

Importing from Tab-Delimited Files

Tab-delimited text files are an easy way to import metadata for multiple files. (For more general information about using and troubleshooting tab-delimited files, see [Using Tab-Delimited Text Files](#).)

This tutorial leads you through the process of creating a tab-delimited file from a MARC file, configuring and cleaning the data for import, and then adding your metadata to CONTENTdm®.

To successfully complete this tutorial:

- The CONTENTdm server must be running
- The Acquisition Station must be installed on your desktop computer
- Microsoft® Excel® must be installed
- Optionally, MarcEdit 5.0 can be installed (MarcEdit, written by Terry Reese of Oregon State University, is sponsored by the Library of Congress and is used free of charge by thousands of libraries to analyze and manipulate MARC data files. To download, see <http://oregonstate.edu/~reese/marcedit/html/>.)

Learn About

- Section I** **Creating a Tab-Delimited File from a MARC File**
 - Exporting a Tab-Delimited File Using MarcEdit 5.0**
- Section II** **Using Microsoft Excel to Prepare File for Import**
- Section III** **Importing Dublin Core Metadata Only from MARC File Output**
- Appendix A** **Using MarcEdit 5.0 to Evaluate MARC Fields**
- Appendix B** **Cleaning Data in Microsoft Excel**

Section I Creating a Tab-Delimited File from a MARC File

In this section you will learn how to export a MARC file from your ILS, and then how to prepare the file for import to CONTENTdm.

1. Create a crosswalk based on the Library of Congress MARC to Dublin Core (DC) crosswalk (<http://www.loc.gov/marc/marc2dc.html>), adjusting the map to fit your data. In CONTENTdm, only one field, DC.Title, must be mapped to a Dublin Core element.

Following is an example of an abbreviated MARC to DC crosswalk.

Generic Library MARC-Dublin Core crosswalk			
CONTENTdm field name	MARC tag	MARC subfields	Dublin Core element
Call number	090	a	Identifier
Local control #	907	a	Identifier
Creator	100	a c d	Creator
Title	110	a b c d	Title
	245	a b c	
Alternative title	246	a	Title-alternative
Contributors	700	a c d t	Creator
	710	a b y	
Publisher	260	a b	Publisher
Date created	260	c	Date-Created
Item description	300	a b c d e	Format-extent
Cartographic data	255	a b	Coverage
Notes	500		Description
Subjects	650	a	Subject
Geographic location	651	a x	Coverage-spatial
Genre	655	a 2 y	Type
Local info	910	a	Identifier

Figure 1. Sample local crosswalk

- In CONTENTdm Administration, configure your DC collection fields according to your MARC to DC crosswalk. (To learn more about field properties, see **Editing Field Properties.**)

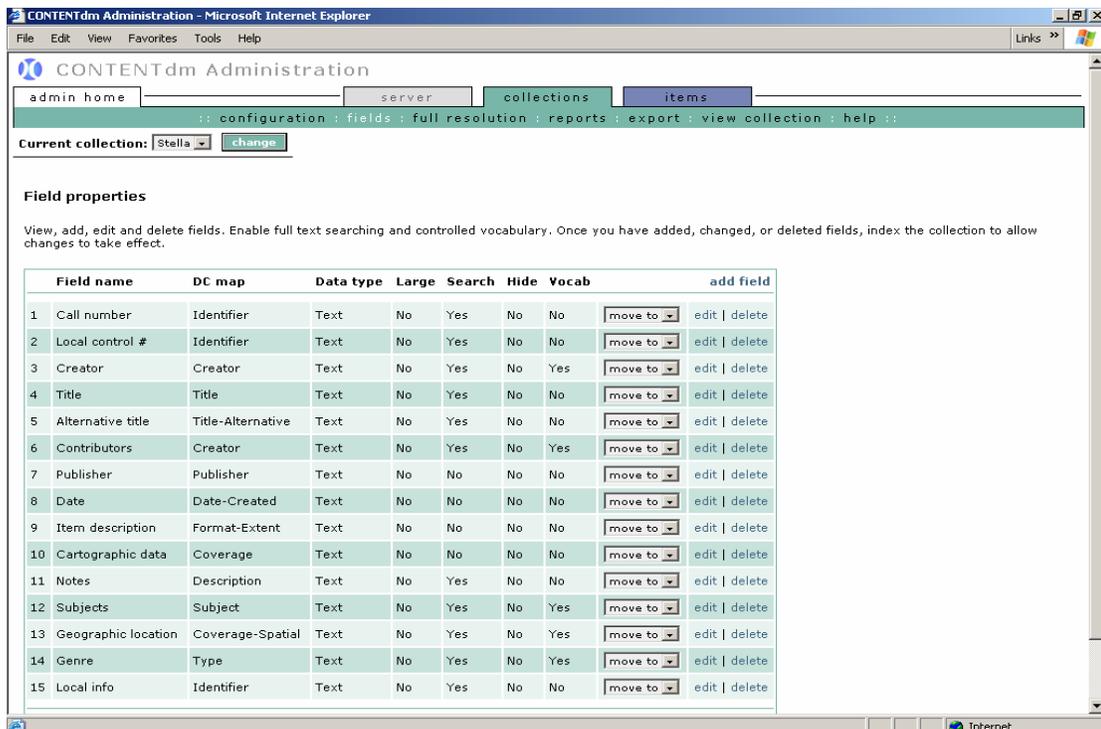


Figure 2. Configure field properties in CONTENTdm Administration

Later, you will need to make sure that the first row of your tab-delimited file contains the field names for your collection. An easy way to ensure that the first row will match your field properties is to export them now while your collection is still empty, as described in the following step.

3. In CONTENTdm Administration, select the **Collections** tab and choose **Export**.

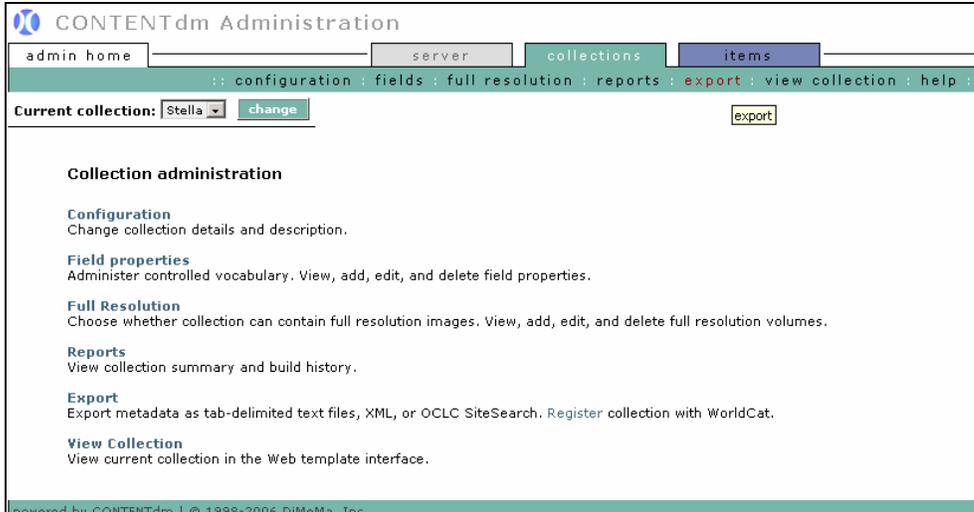


Figure 3. CONTENTdm Administration **Collections** tab

4. On the next screen, select the **Tab-delimited** option and confirm that the **Return field names in first record** option is checked. Click **Next** in the lower-right.

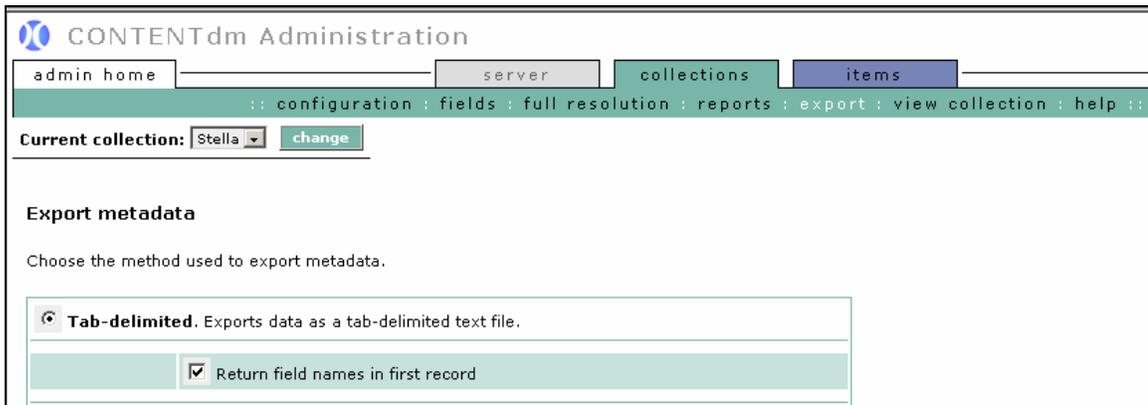


Figure 4. Export your empty collection with the field names in the first record

5. Right-click the link to the file, and select **Save Target As**. Save the file to a directory of choice.



Figure 5. Save the tab-delimited text file

To export a tab-delimited file using MarcEdit 5.0:

For import into CONTENTdm, you will need to create a tab-delimited file using MarcEdit. Using the crosswalk as a guide, choose the MARC fields (and subfield[s] as appropriate to the map) and export the resulting MARC file to a tab-delimited text (.txt) file.

Tip: You may be able to export entire MARC records from your online public access catalog (OPAC) using your integrated library system's **Export** command, and then use MarcEdit to remove unnecessary fields from the tab-delimited file. Alternatively, you can use MarcEdit to export records with only the required fields and subfields from your OPAC directly to the tab-delimited output file. Using the MarcEdit tools MarcBreaker and MarcMaker, you can evaluate which fields are actually used in your catalog. For step-by-step instructions on this alternative approach, see **Appendix A: Use MarcEdit 5.0 to Evaluate MARC Fields.**)

1. Open MarcEdit 5.0.

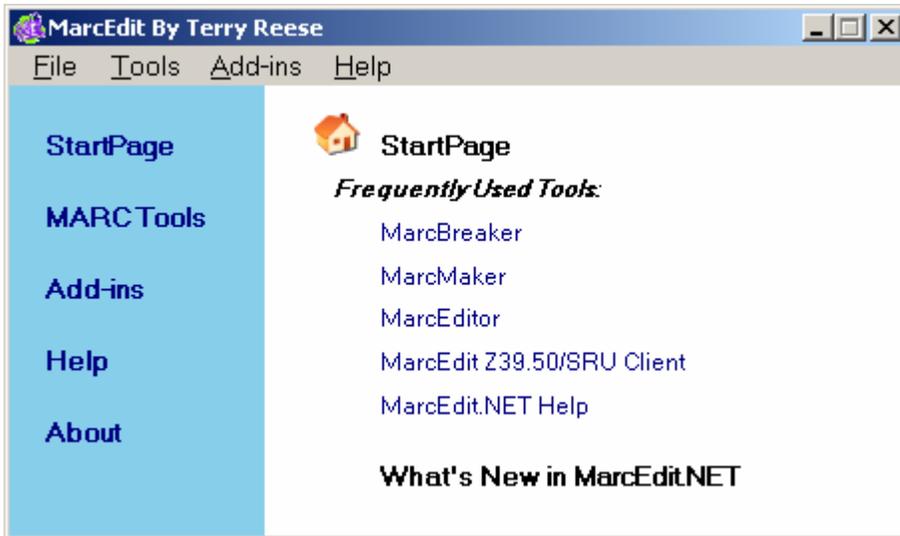


Figure 6. StartPage for MarcEdit 5.0

2. From the **Tools** menu, select **Export Tabbed Delimited Records.**



Figure 7. Tools menu with **Export Tabbed Delimited Records** selected

3. Choose the input file (.mrc, .out, and so on) and output file (tab-delimited .txt) by selecting the **Normalize field data** option. This option removes the subfield and indicator characters from the MARC field. Then select the MARC tags to export using your crosswalk as a guide. Add each field and then click **Export**.

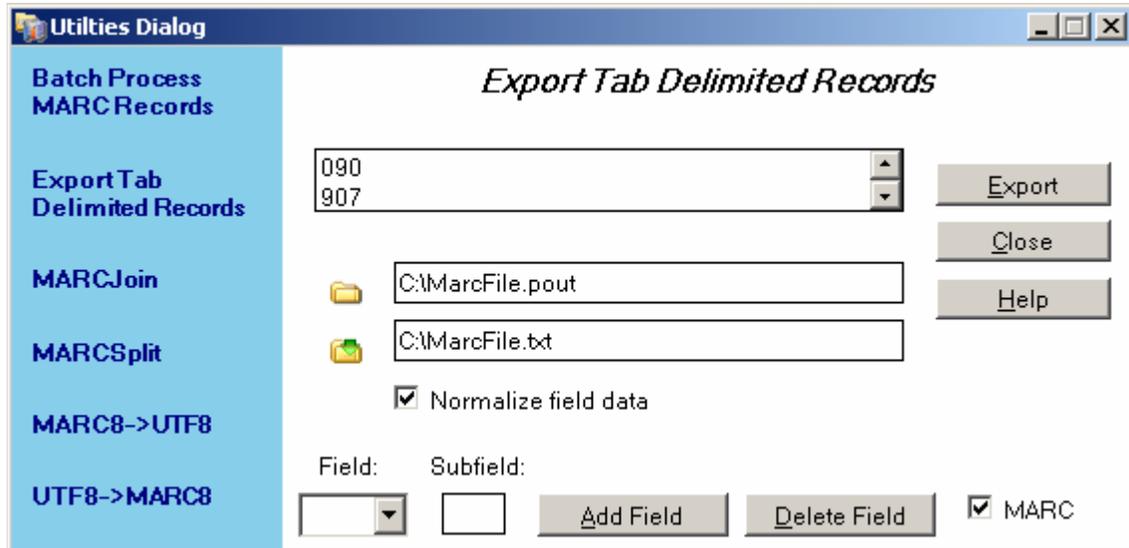


Figure 8. Confirm Normalize field data is checked

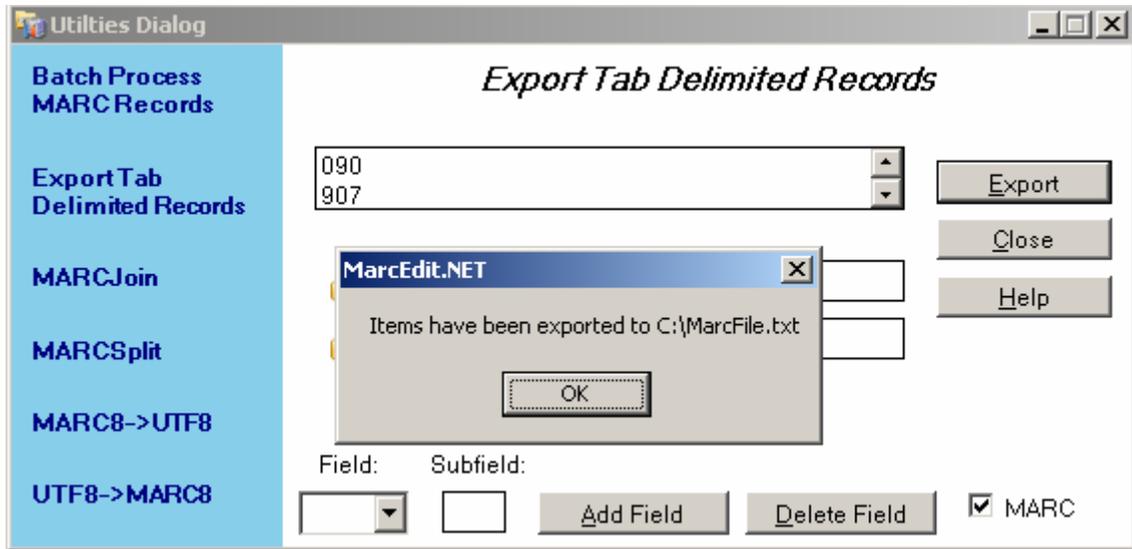
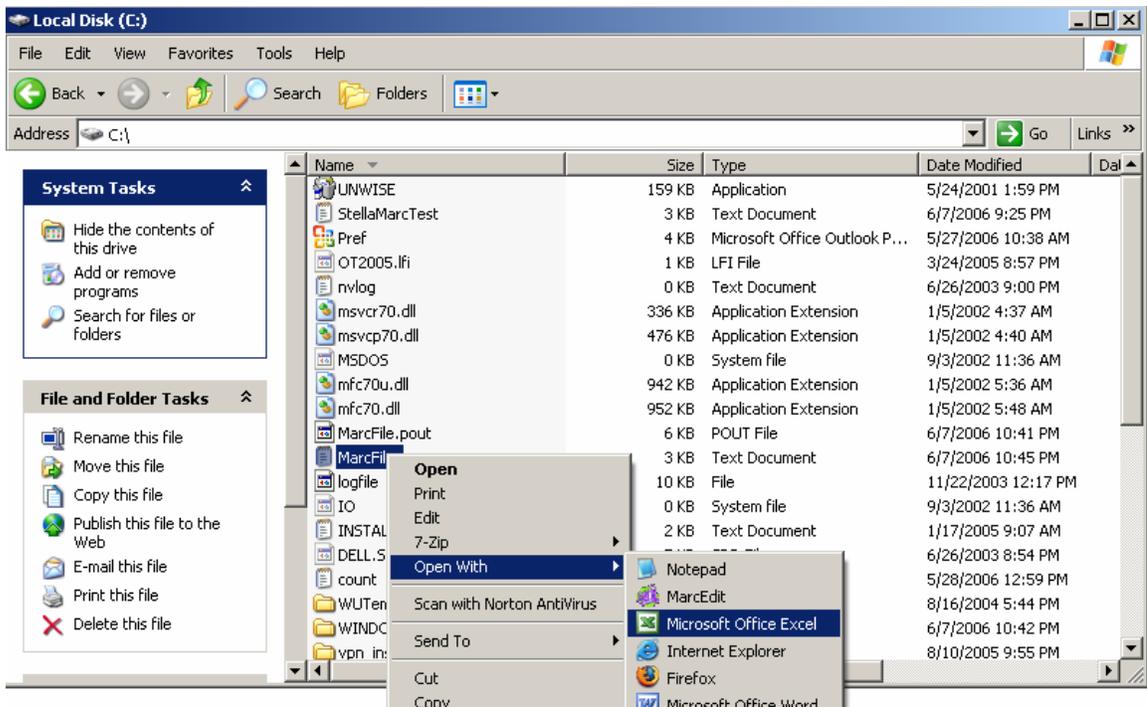


Figure 9. Tab-delimited file MarcFile.txt is created

Section II: Using Microsoft Excel to Prepare File for Import

After you have created your tab-delimited file from a MARC file, you need to configure and clean the data for importing into CONTENTdm.

1. To remove unwanted characters from the MARC file, locate the MarcFile.txt file on your computer. Right-click on the file name, and choose **Open With**, and then select **Microsoft Office Excel**.



Importing from Tab-Delimited Files

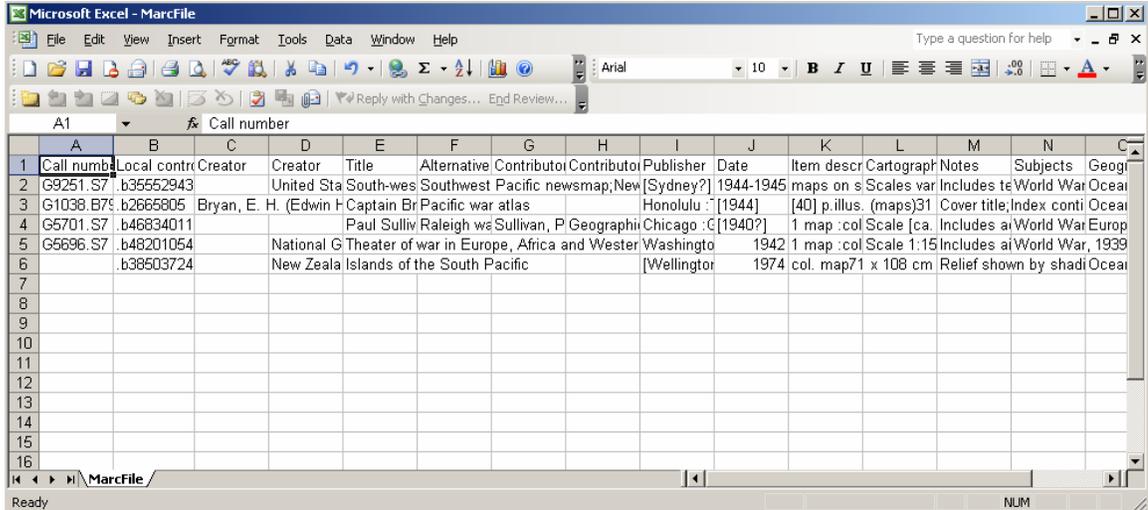


Figure 10. Open the file with Excel

Note: MarcEdit places comma-separated MARC tags in row 1. You can map these during the CONTENTdm import process, or you can edit them now to reflect exactly the CONTENTdm fields to which they are mapped.

2. If you choose to edit now, right-click on the export.txt file that you exported from CONTENTdm and open with Excel.

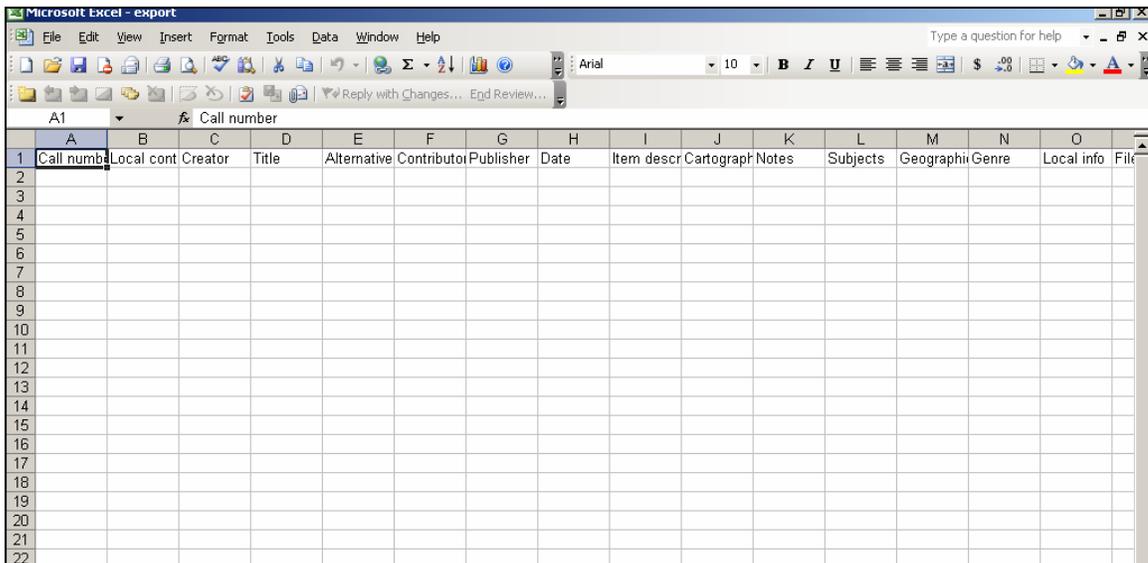


Figure 11. The file contains your collection field names on the first row

3. With both files open in Excel, from the **Window** menu, select **Arrange...** to view the open files horizontally at the same time. You can use the field names as a reference as you format the file for import. Copy the field names and paste them in the first row as appropriate for the column.

Importing from Tab-Delimited Files

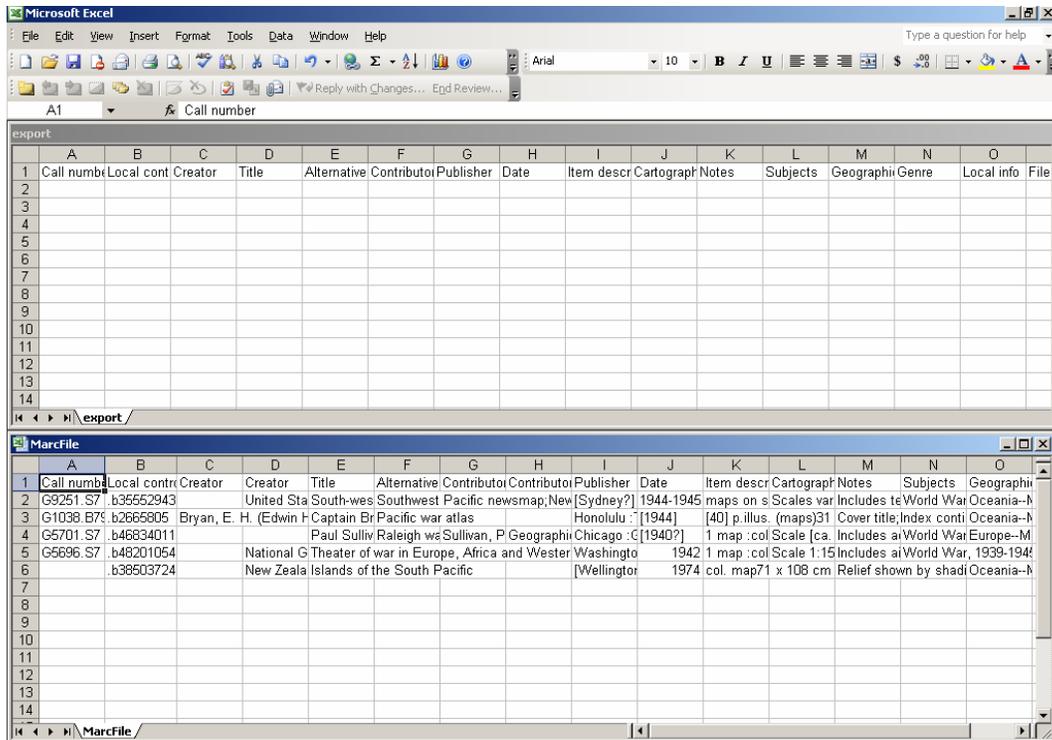


Figure 12. Open MarcFile.txt and export.txt and view in separate windows for reference

4. Working with your MARC tab-delimited text file, make sure the columns are correctly aligned and clean up the text as needed (for example, check for multiple tags mapped to the same field name, remove double quotation marks, and so on).

For more information about using the Excel CLEAN function, see **Appendix B: Cleaning Data in Microsoft Excel**.

After cleaning up the file, save the file as a tab-delimited text file (.txt), and then you can import it into CONTENTdm.

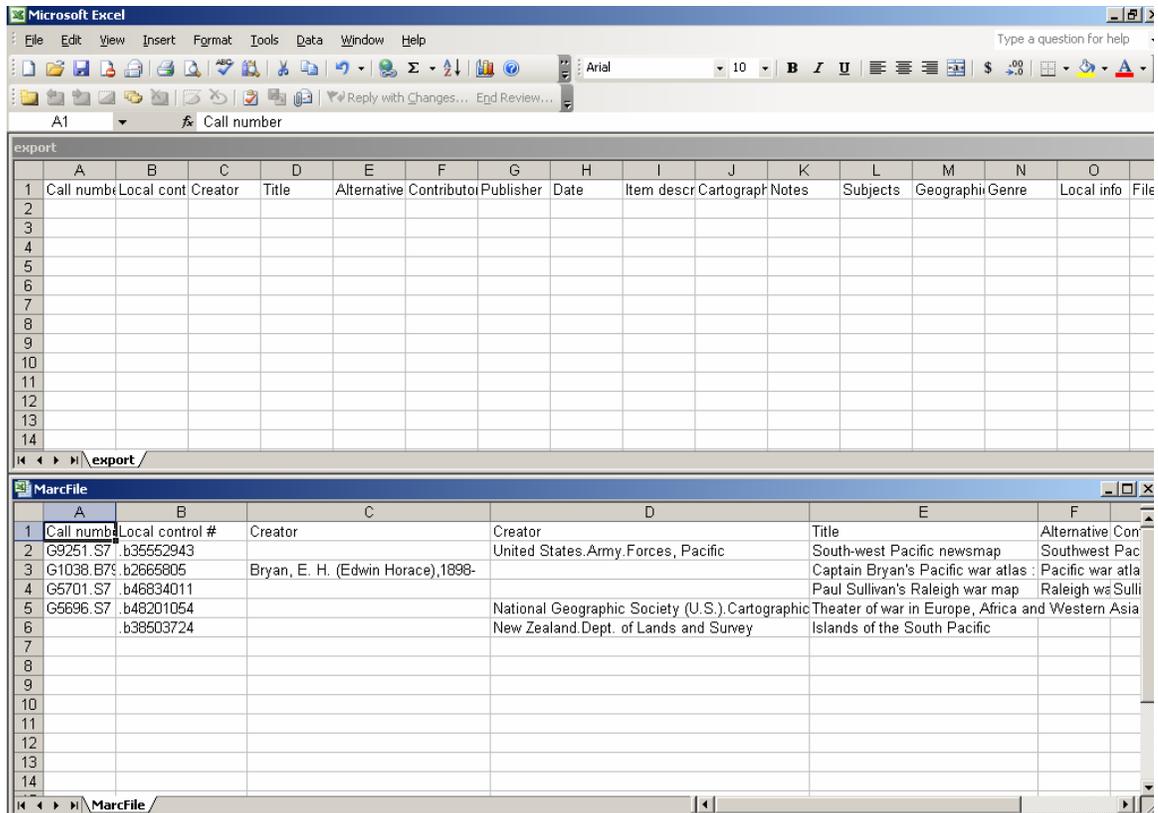


Figure 13. Save file as a tab-delimited text file after cleaning up

Section III: Importing Dublin Core Metadata Only from MARC File Output

When you have metadata (cataloging) already created but the digital objects are not ready to import into CONTENTdm, you can create CONTENTdm items using the metadata only. Digital media source files can be added at a later time.

Note: CONTENTdm can be used to manage URLs and other metadata-only records. You can add images later, if appropriate. (More typically, you will have the digital files ready to import when importing the metadata—CONTENTdm can accommodate that, as well.) The root name of the digital file becomes the Object filename during a mapping step in the wizard.

To import your cleaned metadata file into your project in the Acquisition Station:

1. From the **Import** menu, select **Multiple Files** (or click the file cabinet icon in the toolbar) and then select the radio button for **Import using a tab-delimited text file**.
2. Browse to the tab-delimited text file that contains the metadata you are importing.



Figure 14. Import a tab-delimited text file containing only metadata

3. On the **Import Type** screen, select **Import metadata only**.

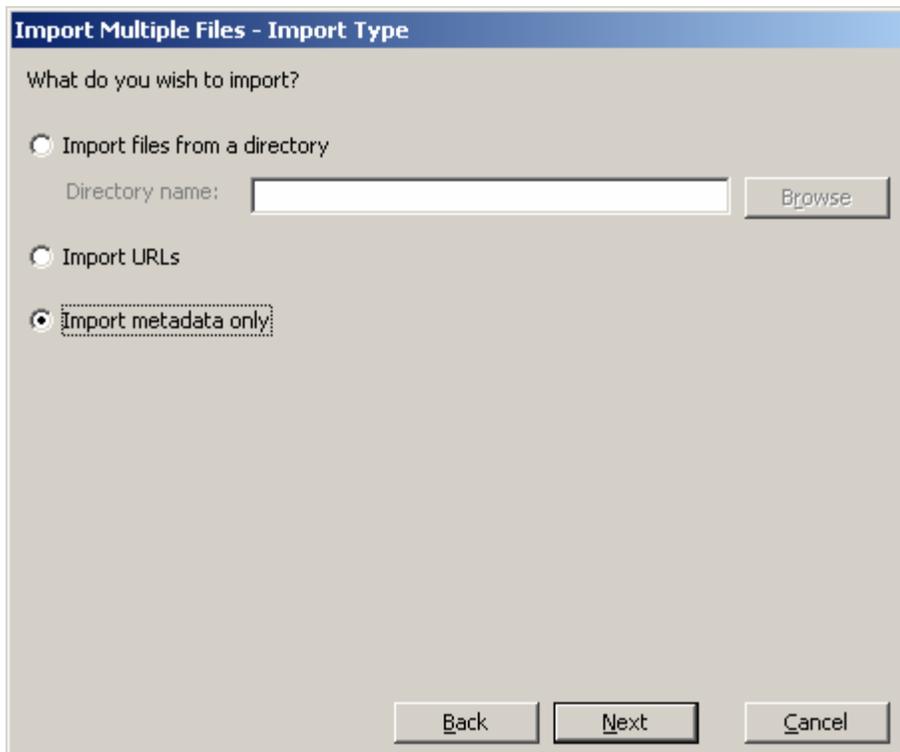


Figure 15. The file MarcFile.txt contains metadata only

4. Map your collection fields to the imported fields, adjusting where necessary. Use the drop-down options on the right side to match the imported fields to your collection field names.

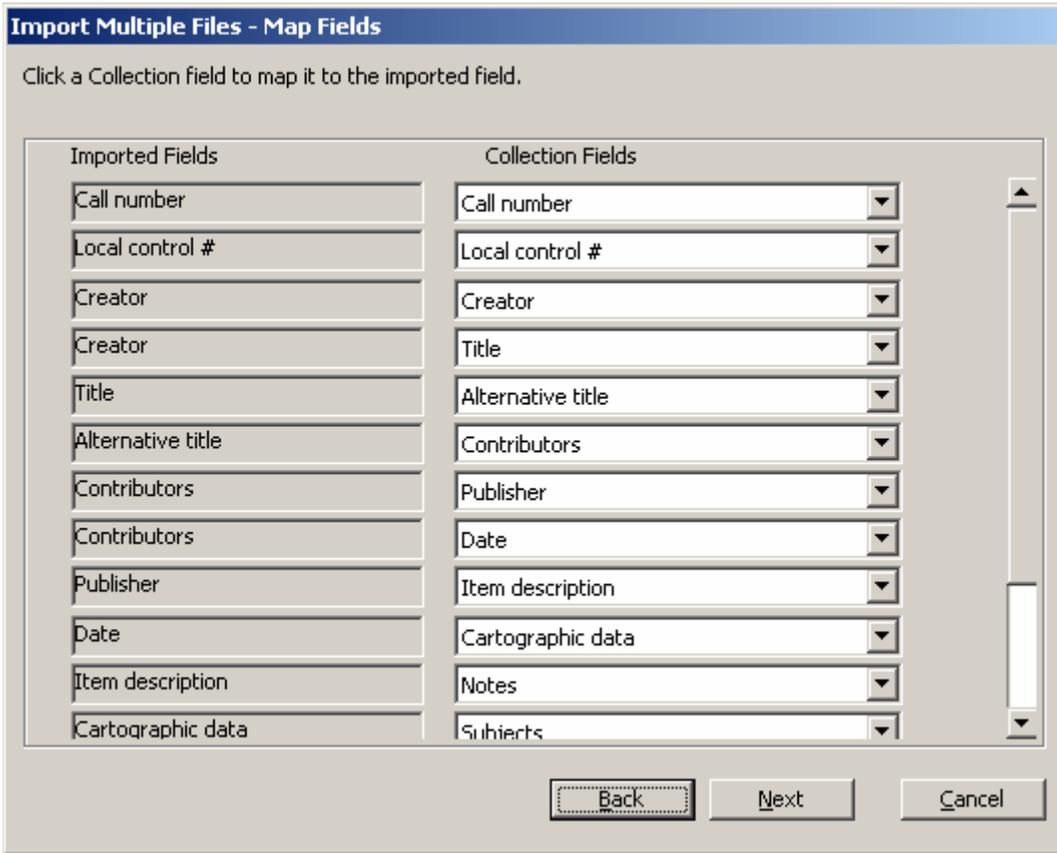


Figure 16. View of **Map Fields** screen before collection fields are mapped to imported fields

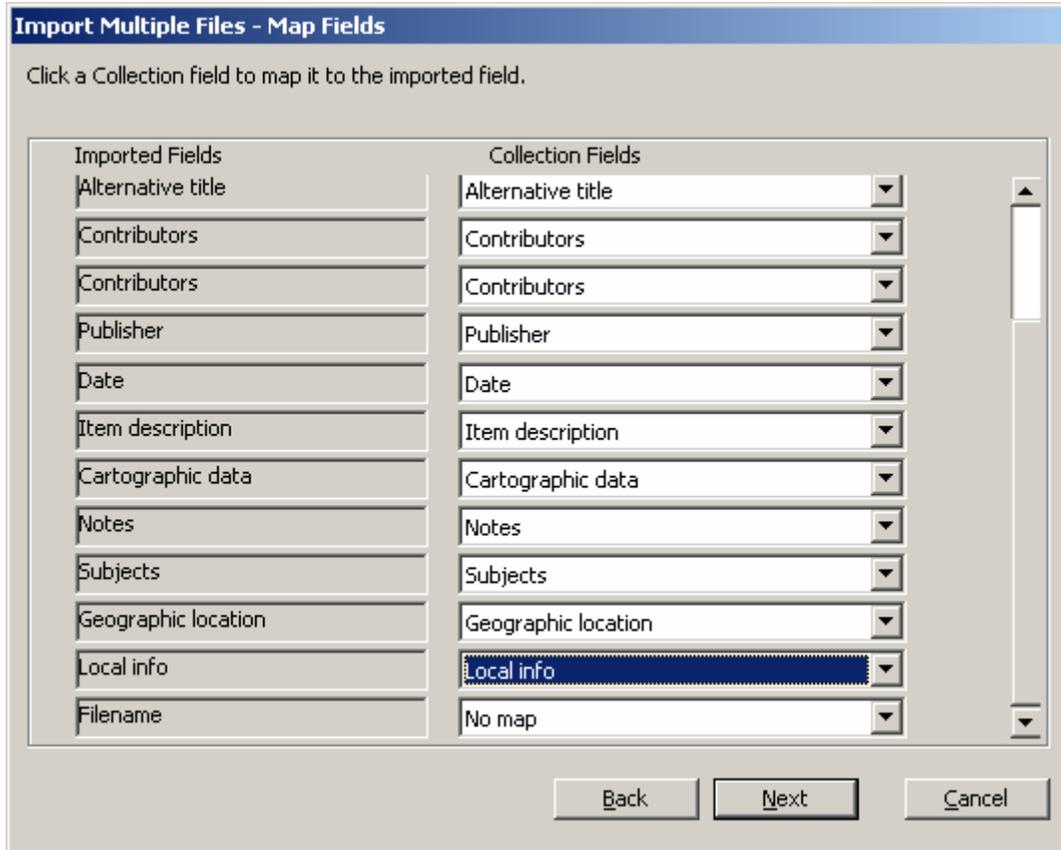


Figure 17. View of **Map Fields** screen after all collection fields are mapped to imported fields

5. In the previous figure, notice that the last field, **Filename**, is mapped to the label **No map**. This indicates that the text file contains no reference to a digital file, and only the metadata will be imported.

Note: If you elected to Import files from a directory earlier in the wizard, the **No map** field would be replaced with **Object file name**. You would use this field to specify the import field with the digital media source file.

Click **Next**. The metadata records are imported, and the default thumbnail for a null value record is used. (The default thumbnail is the same thumbnail as is used by default for a URL.)

Importing from Tab-Delimited Files

Icon	Call number	Local control #	Creator	Title	Alternative title	Contributors	Publsh
	G9251.57 svarU5	b35552943	United States Army, Forces, Pacific	South-west Pacific newsmap	Southwest Pacific newsmap;Newsmap;SWPA newsmap		[Sydney Informa USAAFF
	G1038.B79 1944	b26658057	Bryan, E. H. (Edwin Horace), 1898-	Captain Bryan's Pacific war atlas :containing 75 detailed maps of Pacific areas and of the world,	Pacific war atlas		Honolulu Publishi Compar
	G5701.57 1940.P3	b46834011		Paul Sullivan's Raleigh war map	Raleigh war map	Sullivan, Paul; Geographical Publishing Co	Chicago Publishi
	G5696.57 1942.N38	b48201054	National Geographic Society (U.S.), Cartographic Division	Theater of war in Europe, Africa and Western Asia /compiled and drawn in the Cartographic Section of the National			Washing Geogra
		b38503724	New Zealand Dept. of Lands and Survey	Islands of the South Pacific			[Wellin

5 Objects

Hold Delete Upload

Figure 18. All records are ready to upload

6. Click **Upload** to upload all of the items in your project. Then, in CONTENTdm Administration, approve all the items, and index the collection.

Your collection now contains the metadata for your items.

CONTENTdm Collection

home : browse : advanced search : preferences : my favorites : about : help

select all : clear all : add to favorites

Browsing 5 item(s) in: Stella

Image:	Title:	Subjects:
<input type="checkbox"/> 1.	Captain Bryan's Pacific war atlas :containing 75 detailed maps of Pacific areas and of the world, also complete index of 5000 place names /by E.H. Bryan, Jr	
<input type="checkbox"/> 2.	Islands of the South Pacific	
<input type="checkbox"/> 3.	Paul Sullivan's Raleigh war map	World War, 1939-1945--Maps
<input type="checkbox"/> 4.	South-west Pacific newsmap	World War, 1939-1945--Maps
<input type="checkbox"/> 5.	Theater of war in Europe, Africa and Western Asia /compiled and drawn in the Cartographic Section of the National Geographic Society for the National Geographic magazine; James M. Darley, chief cartographer	World War, 1939-1945--Maps

select all : clear all : add to favorites

powered by CONTENTdm © | contact us

Figure 19. Browse view of metadata-only records in a collection



Figure 20. Item-level view of one metadata-only record

Appendix A: Using MarcEdit 5.0 to Evaluate MARC Fields

An alternative workflow enables you to determine whether you need all of the fields present in your MARC records before you create the tab-delimited output. You can use MarcBreaker first to break the MARC file into a more readable format and then evaluate the data.

To evaluate MARC fields:

1. Open MARCEdit 5.0, and click **MarcBreaker**. Supply the input and output files.

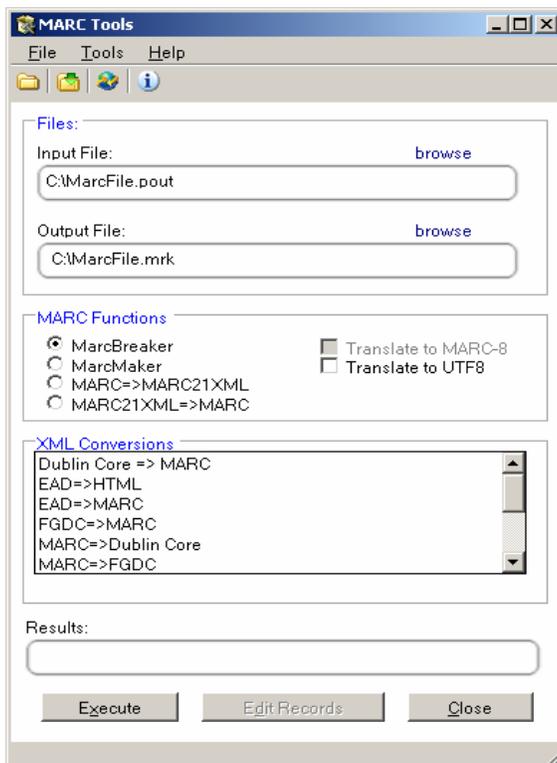
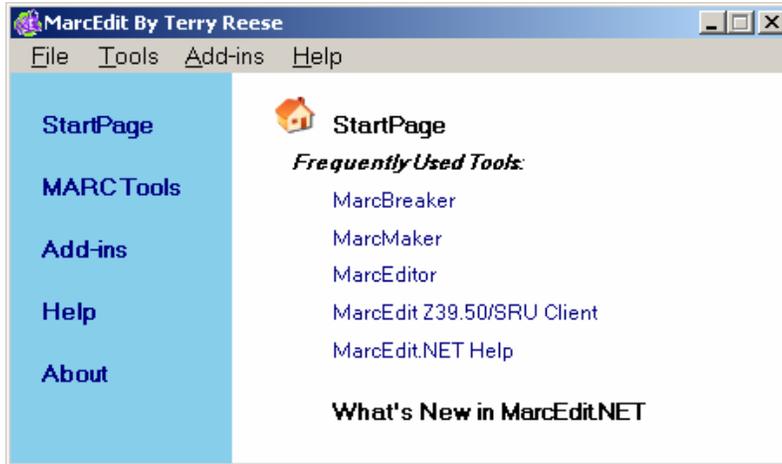


Figure 1. MarcBreaker screen

2. Click **Execute** to process your records.

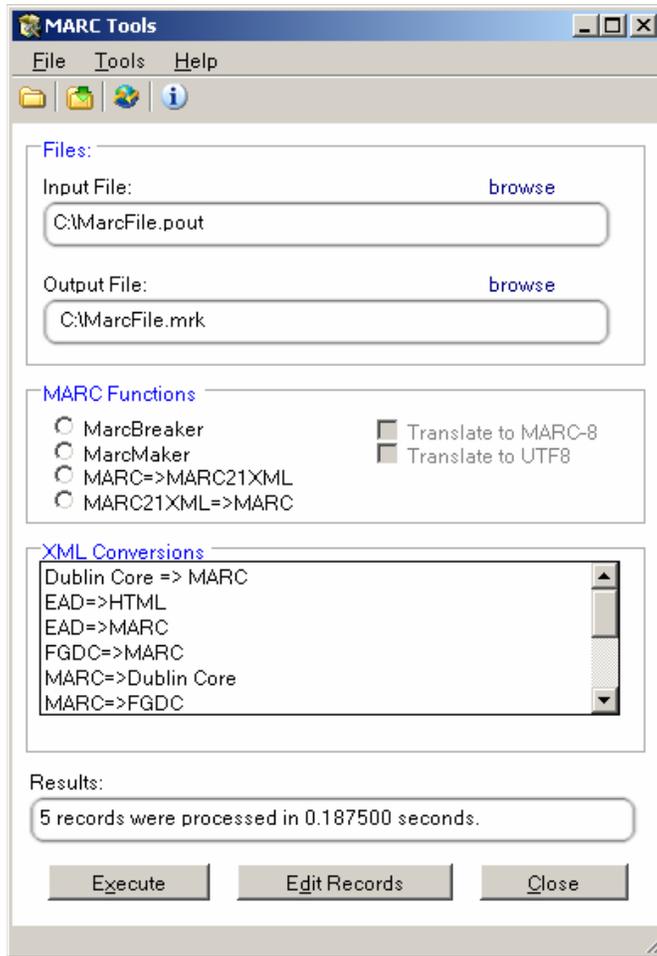


Figure 2. File processed for edit in MarcEditor

3. Click **Edit Records** to display records in MarcEditor. You can edit the records as needed.

Importing from Tab-Delimited Files

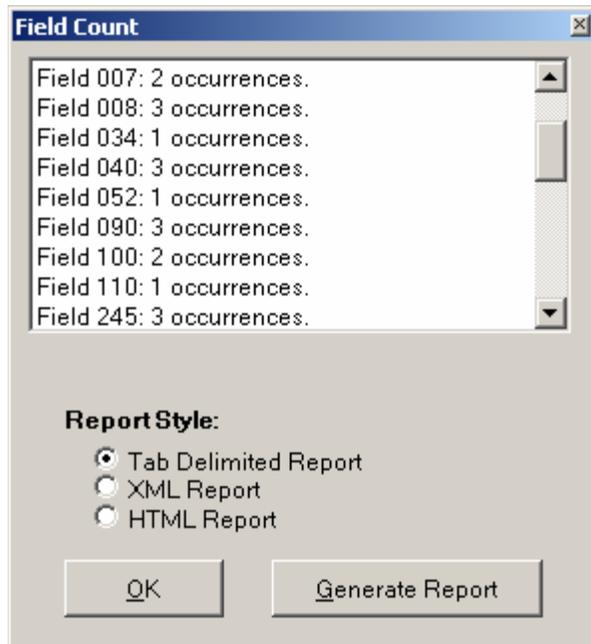
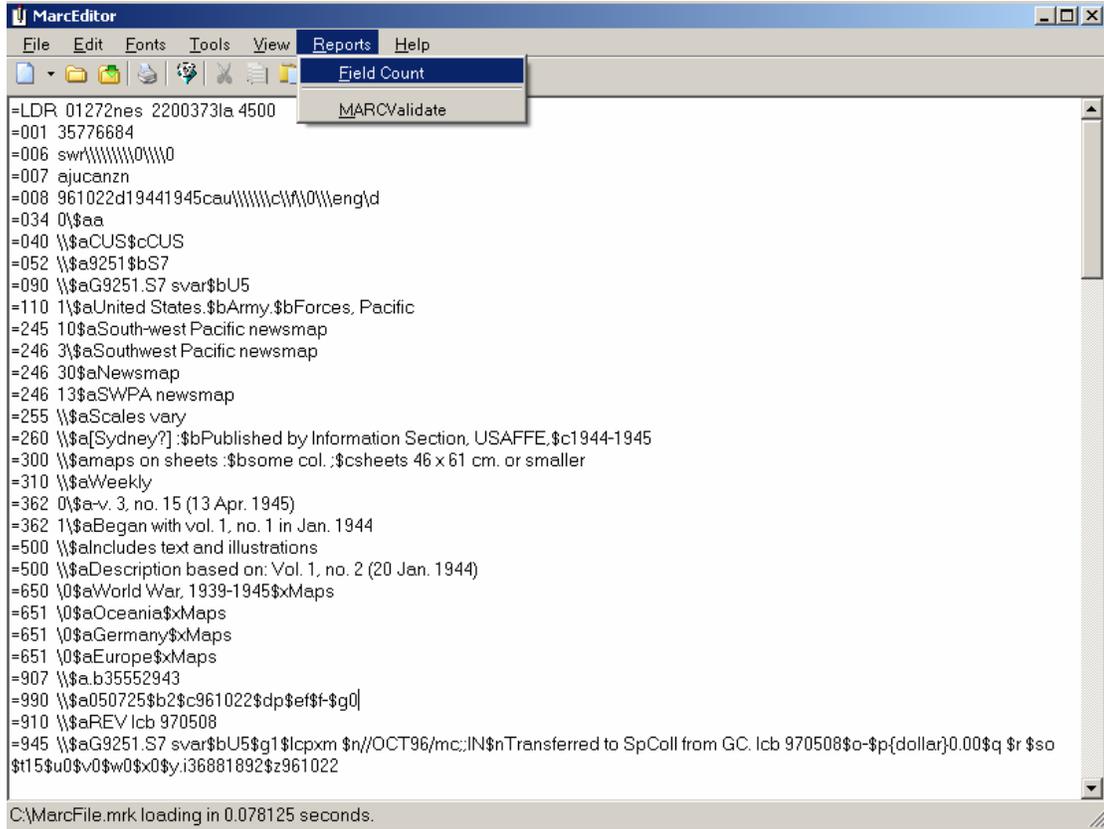


Figure 4. Field count report

5. After evaluating and editing, use the MarcMaker tool to convert the file back into MARC. Once again, supply the input and output files, and click **Execute**.

Appendix B: Cleaning Data in Microsoft Excel

To clean a column or row, you must have your data open in Microsoft Excel.

1. Create or choose a destination cell for the cleaned data. (In the following example, a new column is inserted after column A.)
2. Place your cursor in the first destination cell to fill in the formula (in this example, cell B2). Click **Insert Function**.
3. Choose **CLEAN**, and then click **OK**.

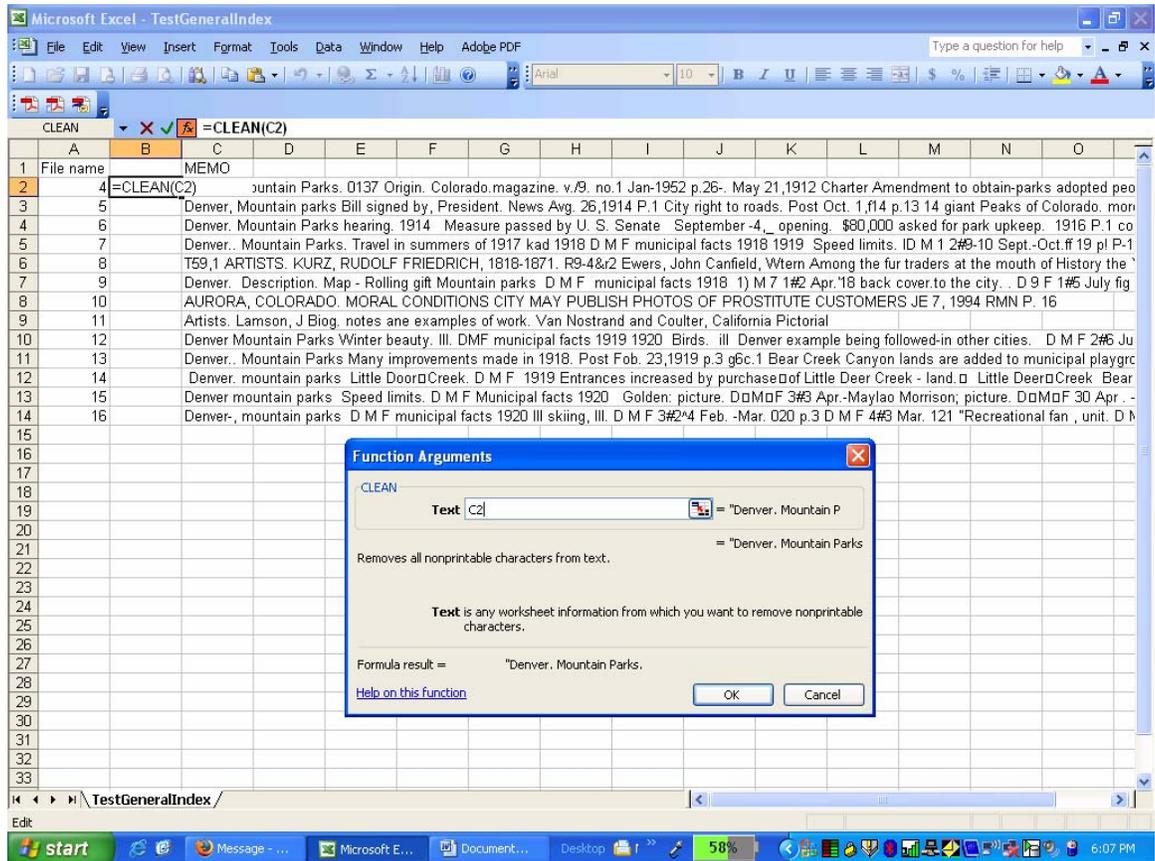


Figure 5. CLEAN function

The newly cleaned text, free of stray blanks, is inserted in cell B2.

Importing from Tab-Delimited Files

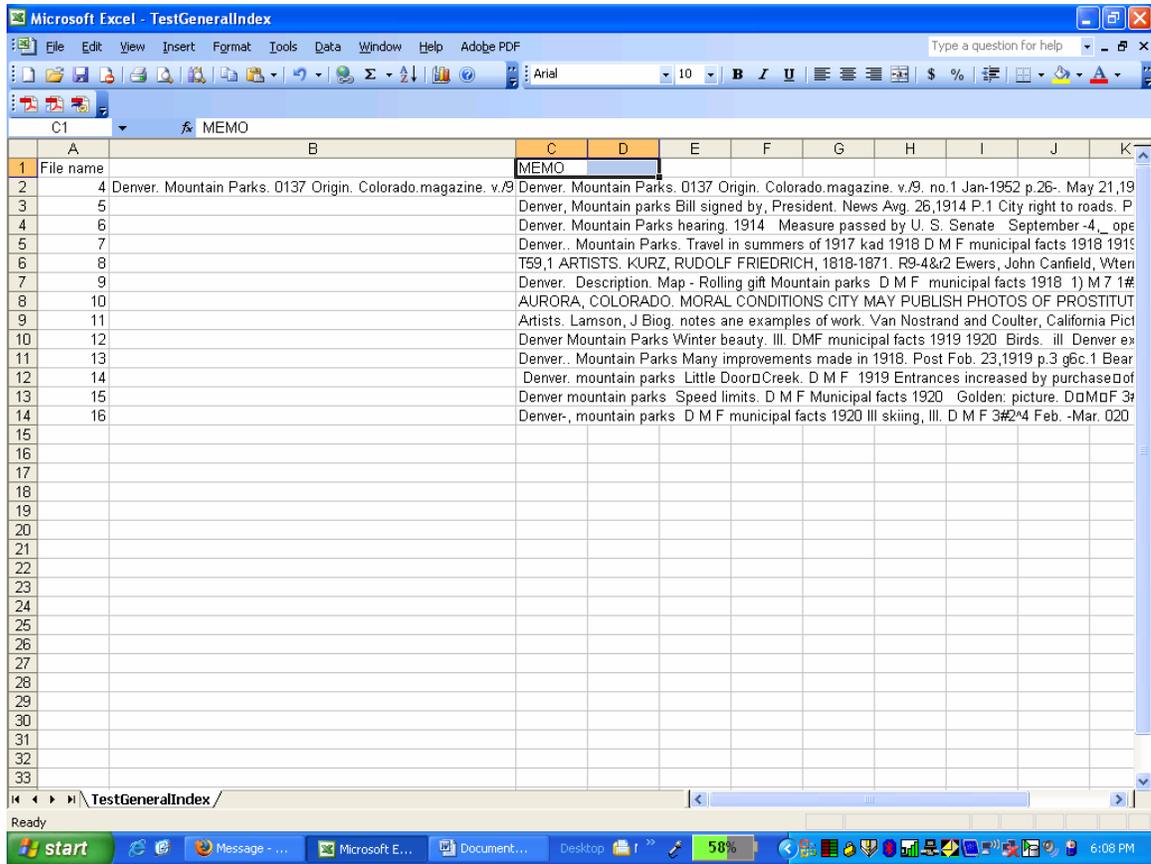


Figure 6. Cleaned text

- Use the fill handle to copy the formula to the rest of the column. The fill handle is the small black square in the lower-right corner of the selection. When you point to the fill handle, the pointer changes to a black cross.
- Click, hold, and drag the fill handle down the column. When you release the handle, column B contains cleaned text.
- Before you delete the original columns, you must make the destination cells contain the actual values. Remember that the destination cells only contain the formula. Highlight the column with the CLEAN formulas applied, then right-click and select **Copy**.

Importing from Tab-Delimited Files

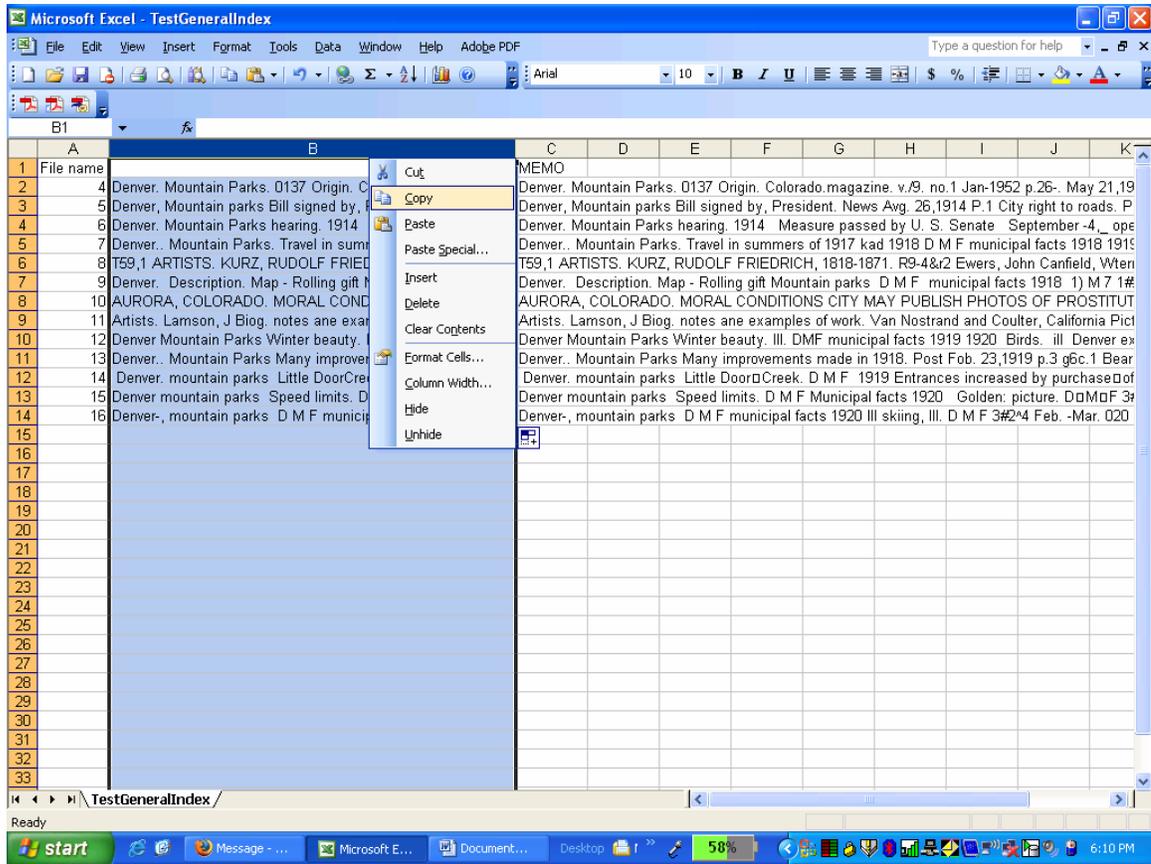


Figure 7. Copy columns to make the destination cells contain the actual values

7. Select the destination column. Then right-click and select **Paste Special**. Select the **Values** option, and then click **OK**.

Importing from Tab-Delimited Files

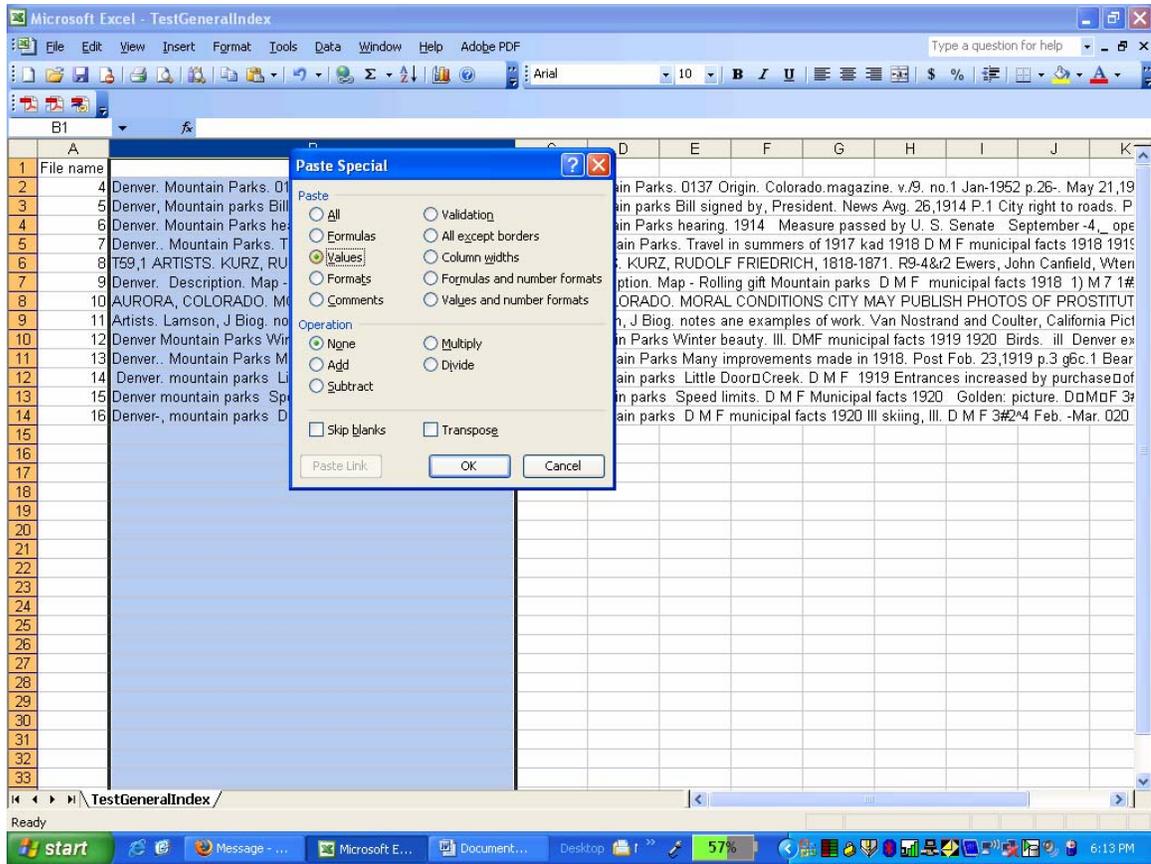


Figure 8. Columns showing the values, not the formulas

The column now contains the actual values and not the formulas.

8. You can now delete the original column (in this example, column C).