

# DIGITAL MAPPING TECHNIQUES 2015

The following was presented at DMT'15  
(May 17-20, 2015 - Utah Geological Survey,  
Salt Lake City, UT)

The contents of this document are provisional

See Presentations and Proceedings  
from the DMT Meetings (1997-2015)

<http://ngmdb.usgs.gov/info/dmt/>

# **A Workflow and Geoprocessing Tool for Automating NCGMP09 Metadata**

By Evan E. Thoms

U.S. Geological Survey  
Alaska Science Center  
4210 University Drive  
Anchorage, AK 99501  
Telephone: (907) 786-7409  
Fax: (907) 786-7401  
email: ethoms@usgs.gov

## **INTRODUCTION**

A workflow and ArcGIS geoprocessing tool were developed to ease the burden of writing Federal Geographic Data Committee (FGDC)-compliant metadata (Federal Geographic Data Committee, 1998) for ArcGIS geodatabases that store geologic map data following the NCGMP09 data model (NCGMP09, 2010). The FGDC mandates that ArcGIS geodatabases, in general, are to be documented according to the Content Standard for Digital Geospatial Metadata (CSGDM), while the National Cooperative Geologic Mapping Program (NCGMP) has a soft requirement that geologic map publications funded by NCGMP will adhere to the NCGMP09 data model (NCGMP09, 2010). The NCGMP09 Metadata Updater tool described here simplifies the process of meeting these requirements by taking values and text from XML files and geodatabase tables and adding them to metadata documents for the feature classes and tables within a geodatabase. In this way, boilerplate metadata elements and values in the various dictionary-like tables in the geodatabase can be programmatically distributed among the data objects while avoiding the use of the slower, form-based ArcCatalog Metadata Editor or other un-scriptable metadata editors.

## **PURPOSE**

### **Programmatically Re-use Boilerplate Metadata Elements**

Many elements within CSDGM metadata document are often re-used from publication to publication. Some elements – Purpose, Access\_Constraints, and Attribute\_Accuracy\_Report, for example – might be copied verbatim from one document to the next while others – child elements within Description or Data\_Quality\_Information, for example – might be edited as appropriate for the particular geologic and digital content. Storing this content in a template XML metadata document allows for the distribution of those elements programmatically rather than manually.

### **Programmatically Write Entity\_and\_Attribute\_Information**

Within the Entity\_and\_Attribute\_Information metadata element required by the CSDGM are child elements that describe the tables (Entities) and fields (Attributes) in a geodatabase. While these could be stored in a template document and then edited as necessary, the approach recommended here is to store the values and definitions in a dictionary-like XML document from which the items can be retrieved modularly depending on the content of the particular geodatabase. A file called NCGMP09\_entity\_definitions.xml is included with the code for the geoprocessing tool. NCGMP09\_entity\_definitions.xml contains table and field names, with definitions taken from the NCGMP09 documentation. Because the NCGMP09 schema is extensible, new tables and fields can be added to the dictionary so they are included during automation.

### **Programmatically Write Attribute\_Domain\_Values**

The NCGMP09 specification requires that some data, which might otherwise have only appeared in metadata, be promoted to tables in the database, for example the Glossary, DescriptionOfMapUnits, and DataSources tables. Glossary is a table of possibly confusing terms used in other attribute tables and their definitions. DescriptionOfMapUnits lists and describes the map units in the database. DataSources lists the sources from which geologic content, analytical results, terms and definitions, etc. have been compiled. These are the same types of values that are required in an Enumerated\_Domain element. Table 1 shows the mapping between CSDGM elements and data table fields. Assuming the tables have been filled out first, which would normally be the case during map production, it makes sense to programmatically write the values appropriate for each feature class and field to the metadata.

Table 1. FGDC Enumerated\_Domain elements and their equivalent NCGMP09 fields

<b>FGDC element</b>	<b>NGCMP09 table / field</b>
Enumerated_Domain_Value	Glossary / Term, MapUnit / Description
Enumerated_Domain_Value_Definition	Glossary / Definition
Enumerated_Domain_Value_Definition_Source	DataSources / Source

### **Batch Process**

While a single “report-level” metadata document is useful for generally describing the geodatabase and its content, the CSDGM is a content standard for a single standalone dataset and, as written, does not permit the detailed description of nested or multiple datasets and their attributes. Thus, a separate detailed metadata document is also required for each feature class and table. At the same time, there are many elements useful in describing the entire geodatabase that the author may wish to migrate to the metadata of all objects. The geoprocessing tool automates the tedious cross-referencing required to write a report-level metadata record as well as a separate document for each object.

## **WORKFLOW AND USE OF THE TOOL**

### **Prepare DescriptionOfMapUnits, Glossary, and DataSources**

As these tables are required by NCGMP09, they should be complete by the time the metadata preparation begins. If Glossary has not been completed, use the Create Glossary Stub tool included with the NCGMP09 Metadata Updater tool (Figure 1) to build a table listing the controlled vocabulary terms that are being used in the geodatabase. Use this as a starting point for adding definitions and definition source ID's either by writing them manually or by joining the table to previously-used Glossary tables or other tables (style file tables, or analysis results spreadsheets, for example) containing the terms and definitions you need to calculate over to your new Glossary. Run the Validate Database tool from the NCGMP09 Toolbox ([http://ngmdb.usgs.gov/Info/standards/NCGMP09/docs/NCGMP09v1.1\\_Tools2\\_Arc10.1.zip](http://ngmdb.usgs.gov/Info/standards/NCGMP09/docs/NCGMP09v1.1_Tools2_Arc10.1.zip)) to find mistakes and omissions.

### **Edit the boilerplate metadata template**

Starting with a copy of an XML metadata file from a previous but similar publication, make changes to the appropriate elements (most likely Identification\_Information, Data\_Quality\_Information, Distribution\_Information, and Metadata\_Reference\_Information) to reflect the content of the new geodatabase (see Available Free Software).

### **Edit NCGMP09\_entity\_definitions.xml**

NCGMP09 feature class, standalone table, and field names are listed and defined in NCGMP09\_entity\_definitions.xml (found in the \docs folder in the data release; see next section). The XML elements follow the same schema as those within the Entity\_and\_Attribute\_Information element. If non-NCGMP09 data objects have been added to the geodatabase or NCGMP09 data objects have been edited, make changes to this file as necessary (see Available Free Software) keeping the schema and file extension (.xml) intact.

### **Download and run the Metadata Updater tool**

The NCGMP09 Metadata Updater tool can be downloaded from <https://github.com/evanthoms/NCGMP09-metadata>. From the lower-right on that page, select 'Download ZIP'. Copy the ZIP file to a location of your choice and unzip it. Browse to the location in ArcCatalog to see the NCGMP09 Metadata toolbox and the tools within.

The Metadata Updater tool takes as input the location of a geodatabase, the location of a boilerplate template, some parameter selections, and the location of a folder into which output files will be written (Figure 2).

XML files for all input objects are written to the output folder, but other formats are available as well. Output files are generated by mp.exe (Schweitzer, 1995b) and are translations of the XML file into more readable forms. Options include a text version where the CSDGM

XML tags have been replaced by longer, more descriptive, text; an HTML version of the text version; and an HTML version where metadata items are presented in a FAQ format with questions at the top of the document linked to answers further below.

## Available Free Software

XML Notepad (<https://xmlnotepad.codeplex.com/>). Best used for copying and pasting XML elements within or between documents, not for editing the text of those elements.

Notepad ++ (<https://notepad-plus-plus.org/>). Best used for editing the text of XML elements.

Tkme (Schweitzer, 1995b). Best used for inserting new CSDGM elements and storing and re-using ‘snippets’, such as a Citation element.

XML elements in the Python code are manipulated using the ElementTree module, which has been a standard library since Python 2.5.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## CONCLUSION

The goal in developing this workflow and tool was to automate the writing of as many lines of metadata as possible from re-usable sources and to batch-process many documents at once. The speed at which the tool runs is encouraging and invites the speculation about how many other metadata elements might be promoted to either geodatabase tables or XML boilerplate templates. For example, a ProcessStep table, which a GIS author might maintain as part of a data management and history plan, could easily be parsed into the Lineage element. Likewise, a personnel or author database which might already exist for other purposes within an agency, could be queried for transcription to Citation elements.

An additional advantage of storing these kinds of data in database tables is that they are more readily viewed within the same table view context as the rest of the GIS data and could even be used in joins and relationship classes. The same data in a separate XML file, even if parsed into a more human-readable form or viewed in the Item Description tab of ArcCatalog, remains outside of the analysis space of the GIS and, therefore, of limited functionality.

## REFERENCES

Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C., <http://www.fgdc.gov/metadata/csdgm>

NCGMP (USGS National Cooperative Geologic Mapping Program), 2010, NCGMP09—Draft standard format for digital publication of geologic maps, version 1.1, in Soller, D.R., ed., Digital Mapping Techniques ‘09—Workshop Proceedings: U.S. Geological Survey Open-file Report 2010–1335, p. 93–146, [http://pubs.usgs.gov/of/2010/1335/pdf/usgs\\_of2010-1335\\_NCGMP09.pdf](http://pubs.usgs.gov/of/2010/1335/pdf/usgs_of2010-1335_NCGMP09.pdf)

Schweitzer, P. N., 1995a, MP: A compiler for formal metadata: U.S. Geological Survey, Reston, Virginia. <http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>

Schweitzer, P. N., 1995b, Xtme and Tkme: Editors for formal metadata: U.S. Geological Survey, Reston, Virginia, <http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html>

## FIGURE CAPTIONS:

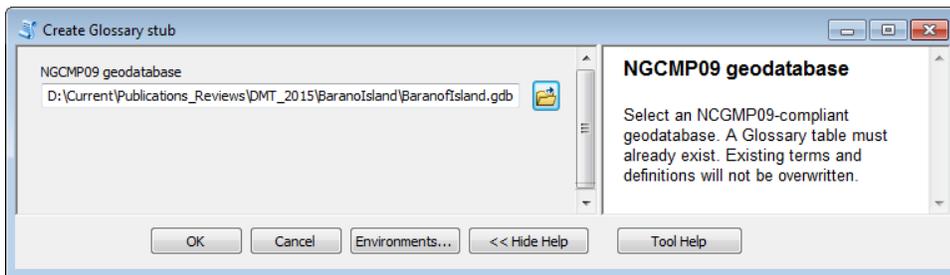


Figure 1. The Create Glossary Stub tool adds Terms to the Glossary table from the controlled fields in all tables in an NCGMP09 database. Use this as a starting point to find all the terms in your database that require defining in the Glossary.

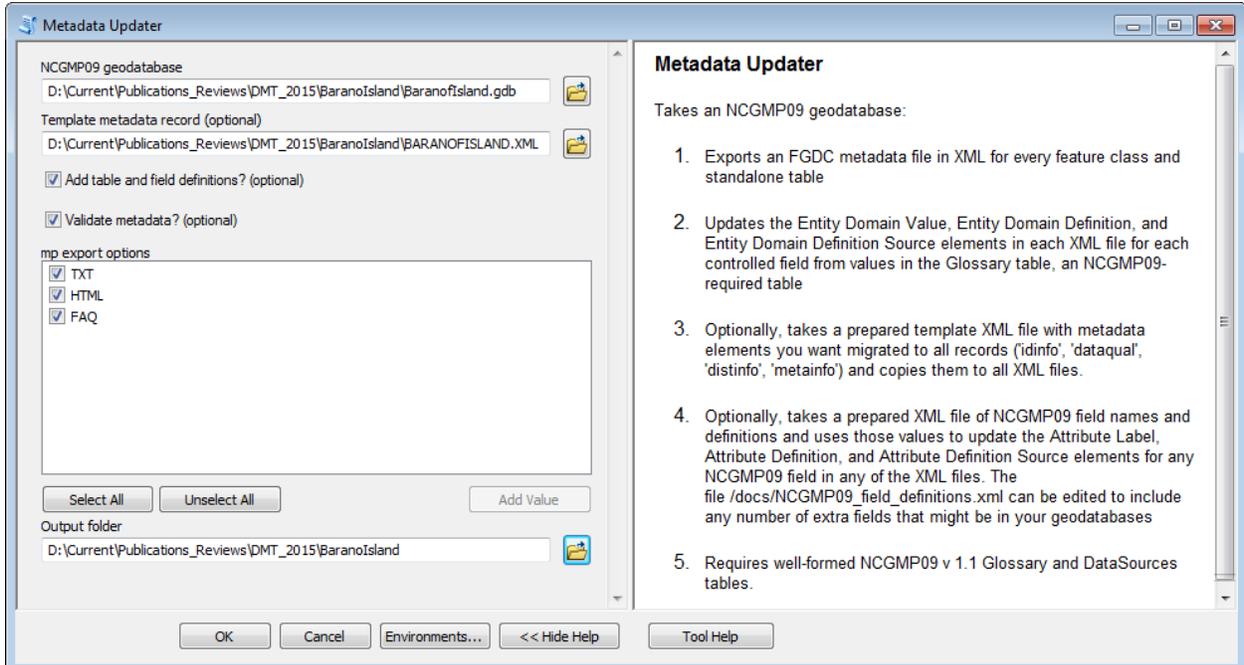


Figure 2. Screenshot of the NCGMP09 Metadata Updater geoprocessing tool with examples of how the parameter controls are filled out.

## The Goal: write as few lines of metadata as possible

FGDC metadata requires the elements shown below in blue.

### Geologic map of Baranof Island, southeastern Alaska

Metadata also available as

#### Metadata:

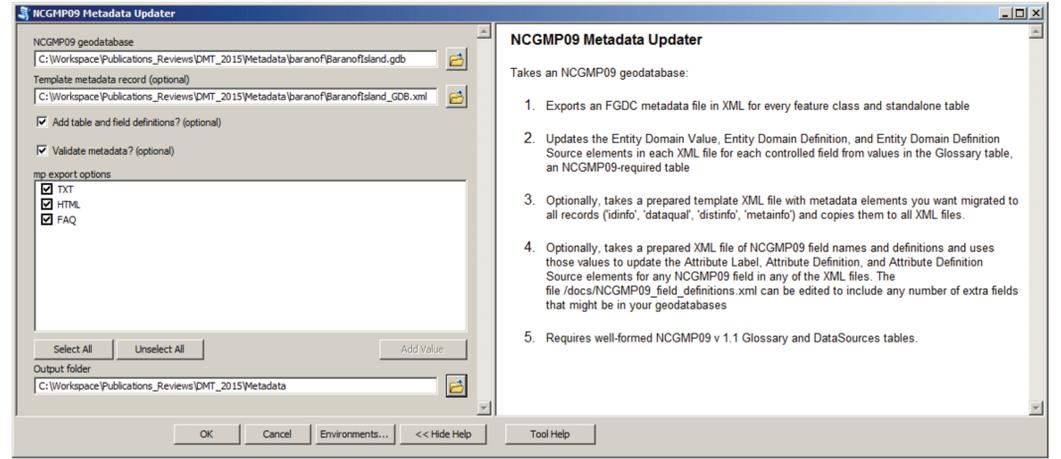
- Identification\_Information
- Data\_Quality\_Information
- Spatial\_Data\_Organization\_Information
- Spatial\_Reference\_Information
- Entry\_and\_Attribute\_Information
- Distribution\_Information
- Metadata\_Reference\_Information

In this workflow the content comes from:

- Template XML
- Template XML
- ArcGIS
- ArcGIS
- DescriptionOfMapUnits, Glossary, and DataSources tables
- Template XML
- Template XML

The Identification\_ Data\_Quality\_ Distribution\_ and Metadata\_Information elements can be re-used for multiple data objects within the same geodatabase. In the preparation of the DescriptionOfMapUnits, Glossary, and DataSources tables for a NCGMP09-compliant geodatabase, the compiler will have already collected information that can go into the Entry\_and\_Attribute\_Information (<eadaata>) element. But copying and pasting the content of these elements in the ArcCatalog metadata editor is tedious at best. While alternatives (Tkme, USGS Metadata Wizard, EPA Metadata Editor) allow for the insertion of stored elements and workflows for creating <eadaata> elements, they are still form-based, slower than necessary, and do not allow for batch processing. My solution is to export an FGDC-compliant XML metadata file out of ArcCatalog after the Spatial\_Data\_Organization\_ and Spatial\_Reference\_Information elements have been updated and then manipulate it through simple 3rd party programs and python scripting.

## Result: NCGMP09 Metadata Updater geoprocessing tool

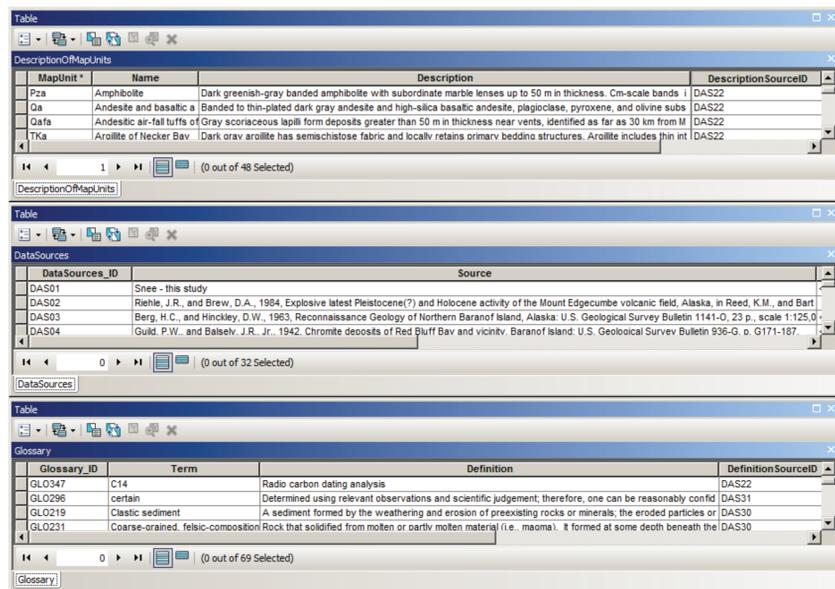


### NCGMP09 Metadata Updater

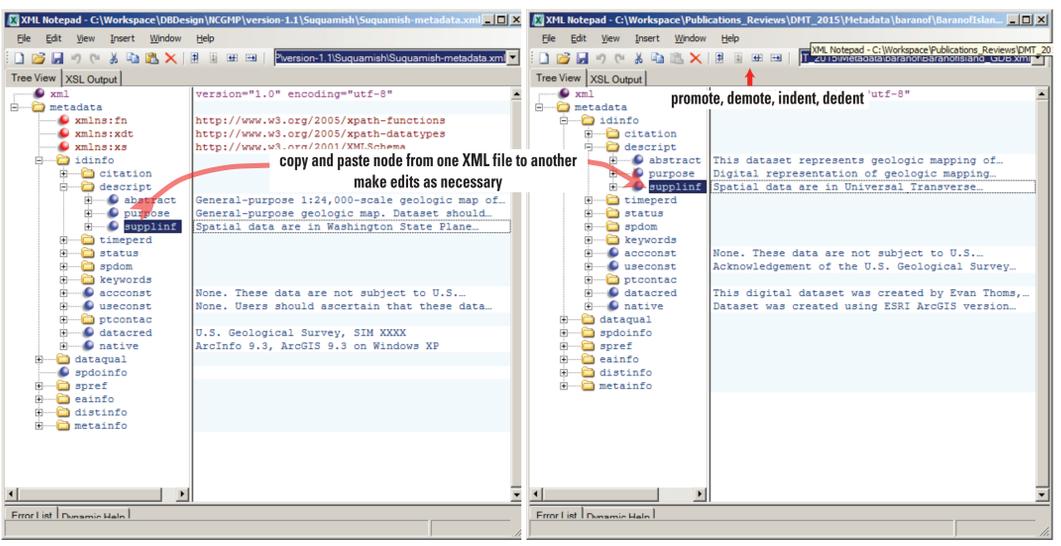
Takes an NCGMP09 geodatabase:

1. Exports an FGDC metadata file in XML for every feature class and standalone table
2. Updates the Entity Domain Value, Entity Domain Definition, and Entity Domain Definition Source elements in each XML file for each controlled field from values in the Glossary table, an NCGMP09-required table
3. Optionally, takes a prepared template XML file with metadata elements you want migrated to all records ('dinfo', 'dataqual', 'distinfo', 'metainfo') and copies them to all XML files.
4. Optionally, takes a prepared XML file of NCGMP09 field names and definitions and uses those values to update the Attribute Label, Attribute Definition, and Attribute Definition Source elements for any NCGMP09 field in any of the XML files. The file /docs/NCMP09\_field\_definitions.xml can be edited to include any number of extra fields that might be in your geodatabases
5. Requires well-formed NCGMP09 v 1.1 Glossary and DataSources tables.

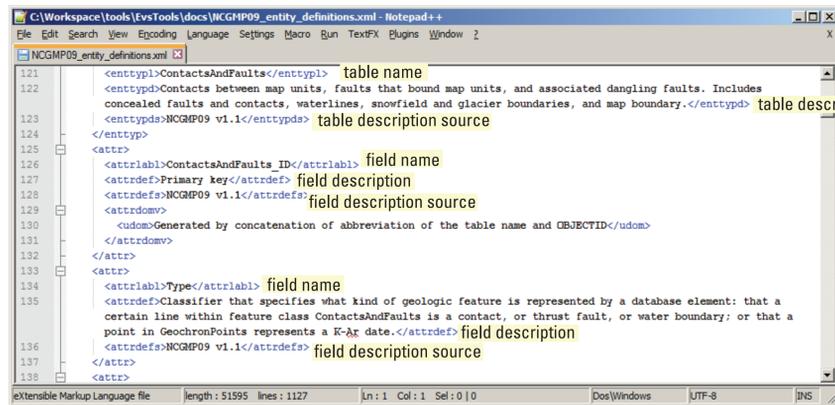
## Step 1- prepare DescriptionOfMapUnits, Glossary, and DataSources tables



## Step 2- prepare geodatabase-level metadata template by 1) exporting XML from geodatabase in ArcCatalog, 2) pasting in elements (if necessary) from a previously used file using XML Notepad or Tkme, 3) editing content using XML Notepad, Tkme, or Notepad++. Keep the metadata in XML format, and 4) validate with mp.exe.



## Step 3- prepare NCGMP09\_entity\_definitions.xml or other table and field description XML file:



## Step 4- Run NCGMP09 Metadata Updater:

For every feature class attribute table and stand alone table in the geodatabase:

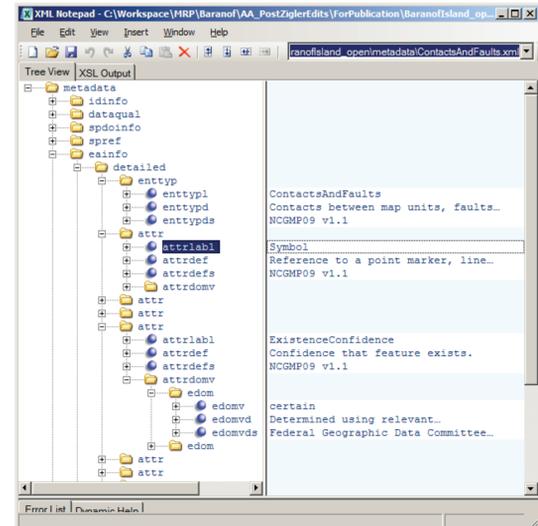
- Export an FGDC-compliant XML metadata file
- Copy elements from the template XML, (idinfo, distinfo, metainfo, etc.) to the XML file.
- For every field in the table that is also in a list of NCGMP09-controlled fields:
  - Write a complete Entity\_Type element based on values in NCGMP09\_entity\_definitions.xml
  - Make a list of the unique values in the field,
  - For every value in that list:
    - Find that value in the Glossary (or DescriptionOfMapUnits in the case of MapUnits)
    - Match up DefinitionSourceID with DataSourceID in DataSources and store the Source
    - Write out an Enumerated\_Domain element based on the Term, Definition, and Source
    - Validate and export the XML in any of the chosen formats (plain text, HTML, or HTML FAQ)

For the future:

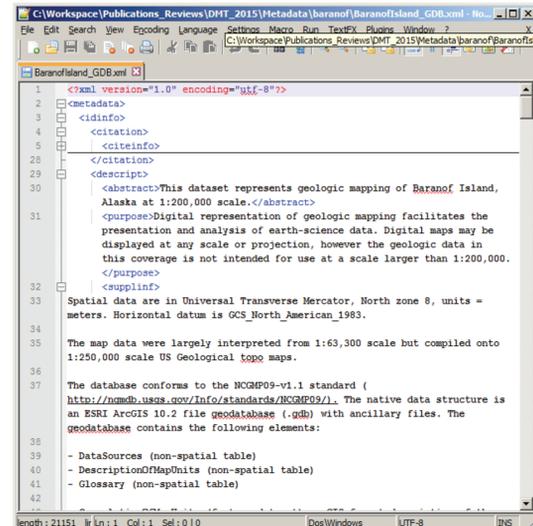
- Import the XML files back in to the geodatabase
- Complete NCGMP09\_entity\_definitions.xml for all required and optional tables and fields. Remember, you can add your own!
- Write in process steps based on entries in a ChangeLog table.
- Store as much metadata information in re-usable XML files or geodatabase tables because migrating that information into a publishable metadata record is FAST with python.

## Software:

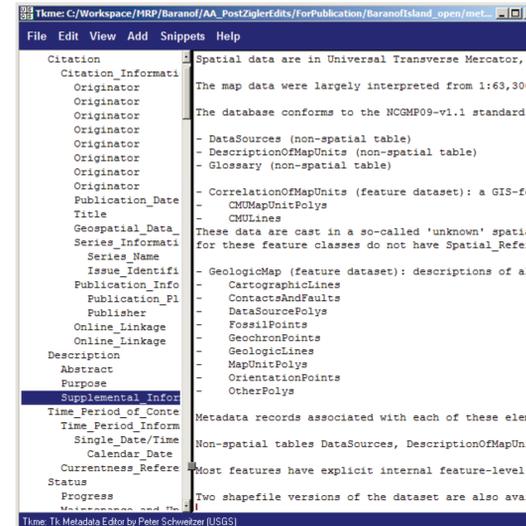
XML Notepad  
<https://xmlnotepad.codeplex.com/>



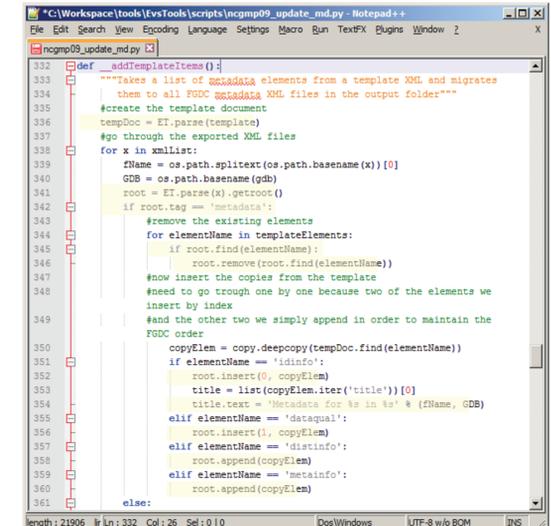
Notepad++  
<http://notepad-plus-plus.org/>



Tkme.exe metadata editor  
<http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html>



ElementTree python module  
 Part of Standard Library since Python 2.5



ArcGIS and ArcPy to export metadata and to access properties of the geodatabase. Metadata Parser (mp.exe) to validate XML files and translate them into other formats (<http://geology.usgs.gov/tools/metadata/>)

**Disclaimer:** Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.