

# General Aspects

## Purpose of these BAWiki Pages

These BAWiki pages do describe all NetCDF conventions required to store baw-specific data in NetCDF data files (see [network common data form](#)). I. e. all *local* conventions are listed, which go beyond the international agreed-upon [CF-metadata convention](#). In many situations where the international agreed-upon CF conventions are insufficient, essentially the *Unstructured Grid Metadata Conventions for Scientific Datasets* (UGRID Conventions) published on [GITHUB](#) are used. Some further widely spread templates known are the [NODC NetCDF Templates](#). The NODC data center has been recently merged with other data centers and is now part of [National Centers for Environmental Information \(NCEI\)](#).

The BAW instance of a NetCDF file developed since 2010 is a file of type [CF-NETCDF.NC](#). Since version NetCDF-4.0 HDF (*Hierarchical Data File*, see [HDF5 Group](#)) is used as the underlying file format. Due to the use of HDF concepts like online compression of data stored in NetCDF files is supported as well as chunking of variables to balance read performance in case of different access to data, e.g. time-series vs. synoptic data set access.

## Important NetCDF Utilities

Important (helpful) NetCDF *Utilities* are:

- [NCDUMP](#) create (selective) text representation of the contents of a NetCDF file;
- [NCCOPY](#) (selective) copy an existing NetCDF file to another, change level of compression, change internal file structure (*File Chunking*); and
- [NCGEN](#) create NetCDF file from a CDL text file; optionally also C or FORTRAN code can be automatically generated.

A good overall view on netCDF is given in [NetCDF documentation](#).

## File Chunking

The chunk size of variables stored in a CF NetCDF file may have significant influence on read performance in case data have to be read along different dimensions, e.g. spatial versus time-series access. Chunk size can be individually tuned using the NetCDF API. As a simple alternative, already helpful in many situations, you can also make use of the [NCCOPY](#) or the [NCCHUNKIE](#) program. For further informations about chunking please read the following informations:

- [Chunking Data - Why it matters](#); und
- [Chunking Data - Choosing shapes](#).

## NetCDF vs. GRIB

Besides NetCDF GRIB is also widely used. Concerning problems of interoperability between NetCDF and GRIB a workshop was held at ECMWF in September 2014 . Further informations can be found on the website of the workshop on [Closing the GRIB/NetCDF gap](#).

## Literature

Biokaghazadeh, Saman, et al. (2015) *Enabling scientific data storage and processing on big-data systems*. Big Data (Big Data), 2015 IEEE International Conference on. IEEE, 2015 Use of data stored in netCDF files in big data analysis system Hadoop.

Lang, G. (2016) "Uniform post-processing of computational results based on UGRID CF netCDF files", 13th International UnTRIM Users Workshop 2016, Villa Madruzzo, Italy, May 30th - June 1st (doi:10.13140/RG.2.1.5059.8000).

Lang, G. (2018) "A few remarks on chunked I/O using netCDF-4/HDF5", 15th International UnTRIM Users Workshop 2018, Villa Madruzzo, Italy, May 28th - May 30th (doi:10.13140/RG.2.2.31262.23368).

Signell, R. P. und Snowden, D. P. (2014) *Advances in a Distributed Approach for Ocean Model Data Interoperability*. J. Mar. Sci. Eng. 2014, 2, 194-208. Describes also the benefits using UGRID CF metadata standard to store data in netCDF files.

## How to acknowledge Unidata

"Software and technologies developed and distributed by the Unidata Program Center are (with very few exceptions) Free and Open Source, and you can use them in your own work with no restrictions. In order to continue developing software and providing services to the Unidata community, it is important that the Unidata Program Center be able to demonstrate the value of the technologies we develop and services we provide to our sponsors ? most notably the National Science Foundation. Including an acknowledgement in your publication or web site helps us do this."

"It helps even more if we are aware of what you're doing. If you're using Unidata technologies and citing them in a paper, poster, thesis, or other venue, we'd be grateful if you would let us know about it by sending a short message to [support@unidata.ucar.edu](mailto:support@unidata.ucar.edu). Thanks!"

## Informal

- This project took advantage of netCDF software developed by UCAR/Unidata ([www.unidata.ucar.edu/software/netcdf/](http://www.unidata.ucar.edu/software/netcdf/)).

## Citation

- Unidata, (year): Package name version number [software]. Boulder, CO: UCAR/Unidata Program Center. Available from URL-to-software-page.

## DOI

- The registered Digital Object Identifier for all versions of netCDF software is <http://doi.org/10.5065/D6H70CW6>.

## Where is NetCDF used?

For an [overview](#) please visit [Where is NetCDF used?](#).

## Quality assurance using NetCDF attributes

[Quality assurance](#) of computed data is supported by programs [NCANALYSE](#), [NCDELTA](#) and [NCAGGREGATE](#) on the basis of NetCDF attributes.

### Attribute *actual\_range*

This attribute stores the actual value range for (geophysical) variables. Execution of **ncdump -h** delivers all metadata stored in a NetCDF file. This output can be searched using **grep** to retrieve **actual\_range**. In doing so a fast and simple [overview](#) is obtained, whether the actual range of a variable is outside or inside a meaningful value range.

### Automatic verification of value range

Before closure of a newly created NetCDF file, all of the above mentioned programs carry through a comparison between actual value range and allowed value range, in case

- attribute *actual\_range* (actual value range),
- attribute *cfg\_bounds\_name* (class name with definition of allowed value range), and
- a file of type [bounds\\_verify.dat](#) (description of valid value ranges for all classes of variables)

exist. \$PROGHOME/cfg/dmqs/bounds/bounds\_verify.dat contains typical valid value range data for all existing classes of variables.

The result of all comparisons done for actual value range vs. allowed value range is stored in a (printer) SDR file. These informations indicate, whether variables ly inside or outside the accepted valid value range. A fast [overview](#) is obtained by means of **grep Pruefergebnis** applied to the SDR file.

For real numbers [machine\\_epsilon](#) can be obtained e. g. from (Fortran) EPSILON:

- single precision data: approx. 1.2E-07;
- double precision data: approx. 2.2E-16.

For real data the tolerance used is given by  $2 * EPSILON * ABS(data)$ .

## Global Attributes

- [NetCDF global attributes](#)

## Locations, Profiles and Grids

- [NetCDF multiple locations](#): several (point) locations, e. g. equivalent to contents of file [location\\_grid.dat](#);

- NetCDF multiple profiles: several longitudinal and cross-sectional profiles, e. g. equivalent to contents of file profil05.bin;
- NetCDF triangular grid: triangular grid, e. g. equivalent to contents of file gitter05.dat and gitter05.bin;
- NetCDF unstructured grid: unstructured grid, e. g. equivalent to contents of file untrim\_grid.dat;
- NetCDF unstructured grid with subgrid: unstructured grid with additional subgrid data, e. g. equivalent to contents of file utrsub\_grid.dat;
- NetCDF aggregation for unstructured grids: aggregated grid and unstructured grid.

## Time Coordinate

- NetCDF time coordinate: date and time, calendar.

## Vertical Coordinate

- NetCDF vertical coordinate: dimensional vertical coordinate (height, depth).

## Horizontal Coordinate Reference System

- NetCDF grid mapping variable

## Reduction of Dataset Size

Traditionally, up to the availability of NetCDF-4 (HDF),

- NetCDF packed data, and
- NetCDF compression by gathering

were the only ways to reduce data set sizes. **Now, with the availability of NetCDF-4 (HDF), it is recommended to use online compression instead.** Online compression can be activated on a per variable basis via the NetCDF API. For existing NetCDF files NCCOPY also allows you to (online-) compress the file after it has been created.

## Data

### Synoptic Data

- NetCDF synoptic data at multiple locations,
- NetCDF synoptic data for multiple profiles,
- NetCDF cross section integral synoptic data for multiple profiles,
- NetCDF synoptic data for triangular grid,
- NetCDF synoptic (morphological) data for triangular grid,
- NetCDF synoptic data for unstructured grid,
- NetCDF synoptic data for unstructured grid with subgrid, and
- NetCDF DelWAQ data.

## Time Series Data

### Analysis Data

- [NetCDF tidal characteristic numbers of water level](#), and
  - [NetCDF differences for tidal characteristic numbers of water level](#).
- 

back to [Standard-Software-Applications \(Add-ons\)](#)

---

[Overview](#)