NetCDF Why and How:

Creating Publication-Quality NetCDF Datasets

Presenters: Michele Thornton, Jack McNelis

With contributions from:

Matt Donovan, Yaxing Wei

ORNL Distributed Active Archive Center (ORNL DAAC)

Environmental Sciences Division

Oak Ridge National Laboratory

Oak Ridge, TN

September 4, 2019

NASA EOSDIS ORNL DAAC

ORNL DAAC*

The ORNL DAAC **mission** is to assemble, distribute, and provide data services for a comprehensive archive of terrestrial biogeochemistry and ecological dynamics observations and models to facilitate research, education, and decision-making in support of NASA's Earth science.

https://daac.ornl.gov/

*Oak Ridge National Laboratory Distributed Active Archive Center

Purpose of this Webinar

1. **Why** we, as a DAAC within NASA EOSDIS, use netCDF file formats w/ CF Conventions and why you as a data provider may want to consider netCDF

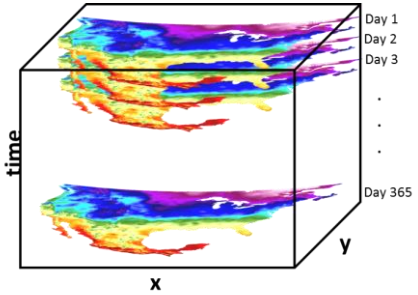
2. **How** we *transform* and *standardize* files to accepted netCDF CF *conventions*
   - Demonstration in:
     1. GDAL and NCO – Command Line Utilities that Manipulate netCDF file formats
     2. Python

This webinar and demonstration material will be available at:
ORNL DAAC Resources Page: https://daac.ornl.gov/resources/learning/

Why netCDF

1. NASA (ESDIS) – Approved!
   https://earthdata.nasa.gov/esdis/eso/standards-and-references/dataset-interoperability-recommendations-for-earth-science

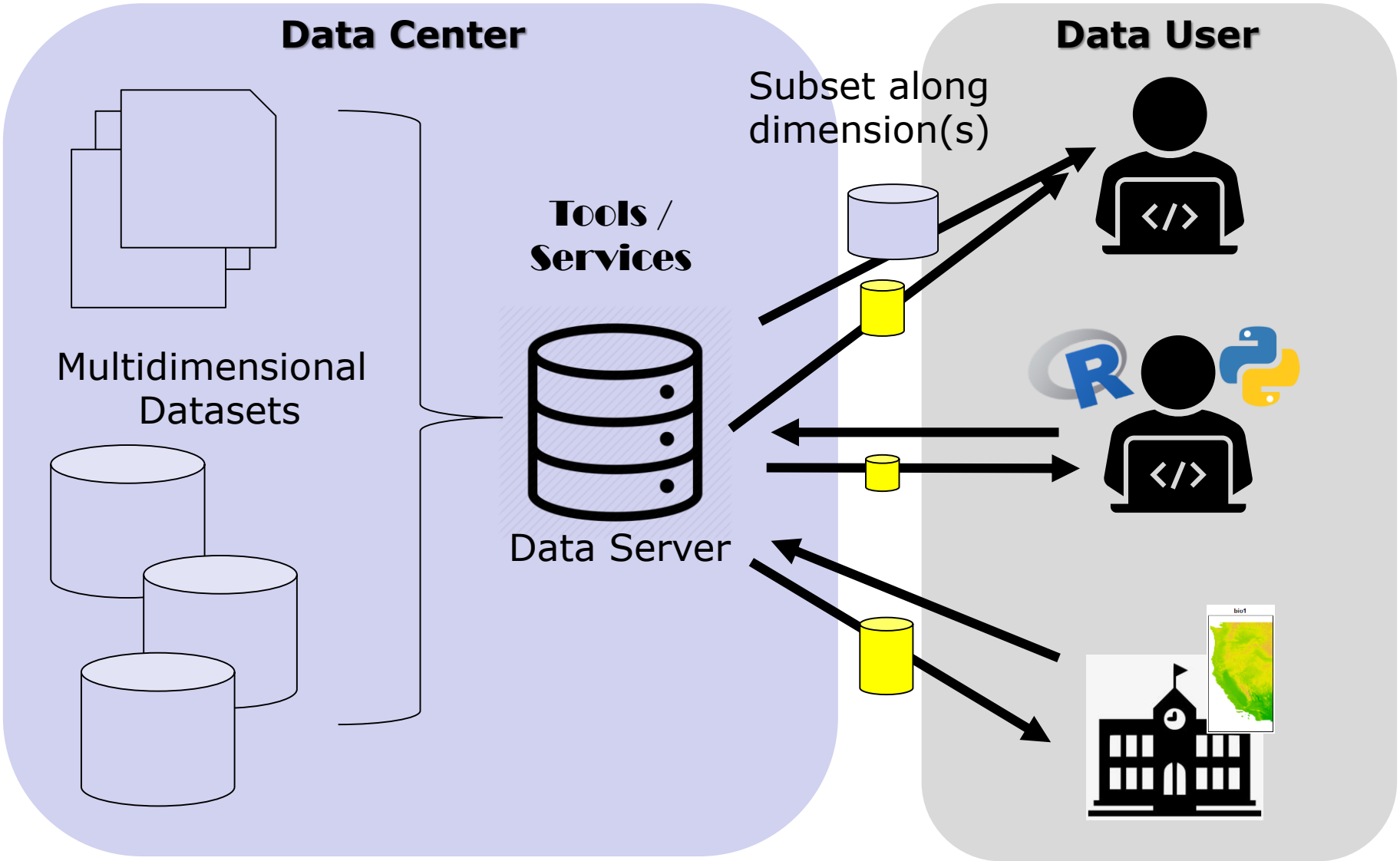
2. netCDF (network common data form) format
   - Created for multi-dimensional data
   - A UCAR Unidata product
   - Supports creation, access, and sharing of array-oriented scientific data
   - Grows in popularity and usefulness



3. It important in the DAAC's mission to distribute and provide data services
   - Users can programmatically access data
   - Users can subset datasets spatially and temporally

4. Improved Data Analysis
   - Software such as R, Python, and ArcPro

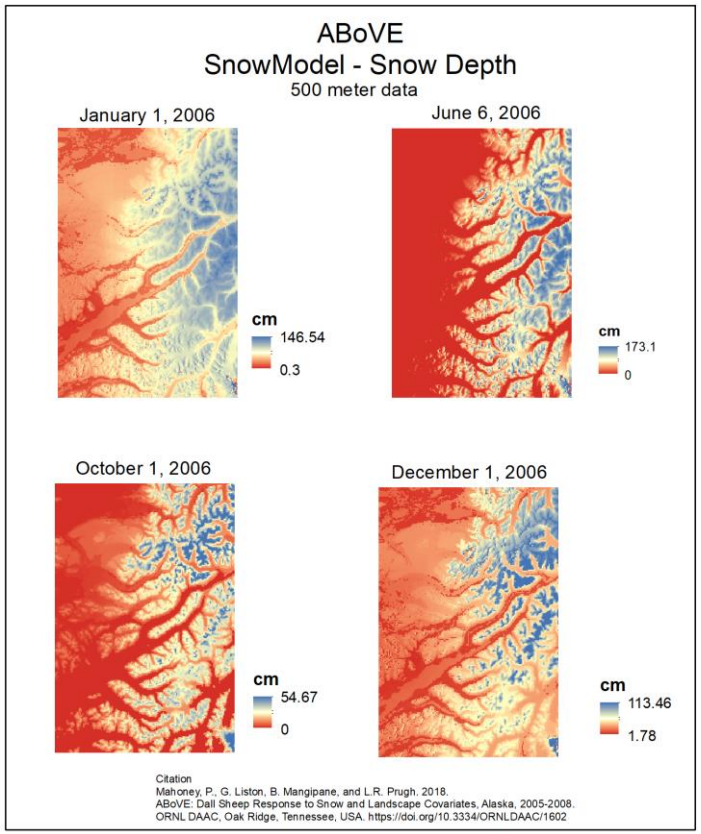
Why netCDF


Data Center
Data User
Subset along dimension(s)
Tools / Services
Multidimensional Datasets
Data Server
R

netCDF How

An Example of a “Typical” Dataset Transformation

ABoVE: Dall Sheep Response to Snow and Landscape Covariates, Alaska, 2005-2008
https://doi.org/10.3334/ORNLDAAC/1602



Original Dataset

Model Data Output - SnowModel

geoTIFF file format

Individual Daily Files for 2005 – 2008

Each file has 2 “Bands”

Band 1 = Snow Depth

Band 2 = Snow Density

5 different spatial resolutions (meter)

25, 100, 500, 2000, 10000

~ 5,480 files (*.tif)

netCDF How

An Example of a “Typical” Dataset Transformation

ABoVE: Dall Sheep Response to Snow and Landscape Covariates, Alaska, 2005-2008
https://doi.org/10.3334/ORNLDAAC/1602

*2006 Data Files

Data File (Granule)
snow_density_00025m_2006.nc4
snow_density_00100m_2006.nc4
snow_density_00500m_2006.nc4
snow_density_02000m_2006.nc4
snow_density_10000m_2006.nc4
snow_depth_00025m_2006.nc4
snow_depth_00100m_2006.nc4
snow_depth_00500m_2006.nc4
snow_depth_02000m_2006.nc4
snow_depth_10000m_2006.nc4

Final Dataset Distribution

Organized files by:

- 1 file/year (365 days)
- Separate Variables
- Separate Spatial Resolution

20 Snow Depth Files

20 Snow Density Files

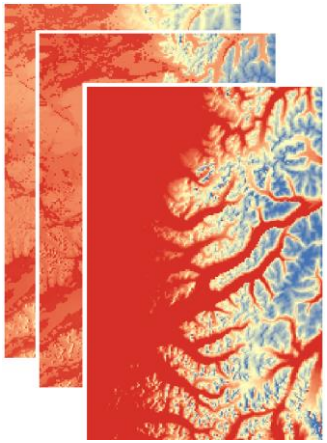
- netCDF file format

---

40 Files (*.nc)

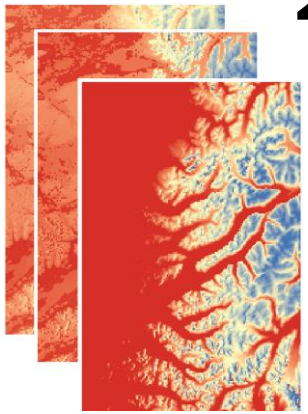
netCDF How

An Example of a "Typical" Dataset Transformation



365, 2-Band geoTIFFs

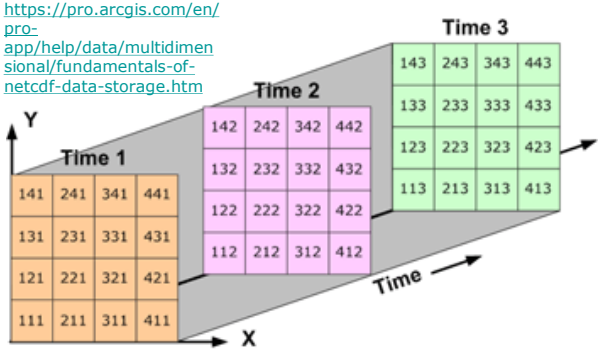
1. Extract Band1 (SnowDepth)



365, 1-Band geoTIFFs

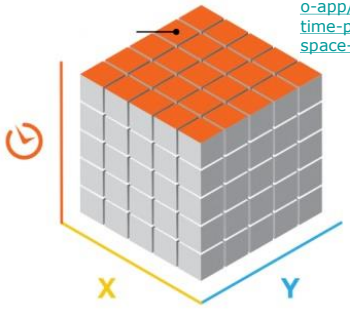
2. Convert to netCDF

https://pro.arcgis.com/en/pro-app/help/data/multidimensional/fundamentals-of-netcdf-data-storage.htm



365, 2-dimensional netCDF

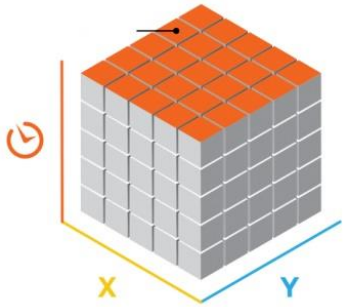
3. Merge into a 3-D Array



1, 3-D Data File

https://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/create-space-time-cube.htm

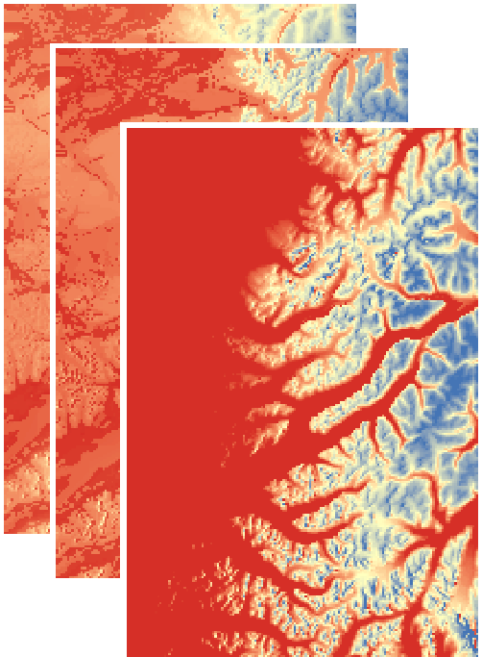
4. Apply CF Standards



Publication Quality

```
netcdf snow_depth_00500m_2006_time_timebn
dimensions:
                time = 365 ;
                x = 160 ;
                y = 240 ;
variables:
float SnowDepth_500m(time, y, x) ;
                SnowDepth_500m:_FillVa
                SnowDepth_500m:coord
                SnowDepth_500m:grid_m
                SnowDepth_500m:long_
                SnowDepth_500m:units
double time(time) ;
                time:units = "days since
                time:calendar = "standa
                time:description = "midd
                time:long_name = "time" ;
                time:standard_name = "time" ;
```

netCDF How



GeoTIFF Files:
Snow Depth, 500 meter,
2006, Daily Data

Parts of a GeoTIFF File

Columns x Rows

- 160 col x 240 row

Individual Files or Multiband / Time

- Daily

Projection (CRS)

- Albers Equal Area

Pixel Variable

- Snow Depth

Attribute/Unit

- cm

```
Coordinate System is:
PROJCS["unnamed",
    GEOGCS["NAD83",
        DATUM["North_American_Datum_1983",
            SPHEROID["GRS 1980",6378137,298.25722,
                AUTHORITY["EPSG","7019"]],
            TOWGS84[0,0,0,0,0,0,0],
            AUTHORITY["EPSG","6269"]],
        PRIMEM["Greenwich",0],
        UNIT["degree",0.0174532925199433],
        AUTHORITY["EPSG","4269"]],
    PROJECTION["Albers_Conic_Equal_Area"],
    PARAMETER["standard_parallel_1",55],
    PARAMETER["standard_parallel_2",65],
    PARAMETER["latitude_of_center",50],
    PARAMETER["longitude_of_center",-154],
    PARAMETER["false_easting",0],
    PARAMETER["false_northing",0],
    UNIT["metre",1,
        AUTHORITY["EPSG","9001"]]]
```

netCDF How

Parts of a NetCDF File - Stored as a single file comprising two parts:

1. Header
   - containing all the information about ***dimensions***, ***attributes***, and ***variables*** except for the variable data

     dimensions
     - Define the structure and record the length of the data array
     - Typically Lon (or X), Lat (or Y),
     - Time (or Altitude, Depth, etc.)

     variables
     - Variables are the actual data in the array.

     attributes
     - provide details about each variable
2. Data part
   - contain the ***data values*** for the variables

netCDF How

Standards (CF Conventions) http://cfconventions.org

- CF Conventions refer to the “rules” of how-to build and label netCDF files, especially with regard to the header metadata content
- The metadata provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data
- Software that reads netCDF is dependent on files following conventions
- Facilitates building applications with powerful extraction, regridding, and display capabilities

netCDF How

SnowDepth GeoTIFF

Columns x Rows

- 160 col x 240 row

Time Step

- Daily

Pixel Variable

- Snow Depth

  Attribute/Unit cm

Projection (CRS)

- Albers Equal Area

SnowDepth netCDF Header

```
netcdf snow_depth_00500m_2006_time_timebnds {
dimensions:
        time = 365 ;
        x = 160 ;
        y = 240 ;
variables:
        float SnowDepth_500m(time, y, x) ;
                SnowDepth_500m:_FillValue = -3.4e+38f ;
                SnowDepth_500m:coordinates = "lon lat" ;
                SnowDepth_500m:grid_mapping = "crs" ;
                SnowDepth_500m:long_name = "snow depth" ;
                SnowDepth_500m:units = "cm" ;
        double time(time) ;
                time:units = "days since 2006-01-01 00:00:00" ;
                time:calendar = "standard" ;
                time:description = "middle of each day" ;
                time:long_name = "time" ;
                time:standard_name = "time" ;
        double x(x) ;
                x:standard_name = "projection_x_coordinate" ;
                x:long_name = "x coordinate of projection" ;
                x:units = "m" ;
        double y(y) ;
                y:standard_name = "projection_y_coordinate" ;
                y:long_name = "y coordinate of projection" ;
                y:units = "m" ;
```

attributes

netCDF How

SnowDepth GeoTIFF

Columns x Rows
- 160 col x 240 row

Time Step
- Daily

Pixel Variable
- Snow Depth

  Attribute/Unit
  cm

Projection (CRS)
- Albers Equal Area

SnowDepth netCDF Header

```
netcdf snow_depth_00500m_2006_time_timebnds {
dimensions:
        time = 365 ;
        x = 160 ;
        y = 240 ;
variables:
        float SnowDepth_500m(time, y, x) ;
                SnowDepth_500m:_FillValue = -3.4e+38f ;
                SnowDepth_500m:coordinates = "lon lat" ;
                SnowDepth_500m:grid_mapping = "crs" ;
                SnowDepth_500m:long_name = "snow depth" ;
                SnowDepth_500m:units = "cm" ;
        double time(time) ;
                time:units = "days since 2006-01-01 00:00:00" ;
                time:calendar = "standard" ;
                time:description = "middle of each day" ;
                time:long_name = "time" ;
                time:standard_name = "time" ;
        double x(x) ;
                x:standard_name = "projection_x_coordinate" ;
                x:long_name = "x coordinate of projection" ;
                x:units = "m" ;
        double y(y) ;
                y:standard_name = "projection_y_coordinate" ;
                y:long_name = "y coordinate of projection" ;
                y:units = "m" ;
```

attributes

netCDF How

SnowDepth GeoTIFF

Columns x Rows

- 160 col x 240 row

Time Step

- Daily

Pixel Variable

- Snow Depth

  Attribute/Unit

  cm

Projection (CRS)

- Albers Equal Area

Header (cont.)

variables: (cont.)

```
        float lat(y, x) ;
                lat:standard_name = "latitude" ;
                lat:long_name = "latitude" ;
                lat:units = "degrees_north" ;
        float lon(y, x) ;
                lon:standard_name = "longitude" ;
                lon:long_name = "longitude" ;
                lon:units = "degrees_east" ;
        char crs ;
                crs:grid_mapping_name = "albers_equal_area" ;
                crs:false_easting = 0. ;
                crs:false_northing = 0. ;
                crs:latitude_of_projection_origin = 50. ;
                crs:longitude_of_central_meridian = -154. ;
                crs:standard_parallel = 55., 65. ;
                crs:long_name = "CRS definition" ;
                crs:longitude_of_prime_meridian = 0. ;
                crs:semi_major_axis = 6378137. ;
                crs:inverse_flattening = 298.257222101004 ;

// global attributes:
                :Conventions = "CF-1.6" ;
                :institution = "University of Washington" ;
                :title = "Navigating Snowscapes: " ;
                :project = " (ABoVE)" ;
                :contact = "pmahoney29@gmail.com" ;
                :references = "Mahoney, P., et al." ;
```

netCDF How

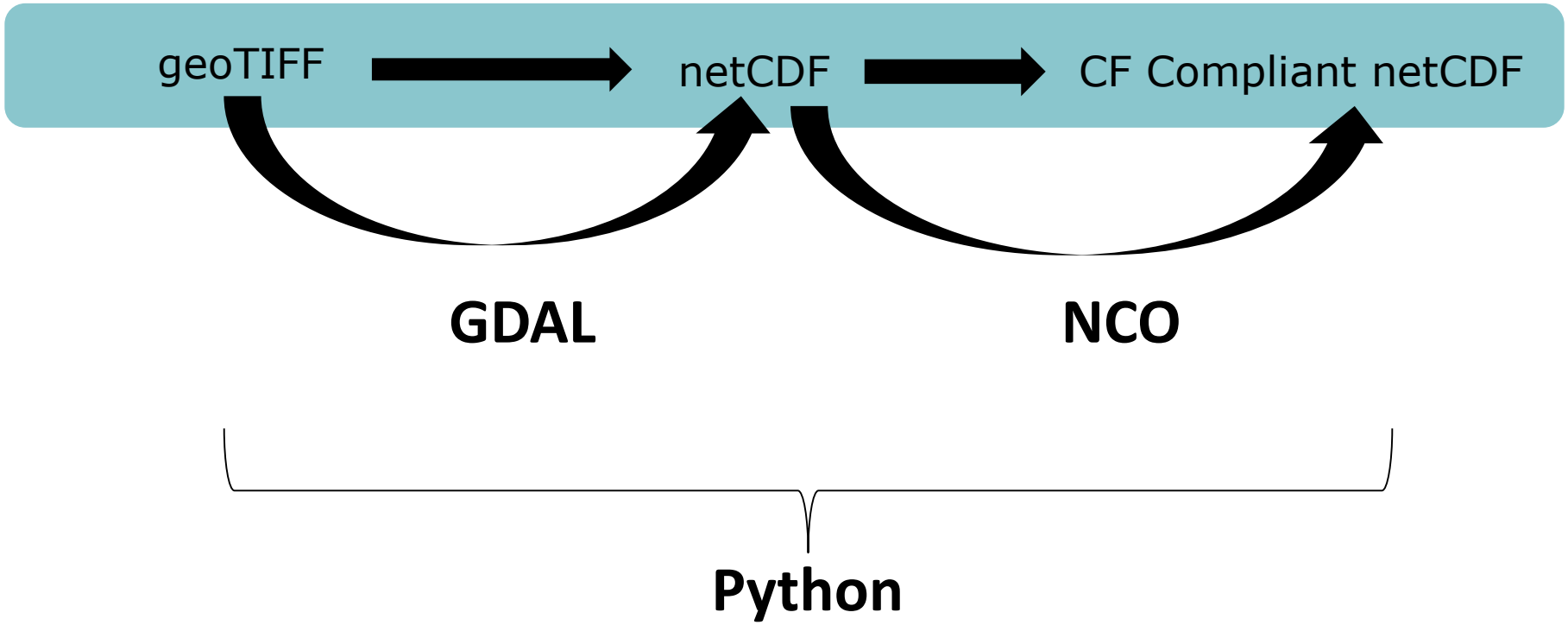
Data Part

ncdump -v variable filename.nc

data:

SnowDepth_500m =
4.14, 4.93, 5.84, 6.47, 6.9, 9.15, 24.42, 25.45, 25.68, 27.52, 29.09,
30.01, 33.62, 38.1, 36.54, 20.24, 31.2, 39.19, 36.21, 26.99, 26.24,
30.99, 24.29, 26.26, 11.75, 7.31, 18.21, 8.07, 8.82, 27.38, 27.75, 29.18,
25.16, 28.1, 28.23, 28.55, 16.18, 16.26, 16.52, 16.8, 16.95, 17.1, 29.7,
30.22, 32.02, 35.23, 36.82, 38.74, 41.44, 38.76, 38.64, 45.14, 48.81,
40.85, 45.08, 52.59, 47.35, 40.96, 40.03, 44.26, 54.95, 49.89, 43.85,
52.49, 40.6, 39.47, 48.32, 47.38, 42.27, 41.93, 49.54, 46.38, 38.94,
37.48, 50.29, 52.52, 56.79, 63.49, 44.64, 41.42, 49.59, 63.27, 54.73,
27.46, 45.41, 55, 59.12, 43.57, 53.41, 69.57, 51.09, 43.31, 48.09, 52.93,
40.27, 32.03, 31.43, 33.34, 36.43, 41.55, 45.76, 47.59, 45.57, 42.19,
42.08, 42.82, 51.15, 56.86, 58.51, 50.62, 43.65, 53.45, 59.99, 68.92,
44.63, 63.05, 58.65, 54.91, 56.68, 55.86, 53.98, 55.48, 55.87, 47.53,
48.95, 50.67, 42.41, 42.35, 40.11, 35.63, 26.62, 26.74, 31.4, 46.73,
54.85, 61.01, 66.35, 53.76, 54.71, 55.3, 56.66, 47.95, 40.11, 41.96,
45.05, 45.94, 47.33, 49.67, 51.3, 64.2, 58.71, 59.98, 62.63, 62.47,
64.68, 67.2, 66.47, 67.21, 66.26, 89.39,
...
...

netCDF How



Software

Command Line Utilities

GDAL – Geospatial Data Abstraction Library https://gdal.org

GDAL is a translator library for raster and vector geospatial data formats

GDAL Programs used in the webinar

- gdalinfo - lists information about a raster dataset
- gdal_translate - converts raster data between different formats

NCO = netCDF Operators http://nco.sourceforge.net/

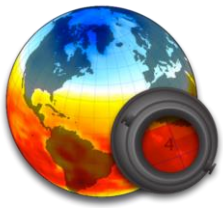
The netCDF Operators (NCO) comprise about a dozen standalone, command-line programs that manipulate netCDF files

NCO Programs used in the webinar

- ncdump - generates text representation of a netCDF dataset
- ncatted – netCDF ATTribute EDitor
- ncrename – netCDF RENAMEer
- ncecat – netCDF Ensemble conCATenator
- ncks – netCDF Kitchen Sink
- nccopy - copies and optionally compresses and chunks netCDF data

Software

Python – Programming Language
https://www.python.org/



Panoply https://www.giss.nasa.gov/tools/panoply/
A very useful visualizer for netCDF Header and Data
Plots geo-referenced netCDF and other datasets