

This is not an ADB material. The views expressed in this document are the views of the author/s and/or their organizations and do not necessarily reflect the views or policies of the Asian Development Bank, or its Board of Governors, or the governments they represent. ADB does not guarantee the accuracy and/or completeness of the material's contents, and accepts no responsibility for any direct or indirect consequence of their use or reliance, whether wholly or partially. Please feel free to contact the authors directly should you have queries.

# Introduction to netCDF Files With CF Metadata Conventions

Dr. Matthias Lankhorst



SCRIPPS INSTITUTION OF  
OCEANOGRAPHY

UC San Diego

# Outline

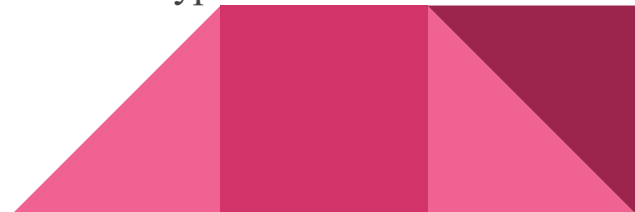
## 1. About netCDF files:

- a. History
- b. Usage


## 2. Importance of metadata

## 3. CF Metadata Conventions:

- a. History
- b. Usage details:
  - i. Identifying physical quantities via their “standard\_name” representation
  - ii. Representing different data geometries through “featureType”
  - iii. Suggested minimum list of metadata fields



# NetCDF History

- “Network Common Data Format”, concepts built on previous “Common Data Format” by NASA but not compatible with it
  - UCAR (“University Corporation for Atmospheric Research”), a US-based nonprofit organization
  - <https://www.unidata.ucar.edu/software/netcdf/>
  - Software libraries to read/write (scientific) data files
  - Original specifications released 1990 (“classic” netCDF format)
  - Current versions have additional features, incl. compatibility with HDF5 format commonly used in astrophysics
- 

# NetCDF Usage

- Data are usually N-dimensional arrays (of numbers)
- Binary file format - requires special software to read/write
- Metadata is included with data - files are “self-explanatory”
- Can access subsets of data (for improved computer efficiency with large datasets)
- Widely used in meteorology and oceanography:
  - Numerical weather/climate models
  - Argo floats, OceanSITES moorings, satellite altimetry, ...



# Importance of Metadata, or: “What’s This?”

You find a file called “SriLanka\_Data.dat” with the following lists of numbers.

What could they mean?

```
[ 23695.083 23695.167 23695.250 23695.333 ... ]
```

```
[ 604.135 604.134 604.134 604.134 ... ]  
[ 3889.274 3889.274 3889.273 3889.273 ... ]  
[ 645.442 645.442 645.442 645.442 ... ]  
[ 4408.656 4408.656 4408.656 4408.655 ... ]
```

# Importance of Metadata, or: “What’s This?”

```
title           = 'Processed Data from PIES Instruments (Pressure-Sensing Inverted Echo Sounders)
                  Collected in the Sri Lankan Boundary Currents between 2014 and 2019'
summary         = 'In 2014/15, PIES instruments (Pressure-Sensing Inverted Echo Sounders) were
                  deployed at four sites that span the boundary currents around Sri Lanka. Two sites
                  are east, and two south of Sri Lanka. PIES are seafloor instruments that make
                  observations of the water pressure and the travel time of a vertically transmitted
                  acoustic signal. Data from these instruments are reported here, in a processed version
                  that is on a common time grid with sensor drifts removed, and tidal signals separated
                  out of the pressure records. The scientific rationale for collecting these data is to
                  monitor the strengths of the boundary current flows.'
```

```
publisher_name  = 'Matthias Lankhorst'
publisher_url   = 'https://orcid.org/0000-0002-4166-4044'
contributor_name = 'Nilanthi Priyadarshani, Uwe Send'
contributor_role = 'principal investigator, principal investigator'
time_coverage_start = '2014-11-15T08:00:00Z'
time_coverage_end   = '2019-11-03T04:00:00Z'
```

```
TIME [ 23695.083 23695.167 23695.250 23695.333 ... ]
Size: 21767x1
Dimensions: TIME
Datatype: double
Attributes:
  units = 'days since 1950-01-01T00:00:00Z'
  standard_name = 'time'
  axis = 'T'
```

```
SEAFLOOR_PRESSURE [ 604.135 604.134 604.134 604.134 ... ]
Size: 4x21767 [ 3889.274 3889.274 3889.273 3889.273 ... ]
Dimensions: NSTATION, TIME [ 645.442 645.442 645.442 645.442 ... ]
Datatype: double [ 4408.656 4408.656 4408.656 4408.655 ... ]
Attributes:
  standard_name = 'sea_water_pressure_at_sea_floor'
  units = 'dbar'
  comment = 'Sensor drift has been removed, and data have been low-pass filtered. Tides
            have been removed and are available in a separate variable.'
```

```
ancillary_variables = 'SEAFLOOR_PRESSURE TIDES'
coordinates = 'LONGITUDE LATITUDE DEPTH TIME SITE_NAME'
```

# Take-Home Message:

Metadata are important. Without metadata, the data are only meaningless numbers.

# NetCDF File Contents

- **Variables:** the structures that hold the “numbers”
- **Dimensions:** the “sizes” of the variables
- **Attributes :** metadata fields
  - Can be associated with a single variable, or
  - with the entire file (“global” attribute)

- **Example:**

```
dimensions:
```

```
    TIME      = 21767 ;
```

```
    NSTATION  = 4 ;
```

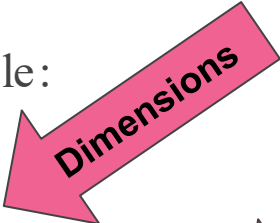
```
double SEAFLOOR_PRESSURE(TIME, NSTATION) ;
```

```
    SEAFLOOR_PRESSURE:standard_name = "sea_water_pressure_at_sea_floor" ;
```

```
    SEAFLOOR_PRESSURE:units         = "dbar" ;
```

```
:title           = "Processed Data from PIES Instruments..." ;
```

```
:acknowledgement = "These data were collected with support from..." ;
```



**Dimensions**



**Variable**



**Attributes**  
(for variable  
and global)



# Take-Home Message:

What do netCDF files contain?

Variables, dimensions, and attributes.

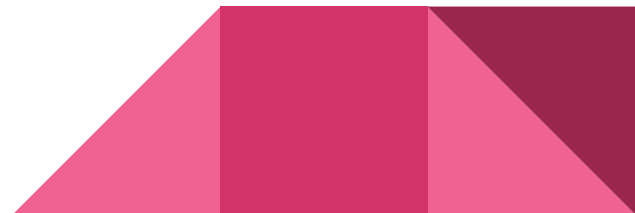
# CF Metadata Conventions

- “Climate and Forecast”, <http://cfconventions.org/>
- Additional reference: Hassell et al., 2017. [doi:10.5194/gmd-10-4619-2017](https://doi.org/10.5194/gmd-10-4619-2017)
- History:
  - Builds on COARDS conventions (1990s, Cooperative Ocean/Atmosphere Research Data Service, a collaboration between US NOAA and universities)
  - Governed by a panel and committees, typically professionals volunteering some of their time to this service, “community-driven”
  - CF Version 1.0 released in 2003
- Assumes (but does not strictly require) netCDF files
- Principles: self-describing files, metadata readable by people and software, ...
- Intended use: climate science (meteorology, oceanography, ...)

# CF Metadata Conventions

Side note: **ACDD**

- Attribute Convention for Data Discovery
- [https://wiki.esipfed.org/Attribute\\_Convention\\_for\\_Data\\_Discovery](https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery)
- Provides additional standards for basic metadata information (title, authorship, ..)
- Compatible with CF
- Many applications in oceanography use both CF and ACDD



# CF Metadata Conventions

- **Global attributes :**
  - Long list of metadata fields to describe file content, authors, institutions, acknowledgments, use restrictions, etc.
  - Fields for overall time and space extent covered by the data (time, lat/lon, depth ranges)
- Assumption that data are located in space/time
  - “**Coordinate variables**” contain time, latitude, longitude, depth/height (or pressure)
  - Regular data variables are linked to their coordinates (several methods exist)
- Physical quantity inside a variable is defined via (a variable-level) attribute called “**standard\_name**” (i.e. not the name of the variable itself)
  - List of names, their definitions and canonical units: <http://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html>
- Units for a variable are defined via an attribute called “**units**”
  - Use only those that the software “UDUNITS” understands, cf.: <https://docs.unidata.ucar.edu/udunits/current/>

# Practice: What are the “global attributes” on this page?

```
title           = 'Processed Data from PIES Instruments (Pressure-Sensing Inverted Echo Sounders)
                  Collected in the Sri Lankan Boundary Currents between 2014 and 2019'
summary         = 'In 2014/15, PIES instruments (Pressure-Sensing Inverted Echo Sounders) were
                  deployed at four sites that span the boundary currents around Sri Lanka. Two sites
                  are east, and two south of Sri Lanka. PIES are seafloor instruments that make
                  observations of the water pressure and the travel time of a vertically transmitted
                  acoustic signal. Data from these instruments are reported here, in a processed version
                  that is on a common time grid with sensor drifts removed, and tidal signals separated
                  out of the pressure records. The scientific rationale for collecting these data is to
                  monitor the strengths of the boundary current flows.'
```

```
publisher_name  = 'Matthias Lankhorst'
publisher_url   = 'https://orcid.org/0000-0002-4166-4044'
contributor_name = 'Nilanthi Priyadarshani, Uwe Send'
contributor_role = 'principal investigator, principal investigator'
time_coverage_start = '2014-11-15T08:00:00Z'
time_coverage_end   = '2019-11-03T04:00:00Z'
```

```
TIME [ 23695.083 23695.167 23695.250 23695.333 ... ]
  Size: 21767x1
  Dimensions: TIME
  Datatype: double
  Attributes:
    units = 'days since 1950-01-01T00:00:00Z'
    standard_name = 'time'
    axis = 'T'

SEAFLOOR_PRESSURE [ 604.135 604.134 604.134 604.134 ... ]
  Size: 4x21767 [ 3889.274 3889.274 3889.273 3889.273 ... ]
  Dimensions: NSTATION, TIME [ 645.442 645.442 645.442 645.442 ... ]
  Datatype: double [ 4408.656 4408.656 4408.656 4408.655 ... ]
  Attributes:
    standard_name = 'sea_water_pressure_at_sea_floor'
    units = 'dbar'
    comment = 'Sensor drift has been removed, and data have been low-pass filtered. Tides
              have been removed and are available in a separate variable.'
    ancillary_variables = 'SEAFLOOR_PRESSURE TIDES'
    coordinates = 'LONGITUDE LATITUDE DEPTH TIME SITE_NAME'
```

# Practice: What defines physical quantities on this page?

```
title = 'Processed Data from PIES Instruments (Pressure-Sensing Inverted Echo Sounders)
       Collected in the Sri Lankan Boundary Currents between 2014 and 2019'
summary = 'In 2014/15, PIES instruments (Pressure-Sensing Inverted Echo Sounders) were
          deployed at four sites that span the boundary currents around Sri Lanka. Two sites
          are east, and two south of Sri Lanka. PIES are seafloor instruments that make
          observations of the water pressure and the travel time of a vertically transmitted
          acoustic signal. Data from these instruments are reported here, in a processed version
          that is on a common time grid with sensor drifts removed, and tidal signals separated
          out of the pressure records. The scientific rationale for collecting these data is to
          monitor the strengths of the boundary current flows.'
```

```
publisher_name = 'Matthias Lankhorst'
publisher_url = 'https://orcid.org/0000-0002-4166-4044'
contributor_name = 'Nilanthi Priyadarshani, Uwe Send'
contributor_role = 'principal investigator, principal investigator'
time_coverage_start = '2014-11-15T08:00:00Z'
time_coverage_end = '2019-11-03T04:00:00Z'
```




```
TIME [ 23695.083 23695.167 23695.250 23695.333 ... ]
Size: 21767x1
Dimensions: TIME
Datatype: double
Attributes:
  units = 'days since 1950-01-01T00:00:00Z'
  standard_name = 'time'
  axis = 't'
```


```
SEAFLOOR_PRESSURE [ 604.135 604.134 604.134 604.134 ... ]
Size: 4x21767 [ 3889.274 3889.274 3889.273 3889.273 ... ]
Dimensions: NSTATION, TIME [ 645.442 645.442 645.442 645.442 ... ]
Datatype: double [ 4408.656 4408.656 4408.656 4408.655 ... ]
Attributes:
  standard_name = 'sea_water_pressure_at_sea_floor'
  units = 'Pa'
  comment = 'Sensor drift has been removed, and data have been low-pass filtered. Tides
            have been removed and are available in a separate variable.'
  ancillary_variables = 'SEAFLOOR_PRESSURE TIDES'
  coordinates = 'LONGITUDE LATITUDE DEPTH TIME SITE_NAME'
```

# CF: More on Coordinate Variables

Different options to provide a **variable** that is a **coordinate variable** :

From previous example:

1. Its dimension has the same name as the variable itself. 
2. It has a “standard\_name” attribute that is known to be geospatial information, such as “latitude”, “longitude”, “time”, “depth”, “height”, “pressure”. 
3. It has an attribute called “axis”, with content “X”, “Y”, “Z”, or “T”. 

```
TIME   
Size: 21767x1  
Dimensions: TIME  
Datatype: double  
Attributes:  
  units = 'days since 1950-01-01T00:00:00Z'  
  standard_name= 'time'  
  axis = 'T'
```

(Note: Not all of these options need to be present!)

# CF: Even More on Coordinate Variables

Data (in regular variables) are linked to the coordinate variables (spatiotemporal) by any of the following:

1. A dimension that is also the name of both a coordinate variable and its dimension,
2. An attribute called “coordinates” that lists names of coordinate variables (example on right has no lat/lon coordinates for option 1. because they do not vary independently).

From previous example:

```
TIME
Dimensions: TIME
Attributes:
  units          = 'days since 1950-01-01T00:00:00Z'
  standard_name = 'time'
  axis          = 'T'

LATITUDE
Dimensions: NSTATION
Attributes:
  units          = 'degree_north'
  standard_name = 'latitude'
  axis          = 'Y'
```

[and similar for LONGITUDE and DEPTH coordinate variables]

```
SEAFLOOR_PRESSURE
Dimensions: NSTATION, TIME
Attributes:
  standard_name = 'sea_water_pressure_at_sea_floor'
  units        = 'dbar'
  coordinates  = 'LONGITUDE LATITUDE DEPTH TIME SITE_NAME'
```



# Take-Home Message:

Coordinate variables locate data in X, Y, Z, T.

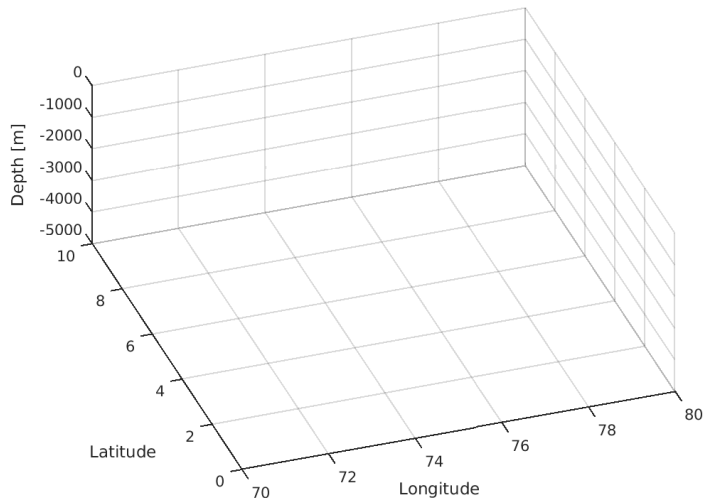
Attributes “standard\_name” and “units” define physical quantity and units of measure.

# Practice: Define Coordinates and Variable Dimensions!

**Use case:** Output from a numerical ocean model on a 4-d (X-Y-Z-T) grid

Variable **LATITUDE** (Dimensions: ???)

...other coordinate variables ???



Variable **SALINITY** (Dimensions: ???)

# Practice: Define Coordinates and Variable Dimensions!

**Use case:** Output from a numerical ocean model on a 4-d (X-Y-Z-T) grid

Variable **LATITUDE** (Dimensions: LATITUDE)

Variable **LONGITUDE** (Dimensions: LONGITUDE)

Variable **DEPTH** (Dimensions: DEPTH)

Variable **TIME** (Dimensions: TIME)

Variable **SALINITY** (Dimensions: LATITUDE x LONGITUDE x DEPTH x TIME)

# Practice: Define Coordinates and Variable Dimensions!

**Use case:** Output from a glider that travels along a path through space

Variable **LATITUDE** (Dimensions: ???)

...other coordinate variables ???

Variable **SALINITY** (Dimensions: ???)



# Practice: Define Coordinates and Variable Dimensions!

**Use case:** Output from a glider that travels along a path through space

Variable **LATITUDE** (Dimensions: TIME)

Variable **LONGITUDE** (Dimensions: TIME)

Variable **DEPTH** (Dimensions: TIME)


Variable **TIME** (Dimensions: TIME)

Variable **SALINITY** (Dimensions: TIME)

```
SALINITY:coordinates = "LATITUDE LONGITUDE DEPTH TIME"
```

# CF: featureType

Current versions of CF have (optional!) specifications for discrete sampling geometries that describe the “logic” in dimensionality and coordinates for different situations.

- Global attribute “**featureType**” defines what the sampling geometry is.
  - Proper use of the “coordinates” attribute for variables links each variable to its spatiotemporal coordinates
  - If there are multiple instances of a feature (e.g. multiple glider trajectories) in one file, there should be a coordinate variable that is designated as the ID for each instance, through the attribute “cf\_role”
  - Representation via ragged (compressed) arrays exist
  - Current definitions: point, timeSeries, trajectory, profile, timeSeriesProfile, trajectoryProfile
- 

# CF: featureType “trajectory”

Example: data from an underwater glider

All data collected along a trajectory in 3d-space, with the vehicle traveling along that trajectory as time progresses.

**Dimensions:** TIME

**Coordinate variables and their dimensions:**

TIME(TIME), DEPTH(TIME), LATITUDE(TIME), LONGITUDE(TIME)

**Data variables and their dimensions:**

DATA(TIME)



# CF: featureType “timeSeries”

Example: seafloor pressure data from a PIES instrument

All data collected at the same location in space as a time series.

**Dimensions:** TIME

**Coordinate variables and their dimensions:**

TIME(TIME), DEPTH(1), LATITUDE(1), LONGITUDE(1)

**Data variables and their dimensions:**

DATA(TIME)





# CF: featureType “profile”

Example: a CTD cast or an Argo float profile

All data collected at the same time and horizontal location, but as a function of depth.

**Dimensions:** DEPTH

**Coordinate variables and their dimensions:**

TIME(1), DEPTH(DEPTH), LATITUDE(1), LONGITUDE(1)

**Data variables and their dimensions:**

DATA(DEPTH)



# CF: featureType “timeSeriesProfile”

Example: repeat CTD casts at the same station, or a mooring with multiple sensors in the water column

Multiple profiles collected after one another at the same location.

**Dimensions:** TIME, DEPTH

**Coordinate variables and their dimensions:**

TIME(TIME), DEPTH(DEPTH), LATITUDE(1), LONGITUDE(1)

**Data variables and their dimensions:**

DATA(TIME, DEPTH)



# Take-Home Message:

“featureType” describes discrete sampling geometries and the corresponding linkage/dimensionality of variables and their coordinates.

# Practice: What Metadata is Useful?

(Students suggest metadata to address questions about a dataset)

What?	Who?	Why?	How?	Where?	When?

## Take-Home Messages:

- Need metadata to understand the data!
- NetCDF files contain variables, dimensions, and attributes.
- Coordinate variables for X, Y, Z, T
- Attributes “standard\_name” and “units” define physical quantity in a variable, and its units
- Attribute “featureType” to describe sampling geometries (profiles, time series) with different coordinate “logic”

# Practice: Inspect Contents of a netCDF File

(Read metadata and data of an example file)