PRODUCT USER MANUAL

For Global Ocean Reprocessed In Situ Observations of Carbon Product

INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050

Issue: 3.1

Contributors : Rocío Castaño-Primo, Benjamin Pfeil, Steve Jones

Approval Date : JUNE 2022

MERCATOR OCEAN
INTERNATIONAL
2 avenue de l'Aérodrome de Montaudran, 31400 Toulouse, FRANCE
Tél : +33 5 61 39 38 02 - Fax : +33 5 61 39 38 99
Société civile de droit français au capital de 2 000 000 € - 522 911 577 RCS Toulouse - SIRET 522 911 577 00016

marine.copernicus.eu
mercator-ocean.eu

RECORD TABLE

Issue	Date	§	Description of Change	Author	Validated By
1.0	2019/01/28	All	Creation of the document	B. Pfeil	
2.0	2019/11/25	All	Re-written text. Added information specific to the product generation and quality control.	R. Castaño-Primo B. Pfeil S. Jones	
3.0	2020/09/10	All	Changed and adapted to the new format and structure. Old Table 1: removed Figure 1: made the figure version-agnostic (e.g. SOCATvxxxx instead of SOCATv2019) Removed references to particular versions Added descriptive tables for the products and datasets. Removed “References” section Added headers from each of the datasets	R. Castaño-Primo	
3.1	2022/06/16	All	New PUM format. New product and dataset names. Changes in FTP folder structure (Figure 1), Table about platform families removed	R. Castaño-Primo	S. Tarot

Table of Contents

GLOSSARY AND ABBREVIATIONS

NetCDF	Network Common Data Form
CF	Climate and Forecast (convention for NetCDF)
PU	Production Unit
DU	Distribution Unit
FTP	File Transfer Protocol
EUROGOOS	European Global Ocean Observing System
ROOS	Regional Ocean Observing System
UiB	University of Bergen
SOCAT	Surface Ocean CO2 ATlas
GLODAP	GLobal Ocean Data Analysis Project
fCO_2	Fugacity of carbon dioxide
CARINA	CARbon dioxide IN the Atlantic ocean
WOCE/JGOFS	World Ocean Circulation Experiment / Joint Global Ocean Flux Study
GEOSECS	GEochemical Ocean SECtions Study

I INTRODUCTION

I.1 Summary

This Product User Manual describes the reprocessed biogeochemical product INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050 distributed by the Copernicus Marine Service In Situ Thematic Centre (INS-TAC): how it is built, what is the content, what data services are available to access them, and how to use the files

The INS-TAC is a distributed system built on the existing activities and services developed within EU-funded projects and support from EuroGOOS Regional alliances. It provides a research and operational framework to develop and deliver in situ observations and derived products to address progressively global and regional needs for monitoring, modeling and downstream service development.

The main objective of the Copernicus Marine Service is to deliver and operate a rigorous, robust and sustainable Ocean Monitoring and Forecasting system to users of all marine applications: maritime safety, marine resources, marine and coastal environment and climate, seasonal and weather forecasting. The INS-TAC prepares reprocessed datasets for reanalysis activities performed by the Copernicus Marine Service forecasting centers (MFCs) and external users.

The carbon REP product described in this manual are the data products SOCAT and GLODAP, reformatted to facilitate the access to Copernicus Marine Service INSTAC users. SOCAT and GLODAP are voluntary efforts of the biogeochemical scientific community that provide harmonized, high-quality data necessary to evaluate and understand the inorganic carbon cycle in the oceans. SOCAT and GLODAP are endorsed by the Global Ocean Observing System (GOOS) and are voluntary contributions to the UN Sustainable Development Goal 14.3: *Life Under Water; minimize and address the impacts of ocean acidification, including through enhanced scientific cooperation at all levels*.

SOCAT (https://socat.info), the Surface Ocean CO2 ATlas, is a synthesis activity for quality-controlled surface ocean fCO_2 measurements by more than 100 contributors from around the globe. It enables quantification of the ocean carbon sink, ocean acidification and the evaluation of ocean biogeochemical models. As such, it is used in the Global Carbon Budget (Friedlingstein *et al.* 2022). Version 1 was released in 2011. Automation has allowed for annual releases since version 4 in 2016.

GLODAP (https://glodap.info) is a compilation of ocean interior observations of key carbon and biogeochemical variables, harmonized, quality controlled and bias-corrected. GLODAP data have been used for model evaluation, inventory calculations and calibration of BGC-Argo sensor observations. The first version was released in 2005, with cruises from the WOCE/JGOFS and GEOSECS programs. The second version included those in the first one and

data from the CARINA and PACIFICA synthesis. The subsequent versions GLODAPv2.xxxx are extensions of GLODAPv2.

The following articles and documents describe in depth the different data products and how they are generated: Bakker *et al.* (2016); Lauvset *et al.* (2018); Lauvset *et al* (2021); Lauvset *et al.* (2016). The reprocessing system is described in the Quality Information Document (QUID), and the references above

https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-INS-QUID-013-046.pdf.

Information on operational issues on products and services can be found on our User Notification Service. If you have any questions, please contact us.

I.2 History of changes

Date	Description of changes and impacted product
2019/04	Release of carbon REP product INSITU_GLO_CARBON_REP_OBSERVATIONS_013_050 with SOCATv6 and GLODAPv2
2020/03	Update with SOCATv2019 and GLODAPv2.2019
2020/11	Update with SOCATv2020 and GLODAP2.2020
2021/12	Update with SOCATv2021 and GLODAP2.2021
2022/11	Product renamed to INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050. Datasets renamed to: cmems_obs-ins_glo_bgc-car_my_socat-obs_irr, cmems_obs-ins_glo_bgc-car_my_socat-gridded_irr, cmems_obs-ins_glo_bgc-car_my_glodap-gridded_irr, cmems_obs-ins_glo_bgc-car_my_glodap-obs_irr Update with SOCATv2022 and GLODAPv2.2022.

II DESCRIPTION OF THE PRODUCT SPECIFICATION

II.1 General Information

Product name	INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050
Geographical coverage	Global
Variables	Fugacity of CO_2 Sea temperature Salinity Dissolved oxygen Chlorophyll-a Nitrate Nitrite Phosphate Silicate Dissolved Inorganic Carbon Total alkalinity pH in situ pH at 25°C 0dbar Total dissolved nitrogen Dissolved organic nitrogen Dissolved organic carbon Bottom depth Number of cruises Number of fCO_2 / temperature / salinity observations Average / max/min/STD fCO_2 unweighted / per cruise weighted Average / max/min/STD sea surface temperature unweighted / per cruise weighted Average / max/min/STD salinity unweighted / per cruise weighted

	Latitude/Longitude offset Number of data in bins Bin averaged input data Standard deviation of bin averaged input data Signal to noise ratio Correlation length Nitrate/phosphate/dissolved inorganic carbon/oxygen/pH at 25°C 0dbar/pH in situ/ salinity/silicate/temperature Nitrate/phosphate/dissolved inorganic carbon/oxygen/pH at 25°C 0dbar/pH in situ/ salinity/silicate/temperature error and relative error
Multi-Year/ Update Frequency	Yearly (SOCAT). Irregular (GLODAP).
Available time series	Since 1957 (SOCAT) / 1972 (GLODAP)
Temporal resolution	Variable (SOCAT and GLODAP OBSERVATIONS) Month, year and decade (SOCAT GRIDDED)
Horizontal resolution	1° (SOCAT and GLODAP GRIDDED) and 0.25° (SOCAT GRIDDED coastal)
Number of vertical levels	Surface / 1 (SOCAT OBSERVATIONS / GRIDDED). Full water column / 33 (GLODAP OBSERVATIONS / GRIDDED)
Format	NetCDF-4 classic model
Delivery mechanisms	FTP

Table 1: INSITU_GLO_BGC_CARBON products

The runtime schedule:

The data products SOCAT and GLODAP are community, voluntary efforts. They have made voluntary commitment to yearly releases. GLODAP gridded climatologies are the exception; they are updated at irregular intervals; the latest version was released in 2016, using the data from GLODAPv2 (see Lauvset *et al.*, 2016).

II.2 Details of datasets

For the gridded fields datasets, the standard_name does not apply

cmems_obs-ins_glo_bgc-car_my_socat-obs_irr
FCO2 [µatm] CO_2 fugacity fugacity_of_carbon_dioxide_in_sea_water
TEMP [°C] Sea temperature sea_water_temperature
PSAL [psu] Practical salinity sea_water_practical_salinity
cmems_obs-ins_glo_bgc-car_my_socat-gridded_irr
count_ncruise Number of cruises
fco2/sst/salinity_count_nobs Number of fCO_2 / temperature / salinity observations
fco2_ave/max/min/std_unwtd/weighted Average / max/min/STD fCO_2 unweighted / per cruise weighted
sst_ave/max/min/std_unwtd/weighted Average / max/min/STD sea surface temperature unweighted / per cruise weighted
salinity_ave/max/min/std_unwtd/weighted Average / max/min/STD salinity unweighted / per cruise weighted
lat/lon_offset_unwtd Latitude/Longitude average offset from cell center
cmems_obs-ins_glo_bgc-car_my_glodap-obs_irr
TEMP [°C] Sea temperature sea_water_temperature
PSAL [psu] Practical salinity sea_water_practical_salinity
DOX2 [µmol kg-1] Dissolved oxygen moles_of_oxygen_per_unit_mass_in_sea_water
CPHL [mg m-3] Chlorophyll-a mass_concentration of_chlorophyll_a_in_sea_water
NTAW [µmol kg-1] Nitrate (NO3-N) moles_of_nitrate_per_unit_mass_in_sea_water

NTIW [µmol kg-1] Nitrate (NO2-N) moles_of_nitrite_per_unit_mass_in_sea_water
PHOW [µmol kg-1] Phosphate (PO4-P) moles_of_phosphate_per_unit_mass_in_sea_water
SLCW [µmol kg-1] Silicate (SIO4-SI) moles_of_silicate_per_unit_mass_in_sea_water
TICW [µmol kg-1] Dissolved inorganic carbon
ALKW [µmol kg-1] Total alkalinity sea_water_alkalinity_per_unit_mass
PHPH pH sea_water_ph_reported_on_total_scale
PH25 pH at 25 degrees and 0 dbar
NT1D [µmol kg-1] Total dissolved nitrogen moles_of_dissolved_total_nitrogen_per_unit_mass_in_sea_water
NODW [µmol kg-1] Dissolved organic nitrogen
CORG [µmol kg-1] Dissolved organic carbon moles_of_dissolved_organic_carbon_per_unit_mass_in_sea_water
BATH [m] Bathymetric depth sea_floor_depth_below_mean_sea_surface
cmems_obs-ins_glo_bgc-car_my_glodap-gridded_irr
Input_N/mean/std Number of data in bins/bin-averaged input data/standard deviation of bin averaged input data
SnR Signal to noise ratio
CL Correlation length
NO3/PO4/TCO2/Talk/oxygen/pHts25p0/pHtsinsitutp/salinity/silicate/temperature *Unit* of *variable* in seawater
NO3/PO4/TCO2/Talk/oxygen/pHts25p0/pHtsinsitutp/salinity/silicate/temperature_error *Variable* error

NO3/PO4/TCO2/Talk/oxygen/pHts25p0/pHtsinsitutp/salinity/silicate/temperature_relerr *Variable* relative error

Table 2: list of the datasets and variable names and unit for the INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050 product

II.3 Production System Description

Within the In Situ TAC organization, the carbon Production Unit at the University of Bergen functions currently as a global PU independent from the global PU managed by IFREMER. Due to the nature of the REP carbon products as already-existing data products, they are distributed as stand-alone, independent datasets, with proper attribution given in the metadata by use of DOIs.

SOCAT relies on the regular contributions of voluntary data producers. Surface CO_2 data and metadata are submitted and the automatic upload system performs a first automatic QC of the data (see the QuID for further details). Later, a thorough manual QC is performed by a group of scientists with expertise within a particular region (https://www.socat.info/index.php/regional-groups/). The QC procedures include checks on metadata to ensure completeness and a proper evaluation of data accuracy. Once the synthesis product is published, once a year, the INSTAC carbon PU at UiB reformats SOCAT to make it coherent with the rest of the INSTAC products, and submits to the DU, with metadata and proper attributions.

GLODAP contains bottle data from different data centres (CCHDO, NCEI, Pangaea) or receives data submitted directly by individual contributors. The data is converted to WOCE exchange format for standardization, and then quality controlled at two levels (see the QuID for further details) by scientists specialized in biogeochemistry (https://www.glodap.info/index.php/group/). Once the synthesis product is finalised, it's published, together with a cruise table that summarizes some of the metadata and with links to the original files. The INSTAC carbon PU at UiB reformats the synthesis files to make it coherent with the rest of the INSTAC products, and submits to the DU, with metadata and proper attributions.

SOCAT provides observations of fCO_2, temperature and salinity. However, the user is advised to use only fCO_2 for research purposes. Temperature and salinity are not explicitly manually quality controlled, and are provided for provenance. In GLODAP, temperature and pressure observations are assumed "good" if they exist.

The gridded datasets are generated by the SOCAT and GLODAP communities, not by the carbon PU. The INSTAC carbon PU provides additional metadata pertaining Copernicus Marine Service INSTAC only.

The SOCAT gridded fields utilize data from 1970, since data before that is scarce. There is no interpolation used in these datasets, only binning and averaging. Additional diagnostic information per cell is also included: number of cruises, total number of observations,

minimum, maximum, standard deviation and the offset between the centre of the observations and the centre of the cell. Two types of fCO_2 values are calculated: unweighted and cruise-weighted means.

GLODAP provides climatology gridded files, generated using the interpolation method DIVA (Data Interpolating Variational Analysis), using data from the period 1972-2013. One climatology is generated per variable: nitrate (NO3), phosphate (PO4), dissolved total inorganic carbon (TCO2), total alkalinity (Talk), oxygen, pH at 25°C and 0 dbar (pHts25p0), pH at in situ temperature and pressure (pHtsinsitutp), salinity, silicate and temperature. Ancillary information per cell includes the number of observations, mean, standard deviation, mapping error and mapping error scaled to the standard deviation of the global input data at that depth level.

II.4 Grid

All gridded files are on a 1x1 degree grid, except for the monthly coastal file of SOCAT, which is on a quarter-degree grid and only data within 400 km of a significant land mass are included.

II.5 Vertical Levels

SOCAT files, both observations and gridded, are at one single depth level, since they are all surface observations.

The DEPTH dimension values in the GLODAP observations files are located at the closest nominal depth. The nominal depths are at 5, 25 and 100 m intervals down to 100, 1000 and 10000 m depth. The GLODAP gridded climatologies have 33 vertical levels (0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2500, 3000, 3500, 4000, 4500, 5000, and 5500 m).

II.6 Processing information

II.6.1 Update Time

The originating community data products (SOCAT and GLODAP) and therefore INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050 product are updated once a year.

II.6.2 Temporal extent

SOCAT observations files extend from 1957 to the first weeks of the year of the release (for example, for v2022, until the first weeks of 2022). The gridded files start at 1970, since data is scarce prior to that date.

GLODAP data start in 1972 and extend until a year-two years prior to the time of release, depending on the processing and availability of the cruises' samples.

II.6.3 Time averaging

Only the gridded datasets cmems_obs-ins_glo_bgc-car_my_socat-gridded_irr and cmems_obs-ins_glo_bgc-car_my_glodap-gridded_irr are averaged over different time periods.

SOCAT provides four gridded files at different time scales: one decadal, one yearly and two monthly averaged files (one global, one, focused on coastal areas). The monthly gridded global dataset is derived by combining all SOCAT data collected within a 1x1 degree box during a specific month and year. The year and decade files are generated by averaging into yearly and decadal bins, and recalculating the diagnostics.

GLODAP provides climatological averages from 1972 to 2013.

III DOWNLOAD A PRODUCT

After registration, you will be able to download our data. To assist you, our HelpCenter is available, and more specifically its section about download.

Information on operational issues on products and services can be found on our User Notification Service. If you have any questions, please contact us.

III.1 Download a product through the Copernicus Marine Service Web Portal Subsetter Service

Copernicus Marine Service Web Portal provides ERDDAP subsetting services for some in situ datasets. The In Situ TAC is working on providing an operational API to enable this in the near future. However, the current setup is still in test phase and therefore, not available yet for users. To facilitate the discovery and downloading processes, In Situ component has developed a dashboard to select, download and visualise the data. It can be accessed from https://marine.copernicus.eu/access-data OCEAN IN SITU.

III.2 Download a product through the Copernicus Marine Service Web Portal FTP Service

Coperncius Marine Service Web Portal FTP service is the one guaranteeing the access to the 100% of the products addressed in this document. This is a secured FTP and therefore users need to request credentials to access it. In here you can find the form to fill in in order to get those: https://resources.marine.copernicus.eu/?option=com_sla..

FTP can be issued also interactively via the In Situ TAC Dashboard viewer interface, available at: http://www.marineinsitu.eu/dashboard/. This viewer is nevertheless subject to periodic maintenance works so that it is better encouraged to issue FTP services programmatically when possible.

III.2.1 FTP portal structure

Due to the nature of the carbon REP product, its FTP organization is slightly different from other REP products. The FTP portal is organized in one product directory, with one dataset directory, see Figure 1. In the dataset directory, there are:

- 4 data directories: SOCAT/GLODAP obs, SOCAT/GLODAP gridded, with subdirectories for each data type (
- 4 index files that enumerate and describe the files in the 4 directories
- 2 index files describing the platforms that generate the data in SOCAT obs and GLODAP obs

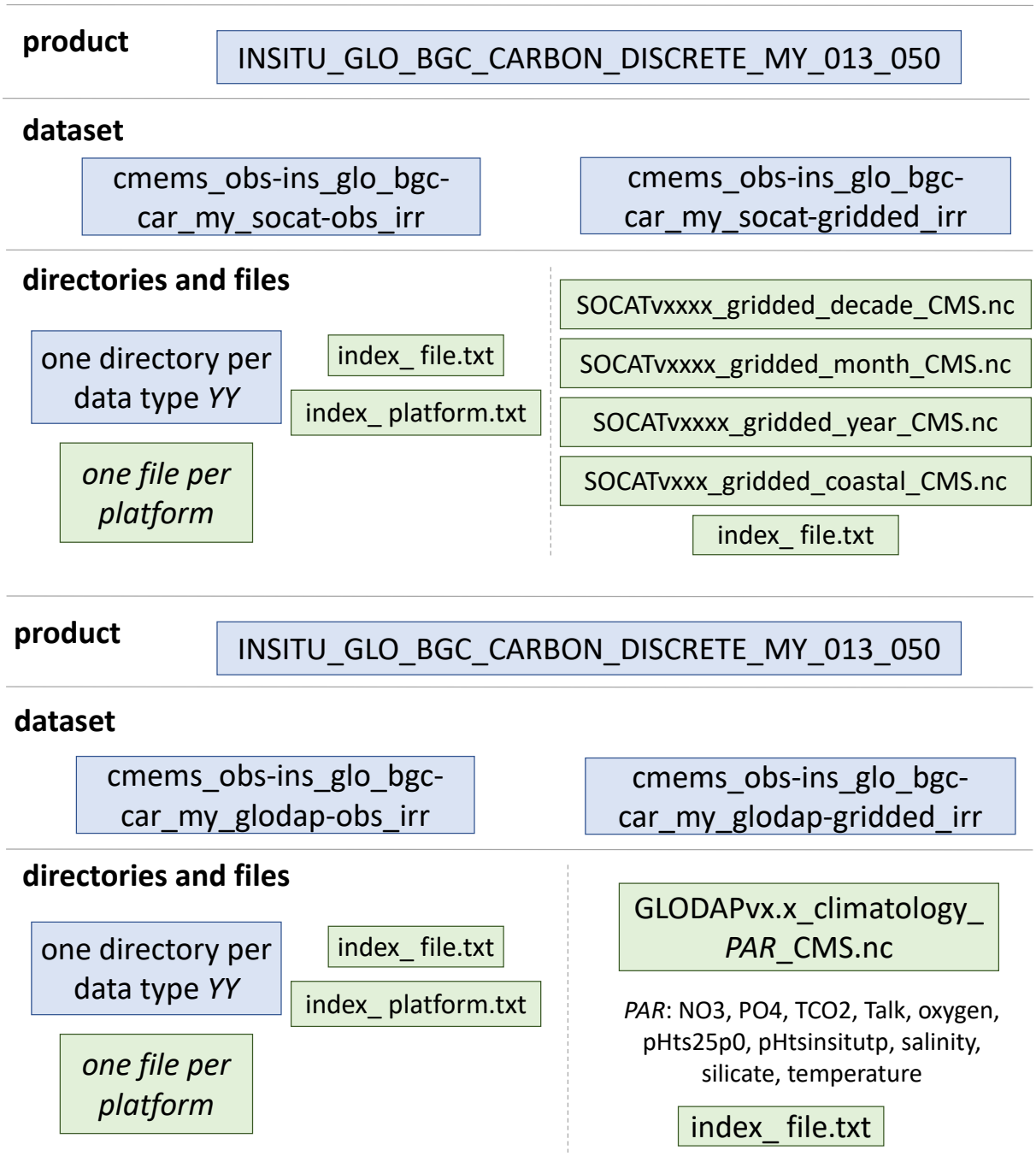

Figure 1: Distribution Unit (DU) FTP server organization

FILES NOMENCLATURE

III.3 Nomenclature of files

The cmems_obs-ins_glo_bgc-car_my_*-obs_irr directories contain the observations from SOCAT / GLODAP, distributed in folders by *data type* (see Table 4), and one file per individual platform. The file naming convention is similar to that of other In Situ TAC products: GL_*FT*_*YY*_*CODE*-*dataproductversion*.nc:

- GL: region bigram (global).
- *FT*: file type. See Table 3. SOCAT provides only trajectory or time series (TS) type of files; GLODAP provides only profiles (PR)
- *YY*: data type. See Table 4. GLODAP provides only BO data.
- *CODE*: platform code. See section III.5.1.
- *dataproductversion*: SOCATvxxx or GLODAPvxxxxx, respectively

File type	Description
TS	time series, trajectories
PR	profiles

Table 3: File types

Data type	Description
BO	bottle samples
CO	autonomous underway pCO_2 data
DB	drifting buoys
GL	glider
MO	fixed buoys, mooring time series, fixed observations
SD	Saildrone

Table 4: Data types

The cmems_obs-ins_glo_bgc-car_my_*-gridded_irr directories contain gridded fields created from the observation datasets, at different time scales, and the file nomenclature is different.

The cmems_obs-ins_glo_bgc-car_my_socat-gridded_irr directory contains 4 gridded datasets: 3 datasets at 1-degree spatial resolution at different time scales (per month, year and decade), and the quarter-degree coastal dataset. The naming convention is particular to this dataset: SOCATvxxxx_gridded_*scale*_CMEMS.nc. Values of *scale* can be month, year, decade or coastal.

The cmems_obs-ins_glo_bgc-car_my_glodap-gridded_irr directory contains 7 files; one climatology dataset per parameter. The naming convention is particular to this dataset: GLODAPv2.1_climatology_*PAR*_CMEMS.nc. The values of *PAR* can be: NO3, PO4, TCO2, Talk, oxygen, pHts25p0, pHtsinsitutp, salinity, silicate or temperature.

III.4 Index files (data)

Each obs directory has an index file associated with it. The index files are updated to describe all the files available in the directory. They contain a header with general information about the product and date of update, and one line per file, ordered alphabetically and by start of the time coverage. Dates and times are in ISO8601 format and position in decimal degrees north/east. They include the following information, comma-separated, about each file:

- catalog_id: COP-GL-01
- file_name: with full path to the file in the ftp server
- geospatial_lat_min
- geospatial_lat_max
- geospatial_lon_min
- geospatial_lon_max
- time_coverage_start
- time_coverage_end
- institution: name, as stated in the EDMO catalogue. Separated by semicolon
- date_update
- data_mode: D (delayed-mode data)
- parameters: list of all the parameters regardless of the quality of data, except *PARAM*_DM, *PARAM*_QC, TIME, LATITUDE, LONGITUDE. Separated by blanks.

Users are advised to use these index files to select the files they are interested in. For example, if a user is interested in retrieving the files that provide data in a specific area, he

just has to select the lines for which the latitude and longitude of interest are included in the box [geospatial_lat_min, geospatial_lon_min: geospatial_lat_max, geospatial_lon_max], store the complete file name in a list and then perform a get of all the files of the list.

An example of an index file:

```
# Title : in-situ files catalog
# Description : catalog of available in-situ files
# Project : Copernicus Marine In Situ TAC
# Format version : 1.4
# Date of update : YYYYMMDDhhmmss
#
catalog_id,file_name,geospatial_lat_min,geospatial_lat_max,geospatial
_lon_min,geospatial_lon_max,time_coverage_start,time_coverage_end,ins
titution,date_update,data_mode,parameters
COP-GL-01,ftp://my.cmems-
du.eu/Core/INSITU_GLO_BGC_CARBON_DISCRETE_MY_013_050/cmems_obs-
ins_glo_bgc-car_my_socat-
obs_irr/VESSEL/GL_TS_TS_DBKV_SOCATv2022.nc,16.9002,64.38,328.33,350.7
6,1991-06-16T00:00:00Z,2010-06-23T04:10:38Z,University of Bergen
Geophysical Institute,2019-03-26T00:00:00Z,D,DEPH FCO2 PSAL TEMP
```

The index files for the gridded files directories are provided to give a quick overview of the files. Spatial and temporal extension are not relevant information, but are kept nonetheless to maintain uniformity through the INSTAC products.

III.5 Platform index files

Only available in the obs directories, index_platform is a list of the individual platforms that are available on the server, updated daily. It is ordered alphabetically by platform_code. It contains a header with some general information and one line per platform with the following information, comma-separated:

- platform_code
- creation_date
- update_date
- wmo_platform_code
- data_source: name of the data streams; i.e. the generic file names with the date info replaced by “X”s. Separated by blanks.
- institution: name of the institution(s) that provide data to the platform. Within quotes and separated by “/”.
- institution_edmo_code: separated by blanks.

- parameter: list of all the parameters regardless of the quality of data, except *PARAM*_DM, *PARAM*_QC, TIME, LATITUDE, LONGITUDE. Separated by blanks.
- last_latitude_observation
- last_longitude_observation
- last_date_observation

An example of an index platform file:

```
# Title : in-situ platforms catalog
# Description : catalog of available in-situ platforms.
# Project : Copernicus Marine In Situ TAC
# Format version : 1.4
# Date of update : YYYYMMDDhhmmss
#
platform_code,date_creation,date_update,wmo_platform_code,data_source
,institution,institution_edmo_code,parameters,last_latitude_observati
on,last_longitude_observation,last_date_
observation
00000,2013-02-10T18:57:49Z,2013-06-
24T16:32:11Z,00000,GL_XXXXXX_PR_TE_00000,"Unknown
institution",,DC_REFERENCE       PRES       TEMP,55.15,-42.01,1999-02-
26T13:08:00Z
```

III.5.1 Platform code convention

An individual platform may have different ways of identification, with different codes. The NetCDF global attributes can provide this information:

- platform_code: The global attribute "platform_code" is mandatory in INSTAC NetCDF data files. It is unique within all Copernicus Marine Service In Situ TAC portals. What constitutes the platform code varies per type of platform:
  - vessels: their maritime call sign, assigned by the ITU via national authorities.
  - moorings, drifting buoys, autonomous vehicles: their WMO number, if they have one. If they don't (e.g. they don't transmit to GTS), the platform name.
- platform_name: not mandatory. The human-readable name of the platform.
- wmo_platform_code: not mandatory.
- ices_platform_code: not mandatory. ICES (International Council for the Exploration of the Seas) assigns codes that can be searched by platform name in the following link: (https://vocab.ices.dk/Request/Login.aspx?ReturnUrl=%2fRequest%2f).

If the non-mandatory attributes do not exist, the global attribute will not appear in the NetCDF file. If they exist, but are unknown, they will be left empty.

IV FILE FORMAT

IV.1 NetCDF

The products are stored using the NetCDF format.

NetCDF (network Common Data Form) is an interface for array-oriented data access and a library that provides an implementation of the interface. The netCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data. The netCDF software was developed at the Unidata Program Center in Boulder, Colorado. The netCDF libraries define a machine-independent format for representing scientific data.

Please see Unidata netCDF pages (http://www.unidata.ucar.edu/software/netcdf/) for more information, and to retrieve netCDF software package.

NetCDF data format is:

* Self-Describing. A netCDF file includes information about the data it contains.

* Architecture-independent. A netCDF file is represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.

* Direct-access. A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.

* Appendable. Data can be appended to a netCDF dataset along one dimension without copying the dataset or redefining its structure. The structure of a netCDF dataset can be changed, though this sometimes causes the dataset to be copied.

* Shareable. One writer and multiple readers may simultaneously access the same netCDF file.

For the OBS files, data and metadata follow the NetCDF CF convention implementation specified in “Copernicus Marine in situ NetCDF format reference manual” https://doi.org/10.13155/59938.

The gridded fields are already provided by the sources (SOCAT/GLODAP) in NetCDF format. They are richly described through variable and global attributes and additional INSTAC attributes are added. Their file formats is netCDF-4 instead of netCDF-4 classic model.

IV.2 Reader Software

NetCDF data can be browsed and used through a number of software packages, including:

- ✓ ncBrowse: http://www.epic.noaa.gov/java/ncBrowse/,
- ✓ NetCDF Operator (NCO): http://nco.sourceforge.net/
- ✓ IDL, Matlab, Panoply, GMT…

IV.3 File format checker

Every Copernicus In Situ TAC data file submitted by a PU (Production Unit) for distribution has its format and data consistency checked by the Copernicus file format checker. It ensures the file formats match the Copernicus In Situ TAC standards precisely, as specified in the SRD.

Files with format or consistency errors are rejected by the PU and are not distributed. Less serious problems may generate warnings and the file will still be distributed on the PU.

The file format checker is publicly available on:

Carval Thierry, Rannou Jean-Philippe, Fontaine Laure (2017). NetCDF file format checker for Argo floats, Copernicus In Situ TAC, EGO gliders, OceanSITES. SEANOE. http://doi.org/10.17882/45538

The reference document for file format checks is available on:

Copernicus Marine In Situ TAC Data Management Team (2016). Copernicus In Situ TAC - Copernicus Marine Service System Requirements Document. http://doi.org/10.13155/40846

V REFERENCES

- SOCAT data product description

Bakker, D. C. E., Pfeil, B. Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., Van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., Xu, S. (**2016**). A multi-decade record of high quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT). Earth System Science Data, 8, 383-413. doi: 10.5194/essd-8-383-2016.

- SOCAT Quality Control cookbook

Lauvset, S., Currie, K., Metzl. N., Nakaoka, S., Bakker, D., Sullivan, K., Sutton, A., O'Brien, K., Olsen, A. (**2018**) SOCAT Quality Control Cookbook -For SOCAT version 7. https://www.socat.info/wp-content/uploads/2019/01/2018_SOCAT_QC_Cookbook_for_SOCAT_Version_7.pdf

- SOCAT v2022 release poster

https://www.socat.info/wp-content/uploads/2022/06/2022_Poster_SOCATv2022_release.pdf

- GLODAPv2.2021 synthesis product description

Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., Álvarez, M., Becker, S., Brown, P. J., Carter, B. R., Cotrim da Cunha, L., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Jones, S. D., Karlsen, M. K., Lo Monaco, C., Michaelis, P., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R., Woosley, R. J., and Key, R. M. (**2021**): An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2021, Earth System Science Data, 13, 5565–5589, https://doi.org/10.5194/essd-13-5565-2021

- GLODAP gridded climatologies description

Lauvset, S. K, R. M. Key, A. Olsen, S. van Heuven, A. Velo, X. Lin, C. Schirnick, A. Kozyr, T. Tanhua, M. Hoppema, S. Jutterström, R. Steinfeldt, E. Jeansson, M. Ishii, F. F. Pérez, T. Suzuki

and S. Watelet. (**2016**). A new global interior ocean mapped climatology: the 1°x1° GLODAP version 2, Earth System Science Data, 8, 325–340, doi: 10.5194/essd-8-325-2016

- Global Carbon Budget 2021

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L. M., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G. R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J. (**2022**): Global Carbon Budget 2021, Earth System Science Data, 14, 1917–2005, https://doi.org/10.5194/essd-14-1917-2022.

Quality Information Document (QUID): https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-INS-QUID-013-050.pdf