

DATA
TERRA

Ecole Thématique DATA SDUE

*Guide de Survie dans la jungle des données
en Sciences de l'Univers et de l'Environnement (SDUE) :
Comment gérer les données pour les valoriser?*

Session «Formats de fichiers : NetCDF»

Maurice Libes – OSU Pytheas

NetCDF un format de fichier Interopérable pour la science ouverte

Mots clés : gestion de données, format de fichier, science ouverte, interopérabilité, environnement, géolocalisation

1/ Généralités sur NetCDF

Le format NetCDF « [Network Common Data Format](https://www.unidata.ucar.edu/) » créé aux US par Unidata (<https://www.unidata.ucar.edu/>), évolue régulièrement depuis sa création en 1988, c'est un projet très suivi, utilisé internationalement.

Le projet est pérenne, financé par la National Science Foundation (NSF). Il regroupe différentes communautés provenant des instituts d'éducation et de recherche dans le but de **mutualiser les données** scientifiques.

- format de fichier libre et ouvert permet la ***standardisation et l'interopérabilité***.
 - Très utilisé en météorologie, climatologie, en océanographie, et dans le spatial...
 - mais s'applique parfaitement de manière générique dès lors qu'on veut représenter des données qui évoluent selon 1, 2 ou plusieurs axes.

NetCDF est à la fois un modèle de représentation de données et un format de fichier. Le format étant « binaire », il nécessite et fournit :

- un ensemble de programmes spécifiques à NetCDF pour accéder et visualiser ces données :
 - pour lire traiter et exploiter les fichiers dans de nombreux langages (créer, lire, modifier, extraire des données)
- une interface de programme d'application (API) utilisable dans de nombreux langages comme *Fortran, C, Java, R, Python etc...*

1/ NetCDF pour qui ?, pour quel usage ?

Dans quel cas NetCDF peut m'être utile ? Quels sont les champs d'utilisation ?

Le format NetCDF est couramment utilisé pour ranger et ***représenter des données sous forme de tableaux (1D, 2D ou plus)***, et ***lorque ces données évoluent selon des variables***.

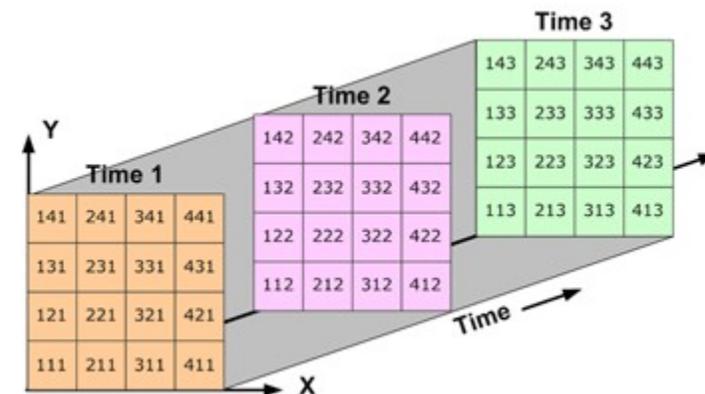
Cette représentation est fréquente par exemple dans les mesures physiques en océanographie, les sciences de l'atmosphère, de climatologie, de météorologie, les cartes satellitaires dans le spatial etc... les applications dans des systèmes d'information géographiques (S.I.G.)

- Le format de fichier NetCDF peut être très utile à tous les chercheurs, ingénieurs, techniciens qui sont en charge de :
 - Collecter des données environnementales sur le milieu naturel puis de les
 - Stocker et de **Représenter les données dans un format de fichier qui est sémantiquement riche en informations**, afin de les
 - Traiter et les Analyser, puis de les
 - Partager et de les diffuser dans le cadre de *l'interopérabilité* demandée par la science ouverte

1/ NetCDF pour qui ? pour quel usage ?

Les données d'un fichier NetCDF sont stockées sous la forme de tableaux. NetCDF est totalement approprié dès lors qu'on veut représenter, par exemple :

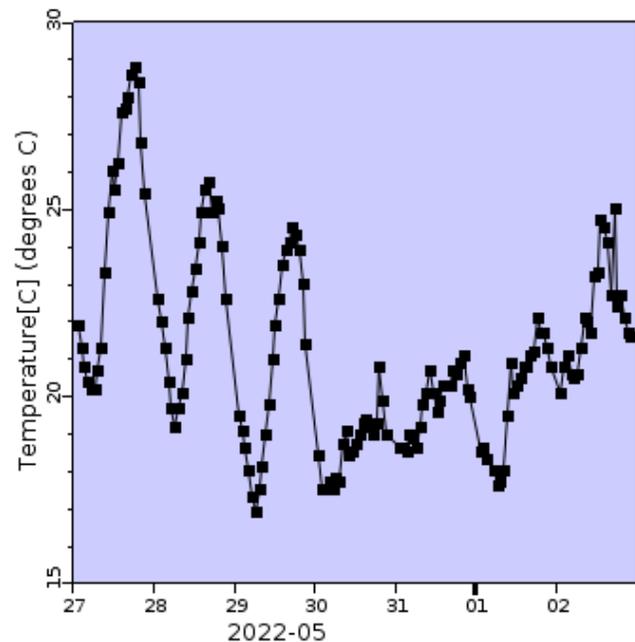
- - l'évolution de mesures effectuées au fil du temps sur un point fixe. Les mesures sont stockées dans un **tableau unidimensionnel** : par exemple la température d'un milieu qui évolue selon le temps
 $Temperature = F(temps)$
- - l'évolution de mesures au fil du temps sur une station fixe, mais à plusieurs altitudes (ou profondeurs) est stockée dans un **tableau bidimensionnel**. $Temperature = F(temps, altitude)$
- - des mesures géolocalisées, par exemple la température mesurée sur divers points géographiques (latitudes, longitudes) est stockée dans un **tableau bidimensionnel** :
 $Temperature = F(latitude, longitude)$
- La température mesurée sur divers points géographiques
- au cours du temps sera stockée en tant que **tableau Tri-dimensionnel** : $Temperature = F(temps, latitude, longitude)$



2/ Généralités sur NetCDF

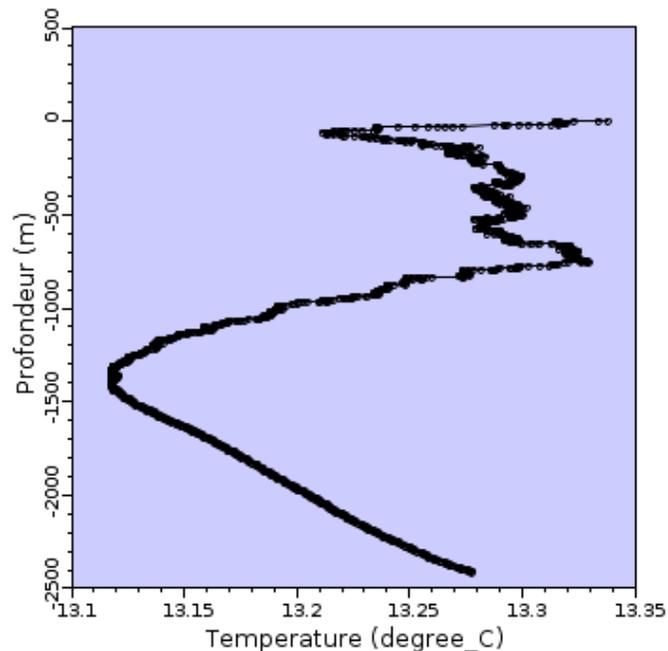
- NetCDF est un format de fichier qui permet de stocker et de représenter des données scientifiques qui évoluent selon certains paramètres qu'on appelle les « **dimensions** »
 - Les **dimensions** sont les axes d'évolution des mesures, qui peuvent être par exemple
 - le temps : on parle alors de séries temporelles (« ou « *timeseries* » en anglais)
 - l'altitude ou la profondeur : on parle alors de "profils" verticaux ("profile")
 - des "trajectoires" lorsque les points de mesures géographiques se déplacent dans le temps et dans l'espace (latitudes, longitudes pour un navire, avion par exemple)
 - Si les mesures se font en 1 seul point : on a des points de mesure
- Les **données** (variables mesurées) sont stockées à l'intérieur du fichier dans des tableaux mono-dimensionnés (vecteurs 1D) , bi-dimensionnés (matrices 2D), 3D ou plus
- Par exemple, pour des données météo on peut ainsi représenter l'évolution de la température, l'humidité, la pression, la vitesse et la direction du vent, etc. selon le temps, ou l'altitude, ou des coordonnées latitudes/longitudes

Exemple de représentations de données dimensionnelles



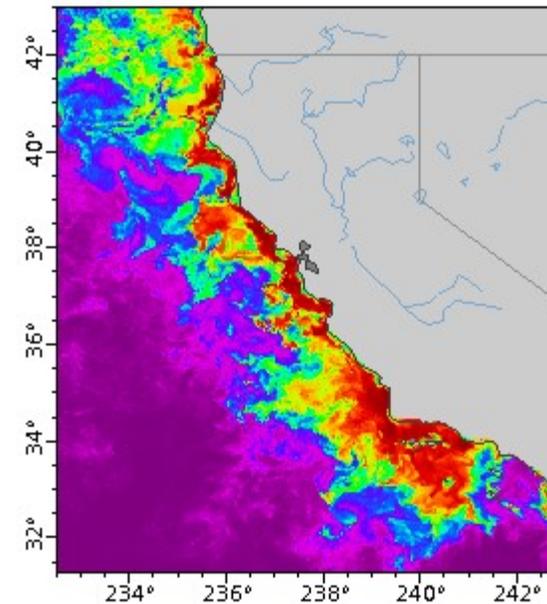
■ **Meteorologie locale en baie de Marseille : site Endoume**
Data courtesy of OSU Pytheas UAR 3470 CNRS

Evolution temporelle de la température (1D)



● **CTD profile ANTARES station (NetCDF files) (lat/long : 42.485/6.06)**
(time=2013-02-19T10:31:00Z, time<=2020-03-09T00:00:00Z)
Data courtesy of MIO UMR 7294 CNRS / OSU Pytheas UMS 3470 CNRS

Profil vertical (1D) de Température en fonction de la profondeur



0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Probability of Pseudo-nitzschia > 10,000 cells/L (1)
C-HARM v1 1-Day Forecast, Pseudo-nitzschia, cellular domoic acid, and particular domoic acid probability, California and Southern Oregon coast, 2018-2022
(2022-03-18T12:00:00Z)
Data courtesy of UCSC, UCSD

Carte 2D d'Abondance géospatialisée d'algues sur la côte ouest des U.S

2/ Généralités sur NetCDF : pourquoi utiliser le format NetCDF ?

- Le format est **auto-descriptif (ou auto-documenté)** : car il inclue un ensemble de **métadonnées** qui fournissent toutes les informations dont on a besoin pour comprendre les données mesurées que contient le fichier,
- Les métadonnées sont par exemple :
 - *Le nom des variables mesurées, les unités de mesure, la date des mesures, les profondeurs, ou les altitudes, la localisation géographique, l'appareil de mesure, le propriétaire, la licence d'accès, la méthodologie etc*
- Le format est **portable, et multi-plateformes**, un fichier netCDF est indépendant d'un système d'exploitation ou d'une machine.
 - Un fichier peut être lu par un même logiciel sur différents systèmes informatiques comme Linux, MacOS, ou Windows
- Il fournit une bibliothèque de dizaines de logiciels libres et gratuits pour lire, écrire et utiliser les fichiers NetCDF
 - Ces programmes permettent différentes actions et traitement sur les fichiers : par exemple accéder à une certaine partie d'un fichier netCDF volumineux sans avoir à lire le fichier entier, découper, assembler, faire des moyennes, visualiser, etc

2/ Généralités sur NetCDF : pourquoi utiliser le format NetCDF ?

- NetCDF est un format qui permet, autant que faire se peut, la **pérennité, la compréhension et la réutilisation des informations** dans le temps
 - Car c'est un format libre et ouvert, supporté depuis les années 80 par une large communauté internationale de scientifiques
 - Netcdf devrait permettre de relire et réutiliser un fichier 10 ans plus tard et de comprendre comment les données ont été produites, mesurées sans avoir recours à un autre support d'information
 - S'il a été bien conçu, le format est auto-documenté, il permet de conserver à l'intérieur du fichier lui même toutes les informations relatives aux données par le biais des métadonnées placées dans l'entête du fichier
- Bien entendu il aura fallu inscrire toutes les métadonnées nécessaires à la compréhension des données dans les sections de métadonnées (**header**) du fichier !
- NetCDF n'impose pas l'utilisation de métadonnées standardisées ! Il est donc utile, préférable et nécessaire de :
 - suivre des **conventions ou des standards** internationaux pour la rédaction de ces métadonnées (voir plus loin la **convention CF**), et d'
 - utiliser des métadonnées provenant de **thésaurus standardisés, disciplinaires**

3/ Comprendre le format NetCDF

- Un fichier NetCDF comprend 4 parties :
 - 1) Une section “***dimensions***” qui spécifie quelles sont les dimensions (nom et valeur) selon lesquelles les mesures évoluent
 - 2) Une section de description de toutes les “***variables***” (paramètres) mesurées, avec de nombreux attributs descriptifs (unités, etc.)
 - 3) Une section d’attributs globaux “***global attributes***” décrivant l’ensemble du fichier avec de nombreuses métadonnées
 - 4) Une section “***data***” : qui contient les valeurs des paramètres mesurés

3/ Comprendre le format NetCDF

Exemple d'entête d'un fichier NetCDF (affiché par le logiciel "panoply") : on y voit le nom et la description des variables mesurées

| Name | Long Name | Type |
|--------------|------------------------------------|------------|
| goro-2.nc | CLAPPP Project : CTD profile (N... | Local File |
| conductivity | conductivity | 1D |
| density | density | 1D |
| depth | Profondeur | 1D |
| fluorescence | fluorescence | 1D |
| irradiance | irradiance | 1D |
| latitude | latitude | — |
| longitude | longitude | — |
| salinity | salinity | 1D |
| stationname | station name | — |
| temperature | Temperature | 1D |
| time | date de prelevement | — |
| turbidity | turbidity | 1D |

File "goro-2.nc"

File type: Hierarchical Data Format, version 5

```
netcdf file:/home/libes/Téléchargements/goro-2.nc {
  dimensions:
    depth = UNLIMITED; // (364 currently)
    lenstation = 17;
  variables:
    int time;
      :long_name = "date de prelevement";
      :standard_name = "time";
      :units = "minutes since 1970-01-01 00:00:00 UTC";
      :origin = "01-JAN-1970 00:00:00";
      :calendar = "standard";

    char stationname(lenstation=17);
      :standard_name = "platform_name";
      :long_name = "station name";
      :cf_role = "profile_id";

    float latitude;
      :units = "degrees_north";
      :standard_name = "latitude";
      :axis = "Y";
      :coverage_content_type = "coordinate";

    float longitude;
      :units = "degrees_east";
      :standard_name = "longitude";
      :axis = "X";
      :coverage_content_type = "coordinate";
```

3/ Comprendre le format NetCDF

3.1/ Le format NetCDF : *les dimensions*

- Dans l'en-tête du fichier, une section “*dimensions*” spécifie quelles sont les dimensions (ou axes) selon lesquelles les mesures évoluent :
- ci dessous par exemple la dimension “*depth*” : 364 points de mesures indique que les mesures présentes dans le fichier évolueront selon la *profondeur*

```
netcdf file:/home/libes/Téléchargements/goro-2.nc {
  dimensions:
    depth = UNLIMITED;    // (364 currently)
    lenstation = 17;
  variables:
    int time;
      :long_name = "date de prelevement";
      :standard_name = "time";
      :units = "minutes since 1970-01-01 00:00:00 UTC";
      :origin = "01-JAN-1970 00:00:00";
      :calendar = "standard";
```

- On peut bien évidemment avoir plusieurs dimensions, lorsque les mesures évoluent selon plusieurs facteurs :
 - comme latitude, longitude, Temps, altitude, profondeur etc...

3/ Comprendre le format NetCDF

3.2) Le format NetCDF : les *variables*

La section “*variable*” : permet de lister exhaustivement toutes les variables mesurées présentes dans le fichier, et de fournir de nombreuses informations à l’aide de divers attributs descriptifs (*métadonnées* sur les variables).

| Name | Long Name |
|--------------|------------------------------|
| goro-2.nc | CLAPPP Project : CTD profile |
| conductivity | conductivity |
| density | density |
| depth | Profondeur |
| fluorescence | fluorescence |
| irradiance | irradiance |
| latitude | latitude |
| longitude | longitude |
| salinity | salinity |
| stationname | station name |
| temperature | Temperature |
| time | date de prelevement |
| turbidity | turbidity |

```
variables:
  int time;
    :long_name = "date de prelevement";
    :standard_name = "time";
    :units = "minutes since 1970-01-01 00:00:00";
    :origin = "01-JAN-1970 00:00:00";
    :calendar = "standard";

  char stationname(lenstation=17);
    :standard_name = "platform_name";
    :long_name = "station name";
    :cf_role = "profile_id";

  float latitude;
    :units = "degrees_north";
    :standard_name = "latitude";
    :axis = "Y";
    :coverage_content_type = "coordinate";
```

```
float temperature(depth=364);
  :standard_name = "sea_water_temperature";
  :units = "Celsius";
  :long_name = "Temperature";
  :source = "Seabird CTD";
  :coverage_content_type = "physicalMeasurement";
  :coordinates = "time longitude latitude depth";
  :_ChunkSizes = 1024U; // uint

float conductivity(depth=364);
  :standard_name = "sea_water_electrical_conductivity";
  :long_name = "conductivity";
  :units = "S.m^-1";
  :coverage_content_type = "physicalMeasurement";
  :source = "Seabird CTD";
  :_ChunkSizes = 1024U; // uint

float salinity(depth=364);
  :standard_name = "sea_water_salinity";
  :long_name = "salinity";
  :station = "CLAPPP";
  :units = "1e-3";
  :coverage_content_type = "physicalMeasurement";
  :coordinates = "time longitude latitude depth";
```

3/ Comprendre le format NetCDF

3.3) Le format NetCDF : les attributs globaux

Une section d'attributs globaux (*global attributes*) fournit les métadonnées globales, donnant de nombreuses informations sur le jeu de données dans son ensemble :

- *Projet, Titre, Résumé*
- *Licence d'utilisation*
- *Contact, propriétaire, créateur,*
- *Distributeur*
- *Méthodologie,*
- *Mots clés*
- *DOI (identifiant pérenne)*
- *etc...*

```
// global attributes:
:description = "CTD profile (NetCDF files) for station ./Clapp5//goro-2";
:title = "CLAPPP Project : CTD profile (NetCDF files) station ./Clapp5//goro-2";
:keywords = "Seabird CTD SBE19plus, temperature, conductivity, fluorimetry, salin";
:history = "Created 02/06/21";
:production = "MIO UMR 7294 CNRS / OSU Pytheas UMS3470 CNRS";
:contact = "martine Rodier (martine.rodier@mio.osupytheas.fr)";
:infoUrl = "https://dataset.osupytheas.fr/geonetwork/srv/fre/catalog.search#/meta";
:references = "https://www.mio.osupytheas.fr/";
:summary = "Campagne ./Clapp5/ : Mesures CTD SBE19plus";
:featureType = "profile";
:cdm_data_type = "profile";
:cdm_profile_variables = "stationname,depth";
:Conventions = "CF-1.6, ACDD-1.3";
:program = "CLAPPP New Caledonian lagoons: Physics and Phytoplankton processes";
:project = "Campagne ./Clapp5/ : New Caledonian lagoons: Physics and Phytoplankton";
:source = "water column measurements observation by Seabird SBE19plus CTD Measure";
:doi = "https://doi.org/10.34930/2b52defe-e5f3-4fe2-9f2f-741d90e624ea";
:processing_level = "1a";
:comment = "Seabird CTD SBE19plus Measurements : NetCDF files";
:acknowledgement = "";
:license = "data are accessible freely and free of cost, for public research and";
:creator_name = "martine Rodier : MIO Marine Institute of Oceanography UMR 7294 (";
:creator_email = "martine.rodier@mio.osupytheas.fr";
:creator_url = "http://www.mio.osupytheas.fr";
:creator_type = "institution";
:creator_institution = "MIO Marine Institute of Oceanography UMR 7294 CNRS / OSU";
:contributor_name = "OSU Pytheas UMS 3470 CNRS";
:contributor_role = "data formatting in netCDF";
:contributor_email = "maurice.libes@osupytheas.fr";
:contributor_url = "https://www.osupytheas.fr/";
```

3/ Comprendre le format NetCDF

4.4) Le format NetCDF : *les valeurs*

- La section “**data**” du fichier : fournit *in fine* les valeurs des variables elles-mêmes, qui ont été référencées dans la section “*variable*” précédente

```
data:
time = 21486822 ;
stationname = "./Clapp5//goro-2" ;
latitude = -22.33 ;
longitude = 167.01 ;

depth = -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9, -1, -1.1, -1.2, -1.3,
-1.4, -1.5, -1.6, -1.7, -1.8, -1.9, -2, -2.1, -2.2, -2.3, -2.4, -2.5,
-2.6, -2.7, -2.8, -2.9, -3, -3.1, -3.2, -3.3, -3.5, -3.6, -3.7, -3.8,
-3.9, -4, -4.1, -4.2, -4.3, -4.4, -4.6, -4.7, -4.8, -4.9, -5, -5.1, -5.2,
-5.3, -5.4, -5.5, -5.6, -5.7, -5.8, -5.9, -6, -6.1, -6.2, -6.3, -6.4,
-6.5, -6.6, -6.7, -6.8, -6.9, -7, -7.1, -7.2, -7.3, -7.4, -7.5, -7.6,
```

Valeur du fichier affichées par la commande “*ncdump*”

```
temperature = 24.169, 24.202, 24.221, 24.225, 24.194, 24.16, 24.137, 24.109,
24.114, 24.128, 24.129, 24.13, 24.131, 24.129, 24.127, 24.128, 24.13,
24.132, 24.131, 24.129, 24.127, 24.129, 24.136, 24.14, 24.138, 24.131,
24.13, 24.127, 24.127, 24.14, 24.147, 24.149, 24.131, 24.126, 24.124,
24.123, 24.122, 24.122, 24.124, 24.127, 24.125, 24.124, 24.125, 24.127,
24.127, 24.125, 24.125, 24.124, 24.122, 24.121, 24.12, 24.123, 24.124,
24.122, 24.121, 24.122, 24.124, 24.13, 24.139, 24.146, 24.148, 24.148,
24.148, 24.146, 24.133, 24.128, 24.126, 24.124, 24.116, 24.118, 24.116,
24.115, 24.117, 24.119, 24.121, 24.12, 24.118, 24.116, 24.117, 24.118,
24.119, 24.117, 24.115, 24.114, 24.115, 24.116, 24.115, 24.116, 24.118,
24.121, 24.121, 24.12, 24.116, 24.109, 24.11, 24.108, 24.107, 24.106,
```

4/ Les outils logiciels de NetCDF

4.1) Les commandes de base de NetCDF

NetCDF est un format qui s'utilise avec des **commandes et programmes spécifiques à NetCDF**. On ne peut pas éditer ou lire un fichier NetCDF binaire avec des outils de bureautiques courants.

Il existe de nombreux outils qui permettent de visualiser le contenu d'un fichier NetCDF.

- Le site <https://www.unidata.ucar.edu/software/netcdf/software.html> liste un vaste ensemble de programmes permettant d'utiliser le format NetCDF

Cependant les **commandes de base sont disponibles directement dans la distribution de base**

- Exemple sous Linux Debian/ubuntu/Mint le paquet ***netcdf-bin*** contient les commandes basiques pour lire et créer des fichiers NetCDF
 - ***netcdf-bin***
 - ➔ ***ncdump*** : pour lire et afficher le contenu d'un fichier NetCDF
 - ➔ ***nccopy*** : pour copier tout ou partie d'un fichier .nc dans un autre
 - ➔ ***ncgen*** : pour créer un fichier NetCDF binaire à partir d'une description texte d'un fichier NetCDF (issu de *ncdump* par exemple)
 - ➔ ***ncgen3*** : génère le code en langage C ou fortran pour produire un fichier NetCDF

4.1/ Les commandes de base

pour lire et écrire un fichier NetCDF

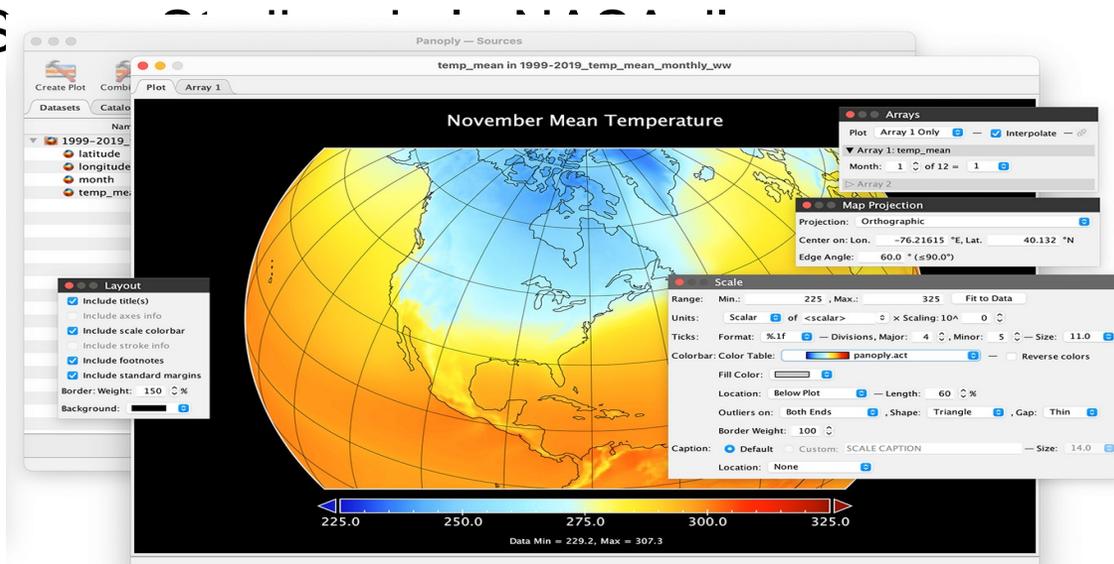
- La commande ***ncdump*** est la commande de base la plus utile pour lire et afficher un fichier NetCDF binaire sous forme textuelle. La sortie “texte” qui s’affiche est alors “lisible” par un être humain.

Cette représentation “texte” du fichier NetCDF binaire s’appelle le **CDL** (Common Data Language). Cette représentation permet d’avoir ***une conversion en texte lisible, à partir du format binaire***

- La commande ***ncgen*** quant à elle, permet à l’inverse de créer un fichier binaire NetCDF à partir de la description texte en CDL
 - Affiche le fichier NetCDF binaire sous forme textuelle CDL
 - ***# ncdump monfichier.nc > monfichier-cdl.txt***
 - Régénère un fichier NetCDF .nc à partir de la description texte .cdl précédente
 - ***# ncgen monfichier-cdl.txt -o monfichier2.nc***

4.3/ Les outils NetCDF avancés et interactifs

- ***En tant que simple utilisateur, si vous devez lire un fichier NetCDF et visualiser les données,***
- ***“**panoply**” est le logiciel le plus simple qui ne nécessite pas de savoir taper des commandes en ligne***
- ***Panoply*** est un visualisateur de fichiers NetCDF.
- Développé en “*java*” par le Goddard Institute for S fonctionne sur tous les systemes d’exploitation et aisée des fichiers NetCDF
- <https://www.giss.nasa.gov/tools/panoply/>



4.2/ Les opérateurs NCO de NetCDF

La suite NCO (NetCDF Operator) permet de manipuler les fichiers NetCDF et leurs contenus et de réaliser des opérations plus complexes sur les données des fichiers : *par exemple, dériver de nouveaux champs, calculer des statistiques, imprimer, manipuler des métadonnées, regriller) et produisent les résultats à l'écran ou dans des fichiers aux formats texte, binaire ou netCDF.*

NCO facilite l'analyse des données scientifiques maillées (gridded). Le style de commande shell de NCO permet aux utilisateurs de manipuler et d'analyser des fichiers de manière interactive, ou à l'aide de scripts qui évitent d'avoir à utiliser des langages, ou environnements de programmation de niveau supérieur.

- <http://nco.sourceforge.net/>
- Le paquet de commande est disponible sur Linux et contient les commandes suivantes
 - ncap2 - arithmetic processor**
 - ncatted - attribute editor**
 - ncbo - binary operator**
 - ncdiff - differencer**
 - ncea - ensemble averager**
 - necat - ensemble concatenator**
 - ncflint - file interpolator**
 - ncks - kitchen sink (extract, cut, paste, print data)**
 - ncpdq - permute dimensions quickly**
 - ncra - running averager**
 - ncrcat - record concatenator**
 - ncrename - renamer**
 - ncwa - weighted averager**

4.3/ Les outils NetCDF avancés et interactifs

Ferret est un outil d'analyse et de visualisation de fichiers netCDF qui répond aux besoins des océanographes et des météorologues qui analysent de grands ensembles complexes de données maillées 2D

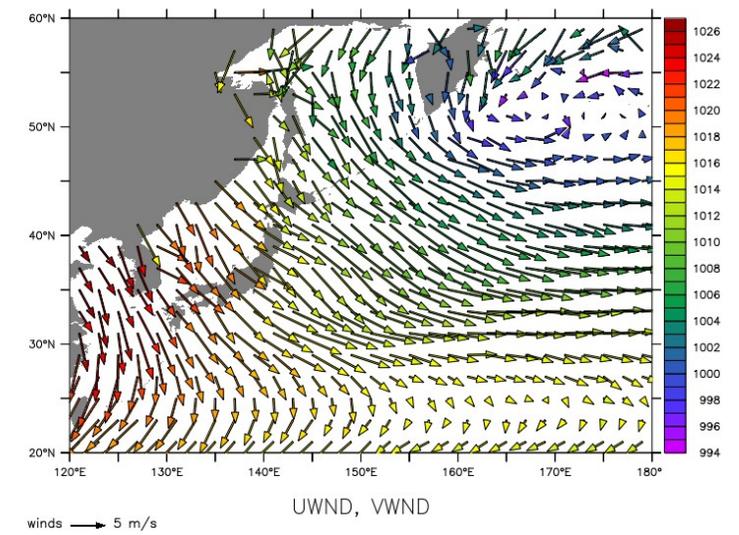
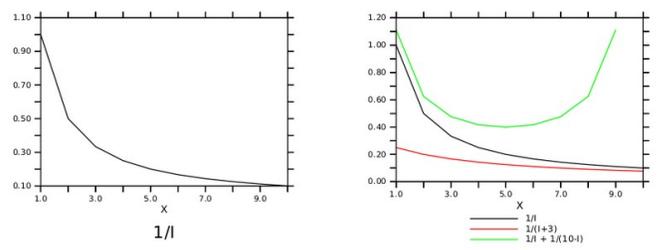
- Disponible sur <https://ferret.pmel.noaa.gov/Ferret/>
- Installation sous Linux Debian
 - `# apt-get install python3-ferret`

```
yes? LET press = YSEQUENCE(slp[X=120:180,Y=20:60,L=1])

! draw the polygons
yes? POLYGON/OVER/NOAX/NOLAB/KEY/LINE/PAL=rainbow plt_x_arrow, plt_y_arrow, press

yes? GO poly_arrow_key "winds", "5 m/s"
```

```
yes? CANCEL MODE LOGO ! clean up: don't draw the Ferret logo in the upper left
yes? SET WINDOW/SIZE=.9 ! (bigger 'cause there will be 4 plots)
yes? SET VIEWPORT UL
yes? PLOT/i=1:10 1/i
yes? SET VIEWPORT LL
yes? PLOT/i=1:10/SYMBOLS 1/i
yes? SET VIEWPORT LR
yes? PLOT/i=1:10/SYMBOLS=2/LINE/VLIMITS=-1:2:0.25 iinverse
yes? SET VIEWPORT UR
yes? PLOT/i=1:10 1/i, 1/(i+3), 1/i + 1/(10-i)
```



4.3/ Les outils logiciels de NetCDF

installation sur différents systemes

- <https://downloads.unidata.ucar.edu/netcdf/>
- <https://docs.unidata.ucar.edu/netcdf-c/current/faq.html#HowdoIgetthenetCDFsoftwarepackage>
- Les distributions netCDF fournissent un ensemble de commandes logicielles qui permettent la création, la lecture et accès, et le partage des données scientifiques NetCDF organisées en tableaux
- Sur linux Debian | Ubuntu | Mint :
 - `# sudo apt-get install netcdf-bin`
 - `# sudo apt-get install nco`
- Sur Windows : récupérer et installer le .exe fourni
 - <https://docs.unidata.ucar.edu/netcdf-c/current/winbin.html>
 - <https://downloads.unidata.ucar.edu/netcdf-c/4.9.0/netCDF4.9.0-NC3-64.exe>
- Sur MacOS : avec l'utilitaire d'installation "brew" à installer depuis <http://brew.sh/index.html>
 - `brew tap homebrew/science`
 - `brew install netcdf`
 - `brew install moffat/sciencebits/cdo`
 - `brew install nco`

5/ NetCDF et la convention CF

- La bibliothèque NetCDF est conçue pour **décrire des données**, structurées selon des règles bien définies dans l'entête du fichier, et qui facilitent l'interopérabilité sur diverses plateformes informatiques.
- Pour autant, le format NetCDF permet mais n'exige pas la création d'ensembles de données autodescriptives.
- L'objectif des conventions CF est de **standardiser la description des données pour qu'ils contiennent suffisamment de métadonnées et soient autodescriptifs** afin de les rendre réutilisable par n'importe quelle autre personne.
- Ainsi CF demande à ce que chaque variable du fichier doit avoir **une description associée de ce qu'elle représente, avec les unités physiques si nécessaire, et la localisation de chaque mesure dans l'espace (par rapport aux coordonnées terrestres) et dans le temps.**

5/ NetCDF et la convention CF

- Les conventions de métadonnées CF répondent à un ***besoin de standardisation*** à faire dans les différentes communautés. Elles sont conçues pour ***promouvoir le traitement et le partage des fichiers*** créés avec l'API NetCDF.
- Les conventions définissent des métadonnées qui fournissent une description standardisée de ce que représentent les données dans chaque variable, ainsi que les ***propriétés spatiales et temporelles des données***.
- Cela permet aux utilisateurs de données provenant de différentes sources de pouvoir comparer des données, et facilite la création d'applications dotées de capacités d'extraction, de recadrage et d'affichage.
- La convention CF utilise ***une table de noms standard***, qui définit les dénominations permettant d'identifier les noms des variables et les quantités physiques, unités de manière standardisée

5/ NetCDF et la convention CF

- La convention CF utilise une table de noms standards qui permet de fournir des préconisations précises pour la dénomination et la description des variables que contient un fichier NetCDF
https://cfconventions.org/Data/cf-conventions/cf-conventions-1.9/cf-conventions.html#_des_cription_of_the_data
- **long_name** : le nom commun d'une variable "temperature" par exemple
- **standard_name** : le nom standardisé par la convention CF permet d'utiliser un nom de variable qui a un sens international dans une communauté disciplinaire
 - <https://cfconventions.org/Data/cf-standard-names/79/build/cf-standard-name-table.html>
- **units** : les unités des mesures du parametre considéré
- **missing_value** : la valeur des mesures manquantes -9999.999 par exemple
- Variables de coordonnées
- Etc ,

5/ NetCDF et la convention CF

Exemples de “standard name” pour la dénomination de la température de l'eau

▼ [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 `sea_surface_skin_temperature`, `sea_surface_subskin_temperature`, and `surface_temperature` 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 `sea_water_temperature` with a vertical coordinate axis should be used.

▶ [sea_water_added_conservative_temperature](#)

▶ [sea_water_added_potential_temperature](#)

▶ [sea_water_conservative_temperature](#)

▶ [sea_water_potential_temperature](#)

▶ [sea_water_potential_temperature_at_sea_floor](#)

▶ [sea_water_potential_temperature_expressed_as_heat_content](#)

alias: `integral_wrt_depth_of_sea_water_potential_temperature_expressed_as_heat_content`
alias: `integral_of_sea_water_potential_temperature_wrt_depth_expressed_as_heat_content`

▶ [sea_water_redistributed_conservative_temperature](#)

▶ [sea_water_redistributed_potential_temperature](#)

▼ [sea_water_temperature](#)

Sea water temperature is the in situ temperature of the sea water. To specify the depth at which the temperature applies use a vertical coordinate variable or scalar coordinate variable. There are standard names for `sea_surface_temperature`, `sea_surface_skin_temperature`, `sea_surface_subskin_temperature` and `sea_surface_foundation_temperature` which can be used to describe data located at the specified surfaces. For observed data, depending on the period during which the observation was made, the measured in situ temperature was recorded against standard "scales". These historical scales include the International Practical Temperature Scale of 1948 (IPTS-48; 1948-1967), the International Practical Temperature Scale of 1968 (IPTS-68, Barber, 1969; 1968-1989) and the International Temperature Scale of 1990 (ITS-90, Saunders 1990; 1990 onwards). Conversion of data between these scales follows $t_{68} = t_{48} - (4.4 \times 10^{-6}) * t_{48}(100 - t_{48})$; $t_{90} = 0.99976 * t_{68}$. Observations made prior to 1948 (IPTS-48) have not been documented and therefore a conversion cannot be certain. Differences between t_{90} and t_{68} can be up to 0.01 at temperatures of 40 C and above; differences of 0.002-0.007 occur across the standard range of ocean temperatures (-10 - 30 C). The International Equation of State of Seawater 1980 (EOS-80, UNESCO, 1981) and the Practical Salinity Scale (PSS-78) were both based on IPTS-68, while the Thermodynamic Equation of Seawater 2010 (TEOS-10) is based on ITS-90. References: Barber, 1969, doi: 10.1088/0026-1394/5/2/001; UNESCO, 1981; Saunders, 1990, WOCE Newsletter, 10, September 1990.

6/ NetCDF et le protocole Data Access Protocol - DAP

- Les fichiers NetCDF facilitent l'interopérabilité et la réutilisation car ils sont bien appropriés pour être accessibles par un protocole comme [DAP \(Data Access Protocol\)](#) et être diffusés en ligne par des logiciels comme
 - [Thredds](#),
 - [Erddap](#),
 - Ferret,
 - Matlab
- Il suffit d'une URL DAP bien formée pour que les données de fichiers NetCDF fournies par un serveur, arrivent dans votre logiciel d'analyse préféré comme Matlab, Jupyter Notebook, R, Ferret etc..

6/ NetCDF et le protocole Data Access Protocol - DAP

Les fichiers NetCDF sont bien appropriés pour être utilisés par le protocole DAP Data Access Protocol -

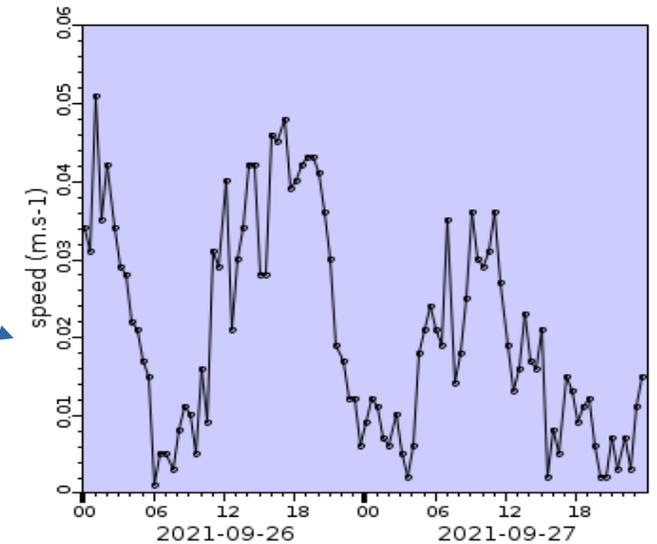
Un utilisateur (ou programme) peut construire une URL avec des expressions de contraintes élaborées qui envoyées à un serveur DAP, renverront précisément les données demandées contenues dans un fichier NetCDF

- `https://erddap.osupytheas.fr/erddap/tabledap/Emso_Ligure_Ouest_Albatross_Aquadopp_NetCDF_2021.graph?time,Speed&depth=500.0&time>=2021-09-26T00:00:00Z&time<2021-09-28T00:00:00Z&.draw=linesAndMarkers&.marker=6|3&.color=0x000000&.colorBar=%7C%7C%7C%7C%7C&.bgColor=0xffccccff`
- Le serveur : `https://erddap.osupytheas.fr/erddap/tabledap`
- Le dataset : `Emso_Ligure_Ouest_Albatross_Aquadopp_NetCDF_2021`
- Les données à extraire : `time,Speed`
- Les filtres : `depth=500.0&time>=2021-09-26T00:00:00Z&time<2021-09-28T00:00:00Z`

6/ NetCDF et le protocole *Data Access Protocol - DAP*

- URL : https://erddap.osupytheas.fr/erddap/tabledap/Emso_Ligure_Ouest_Albatross_Aquadopp_NetCDF_2021.graph?time,Speed&depth=500.0&time>=2021-09-26T00:00:00Z&time<2021-09-28T00:00:00Z&.draw=linesAndMarkers&.marker=6|3&.color=0x000000&.colorBar=%7C%7C%7C%7C%7C&.bgColor=0xffccccff
- **Le serveur :**
<https://erddap.osupytheas.fr/erddap/tabledap>
- **Le dataset :**
Emso_Ligure_Ouest_Albatross_Aquadopp_NetCDF_2021
- **Les données à extraire :** time, Speed
- **Les filtres :** depth=500.0&time>=2021-09-26T00:00:00Z&time<2021-09-28T00:00:00Z

- Les fichiers NetCDF sont bien appropriés pour être utilisés par le protocole **DAP Data Access Protocol**
- Un utilisateur (ou programme) peut construire une URL avec des expressions de contraintes élaborées qui, envoyées à un serveur DAP, renverront précisément les données demandées contenues dans un fichier NetCDF



—○— EMSO Ligure Ouest : ALBATROSS capteur AQUADOPP (NetCDF files from 2021-07) (depth=500.0)
Data courtesy of MIO UMR7294 CNRS / OSU Pytheas

7/ NetCDF dans le contexte de la science ouverte

Dans ce contexte de science ouverte, **la diffusion, le partage et la réutilisation des données** sont au coeur de la démarche.

- On parle de gestion de données "FAIR" afin que les données soient
 - "F"aciles à trouver
 - "A"ccessibles
 - "I"nteropérables
 - "R"éutilisables

7/ NetCDF dans le contexte de la science ouverte

- C'est dans ce contexte d'interopérabilité demandé par la Science Ouverte, que **[le format NetCDF](#) est préconisé par rapport à d'autres formats.**
- Le format est recommandé dans l'Infrastructure de Recherches "Data Terra" par le [pôle de données Odatis](#)
- <https://www.odatis-ocean.fr/donnees-et-services/principes-de-gestion-des-donnees/formats-attributs-conventions>
- —

Example of NetCDF gridded file:

```
netcdf nrt_mozambique_alust_phy_i4_20190130_20190205 {
  dimensions:
    time = 1 ;
    latitude = 340 ;
    longitude = 340 ;
    nv = 2 ;
  variables:
    int cns ;
    cns:comment = "This is a container variable that describes the grid_mapping used by the data in this file. This variable does not contain any data; only information about the geographic coordinate system." ;
    cns:grid_mapping_name = "latitude_longitude" ;
    cns:inverse_flattening = 298.257 ;
    cns:semi_major_axis = 6378136.3 ;
    float time(time) ;
    time:axis = "T" ;
    time:calendar = "gregorian" ;
    time:long_name = "Time" ;
    time:standard_name = "time" ;
    time:units = "days since 1950-01-01 00:00:00" ;
    float latitude(latitude) ;
    latitude:axis = "Y" ;
    latitude:bounds = "lat_bnds" ;
    latitude:long_name = "Latitude" ;
    latitude:standard_name = "latitude" ;
    latitude:units = "degrees_north" ;
    latitude:valid_max = -0.125 ;
    latitude:valid_min = -30. ;
    float lat_bnds(latitude, nv) ;
    lat_bnds:comment = "latitude values at the north and south bounds of each pixel." ;
    lat_bnds:units = "degrees_north" ;
    float longitude(longitude) ;
    longitude:axis = "X" ;
    longitude:bounds = "lon_bnds" ;
    longitude:long_name = "Longitude" ;
    longitude:standard_name = "longitude" ;
    longitude:units = "degrees_east" ;
    longitude:valid_max = 59.675 ;
    longitude:valid_min = 30. ;
    float lon_bnds(longitude, nv) ;
    lon_bnds:comment = "longitude values at the west and east bounds of each pixel." ;
    lon_bnds:units = "degrees_east" ;
    int nv(nv) ;
    nv:comment = "Vertex" ;
    nv:units = "1" ;
```

Références utiles

- <https://www.unidata.ucar.edu/software/netcdf/>
- <https://fr.wikipedia.org/wiki/NetCDF>
- <https://www.unidata.ucar.edu/software/netcdf/software.html>
- <https://cfconventions.org/>
- https://www.odatis-ocean.fr/fileadmin/documents/activites/ateliers/atelier_201910/201910_ODATIS_Atelier_JSudre_netcdf.pdf
- Outils de visualisation
 - Ferret: <https://ferret.pmel.noaa.gov/Ferret/>
 - Panoply: <https://www.giss.nasa.gov/tools/panoply/download/>