



# **BODC requirements for climate and forecast (CF) NetCDF metadata from numerical models**

## **Introduction**

BODC aims to ensure that all numerical model data incorporated into our databases are adequately documented to allow long term viability and future access. Therefore, key data set information (metadata) are important for each data set we receive.

Climate and forecast (CF) metadata conventions have been adopted by a wide range of organisations as a primary standard for self-describing NetCDF files.

BODC are incorporating CF metadata into our procedures, and will therefore accept CF compliant NetCDF format files for numerical model data, but only with some stipulations on how variable attributes are defined. These are necessary to ensure that any CF-compliant NetCDF files being sent to BODC have a standardised metadata structure, which will make BODC processing more efficient.

Please note that the BODC requirements are in addition to the requirements set down in the CF Conventions themselves.

## **BODC CF and other attribute requirements (requirements are necessary for each data file)**

1. **Time calendar reference.** All numerical model sent to BODC must have a time channel that is based on time elapsed in days since 00:00 01 January 4713 BC. This ensures that the time channel is interoperable with BODC observations, and is in keeping with SeaDataNet requirements (see 7 below).
2. **Presence of attributes.** All variables in the NetCDF file must have at least one attribute in the CF header.
3. **Standard\_name and long\_name attributes.** Each variable being included in the NetCDF file must have a long\_name attribute in the CF header in order to provide a description of what it represents. If one exists, the appropriate standard\_name attribute for the variable should also be selected from the CF Conventions Standard Names table and added to the CF header.
4. **Units.** Each physical variable being included in the NetCDF file must have a units attribute in the CF header with the attribute content being a match to the relevant standard\_name (allowing for conversions acceptable to the UNIDATA UDUNITS software).
5. **Valid\_min, valid\_max or valid\_range.** Where possible, each variable in the file should have the attribute pair valid\_min and valid\_max present in the CF header, or the attribute valid\_range. This defines the theoretical limits of the variable. Attribute values should therefore lie outside the range of values in the data.
6. **\_FillValue or missing\_value.** Where possible, a variable should have a \_FillValue or missing\_value attribute present in the CF header. Either of these attributes describes the value that represents absent data in the file. If possible, these should be the most extreme values in the file, either the very highest or very lowest. Common values are -9999 or 2e20. Dimensional variables should not require a \_FillValue or missing\_value attribute as every dimensional space should be populated in the data file.
7. **Actual\_range.** Where possible, each variable in the file should have the attribute actual\_range present in the CF header. This defines the range of data values present in the file for that variable, excluding the value of the \_FillValue/missing\_value attribute. It should always be equal to or inside the range of values defined in valid\_min and valid\_max or

valid\_range and should also not be outside the value specified by \_FillValue or missing\_value.

8. **SeaDataNet vocabulary extensions.** In addition to having (where possible) a CF standard\_name attribute for each variable, BODC require that SeaDataNet vocabulary terms be added. These extensions are:

- a. Attribute **sdn\_parameter\_urn**

This is the SeaDataNet vocabulary code for the variable in question. The BODC Parameter Usage Vocabulary (List key='P011') vocabulary term required is the entry in the column 'Entrykey' for the relevant parameter and needs to be input as an attribute with the formatting 'SDN:P01::XXXXXXXX' where the XXXXXXXX is the Entrykey term.

The Entrykey term can be obtained in two ways from either search interface (hosted by the Marine Information Service, The Netherlands (MARIS)) linked below:

1. The [SeaDataNet hierarchical search](#) of the BODC parameter Usage Vocabulary (List key='P011'). This allows you to drill down (via the '+' buttons) through the vocabulary to find the appropriate variable. At each hierarchical level, definitions of the terms presented are also accessible via the 'i' button to enable you to ascertain which path to take. For example: following the path 'Physical oceanography> Water column temperature and salinity>Temperature of the water column>Temperature of the water body' would result in a **sdn\_parameter\_urn** equal to 'SDN:P01::TEMPPR01'.

2. The [SeaDataNet Vocabulary Web Services](#) for the BODC vocabularies. This allows you perform a free text search on any of the vocabularies maintained by BODC. The vocabularies are listed alphabetically by their list key value. To ascertain a **sdn\_parameter\_urn** you must search for variables in the BODC Parameter Usage Vocabulary (List key = 'P011').

Follow the link to open a free-text search window and enter text in the free search box.

#### **Free text search help**

Please be aware that the search is not particularly intelligent. You may need to use the wildcard option or counter-intuitive sequencing ('temperature of water' rather than 'water temperature') to find the EntryKey term for the variable you require.

The wildcard is '%' and '%emper%' would return any entries which contain the letters 'emper' within any vocabulary fields. Currently, it would return 285 entries, one of which is 'Temperature of the water body' which has the Entrykey = 'TEMPPR01' and would result in a **sdn\_parameter\_urn** equal to 'SDN:P01::TEMPPR01'. Employing a search such as '%emper%ater%' would reduce the number of returns significantly.

If it is difficult to identify which vocabulary code best suits the data, please contact BODC for advice.

- b. Attribute **sdn\_parameter\_name**

This is the SeaDataNet name (the Entryterm value) of the variable in the BODC Parameter Usage Vocabulary (List key = 'P011'). For the examples quoted in section 7a above the **sdn\_parameter\_name** would equate to 'Temperature of the water body'.

- c. Attribute **sdn\_uom\_urn**

This attribute describes the SeaDataNet code for the units being used to quantify the variable. The vocabulary term required is the entry in the column 'Entrykey' for the relevant unit from the vocabulary BODC Data Storage Units (List key='P061') and needs to be input as an attribute with the formatting 'SDN:P06::XXXX' where the XXXX is the Entrykey term.

Find the code via the [SeaDataNet Vocabulary Web Services](#) interface. Simply follow the link for List key = 'P061' and enter text in the free search window.

Wildcard options are allowed (see free text search help above) and may assist you in finding the appropriate code. For example, '%els%' may be used to find the Entrykey for Degrees Celsius and would result in a **sdn\_uom\_urn** equal to 'SDN:P06::UPAA'

d. Attribute **sdn\_uom\_name**

This is the SeaDataNet name (the Entryterm value) of the units in the BODC Data Storage Units (List key='P061') vocabulary. For the example quoted in 7c above the **sdn\_uom\_name** would equate to 'Degrees Celsius'.

EXAMPLE of SeaDataNet extensions:

Model has a variable with a standard\_name : "sea\_water\_temperature".

Model variable measures in units of degrees Celsius

How do the SeaDataNet extensions describe this?

Attributes:

sdn\_parameter\_urn: SDN:P01::TEMPPR01

sdn\_parameter\_name: Temperature of the water body

sdn\_uom\_urn: SDN:P06::UPAA

sdn\_uom\_name: Degrees Celsius