**QUICK AND DIRTY CONVERSION OF ARCGIS VS 9.3 FGDC METADATA FILE TO EML VS 2.1 – DRAFT**

**Leanne Lestak, 7/16/2013**

***Introduction***

This is a document in progress. It contains a manual method that could be automated. This method builds on Teresa Valentine’s *Best Practices for documenting geospatial data, version 5.0 3/8/2012*. I would suggest using both documents, though I have added things from Teresa’s document to this one for ease of use.

If you get stuck and need EML assistance Margaret O’Brien and Sven Bohm are getting paid by the LTER to assist with EML questions, problems, enhancements, etc. Mark Servilla and Duane Costa at the LTER Network office can help with catalog problems/questions. Contact info is here:

Margaret O’Brien mob@msi.ucsb.edu

Sven Bohm bohms@msu.edu

Mark Servilla servilla@lternet.edu

Duane Costa dcosta@lternet.edu

It seems that ArcGIS FGDC vs 9.3 XML metadata is easier to convert than arcgis vs 10.1 XML metadata. That said use a discerning eye to compare input FGDC metadata with output EML to make sure everything is converting properly.

Start with the .xml file produced by ArcGIS vs 9.3 using the FGDC classic style sheet. Use an XML editor like Oxygen to edit XML files.

***Oxygen***

Download Oxygen, instructions are here:

http://im.lternet.edu/node/987

***Customize Stylesheet***

This section is copied from Valentine (2012).

Modify the arc2esri\_ver4\_2.zip stylesheet for your site. Available here: (<http://im.lternet.edu/node/243/release>) look for the latest version:

* 1. If you don’t have your own site data policies, the LTER Network Defaults will be used to populate that section of your EML document. Policies are here in stylesheet: You may want to enter here ...
  2. Search for “AND” in the stylesheet, and adjust for your site.
  3. Change package ID root to reflect your site:

<xsl:choose>  
<xsl:when test="metadata/packageId">  
<!-- an ad-hoc packageID was added a posteriori to the original ESRI record-->  
<xsl:value-of select="metadata/packageId"/>  
</xsl:when>  
<xsl:otherwise>  
<xsl:value-of select="'knb-lter-nwt.'"/>  
</xsl:otherwise>  
</xsl:choose>

***Edit ArcGIS vs 9.3 XML Metadata File***

First create a copy of the arcgis xml metadata file.

Then edit this copy as this will be used to create the EML, the original ISO or FGDC metadata file must be kept intact with the data.

Use Oxygen to delete these lines (we will be using the liability statement in the stylesheet). If the data is from another source and we're just sharing it, then keep all below, it will overwrite what's in the stylesheet.

*Delete:*

<accconst>None</accconst>

<useconst>Dataset credit required.... .</useconst>

<distliab>The Institute of Arctic and Alpine Research.....</distliab>

*Replace:*

Replace spatial reference with this (if it’s UTM zone 13, nad83). Spatial reference tags must be similiar to those below, the ones in FGDC vs 9.3 don’t work:

<spref>

<horizsys>

<planar>

<mapproj>

<mapprojn>Transverse Mercator</mapprojn>

<transmer>

<sfctrmer>0.9996</sfctrmer>

<longcm>-105.0</longcm>

<latprjo>0.0</latprjo>

<feast>500000.0</feast>

<fnorth>0.0</fnorth>

</transmer>

</mapproj>

<planci>

<plance>coordinate pair</plance>

<coordrep>

<absres>0.8</absres> **PUT PROPER VALUE IN FROM EACH DATASET**

<ordres>0.8</ordres> **PUT PROPER VALUE IN FROM EACH DATASET**

</coordrep>

<plandu />

</planci>

</planar>

<geodetic>

<horizdn>D North American 1983</horizdn>

<ellips>GRS 1980</ellips>

<semiaxis>6378137.0</semiaxis>

<denflat>298.257222101</denflat>

</geodetic>

</horizsys>

</spref>

*For entries w/ NWT LTER Data Manager change to this, or appropriate site data manager:*

<cntper>Niwot Ridge LTER Data Manager</cntper>

<cntpos>Niwot Ridge LTER Data Manager</cntpos>

<cntorg>University of Colorado, Institute of Arctic and Alpine Research (INSTAAR), Niwot Ridge LTER</cntorg>

</cntperp>

<cntaddr>

<addrtype>mailing and physical address</addrtype>

<address>1560 30th Street</address>

<city>Boulder</city>

<state>CO</state>

<postal>80309-0450</postal>

<country>USA</country>

</cntaddr>

<cntvoice>303-492-2594</cntvoice>

<cntfax>303-492-6388</cntfax>

<cntemail>lternwt@colorado.edu</cntemail>

*Add this for satellite rasters, no need to add for vectors:*

After </spref>

<eainfo>

<detailed>

<enttyp>

<enttypl>nwt\_ortho\_1972.tif</enttypl> **PUT PROPER VALUE IN FROM EACH DATASET**

<enttypd>raster data set</enttypd>

<enttypds>dataset</enttypds>

</enttyp>

<attr>

<attrlabl>value</attrlabl>

<attrdef>Digital number (DN) which records the intensity of radiation (or amplitude) falling on each pixel.</attrdef>

<attrdefs>ESRI</attrdefs>

<attrmfrq>011</attrmfrq>

</attr>

</detailed>

</eainfo>

*Add this for image photo rasters, no need to add for vectors:*

After </spref>

<eainfo>

<detailed>

<enttyp>

<enttypl>nwt\_ortho\_1972.tif</enttypl> **PUT PROPER VALUE IN FROM EACH DATASET**

<enttypd>raster data set</enttypd>

<enttypds>dataset</enttypds>

</enttyp>

<attr>

<attrlabl>value</attrlabl>

<attrdef>Saturation measured in 8 bits, between 0-255, where 0 is no saturation and 255 is full saturation.</attrdef>

<attrdefs>ESRI</attrdefs>

<attrmfrq>011</attrmfrq>

</attr>

</detailed>

</eainfo>

***For Raster Data:***

From Valentine, 2012.

“Adding the entity information: entity and attribute detail. If you have a VAT or an RAT (value or raster attribute table), you would add them in here as additional entity and attribute details.”

***Only need this one time in metadata file, put your zip package URL here:***

<networkr>ftp://niwotftp.colorado.edu/pub/Spatial/packages/nwt7031.zip</networkr>

Save XML file.

***Use XALAN (DOS program) to Create EML vs 2.1 File***

Grab a combined xalan/xerces package, they must be compatible versions. I used XALANCPKG-11-31-VC90 and XERCESCPKG-31-VC90. You can get info here:

http://xalan.apache.org/index.html

Install on your PC. Copy xerces .dll binary files to xalan bin directory.

Then copy edited vs 9.3 metadata file (file you edited above) to XALAN install bin directory, and make sure customized site stylesheet in same directory:

C:// ... XALANCPKG-11-31-VC90\bin>

Type this command at DOS prompt:

C://... XALANCPKG-11-31-VC90\bin>xalan -o <output\_file>.xml <arcGIS vs 9.3 edited file>.xml <customized site stylesheet>.xsl

***Check Output EML vs 2.1 XML from Xalan Against Original ArcGIS vs 9.3 XML Metadata File***

There are some things repeated many times. There will be lots of stuff missing. It really doesn't matter as the tags that are needed for the LTER Data portal should be preserved.

*Data credit will stay as is, should look like this in EML vs 2.1 XML:*

<section>

<title>Data Credit</title>

<para>U.S. Geological Survey (USGS), Denver Regional Council of Governments (DRCOG), Bohannan Huston Inc. (BHI), Boulder Creek Critical Zone Observatory (CZO), the University of Colorado, Institute of Arctic and Alpine Research (INSTAAR) and Niwot Ridge Long-Term Ecological Research (LTER) project.</para>

</section>

*Make sure entities are like this for vector:*

<spatialVector id="nwt\_project\_extent.zip">

<alternateIdentifier>nwt\_project\_extent</alternateIdentifier>

<entityName>nwt\_project\_extent.zip</entityName>

<entityDescription>This vector shapefile is a polygon shapefile....</entityDescription>

<physical>

<objectName>nwt\_project\_extent.zip</objectName>

<dataFormat>

<externallyDefinedFormat>

<formatName>ESRI Shapefile</formatName>

</externallyDefinedFormat>

</dataFormat>

<distribution>

<online>

<url>ftp://niwotftp.colorado.edu/pub/Spatial/packages/nwt9995.zip</url>

</online>

</distribution>

</physical>

**OR**

*Make sure entities are like this for raster:*

<spatialRaster>

<alternateIdentifier>nwt\_ortho\_1972.tif</alternateIdentifier>

<entityName>nwt\_ortho\_1972.tif</entityName>

<entityDescription>This image is a mosaic of orthorectified aerial photography .....</entityDescription>

<physical>

<objectName>nwt\_ortho\_1972.zip</objectName>

<dataFormat>

<externallyDefinedFormat>

<formatName>geoTIF file</formatName>

</externallyDefinedFormat>

***Edit EML vs 2.1 XML File***

Use Oxygen to edit EML vs 2.1 XML file.

*Insert package ID, do a find, or look for it on the first line:*

From: packageId="knb-lter-nwt."

To: packageId="knb-lter-nwt.703.1" *Add metacat number, unique to each dataset, and version number, start with “1”.*

*Change creator surname, parse out to*:

givenName

surName

role

organizationName

*Should look like this:*

<creator>

<individualName>

<givenName>Eric</givenName>

<surName>Parrish</surName>

</individualName>

<address>

<deliveryPoint>University of Colorado, Institute of Arctic and Alpine Research (INSTAAR), UCB 450</deliveryPoint>

<city>Boulder</city>

<administrativeArea>CO</administrativeArea>

<country>USA</country>

</address>

<electronicMailAddress>Eric.Parrish@colorado.edu</electronicMailAddress>

</creator>

*Replace these lines*:

<alternateIdentifier>Please cite as: M ... </alternateIdentifier>

<entityName>Please cite as: M ... </alternateIdentifier>

<objectName>Refer to Title of dataset</objectName>

*With these lines, inserting proper dataset name:*

<alternateIdentifier>nwt\_ortho\_1972.tif</alternateIdentifier> <entityName>nwt\_ortho\_1972.tif</entityName>

<objectName>nwt\_ortho\_1972.zip</objectName>

*For rasters search for:*

<numberOfBands>unknown</numberOfBands>, and add number of bands.

***Additions to EML***

It’s good to include a text version of the FGDC metadata file, which will show up in the LTER NIS portal metadata as a separate download. A text version (use FGDC CSDGM (TXT) stylesheet) can be exported from ArcCatalog vs 9.3 do this after final edits are made below.

*For vector (insert after the tag </spatialVector>):*

<otherEntity>

<entityName>Shapefile\_FGDC\_metadata</entityName>

<entityDescription>Shapefile metadata in text format</entityDescription>

<physical>

<objectName>nwt7391\_FGDC\_metadata.txt</objectName>

<dataFormat>

<externallyDefinedFormat>

<formatName>Natural language text</formatName>

</externallyDefinedFormat>

</dataFormat>

<distribution>

<online>

<url>ftp://niwotftp.colorado.edu/pub/Spatial/packages/nwt7391\_FGDC\_metadata.txt</url>

</online>

</distribution>

</physical>

<entityType>Shapefile metadata in text format</entityType>

</otherEntity>

*For raster (insert after the tag </spatialRaster>):*

<otherEntity>

<entityName>geoTIF\_FGDC\_metadata</entityName>

<entityDescription>GeoTIF metadata in text format</entityDescription>

<physical>

<objectName>nwt7001\_FGDC\_metadata.txt</objectName>

<dataFormat>

<externallyDefinedFormat>

<formatName>Natural language text</formatName>

</externallyDefinedFormat>

</dataFormat>

<distribution>

<online>

<url>ftp://niwotftp.colorado.edu/pub/Spatial/packages/nwt7001\_FGDC\_metadata.txt</url>

</online>

</distribution>

</physical>

<entityType>GeoTIF metadata in text format</entityType>

</otherEntity>

*“Save as”* *to final EML metadata file name:*

knb-lter-nwt.**700.1**.xml

***Check Output in KNB Parser***

Run final EML vs 2.1 file thru EcoInformatics Parser to check for EML and XML errors:

http://knb.ecoinformatics.org/emlparser/

***Create Zipped Package***

Put all the raster or vector files, including ArcGIS metadata into one directory.

*Update ArcGIS vs 9.3 FGDC metadata:*

Vs 9.3 metadata that goes with spatial layer is now out of date, it has to be updated, can do that in Oxygen.

*Update URL to package zip file:*

Can be in multiple places, but only one URL for 1 package ID per EML file.

<networkr>ftp://niwotftp.colorado.edu/pub/Spatial/packages/nwt**7391**.zip</networkr>

Save the file.

*Export final vs 9.3 metadata file****:***

Using arccatalog vs 9.3, view tif or shapefile, choose FGDC classic stylesheet, export out to FGDC CSDGM (TXT).

Create readme text file to put in directory if desired.

*Zip up directory:*

Zip up directory, then rename zip files with catalog number. For example, directory name would be nwt\_ortho\_1938 and zip file would be named nwt7001.zip. Name it the same as URL link in EML file: nwt7391.zip.

*Zip up:*

*shapefile:*

.dbf, .prj, .sbn, .sbx, .shp, .shp.xml, .shx

*OR geotif:*

.tif, .tif.xml

***Harvest XML***

A harvest list XML can be used to do multiple uploads at the same time. Format is:

<document>

<docid>

<scope>knb-lter-nwt</scope>

<identifier>739</identifier>

<revision>1</revision>

</docid>

<documentType>eml://ecoinformatics.org/eml-2.0.1</documentType>

<documentURL>ftp://niwotftp.colorado.edu/pub/Spatial/packages/knb-lter-nwt.739.1.xml</documentURL>

</document>

***Put Everything in Packages Area on Server***

nwt\_harvestList.xml

knb-lter-nwt.**739.1**.xml

nwt**7391**.zip

***PASTA Staging Area (portal-s.lternet.edu vs portal for actual site)***

First evaluate the EML, click "tools", then "Evaluate Data Package":

https://portal-s.lternet.edu/nis/dataPackageEvaluate.jsp

If there are no errors then upload the package (either using the harvest XML or just a single EML) to the PASTA staging area, you need a login for this, which I requested from:

Tech-Support <tech-support@lternet.edu>

*Then a report is produced here:*

Check for errors, look at metadata entry, check that data downloads properly.

https://portal-s.lternet.edu/nis/harvestReport.jsp

***Post Data to PASTA Distribution Portal (portal.lternet.edu)***

Once all errors have been corrected and everything looks correct in the PASTA staging area, then upload the package (either using the harvest XML or just a single EML) to the PASTA area, you need a login for this, which I requested from:

Tech-Support <tech-support@lternet.edu>

*Then a report is produced here:*

Check for errors, look at metadata entry, check that data links are correct and that data downloads properly.

https://portal.lternet.edu/nis/harvestReport.jsp

***References***

Theresa Valentine, 2012, *Best Practices for documenting geospatial data, version 5.0 3/8/2012,* http://im.lternet.edu/project/GIS\_document.