SEVENTH FRAMEWORK PROGRAMME

Capacities Specific Programme

Research Infrastructures

Project acronym: **EUFAR**

Project full title: **European Facility for Airborne Research in Environmental and Geo-sciences**

Proposal no.: **227159**

EUFAR FP7

N6SP – Standards and Protocols

DN6.1.3A3 – Version 1.0 of NetCDF data and metadata protocols

Stefanie Holzwarth, Beneficiary-No. 6, DLR

Matt Freer, Beneficiary-No. 1, MétéoFrance

March 2011

1 Introduction

NetCDF, originally developed by Unidata, is currently in wide use by the atmospheric science community. It is a binary, array-oriented format, and is self-describing, architecture independent, compatible with multi-rate data and has well-developed interfaces for many programming languages including C++, FORTRAN, Python, Matlab, and IDL. EUFAR N6SP recommends NetCDF as the primary storage format for airborne data due to its flexibility in handling large, complex datasets, and its built-in metadata capabilities. See http://www.unidata.ucar.edu/software/netcdf/ for more information.

For data in NetCDF format, EUFAR N6SP and N7DB recommend that at the very minimum, users follow the Climate and Forecast (CF) NetCDF conventions found at http://cf-pcmdi.llnl.gov/. However, in this document, we propose a set of conventions which extend the CF conventions and incorporate features from the NCAR EOL-RAF metadata conventions. Also included in the present conventions are data discovery fields proposed by the US-based Interagency Working Group for Airborne Data and Telemetry Systems (IWGADTS). We highly recommend that the EUFAR N6SP NetCDF conventions outlined in this document are followed whenever possible.

2 NetCDF File Structure

2.1 Dimensions

For airborne instrument data, the first dimension should always be the time dimension. We recommend the time dimension be named 'Time' for files with a single time. Multiple time dimensions are allowed per file, and they should be labeled properly, with the 1Hz time dimension named 'Time', and other sampling rates defined using numeric suffixes - for example, 'Time25hz' for the 25Hz signal, or 'Time100hz' for the 100Hz signal, etc. For time dimensions slower than 1hz, the 'sec' suffix should be used: 'Time5sec', 'Time10sec', etc.

Additional dimensions will represent vector length of a dataset, for example, size distributions from 1D/2D probes, lidar/radar range bins, etc.

2.2 Time Variable

The time variable should follow the specifications described in the CF conventions, section 4.4 (http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.4/ch04s04.html). From this convention, the time variable *must* include a 'units' attribute in the following format:

'x since *y'*

where *x* is the unit (seconds, milliseconds, hours, days, etc) and *y* is an ISO8601 string specifying the reference date/time. For example:

'seconds since 1970-01-01 00:00:00Z'
'days since 2011-03-01 00:00:00Z'

It is recommended, where feasible, to name this variable after the corresponding dimension, i.e., 'Time'. For files with multiple time, a time coordinate should be provided for each corresponding time dimension, and named accordingly. For example, the corresponding time coordinate for the Time25hz dimension would be 'Time25hz'.

2.3 Variables

Variables can be named as the user pleases, following CF conventions. However, we suggest following a naming structure to quickly guide users to the relevant data. The naming structure should be descriptive, human readable, and brief, since the objective is simply to guide the user to the correct variable but not completely describe the variable. The majority of variable details can be stored in the associated metadata. Thus, we propose the following structure:

`[measurement_family]_[measurement_detail]_[source/instrument/number]_[sampling_rate]`

`[measurement_family]` represents the broad measurement type or category, e.g. humidity, temperature, pressure, number_concentration, velocity, etc. Abbreviations may be used where clear and unambiguous, e.g. temp, pres, hum, num_conc, vel, etc.

`[measurement_detail]` is used to further clarify the type of measurement represented by the variable. For measurements where further detail is not necessary, this parameter may be skipped. Examples include:

- temperature_**static**_...
- temperature_**potential**_...
- humidity_**relative**_...
- velocity_**tas**_...
- diameter_**mean**_...

`[source/instrument/number]` describes the source of the measurement. If multiple probes of the same type are present, a probe number can be added to the field. Not required for all variables. Examples:

- humidity_relative_**capacitive**_...
- diameter_mean_**fssp**_...
- diameter_mean_**fssp2**_...
- latitude_**ins**_...
- latitude_**gps**_...

`[sampling_rate]` specifies the rate at which the variable is sampled. This parameter is optional, but useful for files that may contain data with different sampling rates. For rates 1 Hz or faster, the numeric sampling value should be used, with a 'hz' suffix. For rates slower than 1 Hz, the numeric amount of seconds between each sample should be used,

with a 'sec' suffix. Decimal points in non-integer rates should be marked with a 'd' character:

- latitude_gps_**10hz** (sampled at 10 Hz)
- latitude_gps_**5d5hz** (sampled at 5.5 Hz)
- latitude_gps_**1hz** (sampled at 1 Hz)
- latitude_gps_**10sec** (sampled at 0.1 Hz)

3 Metadata

EUFAR proposes to extend the CF conventions vocabulary with features from the RAF conventions and IWGADTS data discovery fields. The proposed list of global and variable metadata attributes can be found in this section.

3.1 Global Attributes

Name	Description	Usage
Conventions	Comma separated list of conventions used (CF, RAF, EUFAR)	**Required**
title	Description of the dataset	
source	Aircraft/satellite/instrument name	
institution	Name of institution or company providing the data	
project	Project name	**Recommended**
date_created	Date of creation of dataset (ISO 8601 string)	
geospatial_lat_min geospatial_lat_max geospatial_lon_min geospatial_lon_max geospatial_vertical_min geospatial_vertical_max geospatial_vertical_positive geospatial_vertical_units	Describes geospatial coverage of the data	
time_coverage_start time_coverage_end time_coverage_duration	Describes temporal coverage of the data (ISO8601 string)	
history	List of modifications to the original data	**Optional**
references	Published or web-based references that describes data or processing methods	
comment	Other relevant comments concerning the data set	

3.2 Variable Attributes

Name	Description	Source	Usage
units	Description of variable units (best practice is to follow udunits names)	CF	**Required**
_FillValue	Specifies the fill value used for missing data.	CF	
long_name	Descriptive name of the variable in English	CF	**Recommended (where applicable)**
standard_name	Standard name for the variable drawn from the CF standard names table	CF	
valid_range -or- valid_max & valid_min	Defines valid range of values for variable	CF	
SampledRate	Original data sampling rate in Hz. This applies to raw data only, not derived parameters.	RAF	
Category	Name(s) of probe category – comma separated list (see examples in section 4.1)	RAF	
CalibrationCoefficient	Coefficients used to convert analog channels to digital counts (raw variables only)	RAF	
InstrumentLocation	Location of instrument on aircraft in human-readable terms, e.g. left wing pod, upper starboard fuselage port, etc. (raw variables only)		
InstrumentCoordinates	x,y,z location of instrument on aircraft x – longitudinal axis, positive to nose of aircraft y – transverse axis, positive to right wing z – vertical axis, positive towards bottom of aircraft		
Dependencies	List of input variables used to produce this derived variable	RAF	
Processor	List of EGADS processors used to produce this variable		**EGADS**
Comments	Free-form string field for variable comments		**Optional**
ancillary_variables	List of other variables in file which further describe current variable (see: CF Conventions Section 3.4)	CF	
flag_values, flag_masks, flag_meanings	Intended to make variables that contain flag values self describing (see: CF Conventions Section 3.5).	CF	
long_name_<xx>	Descriptive name of the variable in language <xx>, where <xx> is the language code (from ISO 639-1). Examples: French long name would be long_name_fr; German long name would be long_name_de; etc.		

3.3 Data discovery

Data discovery is a scheme to create a standard vocabulary which can be used to find essential data in a large NetCDF dataset. In a file with multiple measurements of the same type and the right data discovery fields,

- programs can easily find the right variable for a given task
- users can decipher cryptic variable names
- 'best' or 'recommended' variables can be highlighted, showing users which to use

IWGADTS has developed a draft recommendation for data discovery variables in aircraft NetCDF files. They recommend a list of parameters to include for data discovery in the global attributes, using a namespace to differentiate the data discovery variables from other global attributes. The following extract of NetCDF global attributes is an example of data discovery:

```
:reference_latitude = "GPS_latitude" ;
:reference_longitude = "GPS_longitude" ;
:reference_altitude = "pres_altitude" ;
```

EUFAR recommends following a similar scheme. The following are the attribute names recommended to be used for EUFAR data discovery. When added to a NetCDF file, the namespace 'reference' should be added to each attribute, as shown in the above example.

Highly recommended:

- latitude
- longitude
- altitude

Recommended:

- wind_speed
- upward_air_velocity
- air_temperature
- platform_speed_wrt_air
- dew_point_temperature
- air_pressure
- platform_roll_angle
- platform_pitch_angle
- platform_orientation
- ground_speed_east_west
- ground_speed_north_south

4 Annex

4.1 Example Category Types

In this section is a non-exhaustive list of category types intended for use in the variable attribute “Category.” These are primarily meant to help users and programs differentiate between different types of variables.

- Time
- Position
- Aircraft State
- Atmos. State
- Analog
- Uncorr Raw
- PMS Probe
- Housekeeping
- Aerosol
- Thermodynamic
- Condensed Water
- Chemistry
- Radiation
- Wind
- Non-Standard