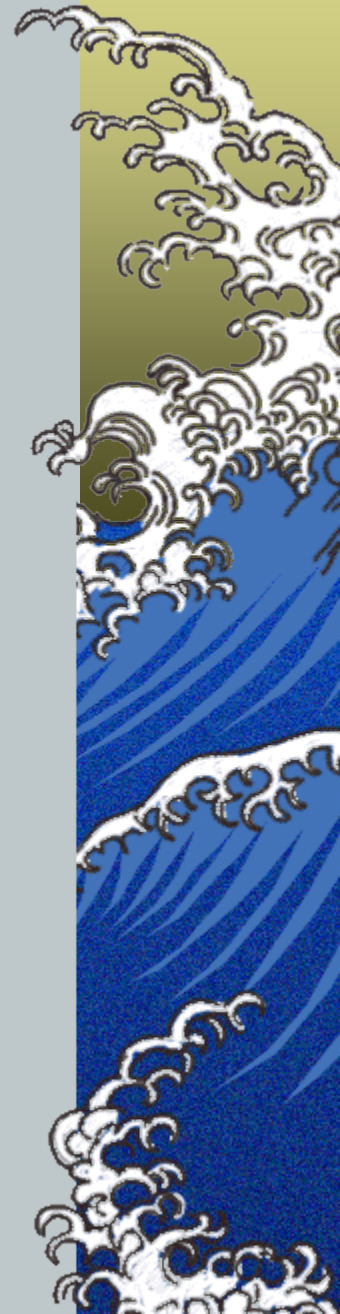


Adventures in Minimal MARC and Bulkimport; Or,

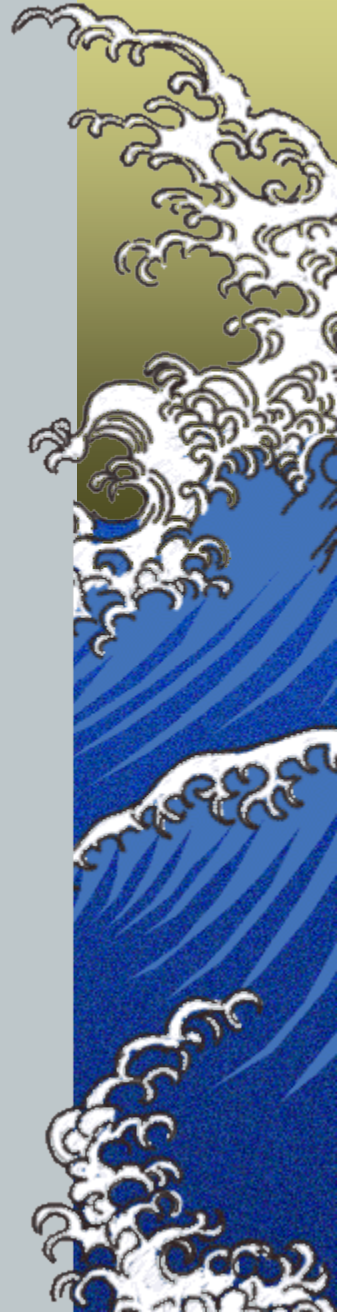
*How to Get OCLC Numbers or URLs or Other
Bits of Information into Lots of Bib Records with
Minimal Fuss*

*Maggie Rioux
Systems Librarian
MBLWHOI Library
Woods Hole, Mass.*



What we'll cover

- *What in the world am I talking about?*
- *Why would anyone want to do this?*
- *Basic principles and overview of process*
- *Serious look at how to do it*
- *Parting advice from one who's been there*



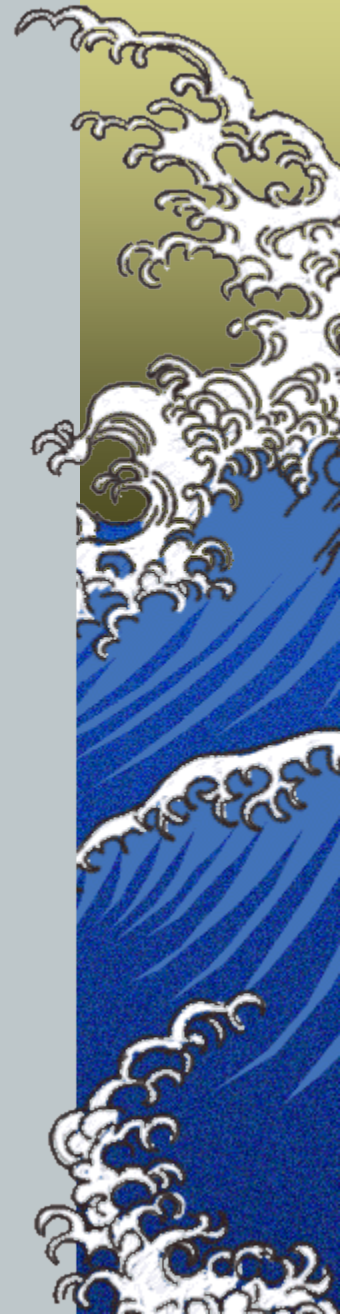
Basic Orientation

★ *You have:*

- 1. Plenty of perfectly good bib records in your database*
- 2. A long list of record-specific data that needs to be added, each piece to its proper record*

★ *You don't want to either:*

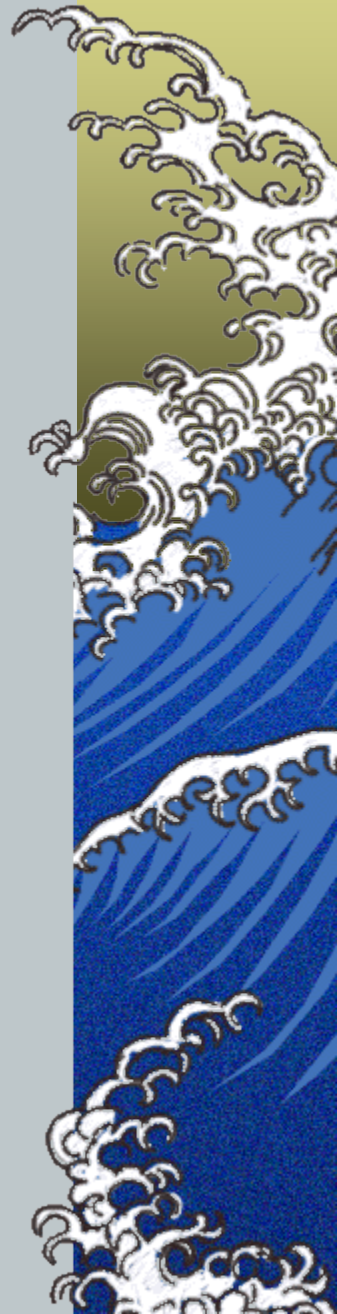
- 1. Destroy your database or,*
- 2. Individually edit half-a-bazillion records*



Basic Orientation

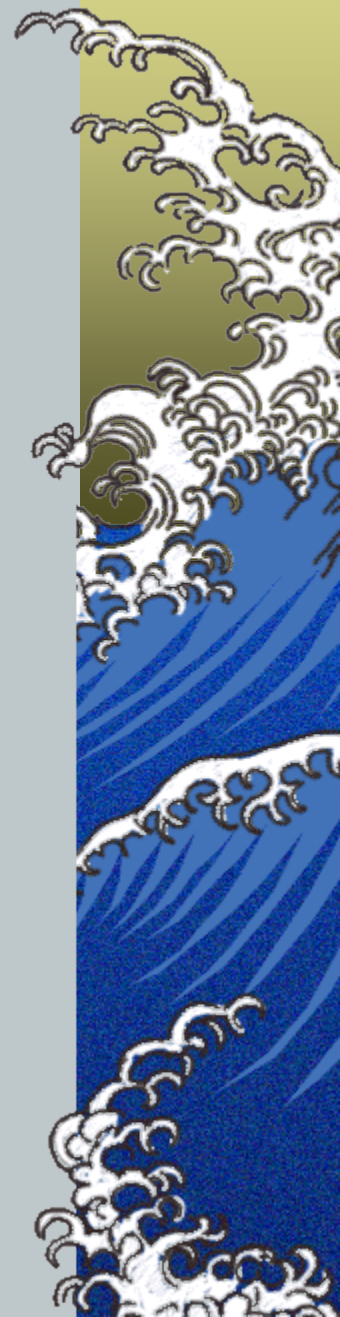
★ *You do want to:*

- 1. Take your list of data and, using batch procedures, insert the proper new field in each specific record*
- 2. Do this with as little blood, sweat, and (especially) tears as possible*



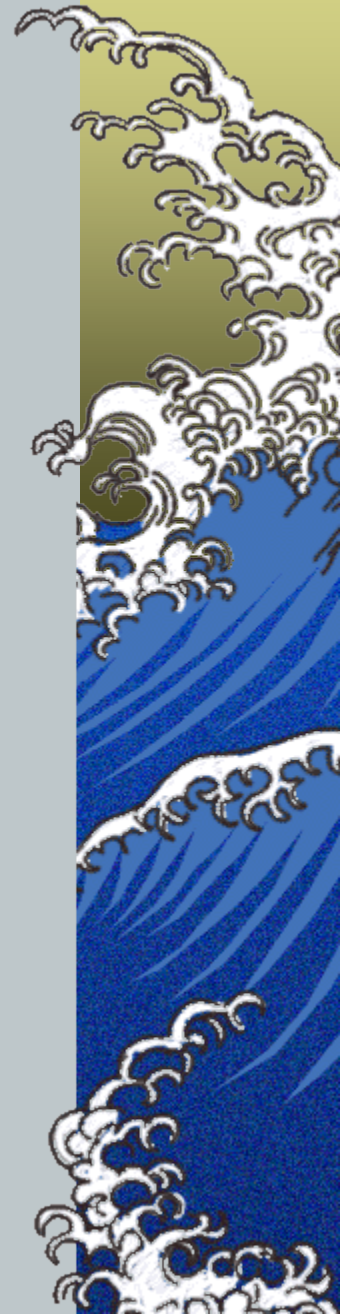
Why specifically would I use this?

- ▶ *OCLC record reload project*
 - ▶ *Need to load each returned OCLC number into an 035 field in the correct bib record*
- ▶ *Bulk scanning project*
 - ▶ *Need to load an 856 link to each scanned item (Google books or Internet Archive?) into the correct bib record*
- ▶ *Or any other data set that needs to be inserted into matching bib records*



Basic terms

- ▶ *Minimal MARC record*
- ▶ *Bulkimport*
- ▶ *MarcEdit*
- ▶ *OCLC Number = 035 field*
- ▶ *URL = 856 field*
- ▶ *Marc Leader*
 - ▶ *Encoding_Level = Elvl = Leader/17*
- ▶ *Bidirectional merge*

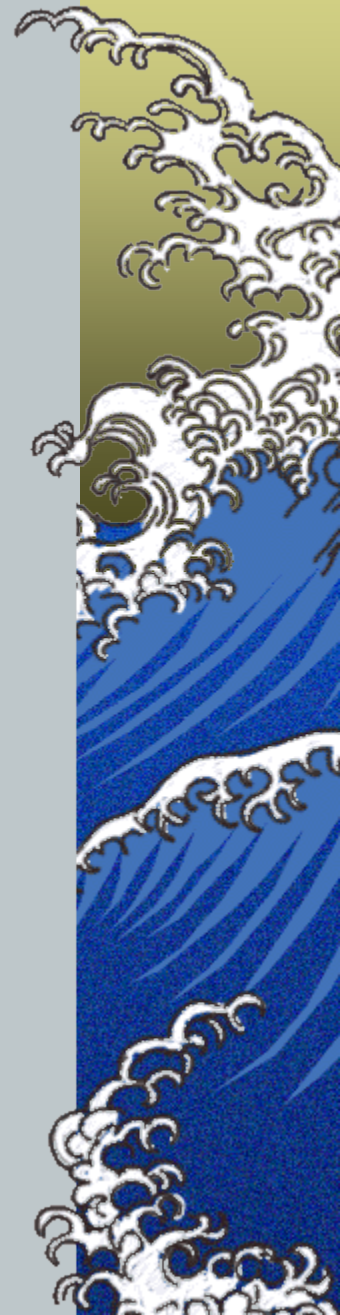


Step 1: Prepare your incoming data

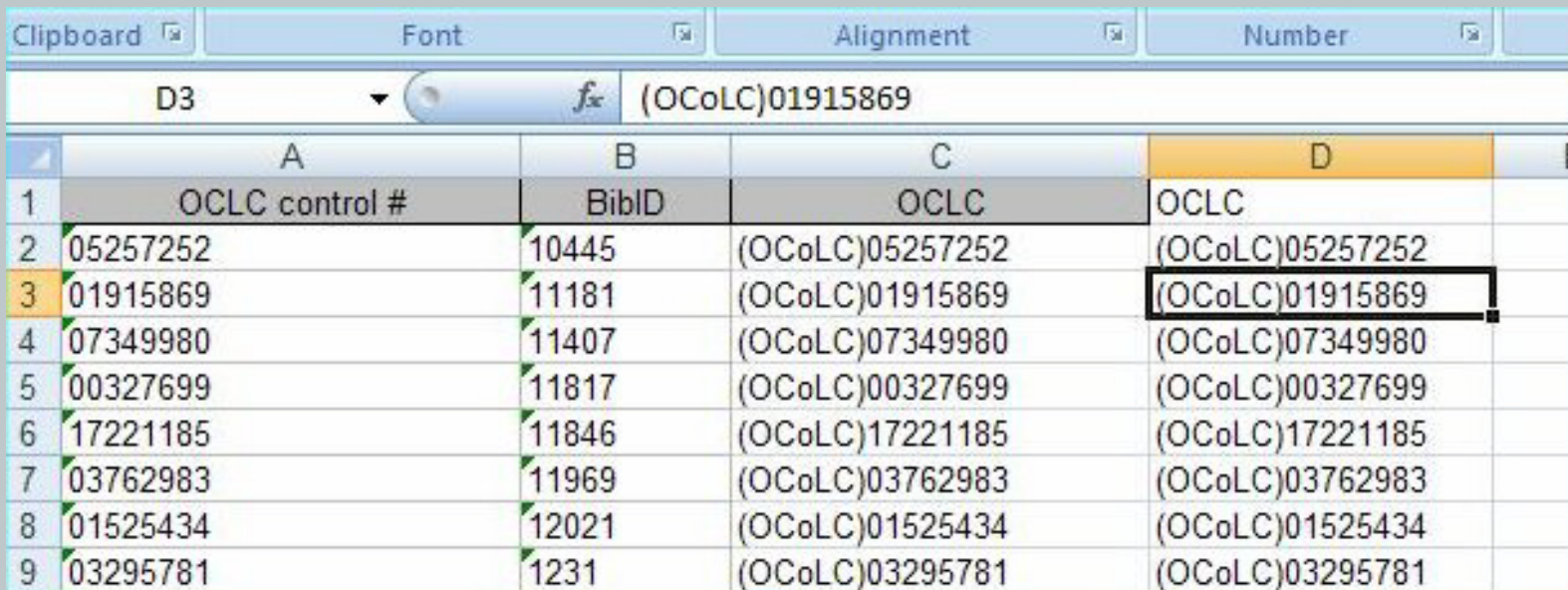
The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
1	OCLC control #	BibID	OCLC	
2	05257252	10445	= "(OCOLC)"&\$A2	
3	01915869	11181		
4	07349980	11407		
5	00327699	11817		
6	17221185	11846		
7	03762983	11969		
8	01525124	12024		

- Add the (OCoLC) to the front of all the OCLC numbers using this formula in a blank column. (Column A needs to be formatted as text in order to preserve the leading zeros.)

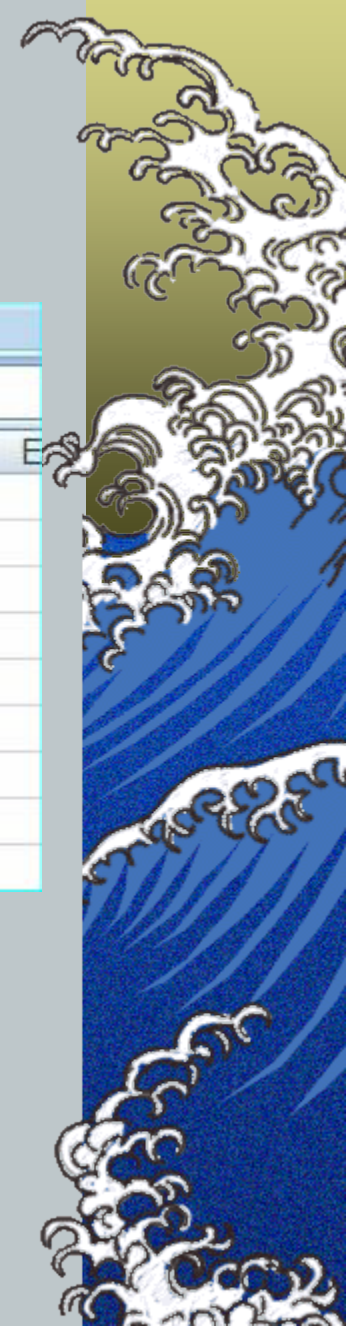


Step 1: Prepare your incoming data



	A	B	C	D
1	OCLC control #	BibID	OCLC	OCLC
2	05257252	10445	(OCoLC)05257252	(OCoLC)05257252
3	01915869	11181	(OCoLC)01915869	(OCoLC)01915869
4	07349980	11407	(OCoLC)07349980	(OCoLC)07349980
5	00327699	11817	(OCoLC)00327699	(OCoLC)00327699
6	17221185	11846	(OCoLC)17221185	(OCoLC)17221185
7	03762983	11969	(OCoLC)03762983	(OCoLC)03762983
8	01525434	12021	(OCoLC)01525434	(OCoLC)01525434
9	03295781	1231	(OCoLC)03295781	(OCoLC)03295781

- b. Propagate the formula down the list, then copy and paste special to get the values into a column.
- c. Lastly, tidy up and export the worksheet as tab-delimited text.



Step 2: Make minimal Marc records

Field 0	Field 1	Field 2	Field 3	Field 4	Field 5
10445	(OCoLC)05257...				

Select:
Field 0
Field 1

Map To:

Indicators:
w

Term. Punctuation:

Constant Data
 Repeatable Subfield

Arguments:
0 887\$a 0
1 035\$a 0

Save Template
 Sort Fields

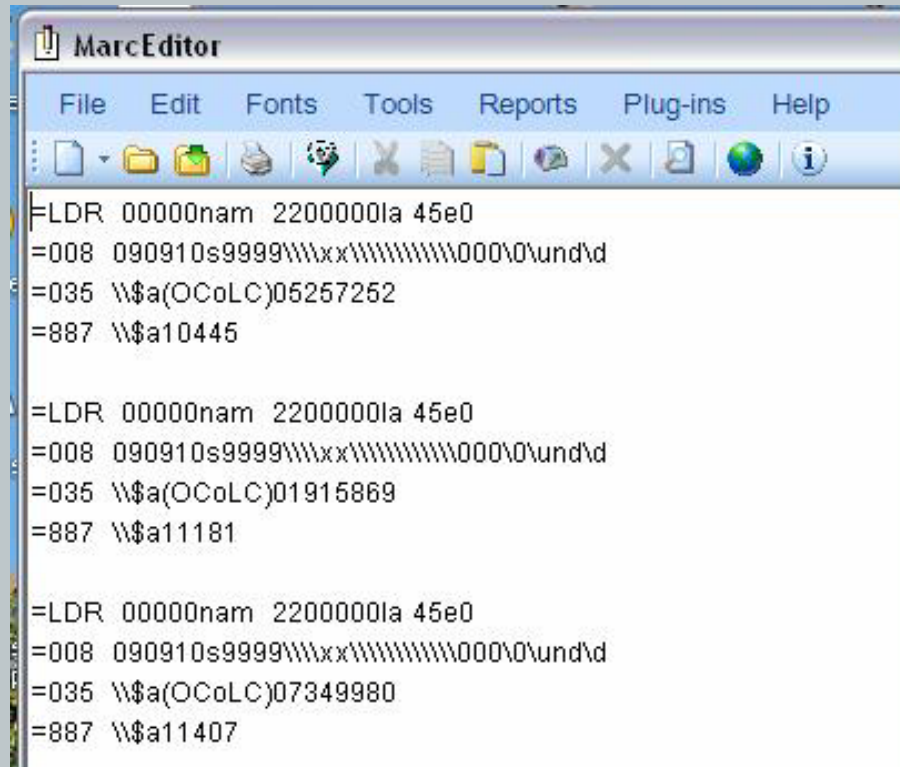
Apply Load Template

Finish Cancel

- Use MarcEdit Delimited Text Translation Wizard to create the records. You will want the BibID mapped to 887\$a (or another non-standard tag below 900) and the OCLC number mapped to 035\$a. Save your template to use on later batches of records
- The actual Marc records are then created in an “mrk” file.



Step 2: Make minimal Marc records



The screenshot shows the MarcEditor application window with a menu bar (File, Edit, Fonts, Tools, Reports, Plug-ins, Help) and a toolbar. The main text area contains three MARC records, each starting with an LDR line and followed by three data lines (008, 035, 887).

```
=LDR 00000nam 2200000la 45e0
=008 090910s9999\\xx\\0000\\und\\d
=035 \\$a(OCOLC)05257252
=887 \\$a10445

=LDR 00000nam 2200000la 45e0
=008 090910s9999\\xx\\0000\\und\\d
=035 \\$a(OCOLC)01915869
=887 \\$a11181

=LDR 00000nam 2200000la 45e0
=008 090910s9999\\xx\\0000\\und\\d
=035 \\$a(OCOLC)07349980
=887 \\$a11407
```

- c. Open the MarcEditor and then open the .mrk file
- d. In all records, change LDR/17 (Elvl) to “X” or another unused value so they will fail the quality check in bi-directional merge



Step 2: Make minimal Marc records

MarcEdit.NET Batch Editing Tools

Edit Subfield

Field: LDR Position: 17 Field Data:

Replace with: X

Replace Text
Remove Text
Close

Search Options:

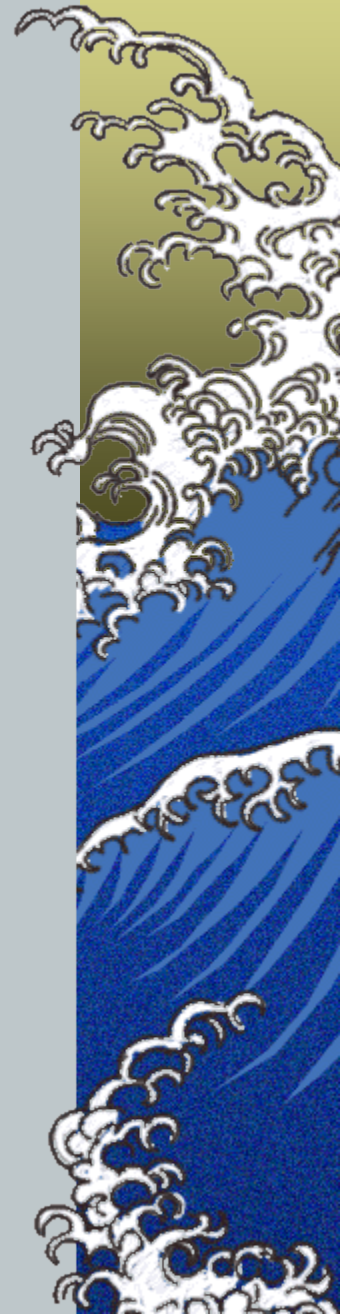
- New subfield only
- Add subfield if not present
- Match case
- Move subfield data
- Use regular expression

DB 00000000 22000000 1500

- e. Choose “Edit subfield data” on the tools menu or press F9
- f. Click “Replace Text” to complete the edit
- g. Compile the file into Marc, saving it as a .mrc file

Step 3: Use Bulk import with bi-directional merge to load the data

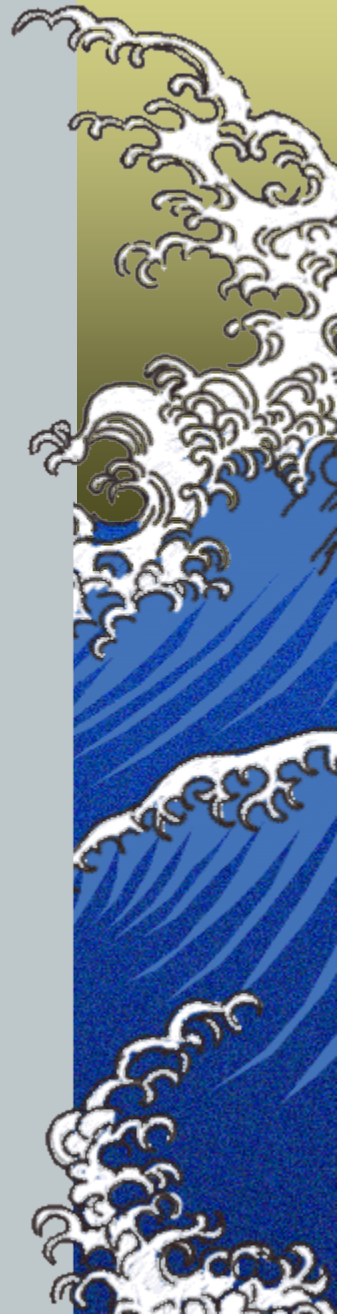
- a. Define a Bulk Import Rule and a Duplicate Detection Profile which will cause records to be matched on the incoming 887\$a mapped against the existing 001 and also to fail the quality check (more on this later)
- b. FTP the .mrc file to your ILS server (or use webadmin)
- c. Use bulk import to load the records, starting with a small test set



The Elephant in the room: Let's talk about bi-directional merge

▲ *How it works:*

- ▲ *Voyager compares the incoming record with what's already there*
- ▲ *If the incoming record is “better,” it overlays the current record, but some parts of the current record are saved*
- ▲ *If the current record is “better,” it stays and some parts of the incoming record get added*



The Elephant in the room: Let's talk about bi-directional merge

- ▶ *We want the incoming record to fail, so that the current one stays intact, but our new 035 field gets added*
- ▶ *This is the reason for the exercise with LDR/17 and X*
- ▶ *We need a Duplicate Detection Profile and a Bulk Import Rule*
 - ▶ *These can be a little hairy to define, but once you get it right you don't need to worry about it anymore*



Duplicate Detection Profile

Cataloging - Bibliographic Duplicate Detection Profiles

Codes	Names
ADDOCLC	Add OCLC Numbers to bibs
ALWYSADD	Unconditional Add
BHLURLS	Add BHL URLs to bibs
OCLCREPL	Check 035a (OCLC) only
REGREPL	Regular Replace

New Edit Delete

Edit Bibliographic Duplicate Detection Profile:

Profile | Field Definitions | Quality Hierarchy | Merge Fields

Profile Name: Add OCLC Numbers to bibs Profile Code: ADDOCLC

Duplicate Handling:

Add-Unconditional Replace Add-Conditional Merge Bi-Directional Merge

Discard incoming records that do not match existing records

Cancellation:

None OCLC RLIN

Duplicate Replace: 100 Duplicate Warn: 100

Save Cancel

Full details are in the handout, but here's a quick look



Duplicate Detection Profile

Cataloging - Bibliographic Duplicate Detection Profiles

Codes	Names
ADDOCLC	Add OCLC Numbers to bibs
ALWYSADD	Unconditional Add
BHLURLS	Add BHL URLs to bibs
OCLCREPL	Check 035a (OCLC) only
REGREPL	Regular Replace

Buttons: New, Edit, Delete

Edit Bibliographic Duplicate Detection Profile:

Profile | **Field Definitions** | Quality Hierarchy | Merge Fields

Code	Available Indexes	Code	Selected Indexes	Field	Subfield
008D	Publication Date(008)	BBID	Bibliographic Record ID	887	a
008L	Language(008 coded)				
008P	Publication Place(008 coded)				
010A	LCCN - valid (010a)				
0100	OCLC Number (up to 12/95) (010o)				
010Z	LCCN - Not valid (010z)				
019A	OCLC Number (019a):				

Field Override: 887 Sub Field Override: a Field Weight: 100

Buttons: Save, Cancel

Taskbar: c Duplicate Detection Profiles | 10:24 AM

Field Definition tab: BIBID code matches against the 887\$a



Duplicate Detection Profile

Cataloging - Bibliographic Duplicate Detection Profiles

Codes	Names
ADDOCLC	Add OCLC Numbers to bibs
ALWYSADD	Unconditional Add
BHLURLS	Add BHL URLs to bibs
OCLCREPL	Check 035a (OCLC) only
REGREPL	Regular Replace

Edit Bibliographic Duplicate Detection Profile:

Profile | Field Definitions | **Quality Hierarchy** | Merge Fields

Leader/06	Leader/17	Original Agency 040#a:	Last Modifying Agency 040#d:
*		*	*
*	1	*	*
*	2	*	*
*	3	*	*
*	4	*	*
*	5	*	*
*	7	*	*
*	8	*	*

Record Type Leader/06: * Original Agency 040#a: *

Encoding Level Leader/17: Last Modifying Agency 040#d: *

Save Cancel

Quality Hierarchy tab: Needs a line for every possible value of LDR/17 **except** X (so that your incoming records will fail)



Duplicate Detection Profile



Merge fields tab: Here's where you tell it to merge the 035\$a into the current record



Bulk Import Rule

Cataloging - Bulk Import Rules

Codes	Names
ADDOCLC	Add OCLC Numbers to bibs
ALWYSADD	Unconditional Add
BHLURLS	Add URLs to BHL records
FLESCHER	Flescher Import
OCLCREPL	Check 035a (OCLC) only

New Edit Delete

Edit Bulk Import Rule:

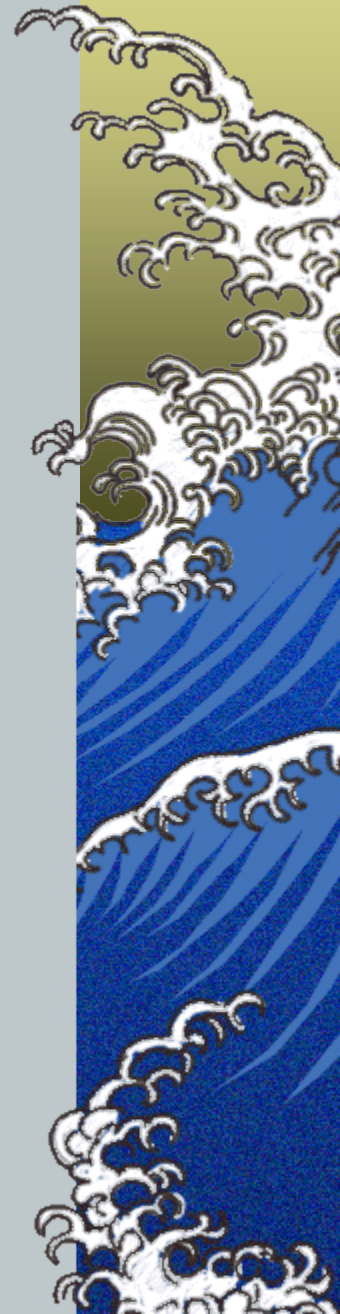
Rule Name Rules Item Type Mapping Barcode

Code: ADDOCLC Orders...

Name: Add OCLC Numbers to bibs

Save Cancel

The Bulk Import Rule is a parameter in the Bulk Import command line and it calls the Duplicate Detection Profile.



Bulk Import Rule

Cataloging - Bulk Import Rules

Codes	Names
ADDOCLC	Add OCLC Numbers to bibs
ALWYSADD	Unconditional Add
BHLURLS	Add URLs to BHL records
FLESCHER	Flescher Import
OCLCREPL	Check 035a (OCLC) only

Edit Bulk Import Rule:

Rule Name: **Rules** | Item Type | Mapping | Barcode

Bib Dup Profile: Add OCLC Numbers to bibs
Auth Dup Profile: LC Check
Owning Library: MBLWHOI Library
Expected Character Set: Latin-1 (non-Unicode)

Mapping of Imported Records:

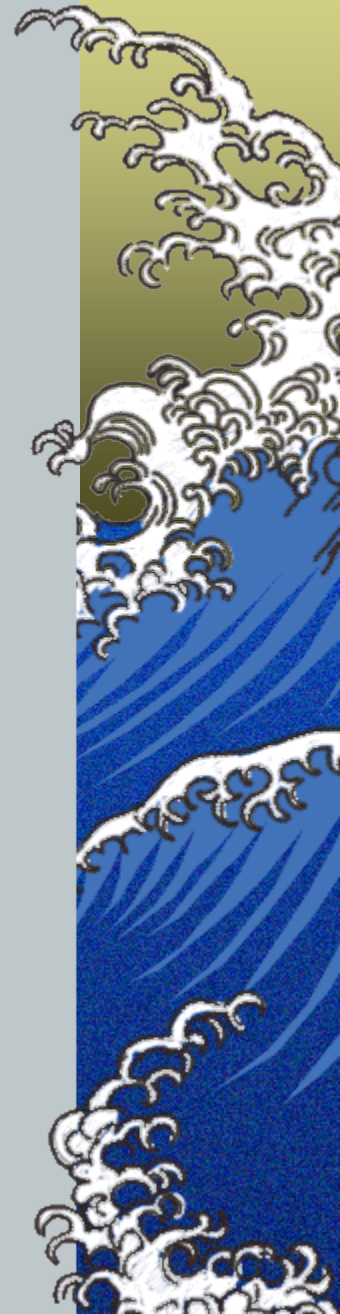
Load Bib / Auth Only
 Bibs, MFHDs
 Bibs, MFHDs, POs
 Bibs, MFHDs, Items
 Bibs, MFHDs, POs, Items

Leave OPAC Suppress Unchanged for Replaced and Merged Records
 Batch Opac Suppress
 Cataloging Review

Loc Field:
Loc Subfield:
Loc Indicator 1: *
Loc Indicator 2: *

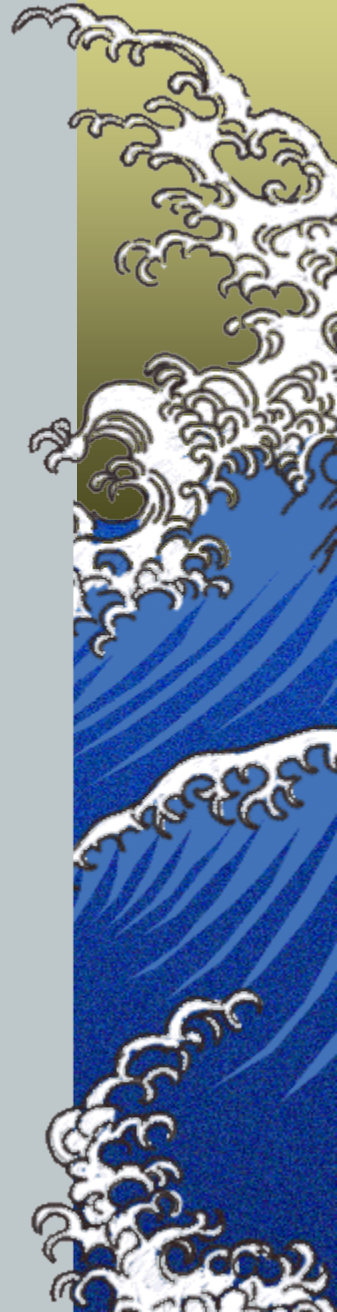
Buttons: New, Edit, Delete, Orders..., Save, Cancel

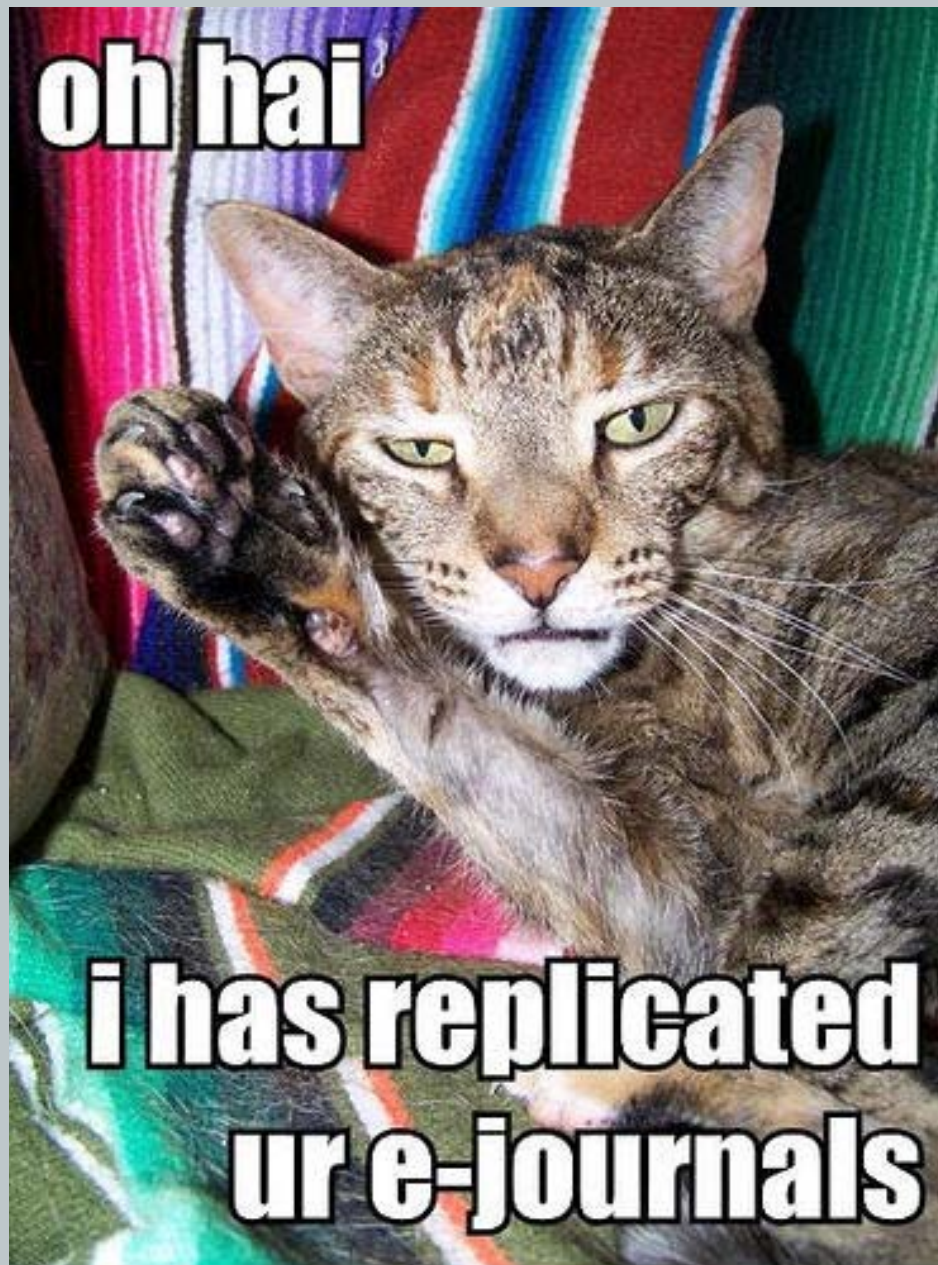
The Rules tab is the most important one to get right.



Maggie's Guidelines for a Successful Project

- ▶ *Read everything through several times*
 - ▶ *These documents and also the Voyager Technical User Guide*
- ▶ *Work with a small batch at first and don't do anything you can't **undo** when it doesn't work right the first (or second or third) time*
- ▶ *Read the documents and manuals while lying down – more blood flow to the brain, no mental effort wasted on sitting up*





Most important: Don't let any cats help you in any way!

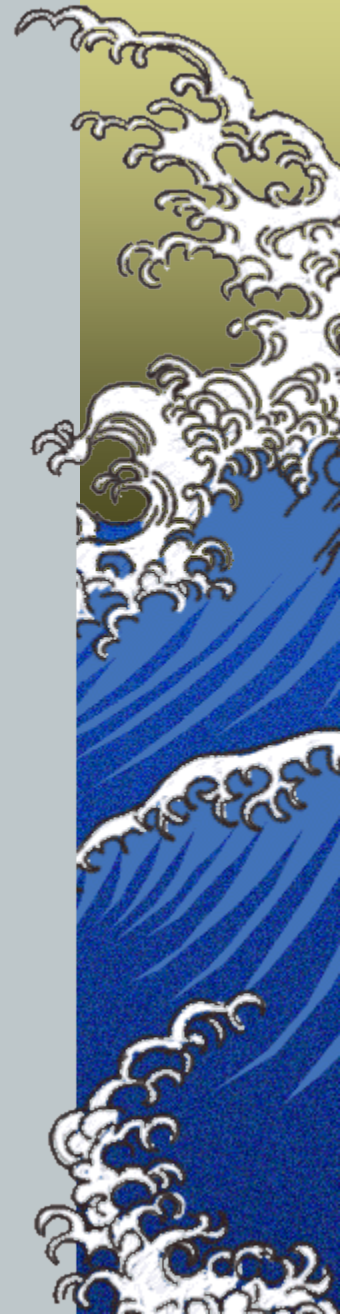
And if you want to get really
carried away with using MarcEdit:

<http://hdl.handle.net/10057/1935>

*Optimized Metadata Repurposing in a Library
Using MarcEdit*

-

Presentation at ALA Midwinter 2009 by Sai
Deng, Wichita State University

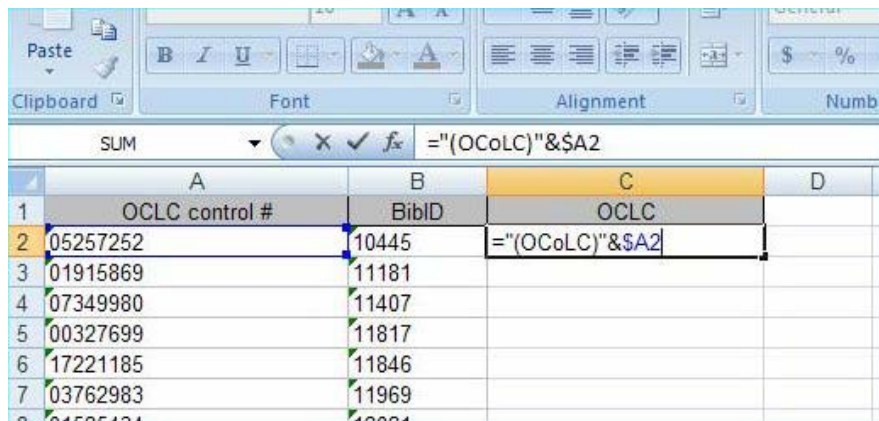


Getting Your Returned OCLC numbers into Your Voyager Records

Once you have sent your Marc records to OCLC for upload, you will get back a spreadsheet containing two columns: OCLC number and BIBid (which will be labeled “submitted 001 field”). Now you need to get the OCLC number into the proper bib record in your catalog as an 035\$a. This document will guide you through the process. There are two main steps: first, creating minimal Marc records from the returned spreadsheet and second, loading those records into the online catalog so that they merge with the records already there, rather than replacing them.

Part I: Creating Minimal Marc Records from the Spreadsheet

1. The XREF spreadsheet returned from OCLC contains 2 columns: “OCLC control #” and “Submitted 001 Field.” Relabel the 2nd one as BibID for clarity. Add another blank column labeled OCLC. The incoming OCLC column should be formatted as text to preserve any leading zeros.
2. In the OCLC column for the first row, enter the formula =“(OCoLC)”&\$A2 where A2 is the cell in that row containing the OCLC control #. Use of the \$ in the formula locks the cell.



	A	B	C	D
1	OCLC control #	BibID	OCLC	
2	05257252	10445	=“(OCoLC)”&\$A2	
3	01915869	11181		
4	07349980	11407		
5	00327699	11817		
6	17221185	11846		
7	03762983	11969		
8	01525424	12034		

3. Once this shows the right value, drag the lower right corner of the cell down the column to copy the formula to all rows.

	A	B	C	D
1	OCLC control #	BibID	OCLC	
2	05257252	10445	(OCoLC)05257252	
3	01915869	11181	(OCoLC)01915869	
4	07349980	11407	(OCoLC)07349980	
5	00327699	11817	(OCoLC)00327699	
6	17221185	11846	(OCoLC)17221185	
7	03762983	11969	(OCoLC)03762983	
8	01525434	12021	(OCoLC)01525434	
9	03295781	1231	(OCoLC)03295781	
10	03146266	12410	(OCoLC)03146266	
11	10588979	12508	(OCoLC)10588979	
12	01098387	12665	(OCoLC)01098387	
13	00849986	12714	(OCoLC)00849986	
14	07549247	13036	(OCoLC)07549247	
15	00522080	1316	(OCoLC)00522080	

- Copy and paste special the column values to another column and delete the one with the formulas and the original "OCLC control #" column.

	C	D	E	F	G	H	I	J
	OCLC							
	(OCoLC)05257252							
	(OCoLC)01915869							
	(OCoLC)07349980							
	(OCoLC)00327699							
	(OCoLC)17221185							
	(OCoLC)03762983							
	(OCoLC)01525434							
	(OCoLC)03295781							
	(OCoLC)03146266							
	(OCoLC)10588979							
	(OCoLC)01098387							
	(OCoLC)00849986							
	(OCoLC)07549247							
	(OCoLC)00522080							
	(OCoLC)00718012							
	(OCoLC)15867022							
	(OCoLC)02976187							
	(OCoLC)02370410							
	(OCoLC)01030543							

Paste Special

Paste

All
 Formulas
 Values
 Formats
 Comments
 Validation

All using Source theme
 All except borders
 Column widths
 Formulas and number formats
 Values and number formats

Operation

None
 Add
 Subtract

Multiply
 Divide

Skip blanks
 Transpose

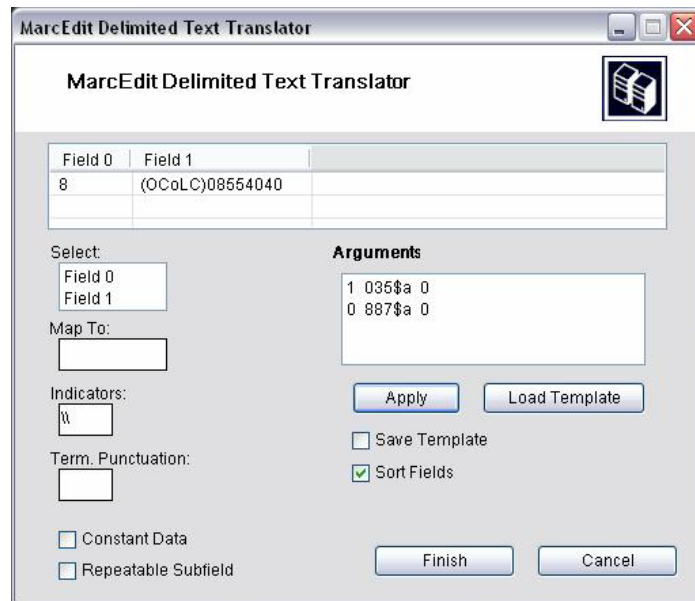
Paste Link OK Cancel

	A	B	C	D	E
1	OCLC control #	BibID	OCLC	OCLC	
2	05257252	10445	(OCoLC)05257252	(OCoLC)05257252	
3	01915869	11181	(OCoLC)01915869	(OCoLC)01915869	
4	07349980	11407	(OCoLC)07349980	(OCoLC)07349980	
5	00327699	11817	(OCoLC)00327699	(OCoLC)00327699	
6	17221185	11846	(OCoLC)17221185	(OCoLC)17221185	
7	03762983	11969	(OCoLC)03762983	(OCoLC)03762983	
8	01525434	12021	(OCoLC)01525434	(OCoLC)01525434	
9	03295781	1231	(OCoLC)03295781	(OCoLC)03295781	

- You should now have 2 columns labeled BibID and OCLC.
- Export the worksheet as tab-delimited text. In order to test the process and also to break up the batch load into manageable pieces, add a new worksheet to the

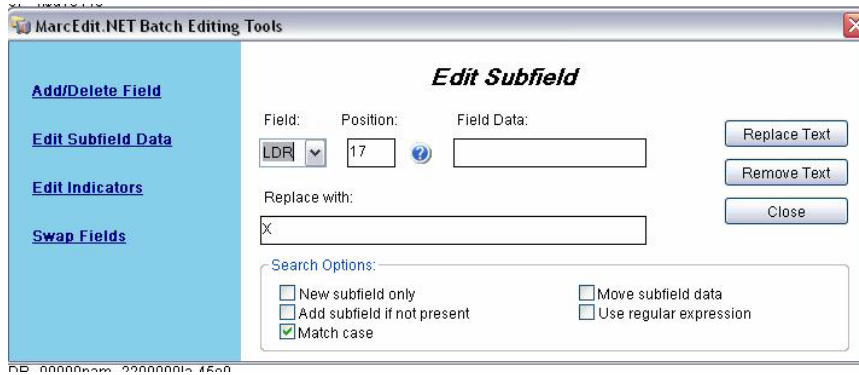
spreadsheet, then copy (or move) a chunk of the data to the new worksheet. Do about 10 rows for the first test chunk and then larger numbers (up to a couple thousand) for succeeding chunks. Export the new worksheet and use the resulting text file for input in the next step. Try to exclude the column labels from the export. If you include them it will make setup easier, but you will get an extra minimal Marc record which you will need to delete.

7. Use the MarcEdit Delimited Text Translator to convert the text file into minimal Marc records containing only an 035 and an 887 field (or use any other Marc tag that you don't usually find in your records and which is below 900). The OCLC field (field 0) should map to the 035 and the BibID field (field 1) to the other tagged field. Leave indicators and terminal punctuation as is. If you save the specifications as a template, you can use them again without re-definition. Your specification screen should look like this, but with different values in the "Field 0" and "Field 1" cells:



Now use MarcEdit's main program to change all LDR/17 (Elv1) to "X" (so that they will fail the quality check in bulkimport),¹ save the .mrk file and then compile the result into Marc (.mrc file). To do this choose "Edit subfield data" on the Tools menu or press F9. Click "Replace Text" to complete the edit.

¹ You need to change the value of this byte to something which is NOT already in your database. To see what values already exist use the following SQL in an Access query: SELECT BIB_TEXT.ENCODING_LEVEL, Count(BIB_TEXT.BIB_ID) AS CountOfBIB_ID FROM BIB_TEXT GROUP BY BIB_TEXT.ENCODING_LEVEL ORDER BY BIB_TEXT.ENCODING_LEVEL; I have used "X" since it is not a legal value and should not be used in your database.



8. Use FTP to transfer the .mrc file to your ILS server (or use webadmin for the bulk import). Use bulk import, utilizing the bi-directional merge feature to import/add the data in the minimal Marc records to your database.²
9. You will need to have both a bulk import rule and a duplicate detection profile defined which cause records to be matched on the incoming 887a mapped against the existing 001 (BIBID) and also to fail the quality check. See Appendix A for parameters on setting this up.
10. I recommend testing the process using the following procedure:
 - a. Copy about 4-10 records to a separate worksheet, export and process them as outlined above.
 - b. Before running the bulk import step, go into cataloging, bring up each of the test records in turn and save them to disk as a backup.
 - c. Run the bulk import then check what happened both in the import log (in /m1/voyager/xxxxdb/rpt) and also by looking at the records again in cataloging.
 - d. If something went wrong, you can “import” your locally-saved records and re-overlay them onto the mistakes. If all is well, delete your local save file.
11. The final result will have your 035\$a with the OCLC number. There may possibly be an extra 035\$a containing the Voyager BibID (I’ve found it only occasionally and am not sure what causes it). If so, this extra field should not present any problems, but as catalogers work with records for other purposes, they can be instructed to delete the extraneous field.

² To do this from the command line, log into your server as user voyager and cd to /m1/voyager/xxxxdb/sbin. Enter the following command: ./Pbulkimport -f<full path to your input file> -o<operator name> -i<bulk import rule code> An example is: ./Pbulkimport -f/export/home/voyager/incoming_data.mrc -omrioux -iADDOCLC

Appendix A: Bulk Import Rule and Duplicate Detection Profile for Bi-Directional Merge

In the Voyager System Administration module, you need to define a Bibliographic Duplicate Detection Profile for Bi-Directional Merge which will cause the incoming minimal Marc records to fail the quality check, thus preserving the record already in the database, but adding the incoming fields. The record matching should occur on BIBID in the existing record matched to 887\$a in the incoming minimal record. You also need a Bulk Import Rule, which is invoked in the bulk import command line and calls your Duplicate Detection Profile. Screenshots follow:

Part I: Duplicate Detection Profile

Basic Profile Screen

Codes	Names
ADDCLC	Add OCLC Numbers to bibs
ALWYSADD	Unconditional Add
BHLURLS	Add BHL URLs to bibs
OCLCREPL	Check 035a (OCLC) only
REGREPL	Regular Replace

Profile Name: Profile Code:

Duplicate Handling:
 Add-Unconditional Replace Add-Conditional Merge Bi-Directional Merge

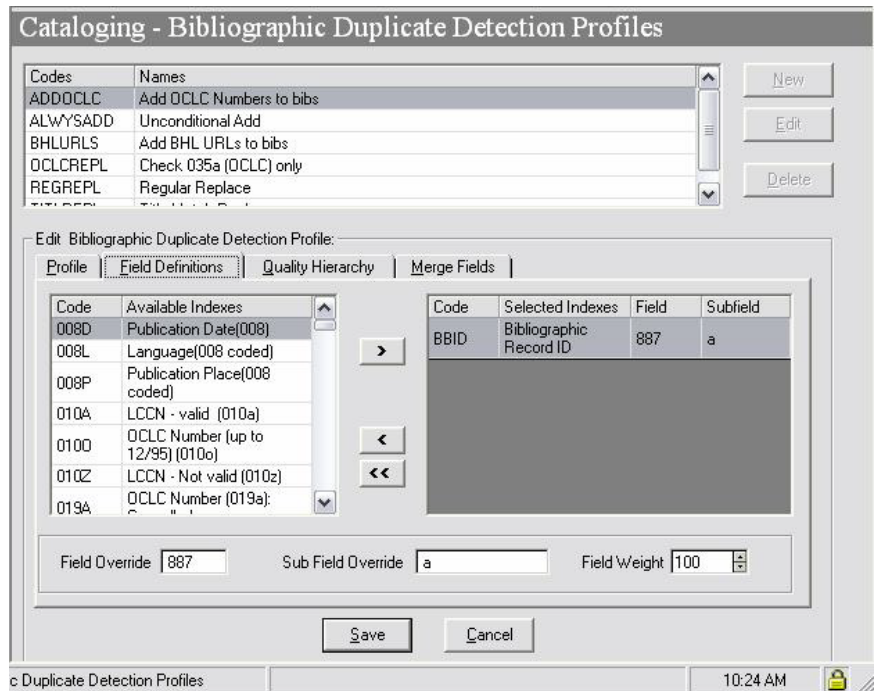
Discard incoming records that do not match existing records

Cancellation:
 None OCLC RLIN

Duplicate Replace: Duplicate Warn:

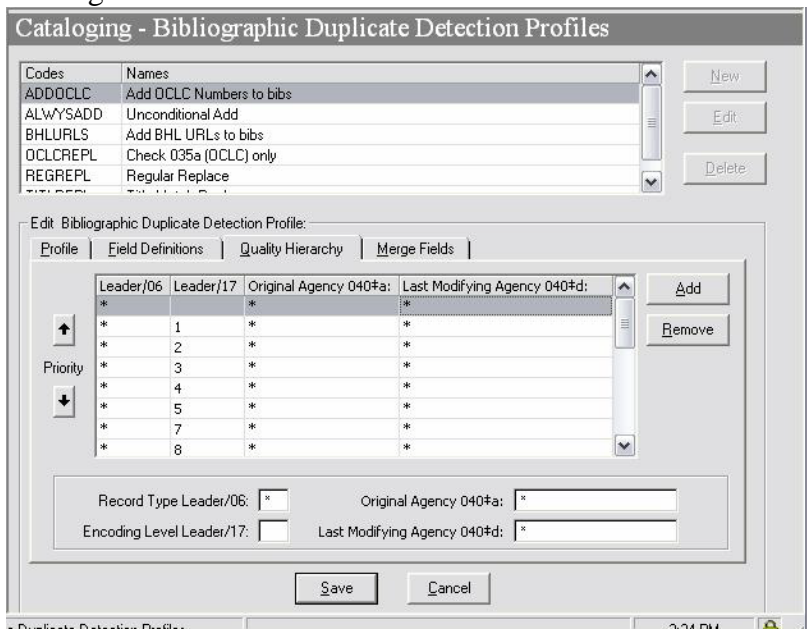
Field Definition Tab

Select BIBID from the Available Indexes and fill in the Field/Subfield Override to match to the 887\$a in the incoming record.



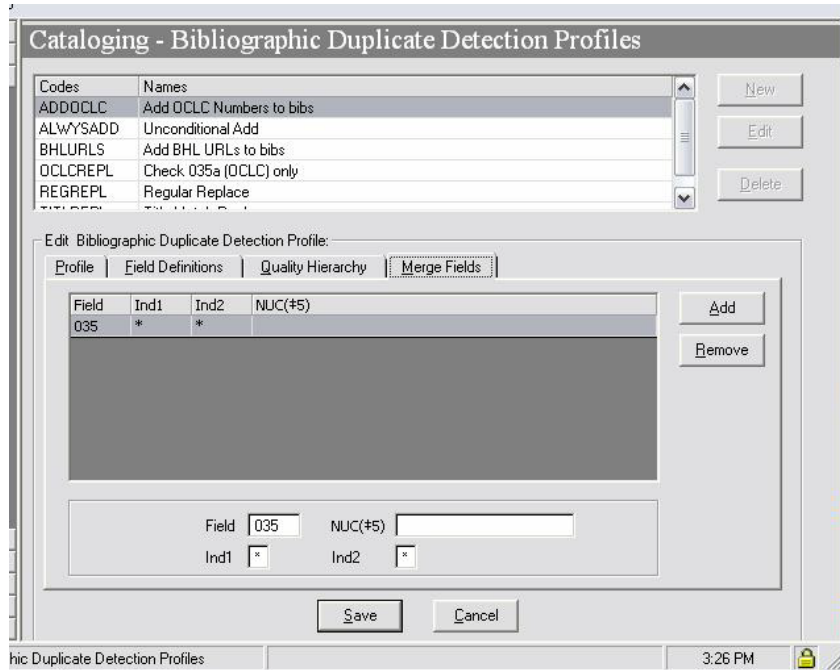
Quality Hierarchy Tab

Enter a line for every value of Leader/17 in your bibliographic database. Make sure you enter a space character to represent blank values – see line one of the grid below – don't just delete the *. Also, make sure to add a line for the “_” character as this can look like a blank character in displays and queries. Having done this, changing Leader/17 to X (which is not a valid value for this byte and which should not appear in your database) in your incoming minimal Marc records will cause them to fail quality check and not overlay your existing records.



Merge Fields Tab

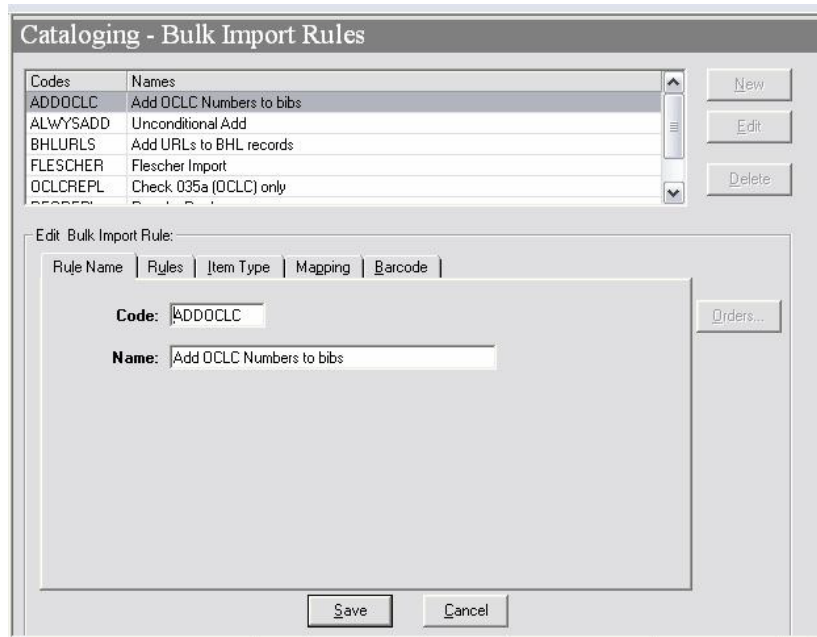
These are the fields in the incoming (minimal) records which will be added to your existing records. You want to enter only 035 so that only the OCLC number will be added.



Part II: Bulk Import Rule

Only the first two tabs (Rule Name and Rules) are relevant.

Basic Rule Name Screen



Rules Tab

This is the most important screen to get right. Auth Dup Profile is irrelevant. The Bib Dup Profile is the one you created in Part I. Make sure that “Load Bib/Auth Only” is checked in the lower half of the screen (it should be the default). Also, “Leave OPAC Suppress Unchanged...” should be checked.

