



SIST19

Toulouse



cnrs

La convention CF pour NetCDF (*Climate & Forecast*)

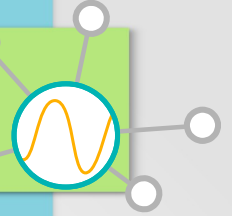
J. Sudre, M. Libes
SIST 2019

Jeudi 7 novembre 2019

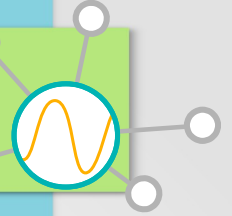
Observatoire Midi-Pyrénées de Toulouse



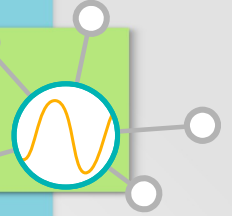
- Contexte – Objectifs
- Attributs des fichiers
- Attributs des données
- Représentation de l'échantillonnage discret (DSG)



- Il existe différentes conventions pour uniformiser les données plus ou moins adaptées selon les domaines :
 - **COARDS** : **C**onvention **C**ooperative **O**cean/**A**tmosphere **R**esearch **D**ata **S**ervice
 - **ACDD** : **A**tttribute **C**onvention for **D**ata **D**iscovery,
 - **CF** : Climate and Forecast Convention
 - CF est compatible avec COARDS
 - **SDN** : SeaDataNet Convention



- Les entetes netCDF permettent de fournir un ensemble de métadonnées avec des attributs permettant de :
 - nommer et décrire les variables et paramètres,
 - décrire la distribution des données dans l'espace et le temps
- Pour autant netCDF est assez ouvert et chaque auteur peut mettre des noms de variables, et des attributs différents pour décrire les données
- Le but de la convention CF... est de proposer une **standardisation des métadonnées embarquées dans les entetes netCDF de manière à les rendre interopérables**
 - *Ex: Erddap suit d'assez près la norme CF*



- La convention CF propose une définition plus rigoureuse des métadonnées nécessaires
 - plus particulièrement adaptée dans les domaines *climat, atmosphere et ocean*
 - Elle permet le traitement et le partage de fichiers NetCDF par une **standardisation de métadonnées** associées aux netCDF
- standardisation des :
- *noms de variables et Attributs des variables*
 - *quelques attributs globaux*
 - *Métadonnées descriptives*
 - *unités*
 - *Géoréférencement lat/long,*
 - *Zones et Régions géographiques*

objectifs

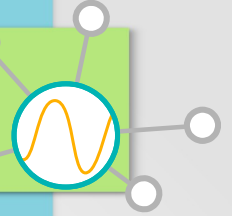
SIST



- Actuellement version 1.7 <http://cfconventions.org/>
- Plus particulièrement adaptée aux types de données *“Océan et Atmosphère”*
 - Chimie et dynamique de l'atmosphère
 - Cycle du carbone
 - Nuage
 - Hydrologie
 - Dynamique océanique
 - Radiation
 - Glace de mer
 - Surface

objectifs

SIST



- Standardisation des noms de variable
 - *air_pressure*
 - *sea_water_potential_temperature*
- Standardisation des régions géographiques
 - *Mediterranean_sea*
 - *atlantic_ocean*
 - <http://cfconventions.org/Data/standardized-region-list/standardized-region-list.html>
- Standardisation des zones naturelles
<http://cfconventions.org/Data/area-type-table/current/build/area-type-table.html>

Description des fichiers : entete global

SIST



- Dans le “header” netCDF : métadonnées “globales” fournissant les informations de base sur l'origine de la donnée

title	Qui a-t-il dans le fichier ?
*institution	Où a-t-il été produit ?
*source	Comment a-t-il été produit?(version du modèle, type d'instrument,etc.)
history	Protocole de traitement
*references	Publication de référence, documentation,...
*comment	Commentaires divers

Description des fichiers : entete global

SIST



- Dans le “header” netCDF : métadonnées “globales” on peut rajouter des informations utiles

license	<i>dataset is licensed under a Creative Commons Attribution 4.0 International license type CC-BY-NC https://creativecommons.org/licenses/by-nc/4.0/ ;</i>
citation	En cas de publication Avec un DOI
aknowledgement	

Description des fichiers : entete global

SIST



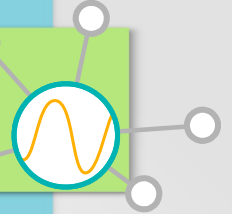
- Exemple: `ncdump -h v_gekco2_usec_20161231.nc`

// global attributes:

```
:conventions = "CF-1.6" ;
:title = "LEGOS Geostrophic - Ekman Velocity" ;
:dsd_entry_id = "GEKCO V2.0" ;
:references = "None" ;
:institution = "FR LEGOS" ;
:contact = "joel.sudre@legos.obs-mip.fr" ;
:gds_version_id = "v1.0-rev1.5" ;
:netcdf_version_id = "3.6.0-p1 of Nov 25 2005 11:29:18" ;
:creation_date = "Thu Nov 16 11:09:10 2017" ;
:product_version = "0.1" ;
:history = "Extracted from gekco2 v_gekco2_20161121.sav" ;
:grid_resolution = "0.25 degree" ;
:source_data = "MADT DUACS V15 - b: NCDC Blended"
:start_date = "20161231 UTC" ;
:start_time = "00:00:00 UTC" ;
:stop_date = "20161231 UTC" ;
```

Description des fichiers : dimensions

SIST



- Une variable peut avoir n'importe quel nombre de dimensions, y compris zéro,
- Les dimensions doivent toutes avoir des noms différents.
- COARDS recommande fortement de limiter le nombre de dimensions à quatre X,Y,Z,T, mais CF permet une plus grande flexibilité.
- Dimensions à afficher dans l'ordre T, puis Z, puis Y, puis X dans la définition CDL

```
dimensions:  
    time = UNLIMITED ; // (44640 currently)  
    nv = 2 ;  
    strlen = 20 ;  
char station_name(strlen) ;  
    :long_name = "station name" ;  
    :cf_role = "timeseries_id" ;  
float ws_040(time) ;  
    :_FillValue = NaNf ;  
    :standard_name = "wind_speed" ;  
    :axis = "T";
```

Description des données : entete des variables

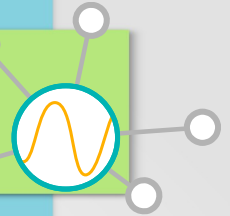
SIST



units	Unité international (degC, Pa, mbar,..) Ne pas utiliser psu, dB, Sv
standard_name	Identifier la donnée Ex : sea_surface_temperature
long_name	Appellation locale Ex : <i>Temperature de surface</i>
_FillValue et/ou missing_value	Valeur des données manquantes:999.9 etc.
valid_min et valid_max	Le min et le max de la variable
add_offset et scale_factor	D= add_offset+scale_factor*variable Ex : add_offset=0.f et scale_factor=1.f
flag_values et/ou flag_meaning	Si la donnée nécessite une valeur de masque ou « flag »

Description des données : entete des variables

SIST



- description du contenu et des unités de mesures pour chaque variable
- ***long_name*** : nom long descriptif "temperature"
- ***standard_name***: nom conventionné obligatoire dans CF
sea_water_temperature
 - *float psl(lat,lon) ;*
 - *:long_name = "mean sea level pressure" ;*
 - ***:units = "hPa" ;***
 - ***:standard_name = "air_pressure_at_sea_level" ;***
- ***Units*** : obligatoire
 - *Deca, hecto, kilo, mega, giga*
 - *Deci; centi, milli, micro*
 - *etc*

Description des données : standard_name

SIST

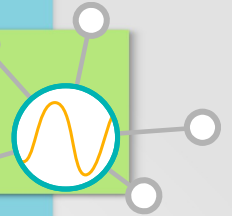


- Table des noms standards "*standard_name*" (V69, Oct. 2019)
 - <http://cfconventions.org/standard-names.html>
 - Fournit les noms standards dans les domaines ocean atmosphere et leurs unités

▼ sea_surface_skin_temperature	The sea surface skin temperature is the temperature measured by an infrared radiometer typically operating at wavelengths in the range 3.7 - 12 micrometers. It represents the temperature within the conductive diffusion-dominated sub-layer at a depth of approximately 10 - 20 micrometers below the air-sea interface. Measurements of this quantity are subject to a large potential diurnal cycle including cool skin layer effects (especially at night under clear skies and low wind speed conditions) and warm layer effects in the daytime.	K
▶ sea_surface_subskin_temperature		K
▼ sea_surface_temperature	Sea surface temperature is usually abbreviated as "SST". It is the temperature of sea water near the surface (including the part under sea-ice, if any). More specific terms, namely <code>sea_surface_skin_temperature</code> , <code>sea_surface_subskin_temperature</code> , and <code>surface_temperature</code> are available for the skin, subskin, and interface temperature. respectively. For the temperature of sea water at a particular depth or layer, a data variable of <code>sea_water_temperature</code> with a vertical coordinate axis should be used.	K
▶ sea_water_added_conservative_temperature		degree_C
▶ sea_water_added_potential_temperature		degree_C
▶ sea_water_conservative_temperature		K
▶ sea_water_potential_temperature		K
▶ sea_water_potential_temperature_at_sea_floor		K
▶ sea_water_potential_temperature_expressed_as_heat_content <i>alias: integral_wrt_depth_of_sea_water_potential_temperature_expressed_as_heat_content</i> <i>alias: integral_of_sea_water_potential_temperature_wrt_depth_expressed_as_heat_content</i>		J m-2
▶ sea_water_redistributed_conservative_temperature		degree_C
▶ sea_water_redistributed_potential_temperature		degree_C
▶ sea_water_temperature		K

Description des données : units

SIST



- Table des noms standards (V69, Oct. 2019)
 - <http://cfconventions.org/standard-names.html>
 - Fournit les noms standards et leurs unités

Units	GRIB	PCMDI	Standard name
K	13	theta	air_potential_temperature
1	71 E164	clt	cloud_area_fraction
kg m ⁻²	79		large_scale_snowfall_amount
kg m ⁻² s ⁻¹			large_scale_snowfall_flux
m s ⁻¹			lwe_large_scale_snowfall_rate
K Pa s ⁻¹		mpwapta	product_of_omega_and_air_temperature
1			region
1	91	sic	sea_ice_area_fraction
1e-3	88	so	sea_water_salinity
W m ⁻²		rlds	surface_downwelling_longwave_flux
W m ⁻²		rls	surface_net_downward_longwave_flux

Description des données

SIST



- ***ancillary_variable*** : variable dépendante d'une autre
- Quand une variable fournit des métadonnées sur les valeurs d'une autre variable : *ancillary_variable* permet d'exprimer cette association en établissant un lien entre les variables.

```
float hum(time) ;
  :standard_name = "specific_humidity" ;
  :units = "g/g" ;
  :ancillary_variables = "h_error_limit h_detection_limit" ;
float h_error_limit(time)
  h_error_limit:standard_name = "specific_humidity
standard_error" ;
  h_error_limit:units = "g/g" ;
float h_detection_limit(time)
  :standard_name = "specific_humidity detection_minimum" ;
  :units = "g/g"
```


Description des données : valeurs manquantes

SIST



- Pour renseigner les valeurs manquantes ou les gammes de données valides
- ***_FillValue*** : pour pré remplir les données avec une valeur de base (valeur manquante par ex.)
- ***missing_value*** : valeur des données manquantes
- ***valid_min, valid_max*** : gamme de valeurs valides entre min et max
- ***valid_range***

Description des données : flag

SIST



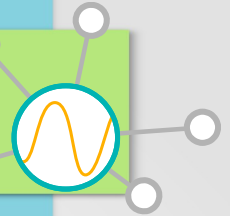
- attributs ***flag***
- Pour identifier des valeurs dans un ensemble “subset” de valeurs codifiées : 0, 1 ,2 , 3,
 - *flag_values, flag_meanings*

```
int bassin(lat, lon);  
    standard_name: region;  
    flag_values: 1, 2, 3;  
    flag_meanings: "atlantic_arctic_ocean  
indo_pacific_ocean global_ocean";
```

```
data:  
    bassin: 1, 1, 1, 1, 2, ..... ;
```

Description des données : coordinate

SIST

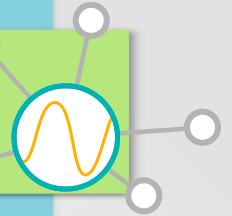


- Variables de **coordonnées** : 4 types de coordonnées spéciales dans la convention : *"latitude"*, *"longitude"*, *"vertical (hauteur/profondeur)"*, *"temps"*
 - définies par l'attribut *"axis"*, *"units"*, *"standard_name"*

```
float longitude;  
  :units = "degrees_east";  
  :standard_name = "longitude";  
  :axis = "X";  
  :coverage_content_type = "coordinate";  
float latitude;  
  :units = "degrees_north";  
  :standard_name = "latitude";  
  :axis = "Y";  
  :coverage_content_type = "coordinate";  
float depth(depth=580);  
  :axis = "Z";  
  :positive = "up";  
  :standard_name = "depth";  
  :long_name = "Profondeur";  
  :units = "m";
```

Description des données : coordinate

SIST



- - Chaque variable de coordonnée a sa propre dimension
 - Ex : Longitude(Longitude), lat(lat)
 - La variable de coordonnée distingue les éléments le long de l'axe (si possible monotone)
- - Associé à la variable par l'attribut coordinates

Exemple :

dimensions:

```
Longitude = 2160 ;
```

variables:

```
float Longitude(Longitude) ;
```

```
Longitude:long_name = "longitude" ;
```

```
Longitude:standard_name = "longitude" ;
```

```
Longitude:units = "degrees_east" ;
```

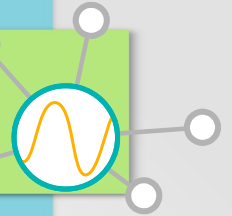
```
Longitude:axis = "X" ;
```

data:

```
Longitude = 25,30,35,... ;
```

Description des données : coordinate

SIST



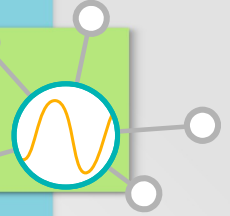
- profondeur/hauteur: Donner le sens positif/négatif
 - Exemple: si valeurs négatives en descendant de 0 à -1000m, on va vers le positif en montant

```
float depth(depth);  
  :axis = "Z";  
  :positive = "up";  
  :standard_name = "depth";  
  :long_name = "Profondeur";  
  :units = "m" ;
```

```
float temperature(depth) ;  
  :standard_name = "sea_water_temperature" ;  
  :units = "Celsius" ;  
  :long_name = "Temperature" ;  
  :source = "Seabird CTD" ;  
  :coverage_content_type = "physicalMeasurement" ;  
  :coordinates = "time longitude latitude depth" ;
```

Description des données : temps

SIST



- Unité de temps = chaîne formatée
 - On doit donner l'origine des valeurs de temps et le calendrier
 - Ex: *seconds since 1992-10-8 15:15:42.5 -6:00*

```
int time;  
    :units = "minutes since 1970-01-01  
00:00:00 UTC";  
    :calendar = "standard"; # gregorian  
    :long_name = "date de prelevement";  
    :standard_name = "time";  
    :origin = "01-JAN-1970 00:00:00";
```
- Comment calculer la valeur de temps selon le format de date ? :
<http://unidata.github.io/netcdf4-python/netCDF4/index.html#netCDF4.date2num>

```
NC_TIME_FMT = 'seconds since 1970-01-01 00:00:00 UTC'  
datenum = nc4.date2num(dateobs, units=NC_TIME_FMT)
```

Echantillonnage discret DSG

SIST



- La convention CF propose de standardiser la représentation de valeurs discrètes échantillonnées dans l'espace et le temps
 - <http://cfconventions.org/cf-conventions/cf-conventions.html#discrete-sampling-geometries>
 - <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#appendix-examples-discrete-geometries>
- Attribut « featureType » obligatoire :
 - *Point*
 - *TimeSeries*
 - *Profile*
 - *Trajectory*
 - *TimeSeriesProfile*
 - *trajectoryProfile*

Echantillonnage discret DSG

SIST



- Pour *timeserie, profiles, trajectories* :
 - une variable avec l'attribut **cf_role** devrait être incluse.
 - Les seules valeurs acceptables de cf_role pour les ensembles de données CF de géométrie discrète sont timeseries_id, profile_id et trajectory_id.
- Il est fortement recommandé qu'il y ait une variable de station (qui peut être de n'importe quel type) avec l'attribut *cf_role="timeseries_id"*, dont les valeurs identifient uniquement les stations.
- Il est recommandé qu'il y ait des variables de station avec les attributs standard_name " nom_plate-forme ", " surface_altitude " et " platform_id " le cas échéant.

Echantillonnage discret DSG

SIST



- Série temporelle : **featureType= «timeSeries»**
 - mesures à des points fixes, en fonction du temps
 - <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.pdf#time-series-dat>

dimensions:

```
station = 1 ;  
obs = UNLIMITED ; // (144 currently)
```

Variables:

```
double time(obs) ;  
    time:units = "minutes since 2013-01-01 00:00:00 UTC" ;  
    time:time_origin = "2013-01-01 00:00:00" ;  
    time:standard_name = "time" ;  
    time:axis = "T" ;
```

```
char station_name(station_name_len=20);  
    :long_name = "station_name";  
    :cf_role = "timeseries_id";
```

```
float tmer(obs) ;  
    tmer:standard_name = "sea_water_temperature" ;  
    tmer:units = "degree_C" ;
```

//global attributes

```
:featureType = "TimeSeries" ;  
:cdm_data_type = "TimeSeries" ## demandé dans autres normes ou API  
:cdm_timeseries_variables = "station_name,latitude,longitude" ;
```

}

Echantillonnage discret DSG

SIST



- Profil vertical : **featureType= «profile»**
 - mesures à des points fixes, temps fixe, en fonction d'un axe vertical (hauteur/profondeur)

dimensions:

```
station = 1 ;  
depth = UNLIMITED ; // (55 currently)
```

Variables:

```
float depth(depth);  
    :axis = "Z";  
    :positive = "up";  
    :standard_name = "depth";  
char stationname(lenstation=7);  
    :standard_name = "platform_name";  
    :cf_role = "profile_id";  
float temperature(depth);  
    :standard_name = "sea_water_temperature";  
    :units = "Celsius";
```

//global attributes

```
:summary = "Mesures CTD : temperature, salinity, oxygen, ;"  
:featureType = "profile";  
:cdm_data_type = "profile"; # demandé dans autres normes ou
```

API

```
:cdm_profile_variables = "stationname,latitude,longitude";
```

Echantillonnage discret DSG

SIST



- trajectoire : **featureType= «trajectory»**
 - mesures en fonction du temps et de l'espace, les coordonnées lat/lon évoluent avec le temps
 - <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.pdf#trajectory-data>

dimensions:

```
obs = 8640;
trajectory=1;
char trajectory(trajectory, name_strlen) ;
    :cf_role = "trajectory_id";
double time(obs);
    :_FillValue = -9999999; // int
    :axis = "T";
    :missing_value = -9999999; // int
    :standard_name = "time";
    :units = "seconds since 1970-01-01T00:00:00Z";
float lon(obs);
    :missing_value = -9999999.0f; // float
    :standard_name = "longitude";
    :units = "degrees_east";

float sal(obs);
    :_FillValue = -9999999.0f; //
float
    :coordinates = "time lat lon"
    :instrument = "thermosalinograph
(TSG)";
    :long_name = "TSG Salinity";
    :standard_name = "..._salinity";

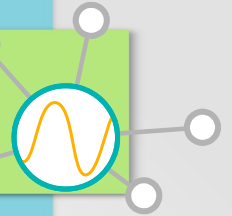
// global attributes:
:cdm_data_type = "trajectory";
:CF:featureType = "trajectory";
```

conclusion

- Documentation CF pas toujours tres claire, mais convention utile à suivre
- Recommendations de certain pôle de données (“Odatis”) de l’ IR Data Terra, pour les données physiques
- Suivre les conventions CF pour
 - Chimie et dynamique de l’atmosphère
 - Cycle du carbone
 - Nuage
 - Hydrologie
 - Dynamique océanique
 - Radiation
 - Glace de mer
 - Surface

Références

SIST



- La convention CF
 - <http://cfconventions.org/>
 - <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html>
- Le Checker netCDF
- <https://pumatest.nerc.ac.uk/cgi-bin/cf-checker.pl>
-