

Environment and organisms - Task #215

Establish file organization and possible 'tiling' scheme for Fused Global Terrain data set

04/18/2011 01:36 PM - Rick Reeves

Status:	Rejected	Start date:	04/18/2011
Priority:	High	Due date:	05/15/2011
Assignee:	Rick Reeves	% Done:	100%
Category:	Terrain	Estimated time:	4.00 hours
Target version:			
Activity type:	Coding/analysis		
Description			
<p>The size and scope of the global fused DEM terrain layer makes it impractical to distribute and use the layer (and possibly the other data layers) as a single physical file.</p> <p>Before generating the global data base,, we need to specify and implement a file naming and organization scheme that stores and serves the data layer(s) as collections of layer 'tiles',each covering a standard area of the Earth surface.</p> <p>Attributes that we need to specify:</p> <ul style="list-style-type: none">- Physical area covered by each file- File format (e.g., ASCII, GeoTIFF)- File naming convention- File 'grouping' convention (if any). Example:- By continent- By Latitude or Longitude band <p>As an example, consider the organization and naming convention of the CGIAR DEM data set: 872 mosaiced 5 degree * 5 degree tiles; with standard names; organized by Longitude-major order; Arc Map GIS/ ASCII or GeoTIFF file format; distributed with a shape file 'index'.</p>			

History

#1 - 04/22/2011 02:28 PM - Rick Reeves

- % Done changed from 0 to 10

I propose dividing the globe into four quadrants:
Northwest Hemisphere, Southwest Hemisphere, Northeast Hemisphere,
Southeast Hemisphere, and dividing/storing most data layers along and in these quadrants.

I would make a small modification to the standard Hemisphere definitions:
do not split up the continents that intersect the Prime Meridian and International Date Line, but keep the entire continent in the West/East (sub)
hemisphere in which most of the continent resides. Two best examples:
Great Britain and West African countries.

Reason for the subdivision: enable researchers to download the (large) portion of the globe in which their area of interest resides, without having
to download the entire global data set.

Something for the group to discuss and decide....

#2 - 05/12/2011 11:05 PM - Rick Reeves

- % Done changed from 10 to 100

I am using this scheme; have created the directory structure, and have constructed
the first prototype fused global terrain layer (1 KM / 30 arcsecond cell size) for
North America.

The shell scripts used to construct this quadrant will be modified (slightly) to
construct the remaining three quadrants.

#3 - 06/08/2011 04:44 PM - Rick Reeves

- Due date changed from 04/21/2011 to 05/15/2011

The complete archive of Digital Elevation Model data to be used for constructing the Fused Terrain Layer and derived data layers has been assembled on /_vulcan_. The root folder is:

/home/reeves/active_work/EandO

Subfolders are:

/asterGdem: ASTER GDEM data files, each covering 1 square degree at nominal 1 arcsecond spatial resolution. Obtained from NASA WIST data portal: <https://wist.echo.nasa.gov>, and the ASTER Japan data portal: <http://www.gdem.aster.ersdac.or.jp/>

/cgiaSrtrm: SRTM data files, each covering 5 square degrees at nominal 3 arcsecond spatial resolution. Obtained from the CGIAR data portal: <http://www.cgia-csi.org>

/CanadaDEM: Canadian DEM data files, with nominal 3 arc second spatial resolution, obtained from the Canadian GeoBase portal: <http://www.geobase.ca>

I have also established a folder hierarchy to contain the terrain data layer products to be produced for this project:

/home/reeves/active_work/OutProducts

To keep production organized, I have divided the globe into eight regions:

Western Hemisphere

Northern quadrant

Southern quadrant

Eastern Hemisphere

Northwestern region

Northern quadrant

Southern quadrant

Northeastern region

Northern quadrant

Southern quadrant

The corresponding folder hierarchy will contain all data layers produced using the incoming DEM data files contained within the active_work/EandO folder. The number of 'product subfolders' will grow as more derived products are generated:

/home/reeves/active_work/WestHemi

NorthWest

SouthWest

/home/reeves/active_work/EastHemi

NorthWestEast

data product 1 folder (e.g., 1 KM fused terrain layer)
data product 2 folder

..... (complete set of product folders TBA)

SouthWestEast

data product 1 folder

.....

NorthEastEast

data product 1 folder

.....

SouthEastEast

data product 1 folder

.....

A set of entries under Redmine Task [#207](#) (Produce global fused DEM Layer) describes the process of obtaining the baseline terrain data stored within /home/reeves/active_work/EandO

#4 - 01/12/2012 10:45 AM - Jim Regetz

- Status changed from New to Rejected

At some point, we'll certainly need to make sure end users can obtain data in convenient ways. However, I'm rejecting this task on the grounds that this issue is separate from our internal approach file organization, especially during the data processing and layer development stage.

For maximum flexibility, we should preferentially tend towards flatter directory hierarchies -- e.g. one directory containing all input ASTER GDEM tiles, without additional arbitrary sub-structure. If extra logical structure is useful at any point (e.g., doing some particular processing step region-by-region), if at all possible this should be implemented only within the analytical layer (i.e., in scripts) or possibly using VRTs specifically in the case of raster data, rather than reorganizing the file system itself.