How to create an XML metadata file from HTML format

Summary

If you have an HTML file containing Federal Geographic Data Committee (FGDC) metadata in <u>outline format</u>, the HTML file can be converted to an XML metadata file. The advantage of this procedure is XML format can be easily modified in ArcCatalog and exported in several types of output (e.g. TXT, FAQ, SGML, HTML, and XML). The HTML file cannot be directly converted to XML, thus a series of procedures must be used to get the desired output. This process requires using Microsoft Word and the U.S. Geological Survey (USGS) metadata utility "cns" to produce a well-formatted text file, which can then be imported into ArcCatalog via the "import metadata tool". Note: simply saving the HTML as a .txt file will not be adequate to import into ArcCatalog due to formatting errors.

Before you begin, you will need to <u>download</u> the program "cns" from the USGS website. This program can be found within the self extracting executable file: "all_win.exe". The USGS website also has an <u>explanation of the programming details</u>. For details on how to install the USGS FGDC metadata utilities follow the <u>instructions</u> outlined by ESRI.

Procedure

1. The process of creating a "well-formatted text file" from HTML is described in detail by ESRI: <u>http://support.esri.com/index.cfm?fa=knowledgebase.techArticles.articleShow&d=23071</u>

2. The image below shows the command line procedure for creating the formatted text file with cns. Notice that you must first type cd <current directory> then "c:\usgs\tools\bin" to use the program.

3. At the c: $\sin > \sin > \s$

The example shows the output file "c:\Data\Metadata\out\text.txt created from the input file "c:\Data\Metadata\in\text.txt. The step is shown in the Figure 1.



Figure 1. Screenshot of program "cns" in command prompt.

Now the most important step!

4. Make sure the new text file shares the exact information (with appropriate formatting) as the original HTML file. In the example above you would open the file c:\Data\Metadata\out\text.txt and compare it to the original HTML file.

5. Open ArcCatalog and navigate to the file of interest (i.e. shapefile, raster layer, etc.). Please refer to Figure 2.

- **6.** Click the metadata tab to preview.
- 7. Click the import metadata button 🔚 .



Figure 2: Screenshot of selecting file of interest and click import metadata button

8. Browse to the appropriate file directory (e.g. C:\Data\Metadata\out\text.txt).

9. Select the FGDC CSDGM (TXT) format (refer to Figure 3).

Import M	etadata	? 🗙
Location:	C:\Data\IMS_Data\Selenium\Data\orphan_sit	Browse
Format:		
🔽 Enable	e automatic update of metadata	
		1
	ОК	Cancel

Figure 3: Browse to appropriate file directory and FGDC CSDGM (TXT) format

10. Check to see that the metadata information is correct and each item is under the appropriate section.

11. If the information is correct, click ok to import the metadata.

Congratulations, you are done!

Created by Cody Schroeder (Metadata Librarian) November 1, 2006. Idaho State University GIS Training and Research Center