Argo data management

Argo workshop in Ghana, December 2006
Argo data flow

**floats deployment and data certification**

- **pi** principal investigators
  - howard freeland, yves desaubies...

- **dm qc**, few months

**prepare data for distribution**

- float data from satellite
  - 4-12 hours

**data assembly centres**

- **dac**
  - aoml, coriolis, jma, meds, cls...
  - 24 hours, real-time qc

- **gdac**
  - global data centers
  - usgodae, coriolis
  - 24 hours

**users**

- scientists, modelers, students...

- **aic**
  - argo information center
  - information about argo

- **long term archive**
  - 1 year

- **long term archive**
  - usnodc

- 1 day to several years

**perform regional analyses**

- **n. atlantic, pacific, indian**
  - 24 hours

**data sent ashore via satellite**

- **ftp**: 24 hours
- **web**: 12 hours
- **gts**: 24 hours

**here is the master copy of argo data**
Argo data management actors

- **PI : Principal Investigators**
  The scientists who deploy the floats, then carry out delayed mode QC and return data to DAC Centres within a few months of observations.

- **DAC: Data Assembly Centres**
  The DACs are facilities set up by many of the nations deploying floats. Their role is crucial in acquiring the raw data from the floats and standardizing the handling.
  Their functions include:
  - collecting data
  - converting to standard exchange formats
  - applying standardized real-time quality control
  - delivering data to the GTS and GDACs within 24 hours of the surfacing and to PIs on a more relaxed schedule
  - coordinating Argo data handling for the floats under their control

- **GDAC : Global Data Centres**
  The GDACs are the data servers where the master copies of the data reside. The Argo global dataset is available from GDACs; central points for data distribution on Internet for all floats.
  Located in Monterrey (US GODAE/FNMOC/USA) and Brest (Coriolis/Ifremer/France).
  A synchronization between the 2 GDACs centres occurs daily.
Argo data management actors

- **RDAC: Regional Data Centres**
  The Regional Centres look at data from ocean basins to verify float data and generate products. They provide basin-wide synthesis of all floats data with other available data. They provide a feedback to Pis on delayed mode quality control.

- **AIC: Argo Information Centre**
  Centre located in Toulouse/France, in charge of informing on the Argo program status and to provide all necessary information to users.

- **Argo long term archive**
  Data centre located in NODC/USA in charge of the long term archive of all Argo data.
Argo workshop in Ghana, December 2006

Argo data management actors

**PI**
- Delayed mode QC
- Interface with DACs

**DAC**
- Data collection
- Real Time QC
- Interface with PIs
- National Distribution
- Archive

**GDAC**
Provide a **unique** access both in **Real Time** and **Delayed mode** to all the Argo data processed by the DACs

**RDAC**
- Delayed mode QC at Basin Level
- Elaboration of Products

**DATA FROM FLOATS**
Argo quality control process

**Data from floats**

**Delayed-mode (DM) data stream**
- **Function:** Apply accepted DM procedures to float data. Provide statistically justified corrections using accepted methods. Provide feedback to RT system.
- **Users:** All needing adjusted data with error estimates.
- **Timeframe:** 6-12 months after transmission.
- **Who/Where:** Perform by PIs with DM Agencies or Regional Centres.

**Real-time (RT) data stream**
- **Function:** Apply agreed RT QC tests to float data. Assign quality flags.
- **Users:** Operational centres, data assimilation, researchers needing timely data.
- **Timeframe:** 24-48 hrs after transmission.
- **Who/Where:** Performed by National Data Assembly Centres.

**Regional Analysis**
- **Function:** Provide basin-wide synthesis of all float data with other available data.
- **Users:** Researchers of climate change and model validation.
- **Timeframe:** Lifespan of float.
- **Who/Where:** Perform by Regional Centres.

All data are available from the two Argo Global Data Assembly Centres: Coriolis at IFREMER, and USGODAE at Monterey.
Argo real-time quality control

Profiles: 17 automatic QC tests performed before gdac and gts distribution

1. Platform Identification
2. Impossible Date
3. Impossible Location
4. Position on Land
5. Impossible Speed
6. Global Range
7. Regional Range
8. Pressure Increasing
9. Spike
10. Top and Bottom Spike: removed
11. Gradient
12. Digit Rollover
13. Stuck Value
14. Density Inversion
15. Grey List
16. Gross salinity or temperature sensor drift
17. Visual QC (not mandatory)
18. Frozen profile
19. Deepest pressure

Trajectories: 7 automatic QC tests performed before gdac and gts distribution

1. Platform Identification
2. Impossible Date
3. Impossible Location
4. Position on Land
5. Impossible Speed
6. Global Range
7. Regional Global Parameter

QC flag scale

0. No QC was performed
1. Good data
2. Probably good data
3. Bad data that are potentially correctable
4. Bad data
5. Value changed
6. Not used
7. Not used
8. Interpolated value
9. Missing value
The GDAC FTP sites provide the master copy of Argo data set (meta-data, profiles, trajectories and technical informations).

The GDAC keeps the highest processing level of floats data. It is not in charge of archiving the different versions of float data.
Argo data transfer from DACs to GDACs FTP sites

Principles

- Each national center has an FTP account on the global data servers

- When a new cycle is available, the National DAC:
  - Applies real-time quality controls to the new profile(s) and trajectory
  - Distributes simultaneously to both GDACs in NetCDF format:
    - The new profile(s)
    - The trajectory file
    - The technical file
  - Distributes the profile(s) to GTS
    (GTS distribution may be performed by CLS-Argos)

- For new floats Metadata must be transferred first
Argo GDAC : content and data formats

On GDAC FTP site, for each float, 4 types of informations are handled.

- **Metadata file**: general informations on the float life.

- **Profile file**: one file per profile. It contains both original data acquired by the float and the best available profile together with quality flags.
  Note: for users convenience, some aggregated profile files are generated by GDAC.

- **Trajectory file**: one file containing the complete trajectory of the float as well as the measurements collected while drifting.

- **Technical file**: one file containing the technical information provided by the float.

Argo NetCDF format is used for metadata, profile, trajectory and technical files.
Argo NetCDF data format is described in « Argo users’s manual »
Argo GDAC FTP structure

- global FTP server
  - dac
    - coriolis
    - meds
    - ... aoml
  - geo
    - atlantic
    - pacific
    - indian
    - latest_data
      - 2002
        - 05
        - 12
        - ...
  - One directory per float containing a file for
    - the profile file
    - the trajectory file
    - the metadata file
    - the technical file
  - Plus a directory containing individual profiles
  - One file per day containing all of the profiles acquired that day
  - One file per day containing all of the profiles processed that day
Argo data management reference documents

Reference documents
- Data management web page: [http://www.coriolis.eu.org/cdc/argo/argo_data_management.htm](http://www.coriolis.eu.org/cdc/argo/argo_data_management.htm)
- Beginner's guide to Argo data: [http://www.coriolis.eu.org/cdc/argo/Argo_data_guide.pdf](http://www.coriolis.eu.org/cdc/argo/Argo_data_guide.pdf)

Data access
- The whole Argo data set is available in real time and delayed mode from the global data centres (GDACs).
  - The internet addresses are:
    - [http://www.usgodae.org/argo/argo.html](http://www.usgodae.org/argo/argo.html)
    - [http://www.coriolis.eu.org/cdc/argo.htm](http://www.coriolis.eu.org/cdc/argo.htm)
  - The FTP addresses are:
- Data discovery: Live Access Server
  - [http://usgodae2.usgodae.org/las/servlets/dataset](http://usgodae2.usgodae.org/las/servlets/dataset)
  - [http://www.usgodae.org/docs/lasget.html](http://www.usgodae.org/docs/lasget.html)
  - [http://www.ifremer.fr/las/servlets/dataset](http://www.ifremer.fr/las/servlets/dataset)
- Data discovery: web interface
  - [http://www.usgodae.org/cgi-bin/argo_select.pl](http://www.usgodae.org/cgi-bin/argo_select.pl)
  - [http://www.coriolis.eu.org/cdc/ArgoZonalDataSelection/cdcArgoZonalDataSelections.asp](http://www.coriolis.eu.org/cdc/ArgoZonalDataSelection/cdcArgoZonalDataSelections.asp)
  - [http://www.coriolis.eu.org/cdc/dataSelection/cdcDataSelections.asp](http://www.coriolis.eu.org/cdc/dataSelection/cdcDataSelections.asp)
- Data access: OpenDAP
  - [http://www.ifremer.fr/cgi-bin/nph-dods/data/in-situ/argo](http://www.ifremer.fr/cgi-bin/nph-dods/data/in-situ/argo)