Figure 1:

EARTH BIG DATA Data Guide

Author: Josef Kellndorfer, October 2017 Latest Update: 2/2019

Naming Conventions

Data stacks produced with the Software for Earth Big Data Processing, Prediction Modeling and Organization (SEPPO) are in general delivered as a bundle of GeoTIFF and Virtual Raster Table (VRT) data structures that are ready to use with GDAL compatible software like QGIS.

Time series data stacks are organized in either custom regions of interest, geographic 1x1 degree or MRGS/Sentinel-2 compatible tiles. Image data are stored in GeoTIFF format containing all georeferencing information. Data stacks (e.g. Timeseries or time series metrics) are organized in VRT formats which can stack the GeoTIFFs in many different ways, e.g. by day of year, seasons, brightness levels, band subsets, etc.

Tile Names

Customary ID's for Region of Interest Stacks

Directories which contain SEPPO preprocessed data have descriptive names of variable length, however typically desribe with the last three characters what satellite sensor that data stack is produced from. For example

BIOsS1 describes a region where some "BIO"-related analysis (likely "biomass!") is done with the sensor (**s**) Sentinel-1 (**S1**)

Within directories filenames are typically prefixed by a code for the spatial reference system identifier (*SRID*) which follows the codes of the EPSG system widely used in geospatial referencing (See spatialreference.org). Following the SRID code, the lower left coordinates of the region of interest in X and Y coordinates in the spatial referencing system are given. Sensor designations typically complete the prefix.

Example:

S32631X398020Y1315440sS1_

- S32631: SRID 32631 (= UTM Zone 31, Northern Hemisphere (326))
- X398020: Easting of lower left coordinate (398020)
- Y1315440: Northing of lower left coordinate (1315440)
- sS1: Sensor Sentinel1

Geographic 1 x 1 degree tiles

Geographic tiles are named based on the upper left corner:

NnnEeee, NnnWwww, SssEeee, or SssWwww

- N/S = North or South
- E/W = East or West
- nn,ss = degrees north or south with leading 0 if needed, e.g. 01 or 10
- eee,www = degrees east or west with leading 0 if needed, e.g. 001 or 100

The thus specified full degree coordinate pair refers to the lower left coordinate of the tile.

Example: N47W078

```
47 ---------
  |         |
  |         |
  |         |
46 ---------
 -78      -77
```

MGRS / Sentinel-2 Tiling Scheme

The MGRS/Sentinel-2 tiling scheme refers to 110x110 sqkm tiles. These tiles are UTM projected data sets. The MGRS tiling scheme explanation and a kml of the Sentinel-2 convention can be found at:

http://earth-info.nga.mil/GandG/coordsys/grids/mgrs.pdf

https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products

The MGRS tiles are identified by a five letter and number code

Example: 18NZJ

- 18 = UTM Zone
- N = UTM Row
- ZJ = Tile identifier within UTM Zone 18N

SAR Data

To idendify image data the following naming conventions are applied.

- Fields are separated by an underscore ’_’
- Tile names are followed by a lowercase ‘s’ (for sensor or satellite) and a two digit code for a sensor or satellite identification e.g. S1 (Sentinel-1), A1 (ALOS-1), A2 (ALOS-2), L8 (Landsat-8), S2 (Sentinel-2), N1 (NISAR-1)
- A series of indentifiers follows. See examples below.

Examples

20NRKsA1_A_HH_0118_mtfil.vrt

- Tile: 20NRK (MRGS)
- Sensor: A1 (ALOS-1)
- Flight direction: ASCENDING
- Polarization: HH
- Path: 0118 (Leading 0s)
- Processing level: multitemporally filtered (mtfil)
- Image format: VRT File

20NRKsS1_D_vv_0083_B_mtfil.vrt

- Tile: 20NRK (MRGS)
- Sensor: S1 (Sentinel-1)
- Flight direction: DESCENDING
- Polarization: vv
- Path: 0083 (Leading 0s)
- Satellite: B (Sentinel-1B)
- Processing level: multitemporally filtered (mtfil)
- Image format: VRT File

20NRKsS1_A_vh_20150520_0164_A_mtfil_amp.tif

- Tile: 20NRK (MRGS)
- Sensor: S1 (Sentinel-1)
- Flight direction: ASCENDING

- Polarization: vh
- Acquisition Date: 2015-05-20
- Path: 0164 (Leading 0s)
- Satellite: A (Sentinel-1B)
- Processing level: multitemporally filtered (mtfil)
- Scaling: Amplitude scaled to 16bit unsigned integer, such that dB = 20 * log10(`amp`) - 83
- Image format: GeoTIFF

Other Identifiers

Often other designators are added to explain the nature of the data in the file name, e.g. a subsetting by bands or months or a regional section of the tile, e.g. _south. There are no fixed standards for this addition, and most should be self explanatory or are otherwise documented.

Time Series Metrics

GeoTIFF Images that are resulting from some standard time series metrics computation are organized in a _tsmetrics subfolder and have the following filename endings:

- 95th Percentile: _p95.tif
- 5th Percentile: _p5.tif
- 95th-5th Percentile: _prange.tif
- Median: _median.tif
- Maximum: _max.tif
- Minimum: _min.tif
- Range (max-min): _range.tif
- Mean: _mean.tif
- Variance: _var.tif
- Coeff. of Variation: _cov.tif
- Sdiff (experimental) _sdiff.tif
- Count: _count.tif

These, or a subset of these metrics are typically organized via a vrt stack named in the parent directory of the subfolder

Example: 21NTEsS1_D_vh_0083_mtfil_26_to_29_tsmetrics.vrt

- Tile: 21NTE
- Sensor: Sentinel-1
- FlightDirection: DESCENDING
- Polarization: vh
- Path: 0083
- Processinglevel: Multi-temporally filtered
- Other: 26_to_29 Subset of bands 26,27,28,29 from 21NTEsS1_D_vh_0083_mtfil

- Time series metrics file: tsmetrics
- Image format: VRT

Special Files

EBD Software stores some special files for input and output of processing steps:

- _imageNum.tif Count of pixles used for multi-temporal filtering
- _imageSum.tif Summation of pixel values used for multi-temporal filtering
- .dates Plain text file with dates for bands in the corresponding vrt file. Format: YYYYMMDD
- EBD_RESULT_FILES Listing of files as output from the last EBD processing step, e.g. multi-temporal filtering
- .log log file for SEPPO processing runs or other logs

SAR Data Types and Scaling Conventions

SAR backscatter data processed with the SEPPO Software are typically available with the following scaling conventions:

Backscatter data:

- float32:
  - SAR power data
  - No data value: 0.
- unsigned 16 bit (default for most SEPPO processing routines):
  - SAR amplitude data, linearly scaled digital numbers (DN)
    - dB = 20*log10(DN) - 83
    - power = DN^2 / 199526231
    - No data value 0
- unsigned 8 bit:
  - dB = 0.15 * DN -31
  - No data value 0

Incidence angle data:

- unsigned 8 bit:
  - radians = DN * 100

Layover/Shadow masks:

- unsigned 8 bit (Gamma Remote Sensing convention):

Value	Effect	Description
0	NOT_TESTED	No effect

Value	Effect	Description
1	TESTED	Neither layover nor shadow
2	TRUE_LAYOVER	Pixel were slope angle is greater than look angle
4	LAYOVER	Pixel in area affected by layover
8	TRUE_SHADOW	Pixel were opposite of slope angle is greater than look angle
16	SHADOW	Pixel in area affected by shadow