Argo data management report 2023

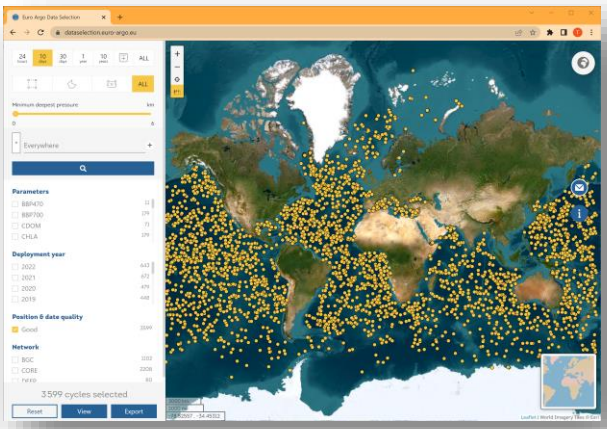
Coriolis DAC & GDAC

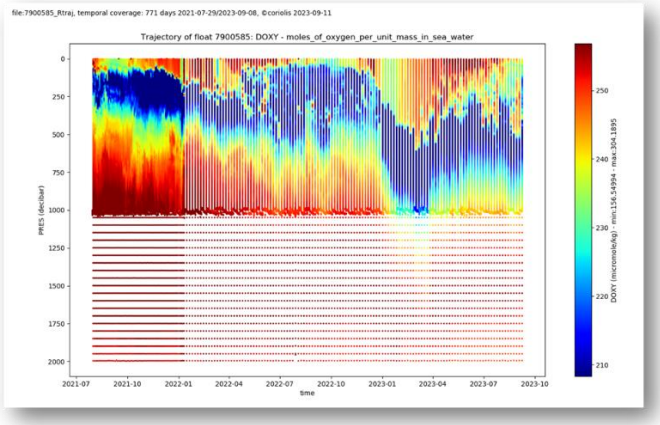
Data Assembly Centre and Global Data Assembly Centre

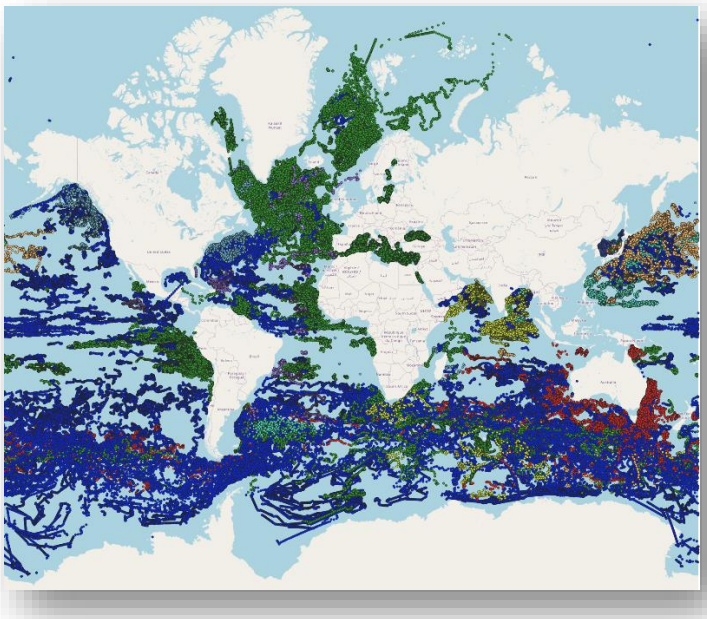
Annual report October 2023

Version 1.1

https://doi.org/10.13155/96772







1 DAC status

This report covers the activity of Coriolis DAC (Data Assembly Centre) for the one-year period from September 1st 2022 to September 30th 2023.

Key accomplishments in the past year include:

- The management of Provor CTS5 jumbo floats, which are substantial floats equipped with a range of BGC sensors, such as the UVP (a camera for identifying and counting zooplankton), the Ramses hyperspectral sensor, pH, chlorophyll, BBP, and Suna nitrate (refer to §1.1.3).
- The Ramses sensor conducts measurements of downward illuminance, upward luminescence, and reflectance. The reflectance data serves as a direct reference for satellite water color data.
- Comprehensive reprocessing of particulate backscattering (BBP) was carried out to implement the new quality control procedure available at https://doi.org/10.13155/60262
- Chlorophyll A (CHLA) also underwent a general reprocessing to adhere to the new quality control procedure, accessible at https://dx.doi.org/10.13155/35385
- The reprocessing of trajectories with format 3.2 was completed.
- An enhanced Argo real-time trajectory product that incorporates 20 additional RTQC procedures was developed. Additionally, a daily Argo deep currents product is now available for distribution.
- Deployment sheets for Coriolis floats, including metadata and calibrations, are shared on a cloud workspace located at cloud.ifremer.fr.

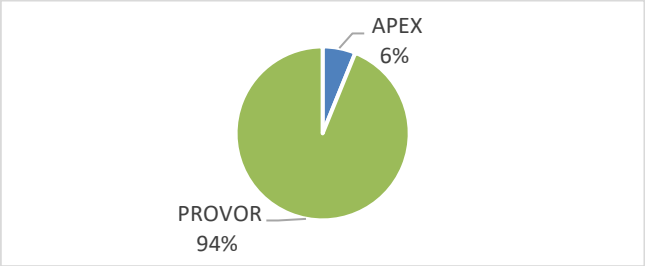
1.1 Data acquired from floats

1.1.1 Active floats for the last 12 months

These last 12 months, **62155 profiles from 923 active floats** were collected, controlled and distributed.
Compared to 2022, **the number of profiles slightly decreasing (-4%), the number of floats increased by 4%.**
These figures illustrate a good momentum in Coriolis DAC activity.

The 923 floats managed during that period had 48 versions of data formats.

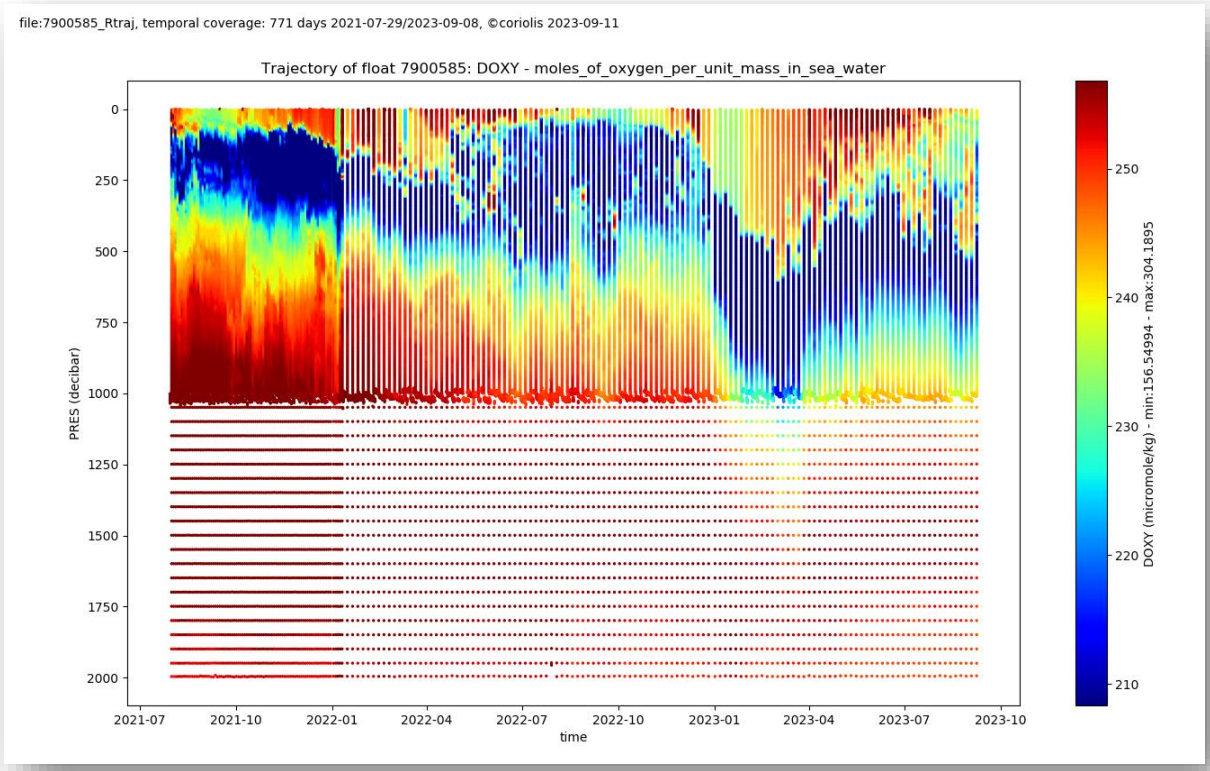
Coriolis DAC, active floats in 2023			
Float family	nb versions	nb floats	nb core p
APEX	15	57	2 710
NOVA	2	5	127
PROVOR	31	861	59 318
Total	48	923	62 155



1.1.2 All floats managed by Coriolis DAC

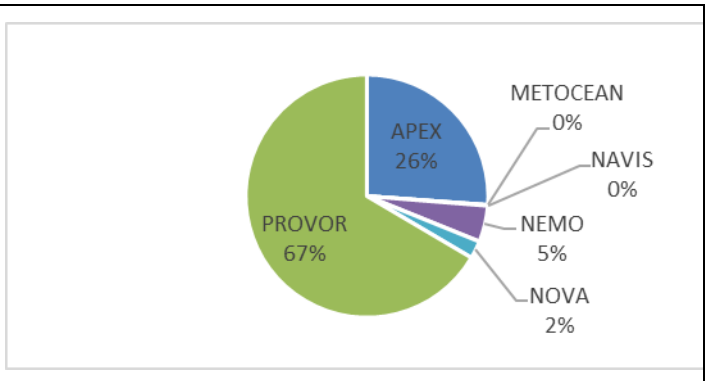
Coriolis DAC manages a total of 3.601 floats with 177 versions, from 6 families.
These floats reported 700.586 core Argo vertical profiles.

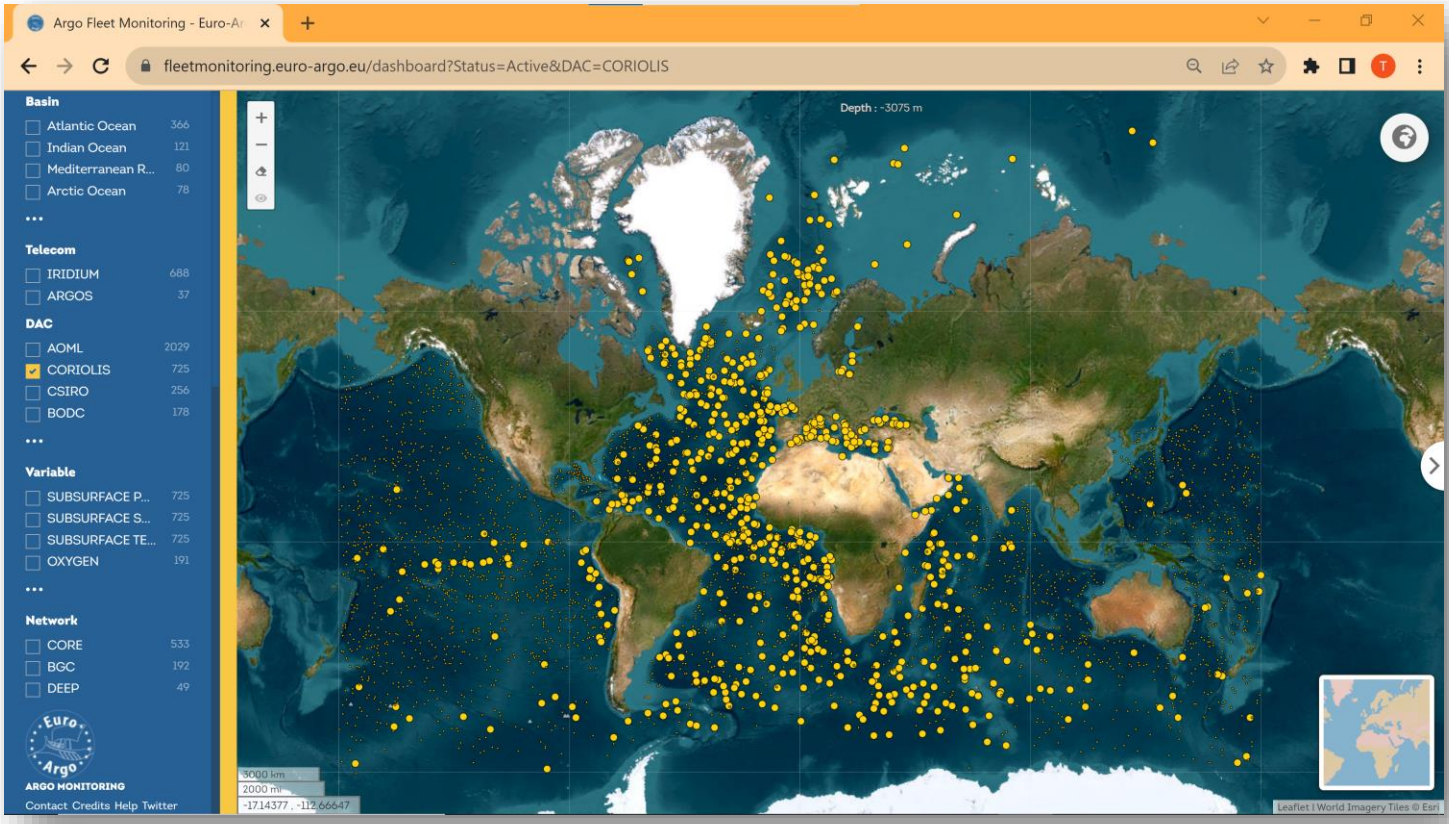
In 2023, most of Coriolis trajectory files were converted into format 3.2; in recent floats such as Provor CTS5 or Apex APF11, every observation has a timestamp and is available in the trajectory file.



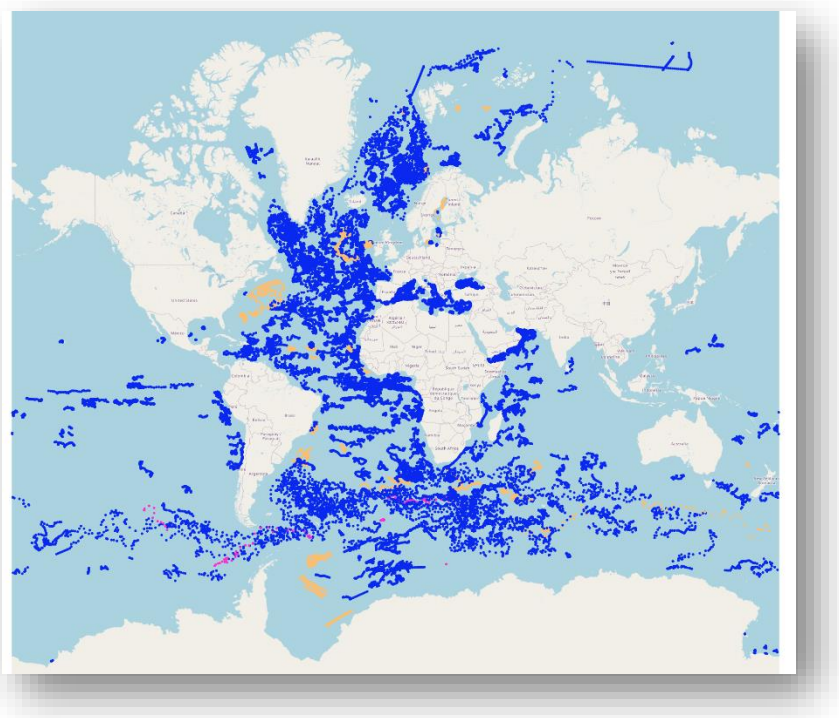
Oxygen observations from float 7900585 trajectory file

Coriolis DAC, all floats			
Float family	nb versions	nb floats	nb core p
APEX	80	942	148 540
METOCEAN	1	1	52
NAVIS	1	3	1 932
NEMO	8	174	10 185
NOVA	3	85	9 715
PROVOR	84	2 396	530 162
Total	**177**	**3 601**	**700 586**

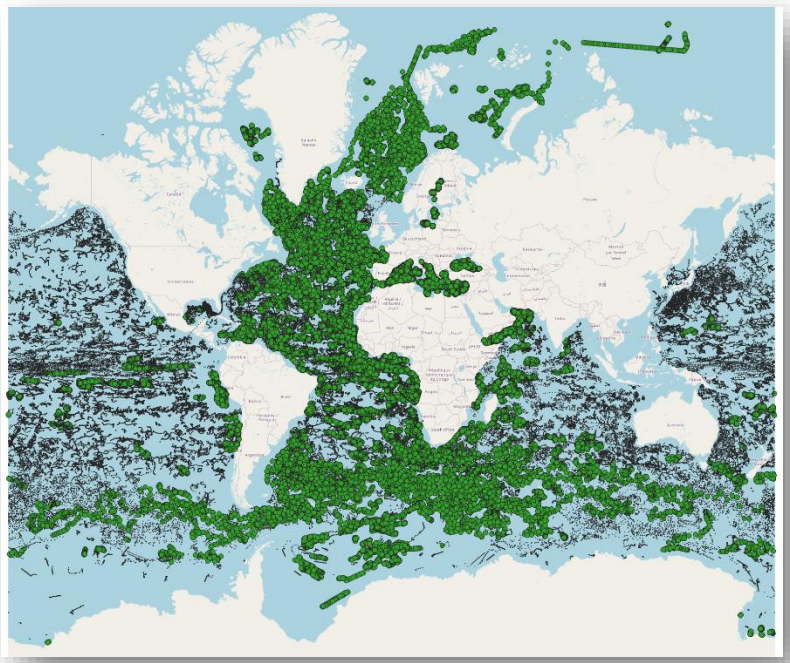




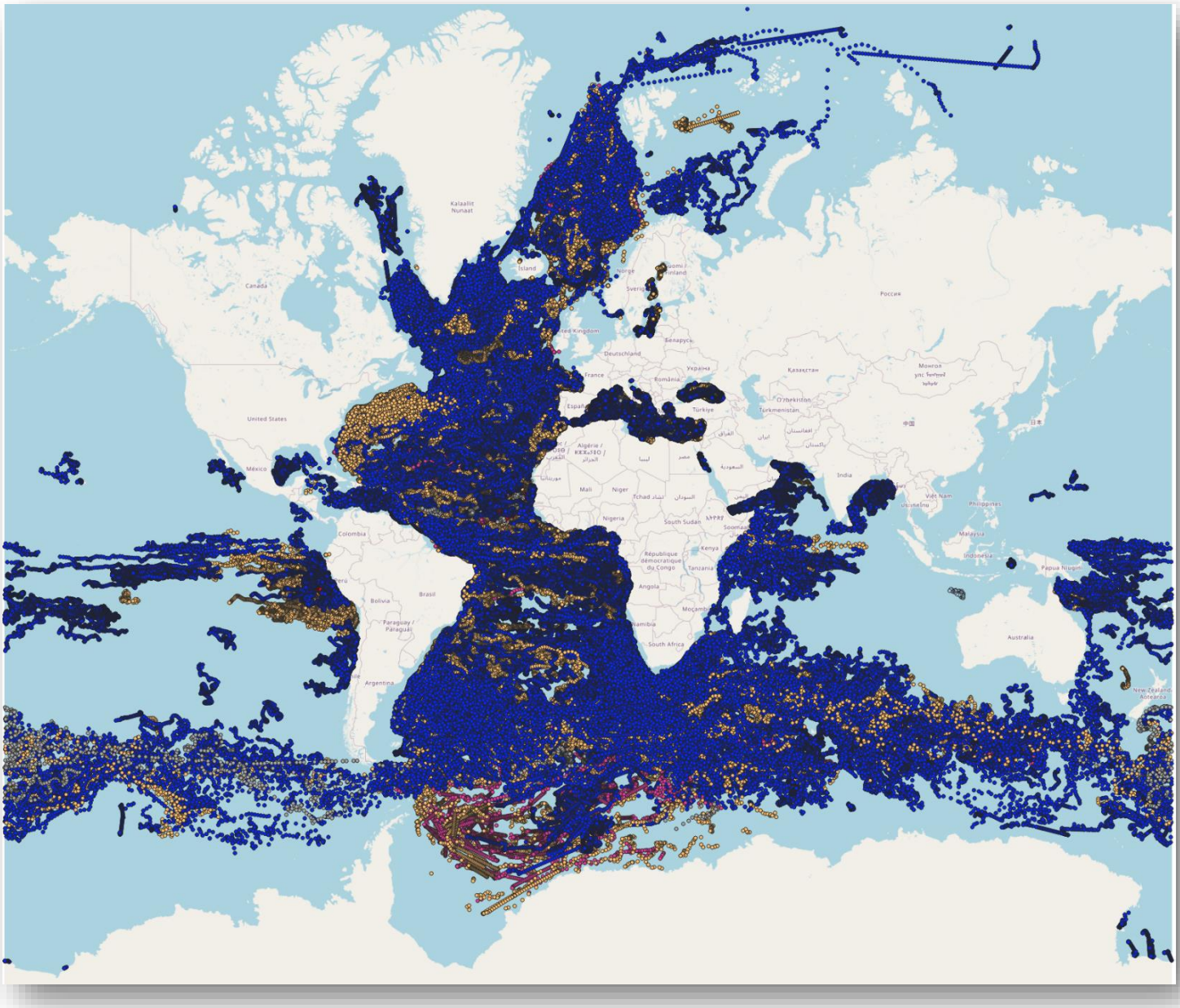
Map of the active floats on October 10th 2023 decoded by Coriolis DAC, among others DACs (small dots) as displayed on Euro-Argo floats dashboard https://fleetmonitoring.euro-argo.eu/dashboard



Map of the 62.155 profiles from 923 active floats active floats decoded by Coriolis DAC this current year
Apex Nova Provor

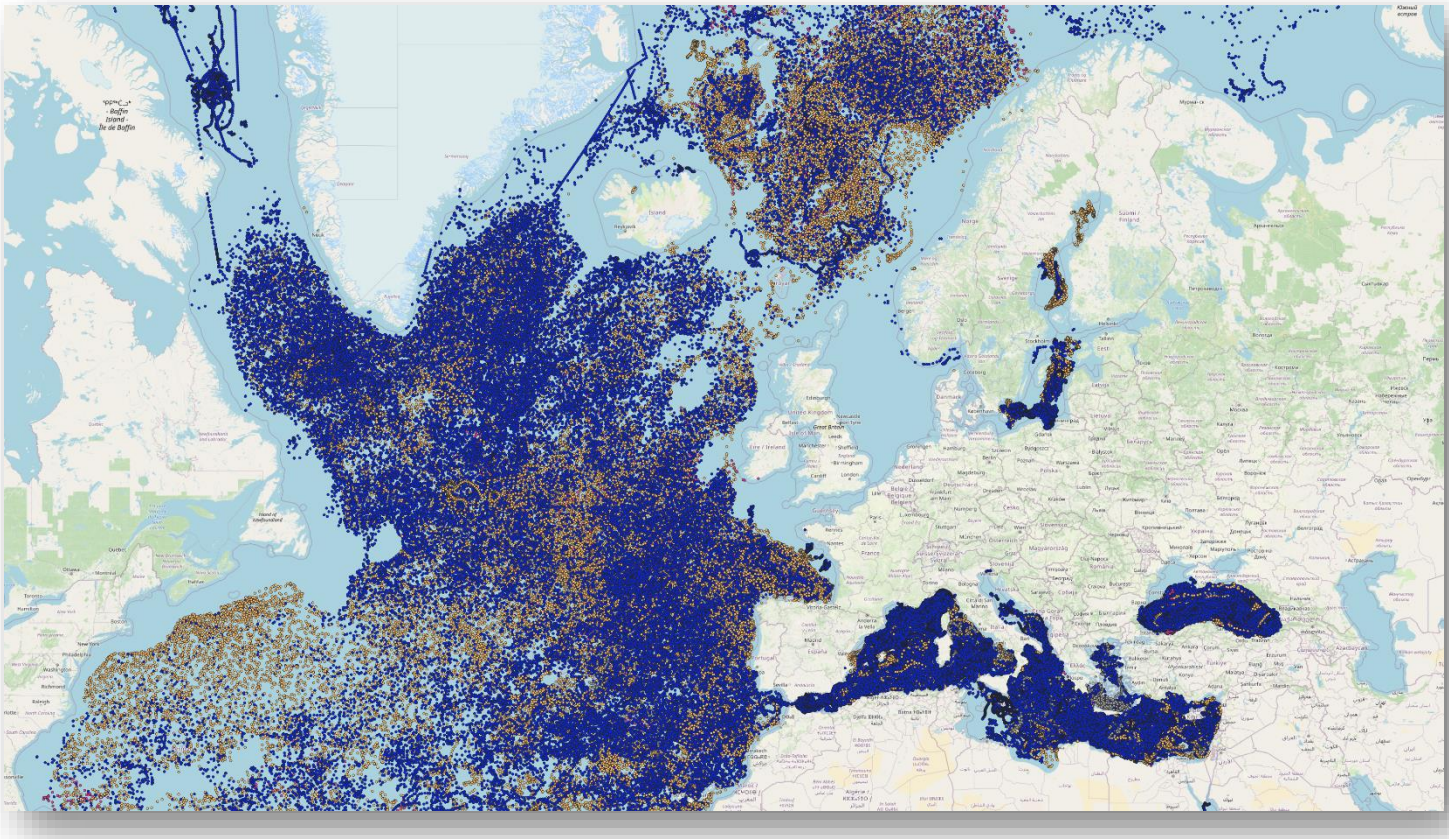


Map of the profiles from active floats decoded by Coriolis DAC this current year, among the other DAC's profiles (Coriolis: green, other DACs: grey)



Map of the 700.000 profiles from 3.600 floats managed by Coriolis DAC

Apex **Metocean** **Navis** **Nemo** **Nova** **Provor**



Map of the profiles floats managed by Coriolis DAC , focus on North Atlantic
Apex **Metocean** **Navis** **Nemo** **Nova** **Provor**

1.1.3 BGC-Argo sensors on Coriolis floats

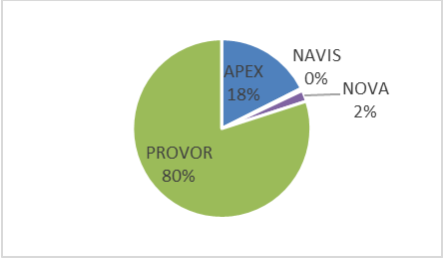
The data processing chain for data and metadata from Coriolis BGC-Argo floats is continuously improved. These are advanced types of floats performing bio-geo-chemical (BGC) measurements.

Coriolis DAC manages 733 BGC-Argo floats from 4 families. They performed 108.444 cycles.

The data processing chain is freely available:

- Coriolis Argo floats data processing chain, http://doi.org/10.17882/45589

Float familly	nb versions	nb floats	nb cycles
APEX	34	129	16 960
NAVIS	1	3	644
NOVA	1	15	1 241
PROVOR	48	586	89 599
Total	**84**	**733**	**108 444**



General characteristics

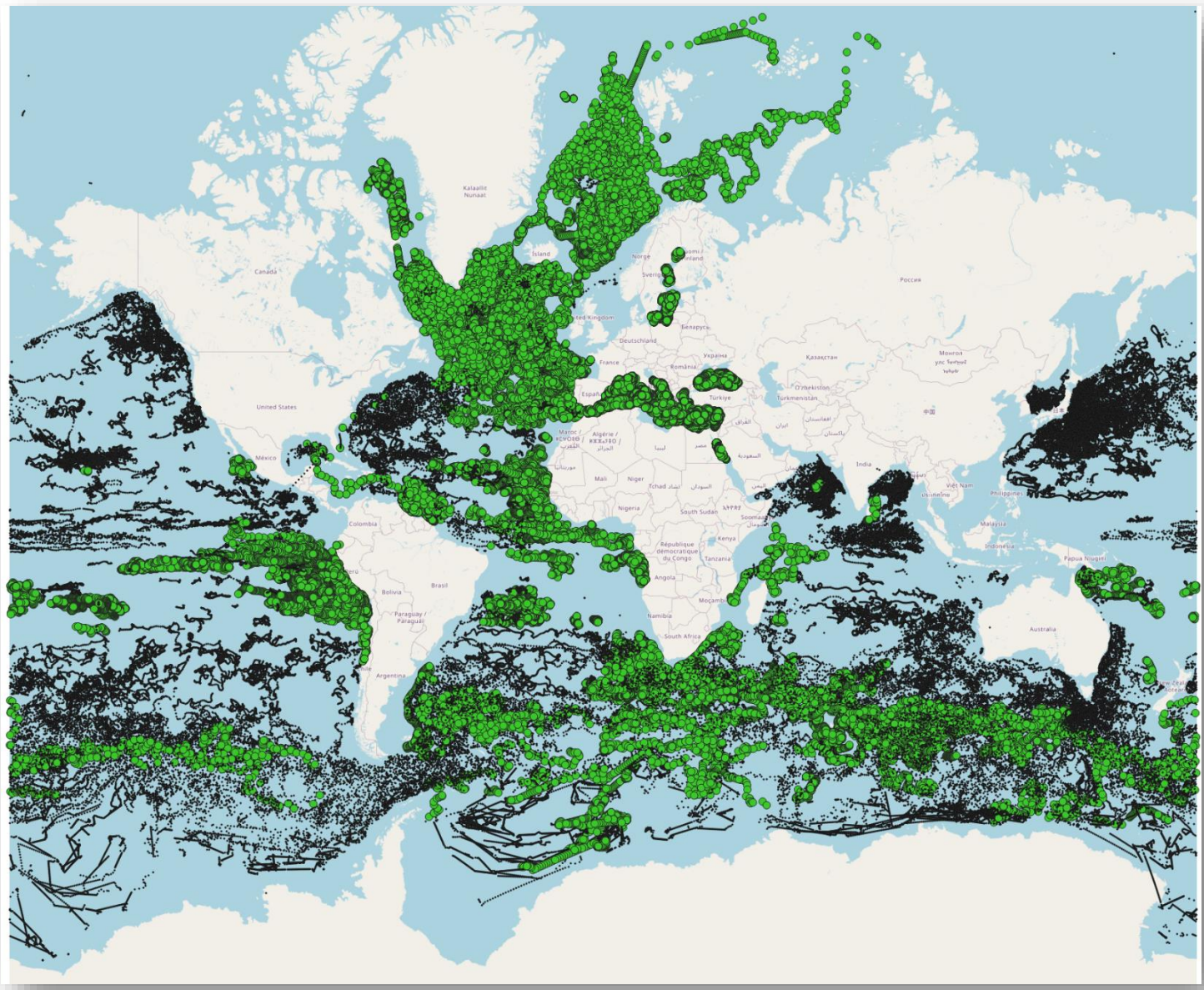
- Iridium sbd or rudics bi-directional communication or Argos
- There are 18 types of sensors fitted on the floats

Coriolis BGC-Argo floats sensor	nb floats	nb profiles
AANDERAA_OPTODE	655	94 631
ECO_FLBB	312	163 019
SATLANTIC_OCR504_ICSW	261	199 680
SUNA_V2	110	19 929
SEAFET	60	6 593
C_ROVER	43	5 724
UVP6-LP	32	1 556
SBE63_OPTODE	20	2 071
RAMSES_ACC	19	1 225
ECO_FLNTU	14	6 176
SBE43F_IDO	13	1 596
9AXIS_IMU	12	317
RAMSES_ARC	12	317
MPE	6	1 004
OPUS_DS	2	792
HYDROC	2	154
CYCLOPS-7_FLUOROMETER	2	106
SEAPOINT_TURBIDITY_METER	2	106

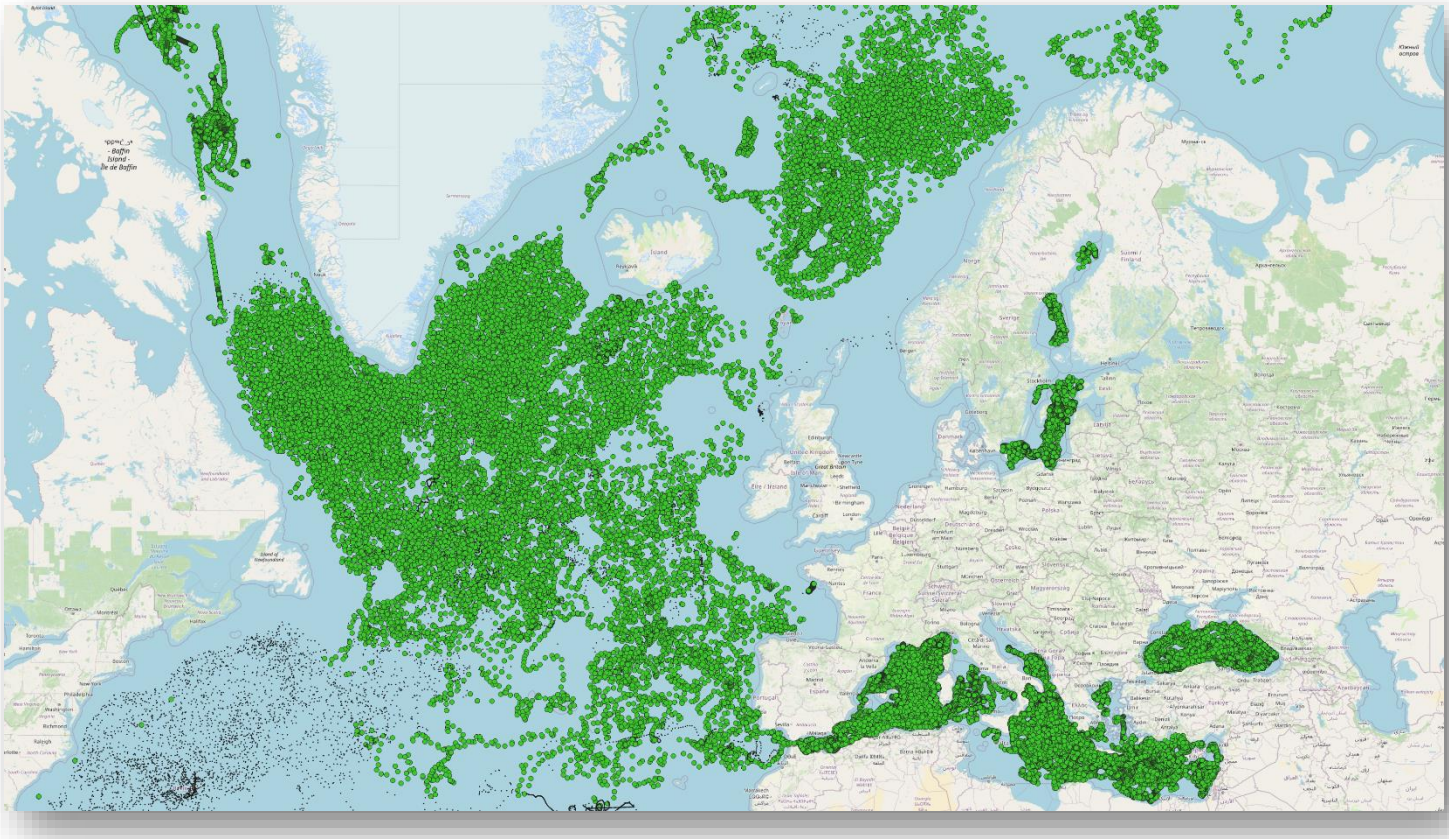
The 18 types of sensors mounted on Coriolis BGC-Argo floats

PARAMETER_CODE	NB_FILE
DOXY	97 201
CHLA	50 512
BBP700	47 998
CDOM	43 094
DOWNWELLING_PAR	42 339
NITRATE	16 509
PH_IN_SITU_TOTAL	5 614
TURBIDITY	2 514
BISULFIDE	1 383

The 9 main BGC parameters reported by Coriolis BGC-Argo floats



Map of the 733 BGC-Argo floats managed by Coriolis DAC (grey dots: the others DACs bio-Argo floats). They measure parameters such as oxygen, chlorophyll, turbidity, CDOM, back-scattering, UV, nitrate, bisulfide, pH, radiance, irradiance, PAR.



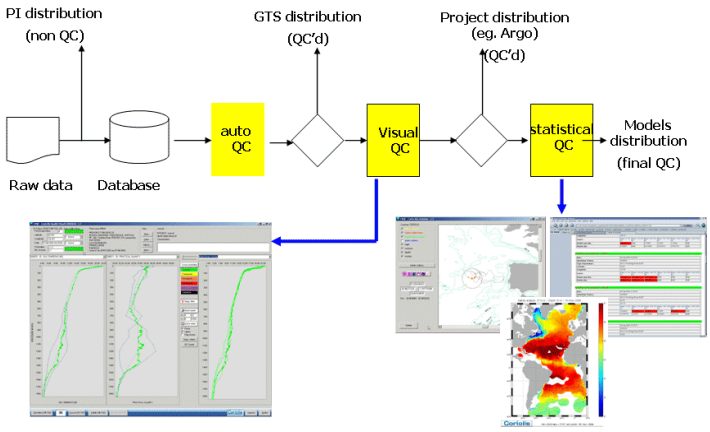
A zoom on North Atlantic of the BGC-Argo floats managed by Coriolis DAC (grey dots: the others DACs bio-Argo floats)

1.2 Data issued to GTS

Vertical profiles processed by Coriolis are distributed on the GTS by way of Meteo-France. This operation is fully automated. After applying the automatic Argo QC procedure, the Argo profiles are inserted on the GTS every hour. The profile files are sent as BUFR messages.

Vertical profiles are distributed on GTS if they are less than 30 days old. Once a day, floats data are checked with ISAS objective analysis that triggers alerts and visual inspection for suspicious observations.
The corrected data are not redistributed on GTS.

In July 2019, Coriolis stopped the TESAC messages distribution; only BUFR messages are now distributed.

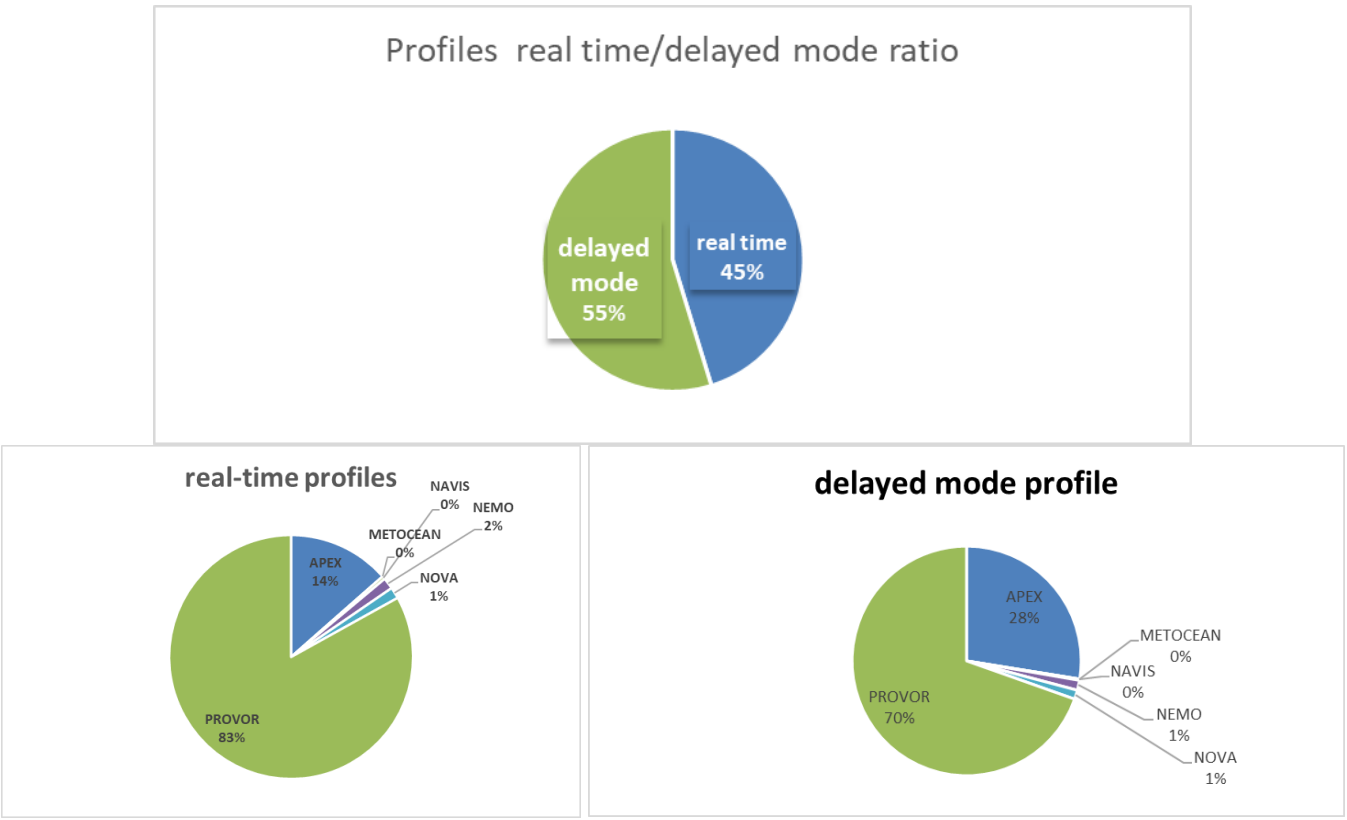


Coriolis DAC Argo data flow

1.3 Data issued to GDACs after real-time QC

All meta-data, profiles, trajectory and technical data files are sent to Coriolis and US-GODAE GDACs. This distribution is automated.

All Coriolis floats, number of profile files on GDAC				
Family	nb floats	nb profile	RT profile	DM profiles
APEX	862	148 540	42 858	105 682
METOCEAN	1	52	-	52
NAVIS	3	1 932	1 288	644
NEMO	110	10 185	4 786	5 399
NOVA	61	9 715	4 776	4 939
PROVOR	1864	530 552	263 360	267 192
Total	**2 901**	**700 976**	**317 068**	**383 908**



Distribution of Coriolis DAC real-time and delayed mode profiles

1.4 Data issued for delayed mode QC

Delayed mode profiles

All profile files are sent to PIs for delayed QC.

1.5 Delayed mode data sent to GDACs

An Argo delayed mode profile contains a calibrated salinity profile (psal_adjusted parameter).

- A total of **108.689 new or updated delayed mode profiles** was sent to GDACs this year.
- **A total of 383.908 delayed mode profiles** were sent to GDACs since 2005.

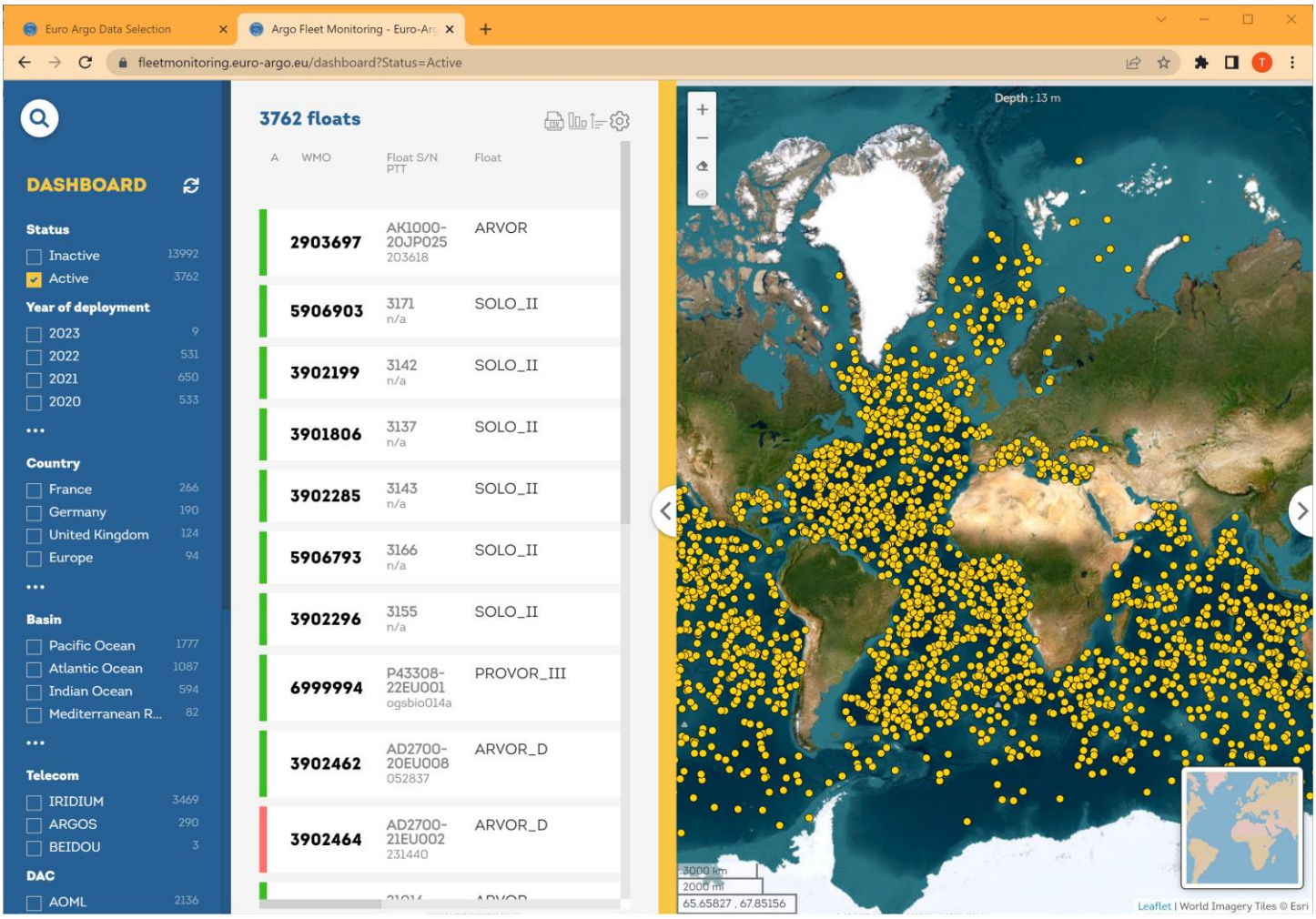
The number of delayed mode profiles increased by 1% this year compared to 2022.

1.6 Web pages

1.6.1 Argo dashboard

The Argo floats dashboard developed in 2019 and regularly improved by Coriolis team is available at:

- https://fleetmonitoring.euro-argo.eu/dashboard

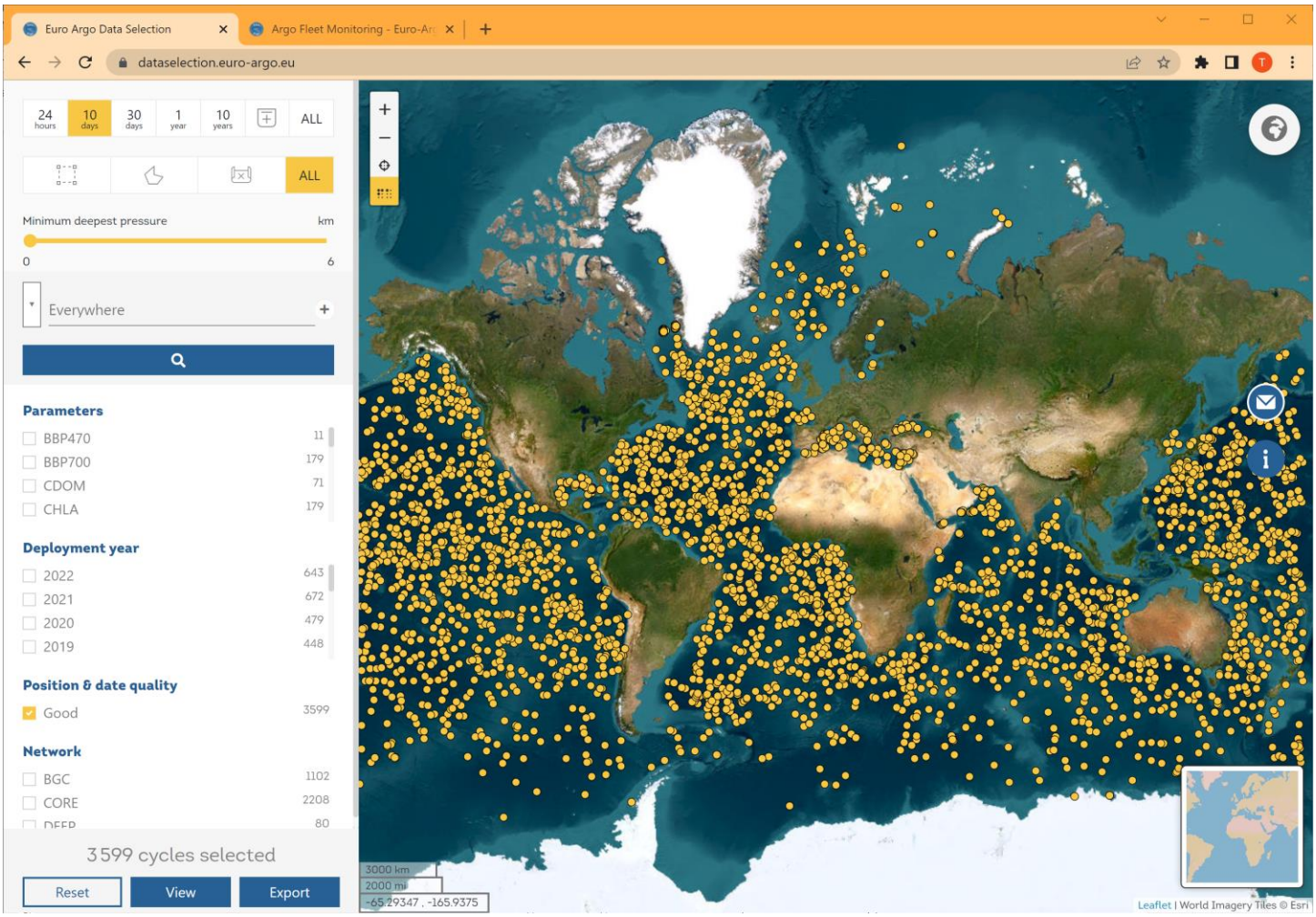


It displays all Argo floats, with facetted interrogations and instantaneous answers. The dashboard is developed on cloud and big-data techniques.

- Cloud techniques: a metadata and a data APIs, opened to internet machine to machine queries
- Big-data techniques: Argo metadata are hourly indexed in an Elasticsearch index, Argo data are hourly indexed in a Cassandra data base. Elasticsearch and Cassandra allows instant answers on dataset having billions of observations.

The Argo data selection was developed in 2020. The initial version is online at https://dataselection.euro-argo.eu/

It proposes data discovery with facetted search on temporal and spatial coverage, parameters, deployment years or quality codes. The selected data are downloadable in NetCDF and CSV formats.



Argo data selection https://dataselection.euro-argo.eu

1.6.2 Interoperability services (ERDDAP API,…)

The APIs used by Argo dashboard and Argo data selection web portals are open and publicly available to interested users at the following endpoints OpenAPI (swagger):

- https://fleetmonitoring.euro-argo.eu/swagger-ui.html
- https://dataselection.euro-argo.eu/swagger-ui.html

More information available on https://www.euro-argo.eu/Argo-Data-access

This web page describes all Argo floats interoperability services from Coriolis:

- http://www.coriolis.eu.org/Data-Products/Data-Delivery/Argo-floats-interoperability-services2
    - Argo data through ERDDAP data server (www.ifremer.fr/erddap)
    - Display an individual float's data and metadata in HTML or XML format
    - Display all Argo floats, display a group of floats
    - Argo profiles and trajectories data selection (HTML or XML)
    - All individual float's metadata, profile data, trajectory data and technical data
    - Argo profiles data on OpenDAP, OGC-WCS and http

- Argo data through Oceanotron data server
- Argo profiles data through GCMD-DIF protocol
- Argo data through RDF and OpenSearch protocols

1.6.3 Data centre activity monitoring

Coriolis operators perform an activity monitoring with an online control board.

Fonction	Description	Etat J	Etat J-1	Etat J-2	Etat J-3	Dernière exécution (TU)
CO-05-08-08	Archivage du GDAC Argo pour DOI (mensuelle)					OK 2023-09-09T03:44:14Z
CO-01-01-13	Argo - Synchronisation des QC de la base et du DAC Coriolis					OK 2023-10-07T16:42:29Z
CO-01-07-08	Collecte Argo Coriolis EDAC					OK 2023-10-07T16:55:11Z
CO-01-07-08-02	Collecte Argo Coriolis EDAC - table index					OK 2023-10-07T16:54:26Z
CO-01-07-01-aoml	Collecte Argo DAC - aoml					OK 2023-10-07T17:01:03Z
CO-01-07-01-bodc	Collecte Argo DAC - bodc					OK 2023-10-07T17:02:02Z
CO-01-07-01-coriolis	Collecte Argo DAC - coriolis					OK 2023-10-07T17:03:02Z
CO-01-07-01-csio	Collecte Argo DAC - csio					OK 2023-10-07T17:04:02Z
CO-01-07-01-csiro	Collecte Argo DAC - csiro					OK 2023-10-07T17:05:42Z
CO-01-07-01-incois	Collecte Argo DAC - incois					OK 2023-10-07T17:06:02Z
CO-01-07-01-jma	Collecte Argo DAC - jma					OK 2023-10-07T17:07:06Z
CO-01-07-01-kma	Collecte Argo DAC - kma					OK 2023-10-07T17:08:12Z
CO-01-07-01-kordi	Collecte Argo DAC - kordi					OK 2023-10-07T17:09:02Z
CO-01-07-01-meds	Collecte Argo DAC - meds					OK 2023-10-07T17:10:02Z
CO-01-07-01-nmdis	Collecte Argo DAC - nmdis					OK 2023-10-07T17:11:02Z
CO-01-07-01-03	Collecte Argo DAC - resubmit files cause meta missing					OK 2023-10-07T14:51:01Z
CO-01-07-01-02	Collecte Argo DAC - table index					OK 2023-10-07T17:22:58Z

Argo GDAC operations monitoring: every working day, an operator performs diagnostics and take actions on anomalies (red or orange smileys)

1.7 Statistics of Argo data usage (operational models, scientific applications, number of National Pis…)

Operational oceanography models; all floats data are distributed to:

- EU Copernicus Marine service models (Mercator, Foam, Topaz, Moon, Noos, Boos)
- French model Soap (navy operational model)

Argo projects: this year, Coriolis data centre performed float data management for **89 Argo scientific projects and 56 PIs (Principal Investigators).**

List of Coriolis scientific PIs and project names

project	nb floats
argo-bsh	108
coriolis	100
argo italy	86
argo germany	63
norargo2	38
argo-france	36
argo bsh	34
ovide	33
mocca	32
mocca-eu	28

Top 10 of Coriolis DAC projects having active floats

pi name	nb active flo
birgit klein	196
pierre-marie poulain	69
christine coatanoan	65
kjell arne mork	55
damien desbruyeres	42
elena mauri	34
andreas sterl	34
herve claustre	26
romain cancouet	25

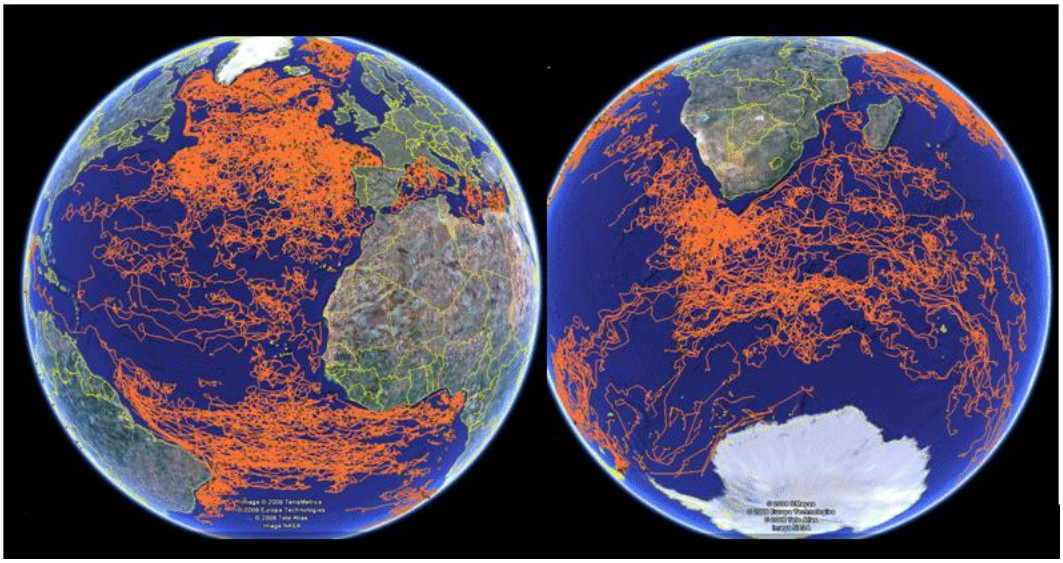
Top 10 of Principal Investigators (PI) in charge of active floats

1.8 Products generated from Argo data

Sub-surface currents ANDRO Atlas

Based on Argo trajectory data, Ifremer and CNRS team are regularly improving the "Andro" atlas of deep ocean currents. The ANDRO project provides a world sub-surface displacement data set based on Argo floats data. The description of each processing step applied on float data can be found in:

- Ollitrault Michel, Rannou Philippe, Brion Emilie, Cabanes Cecile, Piron Anne, Reverdin Gilles, Kolodziejczyk Nicolas (2022). **ANDRO: An Argo-based deep displacement dataset**. SEANOE. https://doi.org/10.17882/47077



Argo trajectories from Coriolis DAC are carefully scrutinized to produce the "Andro" atlas of deep ocean currents.

Sub-surface currents real time data

The Argo current product produced by Copernicus marine in situ is derived from the original trajectory data from Argo GDAC (Global Data Assembly Center). The Argo currents are calculated from Argo trajectories format version 3.1 or higher; the previous formats are ignored (2.*, 3.0).

It is daily updated and available from https://doi.org/10.48670/moi-00041

In November 2023 release, two significant improvements are implemented:

- A series of 20 quality control tests is applied on each Argo trajectory file documented in *Herbert Gaelle (2020). Qualification temps réel des données trajectoire des flotteurs Argo. https://doi.org/10.13155/95169*
- The currents are calculated with the Ollitrault-Rannou method documented in *Ollitrault Michel, Rannou Jean-Philippe (2013). ANDRO Dataset contents and format. https://archimer.ifremer.fr/doc/00360/47126*

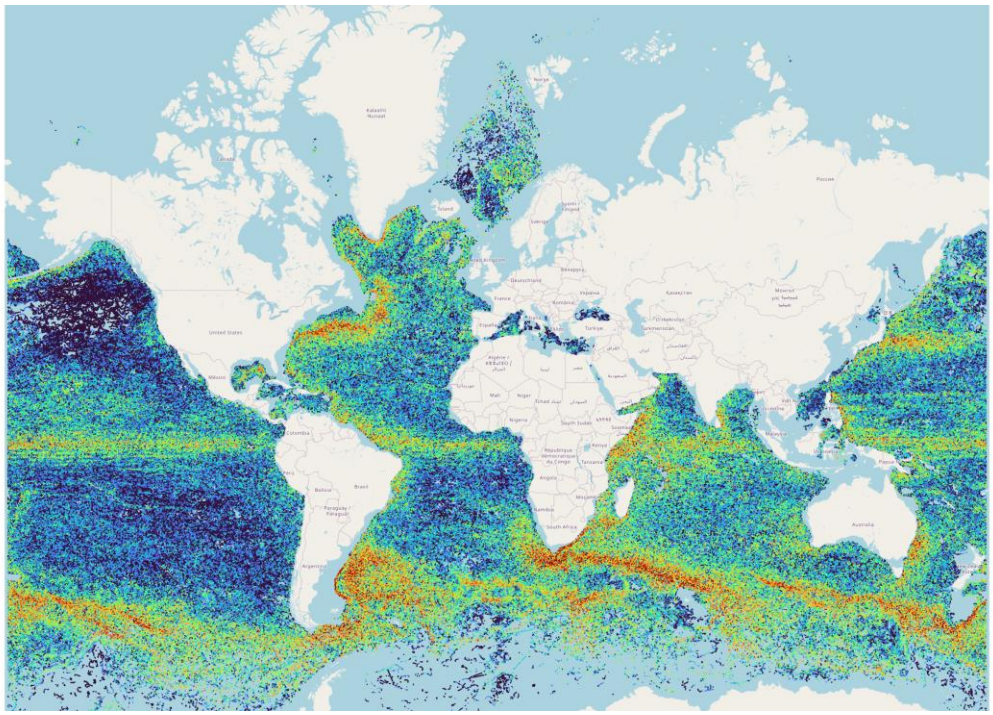


Figure 20: Map of Argo deep ocean currents, each dot represents the deep ocean current from one cycle (typically 10 days) from one float
From dark blue dot: 0 meter/second, to red dot: 2 meter/second

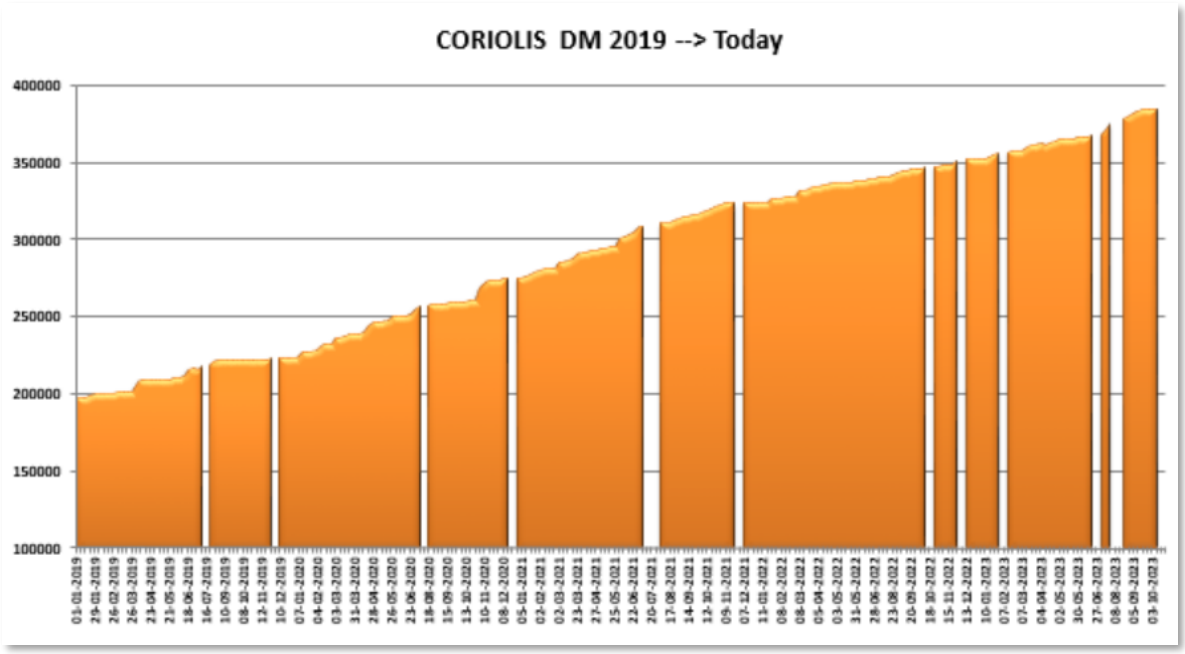
2 Delayed Mode QC

At the Coriolis data centre, we process the delayed mode quality control following four steps. Before running the OW method, we check carefully the metadata files, the pressure offset, the quality control done in real time and we compare with neighbor profiles to check if a drift or offset could be easily detected. By working on this way with PIs, the delayed mode quality control is strengthen.

Some floats have been deployed from some projects, meaning a lot of PIs and a lot of time for explaining the DM procedure to all of them. A few PIs are totally able to work on DMQC following the four steps but this is not the case for most of them. Since the unavailability of the PIs leads to work by intermittence and then extend the period of work on the floats, we did the work with a private organism (Glazeo) to improve the realization of the DMQC, exchanging only with the PIs to validate results and discuss about physical oceanography in studied area. Working in this way, we largely improve the amount of delayed mode profiles

A lot of work is always done from BSH (Birgit Klein) taking into account also floats from other German institutes and OGS (Antonella Gallo/Massimo Pacciaroni/Giulio Notarstefano) for the MedSea as well as Alberto Gonzalez Santana for IEO.

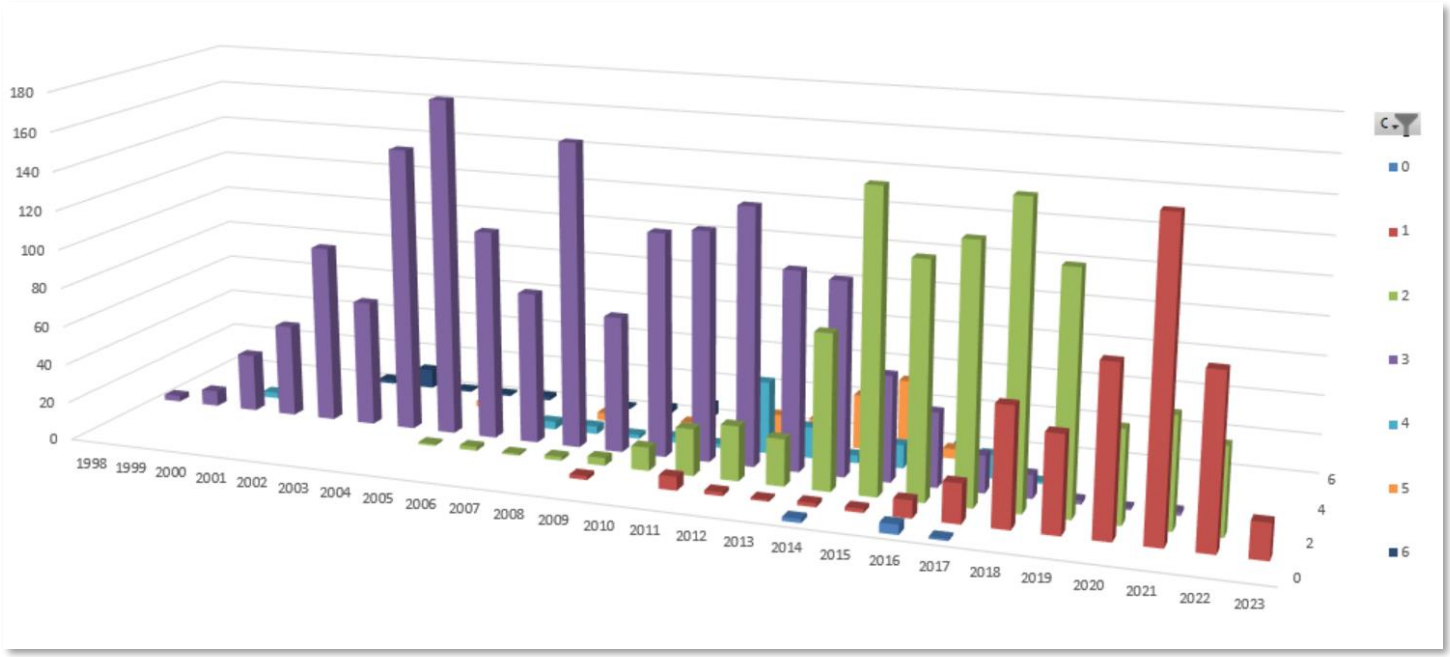
Over the past 5 years, a major effort has been made to steadily improve the quality control status of the delayed mode.



Evolution of the DM profiles' submission versus dates in last 5 years

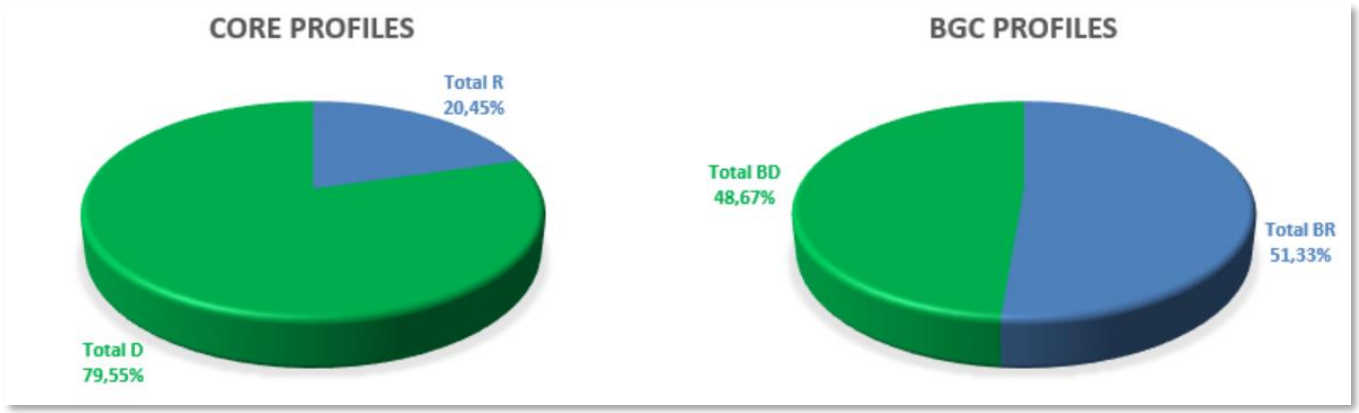
During the last year (from October 2022 to October 2023), 37524 new delayed mode profiles where produced and validated by PIs. A total of 383908 delayed mode profiles where produced and validated since 2005.

The next figure shows the distribution of float by year and by quality level.



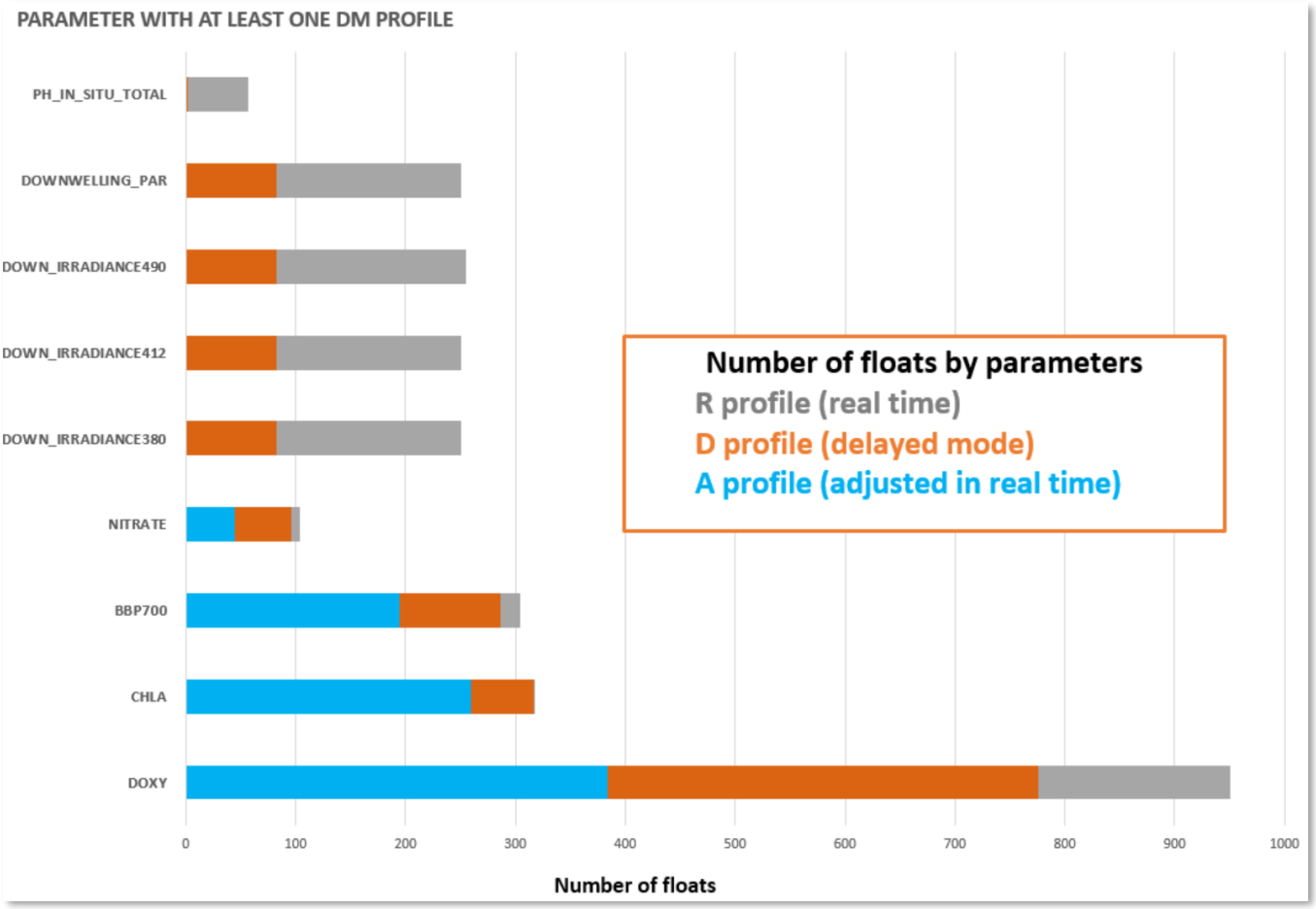
Status of the quality control done on profiles sorted by launch's year, code 1: young float, code 2: active float, DM done, code 3 : dead float, DM done; code 4 : DM in progress, code 5 : waiting for DM, code 6 : problems with float

In the following float, the status of quality control on floats is presented. For the BGC floats, the BD percent means that at least one parameter has been processed in delayed mode.



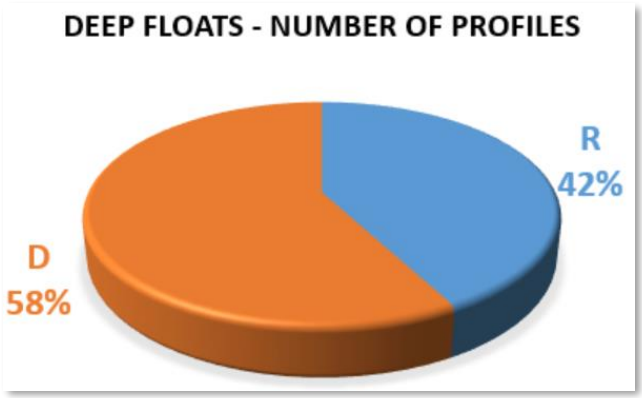
Status of the floats processed by Coriolis DAC.
Left: CORE profiles percent and right: BGC profiles percent (D/BD : delayed mode – R/BR : real time).

The status of the quality control done on the Coriolis BGC floats is presented in the following plot for some BGC parameters. Some parameters are regularly updated in DM mode (Doxy, ChlA, BBP).

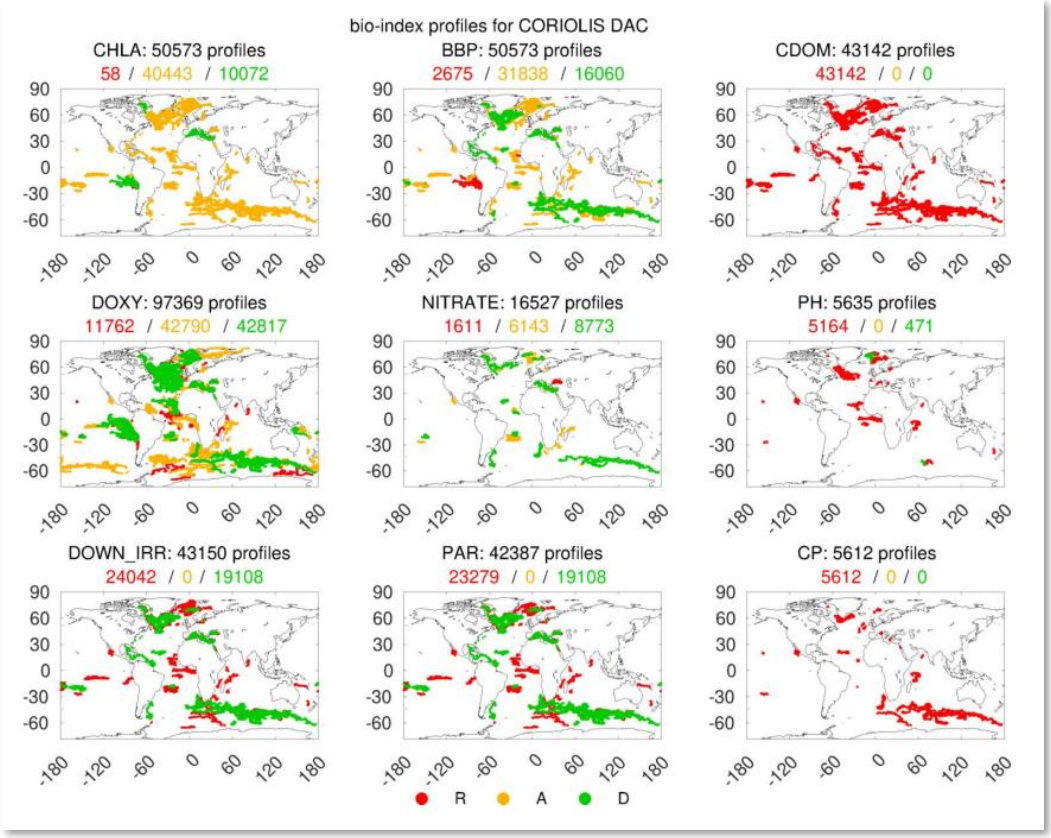


Status of the quality control done on BGC floats.
Float for which at least one profile has been performed in delayed mode for the parameter.

Looking in more detail to focus Deep Argo data, a great effort has also been made to increase the count of delayed mode profiles: 58% of Deep Argo profiles have been processed in delayed mode.



Regarding the BGC data, some information can be found on the document provided by the audit of Henry Bittig (*https://biogeochemical-argo.org/cloud/document/implementation-status/BGC_summary_coriolis.pdf*).

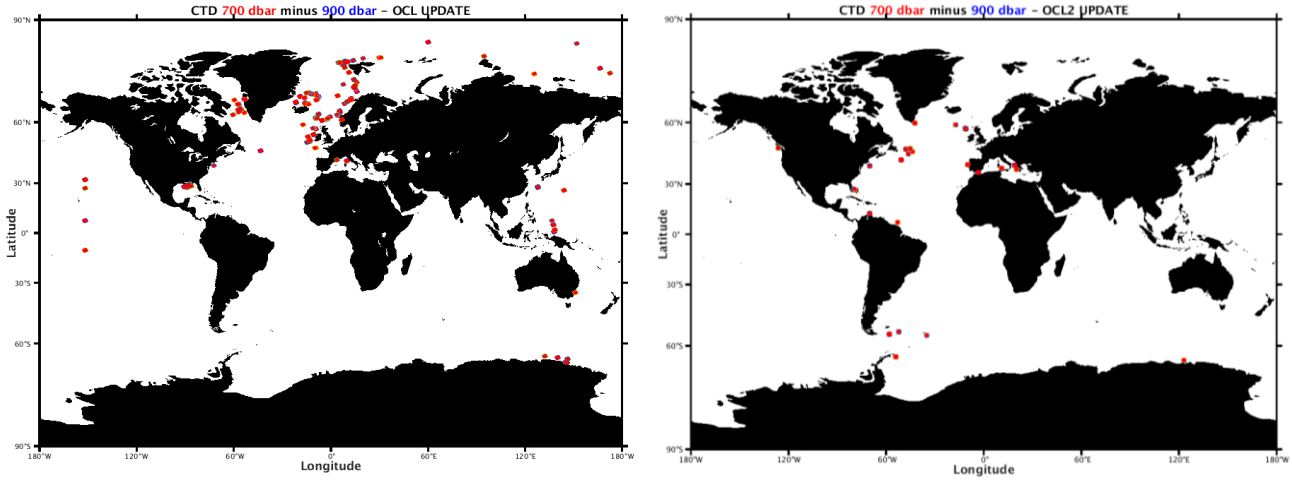


Location and number of R,A,D profiles per parameters.

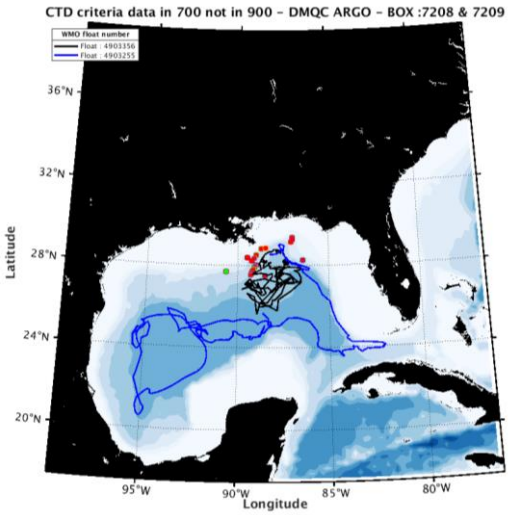
CTD Reference database

A work is in progress to take into account the criteria 700 dbar in shallower regions, as decided at the last ADMT. A new version 2023V01 is currently being prepared and will include updates from OCL, data provided by CCHDO (confidential, and GO-SHIP data from Website), as well as some corrections of anomalies detected by users.

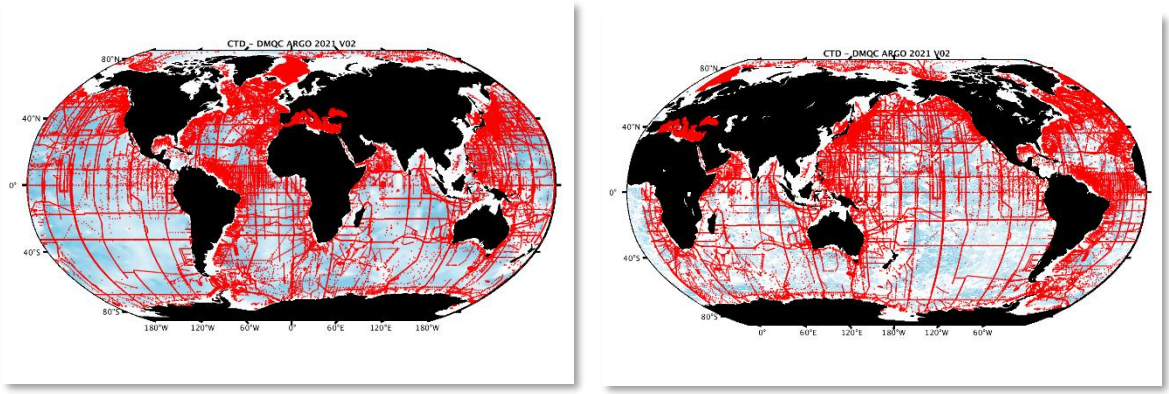
Some tests have been made to study differences between criteria 700 dbar or 900 dbar for selecting profiles, which be included in the CTD reference database. Starting with the two OCL updates datasets, boxes have been created using the both criteria and differences have been studied to vizualise on which part of the ocean, new stations can increase the CTD dataset for the calibration.



In the Gulf of Mexico, it seems that using 700 dbar as criteria will improve the number of CTD to perform the calibration method.



The last version is available on the Ifremer ftp site (ask login/password at codac@ifremer.fr) and is divided in smaller tar balls, one by wmo box area (1-3-5-7): for instance, CTD_for_DMQC_2021V02_1.tar.gz for all boxes starting with wmo 1, then we will have 4 tar files.



3 GDAC Functions

A significant achievement for the last 12 months is the daily distribution of Argo data on Amazon cloud infrastructure, see §3.9 .

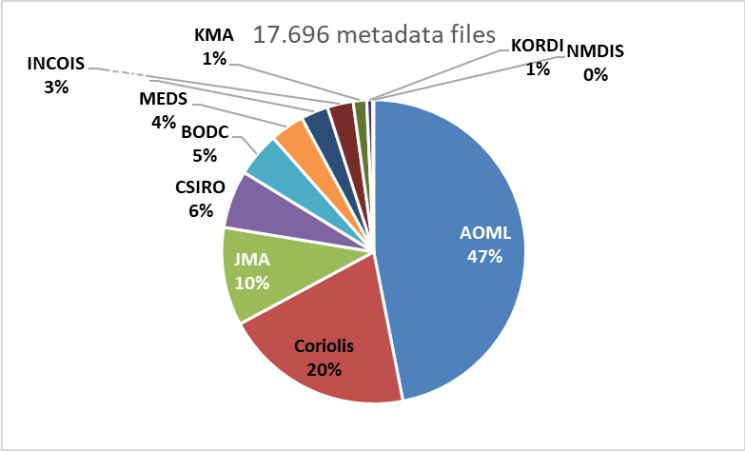
3.1 National centres reporting to you

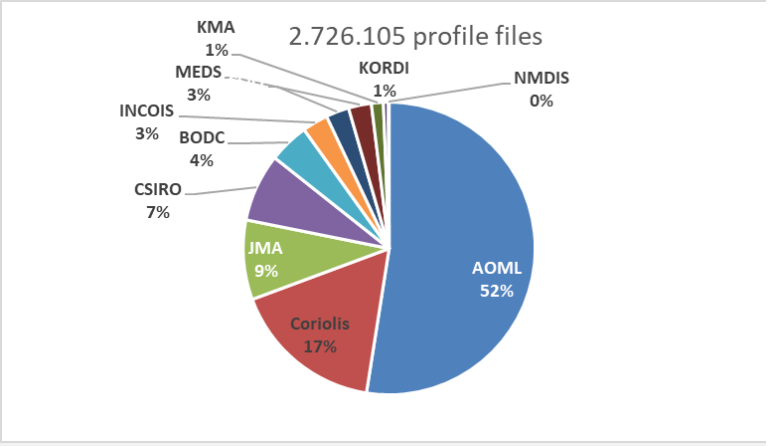
Currently, 11 national DACs submit regularly data to Coriolis GDAC. On October 2023, the following files were available from the GDAC FTP site.

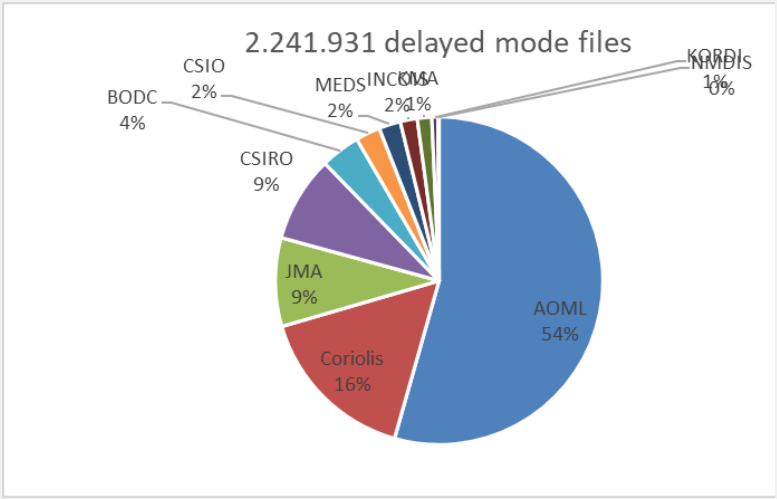
Compared to 2022, the number of floats (metadata) increased by 4%, the number of profile files increased by 5%.

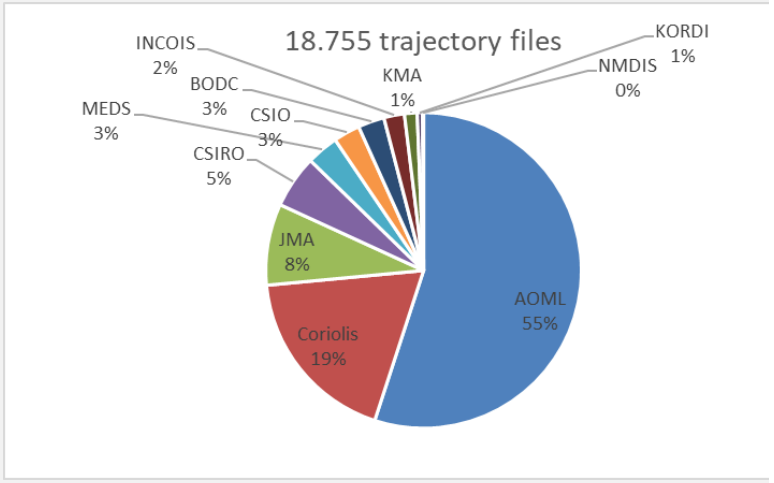
3.1.1 GDAC files distribution in October 2023

DAC	metadata files	increase	profile files	increase2	delayed mode profile files	increase3	trajectory files
AOML	8 612	4%	1 507 278	5%	1 296 336	6%	10 782
BODC	871	6%	127 722	7%	93 122	2%	530
Coriolis	3 710	5%	483 866	8%	383 908	10%	3 624
CSIO	534	2%	73 594	4%	57 822	4%	533
CSIRO	1 130	3%	215 407	5%	201 710	6%	1 062
INCOIS	505	3%	81 813	2%	39 996	10%	412
JMA	1 921	2%	252 772	4%	208 454	4%	1 635
KMA	264	2%	38 064	2%	34 052	1%	255
KORDI	117	2%	15 624	1%	14 504	0%	107
MEDS	672	3%	72 768	7%	51 012	2%	640
NMDIS	19	0%	2 460	0%	2 381	-	19
Total	**18 355**	**4%**	**2 871 368**	**5%**	**2 383 297**	**6%**	**19 599**









Number of files available on GDAC, October 2023

3.1.2 Argo Semaphore dashboard: give credit to data providers

This chapter will be updated in a next version 1.1 of the report.

Ifremer manage a dashboard (Semaphore) to monitor data distribution and give credit to data providers such as Argo floats.

- *https://audience-argo.ifremer.fr*

FTP, HTTPS and ERDDAP downloads log files are ingested in an Elasticsearch index. A link between downloaded files, download originators, floats included in the downloaded files and institution owners of the floats is performed. These links are displayed in a Kibana dashboard.

This dashboard offers the possibility to give credit to Floats owner institutions such as how many data from one particular institution was downloaded, by whose data users.

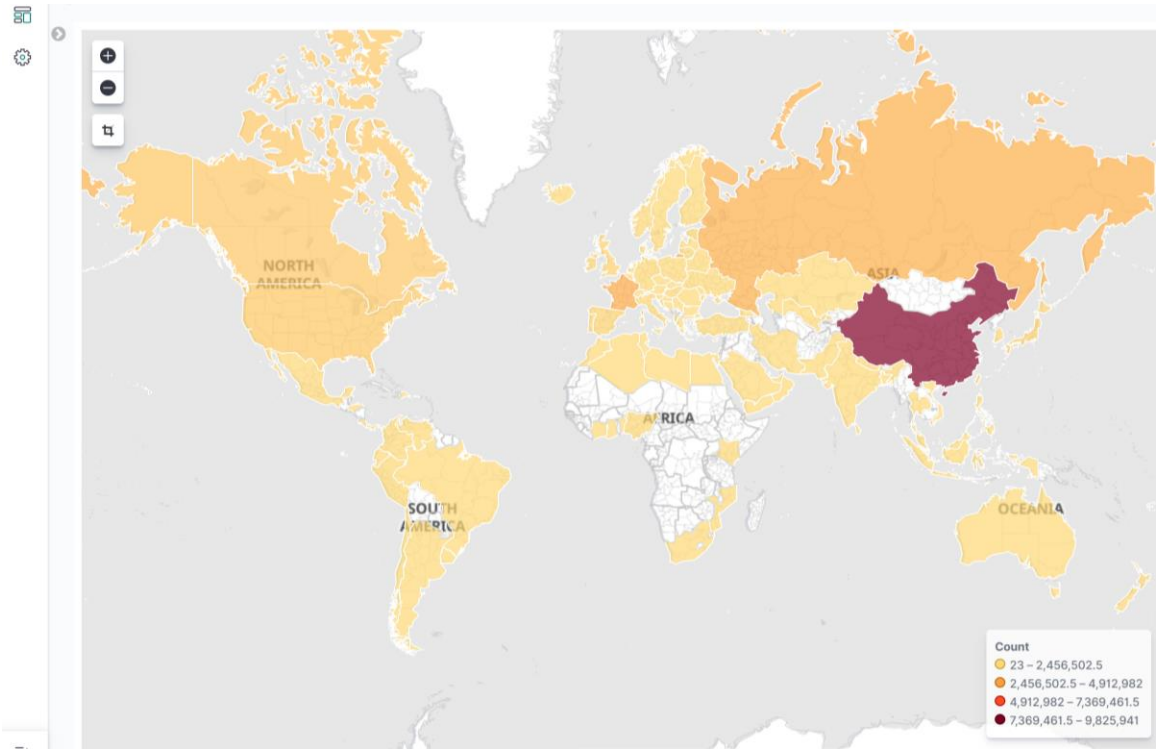
Semaphore key figures for 2022:

- ***2 million sessions for Argo data downloads***
- ***655 million of files downloaded***
- ***80% of ftp downloads, 20% of https downloads***
- ***20 terabytes daily downloads***

Download service	*Nb hits*	*percent*
ftp	*525 899 105*	*80%*
https	*127 977 145*	*20%*
thredds	*816 041*	*0,12%*
cms-ezpublish	*536 706*	*0,08%*
erddap	*72 709*	*0,01%*

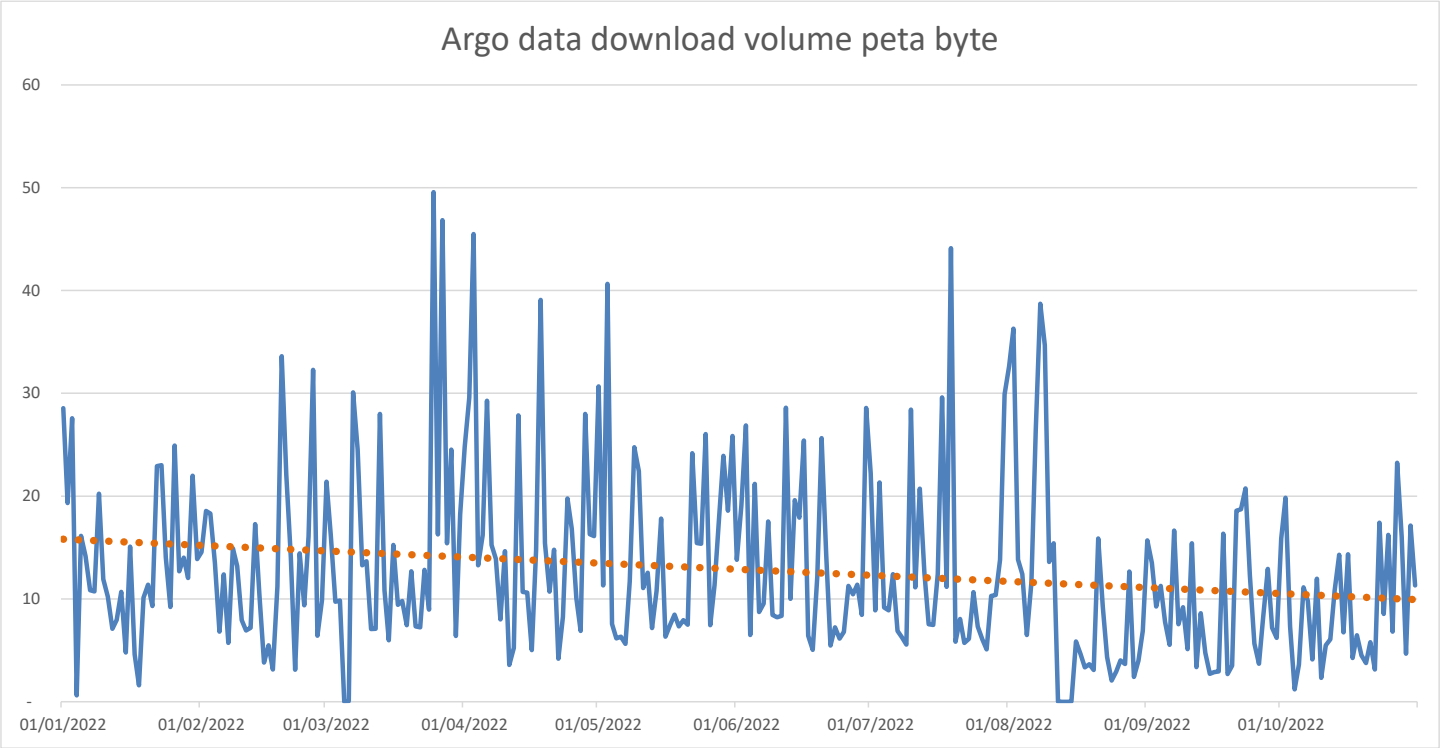
Total	*655 301 706*	

The vast majority of downloads is with ftp (80%), followed by https (20%) and a tiny fraction with Thredds or ERDDAP data services.



Distribution by countries of GDAC ftp, https erddap downloads in 2022

The majority of downloads are from China, Russia, Europe and North America.



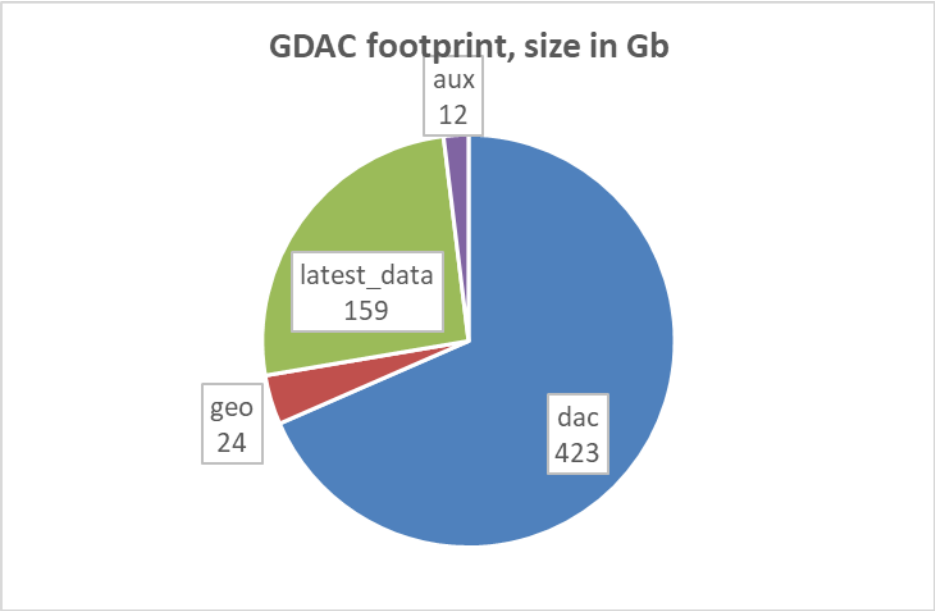
Argo FTP, HTTPS and ERDDAP downloads in 2022, an average of 20 terabytes per day with spikes up to 50 terabytes a day

3.1.3 GDAC files size

- The total number of NetCDF files on the GDAC/dac directory was 3.535.214 (+6% in one year)
- The size of GDAC/dac directory was 423 Go (+11%)
- The size of the GDAC directory was 931 Go (+26%)

More on: http://www.argodatamgt.org/Data-Mgt-Team/News/BGC-Argo-M-prof-files-no-more-distributed-on-GDAC

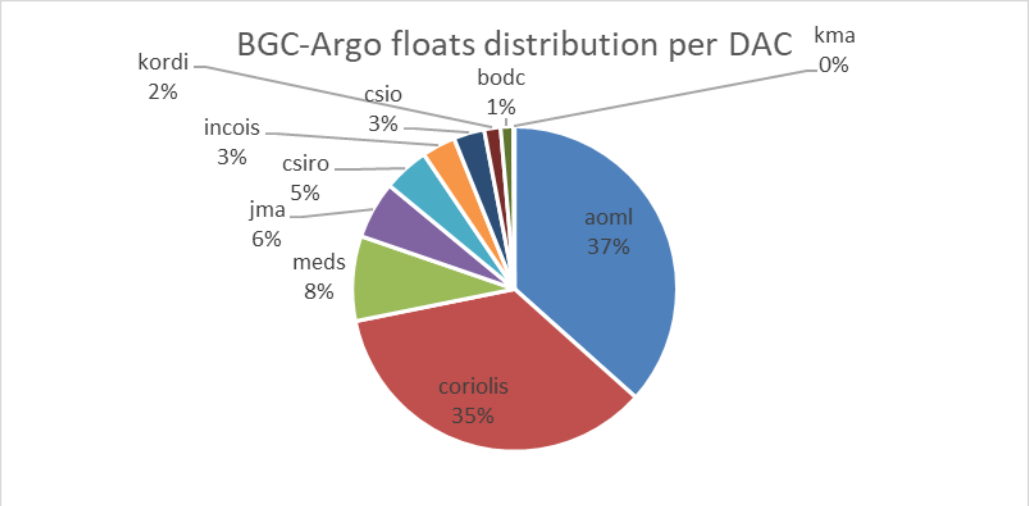
branch	GDAC size in Gb	yearly increase	N-1
dac	423	11%	382
geo	24	33%	18
latest_data	159	7%	149
aux	12	179%	4,3
gdac total	931	26%	740

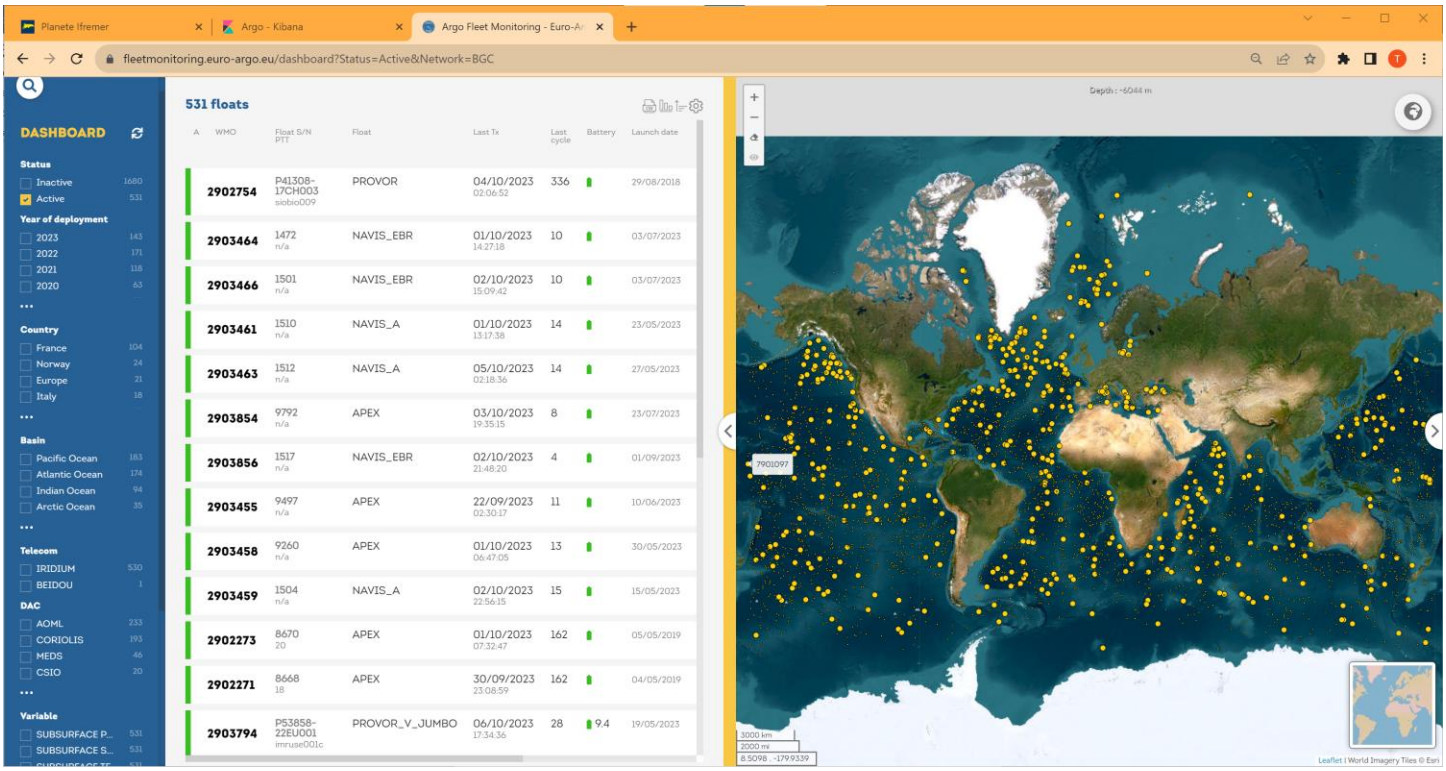


3.1.4 BGC-Argo floats

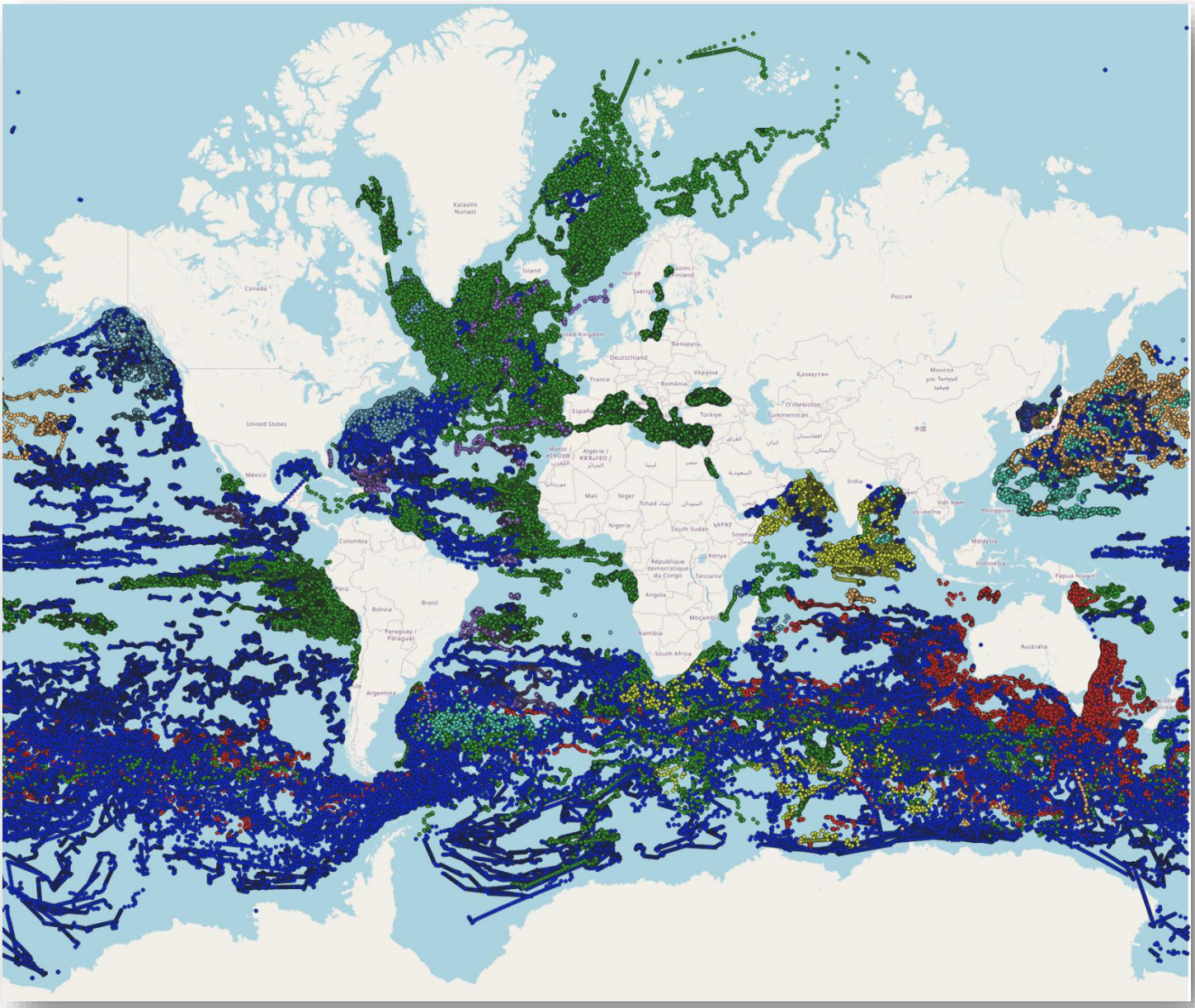
In October 2023, 294.950 BGC-Argo profiles from 2.084 floats were available on Argo GDAC. This is a fair increase compared to 2022: +12% more floats and +9% more profiles.

DAC	nb bgc float	nb bgc file
aoml	763	102 177
coriolis	734	108 574
meds	176	6 969
jma	119	20 470
csiro	95	23 271
incois	70	12 736
csio	64	11 285
kordi	34	3 426
bodc	26	5 574
kma	3	468
Total	**2084**	**294 950**



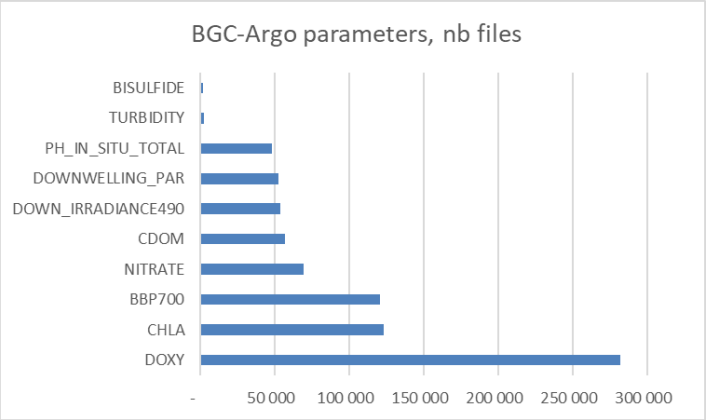


Map of 531 BGC-Argo active floats (yellow) among 2084 BGC-Argo floats (other: grey dots) from https://fleetmonitoring.euro-argo.eu/dashboard



BGC-Argo profiles, colored by DACs

BGC parameter	nb files
DOXY	281 767
CHLA	122 800
BBP700	120 284
NITRATE	69 287
CDOM	56 737
DOWN_IRRADIANCE490	53 371
DOWNWELLING_PAR	51 984
PH_IN_SITU_TOTAL	48 049
TURBIDITY	2 514
BISULFIDE	1 383

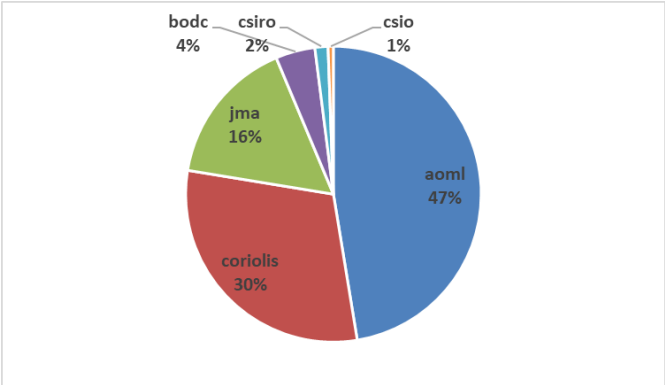


Main BGC-Argo physical parameters, number of s-profiles

3.1.5 Deep-Argo floats

The deep-Argo component of the One-Argo program is under development. A deep Argo float performs high precision observations from ocean surface to bottom, at up to 6000 meters deep.

dac	nb float	nb deep pro
aoml	165	20 318
coriolis	105	6 516
jma	56	3 870
bodc	15	850
csiro	5	478
csio	2	174
Total	348	32 206



3.2 Operations of the ftp, https and erddap servers

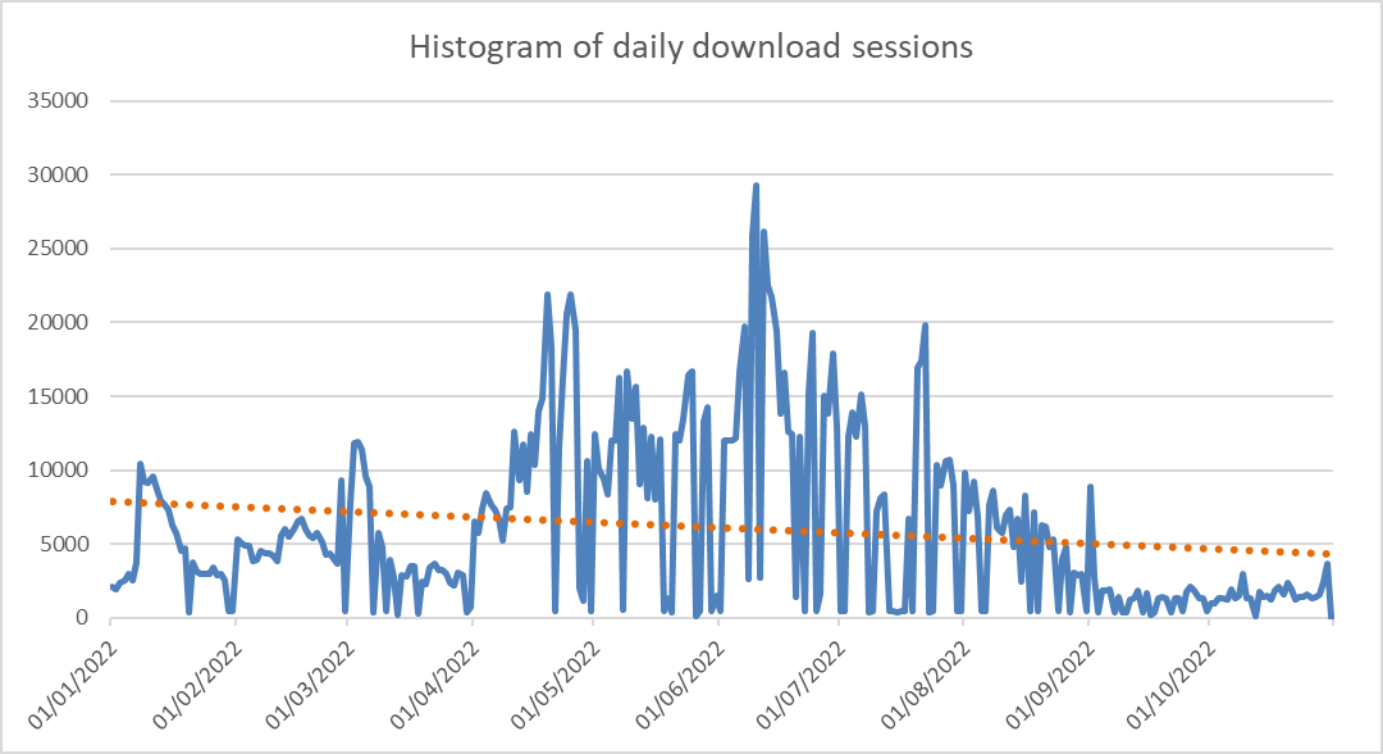
For each individual DAC, every 30 minutes, meta-data, profile, trajectory and technical data files are automatically collected from the national DACs. The 11 DACs are processed in parallel (one process launched every 3 minutes).

Index files of metadata, profiles, trajectories, technical and auxiliary data are hourly updated.

GDAC download services

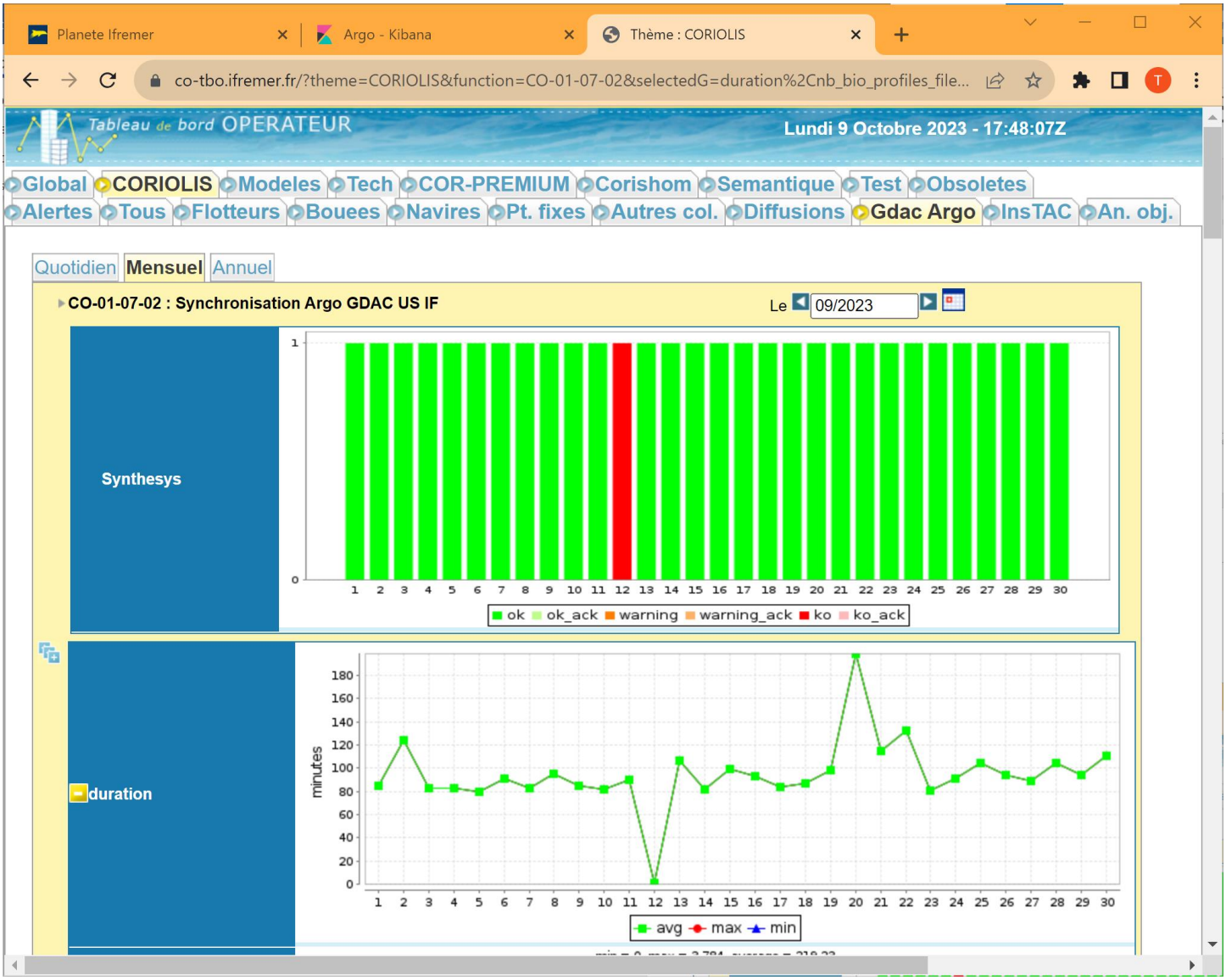
- ftp ftp://ftp.ifremer.fr/ifremer/argo
- https https://data-argo.ifremer.fr
- erddap https://erddap.ifremer.fr

There is a daily average of 6000 sessions and downloading 20 terabytes of data files. There was a huge variability in number of sessions between May and August 2022.



3.3 GDAC files synchronization

The synchronization with US-GODAE server is performed once a day at 03:55Z



Synchronization dashboard in September 2023: the daily sync. time takes on average 80 minutes, with a failure on September 12th

3.4 Download services monitoring

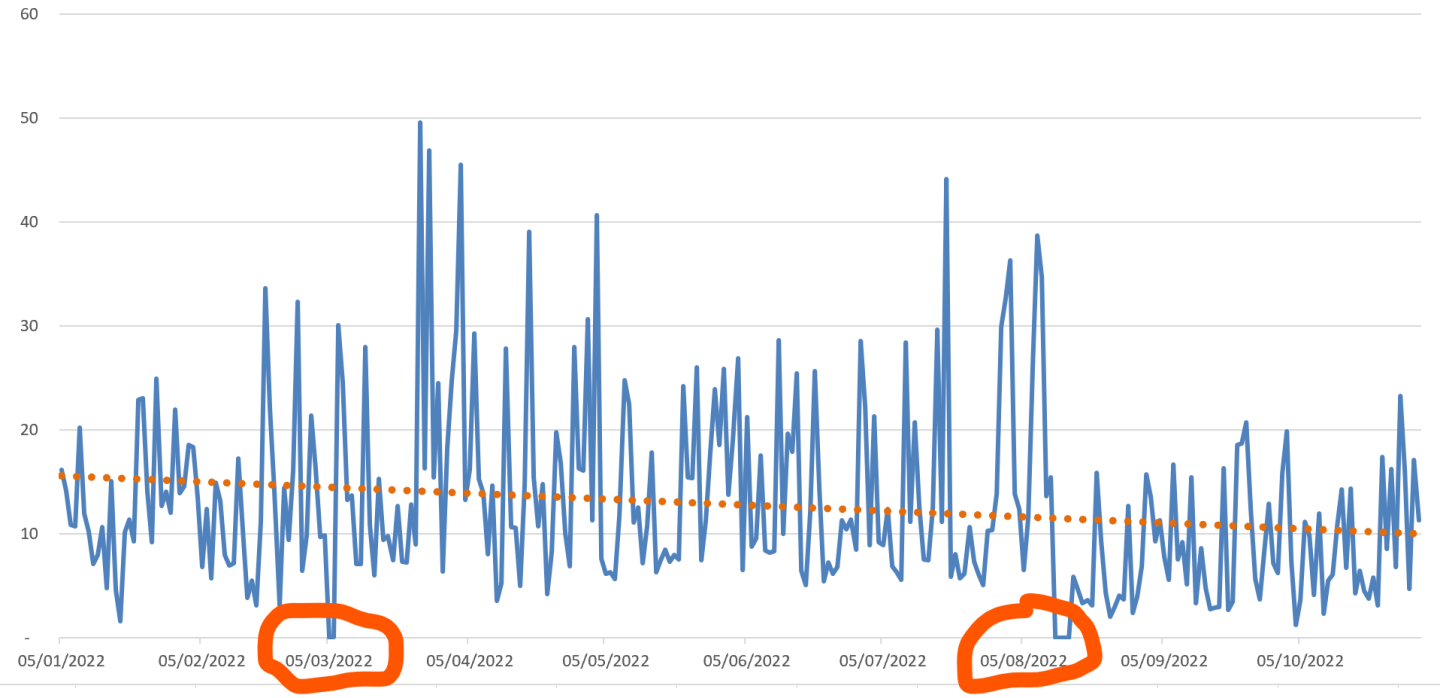
This chapter will be updated in a version 1.1 of the report.

Semaphore is used to monitor the data distribution activity.

In 2022, according to the daily data download volume statistics, there were two period of poor performances: 2 days in March 2022 and 4 days in August 2022.

A poor performance is an abnormally low volume of data downloaded by Argo users (less than 0,1 terabyte of data).

Day	Volume terabyte
05/03/2022	0,013
06/03/2022	0,069
12/08/2022	0,007
13/08/2022	0,005
14/08/2022	0,003
15/08/2022	0,010



Six days of poor data download service in year 2022, less than 0,1 terabyte of data

3.5 Grey list

The GDAC hosts a grey list of the floats which are automatically flagged before any automatic or visual quality control. **The greylist has 1659 core-Argo entries** (October 2023), compared to 1519 entries one year ago.

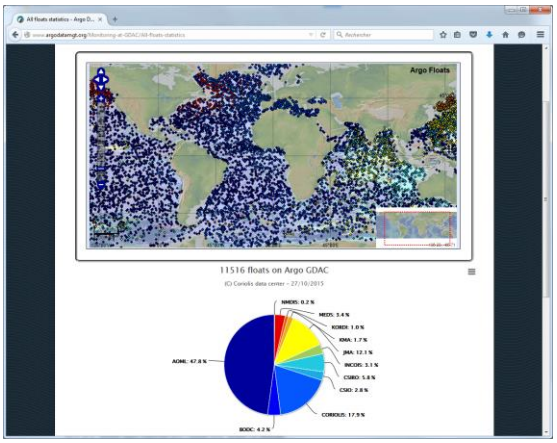
All floats core-Argo	
DAC	NB_FLOAT
aoml	964
coriolis	192
jma	127
bodc	124
csiro	62
csio	28
meds	24
incois	24
kma	13
kordi	3
Total	1561

Distribution of greylist entries per DAC and per parameter

Parameter	nb entries
PSAL	1560
TEMP	160
PRES	121
DOXY	71
BBP700	51
CHLA	37
CDOM	24
BBP532	11
NITRATE	9
DOWN_IRRADIANCE380	4
DOWNWELLING_PAR	4
DOWN_IRRADIANCE490	4
DOWN_IRRADIANCE412	4
PH_IN_SITU_TOTAL	3
CP660	3
PH_IN_SITU_FREE	1
TURBIDITY	1

3.6 Statistics on GDAC content

The following graphics display the distribution of data available from GDAC, per float or DACs. These statistics are daily updated on: http://www.argodatamgt.org/Monitoring-at-GDAC



3.7 Mirroring data from GDAC: rsync service

In July 2014, we installed a dedicated rsync server called vdmzrs.ifremer.fr described on:

- http://www.argodatamgt.org/Access-to-data/Argo-GDAC-synchronization-service

This server provides a synchronization service between the "dac" directory of the GDAC with a user mirror. From the user side, the rysnc service:

- Downloads the new files
- Downloads the updated files
- Removes the files that have been removed from the GDAC
- Compresses/uncompresses the files during the transfer
- Preserves the files creation/update dates
- Lists all the files that have been transferred (easy to use for a user side post-processing)

Examples

Synchronization of a particular float

- rsync -avzh --delete vdmzrs.ifremer.fr::argo/coriolis/69001 /home/mydirectory/...

Synchronization of the whole dac directory of Argo GDAC

- rsync -avzh --delete vdmzrs.ifremer.fr::argo/ /home/mydirectory/...

3.8 Argo DOI, Digital Object Identifier on monthly snapshots

A digital object identifier (DOI) is a unique identifier for an electronic document or a dataset. Argo data-management assigns DOIs to its documents and datasets for two main objectives:

- Citation: in a publication the DOI is efficiently tracked by bibliographic surveys
- Traceability: the DOI is a direct and permanent link to the document or data set used in a publication
- More on: http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier

Since July 2019, the DOI monthly snapshot of Argo data is a compressed archive (.gz) that contains distinct core-Argo tar files and BGC-Argo tar files. A core-Argo user can now ignore the voluminous BGC-Argo files.

Argo documents DOIs

- Argo User's manual: http://dx.doi.org/10.13155/29825

Argo GDAC DOI

- Argo floats data and metadata from Global Data Assembly Centre (Argo GDAC) http://doi.org/10.17882/42182

3.9 Argo GDAC available from Amazon S3 cloud service

As part of "action 48 - admt-21 Argo data on the cloud" we contacted Amazon and its ASDI program (Amazon Sustainability Data Initiative). There is now a prototype of GDAC on the cloud, updated daily in an S3 bucket.

- https://registry.opendata.aws/argo-gdac-marinedata/

Argo GDAC data is directly available, it is no more necessary to download files before using them. Anyone from anywhere can run something like a python script with direct access to an Argo cloud prototype hosted on Amazon.