Environment and organisms - Task #225

Organize and document SRTM input DEM files

05/17/2011 05:33 PM - Jim Regetz

Status:	In Progress		**Start date:**	05/17/2011
Priority:	Normal		**Due date:**	06/04/2011
Assignee:	Jim Regetz		**% Done:**	80%
Category:	Terrain		**Estimated time:**	40.00 hours
Target version:				
Activity type:	Infrastructure			

Description

[Updated ticket description on 04-Aug-2011 to focus specifically on SRTM]

SRTM DEM tiles and associated files have proliferated in jupiter:~organisms/, including:

- tiles for the same areas, but at different resolutions (1km, 90m, 30m?)
- two *different* versions at 90m?? (see more about this below)
- the same tiles stored in multiple raster formats
- multiple copies of tiles in different directories (presumably copied for testing/exploratory purposes?)
- assorted resampled and/or mosaicked outputs of the above

Beyond just the good housekeeping issue, it is especially worrisome that there appear to be two different versions of some SRTM data, with identical filenames but offset from each other by 1/2-pixel. Note differences in the corner coordinates below (and see here for clues):

```
$ gdalinfo ~organisms/CgiarSrtmAll/5_5x5_ascii/srtm_13_02.tif
...
Size is 6000, 6000
...
Upper Left  (-120.0000000,  55.0000000) (120d 0' 0.00"W, 55d 0' 0.00"N)
Lower Left  (-120.0000000,  50.0000000) (120d 0' 0.00"W, 50d 0' 0.00"N)
Upper Right (-115.0000000,  55.0000000) (115d 0' 0.00"W, 55d 0' 0.00"N)
Lower Right (-115.0000000,  50.0000000) (115d 0' 0.00"W, 50d 0' 0.00"N)
...

$ gdalinfo ~organisms/SRTM_90m_ASCII_v4.1/srtm_13_02.tif
...
Size is 6001, 6001
...
Upper Left  (-120.0004167,  55.0004171) (120d 0' 1.50"W, 55d 0' 1.50"N)
Lower Left  (-120.0004167,  49.9995838) (120d 0' 1.50"W, 49d59'58.50"N)
Upper Right (-114.9995833,  55.0004171) (114d59'58.50"W, 55d 0' 1.50"N)
Lower Right (-114.9995833,  49.9995838) (114d59'58.50"W, 49d59'58.50"N)
...
```

Tasks

- Figure out (if not already known) exactly what everything is, and where it came from
- Differentiate downloaded input tiles from processed files and other outputs
- Document everything clearly in READMEs stored along with the files on the server
- Delete any clearly obsolete/redundant/misleading files
- Move any marginally relevant (but not delete-worthy) files into directories with names that clearly indicate this status ('archive', etc)

History

#1 - 05/19/2011 10:26 AM - Rick Reeves

Yes, last week I created a new folder structure to contain the batch files and output files
for the next set of 'production' mosaiced files:

```
/data/project/rcr/OutProducts/
```

```
                                    /EastHemi/
                                        /NorthWest
                                        /SouthWest
```

Also, I have culled through the files in the ./rcr/AsterCgiarMerge and /ValidateBoundary folders
to remove old and obsolete files, and added a new folder. /ValidateBoundary/offsetCheck, to contain
new files generated during review of the boundary 'edge' issue.

#2 - 05/20/2011 01:20 PM - Jim Regetz

Re: Different CGIAR SRTM 3" source data

Rick and I have been discussing the two different sets of CGIAR SRTM rasters that are offset from each other by 1/2 pixel, as documented in the
issue description. Unfortunately, the following README files are all identical, so that doesn't help:

- ~organisms/SRTM_90m_ASCII_v4.1/readme.txt
- ~organisms/CgiarSrtmAll/readme.txt
- ~organisms/CgiarSrtmAll/5_5x5_ascii/readme.txt

I still want to see an authoritative statement of *where each of these sets of data came from* and exactly what they are. If we don't know that, we can't
use them as inputs.

Rick says that all of the work he has done thus far uses the tiles in the third of these directories (i.e., 5_5x5_ascii).

My sense is that the SRTM_90m_ASCII_v4.1 files can be treated as-is, but the 5_5x5_ascii files may need to be shifted by 1/2 pixel to accurately
position them. I base this on the fact that this [CGIAR SRTM FAQ entry about pixel placement](#) jives with what I see in SRTM_90m_ASCII_v4.1.
Moreover, I'll again highlight this [blog post](#) that describes the 1/2-pixel shift between different CGIAR releases. My inspection of the two different
versions of srtm_13_02.asc reveals that the northern row and eastern column have been excluded from the 5_5x5_ascii tiles compared with
SRTM_90m_ASCII_v4.1, but the values themselves are otherwise identical. This quick R code illustrates the pattern:

```
library(raster)

# read a tile extracted from SRTM_90m_ASCII_v4.1
v4.1 <- as.matrix(raster("srtm_13_02_SRTM_90m_ASCII_v4.1.asc"))
dim(v4.1)
## [1] 6001 6001

# read same tile extracted from 5_5x5_ascii
v5x5 <- as.matrix(raster("srtm_13_02_5_5x5_ascii.asc"))
dim(v5x5)
## [1] 6000 6000

# drop northern-most row and eastern-most column of the v4.1 file, then compare
identical(v4.1[-1, -6001], v5x5)
[1] TRUE
```

So if we believe the FAQ implications that the SRTM_90m_ASCII_v4.1 positioning is correct, then the proper correction for the 5_5x5_ascii data
would be to shift them 1.5" south and west --- i.e. subtract 1.5" from all stated extent coordinates. But taking this further, if the only difference between
these two sets of tiles is that the SRTM_90m_ASCII_v4.1 tiles already have this correction, but otherwise the corresponding pixel values are all
identical, then by all means we should just use the corrected data, especially because the 5_5x5_ascii version is evidently missing the northern-most
row of available SRTM data.

As another supporting piece of evidence, note the following about the contents of the zip files obtained from GCIAR in each case:

```
$ unzip -l ~organisms/CgiarSrtmAll/5_5x5_ascii/srtm_13_02.zip
...
     2479  2008-09-19 15:05   readme.txt
170529365  2008-09-19 17:50   srtm_13_02.asc
      656  2008-09-19 17:50   srtm_13_02.prj
...

$ unzip -l ~organisms/SRTM_90m_ASCII_v4.1/srtm_13_02.zip
...
     2479  2008-09-19 15:05   readme.txt
170580676  2008-11-24 16:40   srtm_13_02.asc
      656  2008-11-24 16:40   srtm_13_02.prj
...
```

Looks like the SRTM_90m_ASCII_v4.1 were produced more recently.

#3 - 06/03/2011 09:19 PM - Rick Reeves

- Due date set to 06/04/2011

- % Done changed from 0 to 80

- Estimated time changed from 2.00 h to 40.00 h

I accumulated the (SRTM and ASTER) DEM files over the past two months, as I began this project. The ~ 6000 files are loosely organized on /Jupiter under the /data/project/organisms directory. But the organization needed improvement prior to production of the terrain data layers. As I transferred the files to the /vulcan server in preparation for production, I made this improvements as I re-organized the files under /home/reeves/active_work/EandO.

As June 3, I have acquired all (873) of the CGIAR/SRTM files, and 96% of the ASTER GDEM files (4886 of 5068) required to build the global fused terrain layers. Over the next three days I expect to obtain the remaining 182 ASTER GDEM files from the Japanese ASTER data portal, which facilitates interactive downloads of individual ASTER image tiles.

#4 - 08/04/2011 11:25 AM - Jim Regetz

- Subject changed from Organize and document DEM files on jupiter to Organize and document SRTM input DEM files

- Description updated

- Status changed from New to In Progress

- Assignee deleted (Rick Reeves)

#5 - 01/16/2012 02:16 PM - Jim Regetz

- Assignee set to Jim Regetz

I removed the directory ~organisms/SRTM_1km_ASCII, which contained a global SRTM 1km ASCII grid obtained by Ming early in the project. We have no real need for this version of SRTM. Ming confirmed by email back on 21-Jun-2011 that it's fine to delete it. I'm nevertheless keeping the compressed RAR archive around (for now at ~organisms/DEM/SRTM_1km_ASCII.rar)

The directory also contained extra cruft that I removed, along with a BIL that Natalie had generated in Dec 2011 by reprojecting the SRTM 1km ASCII grid to MODIS sinusoidal and clipping out the Oregon case study; Natalie moved this output elsewhere (currently ~organisms/Oregon_SRTM), although eventually she should be able to aggregate, resample, and clip our fused global DEM to obtain a replacement 1km DEM for Oregon.

#6 - 01/17/2012 03:30 PM - Jim Regetz

Just committed a script (source:terrestrial/terrain/dem/srtm-check.R@176) that I wrote in May 2011 to assess our different SRTM holdings.

Note that the 5_5x5_ascii directory referenced by the script no longer exists, but the direct comparison of ~organisms/SRTM_90m_ASCII_v4.1/ to the CGIAR FTP site (ftp://srtm.csi.cgiar.org/SRTM_v41/SRTM_Data_ArcASCII/) is still useful, and should provide some clues about how (and maybe why?) contents of this directory differ from the directory ~organisms/DEM/cgiarSrtm/SRTM_90m_ASCII_4_1/ that Reeves created.

#7 - 01/18/2012 06:47 PM - Jim Regetz

The purpose of this update is to summarize differences between the following two directories, for the purposes of ultimately removing the first directory and (after necessary cleanup) retaining the second directory as our authoritative CGIAR SRTM global 90m DEM tile repository:

1. ~organisms/SRTM_90m_ASCII_v4.1/ascii-grids: Extracts from original zips downloaded by Ming, but with some later additions (and replacements?) by Reeves
2. ~organisms/DEM/cgiarSrtm/SRTM_90m_ASCII_4_1/: Created by Reeves, with many files (but clearly not all) presumably copied from the other directory

Both directories contained the expected 874 srtm_*_*.asc files, and the filenames matched exactly. MD5 checksums revealed that 862 were identical between the two directories, but 12 differed. Further inspection revealed that the difference in each case was due to the nodata value; those in the first directory all had the expected -9999 (and the original CGIAR timestamps), but those in the second directory had a different nodata value (and newer 2011-06-02 timestamps):

```
     filename nodata1 nodata2
srtm_44_22.asc   -9999  -32768
srtm_46_14.asc   -9999  -32768
srtm_47_22.asc   -9999  -32768
srtm_49_16.asc   -9999  -32768
srtm_50_22.asc   -9999  -32768
srtm_50_23.asc   -9999     255
srtm_51_14.asc   -9999  -32768
srtm_51_22.asc   -9999  -32768
srtm_51_23.asc   -9999  -32768
srtm_68_23.asc   -9999  -32768
srtm_70_23.asc   -9999  -32768
srtm_72_16.asc   -9999  -32768
```

The versions in the second directory are plainly in error, as the source CGIAR SRTM ASCII tiles should all have -9999 for nodata values.

Action: I replaced these 12 asc/prj pairs in the second directory with those taken from the first directory.

Additionally, there were 5 ASCII tiles that were identical between the two directories, but had -32768 for nodata in both cases. All had 2011-06-01

timestamps (same in both directories) rather than the original CGIAR dates:

```
                         mdate  nodata
srtm_65_01.asc  2011-06-01  -32768
srtm_65_03.asc  2011-06-01  -32768
srtm_65_04.asc  2011-06-01  -32768
srtm_65_05.asc  2011-06-01  -32768
srtm_65_13.asc  2011-06-01  -32768
```

Action: I replaced these 5 asc/prj pairs by re-extracting from the corresponding zip files from CGIAR.

All 784 tiles in ~organisms/DEM/cgiarSrtm/SRTM_90m_ASCII_4_1/ now have the expected nodata values of -9999, and no discrepancies remain between the two original directories.

Remaining todos:

1. replace several prj files that appear to differ superficially from the original CGIAR source
2. doublecheck 18 asc files that have newer timestamps than the original CGIAR source (I'm pretty sure the files themselves are unchanged, but want to make sure)
3. remove the first directory (which should no longer be used)

#8 - 01/25/2012 04:42 PM - Jim Regetz

For the sake of consistency, I restored the original CGIAR timestamps on newer dated files by re-extracting the following from our stored zips, and replacing the corresponding copies in ~organisms/DEM/cgiarSrtm/SRTM_90m_ASCII_4_1/. For all asc files and some prj files, the file contents themselves were identical between the directories, so this amounted to nothing more than a timestamp change. Asterisks in the list below indicate prj files for which those in the directory Reeves had created were in ESRI WKT format rather than the old Arc/Info format that the original CGIAR-distributed prj files use, though in any case the projection information itself was identical.

```
srtm_03_16.{asc,prj} *
srtm_06_16.{asc,prj} *
srtm_07_16.{asc,prj} *
srtm_08_16.{asc,prj} *
srtm_09_14.{asc,prj} *
srtm_18_13.{asc,prj} *
srtm_23_23.{asc,prj} *
srtm_24_23.{asc,prj} *
srtm_25_23.{asc,prj} *
srtm_29_23.{asc,prj} *
srtm_35_16.{asc,prj} *
srtm_65_02.{asc,prj}
srtm_65_06.{asc,prj}
srtm_65_07.{asc,prj}
srtm_65_08.{asc,prj}
srtm_65_10.{asc,prj}
srtm_65_11.{asc,prj}
srtm_65_14.asc  [the original CGIAR prj was already there in this case]
```

This cleanup leaves us with a complete set of 874 asc/prj pairs, all with timestamps conforming to the contents of the corresponding 874 zips obtained from CGIAR and stored on our server; the 874 prj files themselves are all identical copies of one another, as expected.

Subject to future renaming, our authoritative SRTM CGIAR source data holdings, including both a directory of the full set of zips and a directory of the full set of extracted asc/prj pairs, are now located in ~organisms/DEM/cgiarSrtm/, along with a new README.txt file that documents their origins to the best of my knowledge.

Lastly, as promised, I deleted ~organisms/SRTM_90m_ASCII_v4.1/, which contained only redundant and/or unnecessary files after completing all of the above.