



DATA
TERRA



ODATIS

Campus de la Mer – 14 mai 2024 !

Une introduction au NetCDF

Joël Sudre (joel.sudre@data-terra.org)



Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Qu'est ce que le Netcdf ?

- **Network Common Data Format (1988)**
 - Un modèle de données
 - Un format de fichier
 - Une interface de programme d'application (API)
 - Une librairie implémentant l'API

D'où vient le NetCDF ?

- Créer aux US par Unidata (<https://www.unidata.ucar.edu/>) :
 - Regroupement de différentes communautés provenant des instituts d'éducation et de recherche
 - But : mutualiser les données de géoscience et des outils pour accéder et visualiser ces données
 - Fourni des données et des outils depuis 30 ans (1983)
 - Financer par National Science Foundation (NSF)
 - Unidata est maintenant un programme communautaire (UCP) de la Corporation Universitaire pour la Recherche Atmosphérique (UCAR)
 - Plan future : Technologie CLOUD, améliorer les mécanismes de distribution et de visualisation

Campus de la Mer, le mardi 14 mai 2024

Pourquoi choisir le NetCDF ?

- Portable
- Données indépendantes de la machine
- **Bibliothèques de procédures** dans de nombreux langages (créer, lire, modifier)
- Bibliothèques NetCDF **libres et gratuites**
- Des **dizaines de logiciels tiers gratuits** permettent de manipuler ce format (découper, assembler, faire des moyennes, visualiser,...)

Campus de la Mer, le mardi 14 mai 2024

Pourquoi choisir le NetCDF ?

- Format riche (métadonnées et données)
- Format flexible et normalisé
- Format **ouvert** (suffixe « .nc »)
- Format bien adapté pour stocker des **tableaux de nombres multidimensionnels**
- Format acceptant la mise en place de conventions
- Format permettant standardisation et interopérabilité
- Énormément utilisé en météorologie et en océanographie, dans le spatial...

Campus de la Mer, le mardi 14 mai 2024

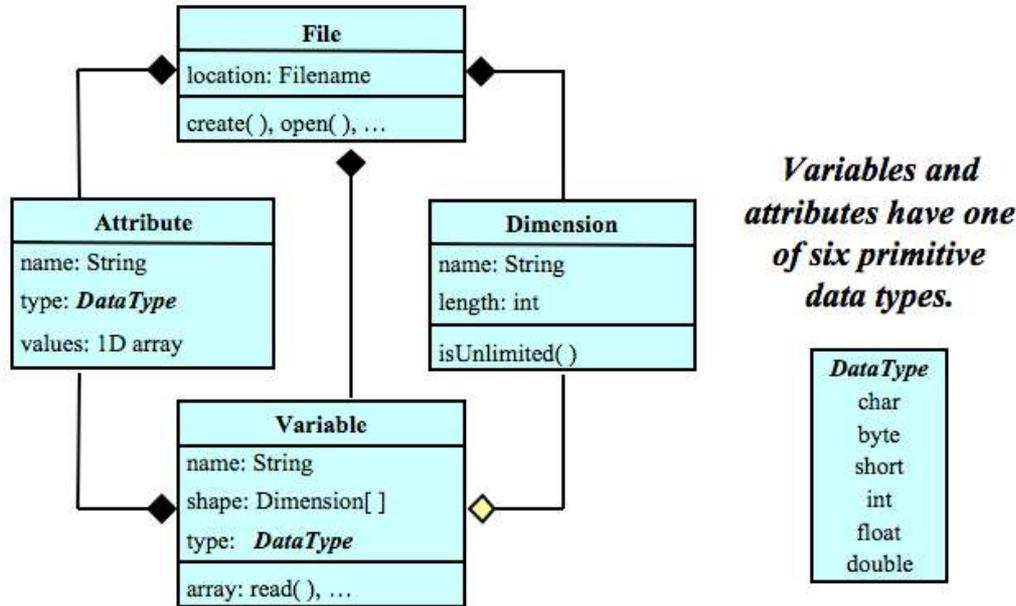
Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Les modèles de données

The Classic Model

The classic netCDF data model consists of variables, dimensions, and attributes. This way of thinking about data was introduced with the very first netCDF release, and is still the core of all netCDF files.



A file has named variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One dimension may be of unlimited length.

NetCDF Classic Data Model

Campus de la Mer, le mardi 14 mai 2024

Les modèles de données

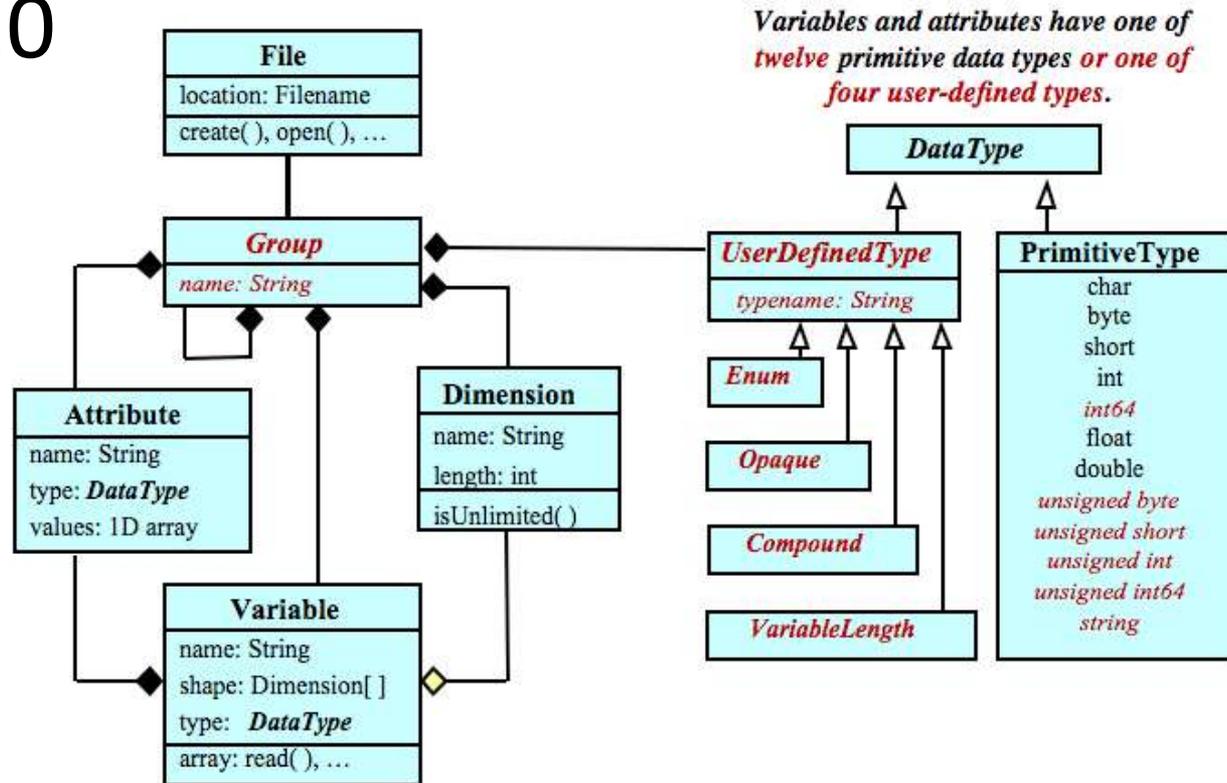
- Les limites du modèle classique (NetCDF 1,2,3)
 - Fichiers de **2GiB** Max (32-Bit)
 - **6 types atomiques** : Char, byte, short, int, float, double
 - **1 seule dimension de longueur illimitée**
 - **Variables en 32-bit**
- Limites rapidement atteintes avec l'explosion du volume des données (ex : fichiers de sortie de modèle numérique, données satellitales, etc.)

Les modèles de données

- Le modèle **NetCDF 64-bit Offset Format** (introduit à partir de la version **3.6.0**)
 - **Même type atomique** que le modèle classique
 - Limitation identique pour l'**unicité de la dimension de longueur illimitée**
 - **64-Bit** → **Fichier de volume quasi illimité**
 - **Variables de 4 GiB** par enregistrement

Les modèles de données

NetCDF 4.0



A file has a top-level unnamed group. Each group may contain one or more named subgroups, user-defined types, variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One or more dimensions may be of unlimited length.

Campus de la Mer, le mardi 14 mai 2024

Les modèles de données

- Le NetCDF 4.0 vs NetCDF 3.0:
- Ajoute un degré d'abstraction au modèle par une **notion de groupe** (conforme au HDF5)
- Permet d'avoir **plusieurs dimensions illimitées**
- Introduit **6 nouveaux types** (atomiques) : int64, unsigned (byte, short, int, int64), string
- Introduit **4 types « Userdefinedtype »** (structure en C)
- Interopérabilité avec le HDF5 (mais **pas** identique!)
- Permet de faire des E/S avec des procédures de calcul parallèle
- **Attention les bibliothèques NetCDF 3 ne peuvent pas lire le NetCDF 4.0**

Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Le NetCDF classique

La **variable** (stockage de la donnée)
va être définie par :

- Un nom (chaîne de caractères)
- Un type de donnée
- Une ou plusieurs dimension(s)
- Des attributs (métadonnées)

Exemple :
variables:

float SLA(Time, Longitude, Latitude) ;

SLA:_FillValue = 9.96921e+36f ;

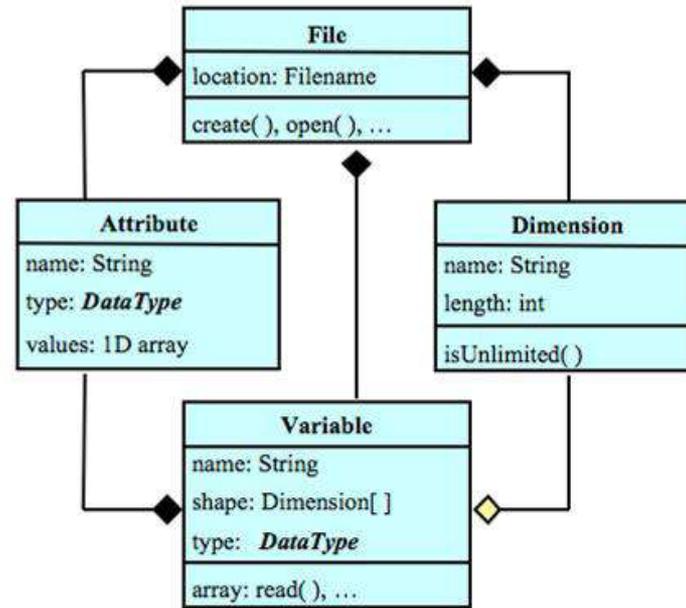
SLA:scale_factor = 1.f ;

SLA:add_offset = 0.f ;

SLA:long_name = "Sea Level Anomaly Estimate" ;

SLA:standard_name = "Sea Level Anomaly Estimate" ;

SLA:units = "m" ;



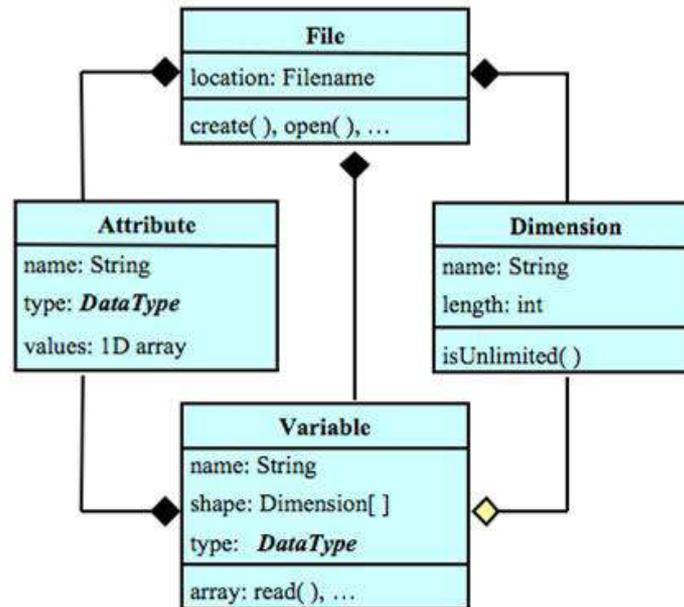
*Variables and
attributes have one
of six primitive
data types.*

| DataType |
|----------|
| char |
| byte |
| short |
| int |
| float |
| double |

Le NetCDF classique

La **dimension** (permet définir la/les forme(s) de la variable)
va être définie par :

- Un nom (chaîne de caractères)
- Une longueur (entier positif ou unlimited)
- Des attributs (métadonnées)



Variables and attributes have one of six primitive data types.

| <i>DataType</i> |
|-----------------|
| char |
| byte |
| short |
| int |
| float |
| double |

Exemple :

dimensions:

Longitude = 2160 ;

Latitude = 960 ;

Time = UNLIMITED ; // (1 currently)

Le NetCDF classique



Une **dimension** et une **variable** peuvent avoir le **même nom** !

Exemple :
dimensions:

Longitude = 2160 ;

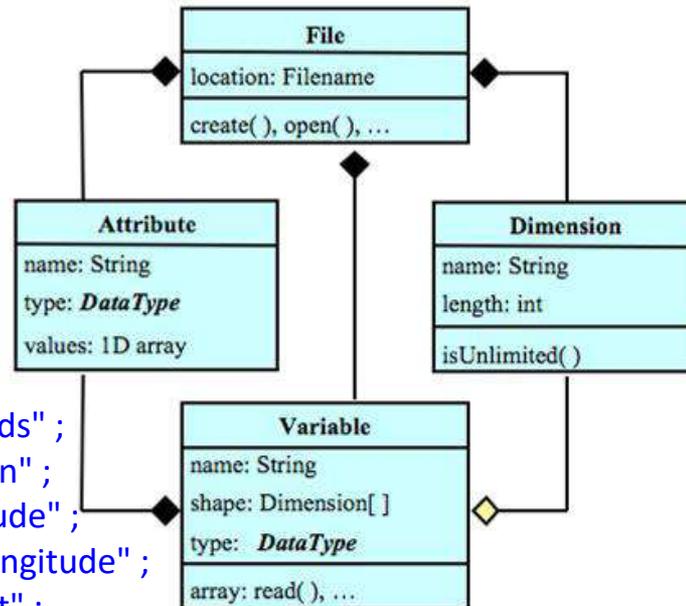
variables:

float Longitude(Longitude) ;

Longitude:bounds = "Lon_bounds" ;
Longitude:point_spacing = "even" ;
Longitude:long_name = "longitude" ;
Longitude:standard_name = "longitude" ;
Longitude:units = "degrees_east" ;
Longitude:axis = "X" ;

data:

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



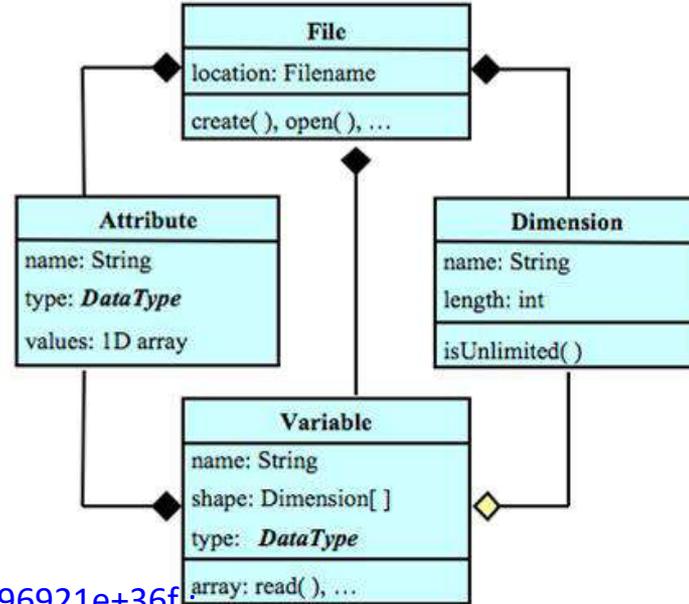
Variables and attributes have one of six primitive data types.

| <i>DataType</i> |
|-----------------|
| char |
| byte |
| short |
| int |
| float |
| double |

Le NetCDF classique

Les **attributs** (permettent d'introduire de la métadonnée ou des données auxiliaires dans le fichier) vont être associés à une variable et définis par :

- Un nom (chaîne de caractères)
- Un type de donnée
- Une valeur (tableau 1D)



Variables and attributes have one of six primitive data types.

| <i>DataType</i> |
|-----------------|
| char |
| byte |
| short |
| int |
| float |
| double |

Exemple :
variables:

```
float SLA(Time, Longitude, Latitude) ;
SLA:_FillValue = 9.96921e+36f ;
SLA:scale_factor = 1.f ;
SLA:coordinates = "Sea Level Anomaly Estimate" ;
SLA:add_offset = 0.f ;
SLA:long_name = "Sea Level Anomaly Estimate" ;
SLA:standard_name = "Sea Level Anomaly Estimate" ;
SLA:units = "m" ;
```

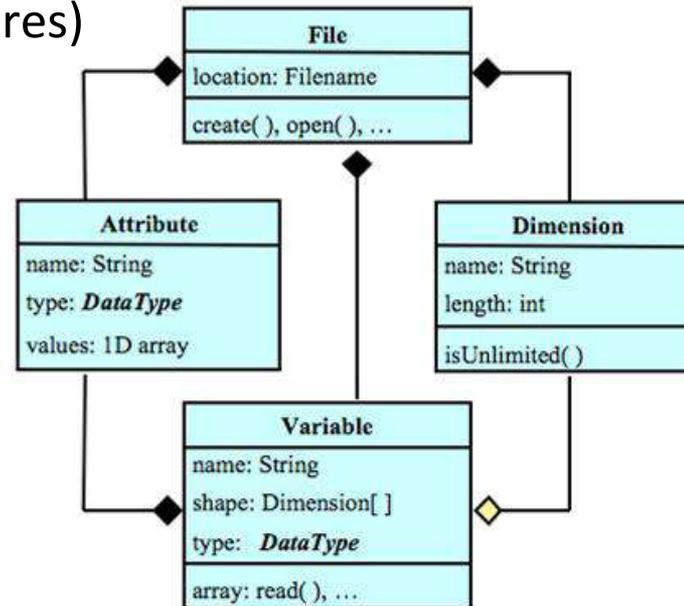
Campus de la Mer, le mardi 14 mai 2024

Le NetCDF classique



Si la variable associée aux **attributs** est nulle les attributs sont globaux :

- Un nom (chaîne de caractères)
- Un type de donnée
- Une valeur (tableau 1D)



Variables and attributes have one of six primitive data types.

| Data Type |
|-----------|
| char |
| byte |
| short |
| int |
| float |
| double |

Exemple :

// global attributes:

```
:geospatial_lon_min = 0.08333334f ;
:geospatial_lat_max = 79.91666f ;
:time_coverage_end = "1998-06-28" ;
:title = "Sea Level Anomaly Estimate based on Altimeter Data" ;
:geospatial_lat_min = -79.91666f ;
:time_coverage_start = "1998-06-28" ;
:Conventions = "CF-1.6" ;
```

Campus de la Mer, le mardi 14 mai 2024

Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

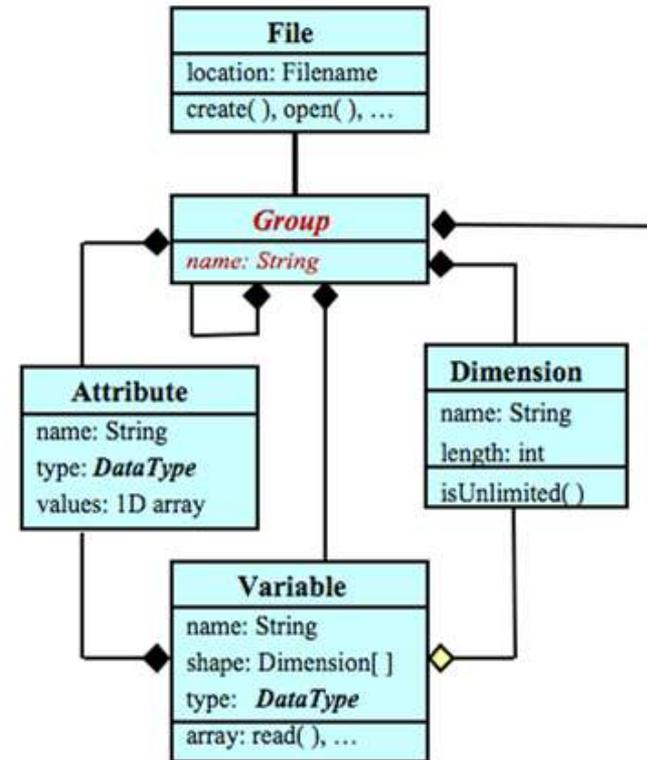
Le NetCDF 4.0

Un **groupe** est associé à:

- Un nom de groupe
- Des attributs de variable et/ou de groupe
- Des dimensions
- Des variables
- Des types

Exemple :

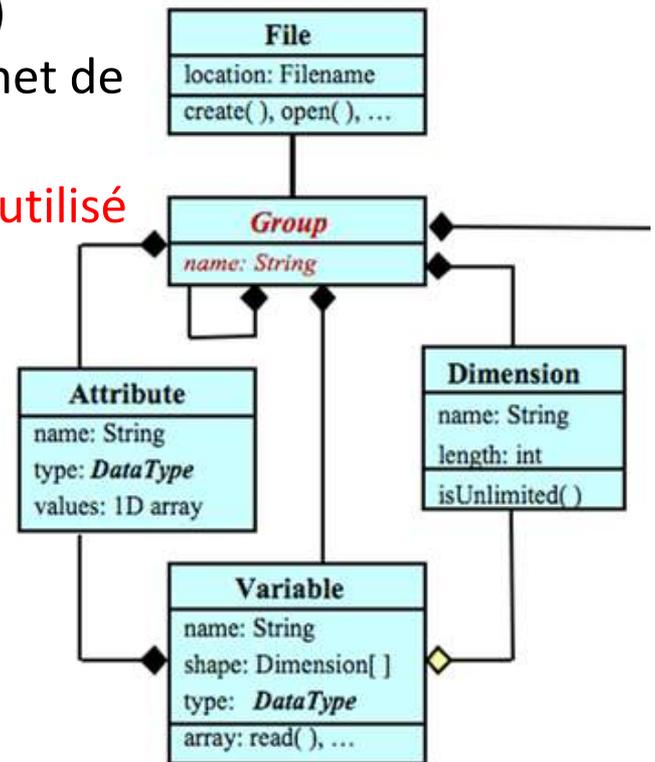
```
group: mozaic_flight_2012030321335035_descent {  
  dimensions:  
    air_press = 76 ;  
  variables:  
    double air_press(air_press) ;  
    air_press:name = "air_pressure" ;  
    air_press:unit = "Pa" ;  
  // group attributes:  
    :airport_dep = "FRA" ;  
    :flight = "2012030321335035" ;  
} // group mozaic_flight_2012030321335035_descent
```



Le NetCDF 4.0

Changement avec l'introduction de la notion de groupe :

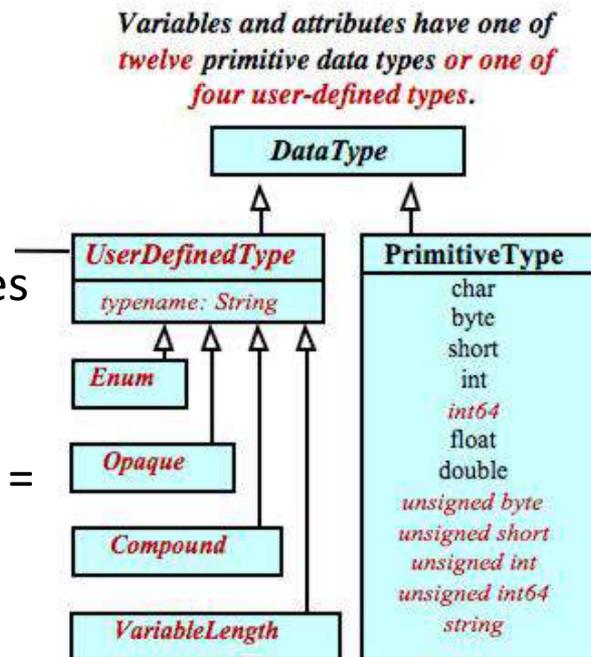
- Les **attributs globaux n'existent plus**
- Création possible d'**attributs de groupe**
- Chaque **groupe est comme un modèle classique** (attributs, dimensions, variables, attributs de groupe)
- Le **groupe par défaut est le groupe racine** (cela permet de faire entrer le modèle classique dans ce modèle)
- Une dimension définie pour le 1^{er} groupe peut-être utilisé dans un second groupe



Le NetCDF 4.0

Changement avec les nouveaux types :

- 6 types atomiques en plus
- 4 UserDefinedType :
 - > **Enum** : associe un entier constant à un nom
 - > **Opaque** : collection d'objets de dimensions connues mais chaque objet à la même taille
 - > **Compound** : (modèle de la structure en C) variable = mélange de plusieurs types
 - > **VariableLength** : permet de créer des tableaux irréguliers d'un même type atomique



Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Les Conventions

- Promouvoir le traitement et le partage de fichiers NetCDF via la **standardisation de l'auto-description** des jeux de **données et métadonnées** associés:
 - Des noms de variables
 - Des attributs
 - Des unités
 - Des métadonnées
 - Du géoréférencement
 - ...

Campus de la Mer, le mardi 14 mai 2024

Les Conventions

- Principes de base :
 - Éviter que ce soit trop pénible pour l'utilisateur et le producteur
 - Minimiser les erreurs « stupides » (ex : Obs T en Celsius ou Fahrenheit!)
 - Le fichier doit contenir toutes les informations nécessaires à l'utilisateur (pas de table et d'information externes)
 - Toutes les métadonnées doivent être lisibles par un humain et analysables par un programme
 - Minimiser les redondances
 - Éviter la multiplication des attributs
 - Informations fournies par variable plutôt que par fichier
 - Rien ne dépend des noms de variables (excepté la convention des variables de coordonnées)

Les Conventions

- Différentes Conventions :
- Convention Cooperative Ocean/Atmosphere Research Data Service (COARDS)
- Climate and Forecast (CF) Convention
- SeaDataNet (SDN) Convention
- ...

Campus de la Mer, le mardi 14 mai 2024

Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

La Convention CF

- Actuellement version 1.11 (<http://cfconventions.org/>)



- Uniquement certains types de données (Océan et Atmosphère):

–Chimie et dynamique de l'atmosphère

–Cycle du carbone

–Nuage

–Hydrologie

–Dynamique océanique

–Radiation

–Glace de mer

–Surface



Gros effort de standardisation à faire dans les différentes communautés

La Convention CF

Origine de la donnée

Insertion de métadonnées fournissant quelques informations basiques 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 |

* attributs qui peuvent aussi être au niveau de chaque variable

La Convention CF

Description de la donnée

Attributs obligatoires à minima

| | |
|---------------------------------------|---|
| units | Unité international (degC, Pa, mbar,..) Ne pas utiliser psu, dB, Sv |
| standard_name | Identifier la donnée Ex : sea_surface_temperature |
| long_name | Pas de standard Ex : Sea Surface Temperature |
| _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 | $D = \text{add_offset} + \text{scale_factor} * \text{variable}$ Ex : add_offset=0.f et scale_factor=1.f |
| flag_values et/ou flag_meaning | Si la donnée nécessite une valeur de « flag » |

La Convention CF

Description de la donnée

Table des noms standards (V47, 19 Sept 2017)(<http://cfconventions.org/standard-names.html>)

| 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 |

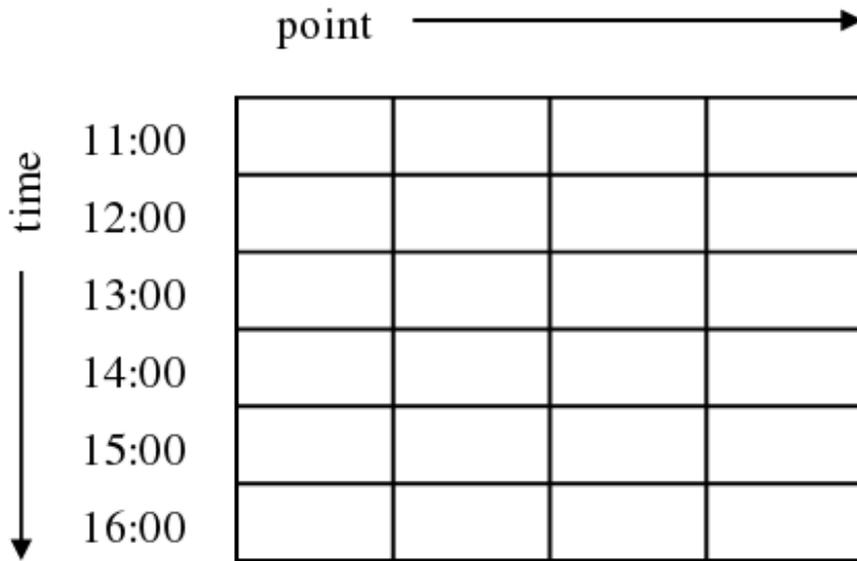
La Convention CF

Dimensions et coordonnées

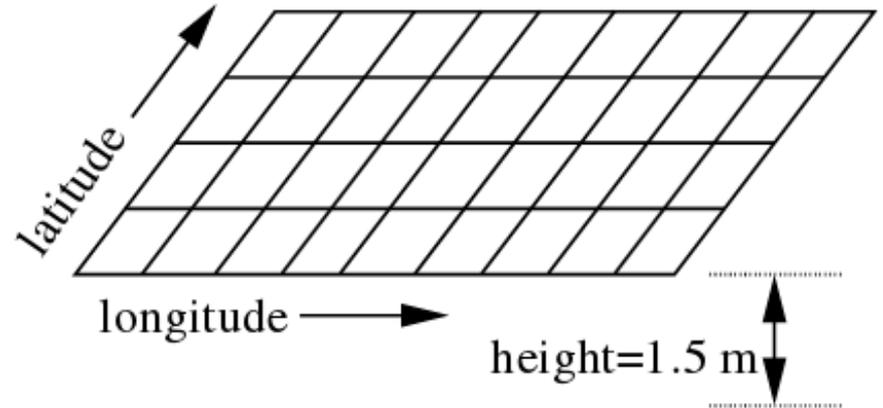
La dimension indexe la donnée Ex : Temperature (46,0)

Ordre des dimensions : tzyx

La coordonnée est une variable indépendante T =26°C à 0°E et 10.0°N



point is a dimension without coordinates



height is a coordinate without dimension or of size one

La Convention CF

Dimensions et coordonnées

Les variables contenant les coordonnées :

- Chaque **variable de coordonnée a sa propre dimension** avec un nom identique (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" ;

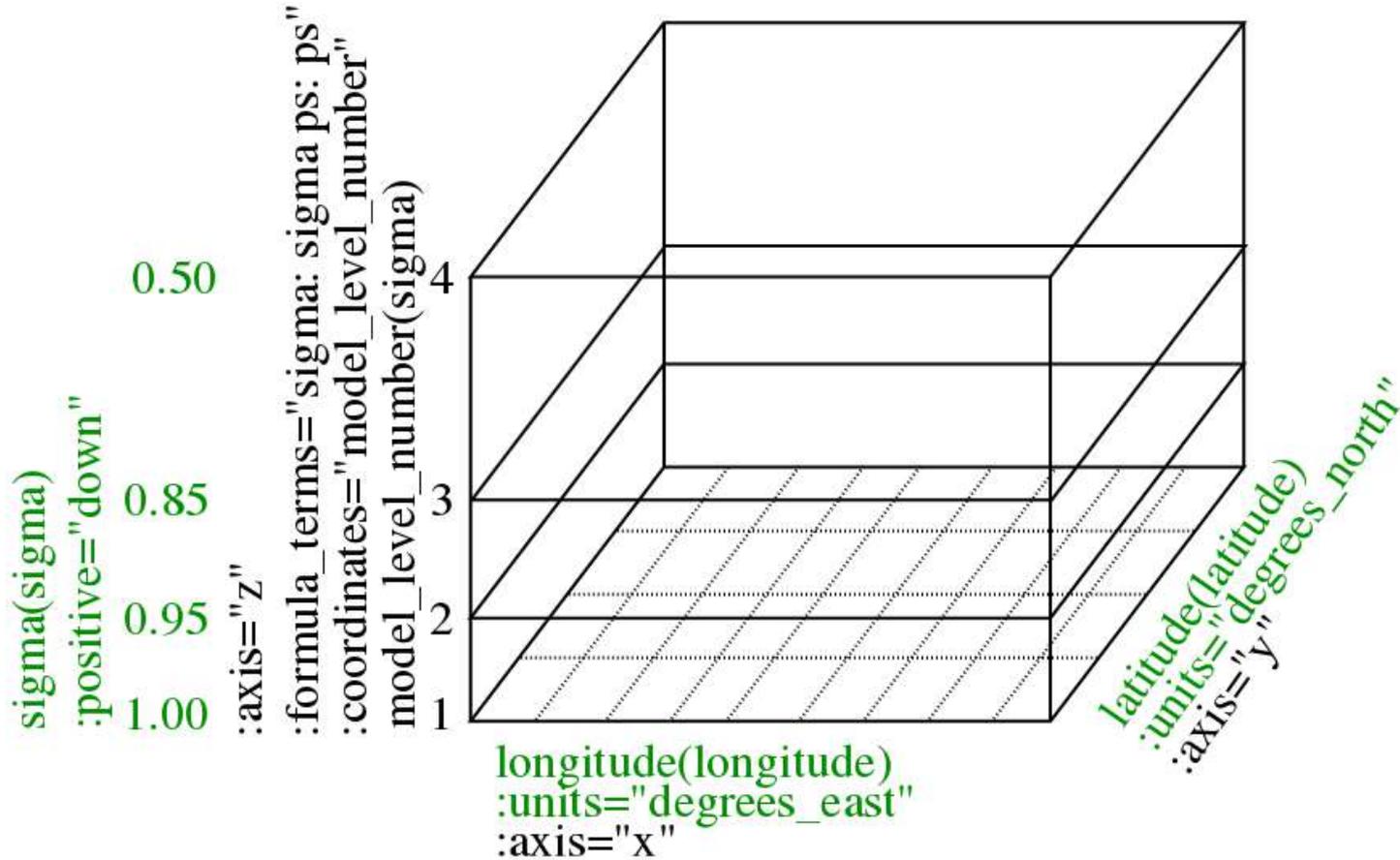
Longitude:axis = "X" ;

data:

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

La Convention CF

Dimensions et coordonnées



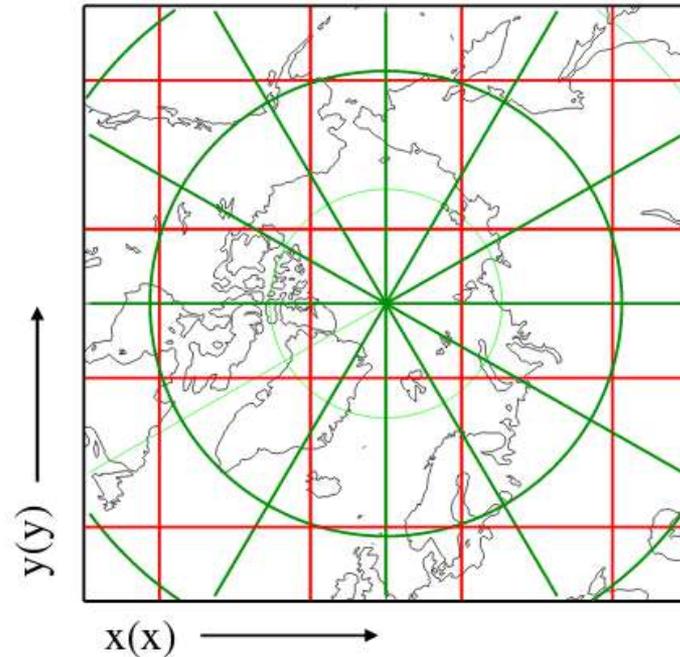
Campus de la Mer, le mardi 14 mai 2024

La Convention CF

Dimensions et coordonnées

| | | name(point,length) | | | |
|------------|-------|--------------------|-----------|-----------|---------|
| | | Hamburg | Livermore | Princeton | Reading |
| lon(point) | | 10 | -122 | -75 | -1 |
| lat(point) | | 54 | 38 | 40 | 51 |
| time ↓ | 11:00 | | | | |
| | 12:00 | | | | |
| | 13:00 | | | | |
| | 14:00 | | | | |
| | 15:00 | | | | |
| | 16:00 | | | | |

float air_temperature(time,point)
:coordinates="lat lon name"



float air_pressure_at_sea_level(y,x)
:coordinates="lat lon"
:grid_mapping="mappinginfo"
float lat(y,x)

La Convention CF

Dimensions et coordonnées

La **variable temporelle** :

- Time (année, mois, jour, heure, minute, seconde,..) doit avoir des attributs obligatoires :

- L'**attribut units** (unité de temps since temps de référence)

Ex time:units = « days since 1950-01-01-00-00 »

- L'**attribut calendar**

(360-days, standard, julian...)

Le calendrier **standard**

est le calendrier **grégorien**

Exemple :

dimensions:

```
time = UNLIMITED ; // (12 currently) ;
```

variables:

```
float time(time) ;
```

```
time:long_name = "time" ;
```

```
time:standard_name = "time" ;
```

```
time :axis = "T"
```

```
time:units = "seconds since 2001-01-01" ;
```

```
time:calendar = "360-days" ;
```

data:

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

Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Représentation de l'échantillonnage discret

- Comment représenter un échantillonnage discret ?
- <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#appendix-examples-discrete-geometries>
- Les « featureType » :
 - Point data
 - TimeSeries
 - Trajectory
 - Profile
 - TimeSeriesProfile
 - trajectoryProfile

Campus de la Mer, le mardi 14 mai 2024

Représentation de l'échantillonnage discret

- Les « featureType » :
 - dimensions:
 - obs = 1234 ;
 - variables:
 - double time(obs) ;
 - float lon(obs) ;
 - float lat(obs) ;
 - float alt(obs) ;
 - float humidity(obs) ;
 - humidity:standard_name = "specific_humidity" ;
 - humidity:coordinates = "time lat lon alt" ;
 - attributes:
 - :featureType = "point";
- Point Data
- observations de paramètres
- sans relation spatio-temporelle

Représentation de l'échantillonnage discret

- Les « featureType » :
- TimeSeries
- observations de paramètres
- à différentes stations
- mais au même temps
- ex : réseaux d'obs

dimensions:

```
station = 10 ; // measurement locations  
time = UNLIMITED ;
```

variables:

```
float humidity(station,time) ;  
double time(time) ;  
float lon(station) ;  
float lat(station) ;  
float alt(station) ;  
char station_name(station, name_strlen) ;
```

attributes:

```
:featureType = "timeSeries";
```



Même temps d'échantillonnage !

Représentation de l'échantillonnage discret

- Les « featureType » :
- TimeSeries
- ex : réseaux d'obs



Temps
d'échantillonnage différents

Matrices creuses :
Privilégier la simplicité au stockage

dimensions:

station = UNLIMITED ;

obs = 13 ;

variables:

float lon(station) ;

float lat(station) ;

float alt(station) ;

char station_name(station, name_strlen) ;

double time(station, obs) ;

float humidity(station, obs) ;

float temp(station, obs) ;

attributes:

:featureType = "timeSeries";

Représentation de l'échantillonnage discret

- Les « featureType » :
- TimeSeries
- ex : réseaux d'obs



Temps
d'échantillonnage dif
férents

échantillonnages très irréguliers :
Privilégier le stockage

dimensions:

station = 23 ;

obs = 1234 ;

variables:

float lon(station) ;

float lat(station) ;

float alt(station) ;

char station_name(station, name_strlen) ;

int row_size(station) ;

row_size:long_name = "number of
observations for this station " ;

row_size:sample_dimension = "obs" ;

double time(obs) ;

float humidity(obs) ;

attributes:

:featureType = "timeSeries";

Représentation de l'échantillonnage discret

- Les « featureType » :
 - Trajectory : featureType= «trajectory» → dimension pertinente time (comme point avec time à la place de obs!)
 - Profile : featureType= «profile» → dimension pertinente z
 - TimeSeriesProfile featureType= «timeSeriesProfile» → (dimensions pertinentes : station, profile, z)
 - TrajectoryProfile featureType= «trajectoryProfile» → (dimensions pertinentes : trajectory, profile, z)
- Voir le poster de l'arbre décisionnel : [netcdf_CF_decision_tree_high_res.pdf](#)

Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Les Outils

Il existe de nombreux outils qui permettent :

- de visualiser simplement dans un terminal le contenu d'un fichier NetCDF (format sémantique :CDL Concept Description Language)

(la commande : **Ncdump**)

- de visualiser la donnée sous forme graphique

(les logiciels : **Ncview** et **Panoply**)

- de manipuler les fichiers NetCDF et leurs contenus

(la suite NetCDF Operator **NCO**)

Les Outils

Quelques exemples pratiques

Ncdump (outil Unidata)

Ncdump -h : permet de voir l'entête d'un fichier

```
ncdump -h testrh.nc
```

```
ncdump -h tos_O1_2001-2002.nc
```

Ncdump -v : permet de voir une variable dans un fichier

```
ncdump -v time tos_O1_2001-2002.nc
```

Ncdump -c : permet de voir les coordonnées

```
ncdump -c tos_O1_2001-2002.nc
```

Les Outils

Quelques exemples pratiques

Ncview et Panoply: (visualisation de la donnée)

Ncview : permet de représenter des données N Dimensions

`ncview testrh.nc` (1D)

`ncview v_gekco2_usec_20161201.nc` (2D)

`ncview tos_O1_2001-2002.nc` (2D, time) anim

`ncview test_hgroups.nc`

Panoply : outil plus complet (mais plus difficile à prendre en main!)

Les Outils

Quelques exemples pratiques

La boîte à outils NCO : (manipulation par lignes de commandes des fichiers NetCDF)

- [ncap2](#) netCDF Arithmetic Processor ([examples](#))
- [ncatted](#) netCDF ATtribute EDitor ([examples](#))
- [ncbo](#) netCDF Binary Operator (addition, multiplication...) ([examples](#))
- [ncclimo](#) netCDF CLIMatOlogy Generator ([examples](#))
- [nces](#) netCDF Ensemble Statistics ([examples](#))
- [ncecat](#) netCDF Ensemble conCATenator ([examples](#))
- [ncflint](#) netCDF FiLe INTerpolator ([examples](#))
- [ncks](#) netCDF Kitchen Sink ([examples](#))
- [ncpdq](#) netCDF Permute Dimensions Quickly, Pack Data Quietly ([examples](#))
- [ncra](#) netCDF Record Averager ([examples](#))
- [ncrcat](#) netCDF Record conCATenator ([examples](#))
- [ncremap](#) netCDF REMAPer ([examples](#))
- [ncrename](#) netCDF RENAMEer ([examples](#))
- [ncwa](#) netCDF Weighted Averager ([examples](#))

Les Outils

Quelques exemples pratiques

La boîte à outils NCO :

Ncap2 : processeur arithmétique

Ex : créer un champs avec la norme d'une vitesse

```
ncap2 -s 'norm[time,lat,lon]=(u^2+v^2)^0.5'
```

Ncatted : manipuler les attributs

Ex : ajouter le long_name sur la norme

```
ncatted -a long_name,norm,o,c,'norme of uv' norme.nc
```

Ex : mettre toutes les missing_value à 0.d

```
ncatted -a missing_value,norm,o,d,0. norme.nc
```

Ex : éliminer toutes les unités

```
ncatted -a units,,d,, norme.nc
```

Les Outils

Quelques exemples pratiques

La boîte à outils NCO :

Ncbo : opération binaire (addition, soustraction, ...)

Ex : différence de toutes les données entre 2 dates

```
ncbo --op_typ="-" fichier1.nc fichier2.nc diff.nc
```

Les fichiers doivent être identiques

Nccat : concaténation (de type Ensemble)

Ex : concaténer dans le temps les 3 premiers fichiers

```
nccat -d time,0, fichier[123].nc all_time.nc
```



Les Outils

Quelques exemples pratiques

Nccat : concaténation (de type Ensemble - record)

Ex : concaténer la variable h sur tous les fichiers

```
nccat -v h -d time,0, fichier*.nc h_time.nc
```

Pour concaténer sur le temps

1) Changer la dimension time en record

Ncks : (la commande « kitchen sink »!)

```
ncks -O -h --mk_rec_dmn time fichier1.nc tmp.nc
```

2) Concaténer tous les fichiers

Nccat : concaténer sur un record

```
nccat -h tmp.nc fichier*.nc c_time.nc
```

Les Outils

Quelques exemples pratiques

Ncks : (la commande « kitchen sink »!)

Extraire une zone géographique lon de 300°E – 310°E
Lat de 28°N-40°N de la variable h

```
ncks -d lon,300.,310. -d lat,28.,40. -d time,0,0 -v h  
fichier1.nc extract.nc
```



Pour **lon** et **lat** les valeurs sont des **floats** →
interprétées comme des **grandeurs géophysiques**

Pour **time** les valeurs sont des **entiers** →
interprétées comme des **indices**

Plan de la présentation

- 1) Introduction générale sur le NetCDF
- 2) Les modèles de données
- 3) Le modèle NetCDF classique
- 4) Le modèle NetCDF 4.0
- 5) Les conventions
- 6) La convention CF (Climate and Forecast)
- 7) Représentation de l'échantillonnage discret
- 8) Les Outils (exercices pratiques)
- 9) Quelques liens utiles

Campus de la Mer, le mardi 14 mai 2024

Quelques liens utiles !

- <https://www.unidata.ucar.edu/software/netcdf/docs/index.html>
- <https://www.unidata.ucar.edu/software/netcdf/docs/modules.html>
- <http://cfconventions.org/>
- <https://www.giss.nasa.gov/tools/panoply/>
- <http://nco.sourceforge.net/>

Objectifs de l'atelier technique

Réunir les partenaires pour:

- Echanger sur:
 - les pratiques
 - les besoins
 - les attentes
- **Définir les orientations techniques du pôle** (Cahier des Charges des CDS).
- **Informé sur les bonnes pratiques** à mettre en œuvre
- **Homogénéiser et coordonner les CDS** et les producteurs de données par l'émission de recommandations (FAIR, Open Science,...)
- **Aide à la certification CoreTrustSeal**
- **Montrer des mises en pratique** (exercices d'application de logiciels)

Campus de la Mer, le mardi 14 mai 2024

Archive de l'Atelier Technique

- L'ensemble des sessions de l'atelier technique sont disponibles sur le site ODATIS (18 sessions)
- <https://www.odatis-ocean.fr/activites/ateliers-techniques>
- Les 3 derniers ateliers:
 - **Atelier de travail « Catalogue ODATIS »**: Atelier de travail, interne aux membres du bureau exécutif du pôle, clarification du périmètre des catalogues ODATIS et Sextant avec l'entrepôt Seanoe; faciliter la découverte et l'accès aux jeux de données du catalogue ODATIS, assurer la traçabilité & visibilité des catalogues et producteurs des données moissonnées et de faciliter la réutilisation des jeux de données du catalogue ODATIS (identification des fonctionnalités à ajouter pour permettre l'intégration des jeux de données avec des services types VDAP (visualisation/analyse/traitement à distance), création d'un catalogue de services ODATIS).
 - **Atelier données de la Flotte Océanographique Française – FOF**: clarifier le workflow des données depuis la préparation de la campagne jusqu'à la bancarisation et la diffusion de ces données; Aspects réglementaires, documentaires et procédures, Définir un plan d'action pour élaborer un plan de gestion de données commun ODATIS / IR*FOF; Identifier des sujets d'intérêt autour des questions de réglementation et définir un plan d'action; Identifier des sujets d'intérêt autour des services supports aux campagnes et définir un plan d'action.
 - **Atelier « Données bioinformatique de diversité »**: organisé conjointement par les pôles ODATIS et PNDB, composantes de l'IR Data Terra; et en collaboration avec ABIMS, SEBIMER et le MIO: Bonnes pratiques FAIR (standards de données et métadonnées), workflow, outils, web services de bio-analyse.

Campus de la Mer, le mardi 14 mai 2024

Quelques recommandations issues de l'atelier technique (1/3)

- **Métadonnées** de découverte à la norme ISO 19115
Moissonnage CSW (éventuellement OAI/PMH)
- Métadonnées d'usage incluses dans les fichiers de données
- **Format** des données (physique, chimie, biogéochimie, déchets, ...):
 - ASCII TSV (ODV spreadsheet normalisé SDN)
 - NetCDF4 (sans groupes) convention Climate and Forecast (V1.6 ou +)
 - Formats spécifiques :Image (JPEG2000, GeoTIFF), géographique (shape)
- Attributs de **paramètres** (double convention):
 - Convention CF (pour les param. Phy. avec attributs SDN liste P01)
 - Convention SeaDataNet (Biogéochimie, chimie, ...) + attributs supplémentaires « long name » en convention CF
- **Biologie** (en cours, en commun avec PNDB et participation du SAR):
 - *Darwin Core Archive, (EML)*
 - *WORMS (en première instance), nécessaires adaptations « locales »*

Quelques recommandations issues de l'atelier technique (2/3)

- Des recommandations d'implémentation
 - Unité des paramètres (P06 SDN) et la discipline (P08 de SDN)
 - Pour les fiches de métadonnées EV obligatoires (P02 de SDN)
 - Nomenclature d'un paramètre: liste P01 de SDN (si découverte P02)
 - Date et heure entier en « long » avec un offset et un scale factor (UTC)
 - Axes et Orientations à bien préciser (Attention à la profondeur positive!)
 - Positionnement EPSG (European Petroleum Survey Group)

- **Services de données** (une liste adaptable)
 - OpenDAP (Thredds Data Server, ERDDAP)
 - OGC-WMS (visualisation)
 - OGC WCS (données « grilles »)
 - ~~➤ OGC WFS (avec modèles INSPIRE), (?), éventuellement via EMODnet~~

Quelques recommandations issues de l'atelier technique (3/3)

Un point important: la traçabilité des données

- Citer la source des données et des informations (indissociable du travail scientifique)
- **Eviter les doublons (nécessité de données interopérables)**
- Identifier l'origine des données (...qualité)
- Reconnaître l'importance de l'effort d'observation
- Etablir un « constat » de l'état de l'art à un instant T
- Respect de certaines obligations réglementaires (APA, DCSMM 19.3)?

- Utilisation d'identificateurs permanents
 - Choix des **Digital Object Identifiers (DOI)**
 - Notion **d'entrepôts** (copies/ images stabilisées des données / Versions)



Merci pour votre attention

