

Argo data management report 2017

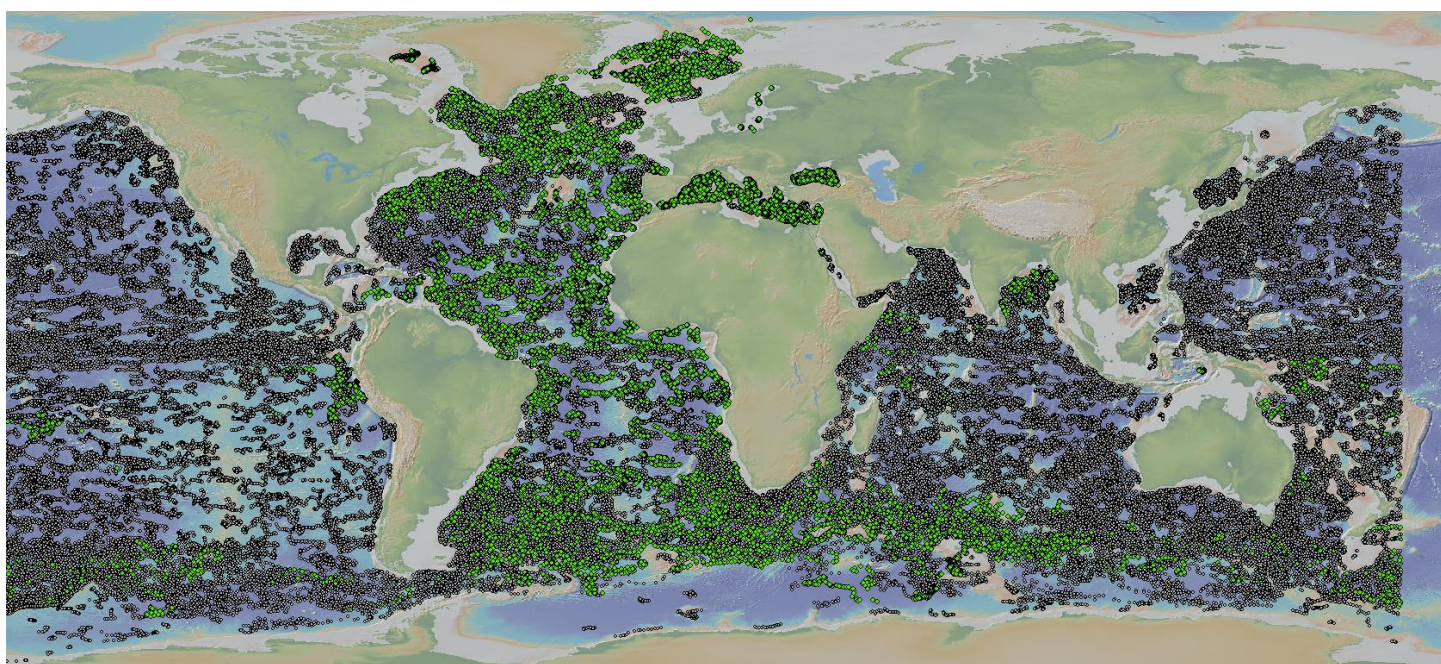
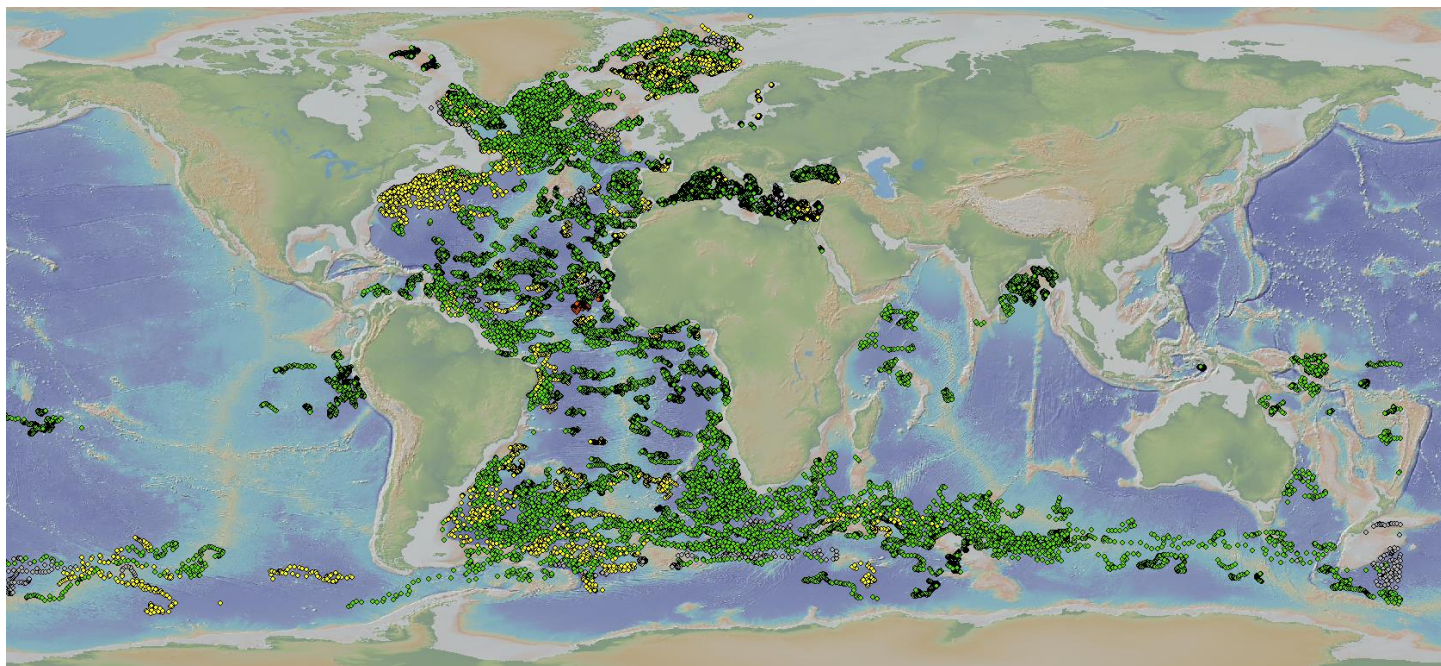
Coriolis DAC & GDAC

Data Assembly Centre and Global Data Assembly Centre

Annual report October 2017

Version 1.1

November 10th, 2017



1 DAC status

This report covers the activity of Coriolis data centre for a one-year period from November 1st 2016 to October 30th 2017.

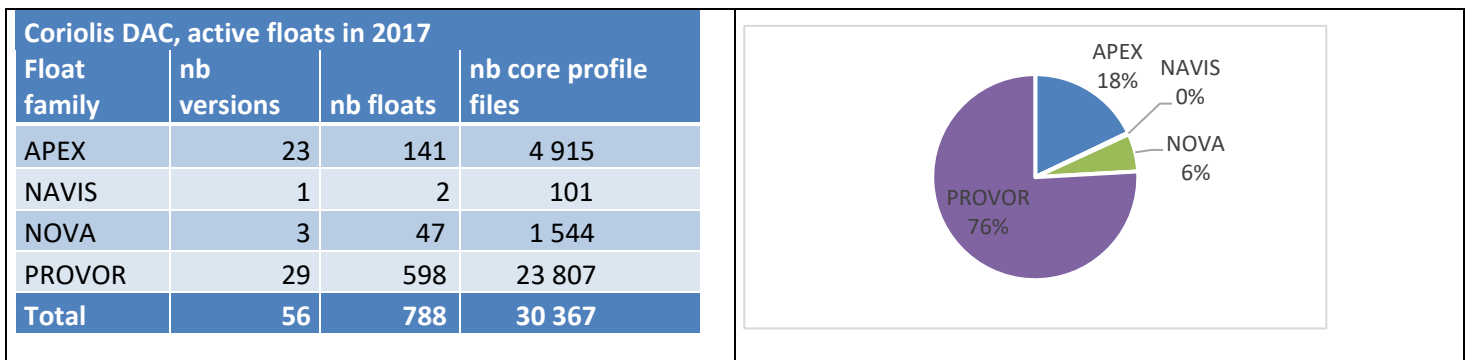
1.1 Data acquired from floats

1.1.1 Active floats for the last 12 months

These last 12 months, **30 367 profiles from 788 active floats** were collected, controlled and distributed.

Compared to 2016, **the number of profiles decreased by 1%, the number of floats increased by 2%**. These figures show a fair stability in Coriolis DAC activity.

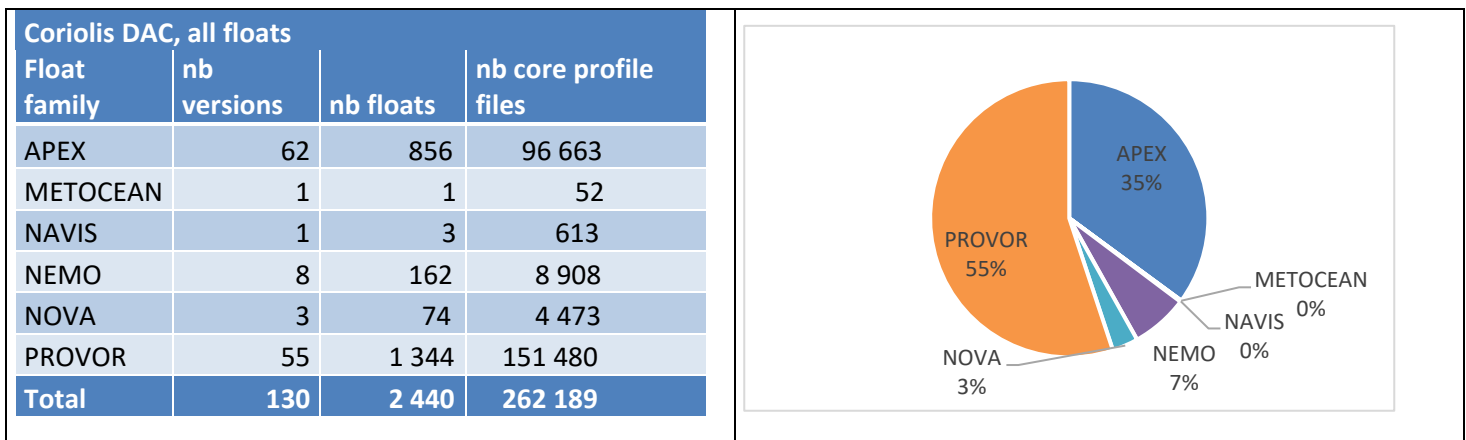
The 788 floats managed during that period had 56 versions of data formats.

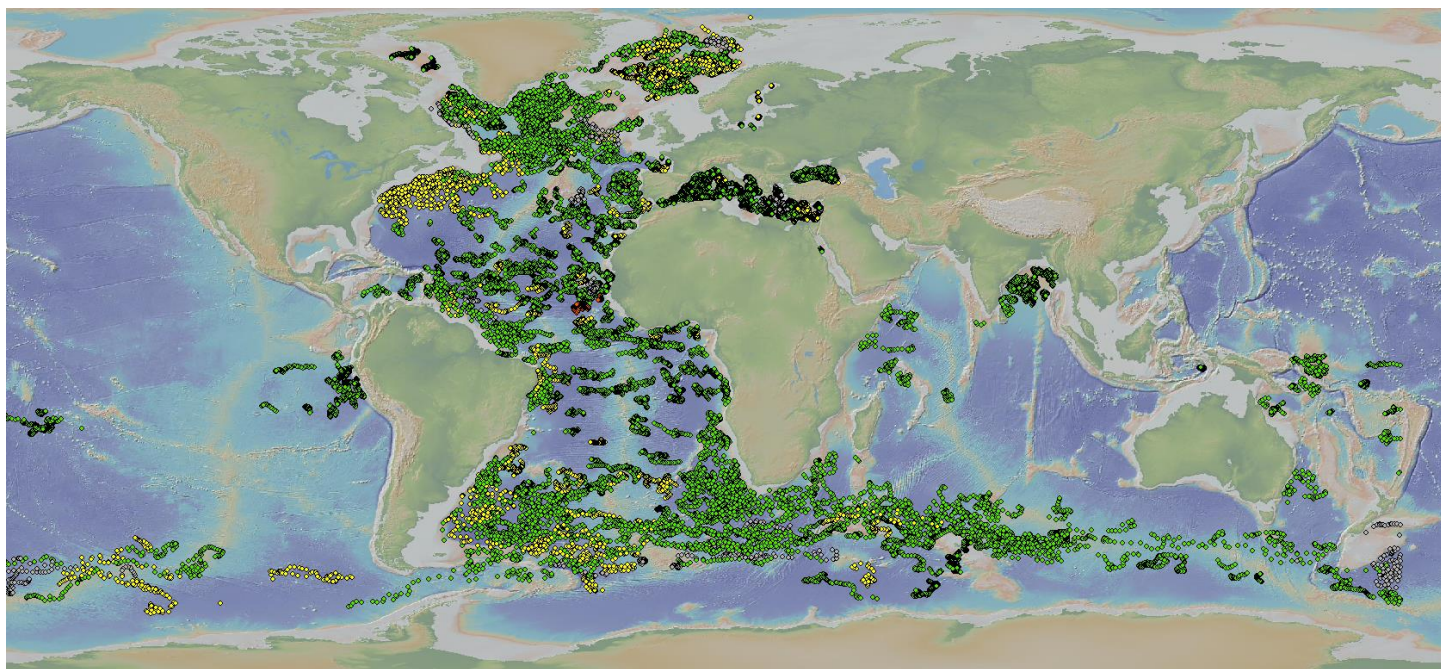


1.1.2 All floats managed by Coriolis DAC

Coriolis DAC manages a total of 2 440 floats with 130 versions, from 6 families.

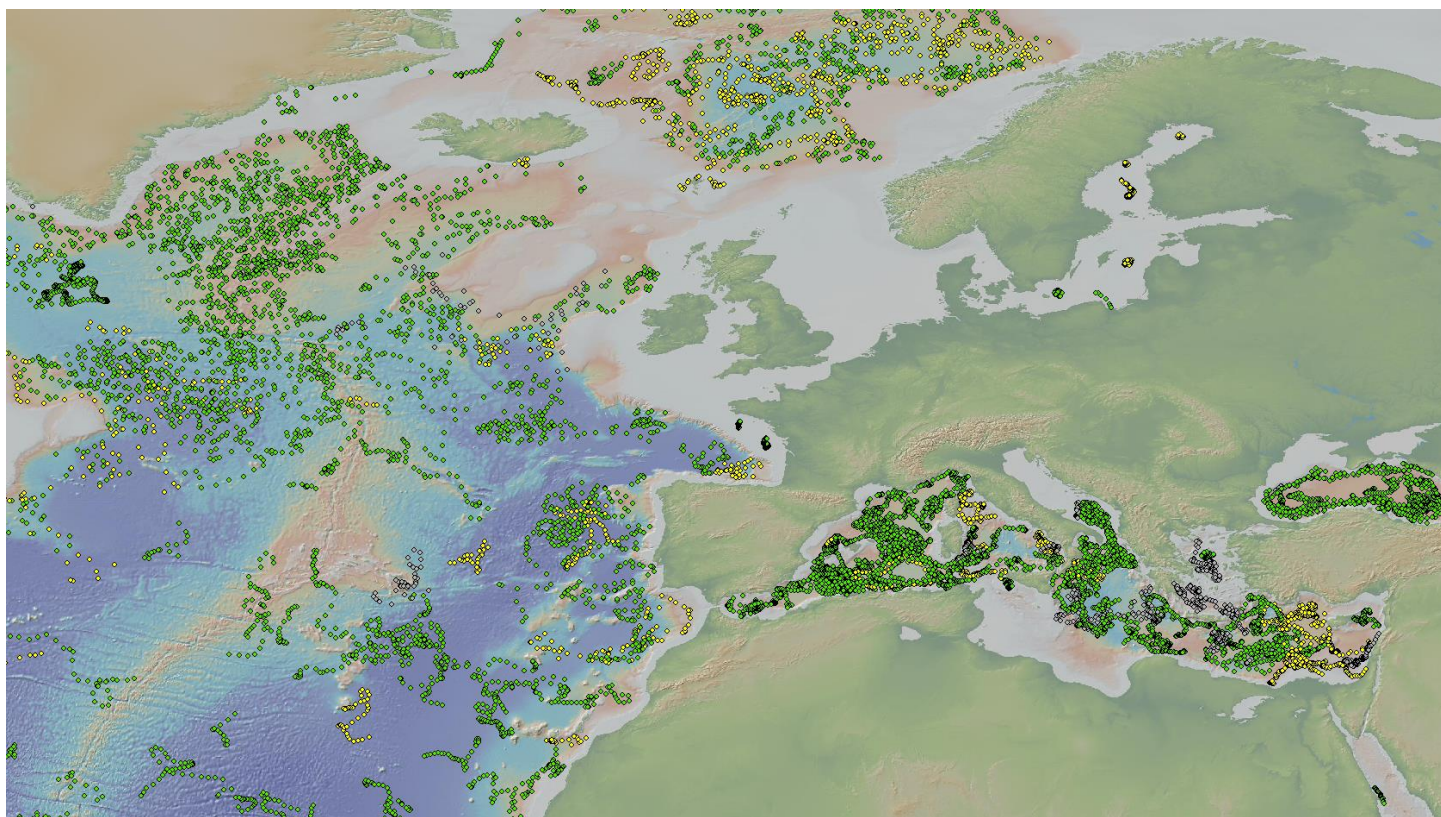
These floats reported 262 198 core Argo vertical profiles.



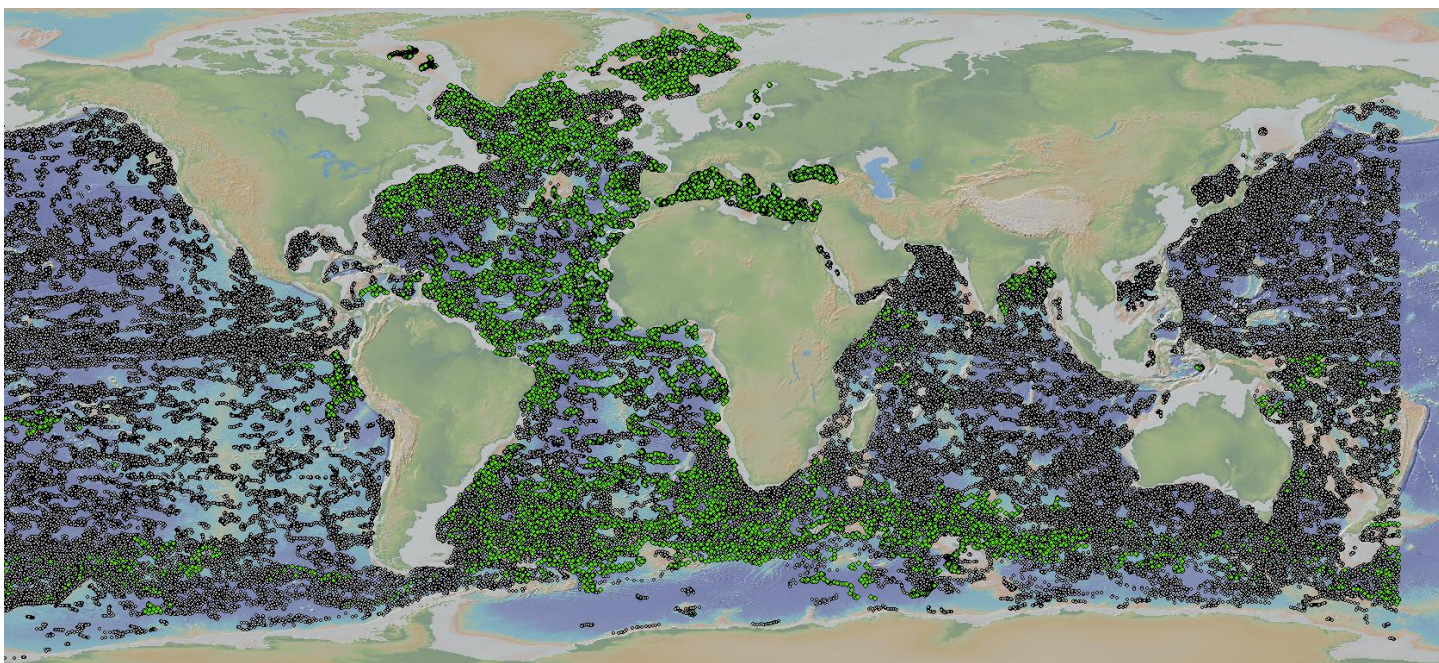


Map of the 30.367 profiles from 788 active floats decoded by Coriolis DAC this current year

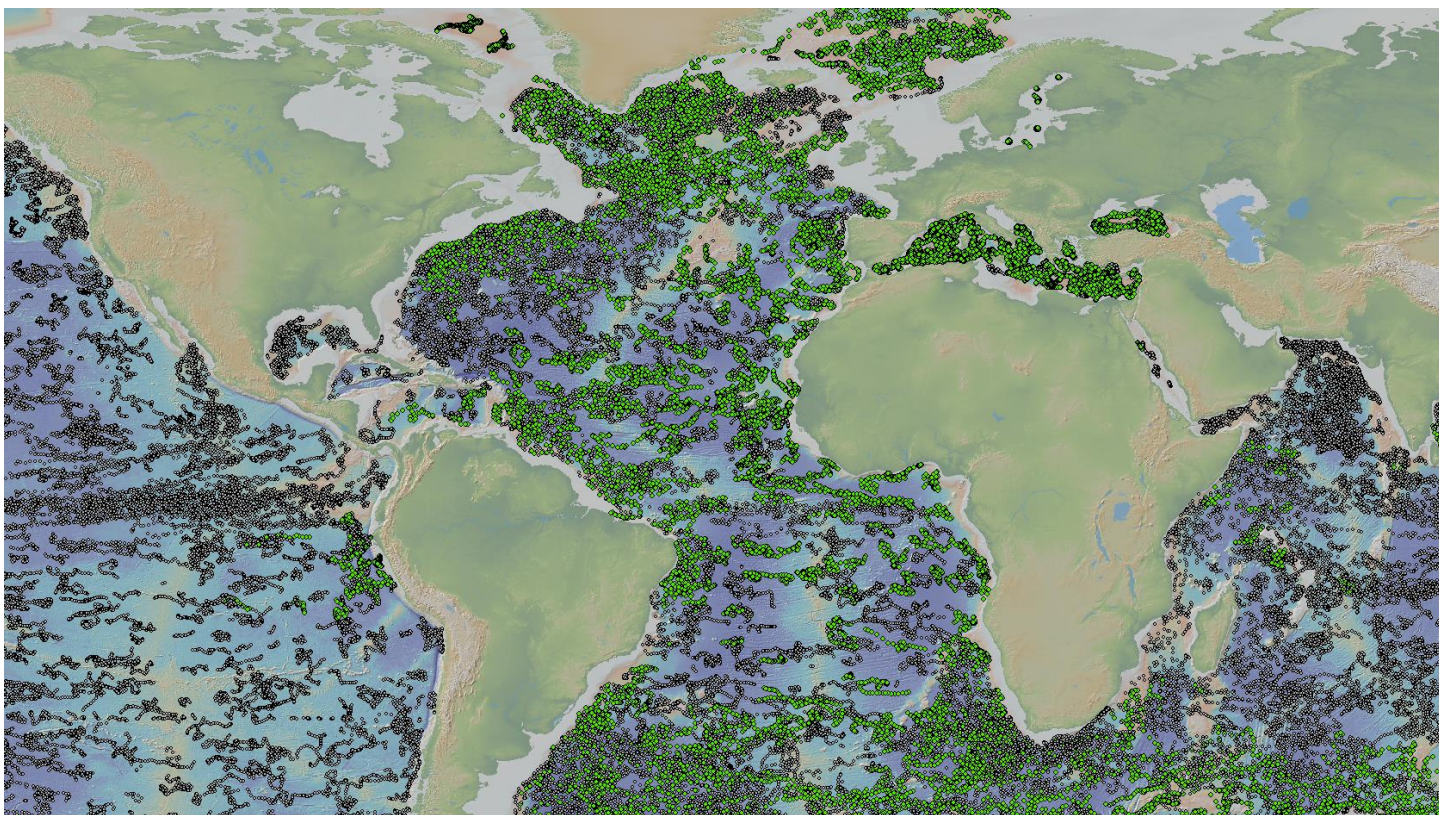
Apex Navis Nemo Nova Provor



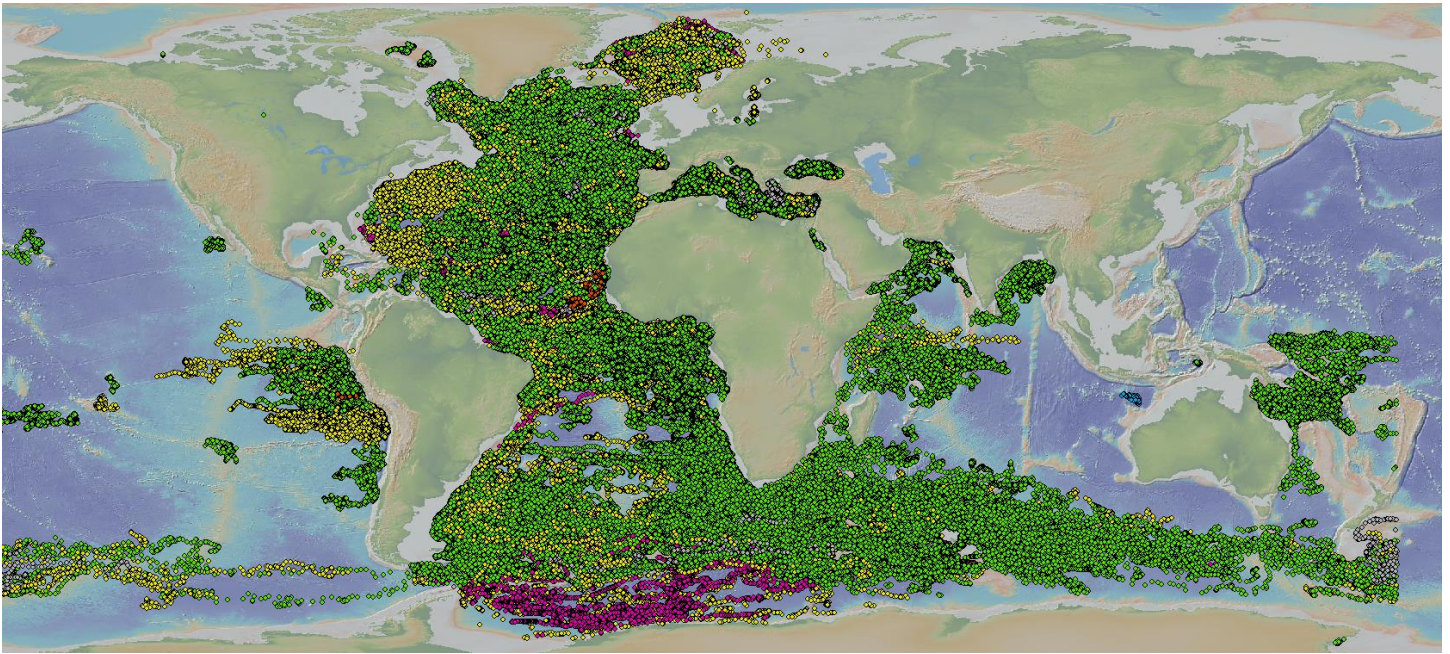
Map of active floats managed by Coriolis this current year, zoom on north Atlantic area



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)



Atlantic map active floats profiles from Coriolis DAC this current year, among the other DAC's profiles (Coriolis: green, other DACs: grey)



Map of the 262.189 profiles from 2.240 floats managed by Coriolis DAC

Apex Metocean Navis Nemo Nova Provior

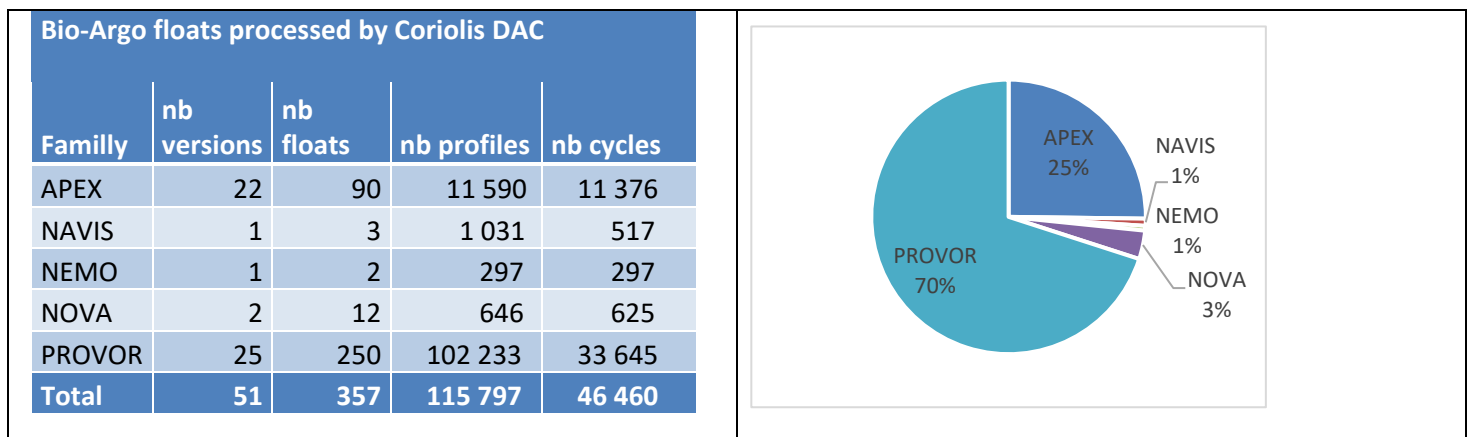
1.1.3 BGC-Argo sensors on Coriolis floats

The data processing chain based on Matlab to manage data and metadata from Coriolis BGC-floats is continuously improved. These are advanced types of floats performing bio-geo-chemical measurements.

Coriolis DAC manages 357 BGC-Argo floats from 5 families and 51 instrument versions. They performed 46.460 cycles.

The data processing chain is freely available:

- *Coriolis Argo floats data processing chain, version 20170328_010a. SEANOE. <http://doi.org/10.17882/45589>*



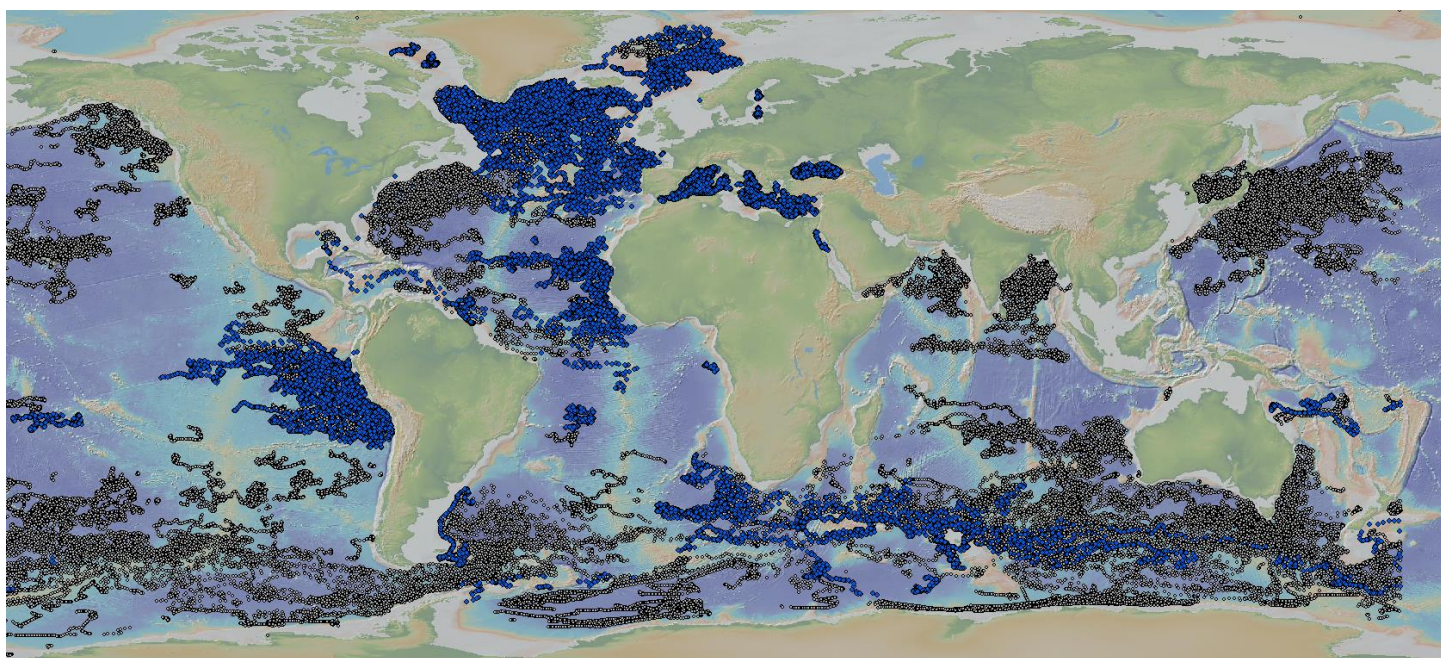
General characteristics

- Iridium rudics bi-directional communication or Argos
- Six sensors are fitted on the floats

- AANDERAA_OPTODE_4330 Aandera oxygen sensor
- C_ROVER Wetlabs transmissiometer
- ECO_PUCK Wetlabs fluorometer turbidity, scattering
- SATLANTIC_OCR504 Satlantic Irradiance sensor
- SBE41CP Seabird CTD sensor
- SUNA_V2 Satlantic nitrate sensor
- 41 parameters managed: core-argo, b-argo, i-argo parameters
bbp532, bbp700, beta_backscattering532, beta_backscattering700, bisulfide, bphase_doxy, c1phase_doxy, c2phase_doxy, cdom, chla, cp660, down_irradiance380, down_irradiance412, down_irradiance490, downwelling_par, doxy, doxy2, fit_error_nitrate, fluorescence_cdom, fluorescence_chla, frequency_doxy, humidity_nitrate, molar_doxy, mtime, nitrate, phase_delay_doxy, phase_delay_doxy2, ppox_doxy, raw_downwelling_irradiance380, raw_downwelling_irradiance412, raw_downwelling_irradiance490, raw_downwelling_par, rphase_doxy, temp_cpu_chla, temp_doxy, temp_doxy2, temp_nitrate, temp_spectrophotometer_nitrate, tphase_doxy, uv_intensity_dark_nitrate, uv_intensity_nitrate

These parameter include chlorophyll, turbidity, CDOM, back-scattering, UV, nitrate, bisulfide, pH, radiance, irradiance, PAR.

| Main parameters | nb profiles |
|---------------------------|-------------|
| DOXY (oxygen) | 32716 |
| BBP700 | 23397 |
| BETA_BACKSCATTERING700 | 23397 |
| CHLA (chrolophyll) | 23397 |
| DOWN_IRRADIANCE380 | 20941 |
| DOWNWELLING_PAR | 20941 |
| CDOM | 20627 |
| NITRATE | 6411 |
| PPOX_DOXY (oxygen in air) | 520 |
| BISULFIDE | 184 |



Map of the 357 bio-Argo floats managed by Coriolis DAC (grey dots: the others DACs bio-Argo floats). They measure parameters such as chlorophyll, turbidity, CDOM, back-scattering, UV, nitrate, bisulfide, pH, radiance, irradiance, PAR.



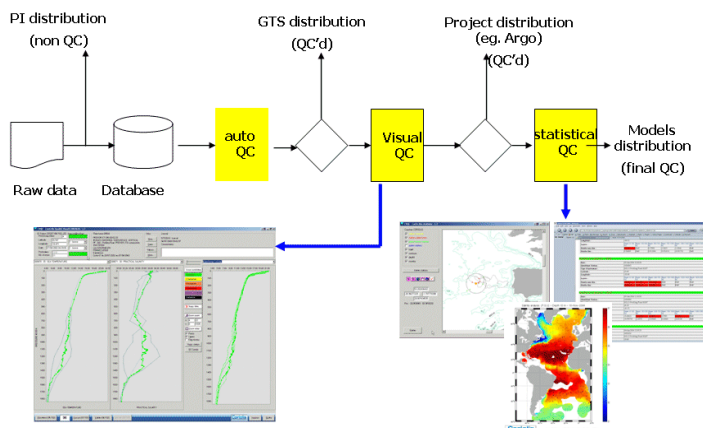
© Antoine Poteau, Observatoire Océanologique de Villefranche (CNRS/UPMC)
Deployments of a bio-argo Provor in Ligurian sea

1.2 Data issued to GTS

All profiles processed by Coriolis are distributed on the GTS by way of Meteo-France. This operation is automatically performed. After applying the automatic Argo QC procedure, the Argo profiles are inserted on the GTS every 2 hours. Argo profiles are inserted on the GTS 365 days per year, 24 hours a day.

The profile files are sent as TESAC and BUFR messages by way of Meteo-France. Meteo-France accept Coriolis as valid BUFR messages and circulate them on neighbouring nodes.

Once a day, floats data that are less than 21 days old are checked in an objective analysis (ISAS) that triggers alert and visual inspection for suspicious observations.

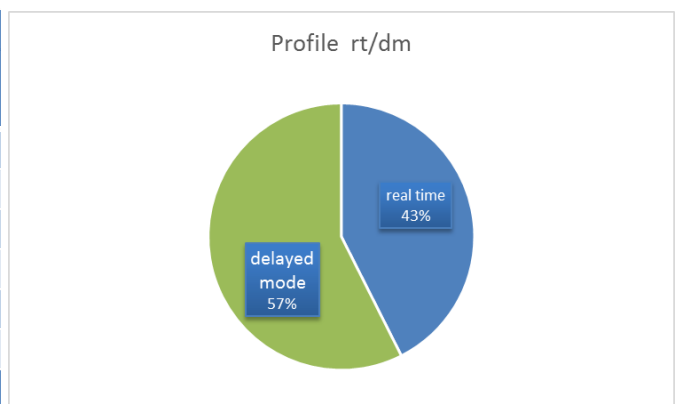


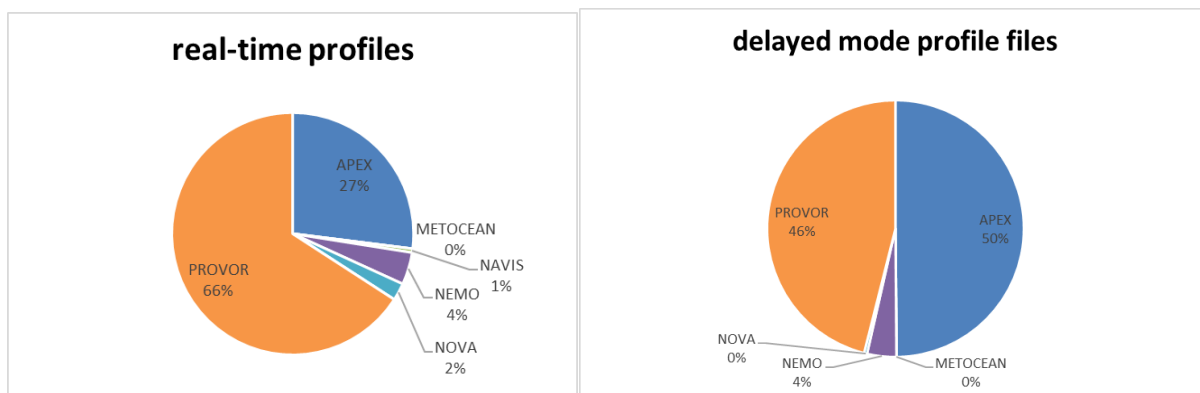
CORIOLIS DAC: Argo data flow

1.3 Data issued to GDACs after real-time QC

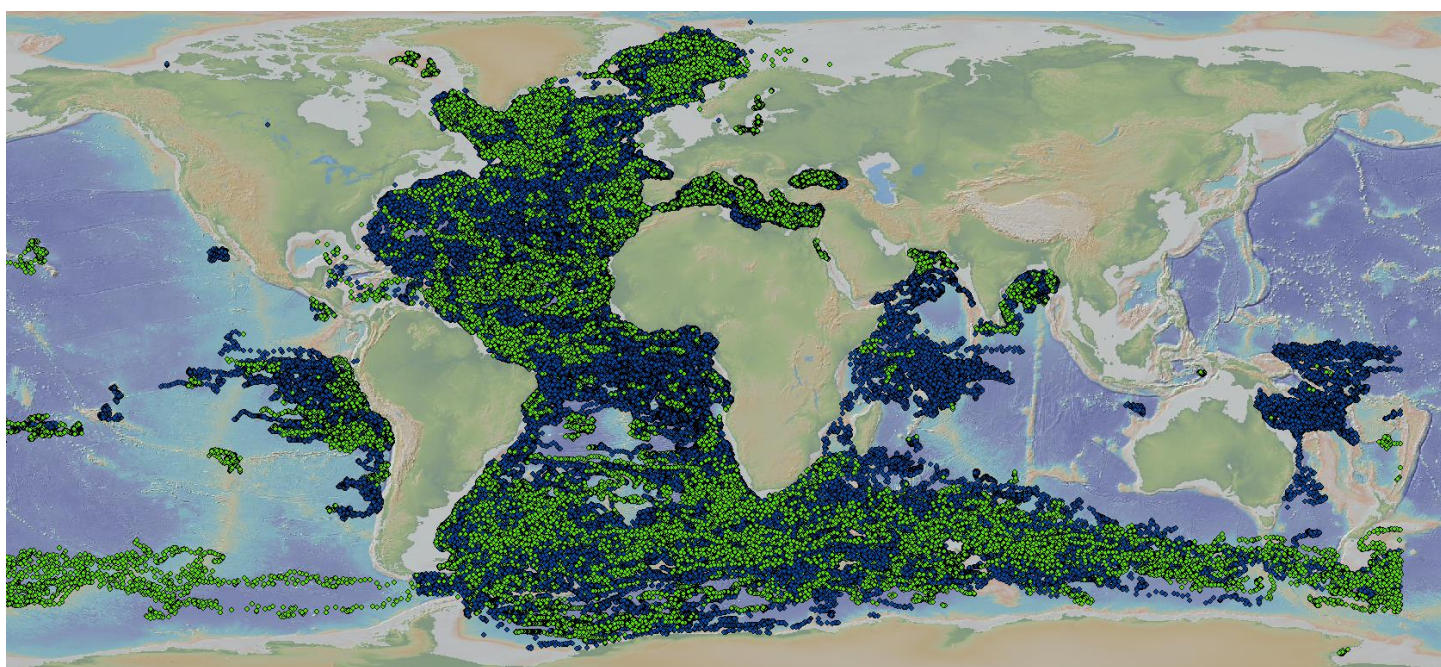
All meta-data, profiles, trajectory and technical data files are sent to Coriolis and US-GODAE GDACs. This distribution is automated.

| All Coriolis floats, number of profile files on GDAC | | | | |
|--|--------------|----------------|---------------|----------------|
| Family | nb floats | nb profil | RT profil | DM profiles |
| APEX | 829 | 90 533 | 25 918 | 64 615 |
| METOCEAN | 1 | 52 | - | 52 |
| NAVIS | 3 | 488 | 488 | |
| NEMO | 163 | 8 911 | 4 141 | 4 770 |
| NOVA | 52 | 2 727 | 2 233 | 494 |
| PROVOR | 1 162 | 122 974 | 63 264 | 59 710 |
| Total | 2 210 | 225 685 | 96 044 | 129 641 |





Distribution of Coriolis DAC real-time and delayed mode profiles (96 044 profiles – 225 685 profiles)



Map of real-time profiles and delayed mode profiles
Real time: green dots, delayed mode: blue dots

1.4 Data issued for delayed mode QC

Delayed mode profiles

All profile files are sent to PIs for delayed QC. Most of the Atlantic data handled by Coriolis are checked by the European project Euro-Argo.

Preparation of Argo delayed mode trajectories

The delayed mode trajectories derived from Andro trajectory product were produced in version 3.0. Their conversion to V3.1 trajectory format will be performed in 2016.

The 1442 delayed mode trajectories files are available from:

- <ftp://ftp.ifremer.fr/ifremer/argo/etc/coriolis-custom/argo-andro-data/data/dac/coriolis/>

The Andro trajectory TRAJ3 files are available for most of the DACs. Each DAC may decide to use these files to provide delayed mode trajectory on GDAC.

Coriolis DAC will use these files as its delayed mode trajectories for old floats versions.

1.5 Delayed mode data sent to GDACs

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

- A total of **57 013 new or updated delayed mode profiles** was sent to GDACs this year.
- **A total of 179 361 delayed mode profiles** were sent to GDACs since 2005.
The number of delayed mode profiles increased by 27% this year.

1.6 Web pages

The web site of the French DAC is available at:

- <http://www.coriolis.eu.org/Observing-the-Ocean/ARGO>

This web page describes all Argo floats:

- <http://www.ifremer.fr/co-argoFloats/>
 - Individual float description and status (meta-data, geographic map, graphics : section, overlaid, waterfall, t/s charts)
 - Individual float data (profiles, trajectories)
 - FTP access
 - Data selection tool
 - Global geographic maps, GoogleEarth maps
 - Weekly North Atlantic analyses (combines Argo data and other measurements from xbt, ctd, moorings, buoys)

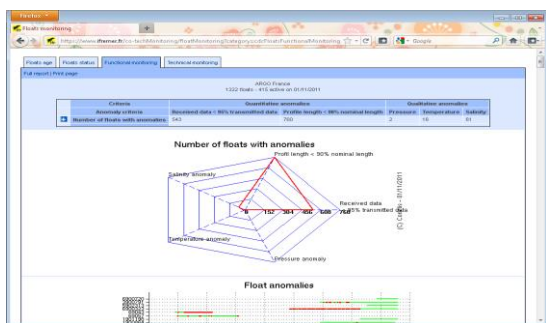
This web page describes all Argo floats interoperability services from Coriolis:

- <http://www.coriolis.eu.org/Data-Products/Data-Delivery/Argo-floats-interoperability-services>

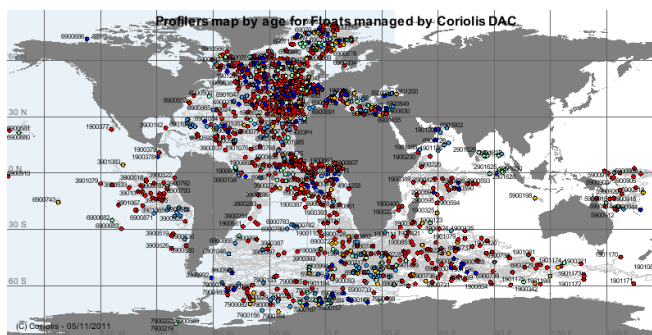
- Display an individual float's data and metadata in HTML or XML format
- Display all Argo floats, display a group of floats
- Argo profiles and trajectories data selection (HTML or XML)
- All individual float's metadata, profile data, trajectory data and technical data
- Argo profiles data on OpenDAP, OGC-WCS and http
- Argo data through RDDAP data server (www.ifremer.fr/erddap)
- Argo data through Oceanotron data server
- Argo profiles data through GCMD-DIF protocol
- Argo data through RDF and OpenSearch protocols
- Display Argo profiles and trajectories with GoogleEarth

Some pages of Coriolis web site are dedicated to technical monitoring:

- <http://www.coriolis.eu.org/Data-Products/At-sea-monitoring>



Example 1: technical monitoring of Argo-France floats



Example 2: age map of floats managed by Coriolis DAC.

Within Euro-Argo ERIC (European Research Infrastructure Consortium), a dashboard with alerts on anomalies was developed: <http://www.ifremer.fr/argoMonitoring/floatMonitoring/385>

| | NB FLOATS (%) | NB CYCLES (%) |
|--|---------------|----------------|
| INFO | 406 (38.05%) | 20692 (13.74%) |
| FLAG_MissingCycle_LOGICAL | 218 (20.43%) | 19471 (12.93%) |
| FLAG_MissingVerticalProfile_LOGICAL | 17 (1.59%) | 17 (0.01%) |
| FLAG_VoltageDrop_LOGICAL | 215 (20.15%) | 1210 (0.80%) |
| DATA TRANSMISSION | 949 (88.94%) | 39250 (24.07%) |
| FLAG_CtdDataTransmissionIncomplete_LOGICAL | 495 (46.39%) | 4138 (2.73%) |
| FLAG_ParameterDataTransmissionIncomplete_LOGICAL | 12 (1.12%) | 22 (0.01%) |
| FLAG_TechnicalDataTransmissionIncomplete_LOGICAL | 191 (17.90%) | 268 (0.18%) |
| FLAG_HydraulicDataTransmissionIncomplete_LOGICAL | 0 (0%) | 0 (0%) |
| FLAG_SatelliteCoverageProblem_LOGICAL | 0 (0%) | 0 (0%) |
| FLAG_PercentageMessagesGoodCRC_LOGICAL | 685 (64.20%) | 24679 (16.39%) |
| FLAG_UnusualTransmissionTime_LOGICAL | 577 (54.08%) | 16828 (11.18%) |
| CONDITIONING | 0 (0%) | 0 (0%) |

Within ENVRIPUS EU project, an Argo big data demonstration: <http://co-discovery-demo.ifremer.fr/euro-argo/>
 All Argo data is indexed and available through an Elasticsearch API.
 More on https://www.youtube.com/watch?v=PKU_JcmSskw

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 | Archive GDAC Argo | | | | | UNDERWAY-LOCKED 2017-11-25T02:07:01Z |
| CO-03-07-01 | Argo files controler | | | | | OK 2017-10-13T08:00:46Z |
| CO-05-08-11 | Argo grey list diffuser | | ● | ● | ● | OK 2017-11-24T11:05:02Z |
| CO-03-07-01-02 | Argo stat controler | | | | ● | OK 2017-11-22T01:02:21Z |
| CO-01-07-08 | Collecte Argo Coriolis EDAC | ● | ● | ● | ● | OK 2017-11-25T09:45:04Z |
| CO-01-07-03 | Collecte Argo DAC - FTP | ● | ● | ● | ● | OK 2017-11-25T10:09:04Z |
| CO-01-07-01-02 | Collecte Argo DAC - Table argo index profiles | ● | ● | ● | ● | OK 2017-11-25T09:58:50Z |
| CO-01-07-01-aoml | Collecte Argo DAC - aoml | ● | ● | ● | ● | OK 2017-11-25T10:00:04Z |
| CO-01-07-01-bodc | Collecte Argo DAC - bodc | ● | ● | ● | ● | OK 2017-11-25T10:01:02Z |
| CO-01-07-01-coriolis | Collecte Argo DAC - coriolis | ● | ● | ● | ● | OK 2017-11-25T10:02:21Z |
| CO-01-07-01-csio | Collecte Argo DAC - csio | ● | ● | ● | ● | OK 2017-11-25T10:03:02Z |
| CO-01-07-01-csiro | Collecte Argo DAC - csiro | ● | ● | ● | ● | OK 2017-11-25T10:04:03Z |
| CO-01-07-01-incois | Collecte Argo DAC - incois | ● | ● | ● | ● | OK 2017-11-25T10:05:02Z |
| CO-01-07-01-jma | Collecte Argo DAC - jma | ● | ● | ● | ● | OK 2017-11-25T10:06:05Z |
| CO-01-07-01-kma | Collecte Argo DAC - kma | ● | ● | ● | ● | OK 2017-11-25T10:07:03Z |
| CO-01-07-01-kordi | Collecte Argo DAC - kordi | ● | ● | ● | ● | OK 2017-11-25T10:08:02Z |
| CO-01-07-01-meds | Collecte Argo DAC - meds | ● | ● | ● | ● | OK 2017-11-25T10:09:03Z |
| CO-01-07-01-nmdis | Collecte Argo DAC - nmdis | ● | ● | ● | ● | OK 2017-11-25T10:10:02Z |
| CO-01-07-06-aoml | Collecte Argo DAC BDD - aoml | ● | ● | ● | ● | OK 2017-11-25T09:42:07Z |
| CO-01-07-06-bodc | Collecte Argo DAC BDD - bodc | ● | ● | ● | ● | OK 2017-11-25T09:42:03Z |

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:

- French model Mercator (global operational model)
- French model MARC (regional operational model)
- French model Soap (navy operational model)
- EU Copernicus models (Foam, Topaz, Moon, Noos)
- EuroGoos projects

Argo projects: this year, Coriolis data centre performed float data management for 55 Argo scientific projects and 55 PIs (Principal Investigators).

List of Coriolis scientific PIs and project names

| Project name | nb floats |
|--------------|-----------|
| euro-argo | 584 |
| coriolis | 397 |

| | |
|-------------|----|
| bsh | 83 |
| argo-bsh | 79 |
| argo italy | 76 |
| naos | 69 |
| remocean | 62 |
| mocca | 60 |
| goodhope | 47 |
| mocca-eu | 43 |
| pirata | 27 |
| gmmc | 21 |
| argomed | 16 |
| rrex asfar | 15 |
| sagar | 12 |
| dap | 11 |
| rrex | 10 |
| argo norway | 10 |
| argo greece | 10 |

List of projects with more than 10 active floats

List of project with less than 10 active floats: argo_fin, geovide, amop, soclim, bwr, argo_spain, ovide, outpace, ticmoc, mocca-germany, naos-canada, gmmc_cnes, brazilian navy argo program, naos-france, mocca-italy, cienperu, argo italy , morsea, argo-poland, moose, vsf, eaims, narval, mocca-poland, mafia, mocca-ger, argo bulgary, sri_lanka, aspex, argo spain, e-aims, argo brazil, cnes, proteusmed, asfar, argo-finland, mocca-eu, dekosim (metu), bioargo-italy, opportunity (sail), upsen, ge moose, naos,pirata, perseus, socib, mocca-ned, argo-italy, dekosim, lefe, bioargo italy, argo bsh, bide, mocca-netherlands, euroargo, peacetime, congas

| PI name | nb floats |
|----------------------|-----------|
| birgit klein | 142 |
| pierre-marie poulain | 95 |
| christine coatanoan | 92 |
| sabrina speich | 64 |
| virginie thierry | 63 |
| bernard bourles | 28 |
| christophe maes | 24 |
| herve claustre | 23 |
| holger giese | 22 |
| fabrizio d'ortenzio | 17 |
| romain cancouët | 16 |
| fabien durand | 15 |
| andreas sterl | 13 |
| jose lluis pelegri | 12 |
| marcel babin | 12 |
| dimitris kassis | 11 |
| waldemar walczowski | 11 |
| kjell arne mork | 10 |

| | |
|----------------------|----|
| laurent coppola | 10 |
| pedro velez belchi | 10 |
| jean-baptiste sallee | 9 |
| peter brandt | 8 |
| tero purokoski | 8 |

List of Principal Investigators (PI) in charge of more than 8 active floats

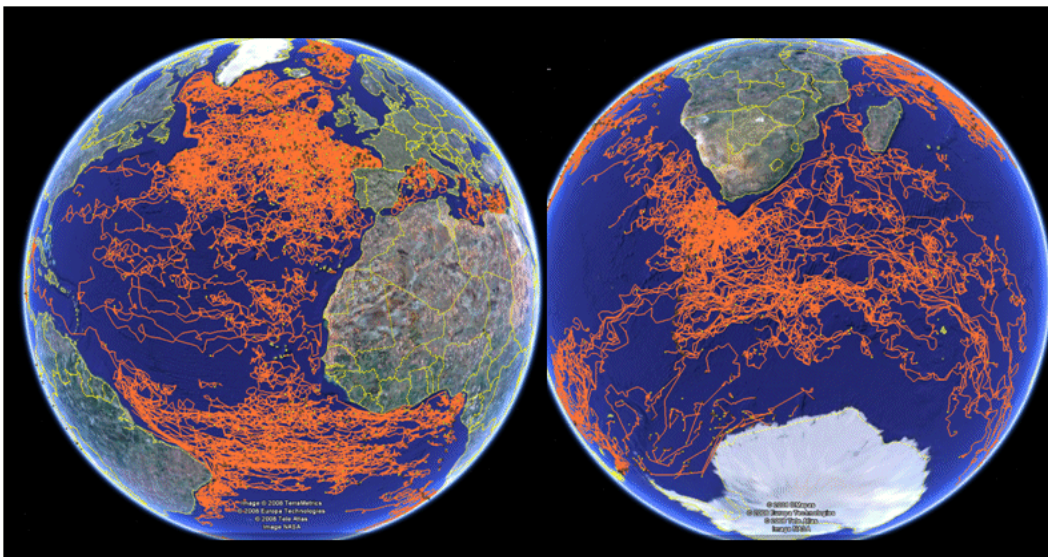
List of Principal Investigators (PI) in charge of less than 8 active floats: sophie cravatte, franck dumas, luis felipe silva santos, vincent echevin, guillaume maze, alban lazar, camille daubord, cecile cabanes, violeta slabakova, bert rudels, louis marié, stephane blain, thierry moutin, xavier andre, arne kortzinger, elodie martinez, katrin latarius, serge le reste, stephanie louazel, sven petersen, tobias ramalho dos santos ferreira, agus atmadipoera, alain serpette, anja schneehorst, bettina fach, jordi font, julia uitz, nathanaele lebreton, pascal conan, s. petersen, sorin balan

1.8 Products generated from Argo data ...

Sub-surface currents ANDRO Atlas

Based on Argo trajectory data, Michel Ollitrault and the Ifremer 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 (2013). *ANDRO: An Argo-based deep displacement dataset*. SEANOE. <http://doi.org/10.17882/47077>
- See also : <http://wwz.ifremer.fr/lpo/Produits/ANDRO>



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

2 Delayed Mode QC

(Please report on the progress made towards providing delayed mode Argo data, how it's organized and the difficulties encountered and estimate when you expect to be pre-operational.)

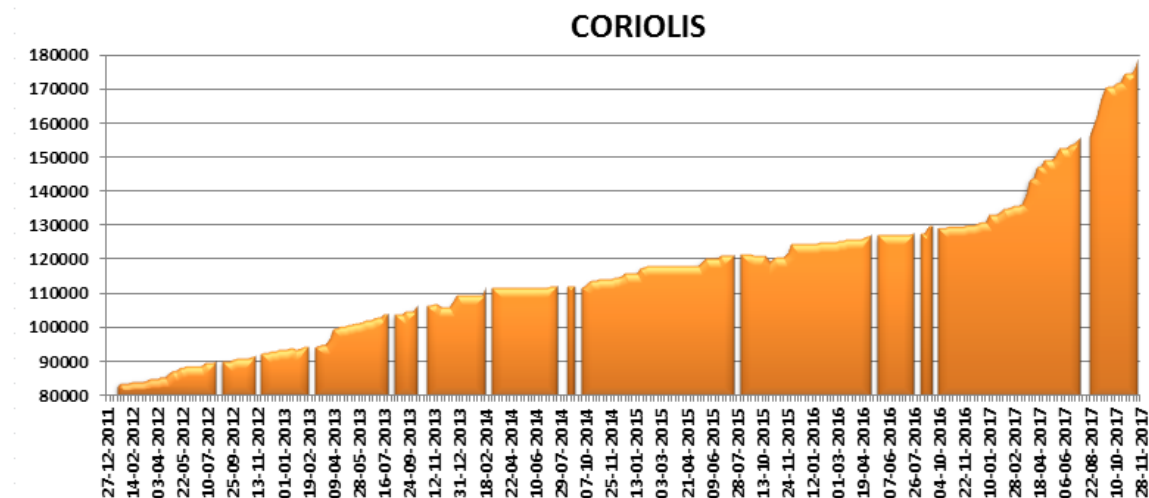
At the Coriolis data centre, we process the delayed mode quality control following four steps. Before running the OW method, we check carefully the metadata files, the pressure offset, the quality control done in real time and we compare with neighbor profiles to check if a drift or offset could be easily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

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.

For a few projects, there are still no identified operators to do DMQC, for instance the first run has been done by students which have now left institutes or are not available to carry on with this work. We have made a lot of progress with BSH (Birgit Klein) taking into account also floats from other German institutes and OGS (Giulio Notarstefano) for the MedSea.

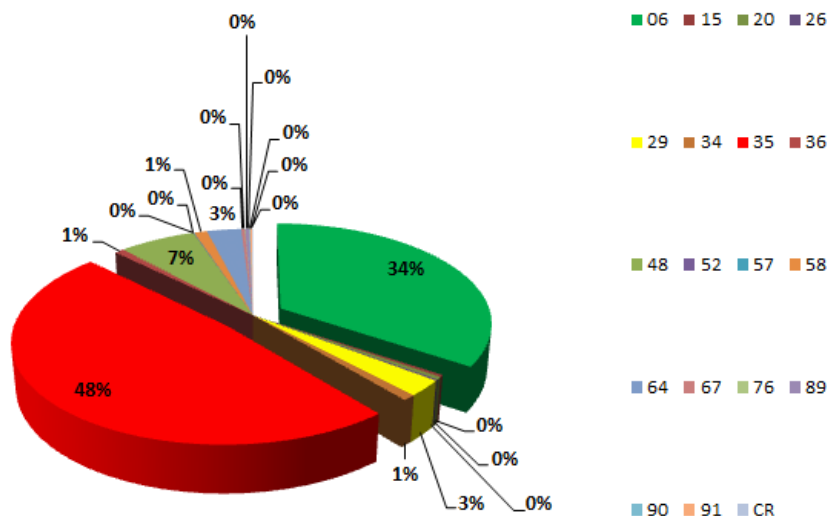
Some DM files have been updated to format version 3.1 taking into account a new decoder (matlab) developed at Coriolis. This work has been done for some Provor and Apex, few files need to be manually updated.

Regular DM files submission is performed each year but an effort has been done during the year 2017 to increase the DM files number.



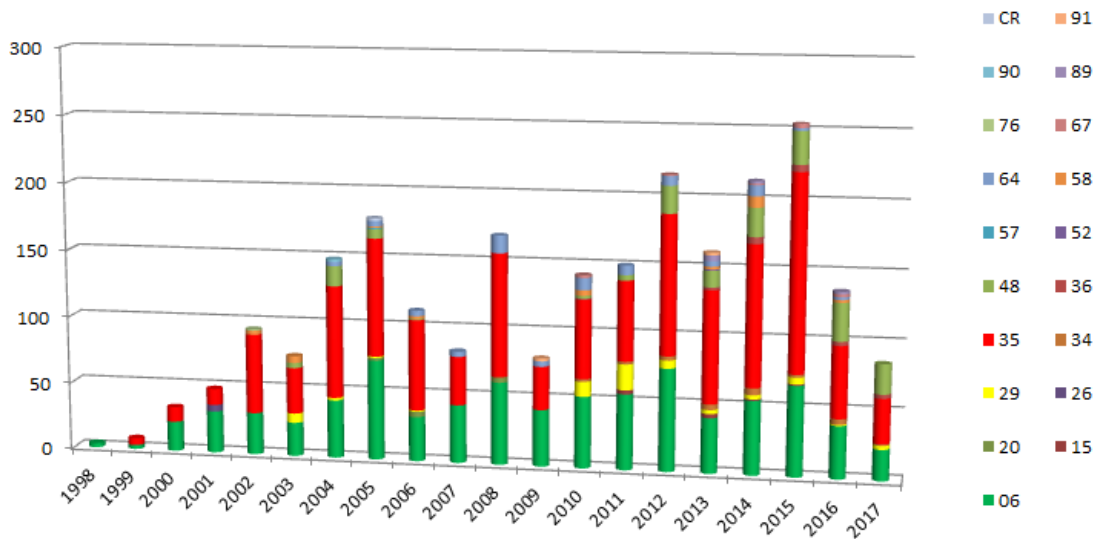
Evolution of the DM profiles' submission versus dates

Floats by country



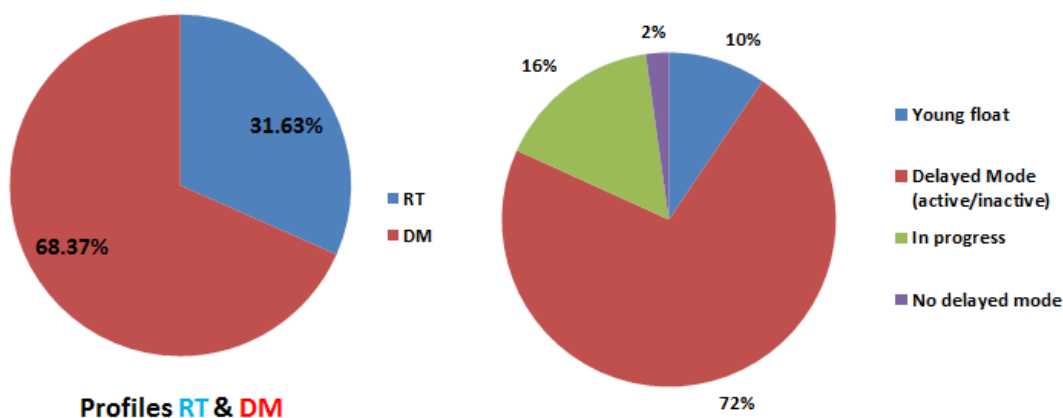
Percentage of floats by country in the Coriolis DAC.

Codes for the countries: 06 : Germany - 15 : Bulgaria - 20 : Chili - 26 : Denmark - 29 : Spain - 34 : Finland - 35 : France - 36 : Greece - 48 : Italy - 52 : Lebanon - 57 : Mexico - 58 : Norway - 64 : Netherlands - 67 : Poland - 76 : China - 89 : Turkey - 90 : Russia - 91 : - South Africa - CR : Costa Rica



Number of floats by country and by launch's year in the Coriolis DAC

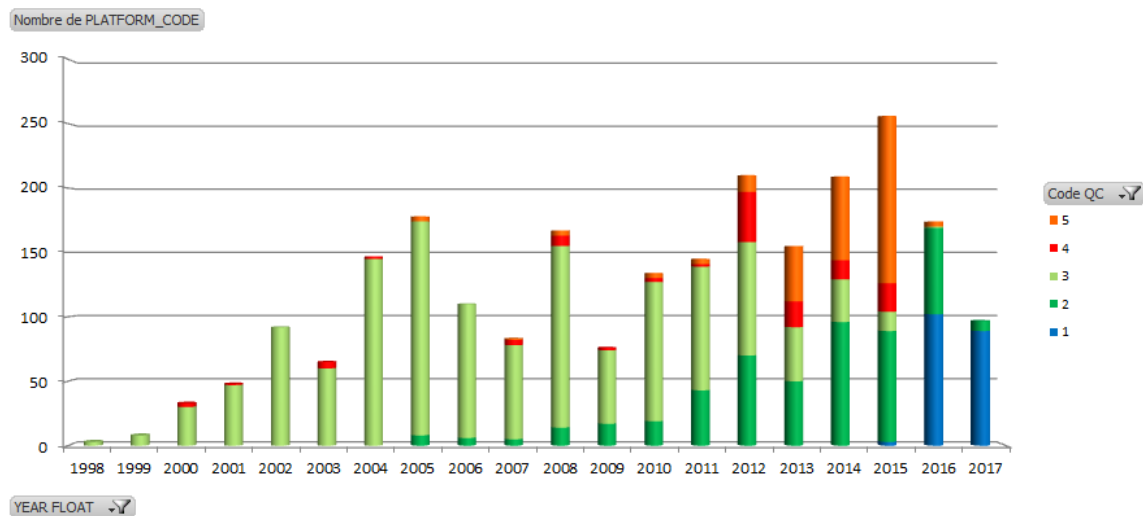
During the last year (from October 2016 to November 2017), 49125 new delayed mode profiles were produced and validated by PIs. A total of 178763 delayed mode profiles were produced and validated since 2005.



Status of the floats processed by Coriolis DAC.

Left: in terms of profile percent and right: in terms of float percent (DM : delayed mode – RT : real time).

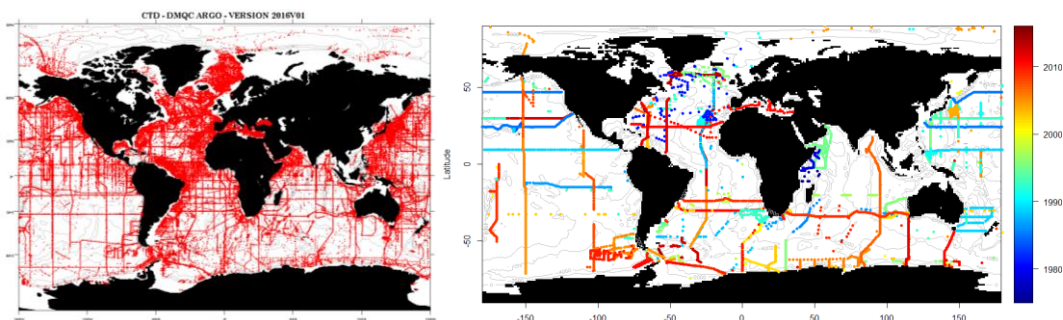
The status of the quality control done on the Coriolis floats is presented in the following plot. For the two last years (2016-2017), most of the floats are still too young (code 1) to be performed in delayed mode. For the years 2012-2013-2014, we are still working on the DMQC of some floats. The codes 2 and 3 show the delayed mode profiles for respectively active and dead floats.



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.

2.1 Reference database

In September 2016, the version CTD_for_DMQC_2016V01 has been provided with a large improvement of the dataset since this version took into account new CTD provided by the CCHDO API (following figure), CTD from scientists as well as feedbacks from users on quality of some profiles.

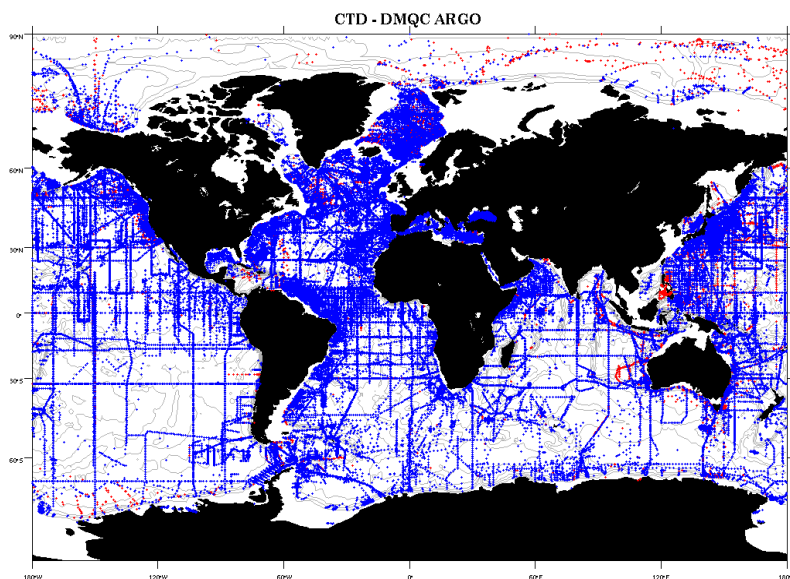


Version 2016 V01 & New CTD datasets downloaded from the CCHDO API

A new version 2017V01 has been provided at the beginning of this year with some updates on a few boxes, following the feedback sent by some scientists.

This last version is available on the ftp site in smaller tar balls, one by wmo box area (1-3-5-7): for instance, CTD_for_DMQC_2017V01_1.tar.gz for all boxes starting with wmo 1, then we will have 4 tar files.

During this year, the OCL updates have been downloaded and some works on quality control have been made before adding them in the reference database. A new version is in preparation and should be delivered at the beginning of 2018. Some new CTDs can be observed in the following plot (blue: last version and red: new CTDs).



3 GDAC Functions

(If your centre operates a GDAC, report the progress made on the following tasks and if not yet complete, estimate when you expect them to be complete)

- National centres reporting to you
- Operations of the ftp server
- Operations of the www server
- Data synchronization
- Statistics of Argo data usage : Ftp and WWW access, characterization of users (countries, field of interest : operational models, scientific applications) ...

3.1 National centres reporting to you

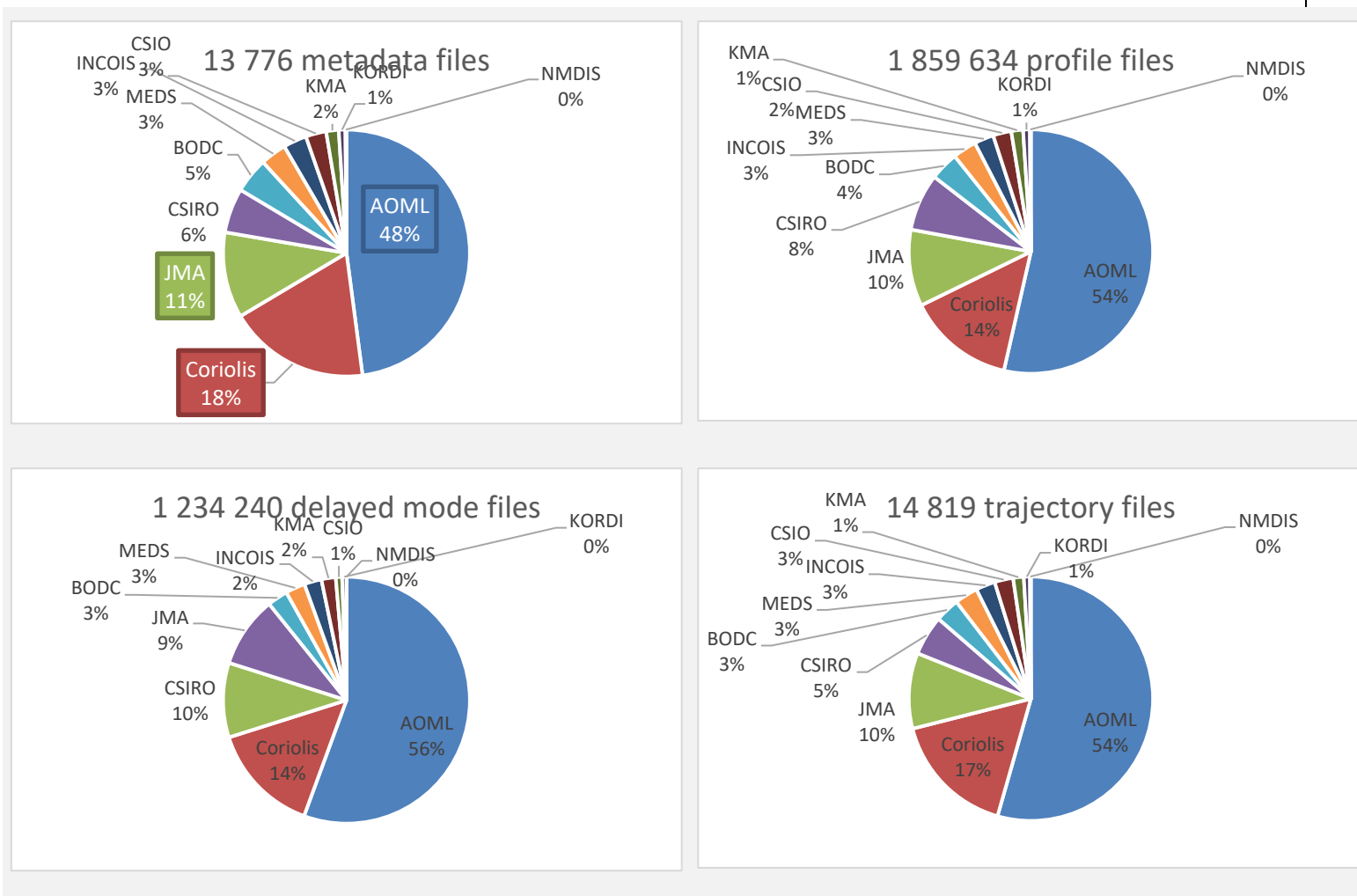
Currently, 11 national DACs submit regularly data to Coriolis GDAC.

The additional GTS DAC contains all the vertical profiles from floats that are not managed by a national DAC. These data come from GTS and GTSP projects. The GTS profiles are quality controlled by the French DAC (Coriolis).

On November 25th 2017, the following files were available from the GDAC FTP site.

3.1.1 GDAC files distribution

| DAC | metadata files 2017 | increase from 2016 | profile files 2017 | increase from 2016 | delayed mode profile files 2017 | increase from 2016 | trajectory files 2017 | increase from 2016 |
|--------------|------------------------|-----------------------|-----------------------|-----------------------|---------------------------------------|-----------------------|--------------------------|-----------------------|
| AOML | 6 601 | 10% | 996 019 | 13% | 686 144 | 20% | 8 064 | 12% |
| BODC | 636 | 18% | 70 637 | 23% | 33 015 | 5% | 479 | 14% |
| Coriolis | 2 554 | 11% | 263 894 | 17% | 179 361 | 38% | 2 465 | 11% |
| CSIO | 370 | 8% | 44 934 | 15% | 10 221 | 0% | 365 | 7% |
| CSIRO | 806 | 8% | 140 870 | 14% | 120 680 | 24% | 781 | 10% |
| INCOIS | 422 | 7% | 58 538 | 14% | 28 008 | 1% | 379 | 2% |
| JMA | 1 550 | 7% | 188 716 | 11% | 115 044 | 20% | 1 484 | 4% |
| KMA | 227 | 5% | 29 005 | 11% | 23 094 | 11% | 208 | 0% |
| KORDI | 119 | 0% | 16 578 | 2% | 6 986 | | 119 | 0% |
| MEDS | 472 | 9% | 47 983 | 8% | 31 687 | 23% | 456 | 8% |
| NMDIS | 19 | 0% | 2 460 | 0% | 0 | | 19 | 0% |
| Total | 13 776 | 9,35% | 1 859 634 | 13,75% | 1 234 240 | 22,02% | 14 819 | 9,93% |



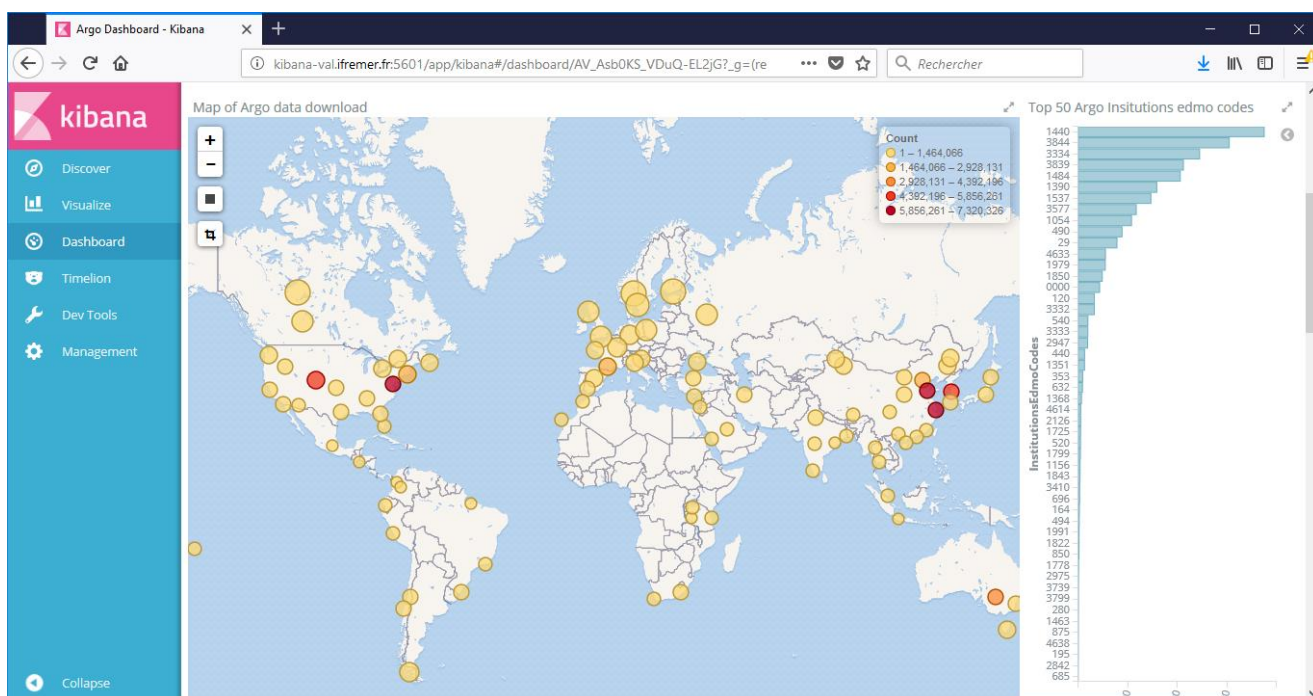
Number of files available on GDAC, November 2017

3.1.2 FTP dashboard: give credit to data providers

Within EU AtlantOS project, Ifremer is setting up a dashboard to monitor data distribution and give credit to data providers such as Argo floats.

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



Geographical distribution of GDAC ftp downloads in 2017

The majority of users (red dots) are located in USA, China, Australia and of course Europe. The right side histogram sorts the floats institution code (1440: PMEL, 3844: WHOI, 3334: INCOIS, 3839: UWA, 1484: CSIRO, ...).

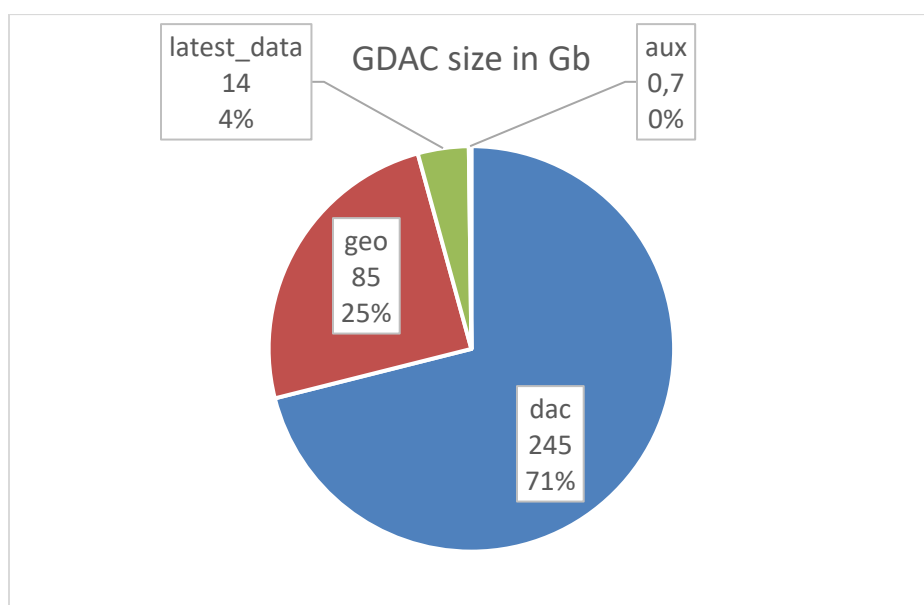


The top 50 of floats institutions downloads and the top 50 of data users's

3.1.3 GDAC files size

- The total number of NetCDF files on the GDAC/dac directory was 2 178 811
- The size of GDAC/dac directory was 245 G (+95%)
- *The size of the GDAC directory was 462Go*

| GDAC branch | GDAC size in Gb | since 2016 |
|-------------|-----------------|------------|
| dac | 245 | 46% |
| geo | 85 | 25% |
| latest_data | 14 | -7% |
| aux | 0,7 | - |

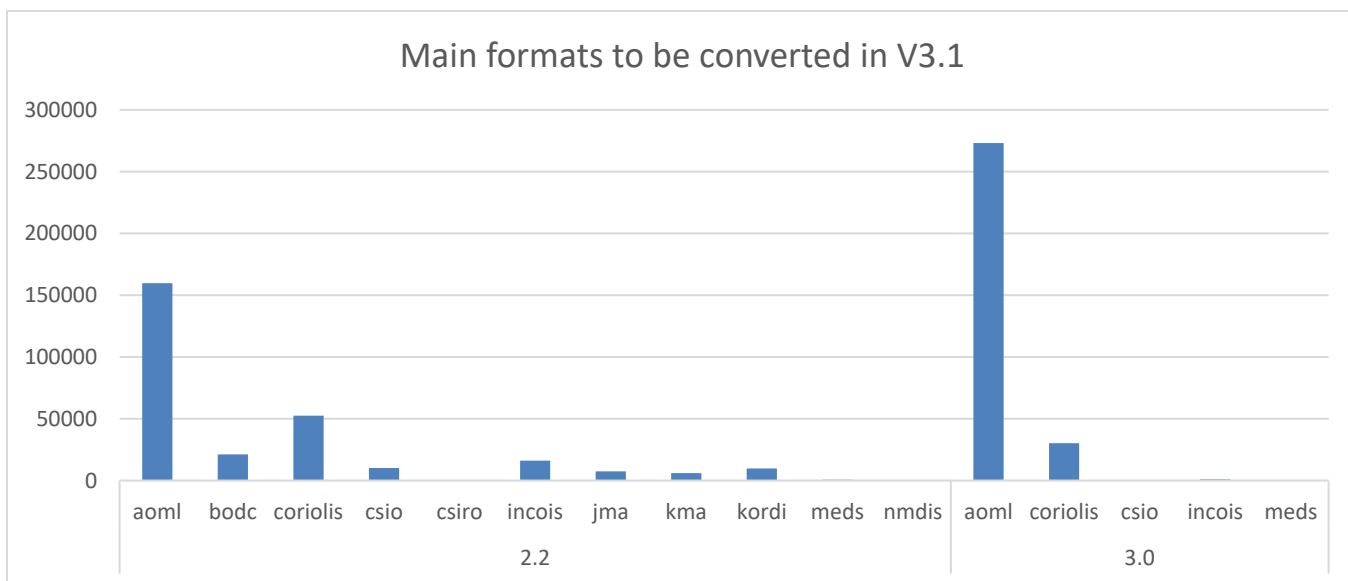
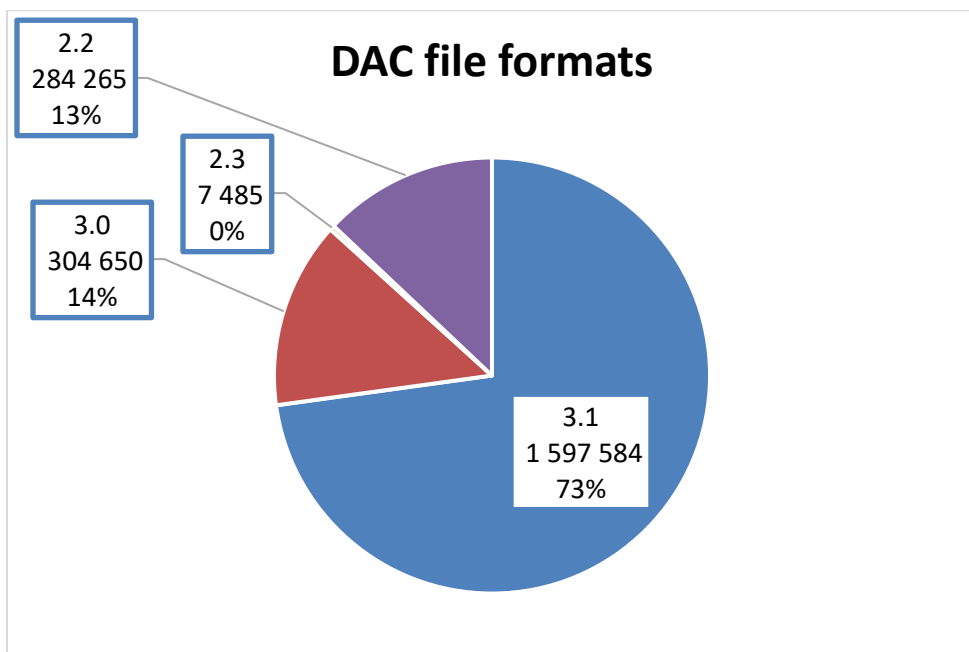


3.1.4 Argo NetCDF transition to format V3.1

The transition from Argo format 2.* and 3.0 toward format 3.1 is underway.

On 2017, the number of files in format version 3.1 reached and passed the 70% threshold.

| format version | nb files | percentage |
|----------------|------------------|-------------|
| 3.1 | 1 597 584 | 73% |
| 3.0 | 304 650 | 14% |
| 2.3 | 7 485 | 0% |
| 2.2 | 284 265 | 13% |
| 2.1 | 12 | 0% |
| Total | 2 193 996 | 100% |



Detailed list of format versions per DAC

| File format | number of files |
|-------------|-----------------|
| 2.1 | 12 |
| aoml | 2 |
| csio | 8 |
| kma | 1 |
| meds | 1 |
| 2.2 | 284265 |
| aoml | 159647 |

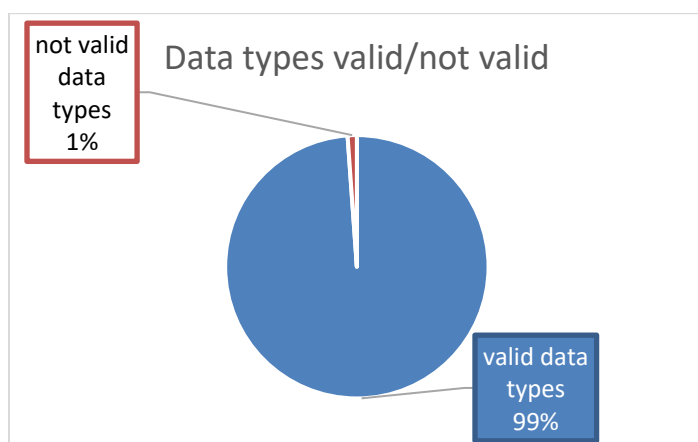
| | |
|----------------------|----------------|
| bodc | 21137 |
| coriolis | 52487 |
| csio | 10097 |
| csiro | 314 |
| incois | 16122 |
| jma | 7447 |
| kma | 6064 |
| kordi | 9789 |
| meds | 713 |
| nmdis | 448 |
| 2.3 | 7485 |
| aoml | 80 |
| bodc | 556 |
| coriolis | 6623 |
| csio | 2 |
| incois | 16 |
| jma | 1 |
| kma | 148 |
| kordi | 55 |
| nmdis | 4 |
| 3.0 | 304650 |
| aoml | 273128 |
| coriolis | 30253 |
| csio | 352 |
| incois | 908 |
| meds | 9 |
| 3.1 | 1577138 |
| aoml | 663282 |
| bodc | 57140 |
| coriolis | 265158 |
| csio | 48531 |
| csiro | 183719 |
| incois | 53261 |
| jma | 216367 |
| kma | 23736 |
| kordi | 7200 |
| meds | 56660 |
| nmdis | 2084 |
| Total général | 2173550 |

The files in format version V3.1 are much more homogeneous than their previous versions. The controls applied by the format checker on V3.1 is much more exhaustive. The controlled vocabulary listed in the 27 reference tables is used for V3.1 format checks. A non-valid content is automatically rejected. Only valid V3.1 content appears on GDAC.

Example of valid content checked by the format checker on V3.1 files

There are 8 valid DATA_FORMAT variables listed in reference table 1 (there are 26 more tables...). A survey on GDAC files shows that 24 779 files (1% of the total) do not have a valid DATA_FORMAT. The V3.1 files are not affected by this kind of problem.

| data_type | nb files | valid type |
|---------------------|----------|------------|
| Argo meta-data | 13748 | yes |
| ARGO profile | 22326 | no |
| Argo profile | 1867361 | yes |
| Argo profile merged | 131347 | yes |
| Argo technical | 556 | no |
| ARGO technical data | 174 | no |
| Argo technical data | 12686 | yes |
| ARGO trajectory | 1237 | no |
| Argo Trajectory | 110 | no |
| Argo trajectory | 13434 | yes |
| B-Argo profile | 130641 | yes |
| B-Argo trajectory | 376 | no |

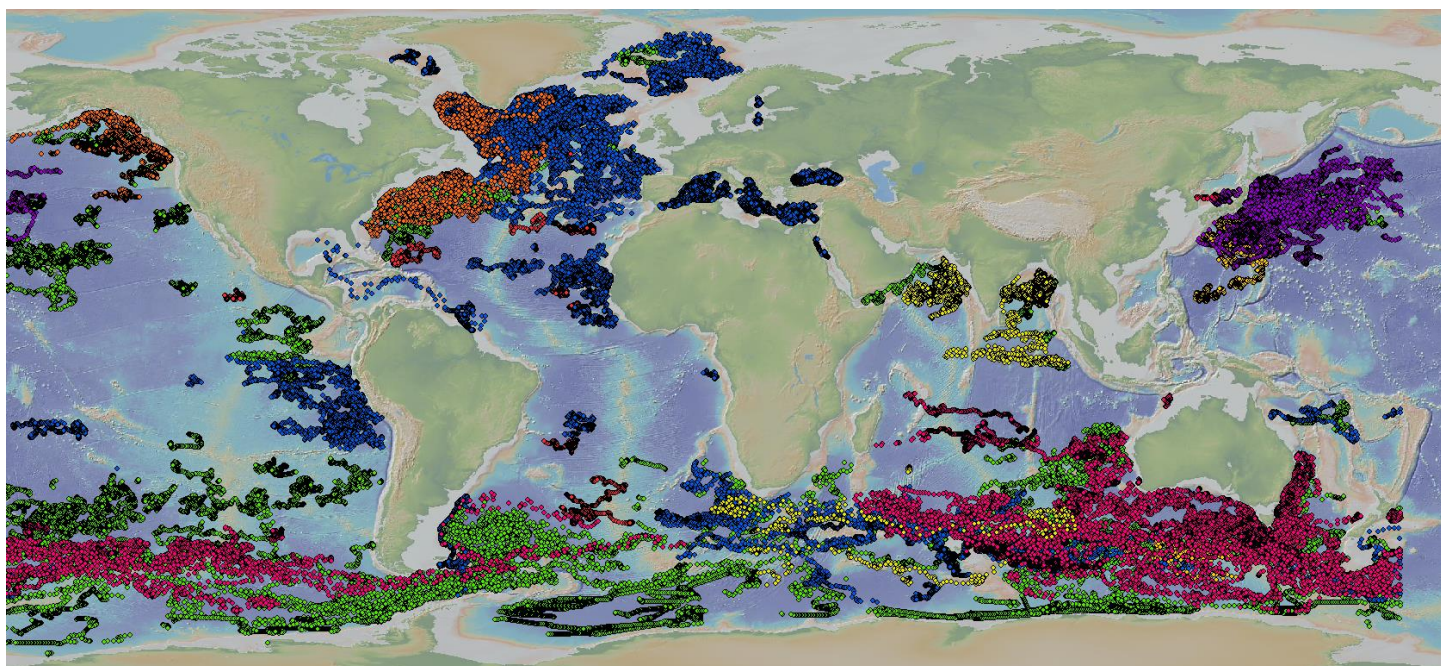
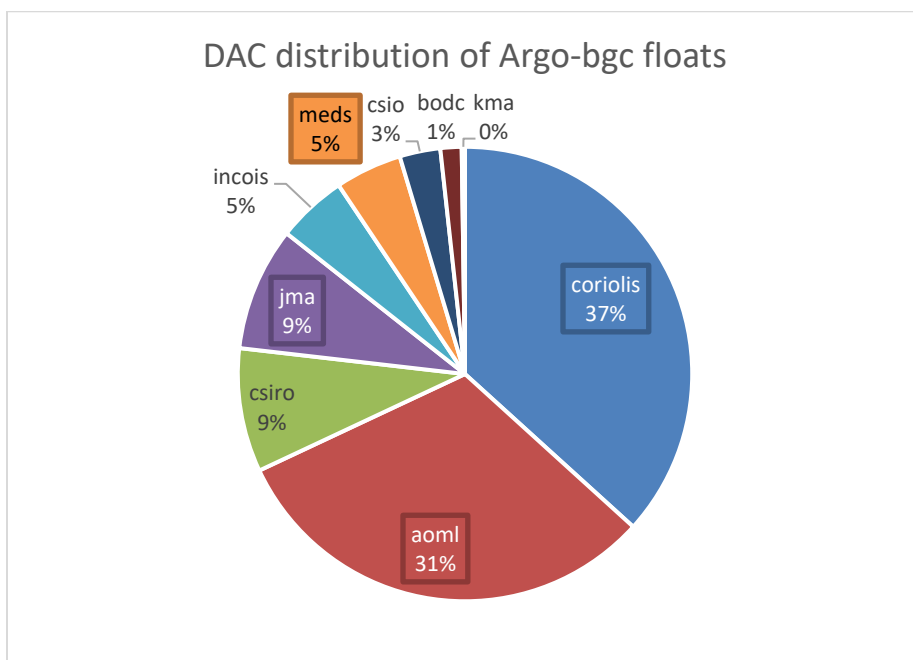


3.1.5 BGC-Argo floats

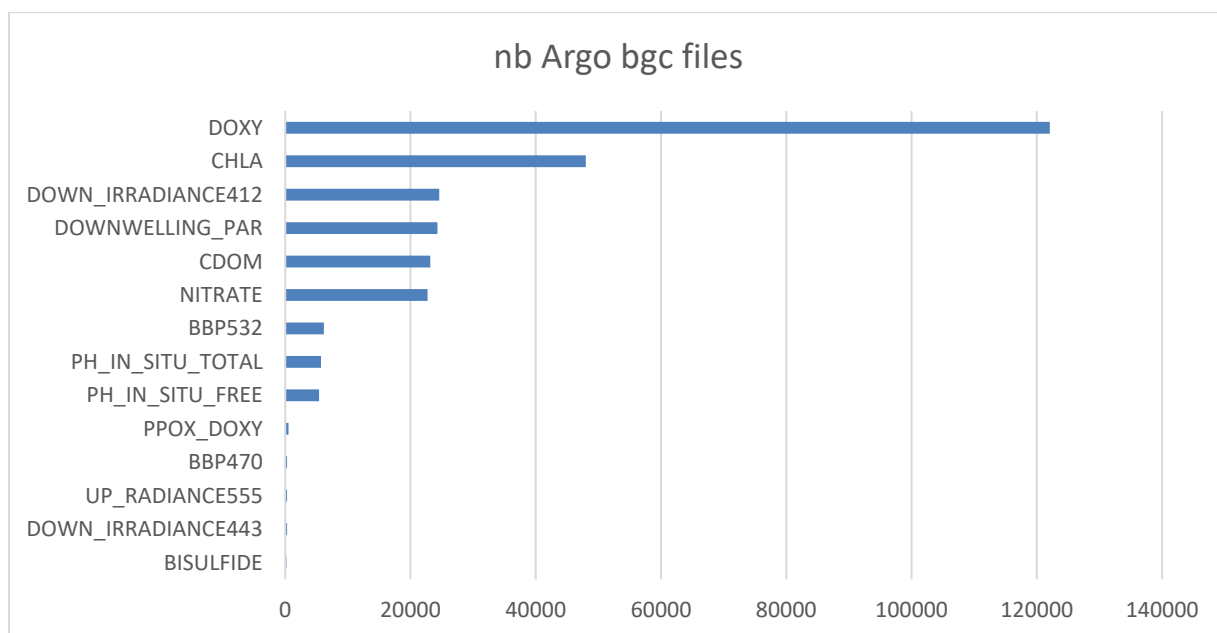
In November 2017, 131 308 BGC Argo profiles from 863 floats were available on Argo GDAC. This is a strong increase compared to 2016 : +65% more profiles and +54% more floats.

| DAC | nb bio floats | nb bio files |
|----------|---------------|--------------|
| coriolis | 317 | 41 147 |
| aoml | 270 | 36 917 |
| csiro | 76 | 20 149 |
| jma | 76 | 14 602 |
| incois | 43 | 5 201 |
| meds | 41 | 3 821 |

| | | |
|--------------|------------|----------------|
| csio | 25 | 6 344 |
| bodc | 13 | 3 041 |
| kma | 2 | 86 |
| Total | 863 | 131 308 |



BGC-Argo profiles, colored by DACs



Main BGC-Argo physical parameters, number of profiles

3.2 Operations of the ftp server

Meta-data, profile, trajectory and technical data files are automatically collected from the national DACs ;

Index files of meta-data, profile and trajectory are daily updated ;

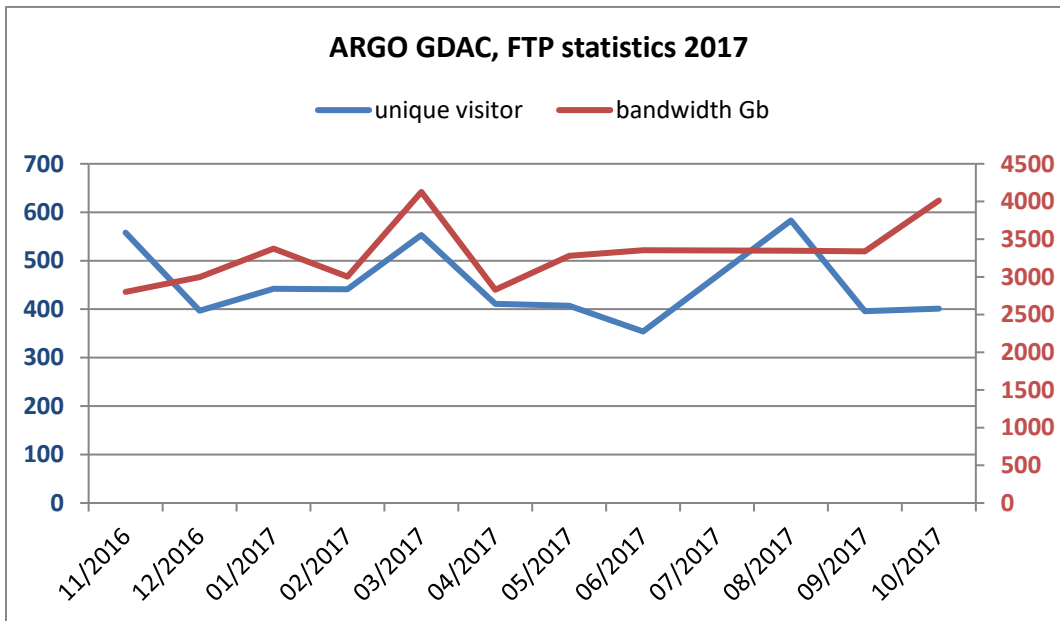
GDAC ftp address: <ftp://ftp.ifremer.fr/ifremer/argo>

Statistics on the Argo GDAC FTP server: <ftp://ftp.ifremer.fr/ifremer/argo>

There is a monthly average of 449 unique visitors, performing 4552 sessions and downloading 3.3 terabytes of data files.

The table below shows a huge spike of visitors in July 2017 on GDAC FTP. This may be an attack of the server, the July unique visitors number is ignored in statistics.

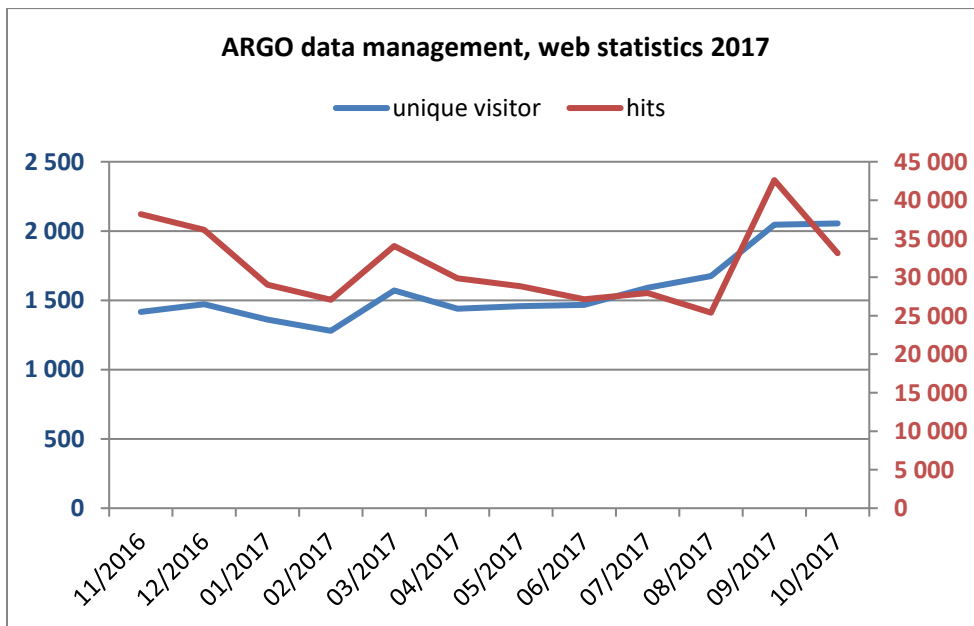
| ARGO GDAC FTP statistics | | | | |
|--------------------------|----------------|------------------|------------------|--------------|
| month | unique visitor | number of visits | hits | bandwidth Gb |
| 11/2016 | 558 | 5 045 | 3 969 591 | 2800,42 |
| 12/2016 | 397 | 3 839 | 4 072 286 | 2997,55 |
| 01/2017 | 442 | 4 778 | 7 213 071 | 3375,6 |
| 02/2017 | 441 | 4 215 | 4 913 490 | 3003 |
| 03/2017 | 553 | 4 619 | 10 495 933 | 4127 |
| 04/2017 | 411 | 4 113 | 3 583 758 | 2829 |
| 05/2017 | 407 | 4 563 | 4 006 245 | 3280 |
| 06/2017 | 354 | 4 011 | 3 718 402 | 3354 |
| 07/2017 | 29386 | 47153 | 3 810 288 | 3334 |
| 08/2017 | 583 | 6 183 | 7 235 786 | 3347 |
| 09/2017 | 396 | 4 345 | 4 923 991 | 3339 |
| 10/2017 | 401 | 4 360 | 8 097 647 | 4013 |
| Average | 449 | 4 552 | 5 503 374 | 3 317 |



Statistics on the Argo data management web site: <http://www.argodatamgt.org>

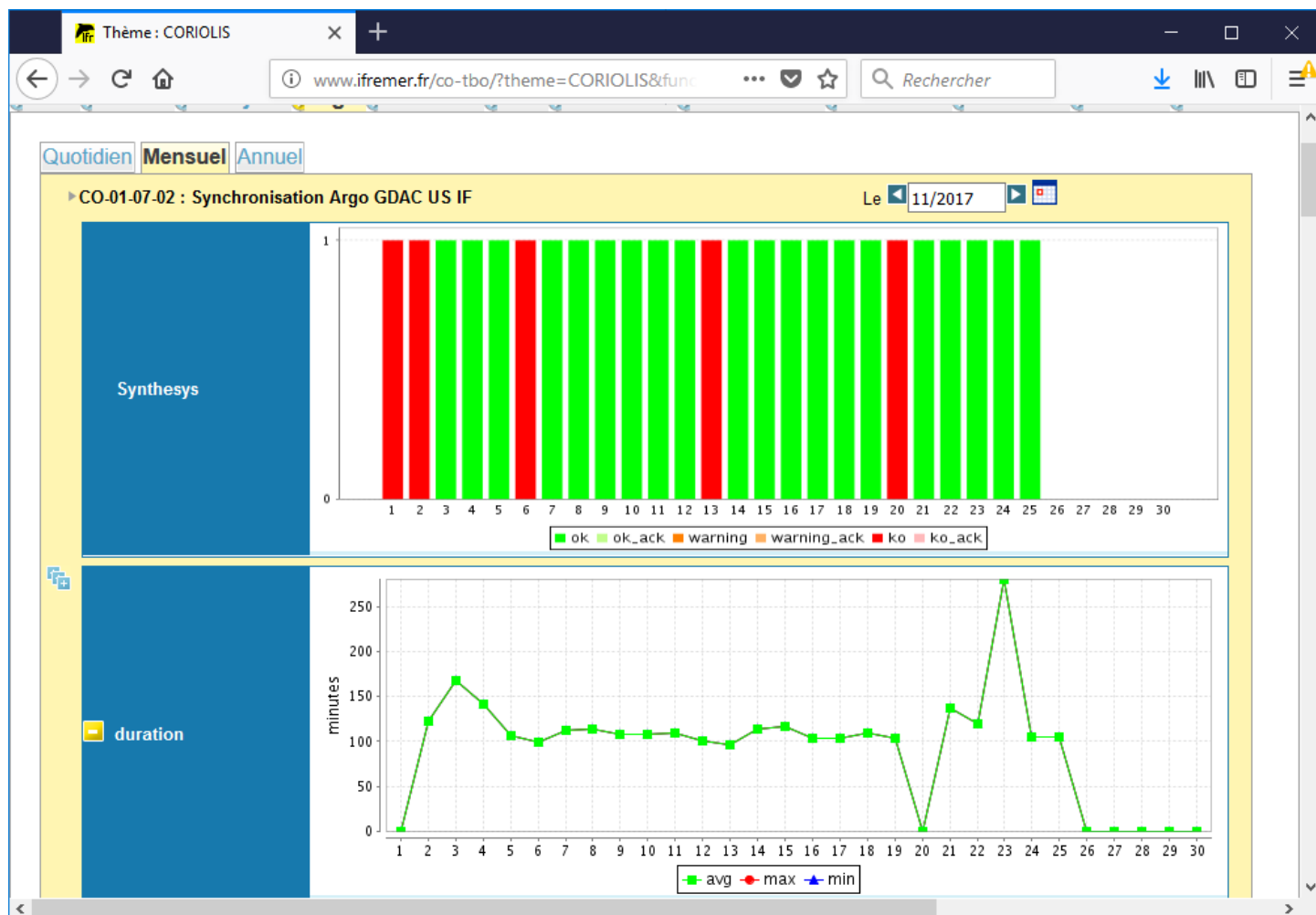
There is a monthly average of 1500 unique visitors, performing 2200 visits and 31000 hits. The graphics shows a slightly increasing number of unique visitors.

| ARGO GDAC web statistics | | | | | |
|--------------------------|----------------|--------------|--------------|---------------|--------------|
| month | unique visitor | visits | pages | hits | bandwidth Go |
| 11/2016 | 1 416 | 2 156 | 4 782 | 38 204 | 0,83 |
| 12/2016 | 1 472 | 2 274 | 5 054 | 36 148 | 1,42 |
| 01/2017 | 1 361 | 2 005 | 3 811 | 29 036 | 1,17 |
| 02/2017 | 1 280 | 1 852 | 3 381 | 27 078 | 0,87 |
| 03/2017 | 1 571 | 2 326 | 4 281 | 34 060 | 1,09 |
| 04/2017 | 1 440 | 2 059 | 4 448 | 29 852 | 0,96 |
| 05/2017 | 1 458 | 2 032 | 3 892 | 28 823 | 1,03 |
| 06/2017 | 1 467 | 2 008 | 3 887 | 27 136 | 1,06 |
| 07/2017 | 1 591 | 2 248 | 4 412 | 27 946 | 1,15 |
| 08/2017 | 1 675 | 2 232 | 3 935 | 25 398 | 1,04 |
| 09/2017 | 2 046 | 3 057 | 5 986 | 42 615 | 1,12 |
| 10/2017 | 2 055 | 2 779 | 5 245 | 33 123 | 1,16 |
| Average | 1 569 | 2 252 | 4 426 | 31 618 | 1,08 |



3.3 Data synchronization

The synchronization with US-Godae server is performed once a day at 01:55Z.



The synchronization dashboard in November 2017: the daily synchronization time takes on average 2 hours.

You may notice on the dashboard that the synchronization process reported 5 errors in November (red bars):

- “Can’t create the ftp connection to usgodae.org”
There was an ftp connection problem between Coriolis and US GDACs

3.4 FTP server monitoring

Note: this chapter is not yet available for 2017.

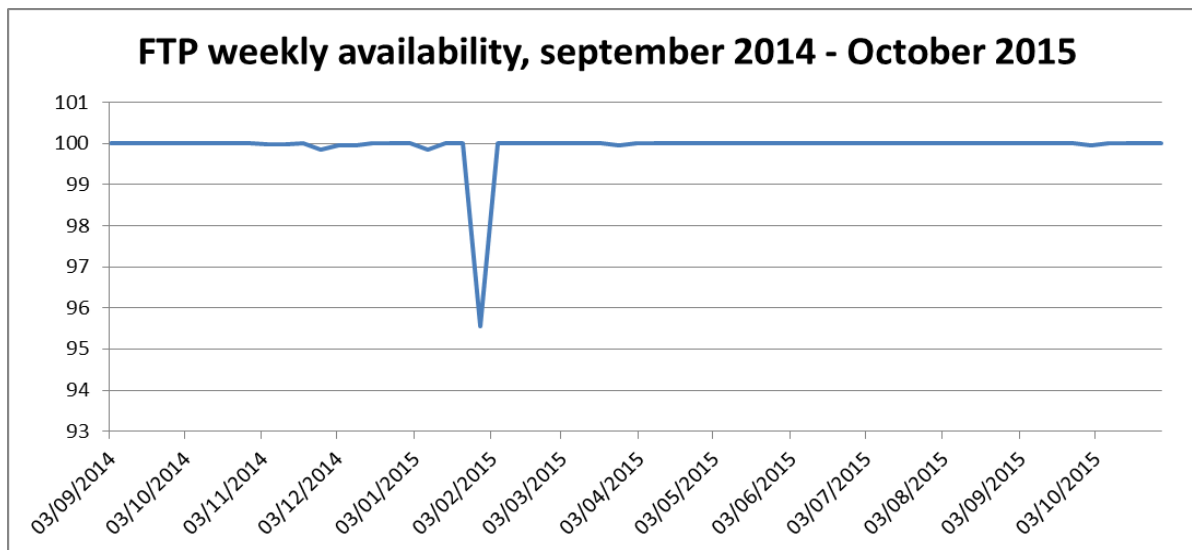
The Argo GDAC ftp server is actively monitored by a Nagios agent (<http://en.wikipedia.org/wiki/Nagios>).

Every 5 minutes, an ftp download test and an Internet Google query are performed. The success/failure of the test and the response time are recorded. The FTP server is a virtual server on a linux cluster.

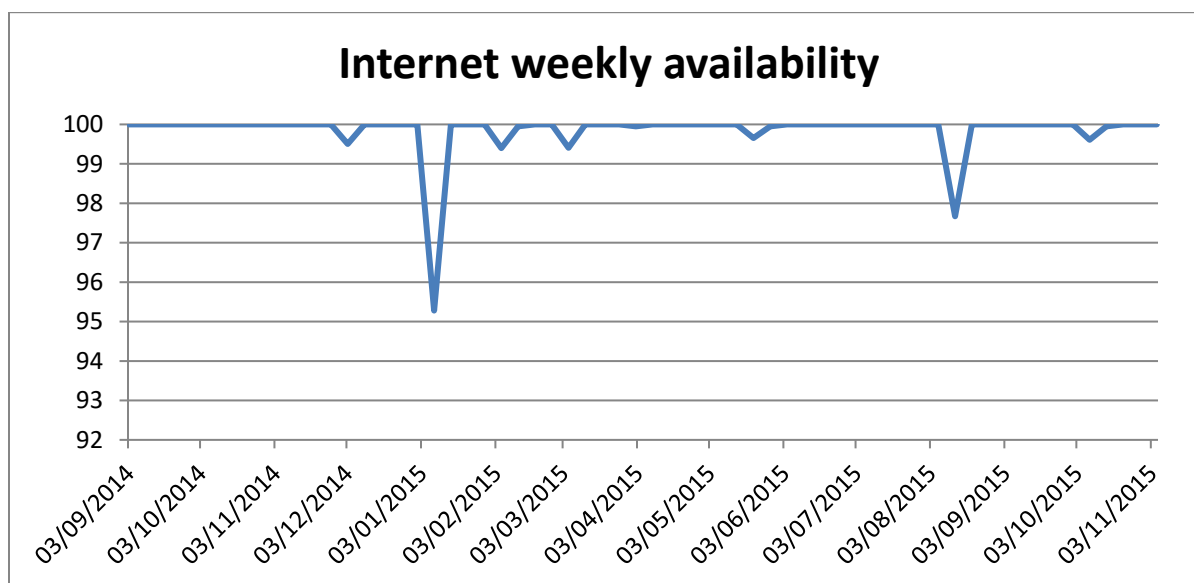
On the last 12 months, the weekly average performance was 99.84%. The 0.16% of poor performances represents 15 minutes for a week. For the last 12 months, the cumulative poor performances period is of 24 hours.

We faced 3 significant events these last 12 months.

- First week of January: 8 hours of Internet poor performances
- Last week of January 2015: disk storage instability: 7 hours and 35 minutes of poor performances of ftp.
- Mid-August 2015 : 4 hours of poor Internet performances



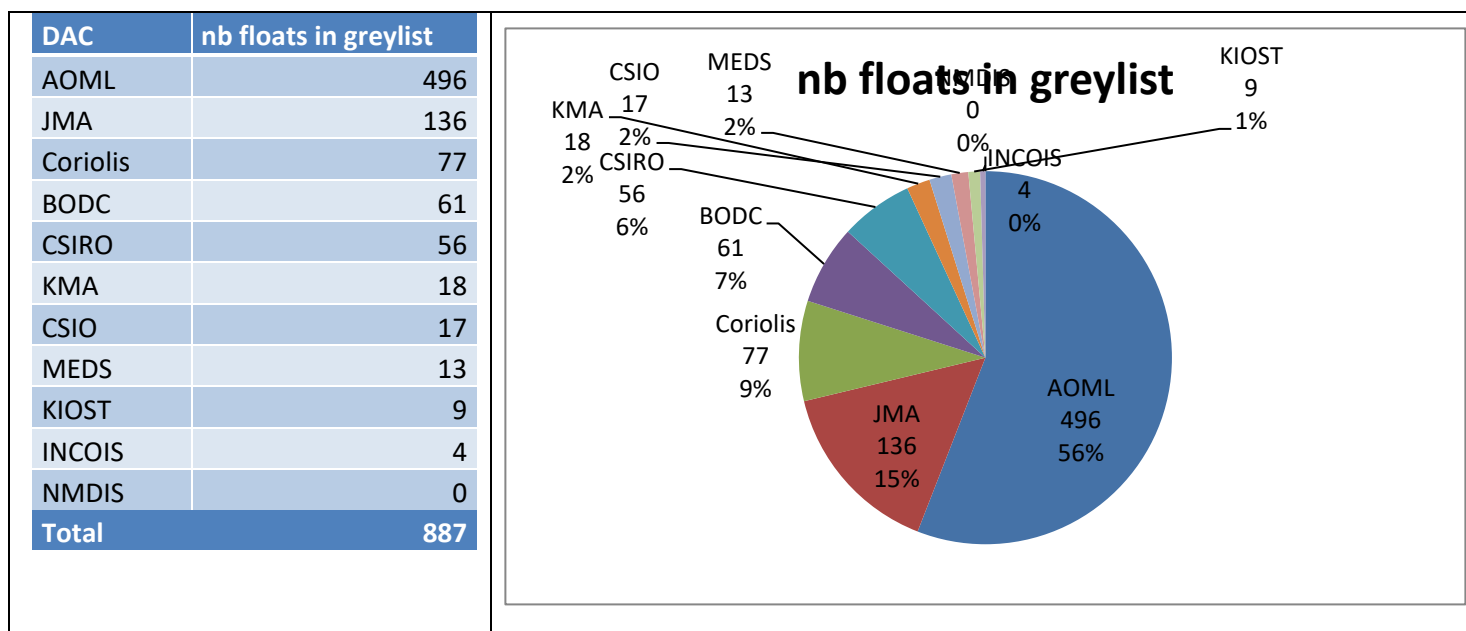
Nagios ftp monitoring: between September 2014 and October 2015



Nagios Internet monitoring: between September 2014 and October 2015

3.5 Grey list

According to the project requirements Coriolis GDAC hosts a grey list of the floats which are automatically flagged before any automatic or visual quality control. **The greylist has 887 entries** (November 25th 2017), compared to 1000 entries one year ago.



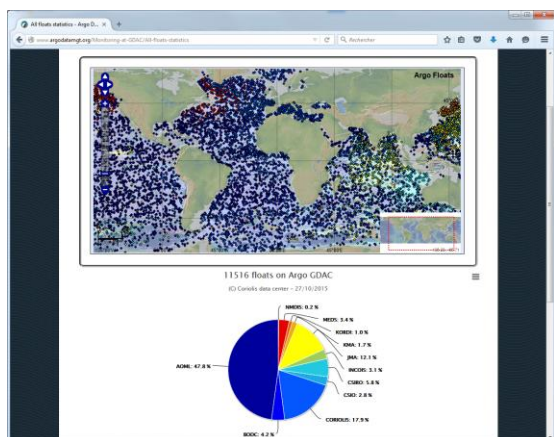
Distribution of greylist entries per DAC and per parameter

AOML reports a high percentage of pressure and temperature in the greylist, compared to other DACs.

| Dac - parameter | Nb floats |
|----------------------|------------|
| AO | 496 |
| PRES | 145 |
| PSAL | 226 |
| TEMP | 125 |
| BO | 61 |
| PRES | 15 |
| PSAL | 29 |
| TEMP | 17 |
| CS | 56 |
| PRES | 10 |
| PSAL | 32 |
| PSAL | 1 |
| TEMP | 13 |
| HZ | 17 |
| PRES | 2 |
| PSAL | 13 |
| TEMP | 2 |
| IF | 77 |
| DOXY | 2 |
| PRES | 5 |
| PSAL | 58 |
| TEMP | 12 |
| IN | 4 |
| PRES | 1 |
| PSAL | 2 |
| TEMP | 1 |
| JA | 136 |
| DOXY | 1 |
| PRES | 22 |
| PSAL | 83 |
| TEMP | 30 |
| KM | 18 |
| PRES | 6 |
| PSAL | 6 |
| TEMP | 6 |
| KO | 9 |
| PRES | 3 |
| PSAL | 3 |
| TEMP | 3 |
| ME | 13 |
| PRES | 2 |
| PSAL | 8 |
| TEMP | 3 |
| Total général | 887 |

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 rsync 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>

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>

Argo GDAC monthly snapshots DOIs

- Snapshot of 2017 October 8th <http://doi.org/10.17882/42182#52113>
- Snapshot of 2014 October 8th <http://doi.org/10.17882/42182#42280>
- Snapshot of 2012 December 1st <http://doi.org/10.17882/42182#42250>