

Data and file management

Digital soil mapping projects generate many files. Developing a system to organize, name, back up, and document data for other soil scientists and end users of the final product is essential. The USDA Geodata Management Document pertains to data requirements for Service Centers, and is a reasonable system to adopt if you are already familiar with it. However, it is a dated system, centered on shapefiles and does not address the variety of products you will be developing and managing.

Develop a folder and file naming convention that works for your office. There will be interim products developed that will need to be named and managed, which will also require adoption of a naming convention. Developing a system to back data up is essential. External drives are inexpensive, and could serve as a backup device. It would be reasonable to have several drives of the same data for redundancy and worst-case planning.

Ideas for naming folders:

1. Quadrangle
2. Watershed (10 or 12 digit HUC)
3. Project_area
4. Climate
5. Biota
6. Elevation
7. Geology

The first three would be “project area” centric, while the last four are “theme” centric. The resulting file names would likely be “theme” or “project area” centric respectively. It is preferable to have unique file names for all data, independent of the folder they are stored in. For example:

Folder

Alpa_Quad

File

Alph_slp
Alph_asp

Folder

Beta_Quad

File

Beta_slp
Beta_asp

Would be preferable to:

Folder

Alpa_Quad

File

Slope

Aspect

Folder

Beta_Quad

File

Slope

Aspect

Why bother with unique names?

It is better than trying to sort out 30 files all named “slope”, or “aspect” when you want to share or trouble-shoot data or processes.

Metadata

Creating and maintaining a minimum level of documentation for metadata is important. As time passes, personnel move on and the metadata should at least serve as a “recipe” of steps and parameters for the derivation of each data layer.

File Structure developed by SSO 12-5, St. Johnsbury, VT

Each quarter-quad has its own file structure, consisting of at least the following file folders:

DbfFiles, Final Map, Geodatabases, GRIDlayers, GRIDresults, Hillshades, Shapefiles

Each set of quarter-quads that is put together for a field season of work has a similar file structure, and this document details the steps to be taken to create all the files that will go into the Final Map, GRIDlayers, and GRIDresults folder. The following is a list of all the files and how they should be named; this example is from the averill_sw_nine area. If there are any files in these folders with any other names, they will be deleted.

More details on how these files are created come later in this document.

Folder	File Name	What is this file?
GRIDlayers	avsw	1m DEM of 9 qquad work area, clipped to 500m outside of 9qquad boundary, after each qquad has been shaved
	avsw99	avsw filtered using a 9 by 9 rectangular neighborhood
	avsw995	The 5m DEM that we use for inference (Filtered and resampled)
	Multiwetmsm	Smoothed wetness index
	Slope30	Slope layer generated using 30m neighborhood
	Slope30re6	30m slope layer reclassified for ablation till
	Slope30re7	30m slope layer reclassified for outwash (includes A slopes)
	Slope60	Slope layer generated using 60m neighborhood

	Slope60re	60m slope layer reclassified to our 5 slope classes
GRIDresults	Cabot	Cabot inference results
	Colonel	Colonel inference results
	Dixfield	Dixfield inference results
	Harden	Hardened map of Cabot, Colonel, and Dixfield
	Hardenent	Uncertainty map
	Hardenexg	Uncertainty map
	Hardph	Hardened map with slope phase
	Hardphmu	Hardened slope phase map combined into map units
	Pdat	Poorly drained ablation till from model
	Allat	Ablation till inference results
Final Map	Allpm	Raster map created from certified parent material layer
GRIDS		
	Btmu	Basal till map units
	Btmu84	Btmu with slivers removed, 8- then 4-connected
	Final	Final raster map (all parent materials put back together)
	Rkymu	Rocky map units
	Rkymu84	Rockymu with slivers removed,8- then 4-connected
	Vrky	Very rocky map units
	Vrky84	Vrky with slivers removed, 8- then 4-connected
	Atmu	Ablation till mapunits
	Atmu84	Atmu with slivers removed, 8- then 4-connected
Final Map Shapefiles	Allpmpolys	Certified parent material exported to shapefile
	Averill_border	Outer border of mapping area
	Polys	Polys before labeling
	FinalPolys	Labeled polys
	SimpPoly	Labeled polys that have been simplified to remove vertices

NAMING CONVENTION FOR ELEVATION TERRAIN AND GENERALIZATION DERIVATIVES PROPOSED by Dwain Daniels, GIS Specialist CNTSC

The following naming convention is proposed to provide consistency in naming elevation derivatives and generalization products. The components of the file name are:

- Area identification
- type of surface (bare earth or first return if applicable)
- cell size
- pyramid used (if applicable)
- derivative identification.

.Examples.

ms052_be_5m_5pyr_fel

This is the file name for a 5meter resolution filled elevation dataset created from a bare earth terrain dataset with a 0.5 meter vertical resolution pyramid applied for Leflore County, Mississippi.

When another dataset is created the extension would be added, as in the example the flow accumulation would have the extension `_acc` added. **ms052_be_5m_5pyr_fel_acc**

ms052_be_5m_5pyr_cir3_slp

This example would be the name for a slope gradient raster dataset that has had a focal mean calculation performed on the elevation raster in a 3 cell radius circle shape.

All raster elevation data is stored in the file geodatabase raster dataset format. This naming convention will **NOT** work with ESRI GRIDs that have a 13 character limit on dataset name. *If you are using ArcSIE for modeling efforts, follow the naming limits for ERSI GRIDs. ArcSIE will implement other raster formats in the future, but is currently based on the ESRI GRID.*

The extension used in the name of the primary and secondary elevation derivatives and selected generalization products that are most commonly used created are:

Table 1. Raster derivative type file extensions.

Extension	Derivative	Value Type	Description
<code>_elev</code>	Elevation	Floating Point 32Bit	Units above Mean Sea Level, default is meters, if units are not meters the unit is identified in the extension, e.g. elevft.
<code>_hsd</code>	Hillshade	Unsigned Integer 8Bit	Sunlight reflection off ground surface.
<code>_slp</code>	Slope Gradient	Floating Point 32Bit	Cells carry a value calculated from the 8 surrounding cells of the maximum rate of change in the ratio of vertical and horizontal distance.
<code>_reggrp</code>	Unique Regions	Unsigned Integer 16 Bit	Cells carry a sequential number of individual regions starting in the upper left corner of the matrix.
<code>_rcls</code>	Reclassification	Unsigned Integer 8Bit	Individual or groups of values in the original raster are converted to designated integer values in the output raster.
<code>_nib#</code>	Nibble	Unsigned Integer 8Bit	Regions of cells determined by a nibble mask are replaced by the cell values surrounding them. A number can be added to indicate what values were replaced.
<code>_cir#</code>	Focal Statistic	Floating Point 32Bit	Cell values are a statistic calculated from values of surrounding cells based on a geometric shape such as a circle, square, rectangle, etc. The number indicates the extent based on the number of cells in the geometric shape. Unless designated, the statistic is assumed to be the mean.
<code>_fel</code>	Filled Elevation	Floating Point 32Bit	Depressions in the surface are filled for continuous flow.
<code>_fdr</code>	Flow Direction	Unsigned Integer 8Bit	Each cell carries a value indicating one direction of flow. D8 flow direction.

_acc	Flow Accumulation	Signed Integer 32Bit	Each cell carries the value indicating how many cells exclusively flow into it.
_pfe#	Partial Filled Elevation	Floating Point 32Bit	Depressions in the surface are filled to a designated threshold depth. A number can be added to indicate the threshold depth value.