



OceanSITES Data Format Reference Manual

NetCDF Conventions and Reference Tables

Version 1.4

July 16, 2020

History

Version	Date	Comment
0.1	March 2003	Creation of the document
1.0	Feb – April 2006	PLATFORM_CODE, SITE_CODE, WMO_PLATFORM_CODE added DATA_MODE set at measurement level (not global) File naming convention, data codes updated
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
1.1	April - June 2008	General revision based on OceanSITES 2008 meeting Epic codes removed, Use ISO8601 for string dates Update global attribute section for CF-1.1 compatibility New dimensions for DEPTH, LATITUDE, LONGITUDE Add an uncertainty attribute
1.2	September 2009 – March 2013	No fill value allowed for coordinates. Use WGS84 datum for latitude and longitude, EPSG coordinate reference for depth. Add optional attribute "reference" for DEPTH Add optional attributes "sensor_mount" "sensor_orientation" Update data file naming convention Add fields to the GDAC index file. Update QC flag scale (6 is not used) Add optional "array" and "network" global attributes
1.3.0 draft	April 2013 – Jan 2015	Seoul: Naming and directory conventions for gridded, product files. Short names no longer standardized Redefine data mode P, correct OGC DEPTH:coordinate_reference_frame syntax
1.4	Dec 2019- May 11, 2020	NRG, JK: Add JCOMMOPS fields. Global variables: deployment and recovery date/ship/cruise, ship ICES code and expocode for cruise. Variable attributes: instruments L22 code, sensor data start and end. Also: some long time series and product file text, and change guidance on 'area' (sea_area) attribute to use C19. Update or remove some broken links. Clarify platform definition, add creator fields. Add theme, <param>_URI.

Table of Contents

Contents

1. Overview	6
1.1 About OceanSITES	6
1.2 Broader Context.....	6
1.3 OceanSITES Organizational model.....	6
1.4 About this document.....	7
1.4.1 Technical Documentation available.....	7
1.5 User Obligations	7
1.6 Disclaimer.....	7
1.7 Feedback.....	8
2. OceanSITES NetCDF Data Format for Primary Observational Data.....	9
2.1 New features in this release.....	9
2.2 Global attributes	9
2.2 Dimensions.....	16
2.3 Coordinate variables.....	16
2.4 Data variables	18
2.5 Quality control variables	20
3. Reference tables	23
3.1 Reference table 1: data_type	23
3.2 Reference table 2: QC_indicator	23
3.3 Reference table 3: Processing level.....	24
3.4 Reference table 4: Data mode	24
3.5 Reference table 5: Data Assembly Center codes.....	25
3.6 Reference table 6: Identifying data variables	25
3.7 Reference table 7: Sensor mount characteristics.....	27
3.8 Reference table 8: Sensor orientation.....	28
4. Data Files.....	29
4.1. Deployment Data Files.....	29
4.1.1 Deployment Data files Naming Convention	29
4.2. Merged, Gridded and Derived Data Files	29
4.2.1 Merged, Gridded and Derived Data File Metadata Conventions.....	30
4.2.2 Merged, Gridded or Product Data File Naming Conventions.....	30
4.2.3 Higher-Level File Locations	31
5. OceanSITES Data Management topics	31

5.1 Global Data Assembly Centers	32
5.2 Index file: GDAC data inventory	32
5.3 Sensor and instrument metadata.....	33
6. Appendices	35
6.1 Appendix 1: Further Information, links, tools.....	35
6.2 Appendix 2: Glossary	35

1. Overview

1.1 About OceanSITES

The OceanSITES programme is the global network of open-ocean sustained time series sites, called ocean reference stations, being implemented by an international partnership of researchers and agencies. OceanSITES provides fixed-point time series of various physical, biogeochemical, ecosystem and atmospheric variables at 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 program is available at: <http://www.oceansites.org>.

1.2 Broader Context

OceanSITES contributes to the Global Ocean Observing System (GOOS) and the Global Climate Observing System (GCOS) which are co-sponsored by the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission of UNESCO (IOC-UNESCO), the United Nations Environment Programme (UNEP), and the International Science Council (ISC). Under the umbrella of the Observations Coordination Group (OCG) of GOOS, OceanSITES observing activities are coordinated with numerous platform-based networks including Argo, GO-SHIP, OceanGliders, and the Data Buoy Cooperation Panel (DBCP), and others. Through OCG a strong link is created to the WMO-IOC Joint Centre for Oceanography and Marine Meteorology Observation Programmes Support (JCOMMOPS), which hosts a metadata base for the OCG ocean observing platforms. In the current revision the OceanSITES metadata conventions seek to accord with or at least to minimize the translations required to allow ingestion of OceanSITES metadata into JCOMMOPS, and thus make OceanSITES more visible and quantifiable in the global context.

1.3 OceanSITES Organizational model

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.

OceanSITES observing objectives are grouped around core themes. Currently, these themes have been defined:

- Transport Moored Arrays: Installations designed to study volume and property transport
- Air/sea flux reference sites: Studies of the ocean/atmosphere interface (long time

- series & boundary layer physics, gas uptake)
- Global Ocean Watch: Multidisciplinary long time series at regions considered “representative” for larger areas (biogeochemical provinces, gyres, etc.)
 - Deep-Ocean T/S Challenge: Platforms collecting deep Ocean observations (below 2000m), especially temperature salinity measurements near the sea floor.

As new themes are developed, they will be incorporated, with coordination between the Science Steering Team and the Project Office.

1.4 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.

1.4.1 Technical Documentation available

Technical documentation of the OceanSITES system consists of **three** pieces

OceanSITES Data Format Reference Manual This manual.

OceanSITES Data Users' Guide For data users, formerly called How to Access OceanSITES Data, this document contains an outline of Global Data Assembly Center (GDAC) data directory structure and ftp/opensdap access, data use policy/license, list of sites, Data Assembly Centers (DACs), etc. It should be used in conjunction with the Data Format Reference Manual.

OceanSITES Data Providers' Guide For data producers: DACs and Principal Investigators (PIs), based on the earlier document How to Work with GDAC. This document contains guidelines for providing metadata and data, file naming scheme, and how to upload data to the system. It should be used in conjunction with the Data Format Reference Manual.

1.5 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 to the program 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 this or an earlier 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 program and the national programs that contribute to it.”

1.6 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.

1.7 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 (projectoffice@oceansites.org).

2. OceanSITES NetCDF Data Format for Primary Observational Data

The primary observational data that OceanSITES reports comes from individual deployments of moorings, or sometimes, repeat ship visits. This section describes the data format specifications for data files containing these primary data sets. Data are typically shown at the native instrumental resolution in time and space. For data that merges multiple deployments, as well as derived products, please refer to the later section.

OceanSITES uses netCDF (Network Common Data Form), a set of software libraries and machine-independent data formats developed by the Unidata program at UCAR. 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 interpret and use appropriately. To the extent possible, netCDF files should be self-describing; 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, including implementing Unidata's netCDF Attribute Convention for Data Discovery (ACDD). Further additions are needed for efficient aggregation by the GDACs, for improved access via the JCOMMOPS metadata portal, and to ensure that the data can be created and understood by basic netCDF utilities. Fields required by ACDD, by the GDACs, or by the JCOMMOPS metadata portal are indicated in the table below.

- 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. In this document, the term 'project' may refer to a single OceanSITES site, or to a group of sites that are managed by a single DAC, or share funding or infrastructure.

2.1 New features in this release

The global scale and cross network coordination provided by the JCOMMOPS metadata portal requires information not previously defined by the OceanSITES netCDF specification. New fields in this release have been added primarily to provide better conformance with the JCOMMOPS metadata database, and these are identified in their descriptions.

2.2 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 the OceanSITES Global Data Assembly Center (GDAC) inventory software or the JCOMMOPS metadata database.

Discovery and identification		
name	example	note
site_code	site_code="CIS" (OceanSITES specific)	Name of the site within OceanSITES program where this platform has been installed. Site codes must be approved by the OceanSITES Project Office to avoid duplication . Required (GDAC)
platform_code	platform_code="CIS-1" (OceanSITES specific)	A unique platform code. This code is either assigned by the site PI (see principle_investigator below) or by the data provider. Required . (GDAC)
data_mode	data_mode="R" (OceanSITES specific)	Indicates if the file contains real-time, provisional or delayed-mode quality controlled data. The list of valid data modes is in reference table 4. Required . (GDAC)
title	title="Real time CIS-1 Mooring Temperatures"	Free-format text describing the dataset, for use by human readers. Use the file name if in doubt. (NUG)
theme	theme="Air/sea flux reference, Global Ocean Watch"	List of OceanSITES theme areas to which this dataset belongs
summary	summary="Oceanographic mooring data from the first deployment of the CIS-1 platform at the CIS site 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="OceanS ITES"	The organization that manages data set names. (ACDD)
id	id="OS_CIS- 1_200502_TS"	The "id" combined with "naming_authority" attributes 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_c ode	wmo_platform_code="4840 9" (OceanSITES specific)	WMO (World Meteorological Organization) identifier. This platform number is unique within the OceanSITES project. (JCOMMOPS)
source	source="subsurface mooring"	Use a term from the SeaVoX Platform Categories vocabulary (L06) list, usually one of the following: "moored surface buoy", "subsurface mooring", "ship" (CF)

principal_investigator	principal_investigator="Alice Juarez" (OceanSITES specific)	Name of the person responsible for the scientific project that produced the data contained in the file. If needed, use a comma-separated list. Guest investigators e.g. adding interdisciplinary studies to an existing site should specify the PI of the site, and use the 'creator' attributes to identify themselves.
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. If needed, use a comma-separated list.
principal_investigator_url	principal_investigator_url="whoi.edu/profile/AJuarez"	URL with information about the project lead.
principal_investigator_id	principal_investigator_id="https://orcid.org/0000-0001-5044-7079"	ID, such as ORCID, of the person responsible for the project that collected the data contained in the file. If needed, use a comma-separated list.
creator_name	creator_name="Alice Juarez"	The name of the person (or other creator type) principally responsible for creating the data in the file. If needed, use a comma-separated list. (ACDD)
creator_email	creator_investigator_email="AJuarez AT whoi.edu"	The email address of the person principally responsible for creating this data. (ACDD)
creator_url	creator_url="whoi.edu/profile/AJuarez"	The URL of the person (or other creator type) principally responsible for creating this data. If needed, use a comma-separated list. (ACDD)
creator_id	creator_id="https://orcid.org/0000-0001-5044-7079"	ID, such as ORCID, of the person principally responsible for creating the data contained in the file. If needed, use a comma-separated list. (ACDD, optional)
creator_type	creator_type='institution'	Specifies type of creator with one of the following: 'person', 'group', 'institution', or 'position'. If not specified, the creator is assumed to be a person. (ACDD, optional)
creator_institution	creator_institution='WHOI'	The institution of the creator; should uniquely identify the creator's institution. This attribute's value should be specified even if it matches the value of publisher_institution, or if creator_type is institution.
array	array="TAO" (OceanSITES specific)	A grouping of sites based on a common and identified scientific question, or on a common geographic location or other rationals..
network	network="EuroSITES" (OceanSITES specific)	A grouping of sites based on common shore-based logistics, funding, 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
sea_area	area="North Atlantic Ocean" (OceanSITES specific)	Geographical coverage. Please use the SeaVox Water Body Gazetteer vocabulary (C19)
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 'duration' convention (examples: P1Y ,P3M, P10D) (ACDD)
time_coverage_resolution	time_coverage_resolution="PT30M"	Interval between records: Use ISO 8601 (PnYnMnDnHnMnS) 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)
platform_deployment_date	platform_deployment_date="2010-02-20T00:00:00Z"	Date and time in ISO format of the deployment of the buoy or other platform (JCOMMOPS)
platform_deployment_ship_name	R/V Melville	Ship names can be found on operators' sites, or on https://ocean.ices.dk/codes/ShipCodes.aspx (JCOMMOPS)
platform_deployment_cruise_name	MV1406	Cruise names may be found on operators' sites, or on rvdata.us (JCOMMOPS)
platform_deployment_ship_ICES_code	platform_deployment_ship_ICES_code='318M'	See Appendix 1 for ICES codes (JCOMMOPS)
platform_deployment_cruise_ExpoCode	318M20100220	ICES ship code, plus cruise start date (JCOMMOPS)
platform_recovery_date	platform_recovery_date="2012-01-13T00:00:00Z"	Date and time in ISO format of the recovery of the buoy or other platform (JCOMMOPS)
platform_recovery_ship_name	R/V Endeavor	Ship names can be found on operators' sites, or at https://ocean.ices.dk/codes/ShipCodes.aspx (JCOMMOPS)
platform_recovery_cruise_name	EN472	Cruise names may be found on operators' sites, or on rvdata.us (JCOMMOPS)
platform_recovery_ship_ICES_code	platform_recovery_ship_ICES_code='32EV'	See Appendix 1 for ICES codes (JCOMMOPS)
platform_recovery_cruise_ExpoCode	32EV2012013	ICES ship code, plus cruise start date (JCOMMOPS)
data_type	data_type="OceanSITES time-series data"	From Reference table 1: OceanSITES specific. (GDAC)
Conventions used		
name	example	note
format_version	format_version="1.5" (OceanSITES specific)	OceanSITES format version; may be 1.1, 1.3, 1.5. (GDAC)
Conventions	Conventions="CF-1.6, OceanSITES-1.5, 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="AI"	Name of the person responsible for metadata and formatting

	Plueddemann	of the data file. (ACDD)
publisher_email	publisher_email="aplueddemann at whoi.edu"	Email address of person responsible for metadata and formatting of the data file. (ACDD)
publisher_url	publisher_url="http://www.who.edu/profile/aplueddemann/"	Web address of the institution or of the data publisher. (ACDD)
publisher_ID	https://orcid.org/0000-0001-5044-7079	unique ID, such as ORCID, of the person responsible for the publication of the data. Available from https://orcid.org/
references	references="http:// www.oceansites.org , http://www.noc.soton.ac.uk/animate/index.php "	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="G EOMAR" (OceanSITES specific)	Data Assembly Center (DAC) in charge of this data file. A partial list of the data assembly centers is 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 Variability and Predictability) standards, cf. http://www.clivar.org/resources/data/data-policy . 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/resources/data/data-policy (ACDD)
citation	citation="These data were collected and made freely available by the OceanSITES program 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, the name of the PI, the site name, platform code, data access date, time, and URL, and, if available, the DOI of the dataset.
acknowledgement	acknowledgement="Principal funding for the NTAS experiment is provided by the US 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 ="2016-04-11T08:35:00Z"	The date on which this file was created. Version date and time for the data contained in the file. See note on time format below. (ACDD)
date_modified	date_modified="2017-03-01T15:00:00Z"	The date on which this file was last modified. (ACDD)

history	history= "2012-04-11T08:35:00Z data collected, A. Meyer. 2013-04-12T10:11:00Z 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, all important 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 collection, editing, or publication of the data in the file. (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

- Format for date and time attributes: Use ISO 8601 combined date and time representations, and always specify UTC with the trailing 'Z' to avoid any confusion. "2007-04-05T14:30Z"
- The file dates, date_created and date_modified, are our interpretation of the file dates as defined by ACDD. 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 acceptable. This information is linked to the site information, and may not be specific to the platform deployment.
- 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 constraints on acceptable coordinate variables; see CF Documentation.
- Note on ExpoCodes:

The ExpoCode is generated using the 4 character ICES platform code followed by the cruise departure date (YYYYMMDD format).

Example: US research vessel Nathaniel B. Palmer (ICES ship code: 3206), starting on 2011-02-19:

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 Geometries of the CF Conventions document, <http://cfconventions.org/cf-conventions/cf-conventions.html#discrete-sampling-geometries>, 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";	Date and time (UTC) of the measurement in days since midnight, 1950-01-01. Example:

<p>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..."</p>	<p>Noon, Jan 2, 1950 is stored as 1.5.</p> <p><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.</p>
<p>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:def:crs:EPSG::4326";</p> <p>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";</p>	<p>Latitude of the measurements. Units: degrees north; southern latitudes are negative.</p> <p>Example: 44.4991 for 44° 29' 56.76" N</p> <p><X>: Text string from reference table 2. Replaces POSITION_QC if constant. <Y>: Text from reference table 3. <Z>: Choose appropriate value.</p>
<p>Float LONGITUDE(LONGITUDE); LONGITUDE:standard_name = "longitude"; LONGITUDE:units = "degrees_east"; LONGITUDE:axis="X"; LONGITUDE:reference="WGS84"; LONGITUDE:coordinate_reference_frame="urn:ogc:def:crs:EPSG::4326"; LONGITUDE:long_name = "Longitude of measurement";</p> <p>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..."</p>	<p>Longitude of the measurements. Unit: degrees east; western latitudes are negative.</p> <p>Example: 16.7222 for 16° 43' 19.92" E</p> <p><X>: Text from reference table 2. Replaces POSITION_QC if constant. <Y>: Text from reference table 3. <Z>: Choose appropriate value.</p>
<p>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:def:crs:EPSG:: <S>"; DEPTH:long_name = "Depth of measurement";</p> <p>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";</p>	<p>Depth of measurements.</p> <p>Example: 513 for a measurement 513 meters below sea surface.</p> <p><Q>: "Positive" attribute may be "up" (atmospheric, or oceanic relative to sea floor) or "down" (oceanic).</p> <p><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.</p>

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:def:crs:EPSG::5831" and for HEIGHT: "urn:ogc:def: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 anchor 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

: 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 = ;	_FillValue: Required
<PARAM>:coordinates = ;	coordinates: Required , if a data variable does not have 4 coordinates in its definition.
<PARAM>:long_name = ;	long_name: text ; should be a useful label for the variable
<PARAM>:URI = ;	URI: text , points to the definition of the parameter

<p><PARAM>:QC_indicator = <A>; <PARAM>:processing_level = <A>;</p> <p><PARAM>:valid_min = ; <PARAM>:valid_max = ;</p> <p><PARAM>:comment = <C>;</p> <p><PARAM>:ancillary_variables = ;</p> <p><PARAM>:uncertainty = ; <PARAM>:accuracy = ; <PARAM>:precision = ; <PARAM>:resolution = ;</p> <p><PARAM>: cell_methods = <A>;</p> <p><PARAM>:DM_indicator = <A>;</p> <p><PARAM>:reference_scale = ;</p> <p><PARAM>:sensor_model = <Y>; <PARAM>:sensor_manufacturer = <Y>; <PARAM>:sensor_SeaVoX_L22_code = ; <PARAM>:sensor_reference = <Y>; <PARAM>:sensor_serial_number = <Y>; <PARAM>:sensor_mount=<A>; <PARAM>:sensor_orientation = <A>;</p> <p><PARAM>:sensor_data_start_date="2006-03-01T00:00:00Z"</p> <p><PARAM>:sensor_data_end_date="2007-03-01T00:00:00Z"</p> <p><PARAM>:sensor_data_file_DOI="http://doi.org/10.1594/PANGAEA.896648"</p>	<p>e.g.http://vocab.nerc.ac.uk/collection/P01/current/HCMR0021/</p> <p>QC_indicator: (OceanSITES specific) text, ref table 2 processing_level: text, ref table 3</p> <p>valid_min: Float. Minimum value for valid data valid_max: Float. Maximum value for valid data</p> <p>comment: Text; useful free-format text</p> <p>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.</p> <p>uncertainty: Float. Overall uncertainty of the observations estimated by a certain technique that considers accuracy, precision and other information for the time series as a whole and in one number. It is preferred to provide uncertainty for each data point (see Float <PARAM>_UNCERTAINTY . accuracy: Float. Nominal accuracy of data. precision: Float. Nominal precision of data. resolution: Float. Nominal resolution of data.</p> <p>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.</p> <p>DM_indicator: Text. Data mode, if constant, as per reference table 4. See note on data modes below.</p> <p>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</p> <p>sensor_*: Text. Use these fields to describe the sensor or instrument, unless the ancillary variable 'instrument' is used. See note on device metadata, below. sensor_SeaVoX_L22_code: from the SeaVoX codes; see Appendix I for the link to the vocabulary (JCOMMOPS) sensor_mount: Text. Deployment characteristics, from ref table 7. sensor_orientation: Text. Deployment characteristics, from ref table 8. sensor_data_start_date: Start date of the data, in UTC. See note on date/time attribute format below. (JCOMMOPS) sensor_data_end_date: End date of the data, in UTC. See note on date/time attribute format below. (JCOMMOPS) sensor_data_file_DOI: if a DOI for the sensor data exists it should be provided here. Use a comma-separated list if needed.</p>
---	---

Notes on data variables

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.

Sensor/instrument metadata:

Complete information about the instrument or the sensor should be provided by one of two methods, which are outlined in Appendix 2. Fields should include model name, manufacturer, serial number, the device code from the SeaVoX L22 vocabulary, 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 when different instruments measure the same data variable, and avoids repetition of information for instruments that measure multiple variables.

Date/time attribute format:

Format for date and time attributes: Use ISO 8601 combined date and time representations, and always specify UTC with the trailing ‘Z’ to avoid any confusion. "2007-04-05T14:30Z"

Uncertainty, Accuracy, Precision, 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 and is not to be confused with the sensor accuracy given by a manufacturer. 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; <PARAM>_QC:flag_meanings = "unknown good_data probably_good_data potentially_correctable_bad_data bad_data nominal_value interpolated_value missing_value"	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 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>"; <PARAM>_UNCERTAINTY:technique_title = "<Y>"; <PARAM>_UNCERTAINTY:technique_DOI = "<Y>";	Uncertainty of the data given in <PARAM>. long_name : type char. Required; fixed value _FillValue : type float. Required. units : type char. Required. Must be the same as <PARAM>:units. technique_title : type char. Optional. Title of the document that describes the technique that was applied to estimate the uncertainty of the data technique_DOI : type char. Optional. DOI of the document that describes the technique that was applied to estimate the uncertainty of the data

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"

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 as given by manufacturer" ;
TEMP_ uncertainty:technique_title = "How to process mooring data? A cookbook for MicroCat, ADCP and RCM
data"
TEMP_ uncertainty:technique_DOI = "DOI:10.13140/RG.2.1.2514.7044"
"

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

This is a partial list.

Data Assembly Centers and institutions	
BERGEN	University Of Bergen Geophysical Institute, NO
CCHDO	CLIVAR and Carbon Hydrographic Office, USA
CDIAC	Carbon Dioxide Information Analysis Center, USA
EUROSITES	EuroSites project, EU
GEOMAR	Helmholtz Centre for Ocean Research Kiel
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: Identifying data variables

When an appropriate CF standard name is available, it is required to be used in the `standard_name` attribute; if no such name exists in the CF standard, this attribute should not be used. In those cases, we have recommended terms to be used in the `long_name` attribute, as well as ‘short names’ or variable names. Standard names in the table below are in bold; recommended long names are not. Please refer to the CF Standard Names table on line for authoritative information (definitions, canonical units) on standard names.

It is recommended that variable names start with a code based on SeaDataNet parameter discovery vocabulary, P02. 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 must be 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	CF Standard name or suggested Long 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 its location to bottom.
upward	ADCP measuring currents towards the surface
horizontal	Optical sensor looking 'sideways' from mooring line or on a ships CTD frame

4. Data Files

4.1. Deployment Data Files

Deployment data files contain data from a single deployment of a platform, and normally hold one type of data; meteorology, salinity, or currents, for example. These files are stored in the directory /DATA/[SiteCode] on the OceanSITES GDAC servers.

4.1.1 Deployment Data files Naming Convention

The deployment data files are named using this convention:

OS_[PlatformCode]_[DeploymentCode]_[DataMode]_[PARTX].nc

- 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.

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.

4.2. Merged, Gridded and Derived Data Files

Building on the individual observations at an OceanSITES site or from an array of OceanSITES sites, a number of higher-level data products can be created:

- A “long time series” version that may simply concatenate multiple deployments into one data file/product for ease of use, but without significant changes to the content of the data from the individual deployment-by-deployment data. The concatenated data will in many cases combine observational data time series acquired at different instrument heights and as such further processing and homogenization may be required.
- A “gridded” version that presents time series created from observational data of single or multiple instruments but interpolated to a space-time grid different from the native instrumental resolution, e.g. by averaging or interpolating the data along one of the coordinate axes;
- “Derived” data products that are computed from the observational data, possibly from one or multiple sites and/or instruments, which contain parameters that are not directly observed, but rather involve some higher-order computations or models in their generation. Informative documented of the computations used, in either a technical document or a publication, is highly recommended by OceanSITES. .

These are not mutually exclusive, and the decision whether to declare a data product as one versus another option rests with the PIs. It is understood that when multiple data files are aggregated, the metadata attributes may not contain all the detailed information of each individual source data file; please refer to the deployment files for complete metadata.

Data in any of these higher-level files are duplicates of the deployment file data, and data aggregation processes should be careful to treat it as such.

4.2.1 Merged, Gridded and Derived Data File Metadata Conventions

OceanSITES recognizes that individual research projects generate different, if any, higher-level data products, and there is no particular set of metadata fields that covers all cases. Therefore, the format specifications of these higher-level data files are only loosely defined as follows:

- The file format for the higher-level data is netCDF. Each file is compliant with the following conventions:
 - CF metadata conventions: Standard names for data variables are required when available, and all other CF conventions should be used when possible.
 - Unidata Attribute Convention for Data Discovery (ACDD)
 - Additional metadata attributes from the deployment-by-deployment files (as specified earlier in this document) are possible and welcome, as long as they make sense for the data product in question.
- The files should contain a list in the metadata that explains what lower-level files they were derived from, including the version of the original data files. This can be done through the global attributes “history” or “comment”.
- For gridded data and derived products, data quality information (such as QC flags) is not strictly required. It is understood that during the gridding and processing, only good source data was used.
- Likewise, information on data mode (delayed-mode versus real-time) is not strictly required for gridded data or derived products. The default assumption is that the best available version was used, and that the metadata records provides a reasonable backtrace to the underlying source data and to the processes used.

4.2.2 Merged, Gridded or Product Data File Naming Conventions

The file names of the original deployment-by-deployment files are encoded as follows (see earlier sections of this document):

OS_[PlatformCode]_[DeploymentCode]_[DataMode]_[PARTX].nc

The file names for the higher-level products have a similar structure, with the following modifications:

- Instead of the [DataMode], a code is inserted that defines the type of data,
- Instead of the [DeploymentCode], a time range is used by default,
- For data from multiple platforms/sites, [PlatformCode] can be replaced with appropriate choices of site, project, array, or network, which are taken from the global attributes of the underlying source data.

The higher-level data files follow this naming convention:

OS_[PSPANCode]_[StartEndCode]_[ContentType]_[PARTX].nc

- OS - OceanSITES prefix
- [PSPANCode] - Deployment, platform, site, project, array, or network code from the underlying source data files. If all data are from one deployment of one platform, the platform and deployment code should be used. Else, move down the sequence terms until one is found that is unique and appropriate for all data in the file.
- [StartEndCode] - A code that describes the time range of the data in the file. Preferred format is e.g. “20050301-20190831” to indicate data from March 2005 through August 2019. Alternatively, if all data are from a single platform, a range of deployment codes can be used (e.g. “01-14” to indicate data from the first through the 14th deployment of this platform).
- [ContentType] - A three-letter code that describes the content of the file (distinguished from the deployment files, which have a one-letter code here), one of:
 - LTS: The data are “long time series” data that are essentially at the native instrumental resolution in space and time. The primary difference from the deployment-by-deployment files is that a single file contains merged data from multiple deployments.
 - GRD: The data are “gridded”, meaning that some sort of binning, averaging, interpolating has been done to format the data onto a space-time grid that is different from the native resolution, and more than a simple concatenation like the “LTS” option.
 - DPR: The data are a “derived product”, which means that there are data that were derived from multiple sites or some other higher-order processing that the data provider distinguishes from the lower-level data.
- [PARTX] - An optional user-defined field for additional identification or explanation of data. For gridded data, this could include the record interval as subfields of ISO 8601 (PnYnMnDTnHnMnS), e.g. P1M for monthly data, T30M for 30 minutes, T1H for hourly.

4.2.3 Higher-Level File Locations

The higher-level data files are found in the GDAC under the directory /DATA_GRIDDED/

To clarify that this includes the long time series and derived products, there are symbolic links called “long_timeseries” and “derived_products” that point back to this directory.

Inside the DATA_GRIDDED directories, there are subdirectories that contain the data files. The names of the subdirectories are either site, array, or network names as defined in the metadata of the underlying files. These do not provide a unique place for a given file; data providers work with the GDAC to identify the preferred location to be used.

5. OceanSITES Data Management topics

The GDACs distribute the best copy of the data files, which means that, if a higher quality

data file (e.g. improved calibrated data) is available, it replaces the previous version of the data file. The data file version is indicated in the netCDF fields.

OceanSITES does not archive data; archive will be implemented by the National Centers for Environmental Information (NCEI, formerly the National Ocean Data Center, or NODC) of the National Oceanic and Atmospheric Administration (NOAA) of the USA. Correct use of our documented data format specification is critical to the archive process.

5.1 Global Data Assembly Centers

Two global data assembly centers (GDACs) provide access points for OceanSITES data. One is in France at Coriolis, Ifremer (<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>

Deployment 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.

For information on uploading data, please see the Data Providers' Guide.

5.2 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_max
- 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
```

5.3 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, the SeaVoX L22 code for the instrument, 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 optical instruments, current meters and profilers.

Method 1 example:

```
variables:
double TEMP(TIME, DEPTH) ;
TEMP:instrument = "T_INST" ;
double PSAL(TIME, DEPTH) ;
PSAL:instrument = "T_INST" ;
int T_INST ;
T_INST:long_name = "instruments" ;
T_INST:ancillary_variables = "T_INST_MFGR T_INST_MOD T_INST_SeaVoX_L22_code
T_INST_SN T_INST_URL T_INST_MOUNT T_INST_CODE" ;
char T_INST_MFGR(DEPTH, strlen1) ;
T_INST_MFGR:long_name = "instrument manufacturer" ;
char T_INST_MODEL(DEPTH, strlen2) ;
```

```

T_INST_MODEL:long_name = "instrument model name" ;
char T_INST_SeaVoX_L22_code (DEPTH, strlen3) ;
T_INST_SeaVoX_L22_code:long_name = "SeaVox Vocabulary L22 code" ;
int T_INST_SN(DEPTH) ;
T_INST_SN:long_name = "instrument serial number" ;
char T_INST_URL(DEPTH, strlen3) ;
T_INST_URL:long_name = "instrument reference URL" ;
char T_INST_MOUNT(DEPTH, strlen3) ;
T_INST_MOUNT:long_name = "instrument mount" ;

```

data:

```

T_INST = _ ; (an empty variable, aka an umbrella)
T_INST_MFGR =
    "RBR-Global ",
    "Seabird Electronics",
    "Seabird Electronics" ;
T_INST_MODEL =
    " TR-1050",
    "SBE37 ",
    "SBE16 " ;
T_INST_SeaVoX_L22_code =
    "TOOL0055",
    "TOOL0018",
    "TOOL0023" ;
T_INST_MOUNT =
    "mounted_on_surface_buoy",
    "mounted_on_mooring_line",
    "mounted_on_seafloor_structure_riser" ;
T_INST_SN = 14875, 1325, 1328 ;
T_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_make = 'RBR-Global, Sea-Bird Scientific, Sea-Bird Scientific'
TEMP:sensor_SeaVoX_L22_code = "TOOL0055","TOOL0018", "TOOL0023";
TEMP:sensor_serial_number = 14875, 1325, 1328
TEMP:sensor_mount="mounted_on_surface_buoy, mounted_on_mooring_line, mounted_on_fixed benthic
node";
TEMP:sensor_orientation = "vertical";
double PSAL(TIME, DEPTH) ;
PSAL:sensor_name = 'RBR-Global TR1060, SBE23,SBE16'
PSAL:sensor_make = 'RBR-Global, Sea-Bird Scientific, Sea-Bird Scientific'
PSAL:sensor_SeaVoX_L22_code = "TOOL0055","TOOL0018", "TOOL0023";
PSAL:sensor_serial_number = 14875, 1325, 1328
PSAL:sensor_mount="mounted_on_surface_buoy, mounted_on_mooring_line, mounted_on_fixed benthic
node";
PSAL:sensor_orientation = "vertical";

```

6. Appendices

6.1 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: Unidata netCDF Attribute Convention for Dataset Discovery, at: http://wiki.esipfed.org/index.php/Category:Attribute_Conventions_Dataset_Discovery
- JCOMMOPS OceanSITES metadata portal <http://oceansites.jcommops.org/>
- ICES Ship Codes, used in platform_deployment_ship_ICES_code, platform_recovery_ship_ICES_code: <https://ocean.ices.dk/codes/ShipCodes.aspx>, or at https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/C17/
- The SeaVoX (SeaDataNet and MarineXML Vocabulary Content Governance Group) vocabularies, served at BODC, contain terms for some of our attributes:
 - Sensors and instruments: use the **L22** Device Catalogue https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/L22/
 - sea_area: use the **C19** Sea Areas vocabulary https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/C19/
 - source: use the SeaVoX Platform Categories vocabulary https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/L06/
- 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>
- NOAA-NCEI (formerly NODC) provides recommended netCDF Templates, available at <http://www.nodc.noaa.gov/data/formats/netcdf/>

6.2 Appendix 2: 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.

Project

A project within the OceanSITES program is a scientific research and observing effort. It may consist of a single platform at a single site, or may include multiple sites and platforms, led by one or more principal investigators.

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. Other prominent examples are OSNAP, RAPID or the TAO array.

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 OceansITES 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' and 'CIS-2' are surface buoys in the Central Irminger Sea, deployed concurrently. 'THETYS II' is 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 CIS-1 deployment performed in May 2009 (200905) is identified as OS_CIS-1_200905. CTD data from this deployment would be distributed in a file called OS_CIS-1_200905_R_CTD.nc.