SeaDataNet NetCDF for Point Data

Roy Lowry

British Oceanographic Data Centre

Presentation Overview

- What are NetCDF and CF?
- SeaDataNet profiling of CF1.6
- Interoperability
- Parameter Naming
- Mapping data types onto SeaDataNet NetCDF
- Further profiling of SeaDataNet NetCDF

NetCDF

- Network Common Data Form (NetCDF) is an interface for array-oriented data access and a library that provides an implementation of the interface.

- What does this mean?
  - NetCDF is a data model rather than a data format
  - Data in NetCDF are accessed by API or tools calling that API

NetCDF

- Practical points
  - Binary data format
  - Details of underlying format depends upon API library version (NetCDF3, NetCDF4, HDF5)
  - API initially developed for C and Fortran
  - Matlab and Java now available
  - NetCDF developed and supported by UniData in Boulder Colorado

CF Conventions

- Climate and Forecast Metadata Conventions (CF) is a ‘profile’ of the NetCDF data model designed to enhance data interoperability
- Developed by the global atmosphere-ocean coupled modelling community to provide interoperability for IPCC intercomparisons
- Governed on an open community model based around e-mail list server and Trac moderated by a committee with voting rights (should they ever be needed)
- Adopted by SeaDataNet

CF Conventions

- A ‘profile’ reduces the number of degrees of freedom in a data model by:
  - Modelling the model
    - CF concepts of co-ordinate, geophysical and ancillary variables
    - Global and parameter attribute naming and status (mandatory, conditional, recommended)
    - Establishing rules (e.g. Geophysical variable must be linked to at least one co-ordinate variable)
  - Standardising semantics through controlled vocabularies
    - Standard Names
    - Cell Methods

CF Conventions

- CF developed by GCM modelling community
  - Versions 1.0 to 1.5 only supported gridded data
  - Strong atmospheric science bias in the design principles and the semantic resources like Standard Names
- CF 1.6 developed by John Carron of UniData to support 'point data'
- CF1.6 based on the concept of 'feature type' defined by co-ordinate variable patterns

SeaDataNet NetCDF

- Two types of NetCDF used in SeaDataNet
  - CF 1.5 profile for gridded data (L24 = CF) - largely out of scope for this presentation
    - CF1.5 plus SeaDataNet P01/P06 parameter attributes and metadata linkage arrays (SDN_CRUISE, SDN_STATION, SDN_LOCAL_CDI_ID and SDN_EDMO_CODE) from CFPOINT
  - CF 1.6 profile for point data (L24 = CFPOINT)
- CFPOINT profile modifies CF 1.6 as follows
  - Point feature type excluded
  - Additional conventions
  - Additional mandatory attributes and metadata arrays
  - Additional optional attributes and metadata arrays

SeaDataNet NetCDF

- Additional conventions
  - Representation of time is UT expressed as Chronological Julian Day
  - Representation of position is actual or assumed EPSG4326 (WGS-84 2D geographic co-ordinates)
  - Mandatory depth/height for time series
  - Names of x, y and z co-ordinate variables are fixed
  - ‘Position’ SeaDataNet QC flag ancillary variable for latitude+longitude
  - SeaDataNet QC flag ancillary variables for all other co-ordinate and geophysical variables named 'variable_SEADATANET_QC'

SeaDataNet NetCDF

- Mandatory extensions

  - Parameter attributes
    - sdn_parameter_urn (P01 URN)
    - sdn_parameter_name (P01 entryTerm)
    - sdn_uom_urn (P06 URN)
    - sdn_uom_name (P06 entryTerm)

SeaDataNet NetCDF

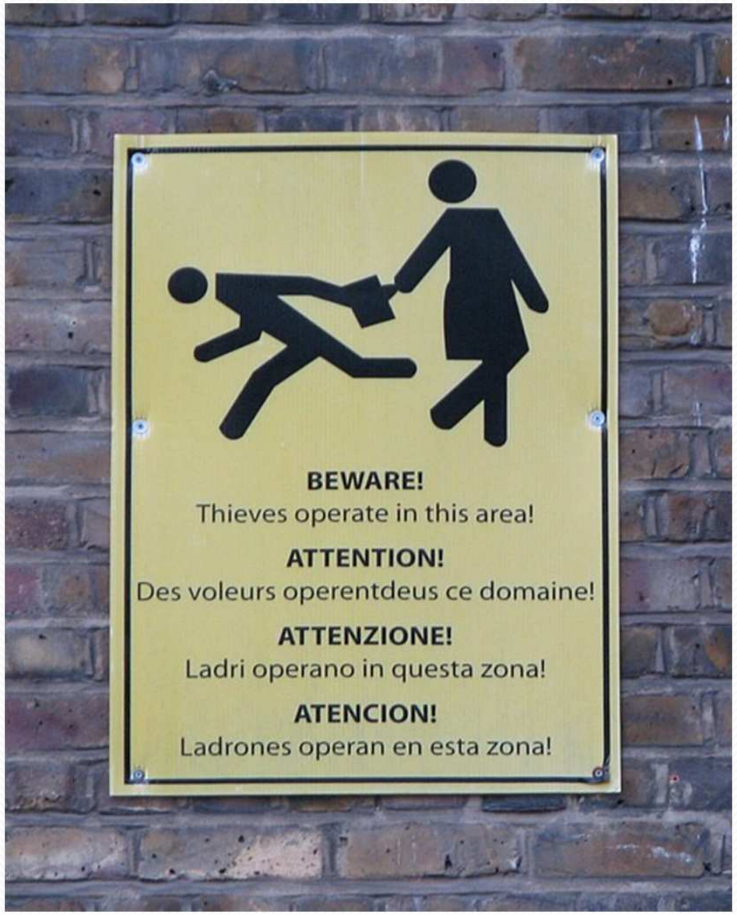
- Mandatory extensions
  - Metadata arrays
    - crs (EPSG4326 declaration)
    - SDN_CRUISE (grouping label e.g. cruise for profiles)
    - SDN_STATION (series label e.g. profile name)
    - SDN_LOCAL_CDI_ID (local CDI identifier)
    - SDN_EDMO_CODE (namespace for local CDI)

SeaDataNet NetCDF

- Optional extensions
  - Parameter attributes
    - sdn_instrument_urn (L22 URN)
    - sdn_instrument_name (L22 entryTerm)
  - Metadata arrays
    - SDN_BOT_DEPTH (water depth: mandatory for profiles)
    - SDN_REFERENCES (single XHTML URI - backwards compatibility)
    - SDN_XLINK (multiple XML/XHTML document linkages following XLINK)

Interoperability

- Design Principle
  - Interoperability through duplication
- Examples
  - Multilingual signs
  - CF Standard Names AND P01 URNs in SeaDataNet NetCDF


BEWARE!
Thieves operate in this area!
ATTENTION!
Des voleurs operentdeus ce domaine!
ATTENZIONE!
Ladri operano in questa zona!
ATENCION!
Ladrones operan en esta zona!

Interoperability

- Flagging conventions are a notorious area of non-interoperability
  - SeaDataNet
  - OceanSites
  - ODV
  - ARGO
  - QARTOD
  - IODE
- 16 conventions documented in L27 vocabulary

Interoperability

- One solution is to attach multiple CF ancillary variable arrays to each co-ordinate or geophysical variable
- Shifts the load from the data consumer to the data provider
- If mapping is involved the data provider is more likely to get it right!
- SeaDataNet NetCDF specification is the LOWEST COMMON DENOMINATOR
- Embellishment encouraged is it make the data more interoperable

Parameter Naming

- CF specifies two parameter attributes for naming each parameter
  - Standard Name (controlled vocabulary)
  - Long Name (free text)
- Standard Names have issues
  - Coverage atmospheric rather than oceanographic
  - Content governance is out of SeaDataNet control and broken

Parameter Naming

- SeaDataNet solution
  - Make the Standard Name optional, but strongly recommend its use for the sake of interoperability
  - Therefore Long Name is mandatory (CF rules)
  - Add SeaDataNet namespace parameter attributes
    - sdn_parameter_urn (P01 URN)
    - sdn_parameter_name (P01 entryTerm)
    - sdn_uom_urn (P06 URN)
    - sdn_uom_name (P06 entryTerm)

Parameter Naming

- SeaDataNet vocabulary URNs
  - Syntax is 'SDN:' || Library || '::' || ConceptID
  - Example - SDN:P01::TEMPPR01
- SeaDataNet names
  - Concept entryTerm
  - Example - Temperature of the water body
- Likewise for units (P06 vocabulary) and instruments (L22 vocabulary)

Mapping Data Types

- There are five types of SeaDataNet NetCDF file
- Correspond to the CF 1.6 Feature Types
  - profile
  - timeSeries
  - trajectory
  - timeSeriesProfile
  - trajectoryProfile
- Data need to be mapped onto one of these
- INSTANCE dimension for CF1.6 compatibility
  - Allows multiple series per file
  - Set to 1 for normal SeaDataNet objects

Mapping Data Types

- profile
    - Measurements in one place, one time, many depths
    - Examples are CTD, XBT, bottle data, radiosondes, etc.
    - SeaDataNet considers z co-ordinate to always be spatial for profiles (depth, pressure, height)

Mapping Data Types

- timeSeries
  - Measurements in one place, one depth, but at many different times
  - Examples are recording current meter record, seafloor pressure gauge, moored temperature logger
  - SeaDataNet requires a height/depth co-ordinate variable with one value per series (nominal covering all time steps)
  - Also possible to include pressure or depth as a geophysical variable (value for each time step), but this is NOT a co-ordinate variable

Mapping Data Types

- trajectory
  - Like timeSeries but measurements in a different place for each time step
  - Can be 2D or 3D. Either can have a depth/height/pressure co-ordinate variable with one value per time step
  - Only one z co-ordinate value per time step
  - Examples are thermosalinograph or AUV-mounted CTD data

Mapping Data Types

- timeSeriesProfile and trajectoryProfile

  - Conceptually a little more difficult
    - Think of timeSeries or trajectory as referring to the platform
    - Tagged onto these are
      - Profile z co-ordinates for each time step
      - Data values for each profile z co-ordinate value at each time step
      - Profile z co-ordinates can be spatial (e.g. ADCP bin depth) but non-spatial parameters (e.g. wavelength) are also permitted

Mapping Data Types

- timeSeriesProfile
  - Platform is in the same place for every time step
  - One depth value per series required
  - Depth array (INSTANCE, MAXT) with one value per time step permitted as a geophysical variable
  - Example is moored ADCP data
    - PROFZ - profile Z co-ordinate array (INSTANCE, MAXT, MAXZ) contains bin depths
    - Geophysical variables
      - For each bin - e.g. current velocity (INSTANCE, MAXT, MAXZ)
      - One per time step - e.g. instrument water temperature sensor (INSTANCE,MAXT)
  - Also fits thermistor chains, particle sizer time series and spectral radiometer time series

Mapping Data Types

- trajectoryProfile
  - Platform in a different place for each time step
  - Platform latitude, longitude and depth (INSTANCE, MAXT) arrays (so will cope with AUV ADCP data)
  - Example is vessel-mounted ADCP data
    - PROFZ - profile Z co-ordinate array (INSTANCE, MAXT, MAXZ) contains bin depths
    - Geophysical variables
      - For each bin - e.g. current velocity (INSTANCE, MAXT, MAXZ)
      - One per time step - e.g. bottom-tracking data (INSTANCE,MAXT)
  - Also fits towed thermistor chains, AUV-mounted optical plankton counter or particle sizer

Mapping Data Types

- Introducing a little bit of controversy
  - The ODV extension for biological data introduces the idea of species as a dimension
  - Therefore PROFZ could be a LSID or numeric equivalent such AphiaID
  - A zoobenthos biomass survey could therefore be mapped onto trajectoryProfile SeaDataNet NetCDF
  - Biological data in a binary format!!!
  - Would need an Excel import tool for it to be accepted

SeaDataNet NetCDF Profiling

- SeaDataNet format specification is a lowest common denominator
- Further restrictions and extensions may be specified to cover requirements for specific types of data
- For example:
  - Making geophysical variables mandatory
    - Platform velocity required for relative wind direction data
  - Mandatory data-type specific metadata
    - SDN_XLINK in XBT data to the MO C3 vocabulary to provide access to the drop rate equation
    - SDN_XLINK in cruise data to CSR document providing access to information like PSO name

SeaDataNet NetCDF Profiling

- Nemo software already profiles the basic SeaDataNet specification through the inclusion of additional global attributes
- Should a manual documenting SeaDataNet format for every type of oceanographic data (CTD, moored ADCP, VMADCP, etc.) be produced?
  - Requires extensive domain knowledge
  - Stretches capabilities of technical expert groups like SeaDataNet TTT
  - Would have to be produced by user communities

SeaDataNet NetCDF Profiling

- Documenting specific data types
  - A better approach might be:
    - Identify community best practice
    - Capture it
    - Build a library of example data files

That's All Folks!!