies the location code for the MFHDs you wish to delete.
- 3 elements
 - Input location
 - Output location
 - Call number hierarchy
- Specify the location to delete as both the input and output location.
- Use 050 and 090 fields for call number hierarchy.
- LOCATION<TAB>LOCATION<TAB>050,090

[CALLTYPES] STANZA

- Specifies the the indicator to be used in 852 field of interleaved MFHD.
- Must be present even though no call number is being created.
- Use 0 as default indicator.
- 050 and 090 on separate lines.
 - 050<TAB>0
 - 090<TAB>0

[MAPPING] Stanza

- Specifies the MARC tag and subfield for location in MFHD in interleaved output file
- L<TAB>852b



Sample prebulk.cfg File

[OVERRIDES]

CREATEMFHD=YES

[MFHDTAG]

XXX

[LOCATIONS]

zLINDAHALL	zLINDAHALL	050,090
------------	------------	---------

[CALLTYPES]

050	0
-----	---

090	0
-----	---

[MAPPING]

1	852b
---	------

Running the Prebulk Utility

- Specify name of input file.
 - MARC file created with MARCEXPORT in previous step.
- Specify name of output file.
 - This is the file you will Bulkimport in the next step
- Specify path and name of prebulk.cfg file.
- Example command line from sbin directory:
`Pprebulk -iinputfile.mrc -ooutputfile.mrc
-c/m1/voyager/xxxdb/local/prebulkx.cfg`

Step 6: Bulkimport the file of records.

- It may seem non-intuitive but you import to delete!
- MFHDs with the Location specified in the Prebulk config file will be deleted, and the BIB record if there are no other MFHDs attached.
- Use the function's delete parameters
- Run the program.
- Check the log files.

Step 6: Bulkimport the file of records.

- ❑ Specify the input file
- ❑ Specify the Bulk Import Rule
- ❑ Specify what you're deleting (BIBs & MFHDs)
- ❑ Import in groups of less than 5,000
- ❑ /m1/voyager/xxxdb/sbin/Pbulkimport
 - f <filename> – i <Bulk Import Rule>
 - b1 –e1000 – r – x

Caveats

- To delete with Bulkimport, you must log in as *voyager* user.
- Test before you go.
- Download the *Batch Deletion for Dummies* file from knowledge Base for More information.
- When in doubt, check the manual.

Questions.

Batch Deletion for Dummies: Removing BIB and MFHD Records from your Database with Bulkimport Supplemental Materials

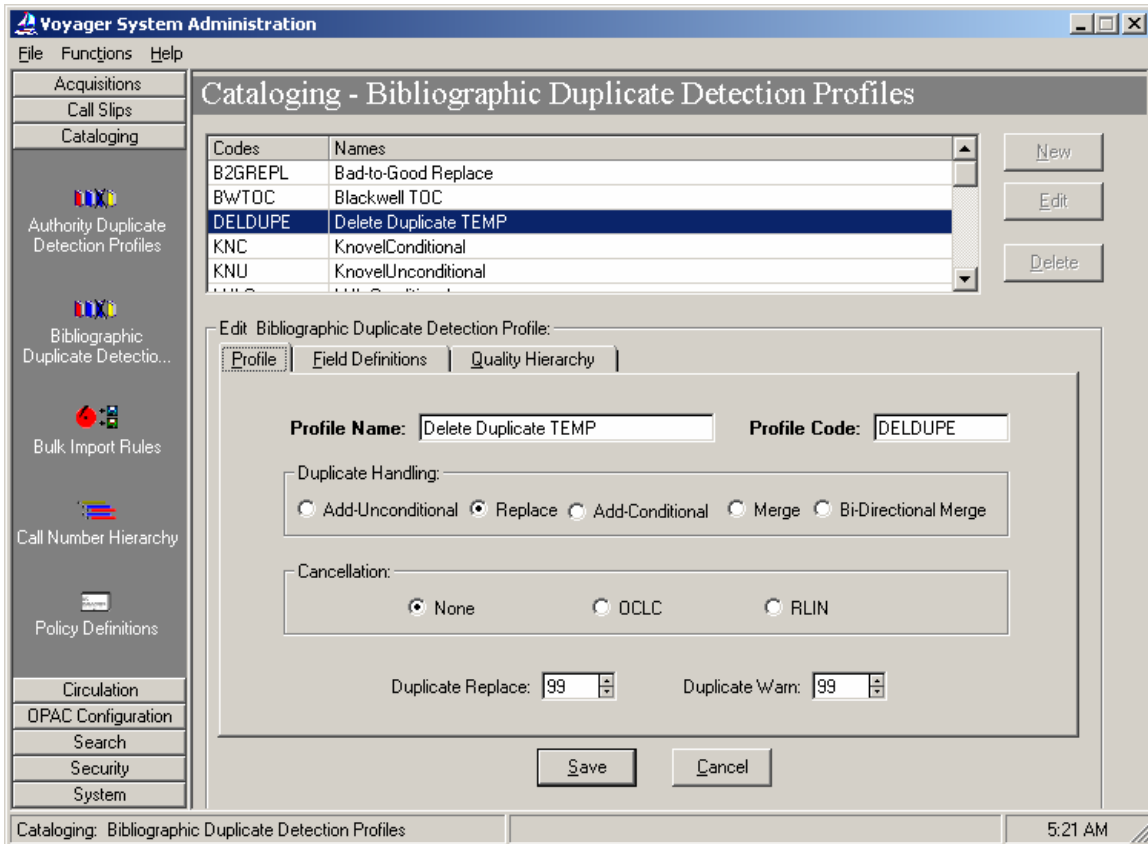
Table of Contents

- I. Bibliographic Duplicate Detection Profile**
- II. Bulk Import Rule**
- III. MARCEXPORT**
- IV. PREBULK**
- V. BULKIMPORT**

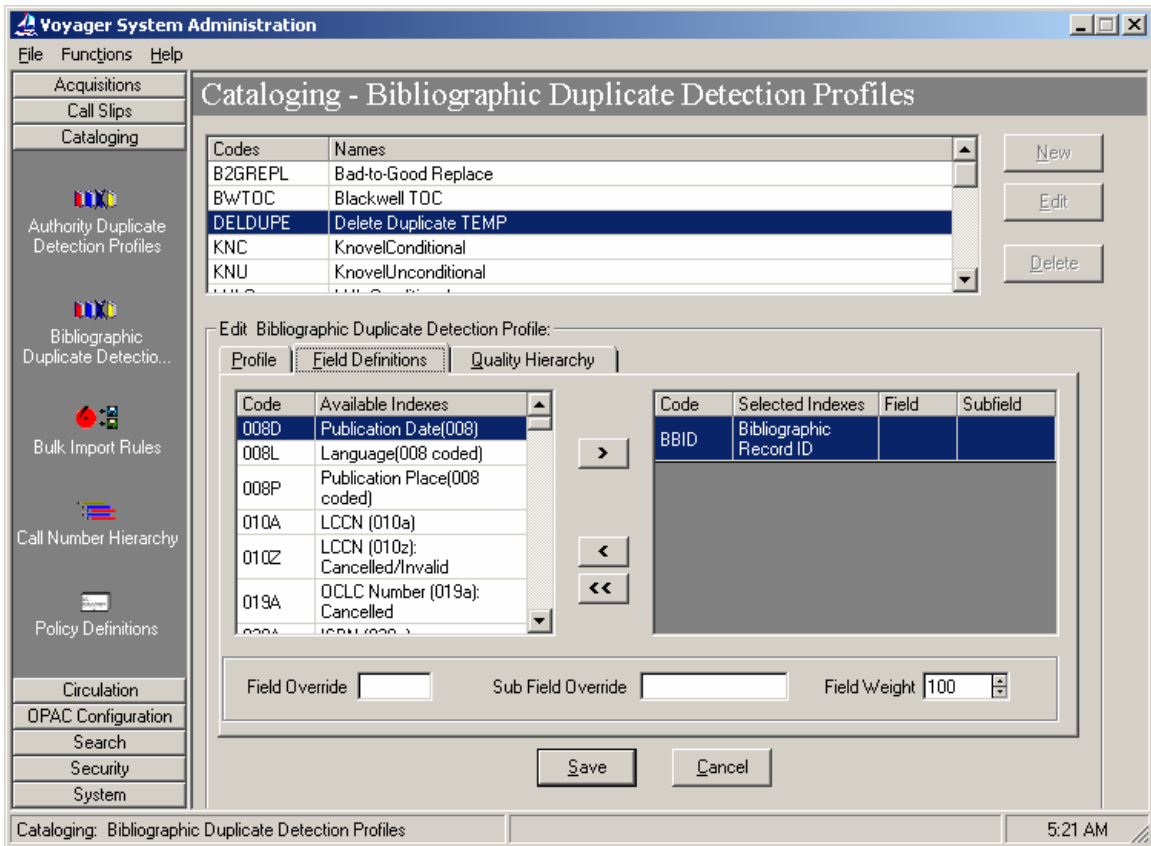
I. Bibliographic Duplicate Detection Profile

When importing bibliographic records this profile is used to determine how the systems should handle the incoming records. It can be used to replace a matching record in the database with a new record.

Note for our purposes Duplicate Handling is set to “Replace”:



Notice that BIBID is used as the duplicate detection key. You'll find it way at the bottom of the available indices:

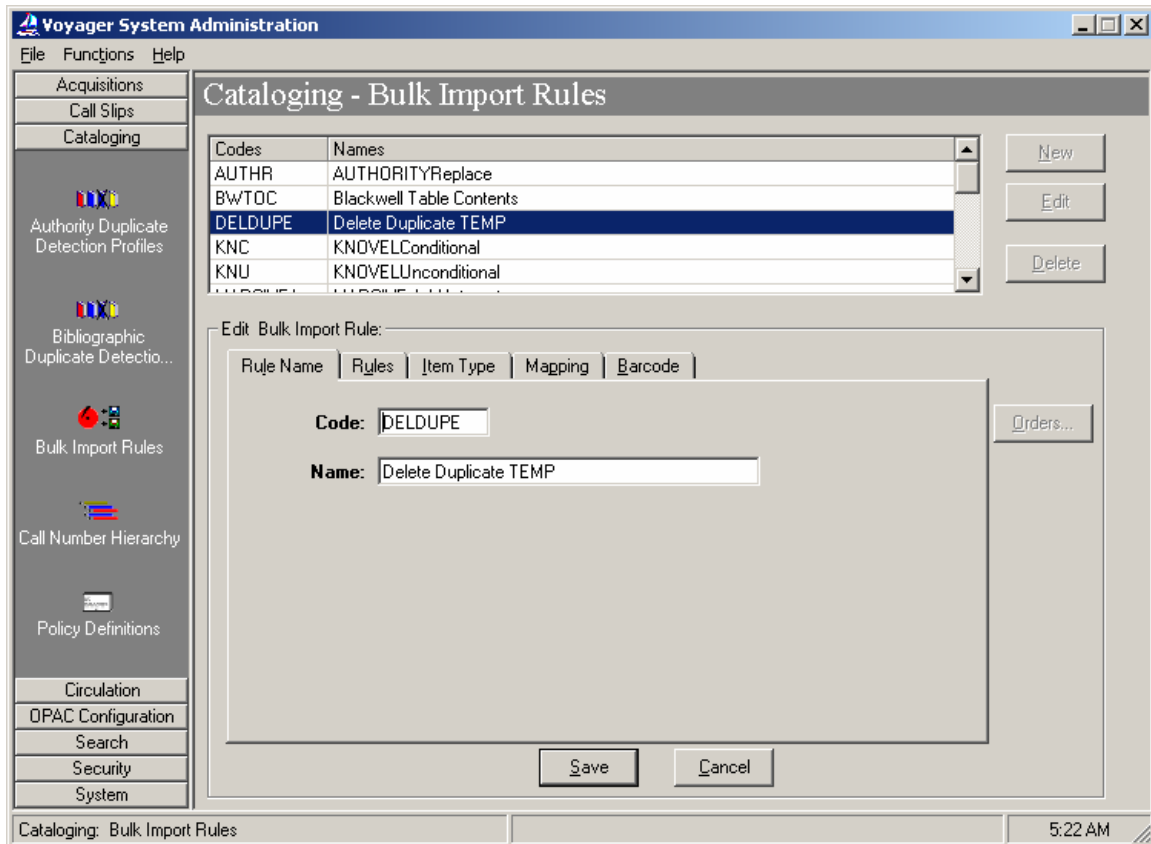


(Quality Hierarchy Tab is unchanged)

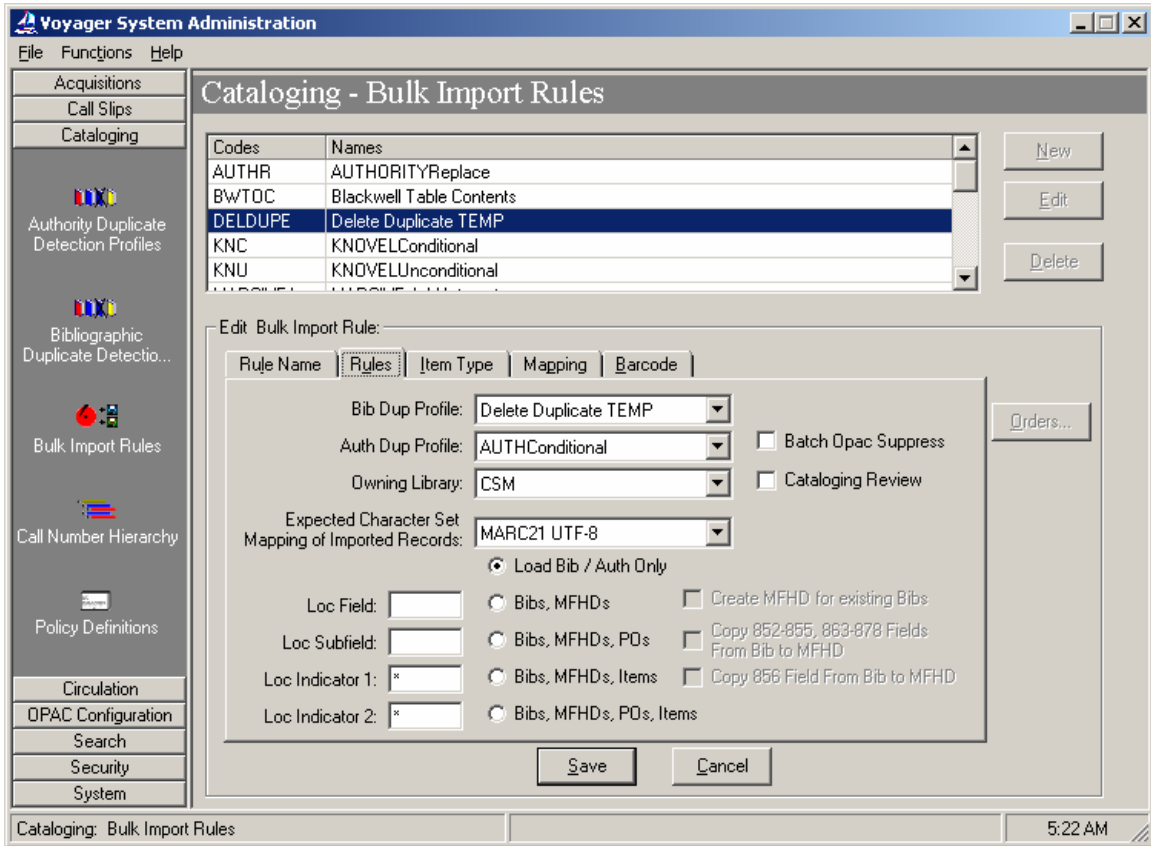
II. Bulk Import Rule

Bulk Import Rules allow you to create the rules that you want followed for importing bibliographic records. Duplicate detection is based on a Duplicate Detection Profile.

Note only the first two Tabs are configured:



Notice that Load Bib/Auth only is selected:



(Item Type, Mapping and Barcode Tabs are unchanged)

III. MARCEXP allows for the export of many MARC records at one time. A variety of criteria can be used to specify the records you want to export. We use the list of BIB Ids generated by our Access report.

Pmarcexport -rB -mM -t/m1/incoming/<filename>

-r = Record type
B = Extracts Bibs only
-m = Export mode
M = Processes files of Marc Ids (aka: BIB Ids)
-t = Input file
<filename> = input file name of Bib Ids
(Character set is left as the default: UTF-8)

Example of log file:

Record Type: BIB
Export Mode: MARC ID Input File
Export Target: /m1/voyager/csmdb/incoming/delete1.txt
Output File Name: /m1/voyager/csmdb/rpt/marc.exp.20050815.1133

Mon Aug 15 11:33:38 2005 EXPORTING...
Mon Aug 15 11:34:06 2005 ...COMPLETED

Records written to Output File: 5999

Note that all records are exported in the Unicode character set unless otherwise specified.

IV. PREBULK is designed to pre-process bibliographic records and create an output file that may be imported into a Voyager database. The manner in which the input file is processed is customized by creating a configuration file. We use it to create an interleaved file of bibliographic records and holdings records: using as input a file of bibliographic records and creating holdings records based on the configuration file.

Example of prebulk.cfg file:

```
[OVERRIDES]
CREATEMFHD=YES

[MFHDTAG]
XXX

[LOCATIONS]
zLINDAHALL    zLINDAHALL    050,090

[CALLTYPES]
050    0
090    0

[MAPPING]
1      852b
```

Pprebulk -i /m1/voyager/csmdb/incoming/delete1.mrc -o /m1/voyager/csmdb/incoming/delete1_out.mrc -c /m1/voyager/csmdb/local/prebulk1.cfg

-i = Input file
-o = Output file
-c = prebulk file

Example of log file:

```
Message: prebulk Version:| <V2.5>
Message: Prebulk Start Time:| <Mon Aug 15 11:40:57 2005>
Bibs:1000 Mfhd: 1000 Time:Mon Aug 15 11:41:15 2005
Bibs:2000 Mfhd: 2000 Time:Mon Aug 15 11:41:34 2005
Bibs:3000 Mfhd: 3000 Time:Mon Aug 15 11:41:52 2005
Bibs:4000 Mfhd: 4000 Time:Mon Aug 15 11:42:12 2005
Bibs:5000 Mfhd: 5000 Time:Mon Aug 15 11:42:31 2005
Bibs:5999 Mfhd: 5999 Time Stamp:Mon Aug 15 11:42:50 2005
Message: Prebulk Completed: | <Mon Aug 15 11:42:50 2005>
Message: Strip Statistics.| <>
```

V. BULKIMPORT imports, replaces or merges many bibliographic records at one time. It acts on a file of records based on the configuration of a bulk import rule and duplicate detection profile. When the proper variables are set *and* the input file is an interleaved file of BIB and holdings records, it can delete both record types.

Pbulkimport -f /m1/voyager/xxxdb/rpt/<filename> -iDELDUPE -b1 -e1000 -r -x

<filename> = Input file you created with Pmarcexport
-iDELDUPE = the bulkimport rule *Code* (case sensitive)
-b = Begin with record 1
-e = End with record 1000
-r = Delete Mfhds
-x = Delete Bibs

Example of log file:

I am 6547. I will be doing 1-1000 from
'/m1/voyager/csmdb/incoming/delete1_out.mrc' for you.
The import code is "DELDUPE" for this run.
The bib dup profile is "Delete Duplicate TEMP" for this run.
The auth dup profile is "AUTHConditional" for this run.
This import is using a rule that does not allow creation of MFHDs or Items.
Mon Aug 15 11:46:56 2005
Expecting Marc21 UTF-8 Records
1(1): Duplicate Bibs above threshold: replace 1, warning 0.
 BibID & rank
 379647 - 100
2(1): Duplicate Mfhds above threshold: replace 1, warning 0.
 MfhID & rank
 510474 - 100
 MFHD 510474 deleted.
 BIB 379647 deleted.
3(2): Duplicate Bibs above threshold: replace 1, warning 0.
 BibID & rank
 379648 - 100
4(2): Duplicate Mfhds above threshold: replace 1, warning 0.
 MfhID & rank
 510475 - 100
 MFHD 510475 deleted.
 BIB 379648 deleted.

<log snipped for brevity – continued on next page>

1997(999): Duplicate Bibs above threshold: replace 1, warning 0.

BibID & rank

403344 - 100

1998(999): Duplicate Mfhd above threshold: replace 1, warning 0.

MfhdID & rank

534386 - 100

MFHD 534386 deleted.

BIB 403344 deleted.

Recs processed: 1000

Mon Aug 15 12:04:07 2005

1999(1000): Duplicate Bibs above threshold: replace 1, warning 0.

BibID & rank

403345 - 100

2000(1000): Duplicate Mfhd above threshold: replace 1, warning 0.

MfhdID & rank

534387 - 100

MFHD 534387 deleted.

BIB 403345 deleted.

BIBLIOGRAPHIC or AUTHORITY Records

Processed: 1000

Added: 0

Discarded: 0

Rejected: 0

Errored: 0

Replaced: 0

Merged: 0

Deleted: 1000

Mfhd created: 0

Items created: 0

MFHD Records

Processed: 1000

Added: 0

Discarded: 0

Errored: 0

Replaced: 0

Deleted: 1000

Mon Aug 15 12:04:08 2005