

EML - Bug #485

eml-physical changes needed

05/01/2002 04:57 PM - Matt Jones

Status:	Resolved	Start date:	05/01/2002
Priority:	Immediate	Due date:	
Assignee:	Dan Higgins	% Done:	0%
Category:	eml - general bugs	Estimated time:	0.00 hour
Target version:	EML2.0.0rc1	Spent time:	0.00 hour
Bugzilla-Id:	485		

Description

Changes as decided upon at the Sevilleta EML meeting, April 24-25, 2002:
Responsible: Dan

- 1) add version and citation of format definition
- 2) add ability to describe BIP and BIL formats for binary raster data -- see the IPW header format for the info needed
- 3) rearrange for better control of required elements when using fixed vs. variable formats. Do this by creating "fixed" and "delimited" elements with proper content models.
- 4) add "objectName" element to contain the filename or other name of the physical object
- 5) add field for pointer for which connection to use to get this physical object (using "objectName"). Question as to how the semantics of that combo work -- how does one add an object name together with connection info for arbitrary connection types?

History

#1 - 05/08/2002 01:45 PM - Dan Higgins

with regard to item 20 ability to describe BIP and BIL formats for raster data;

there is a white paper on the ESRI site that describes the header information used for these types of files ("Extendable Image Formats for ArcView GIS 3.1 and 3.2").

The header information is in the form of keywords/values. 14 keywords are defined as follows:

nrows - The number of rows in the image. Rows are parallel to the x-axis of the map coordinate system. There is no default.

ncols - The number of columns in the image. Columns are parallel to the y-axis of the map coordinate system. There is no default.

nbands - The number of spectral bands in the image. The default is 1.

nbits - The number of bits per pixel per band. Acceptable values are 1, 4, 8, 16, and 32. The default value is eight bits per pixel per band. For a true color image with three bands (R, G, B) stored using eight bits for each pixel in each band, nbits equals eight and nbands equals three, for a total of twenty-four bits per pixel. For an image with nbits equal to one, nbands must also equal one.

byteorder - The byte order in which image pixel values are stored. The byte order is important for sixteen-bit images, with two bytes per pixel. Acceptable values are

I - Intel byte order (Silicon Graphics, DEC Alpha, PC) Also known as littleendian.

M - Motorola byte order (Sun, HP, etc.) Also known as big-endian.

The default byte order is the same as that of the host machine executing the software.

layout - The organization of the bands in the image file. Acceptable values are

bil - Band interleaved by line.

bip - Band interleaved by pixel.

bsq - Band sequential.

The default layout is bil.

skipbytes - The number of bytes of data in the image file to skip in order to reach the start of the image data. This keyword allows you to bypass any existing image header information in the file. The default value is zero bytes.

ulxmap - The x-axis map coordinate of the center of the upper-left pixel. If you specify this parameter, set ulymap, too, otherwise a default value is used.

ulymap - The y-axis map coordinate of the center of the upper-left pixel. If this parameter is specified, ulxmap must also be set, otherwise a default value is used.

xdim - The x-dimension of a pixel in map units. If this parameter is specified, ydim must also be set, otherwise a default value is used.

ydim - The y-dimension of a pixel in map units. If this parameter is specified, xdim must also be set, otherwise a default value is used.

bandrowbytes - The number of bytes per band per row. This must be an integer. This keyword is used only with BIL files when there are extra bits at the end of each band within a row that must be skipped.

totalrowbytes - The total number of bytes of data per row. Use totalrowbytes when there are extra trailing bits at the end of each row.

bandgapbytes - The number of bytes between bands in a BSQ format image. The default is zero.

#2 - 05/16/2002 09:44 AM - Dan Higgins

proposed changes to eml-physical-2.0.0beta8; partially completed (16May2002)

1) 'version' and 'citation' attributes have been added to 'format' element. It was assumed that the 'citation' is a simple reference rather than the full 'citation' element that used elsewhere.

2) a proposed set of elements for describing binary raster data is included. All are included as children of a new element called 'BinaryRasterInfo'

3) No changes have been made in handling 'fixed' vs 'delimited' field Delimiters. I am not sure what to do here. The current system seems to work for me.

4) 'objectName' element has been added. In my mind this is usually simply a file name that can be restored (if desired) when a object is returned

5) field for pointer to connection - Don't know how to handle

#3 - 05/16/2002 09:49 AM - Dan Higgins

Current (16 May 2002) status is reflected in the attached proposed eml-physical changes document. I have received no comments on proposed raster image parameters and need some feedback of proposed changes before they can be completed. (Dan Higgins)

#4 - 05/16/2002 11:07 AM - Matt Jones

Dan,

1) I think 'citation' should be the full citation reference (type cit:LitCitation).

2) I'll review BIP/BIL stuff separately.

3) The intention of rearranging the delimiters was to make it clear when each was required. I think we still need to make these changes.

4) Good.

5) This is tightly bound to the resolution of the "distribution" discussion for eml-resource. What this field looks like, and even whether one is needed, is determined by whether the top level distribution element represents a generalized connection or a connection to a particular resource.

You should feel free to check this into CVS when you are ready, even if it is not complete. The only reason Owen and Dan are using Bugzilla attachments is because they don't have write access to the eml module, which is a side-effect of moving to the ecoinfo cvs server.

#5 - 05/16/2002 11:42 AM - Matt Jones

About raster metadata -- looks good. A few comments:

- 1) nrows & ncols should be required. The rest of the fields should be optional, with the default values explicitly encoded in the schema.
- 2) you are missing all of the documentation tags. Please add them as I have described in other bugs.
- 3) use camel caps for element names as described in other bugs. Elements should be initially lowercase. Types should be initially uppercase. So "BinaryRasterInfo" should be "binaryRasterInfo"
- 4) I think we need to reorganize the placement of the binaryRasterInfo element. Right now it is possible to provide a field delimiter and raster info, which is inappropriate. Maybe we should cluster these into a top-level choice. What happens if we want to add other physical descriptors later? Right now we support various text character encodings for tabular data, and binary raster data. What about text-encoded raster data? I think we need to figure out how physical can be extensible like entity is. Not sure how this should happen.
- 5) Could you also review the Image Processing Workbench (IPW) to make sure that we accommodate everything it can handle as well. IPW allows raster images to be viewed in standard programs like xv, and is well-used in the remote sensing community. IPW information can be found at: <http://www.icess.ucsb.edu/~ipw2/> Look in particular at the "mkbih" command. Thanks.

#6 - 05/17/2002 11:06 AM - Peter McCartney

The raster parsing info looks good. I have some notes that we put together based in the Erdas import tools that i will check this against to see if theres anything else. offhand i dont see where we indicate the origin or whether to read rowfirst or columnfirst.

I hope we are still going to see some struture in this module so that we can start setting elements to required when they should be. I think there should be a primary division at the top between the description of the physical object, description of its format, and the reference to the connection it is found at.

the physical object description would include its name, size, owner?, etc.

the connection is merely a pointer by name, or idref or whatever we decide.

the format section needs to be further subdivided into at least three choices thus far: a named format (with optional version and citation), an ASCII format description (im willing to try working with your mixed model for delimited/fixedlength), and the binary raster. Others are likely to be defined as time goes on.

Ill leave you with an ever further leap to say that i would like to see the object description repeat within a single physical module and attach an optional extent or coverage description to each. this allows you to deal easily with multiple files produced by cutting a single data entity into tiles or series PURELY for the purposes of storage/transport considerations and it is expected that it would be reassembled prior to making use of it. the coverage or extent module would provide the guidelines as to how to reassemble the pieces. In our ASU use of this, it is assumed in this case that the coordinates provided are in the projected units as they have to be quite precise in order to properly put the images together.

I will attach a copy of our earlier draft just so you can see what i mean.

#7 - 05/17/2002 11:09 AM - Peter McCartney

This attachment is NOT a proposed draft, but merely included to illustrate the upper level organization i suggest in the previous comment on this bug.

#8 - 05/21/2002 04:14 PM - Dan Higgins

revised version checked into cvs. This new version has considerably more 'structure' than previous version (borrowed from asu draft). 5/21/2002
Dan Higgins

#9 - 06/10/2002 10:21 PM - Chris Jones

The distribution element underneath the individual entities such as dataTable is redundant since Matt included distribution in the ResourceGroup after the notes from the Sevilleta Meeting. I suggest we remove this and keep it in Resource to minimize confusion as to where it goes.

#10 - 06/10/2002 10:33 PM - Matt Jones

We had agreed that distribution would go in resource, and in physical. I don't understand why you removed it. We need it there.

#11 - 06/13/2002 06:58 PM - Matt Jones

Changes completed, but need to check on the "orientation" concept for binary image files that mccartney mentioned. Moving milestone for this one issue. The rest of the changes are done and in CVS.

#12 - 06/14/2002 10:37 AM - Peter McCartney

I just checked out the files and here are my lingering comments.

- a. I still suggest changing dataFormat. FormatName is only needed if you are NOT providing the parsing information inline. This structure is confusing because someone could enter ascii fixed info but also enter dbase under format.
- b. Distribution element repeats and contains a repeating choice. This doesn't make sense unless there's something about the inline element that might occur more than once per entity.
- c. Ascii fixed won't work as it is. Start column needs to repeat with field length and you need to add physical record information.
- d. drop genericBinary unless someone has a definition for it

#13 - 06/27/2002 01:52 PM - Owen Eddins

I'm passing the following comments from Tim Bergsma the data manager at Kellogg Biological Station in Michigan. He made them in a eml-dev email. I posting to bugzilla just to make sure they don't fall through the cracks.

11. It looks from my printout as though <distribution> is defined somewhat differently under <resourceGroup> vs. <physical>, i.e. no <inline> option.

#14 - 09/12/2002 01:47 PM - Matt Jones

Distribution has now been changed in resource and physical. They are essentially the same now (both include an inline element), but the resource DistributionType allows for an online/connectionDefinition to stand by itself, whereas the PhysicalDistributionType only allows connectionDefinition inside of connection (as in "online/connection/connectionDefinition"). All issues in this bug are resolved with these changes. FIXED.

#15 - 03/27/2013 02:14 PM - Redmine Admin

Original Bugzilla ID was 485

Files

eml-physicalDFH.xsd	27.2 KB	05/16/2002	Dan Higgins
physical.xsd	9.12 KB	05/17/2002	Peter McCartney