Guidelines on making your data interoperable and reusable

The UK Polar Data Centre (PDC) strives to publish data in accordance with international standards and conventions where and when possible. Disciplines vary in their community guidelines, with some establishing written frameworks for how to publish data. We highlight some below for you to explore and have pulled out common use-cases for data published with the PDC.

Please refer to the respective websites before submitting any data to make sure it is compliant where appropriate. The PDC are also happy to provide assistance in ensuring data is compliant. Please email us at PDCServiceDesk@bas.ac.uk with any questions, concerns or assistance. All hyperlinks included in this document will open in a new tab, and the full URLs are provided in the 'References' section at the bottom of the document.

Variables

CF conventions [1]: Established initially for climate forecasting, CF conventions enable standardised processing of NetCDF files. To comply with NetCDF conventions, variables that are included in the CF vocabulary should match the standard names and units:

Variable	Standard_name	unit
Latitude	latitude	degree_north
Latitude (rotated North Pole)	grid_latitude	degree
Longitude	longitude	degree_east
Longitude (rotated North Pole)	grid_longitude	degree
Air pressure	air_pressure	Pa
Altitude	altitude	m
Depth	depth	m

Please note that the CF conventions standard name table [2] has variables in more detail and may include variables more suitable to the dataset. Please refer to the website before submitting the dataset to the PDC.

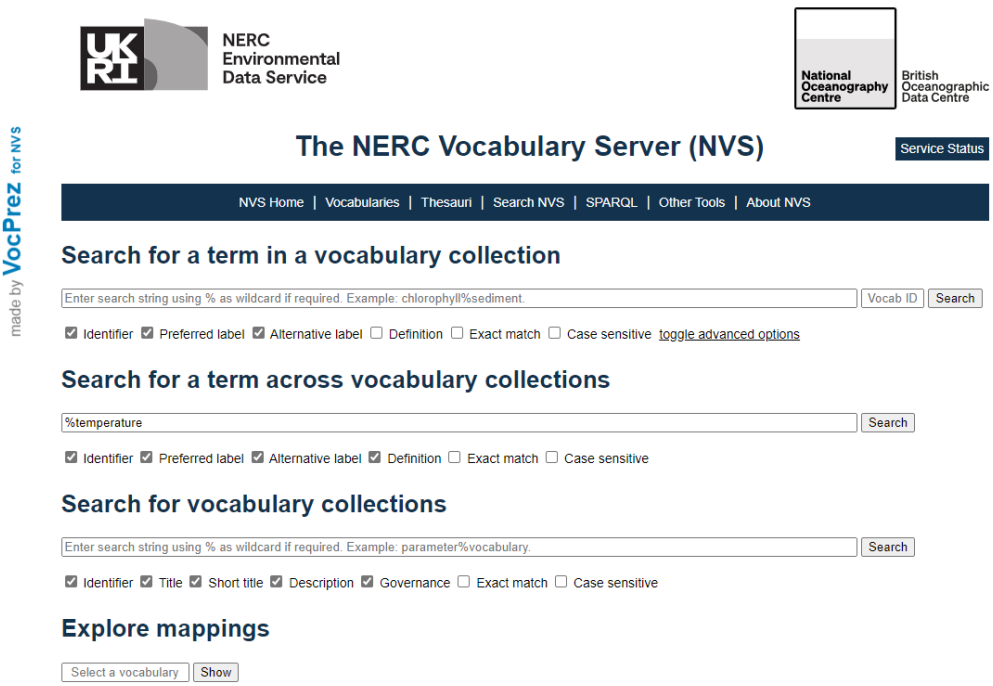
Please also note that for spatial data, we aim to publish datasets using decimal degrees with units as 'degree_north' and 'degree_east' unless in a projection other than ESPG:4326 [3]. Please contact the PDC if you have any questions or concerns about this.

Should you want to include another name in addition to the standard_name, you can populate long_name with another name for this variable. If there is no suitable standard_name available, place a name (of your choosing) for the variable in the long_name attribute.

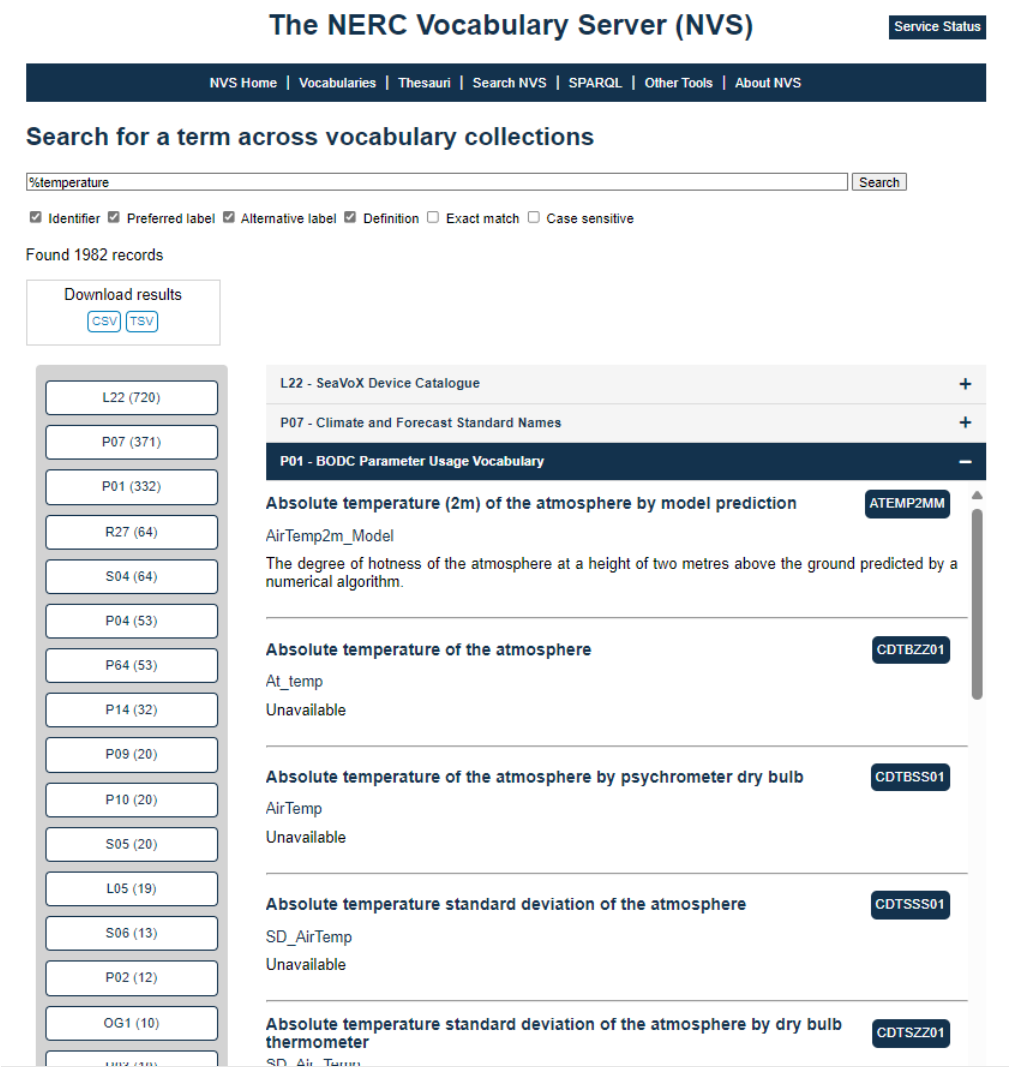
Tabular data should have variables in the first row and for each column and use standard names, where possible. If there is no suitable standard name, use short, unique, meaningful names without spaces and special characters. Units should be included, where applicable (see more details on units below).

The NERC Vocabulary Server [4] provides standardised and hierarchically-organised vocabularies and is managed by the British Oceanographic Data Centre (BODC). Controlled vocabularies provide well-defined terms to standardise data and metadata. Commonly used vocabularies are L05 (SeaDataNet Device Categories) [5], L22 (SeaVox Device Catalogue) [6], B76 (BODC Platform Models) [7], P01 (BODC Parameter Usage Vocabulary) [8], P06 (BODC-approved Data Storage Units) [9] and P07 (Climate and Forecast Standard Names) [10].

To search for a term, head to the NVS [4] website and click 'Search NVS'. If you're familiar with which vocabulary collection you'd like to use, search within a vocabulary collection directly. If not, you can search across the vocabulary collections. Using % with your search string will make it a 'wildcard' when searching within a collection, and potentially expand your search results.



When you click ’Search’, you will be redirected to the results page, which will provide you with various vocabularies to choose from. Please choose the appropriate vocabulary for your scientific activity. Within a vocabulary, choose the parameter that best represents what you’ve measured to describe your data:



Please explore the NVS website fully to utilise its potential for datasets you’re working with.

Units

Attaching units as often as possible will make your data more understandable and reusable. The Unified Code for Units of Measurement (UCUM) [11] should be utilised when possible.

The table below shows some examples of units you may use frequently:

Unit	Quantity type	Printed_as
metre	length	m
metres per second (m/s)	velocity	m.s-1
centimetre	length	cm
kilogram	mass	kg
gram	mass	g
kelvin	temperature	K
parts per million	parts per million	[ppm]
hertz	frequency	Hz
watt	power	W
volt	electric	V
degree	plane angle	deg
minute	plane angle	'
second	plane angle	''
liter	volume	l
degree Celsius	temperature	Cel
minute	time	min
hour	time	h
day	time	d
knot	velocity	[kn_i]
pH	acidity	[pH]
bar	pressure	bar

This following table includes SI prefixes you may be using:

Prefix	Printed as
micro	u
nano	n
mega	M
giga	G

On the UCUM website, when extracting relevant unit information, the following guidance and example can be used:

name	kind of quantity	print	c/s	c/i
meter	length	m	m	M
second	time	s	s	S
gram	mass	g	g	G
radian	plane angle	rad	rad	RAD
kelvin	temperature	K	K	K
coulomb	electric charge	C	C	C
candela	luminous intensity	cd	cd	CD

'name' is the regular textual representation of the unit (for example 'gram')

'kind of quantity' provides information about the unit ('mass for gram)

'c/s' is how the unit should be printed in your file (for grams, 'g')

Values

Data values should be quality checked by data creators and have no obvious errors. Missing values should be distinguished from true zero values, i.e. designated as 'null', 'NaN' (not-a-number), 'NA' (not-applicable), '-99999'.

Values can also make use of standardised terminologies. For example, use a community-recognised taxonomic index for organism names (such as WoRMS [12] for marine species, or Plants of the World Online [13] for plants or Catalogue of Life [14] for all species).

Any units should be included in the header, rather than with each individual value. Please see the units section and variables section above for more details.

Date and time

Dates and times can be represented using the ISO standard 8601 [15]; which recommends using YYYY-MM-DDTHH:MM:SS.F with the header name either conforming to a controlled vocabulary (for example, CF conventions recommend using the header 'time' for these values), or using a clear header name such as 'DateTime' or 'datetime'

XCSV format

If you are submitting tabular data, we recommend using CSV format with the header and unit description guidance we've provided above. PDC data managers are able to generate XCSV versions of your CSV files once they are submitted, which include metadata associated with the deposit in the header. This has the benefit of increasing machine readability. Please discuss with a PDC data manager at the time of data deposit what your individual XCSV needs might be. Below is an example of a published XCSV formatted file.

```
# id: GB/NERC/BAS/PDC/01750
# title: Kinematically corrected GNSS/INS buoy data during the MOSAiC expedition 2019-2020
# summary: This dataset contains the post-processed GNSS/INS buoy data for a kinematic correction of a moving base station. The GNSS/INS buoys were deployed on sea ice during the 2019-20 MOSAiC expedition. These buoys recorded raw
GNSS/INS data at a sampling rate of 10 Hz. For the kinematic correction, two buoys (with overlapping measurements of each other) were selected, and one of the buoys was used as a moving "base" and the other as the "rover". The post-
processed dataset contains kinematically corrected latitude, longitude and velocity of the rover, as well as the baseline distance between the rover and base. The main objective of the kinematic correction is to create high-precision
and high-frequency data to measure ice dynamics at a few centimetre accuracies. The buoys were assembled by the University of Huddersfield team and the deployment was done by the MOSAiC ice team throughout the expedition.
# acknowledgement: This work was funded by NERC MOSAiC program NE/S002545/1.
# keywords: Arctic,GNSS,INS,MOSAiC,crack,kinematic positioning,lead,ridging,sea ice
# citation: Hwang, B., Hagan, B., & MOSAiC ice team (2023). Kinematically corrected GNSS/INS buoy data during the MOSAiC expedition 2019-2020 (Version 1.0) [Data set]. NERC EDS UK Polar Data Centre.
https://doi.org/10.5285/e718a95d-76f0-46bf-ad93-ef02b0ba1f9f
# license: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
# history: GNSS/INS buoy system was assembled by the University of Huddersfield. Each buoy system consists of a miniature inertial system (Spatial), a data storage (ILU), a GNSS antenna and a power supply.  These sensor components were
supplied from Advanced Navigation (advancednavigation.com) and Innovelec (innovelec.co.uk) (see Instrumentation for detailed information on the sensors), and were then assembled in a weather-proof plastic case for deployment. A total
of six sets of the GNSS/INS buoys were shipped to the expedition, but the number of the buoys deployed on the ice varied throughout the expedition due to the logistical and environmental circumstances. The deployment of the buoys was
made in the vicinity (within 2 km) of RV Polarstern at the MOSAiC drifting site. Kinematic correction requires the base station measurement should overlap the period of the rover measurement, so the availability of the kinematically
corrected data is affected by environmental and operational limitations (see Temporal coverage for the data availability). This data set provides high-precision (centimetre accuracy) and high-frequency (10 Hz) ice dynamic data to study
ice drift, lead/ crack opening/ closing and ridging of sea ice with the precision of a few centimetre scales.
# comment: The kinematic correction of the raw GNSS/INS data was conducted using Kinematica from Advanced Navigation (advancednavigation.com). The raw GNSS/INS data (in ANPP format) was converted to RINEX format for a kinematic
correction with a moving base station. The processed data contain Moving Base Fix Type (i.e., the type of GNSS fix between the rover and the base station). For the data quality, the data with Fix Type* = 7 (i.e., RTK fixed - Carrier
phase ambiguities have been resolved) were selected in this dataset.
# : Note that Leg5 data are not included in this dataset.  No changes were observed in Leg5 and some errors were found during the data processing.
# : *Fix Type: Fix type 0 - No Fix; 1 - 2D; 2 - 3D fix (no base station data); 3 - SBAS; 4 - Differential (Base station data does not include carrier phase corrections); 5 - Omnistar; 6 - RTK Float (Carrier phase ambiguities cannot be
resolved); 7 - RTK Fixed (Carrier phase ambiguities have been resolved).
# : Each GNSS sensor unit consists of Spatial, Data Interface Logging Unit (ILU), and antenna.
# : Spatial is a rugged miniature GPS aided inertial navigation system and AHRS that provides accurate position, velocity, acceleration and orientation. It combines temperature calibrated accelerometers, gyroscopes, magnetometers and a
pressure sensor with advanced GNSS receiver to deliver accurate and reliable navigation and orientation (see advancednavigation.com/solutions/spatial/). The position accuracy can be a few centimetres as long as the antenna has a good
sky view without any obstructions.
# : Data Interface Logging Unit (ILU) is a device server that interfaces to Spatial and provides data logging, a web interface, a time synchronisation server and different data input/output option.
# : Two types of GNSS antennas were used: Antcom (G5ANT-53A4T1) and Tallysman (TW3972).
# institution: University of Huddersfield
# authors: Byongjun Hwang,Bernard Hagan,MOSAiC ice team
# project: MOSAiC: Floe-scale observation and quantification of Arctic sea ice breakup and floe size during the autumn-to-summer transition (MOSAiCFSD) https://gtr.ukri.org/projects?ref=NE/S002545/1
# : MOSAiC https://mosaic-expedition.org/
# references: https://online.ucpress.edu/elementa/collection/269/Special-Feature-The-Multidisciplinary-Drifting
# publisher_name: NERC EDS UK Polar Data Centre
# publisher_type: institution
# publisher_email: pdc@bas.ac.uk
# publisher_url: https://www.bas.ac.uk/data/uk-pdc/
# time_coverage_start: 2019-11-30
# time_coverage_end: 2020-07-28
# geospatial_lon_min: -30 (degree_east)
# geospatial_lon_max: 120 (degree_east)
# geospatial_lat_min: 75 (degree_north)
# geospatial_lat_max: 90 (degree_north)
# geospatial_bounds_crs: EPSG:4326
# [a]: Estimated error in position in North-East-Down frame; estimated error values are given to 1 sigma of standard deviation.
# [b]: Estimated error in IMU velocity in North-East-Down frame; estimated error values are given to 1 sigma of standard deviation.
# [c]: Estimated error in relative distance in North-East-Down frame; estimated error values are given to 1 sigma of standard deviation.
TimeStamp (datetime),Unix Time (s since 1970-01-01 UTC),Microseconds (us),Latitude (degree_north),Longitude (degree_east),Height (m),Latitude Standard Deviation (degree_north) [a],Longitude Standard Deviation (degree_east) [a],Height
Standard Deviation (m) [a],Velocity North (m/s),Velocity East (m/s),Velocity Down (m/s),Velocity North Standard Deviation (m/s) [b],Velocity East Standard Deviation (m/s) [b],Velocity Down Standard Deviation (m/s) [b],Base to Rover
North (m),Base to Rover East (m),Base to Rover Down (m),Base to Rover North Standard Deviation (m) [c],Base to Rover East Standard Deviation (m) [c],Base to Rover Down Standard Deviation (m) [c],Velocity (m/s),Baseline (m)
2020-02-18 06:58:34.290,1582009114,290000,88.093008947684098,77.599637094348793,8.78172,0.19047,0.15833,0.24768,0.09449,-0.07789,0.07976,0.45254,0.48983,1.49565,404.79927,-28.60191,0.00888,0.00266,0.00323,0.01112,0.12245,405.80848
2020-02-18 06:58:34.390,1582009114,390000,88.093008653323267,77.599631423268420,8.84539,0.19047,0.15833,0.24768,0.09476,-0.12794,-0.03369,0.45254,0.48982,1.49564,404.80011,-28.60142,0.00946,0.00271,0.00323,0.01112,0.15921,405.80928
2020-02-18 06:58:34.490,1582009114,489999,88.093008516724083,77.599633427626060,8.86462,0.19047,0.15833,0.24768,0.09001,-0.10345,0.05998,0.45254,0.48982,1.49563,404.79997,-28.60266,0.00815,0.00268,0.00320,0.01112,0.13713,405.80923
2020-02-18 06:58:34.590,1582009114,589999,88.093008362776089,77.599632389601311,8.85586,0.19047,0.15833,0.24768,0.09101,-0.08461,-0.02522,0.45254,0.48982,1.49563,404.79972,-28.60171,0.01017,0.00272,0.00325,0.01111,0.12427,405.80891
```

Ln 1, Col 28 | 369,447,835 characters | 80% | Unix (LF) | UTF-8

References:

[1] https://cfconventions.org/

[2] https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html

[3] https://spatialreference.org/ref/epsg/4326/

[4] https://vocab.nerc.ac.uk/

[5] https://vocab.nerc.ac.uk/collection/L05/current/accepted/

[6] https://vocab.nerc.ac.uk/collection/L22/current/accepted/

[7] https://vocab.nerc.ac.uk/collection/B76/current/accepted/

[8] http://vocab.nerc.ac.uk/collection/P01/current/accepted/

[9] http://vocab.nerc.ac.uk/collection/P06/current/

[10] http://vocab.nerc.ac.uk/collection/P07/current/

[11] https://ucum.org/ucum

[12] https://www.marinespecies.org/

[13] https://powo.science.kew.org/

[14] https://www.catalogueoflife.org/

[15] https://www.iso.org/iso-8601-date-and-time-format.html