OceanSITES Data Format Reference Manual

NetCDF Conventions and Reference Tables

Version 1.3

May 19, 2014

Contents

History

Version	Date	Comment
0.1	20/03/2003	TC: creation of the document
0.3	20/02/2004	TC: updates on locations, mooring name, data state indicator, parameters table, epic codes, history information
0.3.2	26/05/2004	NG: make more flexible, add dataset (metadata) file
0.4	01/06/2004	TC: separate data set description and data file, merge with Steve Hankins's straw man
0.6	28/06/2004	TC: updates from Nan Galbraith, Steve Hankins, Jonathan Gregory, Brian Eaton
0.7	23/05/2005	Maureen Edwards: NOCS data centre, new GF3 parameters
0.7	24/05/2005	Roy Lowry: physical parameters from BODC Data Markup Vocabulary
1.0	18/02/2006	TC: updates following OceanSITES data management meeting 2006, Hawai'i §2.1: LEVEL dimension replaces DEPTH to accomadate depth or pressure §2.2: QC_MANUAL field created §2.2: CONVENTION field removed §2.2: PLATFORM_CODE added §2.2: SITE_CODE added §2.2: WMO_PLATFORM_CODE added §2.3: DEPTH renamed DEPH to comply to GF3 §2.3: DATA_MODE set at measurement level §3: metadata file description transferred to "OceanSITES metadata proposal" until approval §5: file naming convention updated
1.0	19/02/2006	NG: data codes in chapter 4.1.2
1.0	28/04/2006	PF & NG: data mode optional
1.0	28/04/2006	TC & JG: §2.2 global attributes
1.1	April-May-June 2008	NG, MM, TC, ML: general revision based on OceanSITES 2008 meeting Epic codes removed Use ISO8601 for string dates Remove general attributes Update global attribute section for CF-1.1 compatibility New dimensions for DEPTH, LATITUDE, LONGITUDE Add an uncertainty attribute New presentation of the document
1.2 draft	September 2009	§1.3 : GDAC distribute the "best data" statement §1.4 : add a "User obligations" paragraph §1.5 : add a "Disclaimer" paragraph §2 : note on format version §2.2.1 : no fill value allowed for TIME, LATITUDE, LONGITUDE, DEPTH §2.2.1 : use WGS84 datum for latitude and longitude §2.2.1 : DEPTH "reference" optional attribute §2.2.3 and §4.7 : use "sensor_mount" optional attribute §2.2.3 and §4.8 : use "sensor_orientation" optional attribute §2.2.3 : use sensor_name and sensor_orientation attributes §4.3 : revisit parameter names §4.4 : update DAC codes §4.6 : add a sentence on OceanSITES site naming policy §5.1 : new data file naming convention §5.2: add GDAC_CREATION_DATE, GDAC_UPDATE_DATE, PARAMETERS in the index file.
1.2 draft	December 7th 2009	§5.1: revisit file naming convention. §5.2: add a data_mode in the index file. §6: add a "Glossary, definition" chapter.
1.2 draft	March 2010	§5.2: add geospatial_vertical_min and geospatial_vertical_min in the index file. §1.7 : useful links chapter created

1.2 draft	April 2010	Last comments received from Matthias Lankhorst, Nan Galbraith, Derrick Snowden, Hester Viola, Andrew Dickson, John Graybeal.§1.6: information and contact on project office §2.2.1: update of Z axis §2.2.1: latitude-longitude reference and EPSG coordinate reference §2.2.1: depth EPSG coordinate reference §2.2.1: note on latitude and longitude WGS84 datum §2.2.1: note on DEPTH reference §2.2.3: all attributes listed in the example §2.2.4: metadata variables: sensors information, calibrations §3: simplify metadata introduction §4.2: QC flag scale, 6 not used (comment) §4.3.1: use DOXY_TEMP instead of TEMP_DOXY §4.4: 4 new centres §4.6: update of OceanSITES catalogue
1.2 draft	June 2010	Updates from 29/06/2010 webex meeting. §2.1: remove “For a mooring with a GPS receiver, use LATITUDE of the same dimension as TIME and provide the actual location.” §2.2: add an optional “array” and “network” optional global attribute Allow multiple axes in a file • §2: remove “Coordinate variables, which describe the dimensions of a data set, are limited to a single set of longitude, latitude, depth and time (X,Y,Z, and T) dimensions in any single file. If data from a reference station cannot all be put on to a single set of axes, then separate files are created for these data.” • §2.3.1: remove “Data with different coordinate variables must be recorded in separate files.“ §2.3.1: empty values are not allowed for coordinate variables.
1.2.3 draft	2013March 2012	NG: allow multiple z coordinates;clarify dimensions and coordinates; clarify requirement for QC flag meanings; change standard name for ATMS (it is NOT surface pressure)
1.3 draft	April 10 2013	NG: Align with ACDD, streamline globals, simplify text
1.3 draft	May 29, 2013	MM, JZ, MP, TC, ML during meeting in Seoul: Highlight OceanSITES specific attributes and affirm compliance with CF, ACDD, udunits, SeaDataNet. Naming and directory conventions for new kinds of files - gridded, products.
1.3 draft	June 1st	TC : cleanup the presentation of the document
1.3 draft	June-Sept2013	NG: address comments and changes to ACDD; remove unused globals. Remove qc_manual. Update crs:EPSG for depth and heioght.
1.3.1	Oct-Dec 2013	NG: Short names now just recommended; multiple variables are now allowed for all dimensions. Geospatial extent global attributes now recommended to be strings. Added source of external attribute terms.
1.3.2	Jan-May 2014	NG,MP,MH: resolve comments, refine required fields, redefine data mode P.
1.3.3	June 2014	NG rewrite sensor attribute options; lower case .QC terms, refine filenames

1 Overview

1.1 About OceanSITES

The OceanSITES program is the global network of open-ocean sustained time series sites, called ocean reference stations, being implemented by an international partnership of researchers. OceanSITES provides fixed-point time series of various physical, biogeochemical, and atmospheric variables at different locations around the globe, from the atmosphere and sea surface to the seafloor. The program's objective is to build and maintain a multidisciplinary global network for a broad range of research and operational applications including climate, carbon, and ecosystem variability and forecasting and ocean state validation.

All OceanSITES data are publicly available. More information about the project is available at: http://www.oceansites.org.

1.2 About this document

The main purpose of this document is to specify the format of the files that are used to distribute OceanSITES data, and to document the standards used therein. This includes naming conventions, or taxonomy, as well as metadata content. Intended users are OceanSITES data providers and users of OceanSITES data.

1.3 User Obligations

An OceanSITES data provider is expected to read and understand this manual and the NetCDF specification it describes. OceanSITES participants are required to submit data in a timely fashion, with the understanding that these are the "best available" versions, and may be updated if improved versions become available. Data files should be in compliance with a published OceanSITES format specification.

A user of OceanSITES data must comply with the requirements set forth in the attributes "license" and "citation" of the NetCDF data files.

Unless stated otherwise, a user must acknowledge use of OceanSITES data in all publications and products where such data are used, preferably with the following standard citation:

"These data were collected and made freely available by the international OceanSITES project and the national programs that contribute to it."

1.4 Disclaimer

OceanSITES data are published without any warranty, expressed or implied. The user assumes all risk arising from his/her use of OceanSITES data.

OceanSITES data are intended to be research-quality and include estimates of data quality and accuracy, but it is possible that these estimates or the data themselves contain errors. It is the sole responsibility of the user to assess if the data are appropriate for his/her use, and to interpret the data, data quality, and data accuracy accordingly.

1.5 Feedback

OceanSITES welcomes users to ask questions and report problems to the contact addresses listed in the data files or on the OceanSITES web page.

2 OceanSITES NetCDF data format version 1.3

OceanSITES uses NetCDF (Network Common Data Form), a set of software libraries and machine-independent data formats. Our implementation of NetCDF is based on the community-supported Climate and Forecast Metadata Convention (CF), which provides a definitive description of the data in each variable, and the spatial and temporal properties of the data. Any version of CF may be used, but it must be identified in the 'Conventions' attribute.

The purpose of the format specification is to make OceanSITES data easy to discover and to use. Any relevant metadata should be included whether it is part of the standard or not. For example, water depth, instrumentation descriptions, and detailed provenance are all useful fields and should be included if available.

OceanSITES adds some requirements to the CF standard, to make it easier to share in-situ data, to make it simpler for the GDACs to aggregate data from multiple sites, and to ensure that the data can be created and understood by basic NetCDF utilities.

- Where time is specified as a string, the ISO8601 standard **"YYYY-MM-DDThh:mm:ssZ"** is used; this applies to attributes and to the base date in the 'units' attribute for time. There is no default time zone; UTC must be used, and specified.
- Global attributes from Unidata's NetCDF Attribute Convention for Data Discovery (ACDD) are implemented.
- Variable names (short names) from a controlled vocabulary are recommended

The components of NetCDF files are described in the following sections.

2.1 Global attributes

The global attribute section of a NetCDF file describes the contents of the file overall, and allows for data discovery. All fields should be human-readable and use units that are easy to understand (e.g. time_coverage_duration should be in days, for a file that spans more than a month). OceanSITES recommends that all of these attributes be used and contain meaningful information, unless there are technical reasons rendering this impossible. Attributes used by our data inventory system are required, and are listed in **bold type**.

Global attribute names are case sensitive.

Attributes are organized by function: Discovery and identification, Geo-spatial-temporal, Conventions used, Publication information, and Provenance. Attributes that are part of the Attribute Convention for Data Discovery (ACDD) or Climate and Forecast (CF) standard, or that appear in the NetCDF Users Guide (NUG) are so indicated, as are those that are used by GDAC inventory software.

Discovery and identification		
name	**example**	**note**
site_code	site_code="CIS" (OceanSITES specific)	Name of the site within OceanSITES project. The site codes are available on GDAC ftp servers. **Required (**GDAC)
platform_code	platform_code="CIS-1" (OceanSITES specific)	The unique platform code, assigned by an OceanSITES project. **Required**. (GDAC)
data_mode	data_mode="R" (OceanSITES specific)	Indicates if the file contains real-time, provisional or delayed-mode data. The list of valid data modes is in reference table 4. (GDAC)
title	title="Real time CIS Mooring Temperatures"	Free-format text describing the dataset, for use by human readers. Use the file name if in doubt. (NUG)
summary	summary="Oceanographic mooring data from CIS observatory in the Central Irminger Sea, 2005. Measured properties: temperature and salinity at ten depth levels."	Longer free-format text describing the dataset. This attribute should allow data discovery for a human reader. A paragraph of up to 100 words is appropriate. (ACDD)
naming_ authority	naming_authority="OceanSITES"	The organization that manages data set names. (ACDD)
id	id="OS_CIS-1_200502_TS"	The "id" and "naming_authority" attributes are intended to provide a globally unique identification for each dataset. The id may be the file name without .nc suffix, which is designed to be unique. (ACDD)
wmo_platform_code	wmo_platform_code="48409" (OceanSITES specific)	WMO (World Meteorological Organization) identifier. This platform number is unique within the OceanSITES project.
source	source="subsurface mooring"	Use a term from the SeaVoX Platform Categories,(L06) list, usually one of the following: "moored surface buoy", "subsurface mooring" (CF)
principal_ investigator	principal_investigator="Alice Juarez"	Name of the person responsible for the project that produced the data contained in the file.
principal_ investigator_email	principal_ investigator_email ="AJuarez AT whoi.edu"	Email address of the project lead for the project that produced the data contained in the file.
principal_ investigator_url	principal_ investigator_url=" whoi.edu/profile/AJuarez"	URL with information about the project lead.
institution	institution="National Oceanographic Centre"	Specifies institution where the original data was produced. (CF)
project	project="CIS"	The scientific project that produced the data.
array	array="TAO" (OceanSITES specific)	A grouping of sites based on a common and identified scientific question, or on a common geographic location.
network	network="EuroSITES" (OceanSITES specific)	A grouping of sites based on common shore-based logistics or infrastructure.
keywords_ vocabulary	keywords_vocabulary =" GCMD Science Keywords"	Please use one of 'GCMD Science Keywords', 'SeaDataNet Parameter Discovery Vocabulary' or 'AGU Index Terms'. (ACDD)
keywords	keywords="EARTH SCIENCE >Oceans >Ocean Temperature"	Provide comma-separated list of terms that will aid in discovery of the dataset. (ACDD)

comment	comment=”Provisional data”	Miscellaneous information about the data or methods used to produce it. Any free-format text is appropriate. (CF)
Geo-spatial-temporal		
name	**example**	**note**
area	area=”North Atlantic Ocean” (OceanSITES specific)	Geographical coverage. Try to compose of the following: North/Tropical/South Atlantic/Pacific/Indian Ocean, Southern Ocean, Arctic Ocean.
geospatial_lat_min	geospatial_lat_min=59.8	The southernmost latitude, a value between -90 and 90 degrees; may be string or numeric. (ACDD, GDAC)
geospatial_lat_max	geospatial_lat_max=59.8	The northernmost latitude, a value between -90 and 90 degrees. (ACDD, GDAC)
geospatial_lat_units	geospatial_lat_units= "degree_north"	Must conform to udunits. If not specified then ”degree_north” is assumed. (ACDD)
geospatial_lon_min	geospatial_lon_min=-41.2	The westernmost longitude, a value between -180 and 180 degrees. (ACDD, GDAC)
geospatial_lon_max	geospatial_lon_max=-41.2	The easternmost longitude, a value between -180 and 180 degrees. (ACDD, GDAC)
geospatial_lon_units	geospatial_lon_units=”degree_east”	Must conform to udunits, If not specified then ”degree_east” is assumed. (ACDD)
geospatial_vertical_min	geospatial_vertical_min=10.0	Minimum depth or height of measurements. (ACDD, GDAC)
geospatial_vertical_max	geospatial_vertical_max=2000	Maximum depth or height of measurements. (ACDD, GDAC)
geospatial_vertical_positive	geospatial_vertical_positive=”down”	Indicates which direction is positive; "up" means that z represents height, while a value of "down" means that z represents pressure or depth. If not specified then “down” is assumed. (ACDD)
geospatial_vertical_units	geospatial_vertical_units=’meter”	Units of depth, pressure, or height. If not specified then “meter” is assumed. (ACDD)
time_coverage_start	time_coverage_start=”2006-03-01T00:00:00Z”	Start date of the data in UTC. See note on time format below. (ACDD, GDAC)
time_coverage_end	time_coverage_end=”2006-03-05T23:59:29Z”	Final date of the data in UTC. See note on time format below. (ACDD, GDAC)
time_coverage_duration	time_coverage_duration=”P415D” time_coverage_duration=”P1Y1M3D”	Use ISO 8601 (examples: P1Y ,P3M, P10D) (ACDD)
time coverage_resolution	time coverage_resolution=”PT30M”	Interval between records: Use ISO 8601 (PnYnMnDTnHnMnS) e.g. PT5M for 5 minutes, PT1H for hourly, PT30S for 30 seconds. (ACDD)
cdm_data_type	cdm_data_type=”Station”	The Unidata CDM (common data model) data type used by THREDDS. e.g. point, profile, section, station, station_profile, trajectory, grid, radial, swath, image; use Station for OceanSITES mooring data. (ACDD)
featureType	featureType=”timeSeries” or ”timeSeriesProfile”	**Optional, and only** for files using the Discrete Sampling Geometry, available in CF-1.5 and later. See CF documents. (CF)

data_type	data_type=”OceanSITES time-series data”	From Reference table 1: OceanSITES specific. (GDAC)
Conventions used		
name	**example**	**note**
format_version	format_version=”1.3” (OceanSITES specific)	OceanSITES format version; may be 1.1, 1.2, 1.3. **(**GDAC)
Conventions	Conventions=”CF-1.6, OceanSITES-1.3, ACDD-1.2”	Name of the conventions followed by the dataset. (NUG)
netcdf_version	netcdf_version=”3.5” (OceanSITES specific)	NetCDF version used for the data set
Publication information		
name	**example**	**note**
publisher_name	publisher_name =”Alice Juarez”	Name of the person responsible for metadata and formatting of the data file. (ACDD)
publisher_email	publisher_email=”ajuarez AT1 ifremer.fr”	Email address of person responsible for metadata and formatting of the data file. (ACDD)
publisher_url references	publisher_url=”http://ifremer.fr” references=”http:// www.oceansites.org, http://www.noc.soton.ac.uk/animate/index.php”	Web address of the institution or of the data publisher. (ACDD) Published or web-based references that describe the data or methods used to produce it. Include a reference to OceanSITES and a project-specific reference if appropriate.
data_assembly_center	data_assembly_center=”EUROSITES” (OceanSITES specific)	Data Assembly Center (DAC) in charge of this data file. The data_assembly_center are listed in reference table 5.
update_interval	update_interval=”PT12H” (OceanSITES specific)	Update interval for the file, in ISO 8601 Interval format: PnYnMnDTnHnM where elements that are 0 may be omitted. Use “void” for data that are not updated on a schedule. Used by inventory software. (GDAC)
license	license =”Follows CLIVAR (Climate Varibility and Predictability) standards, cf. http://www.clivar.org/data/data_policy.php. Data available free of charge. User assumes all risk for use of data. User must display citation in any publication or product using data. User must contact PI prior to any commercial use of data.”	A statement describing the data distribution policy; it may be a project- or DAC-specific statement, but must allow free use of data. OceanSITES has adopted the CLIVAR data policy, which explicitly calls for free and unrestricted data exchange. Details at: http://www.clivar.org/data/data_policy.php (ACDD)
citation	citation=”These data were collected and made freely available by the OceanSITES project and the national programs that contribute to it.” (OceanSITES specific)	The citation to be used in publications using the dataset; should include a reference to OceanSITES but may contain any other text deemed appropriate by the PI and DAC..
acknowledgement	acknowledgement=”Principal funding for the NTAS experiment is provided by the NOAA Climate Observation Division.”	A place to acknowledge various types of support for the project that produced this data. (ACDD)

Provenance		
name	**example**	**note**
date_created	date_created =”2006-04-11T08:35:00Z”	The date on which the data file was created. **Version date and time for the data contained in the file. (UTC). See note on time format below.** (ACDD)
date_modified	date_modified=”2012-03-01T15:00:00Z”	The date on which this file was last modified. (ACDD)
history	history= “2005-04-11T08:35:00Z data collected, A. Meyer. 2005-04-12T10:11:00Z OceanSITES file with provisional data compiled and sent to DAC, A. Meyer.”	Provides an audit trail for modifications to the original data. It should contain a separate line for each modification, with each line beginning with a timestamp, and including user name, modification name, and modification arguments. The time stamp should follow the format outlined in the note on time formats below. (NUG)
processing_ level	processing_level =” Data verified against model or other contextual information” (OceanSITES specific)	Level of processing and quality control applied to data. Preferred values are listed in reference table 3.
QC_indicator	QC_indicator =”excellent” (OceanSITES specific)	A value valid for the whole dataset, one of: ‘unknown’ – no QC done, no known problems ‘excellent’ - no known problems, some QC done ‘probably good’ - validation phase ‘mixed’ - some problems, see variable attributes
contributor_name	contributor_name = “Jane Doe”	A semi-colon-separated list of the names of any individuals or institutions that contributed to the creation of this data. (ACDD)
contributor_role	contributor_role = “Editor”	The roles of any individuals or institutions that contributed to the creation of this data, separated by semi-colons.(ACDD)
contributor_email	contributor_email = “jdoe AT ifremer.fr”	The email addresses of any individuals or institutions that contributed to the creation of this data, separated by semi-colons. (ACDD)

Notes on Global Attributes

- The file dates, date_created and date_modified, are our interpretation of the ACDD file dates. Date_created is the time stamp on the file, date_modified may be used to represent the ‘version date’ of the geophysical data in the file. The date_created may change when e.g. metadata is added or the file format is updated, and the optional date_modified MAY be earlier.
- Geospatial extents (geospatial_lat_min, max, and lon_min, max) are preferred to be stored as strings for use in the GDAC software, however numeric fields are acceptible.
- cdm_data_type is acceptable in any file; the use of a featureType attribute indicates that this is a Discrete Sampling Geometry file that adheres to rules for such files, including some contraints on acceptable coordinate variables; see CF Documentation.

2.2 Dimensions

NetCDF dimensions provide information on the size of the data variables, and additionally tie coordinate variables to data. CF recommends that if any or all of the dimensions of a variable have the interpretations of "date or time" (T), "height or depth" (Z), "latitude" (Y), or "longitude" (X) then those dimensions should appear in the relative order T, Z, Y, X in the variable's definition (in the CDL).

Name	Example	Comment
TIME	TIME=unlimited	Number of time steps. Example: for a mooring with one value per day and a mission length of one year, TIME contains 365 time steps.
DEPTH	DEPTH=5	Number of depth levels. Example: for a mooring with measurements at nominal depths of 0.25, 10, 50, 100 and 200 meters, DEPTH=5.
LATITUDE	LATITUDE=1	Dimension of the LATITUDE coordinate variable.
LONGITUDE	LONGITUDE=1	Dimension of the LONGITUDE coordinate variable.

Notes on Dimensions

- CF v 1.5 introduced Discrete Sampling Geometries; these are permitted in OceanSITES but are not described in this manual; they may require different sets of dimensions from those documented here. Please see Chapter 9. Discrete Sampling Geometriesof the CF Conventions document, http://cf-pcmdi.llnl.gov/documents/cf-conventions/latest-cf-conventions-document-1 for details.

2.3 Coordinate variables

NetCDF coordinates are a special subset of variables. Coordinate variables orient the data in time and space; they may be dimension variables or auxiliary coordinate variables (identified by the 'coordinates' attribute on a data variable). Coordinate variables have an "axis" attribute defining that they represent the X, Y, Z, or T axis.

As with data variables, OceanSITES recommends variable names and requires specific attributes for coordinate variables: units, axis, and, where available, standard_name are required. Missing values are not allowed in coordinate variables.

All attributes in this section are highly recommended. The attribute "QC_indicator" may be omitted for any parameter if there is a separate QC variable for that parameter.

Type, name, dimension, attributes	Comment
Double **TIME**(TIME); TIME:standard_name = "time"; TIME:units = "days since 1950-01-01T00:00:00Z"; ' TIME:axis = "T"; TIME:long_name = "time of measurement"; TIME:valid_min = 0.0; TIME:valid_max = 90000.0; TIME:QC_indicator = <X>; TIME:Processing_level = <Y>; TIME:uncertainty = <Z>; or TIME:accuracy = <Z>; TIME:comment = "Optional comment..."	Date and time (UTC) of the measurement in days since midnight, 1950-01-01. Example: Noon, Jan 2, 1950 is stored as 1.5. <X>: Text string from reference table 2. Replaces the TIME_QC if constant. Cf. note on quality control in data variable section,. <Y>: Text from reference table 3. <Z>: Choose appropriate value.
Float **LATITUDE**(LATITUDE); LATITUDE:standard_name = "latitude"; LATITUDE:units = "degrees_north"; LATITUDE:axis="Y"; LATITUDE:long_name = "latitude of measurement"; LATITUDE:reference="WGS84"; LATITUDE:coordinate_reference_frame="urn:ogc:crs:EPSG::4326"; LATITUDE:valid_min = -90.0; LATITUDE:valid_max = 90.0; LATITUDE:QC_indicator = <X>; LATITUDE:Processing_level= <Y>; LATITUDE:uncertainty = <Z>; or LATITUDE:accuracy = <Z>; LATITUDE:comment = "Surveyed anchor position";	Latitude of the measurements. Units: degrees north; southern latitudes are negative. Example: 44.4991 for 44° 29' 56.76'' N <X>: Text string from reference table 2. Replaces POSITION_QC if constant. <Y>: Text from reference table 3. <Z>: Choose appropriate value.
Float **LONGITUDE**(LONGITUDE); LONGITUDE:standard_name = "longitude"; LONGITUDE:units = "degrees_east"; LONGITUDE:axis="X"; LONGITUDE:reference="WGS84"; LONGITUDE:coordinate_reference_frame="urn:ogc:crs:EPSG::4326"; LONGITUDE:long_name = "Longitude of each location"; LONGITUDE:valid_min = -180.0; LONGITUDE:valid_max = 180.0; LONGITUDE:QC_indicator = <X>; LONGITUDE:processing_level = <Y>; LONGITUDE:uncertainty = <Z>; or LONGITUDE:accuracy = <Z>; LONGITUDE:comment = "Optional comment..."	Longitude of the measurements. Unit: degrees east; western latitudes are negative. Example: 16.7222 for 16° 43' 19.92'' E <X>: Text from reference table 2. Replaces POSITION_QC if constant. <Y>: Text from reference table 3. <Z>: Choose appropriate value.

Float **DEPTH**(DEPTH); DEPTH:standard_name = “depth”; DEPTH:units = “meters”; DEPTH:positive =<Q> DEPTH:axis=”Z”; DEPTH:reference=<R>; DEPTH:coordinate_reference_frame="urn:ogc:crs:EPSG:: <S>”; DEPTH:long_name = “Depth of measurement”; DEPTH:_FillValue = -99999.0; DEPTH:valid_min = 0.0; DEPTH:valid_max = 12000.0; DEPTH:QC_indicator = <X>; DEPTH:processing_level = <Y>; DEPTH:uncertainty = <Z>; or DEPTH:accuracy = <Z>; DEPTH:comment = “Depth calculated from mooring diagram”;	Depth of measurements. Example: 513 for a measurement 513 meters below sea surface. <Q>: “Positive” attribute may be ”up” (atmospheric, or oceanic relative to sea floor) or ”down” (oceanic). <R>: The depth reference default value is “sea_level”. Other possible values are : “mean_sea_level”, “mean_lower_low_water”, “wgs84_geoid” <S>: Use CRF 5831 for depth, or 5829 for height; relative to instantaneous sea level <X>: Text from reference table 2. Replaces DEPTH_QC if constant. <Y>: Text from reference table 3. <Z>: Choose appropriate value.

Notes on coordinate variables

- Time: By default, the time word represents the center of the data sample or averaging period. The base date in the ‘units’ attribute for time is represented in ISO8601 standard "YYYY-MM-DDThh:mm:ssZ"; note that UTC (Z) must be explicitly specified. This requirement is an extension to ISO8601.
- DEPTH: The depth variable may be positive in either upward or downward direction, which is defined in its “positive” attribute. The Z axis may be represented as pressure, if, for example pressure is recorded directly by an instrument and the calculation of depth from pressure would cause a loss of information. Depth is strongly preferred, since it allows data to be used more directly. Meteorological data should include a HEIGHT coordinate that is otherwise identical to DEPTH.
- The default depth reference is "sea_level" (free sea surface). In EPSG coordinate reference system, the default reference for DEPTH is: "urn:ogc:crs:EPSG::5831" and for HEIGHT: "urn:ogc:crs:EPSG::5829".
- The latitude and longitude datum is WGS84. This is the default output of GPS systems.
- Many coordinate variables for ocean data are nominal; an achor position, or a vertical position on a mooring chain. When there is supplemental data, like a GPS time series or a pressure measurement from one instrument, it may be provided as a data variable, and may be given an ‘axis’ attribute, but does not need to be specified as a coordinate.

2.4 Data variables

Data variables contain the actual measurements and information about their quality, uncertainty, and mode by which they were obtained. Different options for how quality indicators are specified are outlined in the notes below the table.

Recommended variable names are listed in Reference Table 6; replace <PARAM> with any of the names indicated there. Required attributes are marked as such, however, OceanSITES requests that all other attributes be used and contain meaningful information, unless technical reasons make this impossible.

<A>: standardized attributes listed in reference tables
<B>: attributes whose values are set following OceanSITES rules
<C>: attributes whose value is free text, set by the data provider

Type, name, dimension, attributes	Comment
Float <**PARAM**>(TIME, DEPTH, LATITUDE,LONGITUDE);	or: Float <**PARAM**>(TIME, DEPTH); or: Float <**PARAM**>(TIME);
<PARAM>:**standard_name** = <A>;	**standard_name: Required, if** there is an appropriate, existing standard name in CF.
<PARAM>:**units** = <A>;	**units:** Required
<PARAM>:**_FillValue** = <B>;	**_FillValue**: Required
<PARAM>:**coordinates** = <B>;	**coordinates: Required, if** a data variable does not have 4 coordinates in its definition.
<PARAM>:long_name = <B>;	l**ong_name**: text; should be a useful label for the variable
<PARAM>:QC_indicator = <A>;	**QC_indicator**: (OceanSITES specific) text, ref table 2
<PARAM>:processing_level = <A>;	**processing_level:** text, ref table 3
<PARAM>:valid_min = <B>;	**valid_min**: Float. Minimum value for valid data
<PARAM>:valid_max = <B>;	**valid_max**: Float. Maximum value for valid data
<PARAM>:comment = <C>;	**comment**: Text; useful free-format text
<PARAM>:ancillary_variables = <B>;	**ancillary_variables**: Text. Other variables associated with <PARAM>, e.g. <PARAM>_QC. List as space-separated string. Example: TEMP:ancillary_variables=”instrument TEMP_QC TEMP_UNCERTAINTY” NOTE: no term may appear in the list of ancillary variables that is not the name of a variable in the file.
<PARAM>:history = <B>;	**history:** Text. A series of entries with one line for each processing step performed on this variable, including the date, person’s name, action taken.
<PARAM>:uncertainty = <B>;	**uncertainty**: Float. Overall measurement uncertainty, if constant.
<PARAM>:accuracy = <B>;	**accuracy**: Float. Nominal accuracy of data.
<PARAM>:precision = <B>;	**precision**: Float. Nominal precision of data.
<PARAM>:resolution = <B>;	**resolution**: Float. Nominal resolution of data.
<PARAM>: cell_methods = <A>;	**cell_methods**: Text. Specifies cell method as per CF convention. Example: TEMP:cell_methods=”TIME: mean DEPTH: point LATITUDE: point LONGITUDE: point”. If all are ‘point’ this may be omitted.
<PARAM>:DM_indicator = <A>;	**DM_indicator**: Text. Data mode, if constant, as per reference table 4. See note on data modes below.
<PARAM>:reference_scale = <B>;	**reference_scale**: Text. For some measurements that are provided according to a standard reference scale specify the reference scale with this optional attribute. Example: ITS-90, PSS-78
<PARAM>:sensor_model = <Y>;	**sensor_*:** Text**.** Use thses fields to describe the sensor, unless the ancillary variable ‘instrument’ is used. See note on sensor metadata, below.
<PARAM>:sensor_manufacturer = <Y>;	
<PARAM>:sensor_reference = <Y>;	
<PARAM>:sensor_serial_number = <Y>;	**sensor_mount:** Text. Deployment characteristics, from ref table 7.
<PARAM>:sensor_mount=<A>	**sensor_orientation:** Text. Deployment characteristics, from ref table 8.
<PARAM>:sensor_orientation = <A>;	

Note on the ‘coordinates’ attribute:

There are two methods used to locate data in time and space. The preferred method is for the

data variable to be declared with dimensions that are coordinate variables, e.g. ATMP(TIME, DEPTH, LATITUDE, LONGITUDE). Alternatively, a variable may be declared with fewer dimensions, e.g. ATMP(TIME). In the latter case, the 'coordinates' attribute of the variable provides the spatiotemporal reference for the data. The value of the coordinates attribute is a blank separated list of the names of auxiliary coordinate variables; these must exist in the file, and their sizes must match a subset of the data variable's dimensions; scalar coordinates do so by default.

The use of coordinate variables as dimensions is preferred, because it conforms to COARDS and because it simplifies the use of the data by standard software. Note that it is permissible, but optional, to list coordinate variables as well as auxiliary coordinate variables in the coordinates attribute.

Note on sensor metadata:

Complete sensor information should be provided by one of two methods, which are outlined in Appendix 2. Fields should include model name, manufacturer, serial number, and a URL or reference that points to an instrument's specifications. This information may be presented in a series of attributes attached to a data variable, or via a single 'instrument' attribute. The instrument attribute points to a group of variables that contain the description of the sensors; the latter method allows two-dimensional information and avoids repetition of information.

Note on accuracy terms:

Accuracy is the closeness of the variable to the actual value; precision is the repeatability of the measurement, and resolution is the fineness to which the value can be displayed. Uncertainty combines accuracy and precision. These terms may be provided as attributes to the target data variables if they are constant over the dataset, or may be provided as ancillary variables if they change over depth or time.

2.5 Quality control variables

Data quality and provenance information for both coordinate variables and data variables is needed. If the quality control values are constant across all dimensions of a variable, the information may be given as text attributes of that variable; if they vary along one or more axes, they are provided as a separate numeric flag variable, with at least one dimension that matches the 'target' variable.

When QC information is provided as a separate flag variable, CF requires that these variables carry attributes 'flag_values' and 'flag_meanings'. These provide a list of possible values and their meanings. When this information is provided in the attributes of the target variables, it should be given in a human-readable form.

Description of QC attributes is provided above in the sections on data variables and coordinates. Below is a description of how to provide this information as a separate variable. Examples are given for coordinate and data variables; data variables are identified by the term <param> which represents a name from our list of variable names.

Type, name, dimension, attributes	Comment
Byte TIME_QC(TIME);	Quality flag for each TIME value.
Byte POSITION_QC(LATITUDE);	Quality flag for LATITUDE and LONGITUDE pairs.
Byte DEPTH_QC(DEPTH);	Quality flag for each DEPTH value.
Byte **<PARAM>_QC**(TIME, DEPTH); <PARAM>_QC:long_name = "quality flag for <PARAM>"; <PARAM>_QC:flag_values = 0, 1, 2, 3, 4, 7, 8, 9;	Quality flags for values of associated <PARAM>. The flag scale is specified in reference table 2, and is included in the flag_meanings attribute. **long_name**: type char. fixed value **flag_values**: type byte. Required; fixed value

<PARAM>_QC:flag_meanings = “unknown good_data probably_good_data potentially_correctable_bad_data bad_data nominal_value interpolated_value missing_value”	**flag_meanings**: type char. Required; fixed value
Char **<PARAM>_DM**(TIME, DEPTH); <PARAM>_DM:long_name = “data mode “; <PARAM>_DM:flag_values = “R”, “P”, “D”, “M”; <PARAM>_DM:flag_meanings = “real-time provisional delayed-mode mixed”;	This is the data mode, from reference table 4. Indicates if the data point is real-time, delayed-mode or provisional mode. It is included when the dataset mixes modes for a single variable. **long_name**: type char. **flag_values**: type char. **flag_meanings**: type char.
Float **<PARAM>_UNCERTAINTY**(TIME, DEPTH): <PARAM>_UNCERTAINTY:long_name = “uncertainty of <PARAM>” <PARAM>_UNCERTAINTY:_FillValue=<Y> <PARAM>_UNCERTAINTY:units = “<Y>”;	Uncertainty of the data given in <PARAM>. **long_name**: type char. Required; fixed value **_FillValue**: Float. Required. **units**: type char. Required. Must be the same as <PARAM>:units.

Example: Sea temperature with QC fields

Float TEMP(TIME, DEPTH);
TEMP:standard_name = “sea_water_temperature”;
TEMP:units = “degree_Celsius”;
TEMP:_FillValue = 99999.f;
TEMP:long_name = “sea water temperature in-situ ITS-90 scale”;
TEMP:QC_indicator = “Good data”;
TEMP:Processing_level =”Data manually reviewed”;
TEMP:coordinates = “TIME DEPTH LATITUDE LONGITUDE”
TEMP:valid_min = -2.0f;
TEMP:valid_max = 40.f;
TEMP:comment = “Provisional data”;
TEMP:uncertainty = 0.01f;
TEMP:accuracy = 0.01f;
TEMP:precision = 0.01f;
TEMP:cell_methods=”TIME: mean DEPTH: point LATITUDE: point LONGITUDE: point”.
TEMP:DM_indicator=”P”;
TEMP:reference_scale = “ITS-90”;

Example: Sea temperature QC variable

If there is no QC_indicator attribute in the TEMP variable,above, there must be a list of ancillary variables, e.g.
TEMP:ancillary_variables = “TEMP_QC TEMP_ uncertainty” ;
as well as the quality indicator variables, e.g.
BYTE TEMP_QC(TIME, DEPTH);
TEMP_QC:long_name = “quality flag of sea water temperature”;
TEMP_QC:conventions = “OceanSITES QC Flags”;
TEMP_QC:coordinates = “TIME DEPTH LATITUDE LONGITUDE”
TEMP_QC:flag_values = 0, 1, 2, 3, 4, 7, 8, 9;
TEMP_QC:flag_meanings = “unknown good_data probably_good_data potentially_correctable bad_data bad_data nominal_value interpolated_value missing_value”

FLOAT TEMP_ uncertainty (TIME, DEPTH);
TEMP_uncertainty:long_name = “uncertainty of sea water temperature”;
TEMP_ uncertainty:units = “degree_Celsius”;
TEMP_ uncertainty:_FillValue = 99999.f;
TEMP_ uncertainty:comment = ”Based on initial accuracy of .002, range of -5 to 35, drift of .0002/month and resolution of .0001 ” ;

3 Reference tables

3.1 Reference table 1: data_type

The data_type global attribute should have one of the valid values listed here.

Data type
OceanSITES profile data
OceanSITES time-series data
OceanSITES trajectory data

3.2 Reference table 2: QC_indicator

The quality control flags indicate the data quality of the data values in a file. The byte codes in column 1 are used only in the <PARAM>_QC variables to describe the quality of each measurement, the strings in column 2 ('meaning') are used in the attribute <PARAM>:QC_indicator to describe the overall quality of the parameter.

When the numeric codes are used, the flag_values and flag_meanings attributes are required and should contain lists of the codes (comma-separated) and their meanings (space separated, replacing spaces within each meaning by '_').

Code	Meaning	Comment
0	unknown	No QC was performed
1	good data	All QC tests passed.
2	probably good data	
3	potentially correctable bad data	These data are not to be used without scientific correction or re-calibration.
4	bad data	Data have failed one or more tests.
5	-	Not used
6	-	Not used.
7	nominal value	Data were not observed but reported. (e.g. instrument target depth.)
8	interpolated value	Missing data may be interpolated from neighboring data in space or time.
9	missing value	This is a fill value

3.3 Reference table 3: Processing level

This table describes the quality control and other processing procedures applied to all the measurements of a variable. The string values are used as an overall indicator (i.e. one summarizing all measurements) in the attributes of each variable in the processing_level attribute.

Raw instrument data
Instrument data that has been converted to geophysical values
Post-recovery calibrations have been applied
Data has been scaled using contextual information
Known bad data has been replaced with null values
Known bad data has been replaced with values based on surrounding data
Ranges applied, bad data flagged
Data interpolated
Data manually reviewed
Data verified against model or other contextual information
Other QC process applied

3.4 Reference table 4: Data mode

The values for the variables "<PARAM>_DM", the global attribute "data_mode", and variable attributes "<PARAM>:DM_indicator" are defined as follows:

Value	Meaning
R	Real-time data. Data coming from the (typically remote) platform through a communication channel without physical access to the instruments, disassembly or recovery of the platform. Example: for a mooring with a radio communication, this would be data obtained through the radio.
P	Provisional data. Data obtained after instruments have been recovered or serviced; some calibrations or editing may have been done, but the data is not thought to be fully processed. Refer to the history attribute for more detailed information.
D	Delayed-mode data. Data published after all calibrations and quality control procedures have been applied on the internally recorded or best available original data. This is the best possible version of processed data.
M	Mixed. This value is only allowed in the global attribute "data_mode" or in attributes to variables in the form "<PARAM>:DM_indicator". It indicates that the file contains data in more than one of the above states. In this case, the variable(s) <PARAM>_DM specify which data is in which data mode.

3.5 Reference table 5: Data Assembly Center codes

Data Assembly Centers and institutions	
BERGEN	University Of Bergen Geophysical Institute, NO
CCHDO	CLIVAR and Carbon Hydographic Office, USA
CDIAC	Carbon Dioxide Information Analysis Center, USA
EUROSITES	EuroSites project, EU
IMOS	Integrated Marine Observing System, AU
INCOIS	Indian National Centre for Ocean Information Services
JAMSTEC	Japan Agency for Marine-Earth Science and Technology
MBARI	Monterey Bay Aquarium Research Institute, USA
MEDS	Marine Environmental Data Service, Canada
NDBC	National Data Buoy Center, USA
NIOZ	Royal Netherlands Institute for Sea Research, NL
NOCS	National Oceanography Centre, Southampton UK
PMEL	Pacific Marine Environmental Laboratory, USA
SIO	Scripps Institution of Oceanography, USA
UH	University of Hawaii, USA
WHOI	Woods Hole Oceanographic Institution, USA

3.6 Reference table 6: Variable names

It is recommended that variable names start with a code based on SeaDataNet-BODC parameter discovery vocabulary. They are not strictly standardized, however; one should use the CF standard_name attribute to query data files. Note that a single standard name may be used more than once in a file, but short names are unique.

For example, if sea temperature on a mooring is measured by a series of 5 Microcats and by a profiler that produces values at 10 levels, it may be reported in a single file with 2 temperature variables and 2 depth variables. TEMP(TIME, DEPTH) could hold the Microcat data, if DEPTH is declared as a 5-element coordinate; and TEMP_prof(TIME, DEPTH_prof) could hold the profiler data if DEPTH_prof is declared as a 10-element coordinate. Both variables would have a standard_name of "sea_water_temperature". The following lists a subset of the OceanSITES recommended variable names.

Parameter	Standard name
AIRT	air_temperature
CAPH	air_pressure
CDIR	direction_of_sea_water_velocity
CNDC	sea_water_electrical_conductivity
CSPD	sea_water_speed
DEPTH	depth
DEWT	dew_point_temperature
DOX2	moles_of_oxygen_per_unit_mass_in_sea_water was dissolved_oxygen
DOXY	mass_concentration_of_oxygen_in_sea_water was dissolved_oxygen
DOXY_TEMP	temperature_of_sensor_for_oxygen_in_sea_water
DYNHT	dynamic_height
FLU2	fluorescence
HCSP	sea_water_speed
HEAT	heat_content
ISO17	isotherm_depth
LW	surface_downwelling_longwave_flux_in_air
OPBS	optical_backscattering_coefficient
PCO2	surface_partial_pressure_of_carbon_dioxide_in_air
PRES	sea_water_pressure
PSAL	sea_water_practical_salinity
RAIN	rainfall_rate
RAIT	thickness_of_rainfall_amount
RELH	relative_humidity
SDFA	surface_downwelling_shortwave_flux_in_air
SRAD	isotropic_shortwave_radiance_in_air
SW	surface_downwelling_shortwave_flux_in_air
TEMP	sea_water_temperature
UCUR	eastward_sea_water_velocity
UWND	eastward_wind
VAVH	sea_surface_wave_significant_height
VAVT	sea_surface_wave_zero_upcrossing_period
VCUR	northward_sea_water_velocity
VDEN	sea_surface_wave_variance_spectral_density
VDIR	sea_surface_wave_from_direction
VWND	northward_wind
WDIR	wind_to_direction
WSPD	wind_speed

3.7 Reference table 7: Sensor mount characteristics

The way an instrument is mounted on a mooring may be indicated by the attribute <PARAM>:”sensor_mount” or by a character variable. The following table lists the valid sensor_mount values.

sensor_mount
mounted_on_fixed_structure
mounted_on_surface_buoy
mounted_on_mooring_line
mounted_on_bottom_lander
mounted_on_moored_profiler
mounted_on_glider
mounted_on_shipborne_fixed
mounted_on_shipborne_profiler
mounted_on_seafloor_structure
mounted_on_benthic_node
mounted_on_benthic_crawler
mounted_on_surface_buoy_tether
mounted_on_seafloor_structure_riser
mounted_on_fixed_subsurface_vertical_profiler

3.8 Reference table 8: Sensor orientation

When appropriate, the orientation of an instrument such as an ADCP should be provided, either as the variable attribute ”sensor_orientation” or as a variable. The following table lists the valid sensor_orientation values.

sensor_orientation	example
downward	ADCP measuring currents from surface to bottom.
upward	In-line ADCP measuring currents towards the surface
horizontal	Optical sensor looking ‘sideways’ from mooring line

4 OceanSITES NetCDF products format

Products derived from OceanSITES data are available on the GDAC servers in the PRODUCT directory of each site. Products derived from multiple sites are available from PRODUCT_MULTIPLE_SITES directory.

This manual does not specify the products data format. However, OceanSITES products should be distributed as NetCDF files and follow the same convention as our standard data files: CF for parameters and attributes, NUG/CF/ACDD for global attributes, Udunits for units.

The NetCDF global attribute “site_code” is required. On multiple sites products, it must contain a blank-separated list of the site codes that contributed to the product.

Example
site_code = “RAPID-1 RAPID-2 RAPID-3 RAPID-4 RAPID-5 RAPID-6 RAPID-7”

5 File naming conventions

Almost all OceanSITES NetCDF files are named using this convention:

- OS - OceanSITES prefix
- [PlatformCode] - Platform code from the OceanSITES catalogue
- [DeploymentCode]- Deployment code (unique code for deployment - date or number)
- [DataMode] - Data Mode
    - R: real-time data
    - P : provisional data
    - D: delayed mode data
    - M: mixed delayed mode and real-time data
- [PARTX] - An optional user defined field for identification of data

Remark: the field separator in the file name is "_". This character must not be used in any of the file name's fields.

5.1 data files

Data files normally contain one type of data, from one deployment.

Data files are found in the directory /DATA/[SiteCode]

Example: /DATA/CIS/OS_CIS-1_200905_R_CTD.nc
This file contains temperature and salinity data from the CIS-1 platform, from the May 2009 deployment.

5.2 gridded data files

Gridded data files are processed from one or more deployment data files. Data may be gridded in time or vertical axes. Be aware that when multiple data files are aggregated, the attributes may not contain all the detailed information of each individual data file.

There may be multiple files derived from the same data but expressed in different scale (e.g.: daily, hourly averages).

Gridded data files are found in the directory /DATA/[SiteCode]/GRIDDED/

Example: /DATA/CIS/GRIDDED/OS_CIS-1_200206_M_CTD-hourly.nc

This file contains hourly averaged temperature and salinity data from the CIS-1 platform, from the June 2002 deployment to present.

5.3 product data files

The product data files are processed from one or more deployment data files. Their variables are derived from one or more files from the DATA directory. Product data files may not contain all the detailed information of each individual data file.

There may be multiple files derived from the same data.

Product data files are found in the directory /DATA/[SiteCode]/PRODUCTS

Example: /DATA/CIS/PRODUCTS/OS_CIS-1_200206_M_AirSeaFluxes-hourly.nc

This file contains hourly air-sea fluxes data from the CIS-1 mooring, from the June 2002 deployment to present.

5.4 products derived from multiple sites

The directory /PRODUCTS_MULTIPLE_SITES contains products from multiple sites; filenames are slightly different from single-site files.

/PRODUCTS_MULTIPLE_SITES/[ProductName]/
OS_[ProductFileIndentification]_[VariablesIndicator].nc

- [ProductName]: name of the product
- [ProductFileIndentification] : file name of the product
- [VariablesIndicator]: optional indicator of the variables

Please indicate the sites used in the site_code attribute.

Example
PRODUCTS_MULTIPLE_SITES/RapidTransportSection/
OS_RapidTransportSection2010.nc

6 OceanSITES Data Management topics

OceanSITES is comprised of three organizational units: PIs, DACs, GDACs.

The **Principal Investigator (PI)**, typically a scientist at a research institution, maintains the observing platform and the sensors that deliver the data. He or she is responsible for providing the data and all auxiliary information to a **Data Assembly Center (DAC)**; a PI may also act as a DAC.

The **DAC** assembles OceanSITES-compliant files from this information and delivers these to the two **Global Data Assembly Centers (GDACs)**, where they are made publicly available.

The **GDACs** distribute the best copy of the data files. When a higher quality data file (e.g. calibrated data) is available, it replaces the previous version of the data file.
The user can access the data at either GDAC, which are synchronized daily.

Data archive will be implemented by the National Ocean Data Center of the USA; our documented data specification is critical to the archive process.

6.1 Global Data Assembly Centers

Two global data assembly centers (GDACs) provide access points for OceanSITES data. One is in France at Coriolis (http://www.coriolis.eu.org), the other is in the US at NOAA's National Data Buoy Center (NDBC , http://www.ndbc.noaa.gov).

The servers at the GDACs are synchronized at least daily to provide the same OceanSITES data redundantly.

The user can access the data at either GDAC's ftp site:

- ftp://data.ndbc.noaa.gov/data/oceansites
- ftp://ftp.ifremer.fr/ifremer/oceansites

Data is organized by site and by resource type in the GDACs' DATA directories:
DATA/site/FileName.nc where site is the OceanSITES site code.

6.2 Index file: GDAC data inventory

To allow for data discovery without downloading the data files themselves, an 'index file' is created by each of the GDACs. The index file is a comma-separated-values text file named oceansites_index.txt, in the root directory of each GDAC. It contains a list of the files on the server, and metadata extracted from those files.

The file contains a header section, lines of which start with # characters, the list of all data files available on the GDAC, and their descriptions.

Each line contains the following information:

- file: the file name, beginning from the GDAC root directory
- date_update: the update date of the file, YYYY-MM-DDTHH:MI:SSZ
- start_date: first date for observations, YYYY-MM-DDTHH:MI:SSZ
- end_date: last date for observations, YYYY-MM-DDTHH:MI:SSZ
- southern_most_latitude
- northern_most_latitude
- western_most_longitude
- eastern_most_longitude
- geospatial_vertical_min
- geospatial_vertical_min
- update_interval: M monthly, D daily, Y yearly, V void
- size: the size of the file in megabytes
- gdac_creation_date: date of creation of the file on the GDAC
- gdac_update_date: date of update of the file on the GDAC.
- data_mode: R, P, D, M (real-time, provisional, delayed mode, mixed; see reference table 5)
- parameters: list of parameters (standard_name) available in the file separated with blank

The fill value is empty: ",,".

GDAC index file: oceansites_index.txt

#OceanSITES Global Data Assembly Center (GDAC) Index File
#Two GDACs FTP servers are on-line at ftp://data.ndbc.noaa.gov/data/oceansites and
ftp://ftp.ifremer.fr/ifremer/oceansites
#Also a THREDDS server is available at
http://dods.ndbc.noaa.gov/thredds/catalog/data/oceansites/catalog.html
#For more information, please contact: http://www.oceansites.org
#
#This OceanSITES index file was last updated on : 2013-04-16T13:30:01Z. Columns are defined as follows:
#FILE (relative to current file directory), DATE_UPDATE, START_DATE, END_DATE, SOUTHERN_MOST_LATITUDE, NORTHERN_MOST_LATITUDE, WESTERN_MOST_LONGITUDE, EASTERN_MOST_LONGITUDE, MINIMUM_DEPTH, MAXIMUM_DEPTH, UPDATE_INTERVAL, SIZE (in bytes),GDAC_CREATION_DATE,GDAC_UPDATE_DATE,DATA_MODE (R: real-time D: delayed mode M: mixed P: provisional),PARAMETERS (space delimited CF standard names)
#
DATA/ANTARES/OS_ANTARES-1_200509_D_CTD.nc,2011-04-06T08:41:10Z,2005-09-15T12:00:13Z,2006-12-31T23:55:21Z,42.7,42.9,6.15,6.19,0,2500,void,3064416,2011-02-22T21:07:27Z,2011-04-08T04:31:05Z,D,time depth latitude longitude sea_water_temperature sea_water_electrical_conductivity sea_water_salinity depth
DATA/ANTARES/OS_ANTARES-1_200701_D_CTD.nc,2011-04-06T08:41:24Z,2007-01-01T00:01:48Z,2007-12-31T23:58:26Z,42.7,42.9,6.15,6.19,0,2500,void,2860400,2011-02-22T21:07:27Z,2011-04-08T04:31:05Z,D,time depth latitude longitude sea_water_temperature sea_water_electrical_conductivity sea_water_salinity depth

7 Appendix 1: Further Information, links, tools

- OceanSITES website: http://www.oceansites.org
- NetCDF: We attempt to follow NetCDF Best Practices, described at unidata.ucar.edu/software/netcdf/docs/BestPractices.html
- CF: We implement and extend the NetCDF Climate and Forecast Metadata Convention, including the CF standard names, available at cfconventions.org
- Udunits: Units are from the Udunits package as implemented by CF unidata.ucar.edu/software/udunits/
- ISO8601: Description available at http://en.wikipedia.org/wiki/ISO_8601
- ACDD: Our global attributes implement the Unidata NetCDF Attribute Convention for Dataset Discovery, at: http://wiki.esipfed.org/index.php/Category:Attribute_Conventions_Dataset_Discovery
- We recommend consulting the NODC NetCDF Templates available at http://www.nodc.noaa.gov/data/formats/netcdf/.
- The SeaDataNet-BODC vocabulary server is the repository for many oceanographic terms:

  Sea Area http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=C16

  Parameter codes http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P02

  Platform category http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=L06
- THREDDS: The 'cdm_data_type' global attribute is used by THREDDS. More information at unidata.ucar.edu/projects/THREDDS/CDM/CDM-TDS.htm
- EPSG, used for the coordinate reference frames: http://www.epsg.org/
- WMO: For information about unique numbering of OceanSITES Moorings and Gliders, see: http://www.wmo.int/pages/prog/amp/mmop/wmo-number-rules.html
- OceanSITES file format checker (java) and file converters are freely available at: http://projets.ifremer.fr/coriolis/Observing-the-ocean/Observing-system-networks/OceanSITES/Access-to-data

8 Appendix 2: Sensor and instrument metadata

There are two methods for providing complete sensor metadata. In method 1, the variable attribute ‘instrument’ points to an umbrella variable that describes an instrument and its sensor suite; the instrument variable ties one or more instruments to one or more data variables.

Instrument variables may include manufacturer, model, serial number, a reference URL that points to a web resource describing the sensor, sensor mount and orientation. Orientation may not be needed for all variables but is highly recommended for current meters and profilers.

Method 1 example:

```
variables:
double TEMP(TIME, DEPTH) ;
        TEMP:instrument = "INST" ;
double PSAL(TIME, DEPTH) ;
        PSAL:instrument = "INST" ;
int INST ;
        INST:long_name = "instruments" ;
        INST:ancillary_variables = "INST_MFGR INST_MOD INST_SN INST_URL INST_MOUNT" ;
char INST_MFGR(DEPTH, strlen1) ;
        INST_MFGR:long_name = "instrument manufacturer" ;
char INST_MODEL(DEPTH, strlen2) ;
        INST_MODEL:long_name = "instrument model name" ;
int INST_SN(DEPTH) ;
        INST_SN:long_name = "instrument serial number" ;
char INST_URL(DEPTH, strlen3) ;
        INST_URL:long_name = "instrument reference URL" ;
char INST_MOUNT(DEPTH, strlen3) ;
        INST_MOUNT:long_name = "instrument mount" ;
data:
INST = _ ;  (an empty variable, aka an umbrella)
INST_MFGR =
        "RBR-Global       ",
        "Seabird Electronics",
        "Seabird Electronics" ;
INST_MODEL =
        "TR1060",
        "SBE37 ",
        "SBE16 ";
INST_MOUNT =
        “mounted_on_surface_buoy”,
        “mounted_on_mooring_line”,
        “mounted_on_seafloor_structure_riser";
INST_SN = 14875, 1325, 1328;
INST_URL =
        "http://www.rbr-global.com/products/tr-1060-temperature",
        "http://www.seabird.com/products/spec_sheets/37smdata.htm",
        "http://www.seabird.com/16plus_ReferenceSheet.pdf" ;
```

In method 2, sensor information is supplied as a series of attributes for a data variable. These attributes may contain comma separated lists to provide different information for multiple instruments.

Method 2 example:

double TEMP(TIME, DEPTH) ;

TEMP:sensor_name = ‘RBR-Global TR1060, SBE23,SBE16’

TEMP:sensor_serial_number = 14875, 1325, 1328

TEMP:sensor_mount=“mounted_on_surface_buoy, mounted_on_mooring_line, mounted_on_seafloor_structure_riser";

TEMP:sensor_orientation = “vertical”;

double PSAL(TIME, DEPTH) ;

PSAL:sensor_name = ‘RBR-Global TR1060, SBE23,SBE16’

PSAL:sensor_serial_number = 14875, 1325, 1328

PSAL:sensor_mount=“mounted_on_surface_buoy, mounted_on_mooring_line, mounted_on_seafloor_structure_riser";

PSAL:sensor_orientation = “vertical”;

9 Appendix 3: Glossary

This chapter gives a definition for the OceanSITES items described in this manual.

Site

An OceanSITES site is a defined geographic location where sustained oceanographic, meteorological or other observations are made. Example: CIS is a site in the Central Irminger Sea.

Note: A site should be thought of as a point in space, i.e. a nominal position, with a small area extent around it, such that successive observations from anywhere within this area reasonably represent conditions at the nominal position for the major scientific questions that the observations address.

Array

An OceanSITES array is a grouping of sites based on a common and identified scientific question, or on a common geographic location.

Example: An IRMINGERSEA array would identify the sites CIS, LOCO-IRMINGERSEA, and OOI-IRMINGERSEA as sharing a common scientific interest and/or geographic location.

Notes: It is valid for a single site to belong to no, one, or multiple arrays.

Documenting the array is recommended only if it identifies commonalities beyond a single project or a single operating institution.

Network

An OceansSITES network is a grouping of sites based on common shore-based logistics or infrastructure.

Example: EuroSITES, although technically a single project, bundles multiple institutional efforts and connects otherwise remote sites to a degree that warrants calling it a network.

Notes: It is valid for a single site to belong to no, one, or multiple networks. Documenting the network is recommended only if it identifies structures beyond a single project or a single operating institution.

Platform

An OceanSITES platform is an independently deployable package of instruments and sensors forming part of site. It may be fixed to the ocean floor, may float or may be self-propelled.

Examples: ‘CIS-1’: a mooring in Central Irminger Sea and ‘THETYS II’: a vessel that performs regular CTDs at DYFAMED site.

Deployment

An OceanSITES deployment is an instrumented platform performing observations for a

period of time. Changes to the instrumentation or to the spatial characteristics of the platform or its instruments constitute the end of the deployment.

Examples: The CTD data for CIS-1 deployment performed in May 2009 (200905) and are distributed as OS_CIS-1_200905_R_CTD.nc file.

Instrument

An OceanSITES instrument is device that provides digital data output.

Examples: CTD, ADCP, Meteorological Package.

Sensor

A device that measures environmental parameter but does not digitize data for transmission, it needs to be connected to an instrument to produce a data stream that a computer can read.
Examples: Transmissiometer, Fluorometer, Oxygen sensor.