Working with the Discard File

following Collection Manager Bulk Loads

John Whisler

Eastern Illinois University

In my recent webinar, I suggested a way to work with the discard file, and suggested there may be better ways, and if such were discovered, it would be good to share them.  I’ve discovered a way to extract the 001 and 019 fields from the full MARC records that appear in the Discard file, which should save a considerable amount of time over the manual procedure shown in the webinar. The steps below are based on Microsoft Office Professional Plus 2010 versions of Word and Excel and Windows 7.

Download the Discard file from the CARLI ftp site.  It will be saved as a plain text file.

Open the file, select all, and copy.

Paste the entire file into a new Word document.

Using the Replace function, change ^p [followed by 10 spaces] to nothing. Replace All.

This will remove line breaks in long 019 fields so you don’t miss any data.

Using the Replace function, change $a to [2 spaces]. Click Replace All.

In the Word document, Select All and copy the entire contents.

Go back to your original plain text file, delete all its contents, and paste in the newly-formatted version you copied from Word. Stretch the plain text file horizontally as wide as you can. Save.

Open the plain text file using Excel and the Fixed Width option.

Click Next.

This should automatically set cell widths appropriately. There should be two dividers.

The first should be between tags and indicators, the second between indicators

and variable fields. Move the dividers to these positions if necessary.

Click Finish.

You should have a spreadsheet of the Discard file with three columns:

  Column A = Tag

  Column B = Indicators

  Column C = Variable field data

Highlight Column C and change its format from General to Text. This will make numbers appear left-aligned instead of right-aligned. Stretch Column C to a width that will fit nicely on a printed page. Click Wrap Text. This will allow all numbers in the 019 field to display in the cell.

Highlight column A.

Open Sort & Filter, then click Filter.

Open the little data selection dropdown box that appears at the top of Column A.

Click to clear the Select All box.

Select values 1 (as in 001 field) and 19 (as in 019 field), then OK.

Highlight and delete Column B. You don’t need its data for anything, and the display will be neater.

Voila—a nice list of  001 fields followed by 019 fields appears, ready to print and work with.