Ifremer

Catalogue of data and platforms at Network GDAC level, including the example of Copernicus In Situ TAC

Table of content

1. Catalogue of data and platforms at Network GDAC level

1.1. Data file driven catalogues

This document presents catalogue techniques used at network GDAC level to facilitate the discovery of platforms and data files.

Some networks are organized as DAC-GDACs that continuously update a catalogue of metadata on observation datasets and platforms:

- A DAC is a Data Assembly Centre operating at national or regional scale. It manages data and metadata for its area with a direct link to Scientifics and Operators. The DAC pushes observations to the network GDAC.
- A GDAC is a Global Data Assembly Centre. It is designed for a global observation network such as Argo, OceanSITES, DBCP, EGO, Gosud, etc… The GDAC aggregates data and metadata of an observation network, in real-time and delayed mode, provided by DACs.

1.2. Example of Copernicus in situ TAC catalogue facilities

Copernicus In Situ TAC (Thematic Assembly Centre) aggregates data and metadata from various networks. It manages a catalogue of data and platforms available data on Copernicus Distribution Units.

The index files of this catalogue are described in chapter 2 of this document. They contain metadata directly extracted from the files (NetCDF global attributes)
1. An index file (basic catalogue), being a simple ASCII file, contains all the relevant metadata to describe each individual file :one line per file described
2. An aggregation on platforms is continuously performed : one line per platform described

At TAC level, the dataset and platform catalogues are continuously updated with new file arrival/update.

The Copernicus in situ TAC catalogue describes more 100 000 data files and 30000 platforms.

1.2.1. Data and metadata portals using Copernicus In Situ catalogues - index files

Data and metadata access applications can efficiently use Copernicus In Situ simple and robust index files to manage sophisticated portals.

Examples of portals using Copernicus index files
- Arctic ROOS http://webprod1.nodc.no/arctic-roos/arctic-roos.html
- IBI Roos
  - http://ibidataportal.puertos.es/
  - http://www.ibiexplorer.eu/
- Mongoos: http://oceanobs.mongoos.eu/
- EMODNET Physics : http://www.emodnet-physics.eu/Map/

1.2.2. Key Performance Indicators calculated from Copernicus In Situ catalogues - index files

Such index files can also be used to generate a series of indicators as it is done periodically from Copernicus In Situ TAC index file or geographical positions contained in the index files can easily be displayed in various geographic maps. See examples hereafter.

Statistics calculated from index files can

Data discovery hit map

Every day, the Copernicus In Situ data discovery hit map is updated with the content of DU index_latest.txt file.
- http://www.ifremer.fr/co/co05010507/home.htm

It displays on a map all the observations available for the last 30 days from Copernicus In Situ Distribution Unit.
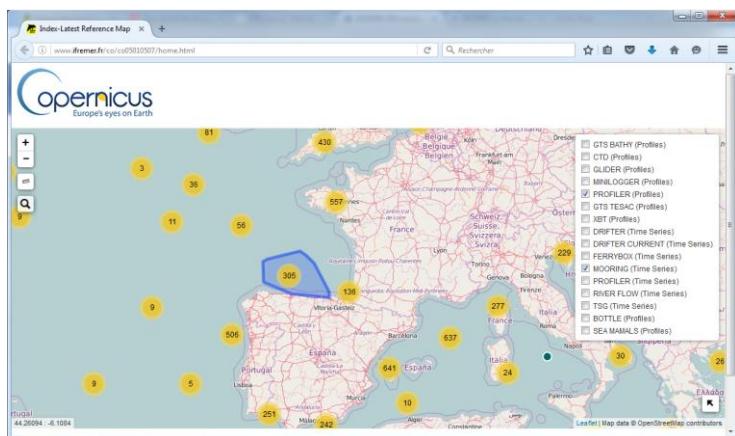

The last 30 days of observations hit map

Platform last position and metadata

Every day, the last position and metadata of the 30 000 platforms available on Copernicus In Situ Distribution Unit are updated from the platform index file.
- http://www.ifremer.fr/co/co05010507/platform.html
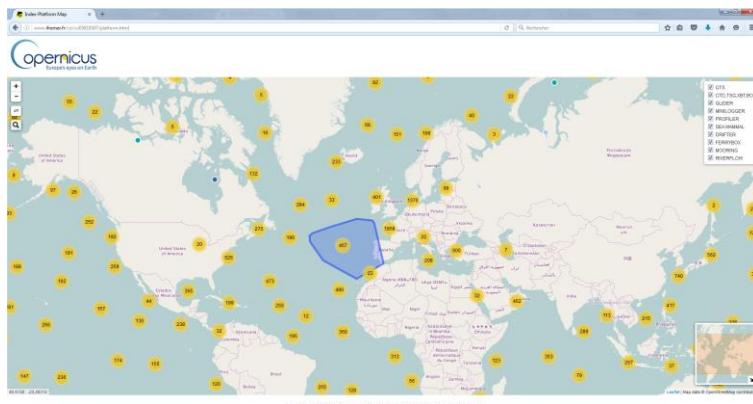
The last position of the 30 000 Copernicus In Situ platforms

Statistics

A series of statistics on Copernicus index files is continuously updated, they provide Key Performance Indicators (KPIs), to monitor the content of the data and metadata.
- http://www.ifremer.fr/co/co05010507/KPI/
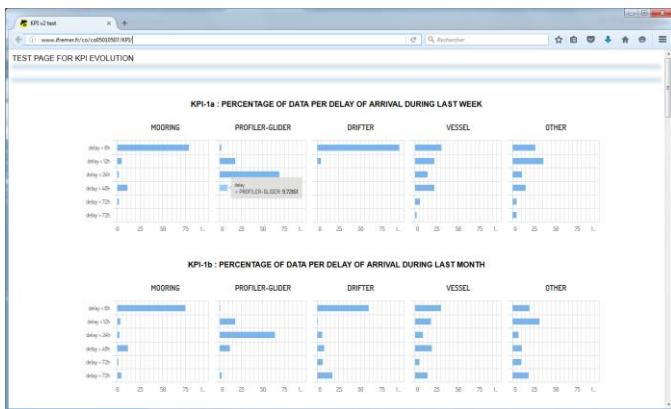
Delay of arrival per platform type (date of availability minus date of observation)
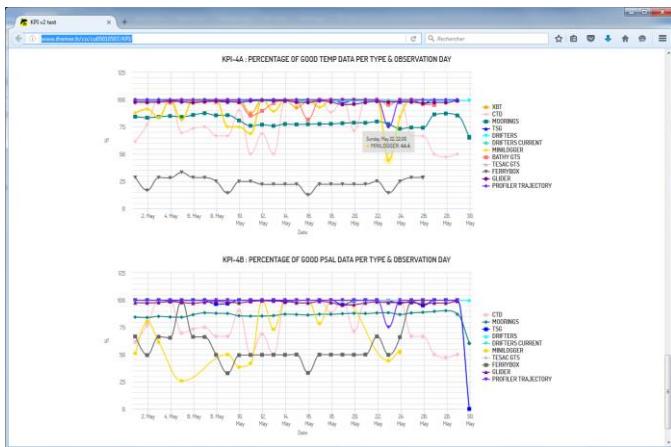
Percentage of good temperature and salinity data per platform type

2. Description of Copernicus In Situ data and platform index files

This chapter is an extraction of the chapter §6.3 "Description of data files index" of Copernicus in situ TAC System Requirements Document http://dx.doi.org/10.13155/40846

2.1. Description of data files index

2.1.1. Index latest/monthly/history/reference_data

The index files are updated to describe all the files available in the latest, the monthly, the history and the reference_data directories.
Each index file contains one line per latest, monthly, history or reference_data file:

- product_id
  COP-XX-YY

  - COP: Copernicus trigram
  - XX: region bigram (see table 2)
  - YY: product version
    Example: COP-NO-01
- file_name
- geospatial_lat_min
- geospatial_lat_max
- geospatial_lon_min
- geospatial_lon_max
- time_coverage_start
- time_coverage_end
- provider
- date_update
- data_mode
  - R: real-time data
  - D: delayed-mode data
  - M: mixed real-time and delayed mode data
- parameters (separator: blank)

Index names
- index_latest.txt
- index_monthly.txt
- index_history.txt
- index_reference_data.txt

Index lines order
The index lines are sorted by file name and time coverage start.

Separator character ","
The fields separator character is "," (comma).
In the "provider" field, each "," (comma) character is replaced by "-" (minus). The "," comma character must not appear within a field.

Date format
OceanSITES format: YYYY-MM-DDThh:mm:ssZ

Copernicus In Situ data file index example
Title : In Situ observations catalog
Description : catalog of available In Situ observations per platform.
Project : Copernicus

```
# Format version : 1.0
# Date of update : 20091105194820
#
product_id,file_name,geospatial_lat_min,geospatial_lat_max,geospatial_lon_min,geospatial_lon_max,time_
coverage_start,time_coverage_end,provider,date_update,data_mode,parameters
MYO-IBIROOS-
01,ftp://ftp.ifremer.fr/ifremer/ibiroos/monthly/drifter/IR_200910_PR_BA_10004.nc,54.17,54.17,6.35,6.35,2
009-10-01T00:43:00Z,2009-10-07T12:43:00Z,BSH,2009-11-05T16:53:32Z,DEPH TEMP
MYO-IBIROOS-
01,ftp://ftp.ifremer.fr/ifremer/ibiroos/monthly/drifter/IR_200910_PR_BA_10007.nc,54.17,54.17,7.43,7.43,2
009-10-01T00:40:00Z,2009-10-25T12:40:00Z,BSH,2009-11-05T17:01:51Z,DEPH TEMP
…
…
MYO-IBIROOS-
01,ftp://ftp.ifremer.fr/ifremer/ibiroos/monthly/vessel/IR_200911_TS_TS_SHIP.nc,37.7667,38.3333,-
24.8333,-23.1,2009-11-02T00:00:00Z,2009-11-02T23:00:00Z,Ifremer,2009-11-05T16:32:58Z,TEMP
```

2.1.2. Index platform

The platform index file is updated daily and registers the list of the individual platforms that are available on the server.

Each platform index contains a line per platform with the following information:

- **platform_code**
- **creation_date** : *creation date of the first file associated with the platform.* 1st of January 2010 if unknown
- **update_date** : *the latest update date of any file associated with the platform*
- **wmo_platform_code** : *list of the different wmo codes of the platform. Can be empty.*
- **data_source** : *list of different data streams of the platform*
- **institution**
- **institution_edmo_code** : *list of the different edmo codes of the platform. Can be empty.*
- **parameters**
- **last_latitude_observation**,
- **last_longitude_observation**
- **last_date_observation**

The triplet (last_latitude_observation, last_longitude_observation, last_date_observation) is associated to the last observation with valid date **and** position: time_qc in (1, 2, 5, 7, 8) and position_qc in (1, 2, 5, 7, 8).
A file with no valid time and position is not distributed on Copernicus In Situ TAC (see §**Erreur ! Source du renvoi introuvable.**).

Index name
- myo_index_platform.txt

Index lines order
The index lines are sorted by platform_code.

Separator character ","
The fields separator character is "," (comma).
The "," comma character must not appear within a field.

List separator " "
The list separator is " " (blank) except for the institution list.
Fields which can contain a list: wmo_platform_code, data_source, Institution_edmo_code, parameter

Institution list

This is a special case where each element can contain spaces.

- The list is in quotes
- The institutions are separated by "/"
- The character "/" of an institution is replaced by a space.

Data_source: name of data streams of the platform

Data streams are identified thanks to the different filenames without date/period value.
Data source list the distinct DUs and data types for each individual platform.

Latest files: RR_LATEST_XX_YY_CODE
Monthly files: RR_XXXXXX_XX_YY_CODE

- RR: region bigram (see table 2)
- XX: TS (timeseries) or PR (profile)
- YY: data type
- CODE: platform code
- XXXXXX : remplaces the date

Example : GL_LATEST_TS_DC_61792 GL_XXXXXX_TS_DC_61792

Parameters list

This list contains all the parameters whatever the quality of data, except PARAM_DM, PARAM_QC, TIME, LATITUDE, LONGITUDE, DATA_MODE and DIRECTION.

Date format

OceanSITES format: YYYY-MM-DDThh:mm:ssZ

Copernicus In Situ data file platform index example
Title : In Situ platforms catalog # Description : catalog of available In Situ platforms. # Project : Copernicus # Format version : 1.0 # Date of update : 20130802051813 # platform_code,creation_date,update_date,wmo_platform_code,data_source,institution,institution_edmo_code,parameter,last_latitude_observation,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 00136,2012-12-10T04:14:54Z,2013-06-24T14:06:31Z,00136,GL_XXXXXX_PR_TE_00000 GL_XXXXXX_PR_BA_00136,"Unknown institution",,DC_REFERENCE DEPH TEMP,32.015,158.862,2008-11-05T10:59:59Z 06GA,2013-06-24T21:41:58Z,2013-06-24T21:43:42Z,,GL_XXXXXX_PR_BA_00136,"BSH",,TEMP PSAL FLUO,54.04425,7.61149,2006-08-29T10:21:00Z 06SW,2013-03-17T20:46:16Z,2013-06-24T14:06:32Z,DCHO,GL_XXXXXX_PR_BA_06SW,"GERMANY (Unknown)",,DC_REFERENCE DEPH TEMP,36.65,-1.75,2010-10-03T12:45:00Z