2.2. Overview of the Netcdf formats

2.2.1. What is NetCDF ?

NetCDF OCO uses the NetCDF (Network Common Data Form) system, a set of software libraries and machine-independent data formats.

NetCDF (Network Common Data Form) is an interface for array-oriented data access and a library that provides an implementation of the interface. The NetCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data. The NetCDF software was developed at the Unidata Program Centre in Boulder, Colorado. The freely available source can be obtained as a compressed tar file or a zip file from Unidata or from other mirror sites.

- Ucar web site address : http://www.ucar.edu/ucar
- NetCDF documentation : http://www.unidata.ucar.edu/packages/netcdf/index.html

2.2.2. NetCDF OCO

Our implementation of NetCDF is based on the community-supported Climate and Forecast (CF) specification, which supplies a standard vocabulary and some metadata conventions.

NetCDF has several more restrictions than the CF standard. These are intended to make it easier to share data, to make it simpler to aggregate data from multiple sites, and to ensure that data can be created and understood by the basic NetCDF utilities.

The requirements are drawn almost exclusively from the NetCDF Style Guide:

- Units are compliant with CF/COARDS/Udunits ;
- The time variable is encoded as recommended by COARDS and CF.
- Parameters are given standard names from the CF table
- Where time is specified as an attribute, the ISO8601 international standard representation is used : yyyy-mm-ddThh:mi:ssZ (i.e. year – month – day T hour : minute : second Z).

Other requirements are:

- All the date and time are given in UTC time, universal time coordinates.
- The _FillValue associated to each variable is not prescribed. However it must be a value out of the span defined by valid_min, valid_max. The reference table 3 suggests a fill value for each type.
- The NetCDF file format should be (in order of preference):
  - NetCDF4 (“netCDF-4” or “netCDF-4 classic model”)
  - NetCDF3 (“classic” or “64 bits”).

For more information on CF, COARDS, NetCDF, Udunits, and ISO8601 see:

- NetCDF: http://www.unidata.ucar.edu/software/netcdf/docs/BestPractices.html
- Udunits: http://www.unidata.ucar.edu/software/udunits/
- CF: http://cf-pcmdi.llnl.gov/
- COARDS: http://www.ferret.noaa.gov/noaa_coop/coop_cdf_profile.html
- ISO8601: http://en.wikipedia.org/wiki/ISO_8601

2.2.3. File structure

NetCDF OCO formats are divided in 4 sections:

- Dimensions and definitions,
- General attributes, meta-data information,
- Coordinate variables,
- Data section as variables.

General attributes, meta-data information

The general attributes section of a NetCDF file contains metadata that describes the contents of the file overall, and allows for data discovery. All fields should be human-readable, and should be of character type, not numeric, even if the information content is a number. It is recommended that all of these attributes be used and contain meaningful information (even “n/a” – not applicable or not available) unless there are technical reasons rendering this impossible.

Global attributes can be thought of as conveying five kinds of information:

- What: what are the data in this dataset;
- Where: the spatial coverage of the data;
- When: the temporal coverage of the data;
- Who: who produced the data;
- How: how were the data produced and made available.

Coordinate variables

The coordinate variables orient the data in time and space. For this purpose, they have an "axis" attribute defining that they point in X, Y, Z, and T dimensions.

Each coordinate variable must be strictly monotonic and must not have any missing data (thus it has no _FillValue attribute).

The <vertical> variable (depth, height, level or layer) may be positive in either upward or downward direction, which is defined in its "positive" attribute.

The level variable’s standard_name could be “ocean_sigma_coordinate” or “ocean_s_coordinate”.

The level variable defines an **ocean sigma coordinate** as :

- z(n, k, j, i) = eta(n, j, i) + sigma(k) * (depth(j, i) + eta(n, j, i))

where z(n, k, j, i) is height, positive upwards, relative to ocean datum (e.g. mean sea level) at gridpoint (n, k, j, i); eta(n, j, i) is the height of the ocean surface, positive upwards, relative to ocean datum at gridpoint (n, j, i); sigma(k) is the dimensionless coordinate at vertical gridpoint(k); and depth(j, i) is the distance from ocean datum to sea floor (positive value) at horizontal gridpoint (j, i).

The level variable defines an **ocean s coordinate** as :

- z(n,k,j,i) = eta(n,j,i)*(1+s(k)) + depth_c*s(k) + (depth(j,i)-depth_c)*C(k)
- C(k) = (1-b)*sinh(a*s(k))/sinh(a) + b*[tanh(a*(s(k)+0.5))/(2*tanh(0.5*a)) - 0.5]

where z(n,k,j,i) is height, positive upwards, relative to ocean datum (e.g. mean sea level) at gridpoint (n,k,j,i), eta(n,j,i) is the height of the ocean surface, positive upwards, relative to ocean datum at gridpoint (n,j,i), s(k) is the dimensionless coordinate at vertical gridpoint (k), and depth(j,i) is the distance from ocean datum to sea floor (positive value) at horizontal gridpoint (j,i). The constants a, b, and depth_c control the stretching.

In the OCO spectra file format theta and frequency are used as pseudo-coordinate variables.

Data variables

Data variables contain the actual data. The variable names are not standardized, but the standard_name attribute of each variable is.

The standard name table is located at http://cf-pcmdi.llnl.gov/documents/cf-standard-names/standard-name-table/current/cf-standard-name-table.html. Standard names by themselves are not always sufficient to describe a quantity. For example, a variable may contain data to which spatial or temporal operations have been applied. Or the data may represent an uncertainty in the measurement of a quantity. These quantity attributes are expressed as modifiers of the standard name. Modifications due to common statistical operations are expressed via the cell_methods attribute (see reference table 4). Other types of quantity modifiers are expressed using the optional modifier part of the standard_name attribute (see reference table 5).

Variable attributes

The attributes “scale_factor” and “add_offset” are optional.

At the current time, the OCO format does not provide for packing data, however a simple packing may be achieved through the use of the optional attributes scale_factor and add_offset.

The aim of this packing is to store data with a type which is smaller than the original type. For example, a Netcdf file containing float values stored with short type (by using scale_factor and add_offset attributes) will be smaller than another file storing the same float values with float type.

A simple way to

real_value = stored_value * scale_factor + add_offset

If both “scale_factor” and “add_offset” are presents, they should not be equal to, respectively, 1 and 0.

When data to be packed contains missing values the attributes that indicate missing values (“_FillValue”, “valid_min”, “valid_max”) must be of the same data type as the packed data.

The missing values (“_FillValue”, “valid_min”, “valid_max”) of a variable with “scale_factor” and/or “add_offset” attributes are interpreted relative to the values stored in the netCDF file.

2.3. Netcdf OCO straight grid 1.4

The OCO straight grid file format is in compliance with the "Mersea IP guidelines for Mersea Data Provider – Gridded Data Product Format" ref MERSEA-WP06-CLS-UMAN-001B_gridded_data_format.doc.

2.3.1. Dimensions and definitions

Name	Value	Comment
time	Unlimited	Number of time steps. This dimension must be set to unlimited.
depth(*)	e.g. : 1	If applicable, for oceanographic data Number of steps along vertical axis.
height(*)	e.g. : 1	If applicable, for meteorological data Number of steps along vertical axis
level(*)	e.g. : 1	If applicable, for oceanographic data with sigma levels Number of steps along vertical axis
layer(*)	e.g. : 32	If applicable, for oceanographic data Number of layers along vertical axis.
latitude	e.g. : 251	Number of latitude steps of the grids
longitude	e.g. : 201	Number of longitude steps of the grids

(*): If the Netcdf file contains only 2D data (along latitude and longitude), none of the vertical dimension (depth, height, level or layer) is required. If the Netcdf file contains at least one 3D data (along latitude, longitude and vertical), one and only one of the vertical dimension (depth, height, level or layer) is required.

2.3.2. General attributes, meta-data information

Name	Value	Comment
WHAT		
data_type	"OCO straight grid"	Type of data contained in the file (reference table 2).
format_version	"1.3"	File format version
title	e.g. : "Meteo-France Aladin analysis and forecast"	A descriptive title for the dataset i.e. a product short name or original product name (habitual reference for users, succinct description of what is in the dataset)
Conventions	"CF-1.5"	Name of the conventions followed by the dataset. *Please note the first capital letter for COARDS compatibility.*
netcdf_version	e.g. : "3.6"	Netcdf version used for the data set
product_name	e.g. : "ww3.2008082011.nc"	Product name for data provider
creation_date	e.g. : "2006-04-11T08:35:00Z"	File creation date (UT)
product_version	e.g. : "1.0"	Release number of the data file
software_version	e.g. : "MFS analysis and forecasting system v5"	Software version (i.e. the software used for the production of the data)
references	e.g. : "http://www.previmer.org/"	Web-based reference that describe the data or methods used to produce it.
history	e.g. : "2006-04-11T08:35:00Z : Creation\n 2008-03-23T11:56:12Z : Merging variable : poc"	Provides an audit trail for modifications to the original data. It should contain a separate line for each modification, with each line beginning with a timestamp, and including modification name and optional modification arguments.
source	e.g. : "MARS V5.14"	The method of production of the original data. If it was model-generated, source should name the model and its version, as specifically as could be useful. If it is observational, source should characterize it (e.g., "surface observation" or "radiosonde")
comment	e.g. : "Any free-format text is appropriate."	Miscellaneous information about the data or methods used to produce it. Any free-format text is appropriate.
WHERE		
area	e.g. : "Western Europe"	Geographical coverage
easting	"longitude"	
northing	"latitude"	
grid_projection	e.g. : "Mercator"	Describe the grid use to store the data
southernmost_latitude	e.g. : "35.5°"	Value between -90° and 90°
northernmost_latitude	e.g. : "53.8°"	Value between -90° and 90°
latitude_resolution	e.g. : "0.1°"	Latitude step between two consecutive latitude values
westernmost_longitude	e.g. : "-11°"	Value between -180° and 180°

Name	Value	Comment
easternmost_longitude	e.g. : "14°"	Value between -180° and 180°
longitude_resolution	e.g. : "0.1°"	Longitude step between two consecutive longitude values
minimum_altitude (*)	e.g. : "-5 m"	Negative value for depth (Zmin= Zmax if no variation)
maximum_altitude (*)	e.g. : "-5 m"	Negative value for depth (Zmin= Zmax if no variation)
altitude_resolution (*)	e.g. : "n/a"	If applicable, vertical distance between two consecutive altitude values
WHEN		
start_date	e.g. : "2006-03-01T00:00:00Z"	Start date of the data (UT)
stop_date	e.g. : "2006-03-05T23:59:29Z"	Final date of the data (UT)
field_type	e.g. : "6-hourly"	If applicable, describe the period between two consecutive timestamp hourly/daily/weekly/monthly/yearly n-hourly/n-daily/n-weekly monthly climatology permanent
run_time	e.g. : "2006-03-01T00:00:00Z"	If applicable, date of the model run when the data were calculated.
forecast_range	e.g. : "2-days forecast"	If applicable, time range of forecast
forecast_type	e.g. : "forecast"	If applicable, "hindcast" or "nowcast" or "forecast"
WHO		
institution	e.g. : "Meteo-France"	Data provider name
institution_references	e.g. : "http://www.meteo.fr/"	Web-based reference for data provider, the place to find all information upon the data set
data_centre	"IFREMER OCO DATA CENTER"	Data centre in charge of this data file
data_centre_references	"http://www.previmer.org/"	Web-based reference for data centre
contact	e.g. : "cdoco-exploit@ifremer.fr"	User desk e-mail
HOW		
distribution_statement	e.g. : "Data restrictions: for registered users only"	Text like "Approved for public release. Distribution unlimited" or "Data restrictions: for registered users only" or link to the place where the rule is described
operational_status	e.g. : "operational"	Statement like : "under development" or "operational" or "experimental"
quality_index	e.g. : "1"	A code value: 0 unknown quality 1 excellent (no known problems, regular quality checking) 2 probably good (occasional problems, validation phase) 3 extremely suspect (frequent problems)

(*) : This attribute is mandatory only if one of the vertical dimension (depth, height, level or layer) is present. If not, the attribute is optional.

2.3.3. Coordinate variables

Type Name(Dimensions), Attributes	Comment
double time(time) long_name = <X_1> standard_name = "time" units = <X_2> conventions = <X_3> axis = "T"	Time stamps of the grid which can be expressed as : • seconds since a reference date <X_1> = "time in seconds (UT)" <X_2> = "seconds since *1900-01-01T00:00:00Z*" <X_3> = "relative number of seconds with no decimal part" e.g. : 320577240.0 : July 25 2001 19:14:00 • julian days since a reference date (the integer part represents the day, the decimal part represents the time of the day) <X_1> = "julian day (UT)" <X_2> = "days since *1900-01-01T00:00:00Z*" <X_3> = "relative julian days with decimal part (as parts of the day)" e.g. : 37095.801388 : July 25 2001 19:14:00 Date and time are in universal time. It is allowed to use another date as the reference date.
<type> latitude(latitude) long_name = "latitude" standard_name = "latitude" units = "degree_north" valid_min = <X_1> valid_max =<X_2> axis="Y"	Latitudes (degree north) of the grid. <type> : double, float <X_1> : depends on <type> (-90. for double, -90.f for float) <X_2> : depends on <type> (90. for double, 90.f for float) e.g. : 44.4991 for 44° 29' 56.76'' N

Type Name(Dimensions), Attributes	Comment
<type> longitude(longitude) long_name = "longitude" standard_name = "longitude" units = "degree_east" valid_min = <X_1> valid_max = <X_2> axis="X"	Longitudes (degree east) of the grid. <type> : double, float <X_1> : depends on <type> (-90. for double, -90.f for float) <X_2> : depends on <type> (90. for double, 90.f for float) e.g. : 16.7222 for 16° 43' 19.92'' E
float depth(depth) long_name = "depth" standard_name = "depth" units = "m" axis="Z" positive="down"	If applicable, for oceanographic data. If and only if depth dimension is defined.
float height(height) long_name = "height" standard_name = "height" units = "m" axis="Z" positive="up"	If applicable, for meteorological data. If and only if height dimension is defined.
float level(level) long_name = "sigma level" standard_name = "ocean_sigma_coordinate" or "ocean_s_coordinate" formula_terms = <X_1> axis="Z" positive="up"	If applicable, for oceanographic data with sigma levels. If and only if level dimension is defined. For "ocean_sigma_coordinate" <X_1> : "sigma: *var1* eta: *var2* depth: *var3*" *var1*, *var2*, *var3* are the names of the variables matching respectively sigma, eta, depth in the equation z=eta+sigma(depth+eta) For "ocean_s_coordinate" <X_1> : "s: var1 eta: var2 depth: var3 a: var4 b: var5 depth_c: var6" *var1*, *var2*, *var3, var4, var5, var6* are the names of the variables matching respectively s, eta, depth, a, b, depth_c in the equations : z(n,k,j,i) = eta(n,j,i)*(1+s(k)) + depth_c*s(k) + (depth(j,i)-depth_c)*C(k) C(k) = (1-b)*sinh(a*s(k))/sinh(a) + b*[tanh(a*(s(k)+0.5))/(2*tanh(0.5*a)) - 0.5]
int layer(layer) long_name = "layer" standard_name = "model_level_number" units = "layer" axis="Z" positive="down"	If applicable, for oceanographic data. If and only if layer dimension is defined.

2.3.4. Data variables

Some examples of grid variables are listed in reference table 6.

Type Name(Dimensions), Attributes	Comment
<type> <var2d>(latitude,longitude) long_name = <X_1> standard_name = <X_2> units = <X_3> _FillValue = <Y> valid_min = <X_4> valid_max = <X_5> scale_factor = <X_6> add_offset = <X_7>	<var2d> contains the values of a (lat, lon) parameter <type> : double, float, int, short, byte <X_1> : a free text which describes <var2d> <X_2> : a CF-1 standard text which describes <var2d> <X_3> : the CF-1 standard unit of <var2d> <Y> : Appropriate chosen value <X_4><X_5> : the min and max valid bounds for <var2d> <X_6> : an optional factor to apply on data <X_7> : an optional offset to add to data
<type> <var3d>(time,latitude,longitude) long_name = <X_1> standard_name = <X_2> units = <X_3> _FillValue = <Y> valid_min = <X_4> valid_max = <X_5> scale_factor = <X_6> add_offset = <X_7>	<var3d> contains the values of a (t, lat, lon) parameter <type> : double, float, int, short, byte <X_1> : a free text which describes <var3d> <X_2> : a CF-1 standard text which describes <var3d> <X_3> : the CF-1 standard unit of <var3d> <Y> : Appropriate chosen value <X_4><X_5> : the min and max valid bounds for <var3d> <X_6> : an optional factor to apply on data <X_7> : an optional offset to add to data

Type Name(Dimensions), Attributes	Comment
<type> <var4d>(time,<vertical>,latitude,longitude) long_name = <X_1> standard_name = <X_2> units = <X_3> _FillValue = <Y> valid_min = <X_4> valid_max = <X_5> scale_factor = <X_6> add_offset = <X_7>	<var4d> contains the values of a (t, z, lat, lon) parameter <type> : double, float, int, short, byte <vertical> : depth, height, level <X_1> : a free text which describes <var4d> <X_2> : a CF-1 standard text which describes <var4d> <X_3> : the CF-1 standard unit of <var4d> <Y> : Appropriate chosen value <X_4><X_5> : the min and max valid bounds for <var4d> <X_6> : an optional factor to apply on data <X_7> : an optional offset to add to data

3. Reference tables

3.1. Reference table 1 : quality control flag scale

n	Meaning
0	No QC was performed
1	Good data
2	Probably good data
3	Probably bad data
4	Bad data
5	Value changed
6	Not used
7	Not used
8	Interpolated value
9	Missing value

3.2. Reference table 2 : data type

The following table contains the list of acceptable contents for data_type attribute.

Name
OCO straight grid
OCO oriented grid
OCO spectra

3.3. Reference table 3 : proposal of fill values

The fill values are not prescribed. The following table suggests a fill value for each type.

Type	Fill value
double	-1.7+308
float	-3.4+38
int	-2147483648
short	-32768
byte	-128
char (for strings)	' ' (space)

3.4. Reference table 4 : cell methods

In the "Units" column, u indicates the units of the physical quantity before the method is applied.

cell_method	Units	Comment
point	u	The data value are representative of points in space or time (instantaneous).
sum	u	The data value are representative of a sum or accumulation over the cell
maximum	u	Maximum
minimum	u	Minimum
median	u	Median
mid_range	u	Average of maximum and minimum
mean	u	Mean (average value)
mode	u	Mode (most common value)
standard_deviation	u	Standard deviation
variance	u^2	Variance

3.5. Reference table 5 : standard name modifiers

In the "Units" column, u indicates the units of the physical quantity before the method is applied.

Modifier	Units	Comment
detection_minimum	u	The smallest data value which is regarded as a detectable signal
number_of_observations	1	The number of discrete observations or measurements from which a data value has been derived
standard_error	u	The uncertainty of the data value.
status_flag		Flag values indicating the quality or other status of the data values. The variable should have flag_values or flag_masks (or both) and flag meanings attributes to show and it should be interpreted. An explanation for the way to use this modifier can be found directly in the CF Metadata convention.

3.6. Reference table 6 : examples of variables definition

Variable	type	long name	standard_name	units	valid min	valid max
Arpege, Aladin						
fllat	float	Flux de chaleur latente	surface_upward_latent_heat_flux	W m-2	-10800000.f	10800000.f
flsen	float	Flux de chaleur sensible	surface_upward_sensible_heat_flux	W m-2	-6480000.f	6480000.f
flsolaire	float	Bilan du rayonnement de courtes longueurs d'onde	surface_net_upward_shortwave_flux	W m-2	-43200000.f	43200000.f
fltherm	float	Bilan du rayonnement de grandes longueurs d'onde	surface_net_upward_longwave_flux	W m-2	-10800000.f	10800000.f
hu2m	float	Humidite relative fraction a 2m.	relative_humidity		0.f	1.f
nebul	float	Nebulosite totale fraction	cloud_area_fraction		0.f	1.f
prlge	float	Precipitations liquide a grande echelle	large_scale_precipitation_amount	kg m-2	0.f	64800000.f
pmer	float	Pression au niveau de la mer	air_pressure_at_sea_level	Pa	80000.f	120000.f
t2m	float	Temperature de l'air a 2m.	air_temperature	K	220.f	330.f
u10m	float	Composante U du vent horizontal a 10m.	eastward_wind	m s-1	-100.f	100.f
v10m	float	Composante V du vent horizontal a 10m.	northward_wind	m s-1	-100.f	100.f
ustress	float	Première composante tension sol intégrée	surface_downward_eastward_stress	Pa	-21600.f	21600.f
vstress	float	Deuxième composante tension sol intégrée	surface_downward_northward_stress	Pa	-21600.f	21600.f
Radar HF Vigicote						
hcdt	float	Direction de projection du courant	direction_of_sea_water_velocity	degree	0.f	360.f
hcsp	float	Vitesse du courant radial	sea_water_speed	m s-1	-100.f	100.f
hcdt_error	float	Precision de la mesure radiale	error_on_sea_water_speed	m s-1	-100.f	100.f
ewct	float	Composante U du courant	eastward_sea_water_velocity	m s-1	-100.f	100.f
nsct	float	Composante V du courant	northward_sea_water_velocity	m s-1	-100.f	100.f
ewct_error	float	Precision de la composante U du courant	error_on_eastward_sea_water_velocity	m s-1	-100.f	100.f
nsct_error	float	Precision de la composante V du courant	error_on_northward_sea_water_velocity	m s-1	-100.f	100.f

"Unitless" quantities such as fraction are left empty (units=" ").