FDM Tutorial – Using MarcEdit to delete MARC records from a record set based on the 'date of publication'

* Note, these instructions are for MarcEdit 7.6.

Background

GPO catalogs, and updates or corrects, catalog records for older publications. These records for older publications will appear in monthly record sets. Some libraries do not wish to catalog older resources.

The MARC record 008 "date" field, character positions 07-10, contain a four-digit year, or a combination of digits and 'u's when the date or start date is unknown. Using regular expressions (regex), MarcEdit can identify dates that are outside of your library's interest. This enables you to delete the unwanted records from your record set.

Process

- Download a FDM .mrc record set.
- Double click on the .mrc file to open it in MarcEdit.
- When the MARC dialog box opens, with the "Select Operation" set to "MarcBreaker," (1), click on **Execute** (2), and then **Edit Records** (3).

🙀 MARC Tools				_		×				
2										
🔦 Tools			What wou	uld you l	ike to	do -				
Select Operation:										
MarcBreaker 🚹			~ 2	Exe	cute					
Select Data to Process:			0	Edit Re	ecords					
Open				Eureria	corus					
ernment Printing Office (G	PO)\Desktop\crdp_nov2024_85	5_utf8.mrc 🛅		Clo	ose					
Save As										
C:\Users\adahlen\OneDrive - Government Printing Office (GPO)\De: 🛅										
Character Encoding Option	-									
Default Character Encoding	: 🗹 Translate	to MARC-8								
MARC8 ~	🗌 Translate	to UTF8								
Results:										

This converts your MARC records into a mnemonic format that MarcEdit can manipulate. The file extension will be .mrk.

MarcEditor: crdp_nov2024_855_utf8.mrk											
File	Edi	it 11ts	Reports	Tools	OCLO	w	orldCat	Plug-ins	Automation	Help	
		Undo			F2						
		Special Undo Cut		Ctrl+Alt+F2 Ctrl+X							
=LDR 0											
=001 0 =003 C		Сору		(Ctrl+C						
=005 2		Paste		(Ctrl+V						
=008 7		Delete Rec	ord(s)		2	•	By	Range	Ctrl+D	el	
=020 \\ =035 9		Duplicate	Lines	(Ctrl+D		Wi	th Invalid U	JTF-8 Characters	5	
=035 \\		Clear			Del		De	lete Selecte			
=040 \\		Find		(Ctrl+F				5		
=110 1		Replace		(Ctrl+R		ittps://io	d.loc.gov/au	uthorities/name	s/n79022809	
=245 1		Edit Short	cuts			•	\$cfrecommended by National Park Service.				

With the .mrk file open, click on Edit (1) > Delete Record(s) (2) > Delete Selected Records (3).

- 1) In the dialog box's Display Field, type 008
- 2) Click Import File.
- 3) In the Search box, type your regular expression to identify records outside of your date range. Regular expressions are described in the Sample Regular Expressions section below.
- 4) Check the box Use Regular Expressions.
- 5) Click the **Search** icon next to the Search box.



It will tell you how many records it selected. Click OK.

Scroll down the records to verify that the selected records are outside of your library's date criteria.

If the records selected are the records you wish to *retain*, then click **Invert Selections** so the records selected are the ones you wish to delete.

When selected records are the records you want to delete from the record set, click **Delete Selected**.



The new record set is in a .mrk file. You will need to covert the file to .mrc.



With the .mrk file open, click on File (1) > Compile File into MARC (2).

It will prompt you to name the file and specify a location for it. Click **Save** when you are done.

Sample Regular Expressions

You will need to use regular expressions to select records by the 4 digits in the MARC 008 field. Your date parameters will differ, depending on your preferences for the age of the material cataloged. Here are some examples. Remember that the regular expression will select the records that you do not wish to retain – leaving the unselected records as the ones you wish to use.

To retain records with a publication year of 1950 or 195u to present:

This will select anything beginning with 190# through 194#, or 19uu, or 18uu, or 18##, or any uuuu:

^.{7}(19[0-4u].|1[0-8u]..|uuuu)

To retain records with a publication year of 1989 to present:

This will select anything beginning with 190# through 1988, or any 19uu, or any 18## or any 18uu, or any uuuu:

^.{7}(19[0-8u][0-8u]|18[0-9u][0-9u]|uuuu)

To retain records with a publication date of 2000 or 20uu to present:

This will select anything beginning with 19##, 19uu, 18##, 18uu, or any uuuu: ^.{7}(19..|18..|uuuu)

To retain records with a publication date of 2020 to present or 202u:

This will select anything beginning with 201#, 200#, 1###, or, or any 20uu: ^.{7}(20[0-1u].|1...|20uu)

Help with Regular Expressions:

If you would like to modify the date parameters, consult the following for information on regular expressions.

- <u>Regular Expression Language Quick Reference .NET | Microsoft Learn</u>
 - There are different types of regular expression syntaxes. This is a cheat sheet to the .NET regular expression syntax, which MarcEdit uses.
- regex101: build, test, and debug regex
 - Get a breakdown of how your regex works by plugging it into the Regular Expression line and looking at the Explanation box.
 - Test your regex by copying over sample 008 fields in the Test String box and seeing if it is 'matched' in the Match Information box as you expected.

*Note: The mention of these above products or online services does not constitute an endorsement or recommendation by GPO or the FDLP. We are not responsible for the quality, safety of these sites and encourage you to conduct your own research by contacting your Libraries IT Department.

Date: 12/31/2024