# Metadata Process Notes (Tips, Tricks, Gotchas)

(revised 2/7/2017)

From there try using the command prompt usgs mp to convert the text file to an xml file that ArcCatalog understands.

Import Type: FROM\_FGDC

Basically workflow is:

* Opened metadata html file in Acrobat Pro. The file is then exported from Acrobat Pro into a Word format which preserved all of the indentation (makes metadata easier to read). If still present in the Word document remove the block of text just under the title that begins with “Metadata also available as…” including the hyperlinked bullet list under the Metadata header. This will get recreated later when the metadata is exported out of ArcCatalog to an html format using the mp tool.
* Make a first pass through the document to make the obvious changes like spatial extent, dates, etc. Turn on ‘Track Changes’ in Word and then open up editing/commenting to other TCC team members. Finalize edits in a team review session so that decisions can be made about lingering edits or comments. Make one more pass through to check document for any issues. Once editing is complete save the Word document as an Acrobat pdf and re-export as a plain text file. The text file will have all lines left-justified.
* Create metadata in word using previous document that has all the fields needed
* In Word export document to a pdf
* In Acrobat Standard (or Pro) export to text (plain). Not sure if this will work with Acrobat Reader.
* Open the plain text file in Notepad++ and carefully go through formatting. One useful trick is to start at the top of the document and do a find on full colons. For the most part there should only be one colon per line corresponding to a separator between an attribute field name and the value associated with the attribute. The exceptions to this will be the References section where a lot of colon can exist, the Corner Coordinate section, and in some cases, the Process Notes where processes are enumerated or URLs are listed. Once this review is complete make another pass through the text document looking for blocks of text, embedded URLs, etc., that have been placed on separate lines. For example, in Notepad++ you will see separate line numbers where a block has been broken by a carriage return.
* Copy the text file to a local drive. Open a command prompt window and navigate to the location of the text file. Run the plain text file through the USGS tool ‘cns’ which stands for chew and spit (<https://geology.usgs.gov/tools/metadata/tools/doc/cns.html>). This tool will attempt to pre-parse the text file using the attribute field names with colons into a formatted file that the USGS tool ‘mp’ (metadata parser) can understand. The basic syntax for cns is: ‘cns inputfilename.txt -o outputfilename.txt’. There are many other flags that can be used with this tool but the command above does what I need it to do. The output from the cns tool needs to be reviewed carefully, looking for the same issues that were present in the plain text file. In this case rather than multiple attribute field names on the same line the typical problems are the separation of attribute field names from their values on to two lines and incorrectly indented attribute field names (use the space bar not tabs to fix this issue). Once editing is complete save the file and run the USGS ‘mp’ tool.
* The USGS ‘mp’ tool is used to parse the cns text file and generate an xml file that can be imported using ArcCatalog. The basic syntax for mp is: ‘mp inputfilename.txt -x outputfilename.xml’. Again there are many flags that can be included but the command above gets the job done. Copy the cns and mp outputs back to the folder containing the raster to be associated with the metadata. Close the command prompt window.
* Open ArcCatalog and navigate to the folder containing the raster and the mp xml output file. Click on the raster file name in the TOC and then the Description tab in the viewer window to open metadata for the raster. Click on Import in the viewer window to open the dialog box. For the Source Metadata select the xml output file from the mp tool. The Import Type should be set to ‘From\_FGDC’. The Target Metadata is automatically set to the current raster image. Leave the box to Enable automatic updates checked which will allow fields like extents and bounding coordinates to automatically be adjusted to match the target raster. The processing box usually blathers a bit about not finding this or that; don’t worry about that for now. When processing finishes click on Edit in the viewer and replace the raster file name in the Title box with the title you want to use for the metadata and then click Save. For some reason anytime you click on save the focus leaves the raster and you think you have lost all of your metadata. Just click on the raster name again in the TOC and all is well; you see the name of the metadata has changed to whatever name you gave it.
* Click on Export in the viewer window to open another dialog box. This time the Source Metadata is the raster you just linked to metadata. The Translator should be correctly set to ‘ArcGIS2FGDC.xml’. Edit the Output file name to some shorter name ending in ‘2html.xml’. (This is just my convention to clearly separate this output from the .xml file we will keep associated with the image.
* ArcCatalog has incorporated a version of the USGS mp tool in its metadata toolbox. In ArcCatalog’s Search window (if not open it can be found under Windows|Search), type ‘usgs\_mp’. The hyperlinked result, USGS MP Metadata Translater, will appear below the search box. Click on the hyperlinked blue text to open the dialog box. The Source metadata will be the file just exported with the 2html.xml (if my naming convention was used). The Conversion type should be set to ‘HTML’. The output file should be the raster file base name with an .html extension. Be careful of where you write the output. Typically the output folder is still looking at the previous raster location and has to be updated to the raster’s location.
* Double-click on the raster.html file to open it in a browser window. Make any edits that need to be made, like fixing the References section, in Notepad++. Carefully review the rest of the file to make sure it looks the way you expect.
* Clean up the folder by removing the intermediate files that are not longer needed.

Other little sinkholes I’ve encountered along the way:

* Check ArcCatalog’s customization settings. By default it appears that Customize/ArcCatalog Options/Metadata/Metadata Style by default is set to “Item Description”. This setting is pretty worthless for viewing/editing the full metadata record. Instead, set the Metadata Style to “FGDC CSDGM Metadata or one of the other ISO options. This is also where you can toggle the automatic update. If someone is unable to view the entire metadata record in their install of ArcGIS this is the setting that needs to be checked.