Coriolis
OPERATIONAL OCEANOGRAPHY

In Situ Observations:
Operational Systems & Data Management

Godae Summer School 2004

S Pouliquen

Goals of this presentation

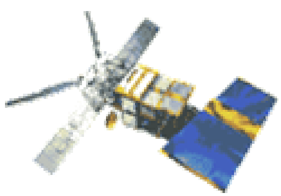
- Show differences between R&D observing systems and operational ones
- Introduce the need of efficient data management procedures
- Show through examples what exist now and show advantages and drawbacks
- Finally highlight what will be the challenges in the future

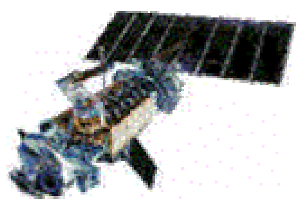
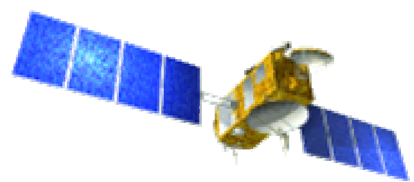
What distinguishes an operational observing system from an R&D one

What do we need to construct an ocean forecast

- **A mathematical model:**
  - ⇒ Would be perfect if we really knew how oceans behave
  - ⇒ But as we don't know how it really varies in time and space thus , we can't put everything in equations ☹

- **Observations from space**
  - ⇒ Give observations at global scale of the surface of the ocean
  - ⇒ Data available in a limited number of professional data centers
  - ⇒ But what about high frequency events that happen between two satellites tracks?
  - ⇒ What about the interior of the ocean for phenomena that have no surface signature ?



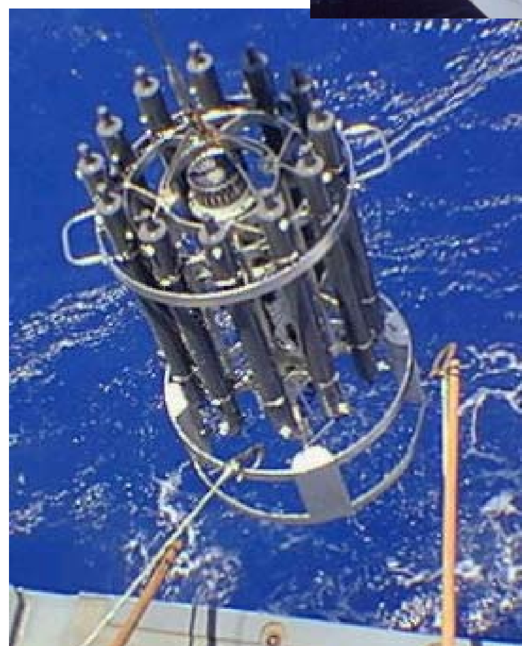
What do we need to construct an ocean forecast

- **In Situ Observations**
  - ⇒ Complementary to satellite but…..data are coming from diverse platforms and sensors stored in an unknown number of places

What happened in the past

- **Since 19th century a lot of measurements have been made by diverse communities for their own needs (Scientists, fishermen, commercial navigators…)**
- **BUT**
  - Not done in an organized way
  - Shared only among small communities
  - Measured over limited periods and areas
  - Not properly archived .

⇒ In Situ data archeology is a hard job providing questionable datasets.



Basic requirements for an operational In Situ Observing system

Instruments

- Autonomous instruments (moorings, drifters, profiling floats, gliders…) to monitor over long periods of time
- Ship measurements to monitor long repeat sections regularly,

Sustainabilty

Infrastructures

- In order to have all these data available for operational models: a well-designed and robust observing system, good communication to shore to deliver data rapidly,
- Operational data centres who work in real-time,

Coverage & Timeliness

Programs

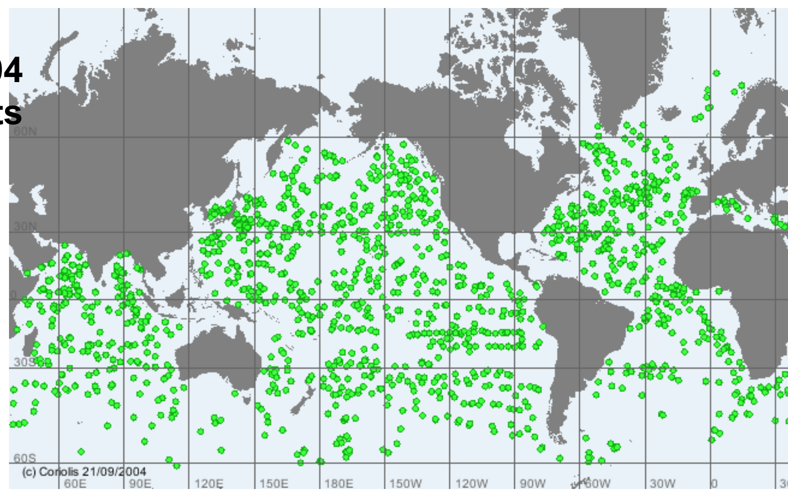
- Suitable data protocols to distribute data to operational centres in a timely way,
- International cooperation to achieve a global coverage, set up an adequate system and maintain it in the long term.

Coordination & Protocols

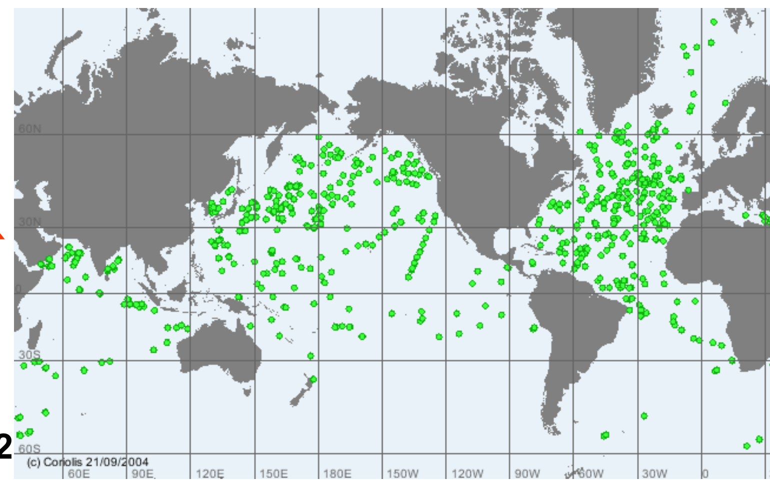
Why do we need sustainability of observing systems

- **Models are tuned to use these data either for initialisation or assimilation or validation. Irregular data provision may lead in the worst case to a real degradation of the forecast…**
  - ⇒ Some impact studies still need to be made to really qualify what are the impact of such or such in-situ data.
  - ⇒ What will happen if suddenly there were funding for maintaining only a fleet of 700 floats out of the 3000 planned?

September 2004
1400 Floats



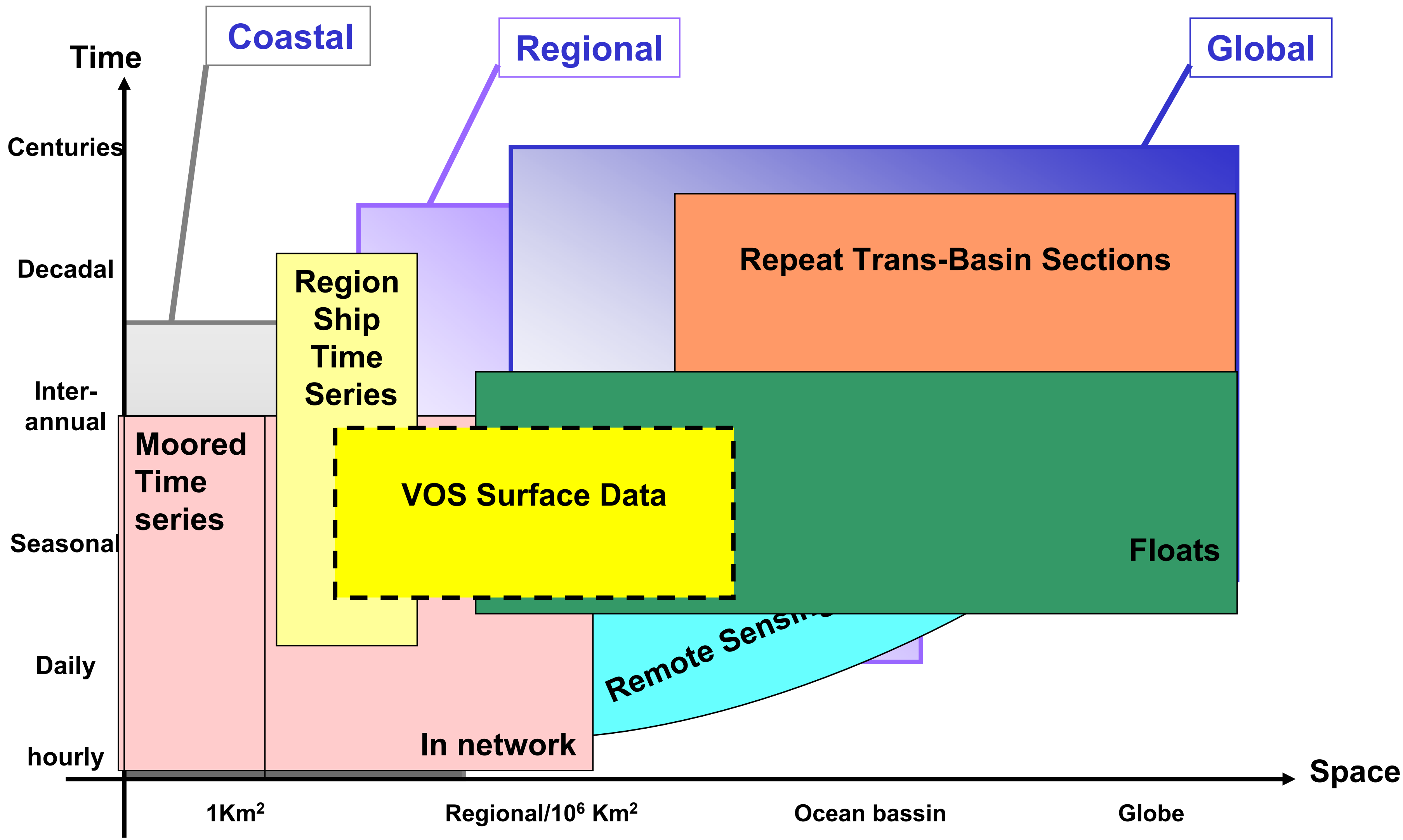
September 2002
700 Floats



Why do we need sustainability of observing systems

- **In search or rescue operations, models have to provide rapidly a product to agencies to help taking decisions. Only few additional measurements can be carried out**

Which System for which need



What is needed by each network

- **Global network:**
  - ⇒ Provide data all over the ocean
  - ⇒ Sampling interval from 10 days to a month
  - ⇒ Build to resolve climate scale phenomena and provide systematic upper ocean observations for a limited number of parameters
  - ⇒ Can only be built on international cooperation

  - ⇒ Examples: Argo , SOOP/VOS, DPCB

What is needed by each network

- **Regional network:**
  - ⇒ Designed to provide data in a specific area to monitor a specific phenomena
  - ⇒ Number of parameters sampled are more important (10 to 20) from physical, biochemical to meteorological measurements
  - ⇒ Sampling interval from hours to days
  - ⇒ Built in collaboration between a few countries

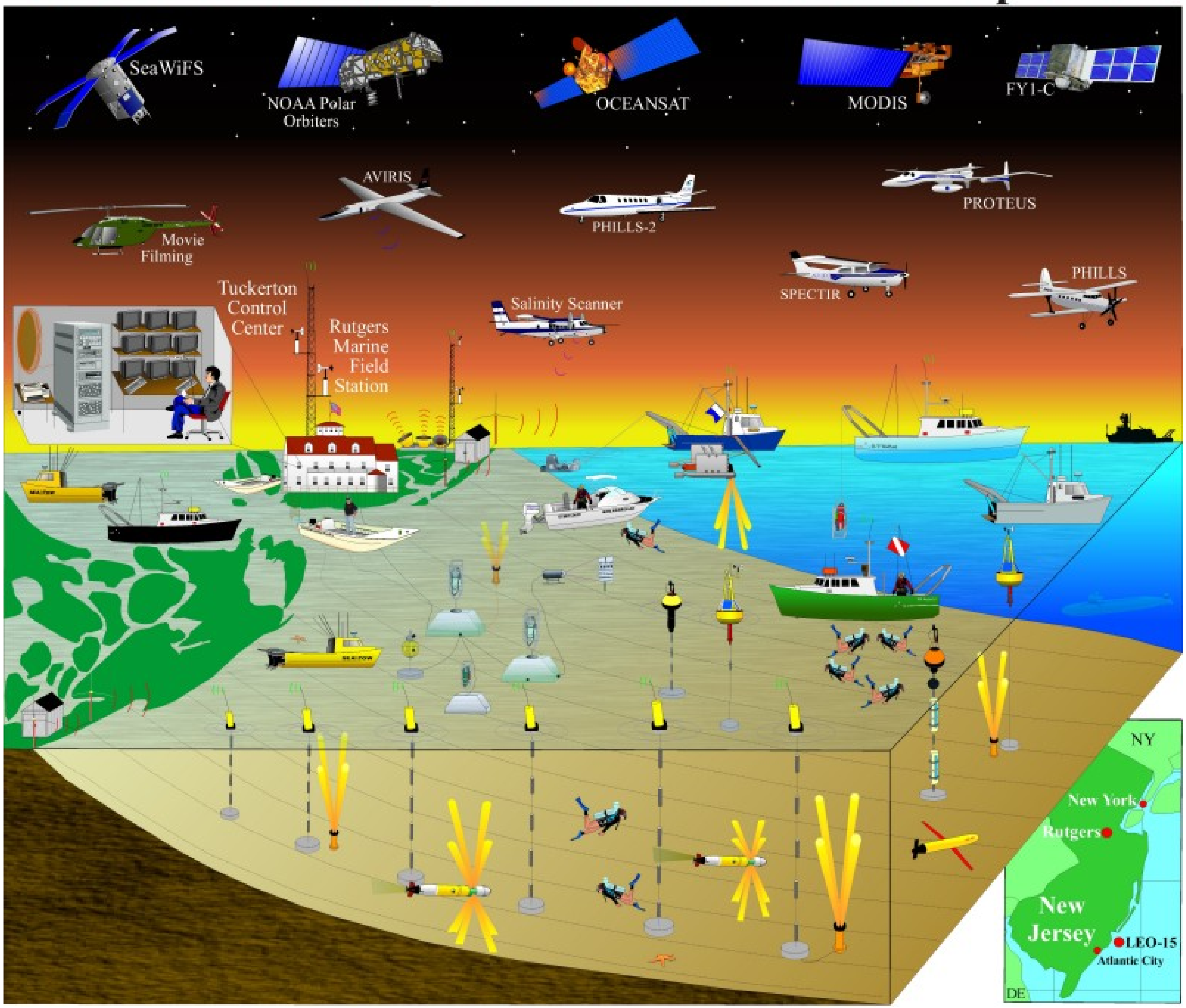
  - ⇒ Examples: Tao/Triton/Pirata array, Artic buoys network

What is needed by each network

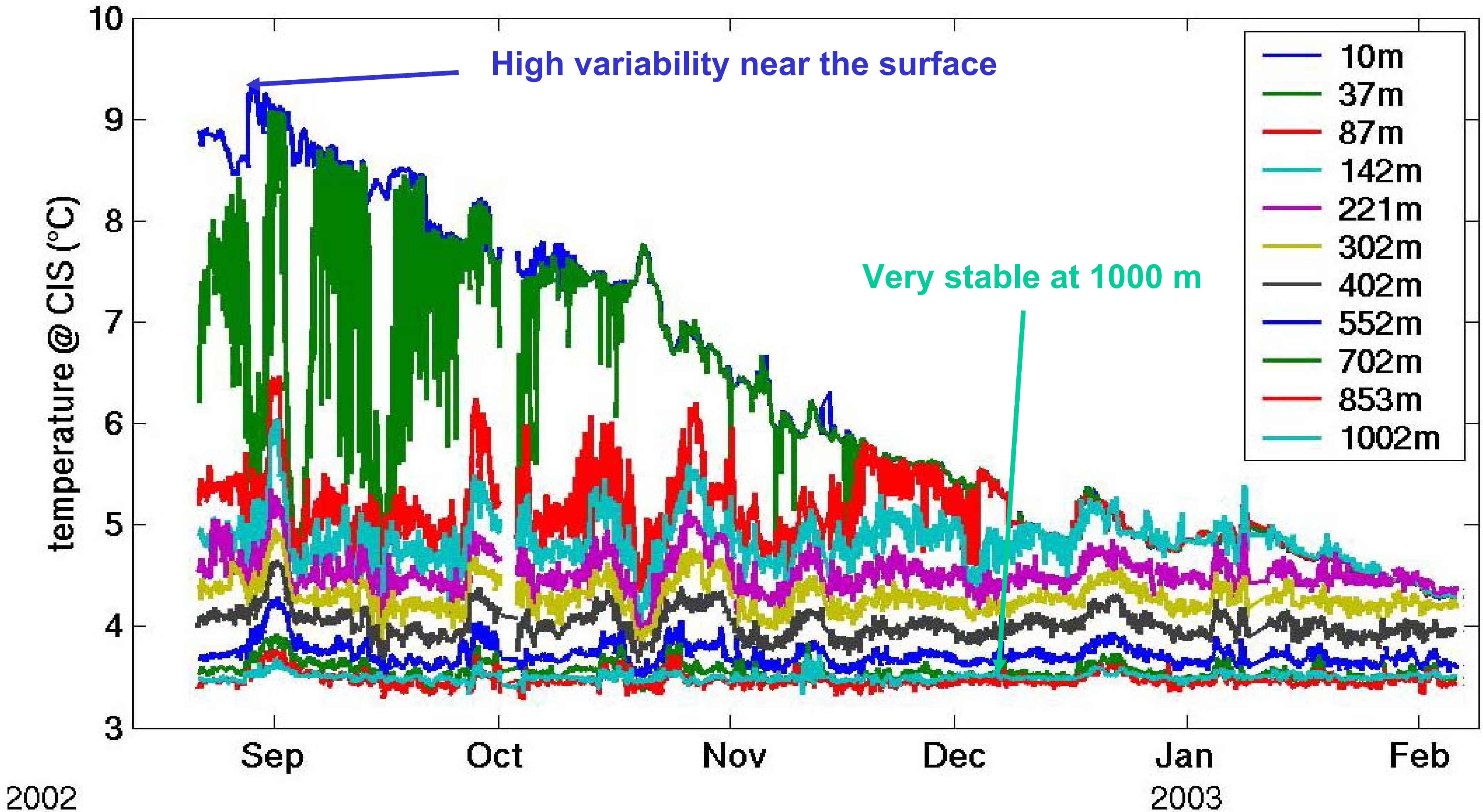
- **Coastal network:**
  - ⇒ Observing system is usually set up for specific applications.
  - ⇒ Number of parameters sampled can be important from physical, biochemical, atmospheric, seismic, biological to…. measurements
  - ⇒ Technical issues to be solved are very important (bio-fouling, interference with fishermen, vandalism …)
  - ⇒ Sampling interval from seconds to months
  - ⇒ Can imply huge volume of data when they are cabled systems with camera, seismograph, AUV etc…
  - ⇒ Build at national level with poor cooperation with other neighboring countries

  - ⇒ Examples: Water quality monitoring, Wind/wave/tide monitoring in harbors, Leo system in USA

LEO Instrumentation Used for the 2000-2001 Experiment



What is timeliness?



6 months of subsurface temperature from the Irminger Sea(Animate project)

None of these should be set up without coordination ….

- **Doing in-situ observations is expensive**
  - ⇒ A float is about 15k€
  - ⇒ A subsurface mooring (Temperature /Salinity/pressure) 250k€
  - ⇒ A Tao like mooring 500k€
    - \+ about 300k€ to maintain it each year…
  - ⇒ Should be set up in collaboration to maximise the benefits
- **Collaboration between countries mandatory to manage to set up and maintain global and regional networks**
  - ⇒ Some important bodies to know : WMO, IOC/Unesco, JCOMM, Pogo, GOOS, GEOS….
- **Collaboration needed to ease data access…**

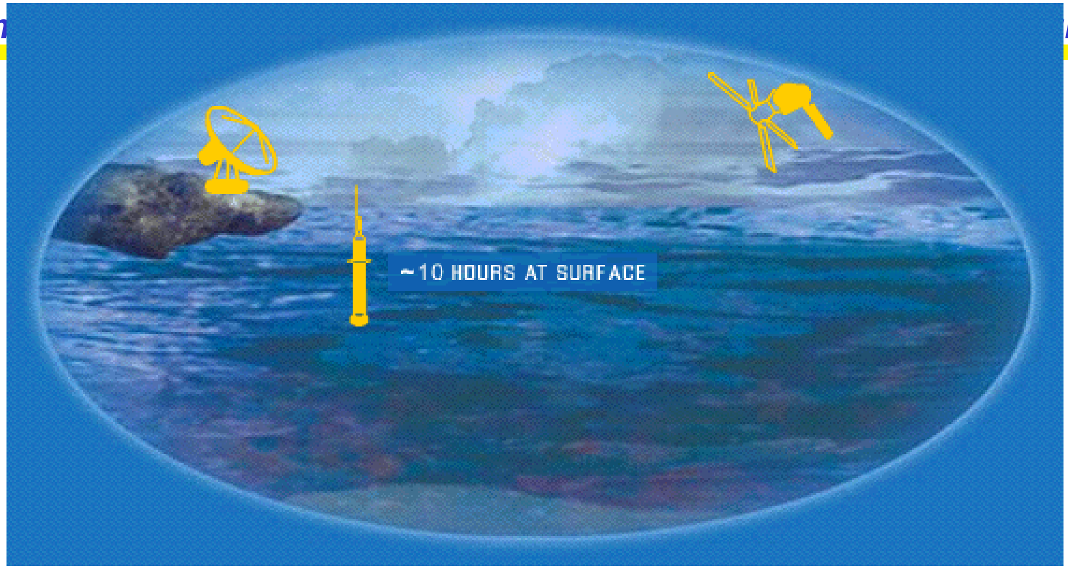
Some examples of observing systems

Argo

- **Rationale**

⇒ monitor and understand ocean circulation and water mass variability on a global scale through systematic observation of temperature and salinity fields

⇒ Provide data complementary to Altimetry to be assimilated in models



Argo

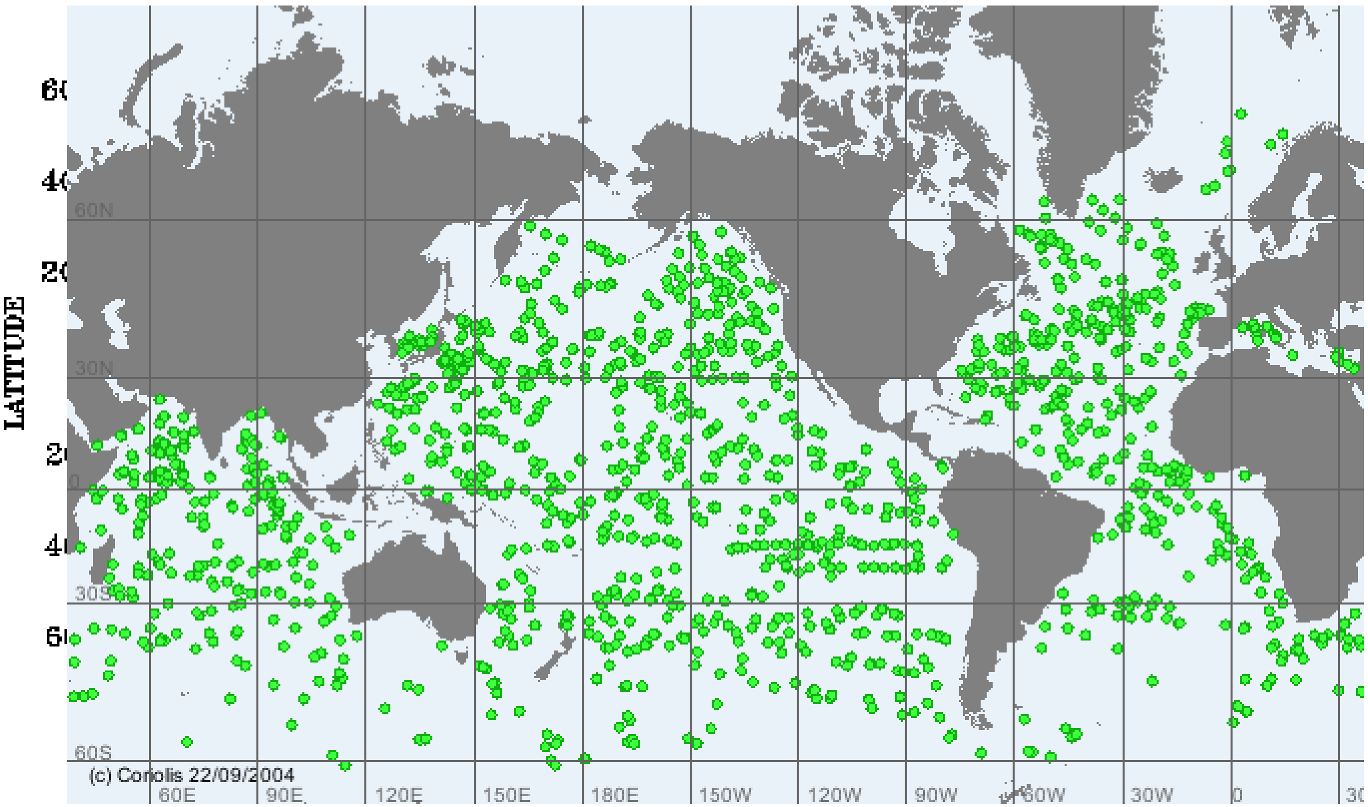
- Status

45% of network deployed by about 20 countries

≈1400 active platforms

North hemisphere pretty well covered BUT first floats are starting to die

Southern Ocean more difficult to populate



Argo

- **Strengths**
  - ⇒ efficient coordination at the implementation level
  - ⇒ collaboration at the scientific and technological levels to improve the quality of the instruments
  - ⇒ an efficient data management system able to distribute the Argo data in real-time within 24h designed early in the project
  - ⇒ Delayed mode process
- **Weaknesses**
  - ⇒ Funding not sustained
  - ⇒ Lifetime of Floats is not long enough yet (reality is more below 3 years than 5 years)
  - ⇒ Vertical sampling is limited by data transmission to shore
  - ⇒ How to correct precisely for drift when not much ground truth…especially for climate use!

OceanSites

➢ Rationale

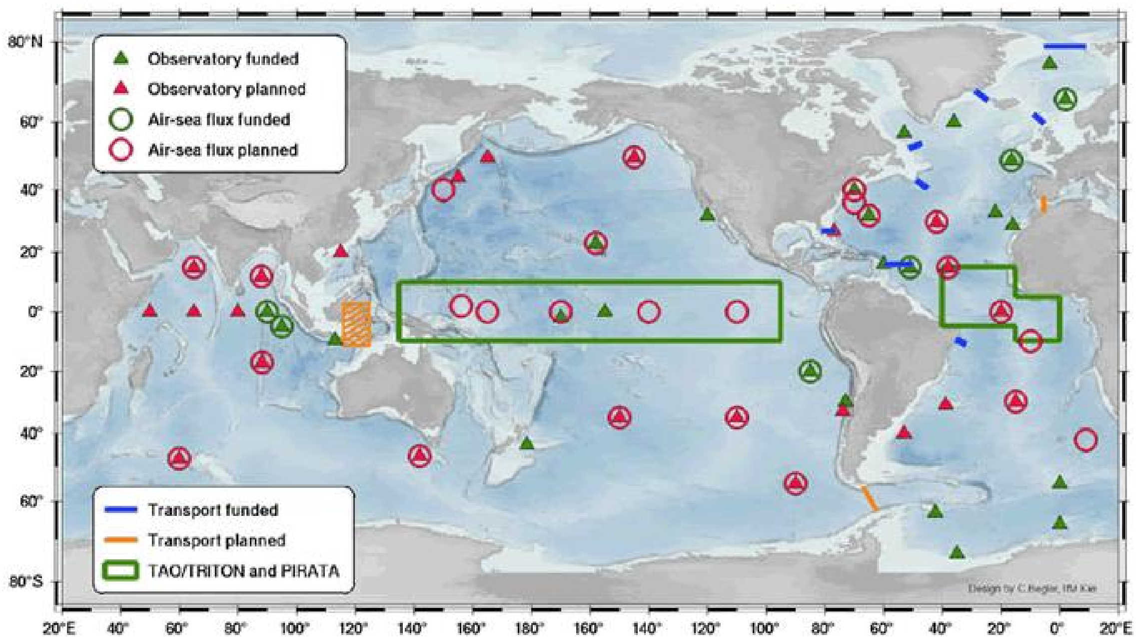
⇒ Provide routine observations in fixed locations, with high temporal resolution of a wide range of variables

⇒ Deploy in important/critical location (water mass formation, straits and passages, major current systems…) or in places expected to be representative of an ocean province/sub-basin

OceanSites

Status

- ⇒ Science team is set up and has worked a lot on network definition and starting on technological issues
- ⇒ Data management is under construction in coherence with Argo and Ocean.US programs



OceanSites

- **Strengths**
  - ⇒ Will provide the biochemical data that will be soon needed by the ocean models
  - ⇒ Provide international framework for national or PI initiatives (may help funding) in link with Clivar and GOOS
  - ⇒ Will create a WWW portal for timeseries data
  - ⇒ Helps to harmonize approaches and technologies, common advocacy and out reach

- **Weaknesses**
  - ⇒ A lot of technological issues to solve for real-time data delivery from everywhere
  - ⇒ International collaboration is hard to put in place especially in data sharing…

Even if an In-situ observing system makes great measurements in a sustained way, if the data are not available easily to the operational users, they will not be used because operational modellers have no resources to chase after these data

What will you do if?

- **Just imagine for a minute that your marvellous model of the North Atlantic is not representing the Mediterranean outflow at the correct depth ?**
- **Your adviser tells that you should assimilate more in-situ data ?**
- **What will you do ???????**

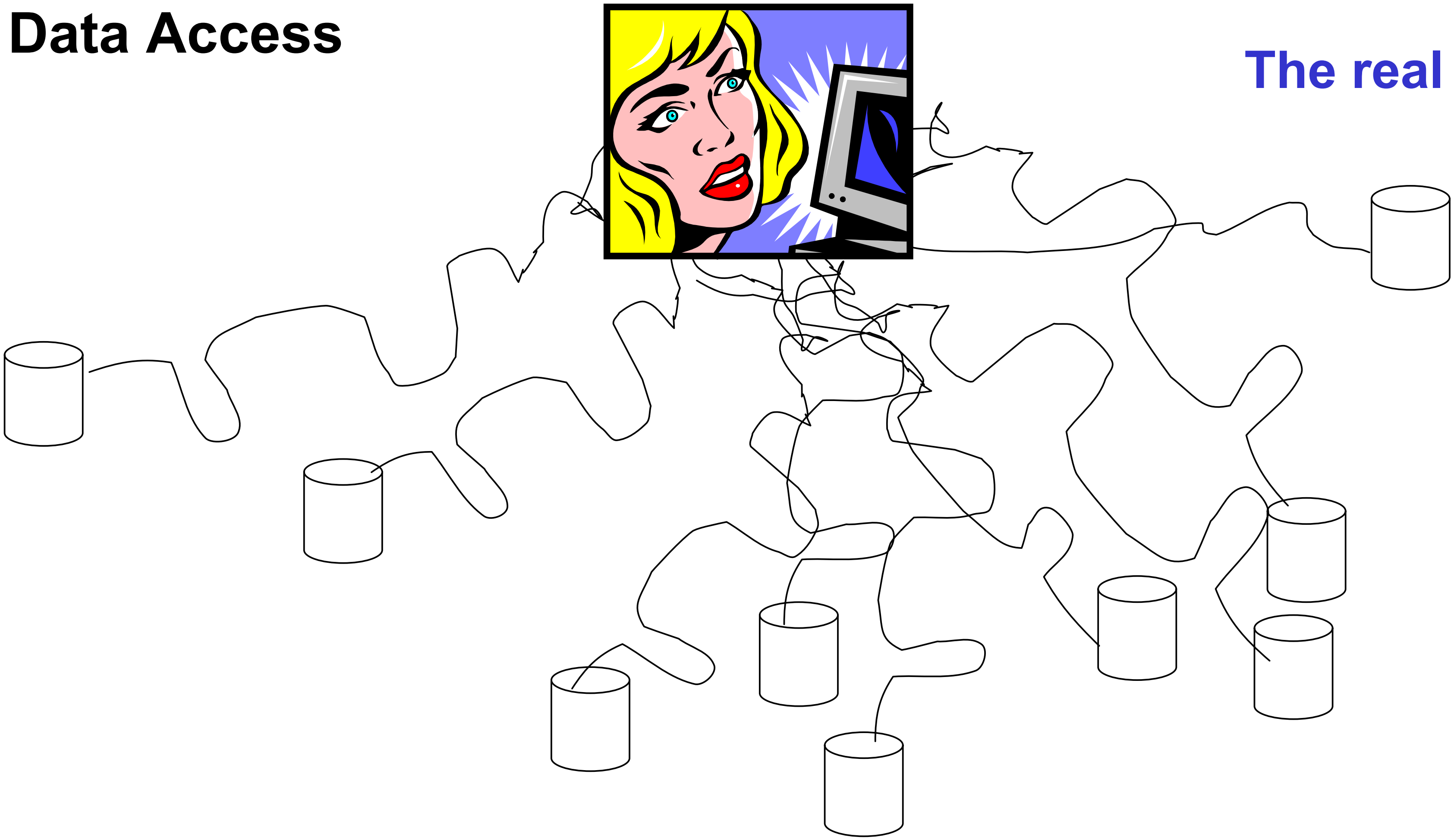
This example shows that we need a good data management for operational observing systems but what should be the goals of such a system?

⇒ a data system for operational oceanography must provide **easy** access to **quality controlled data**, in a **timely** way, on a regular basis, according to **procedures** that are clearly documented

What are the key elements of an efficient Data management network

- Data Access

The real World



What are the key elements of an efficient Data management network

- **Data Access**

The Ideal World

What are the key elements of an efficient Data management network

- Data Quality: Quality depends on the time you have to “clean the data ”
-

What are the key elements of an efficient Data management network

- **Data Format and Metadata**

What Are the Data system architectures used for data distribution

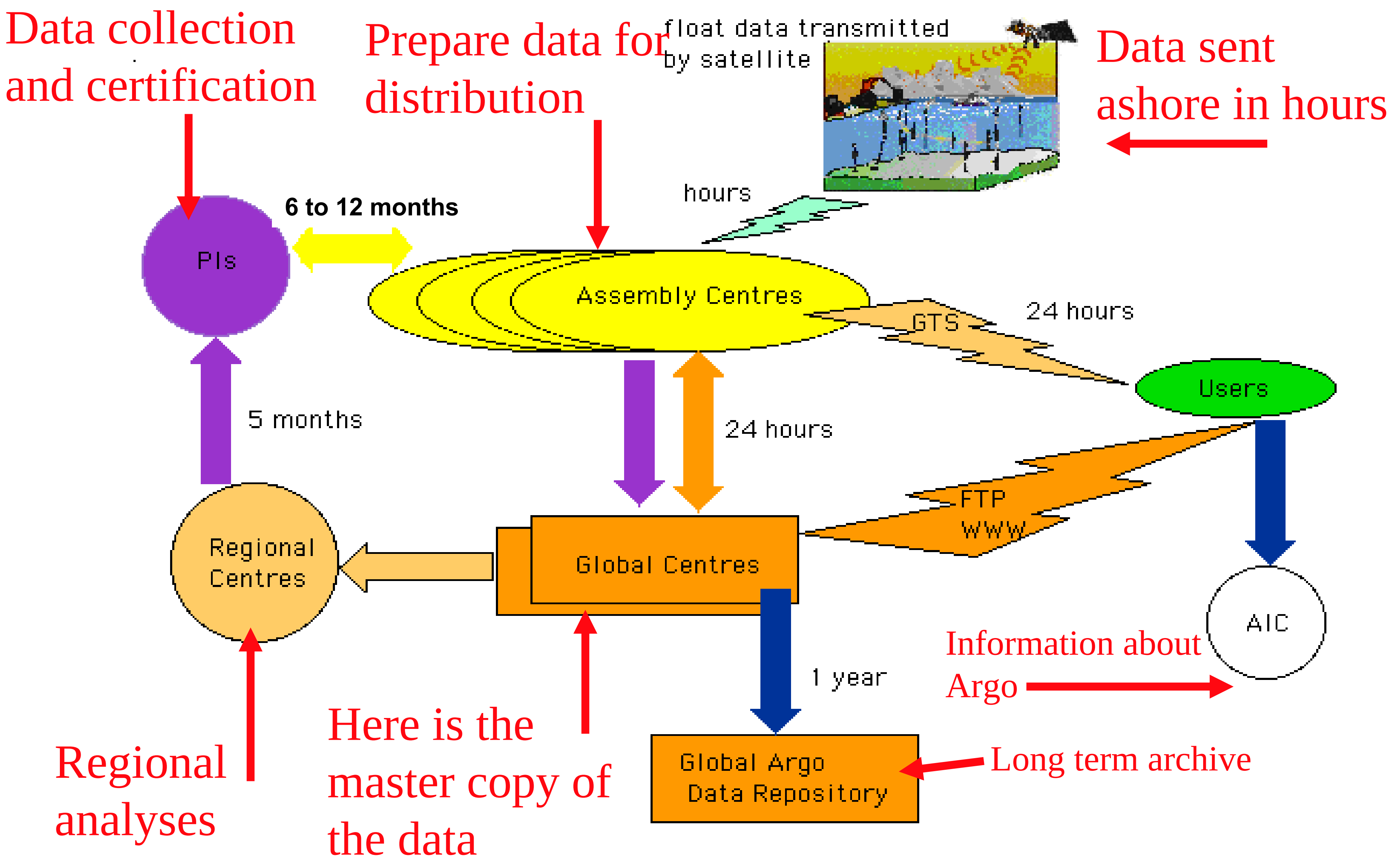
- **Distributed processing but centralized distribution**

  ⇒Argo

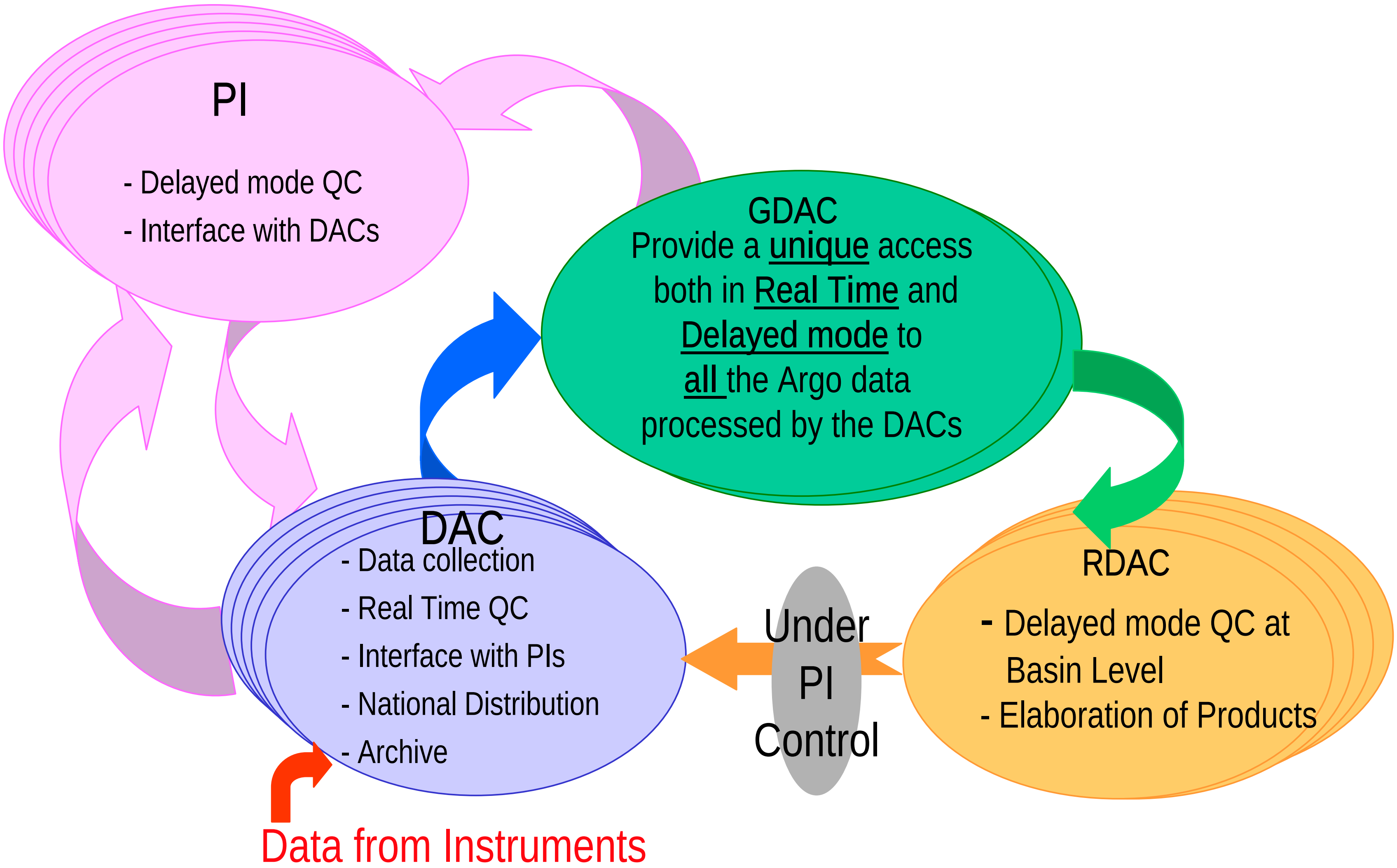
- **Distributed processing and distribution**

  ⇒Ocean.Us

The Argo Data System



Who does what?



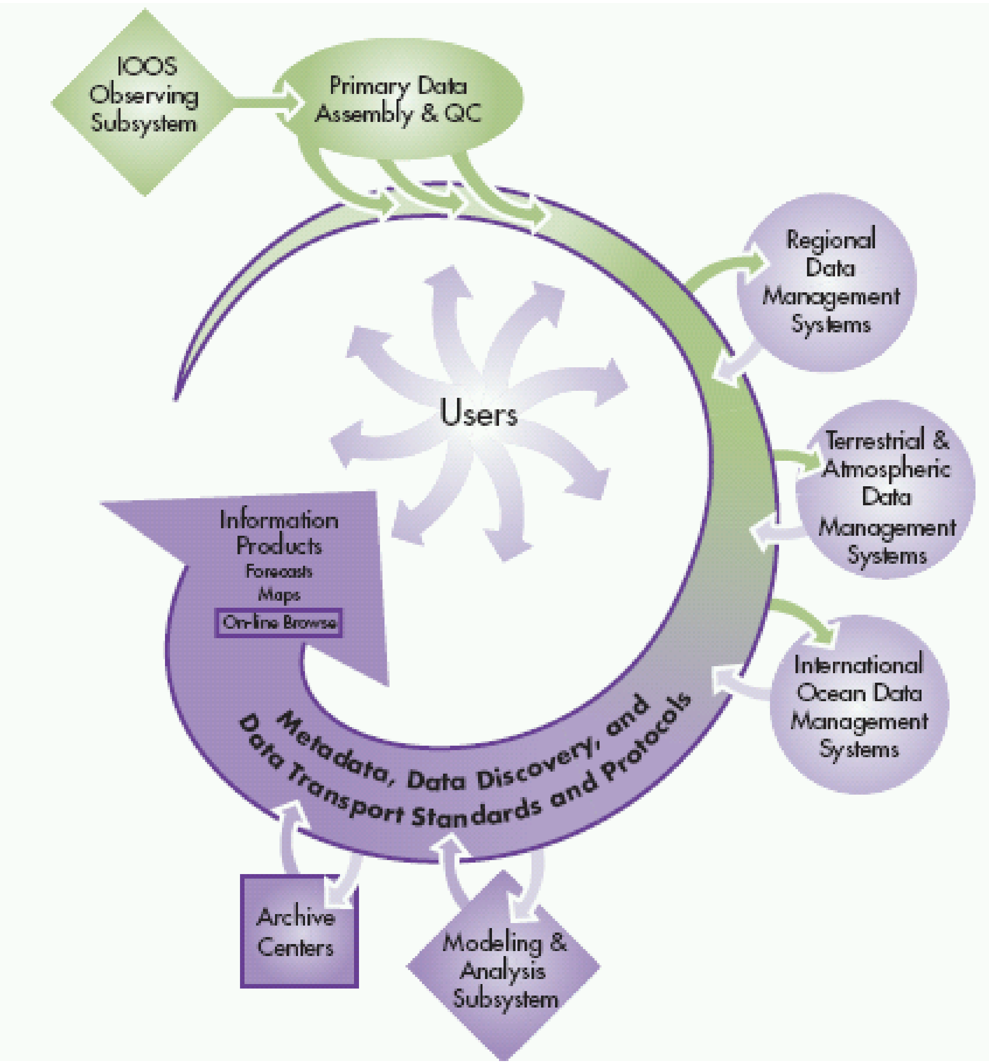
Advantages of such a system

- **A one stop shopping for the users where they get the best available data for ARGO in an unique format**
- **Data discovery and sub-setting tools are easy to implement as all the data are in the same place**
- **A robust system, as the probability that both GDACs fail is very small**
- **Easy to guaranty a quality of service in data delivery because GDAC have the control of all the elements in-house**

Drawbacks of such a system

- **Data are moved around the network and must rely on the "professionalism" of the DAC involved in the system to be sure that GDACS have the best profiles available**
- **Additional work at DAC level to convert their data from their home format to the Argo format. This may be hard to do for small entities**
- **Data format used for data exchange cannot evolve easily as it requires coordination among all actors before implementation. Since users, especially operational ones, do not like format changes it is not such a big problem**
- **If only one main server is set up than the system is fragile. Setting up a mirroring system can over pass this problem with additional synchronisation mechanisms**

Ocean.US : a concept under development at present

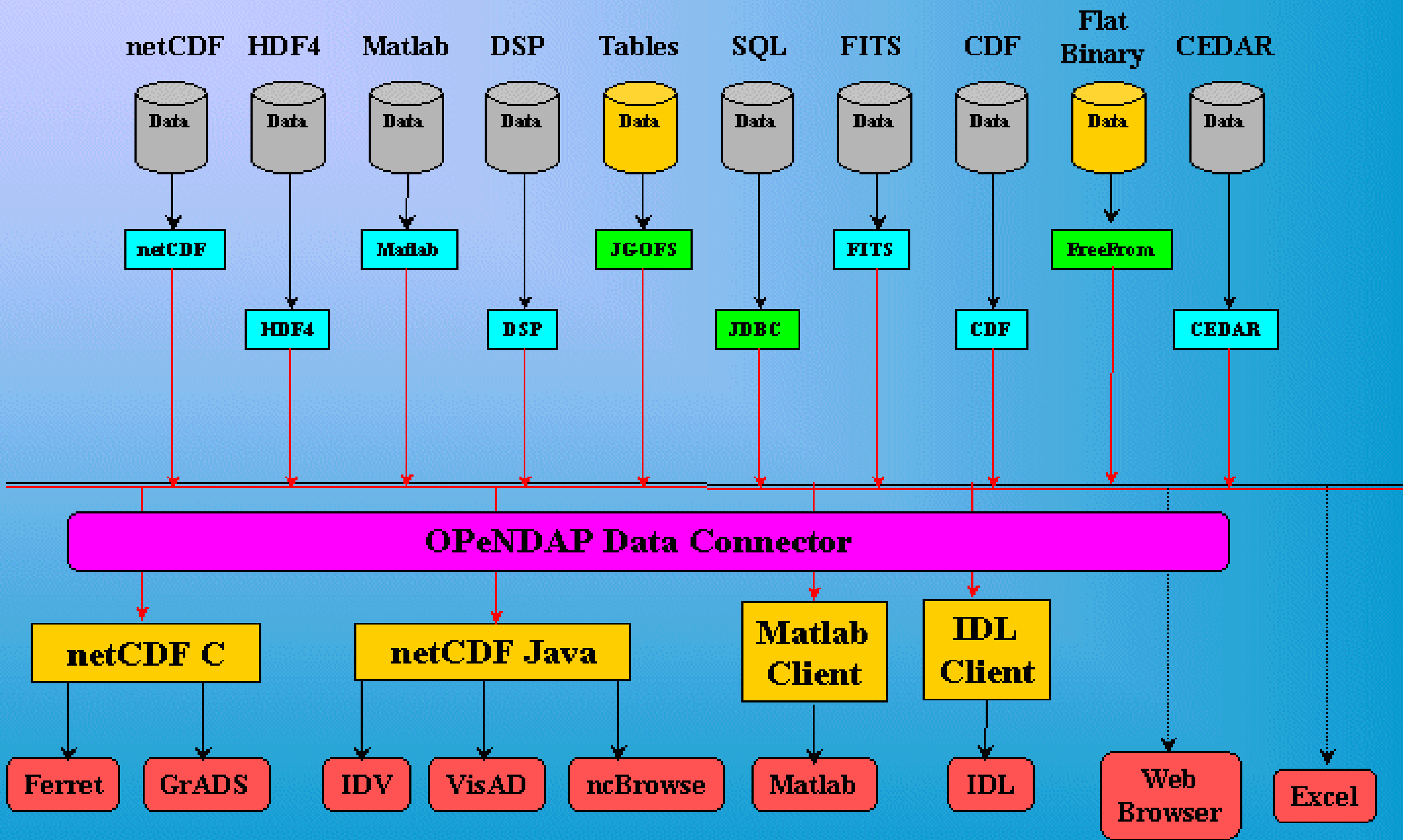


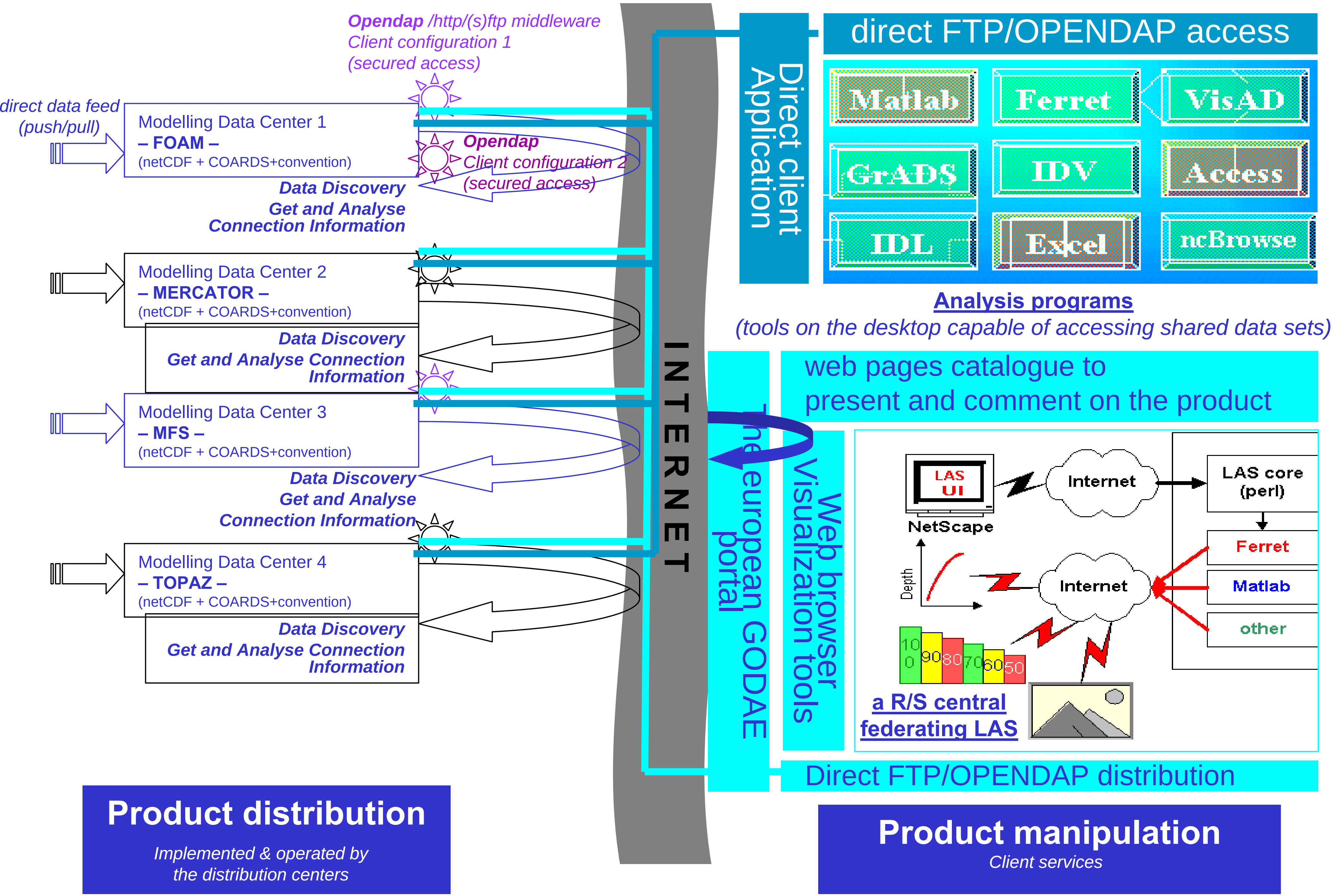
- **Data stay at processing data centers**
- **System implements a metadata catalogue to describe the data available**
- **Implements a data discovery protocol based on the internet network to find the data requested by the user**
- **Uses open source technology to implement data transport between the server and the client**

Part of it already exists now : OpenDAP (Open-source project for a network Data Access Protocol)

- ✓ **an open design to facilitate access to ocean data via the Internet** (http://www.opendap.org, since 1995, supported by the UNIDATA community)
- ✓ **Provides strong support for data stored in various formats and greatly reduces the volumes of data to be transfered across the Internet**
- ✓ **Compatible /connectable with many existing client applications** (e.g. NCbrowse, OpenDAP data connector, Excel, Matlab, IDL, Ferret, GRADS, IDV, LAS, Map Server) **- Acccess to the data can also be FTP enabled –**

OPeNDAP Client and Server Status





Courtesy F Blanc/CLS

Advantages of such a system

- Optimization of the resources (network, CPU, Memory,…) among the contributors,
- Data stay where they are generated preventing the generation non compatible duplicates among the network
- Built on internationally agreed standards that guaranty its efficiency in the long term and its adaptability because it will benefit from international shared developments

Drawbacks of such a system

- The system is not easy to set up because it needs a lot of international coordination, especially for metadata
- Even more work for small contributors because it requests important computer expertise
- It can be unreliable if some data providers cannot guaranty data serving on the long term. To be reliable such a system must rely on sustained data centres

Data quality: the cleaner you want the longer it takes!!!!!

- **It’s Fundamental because**
  - ⇒ accepting erroneous data can cause erroneous forecast
  - ⇒ BUT rejecting extreme good data can lead to miss important events in forecast…

- **It’s a challenge because no « ground truth » really exists**

- **For forecast data must be delivered with one day.. For reanalysis modelers request higher quality data set and even corrected data if possible.**

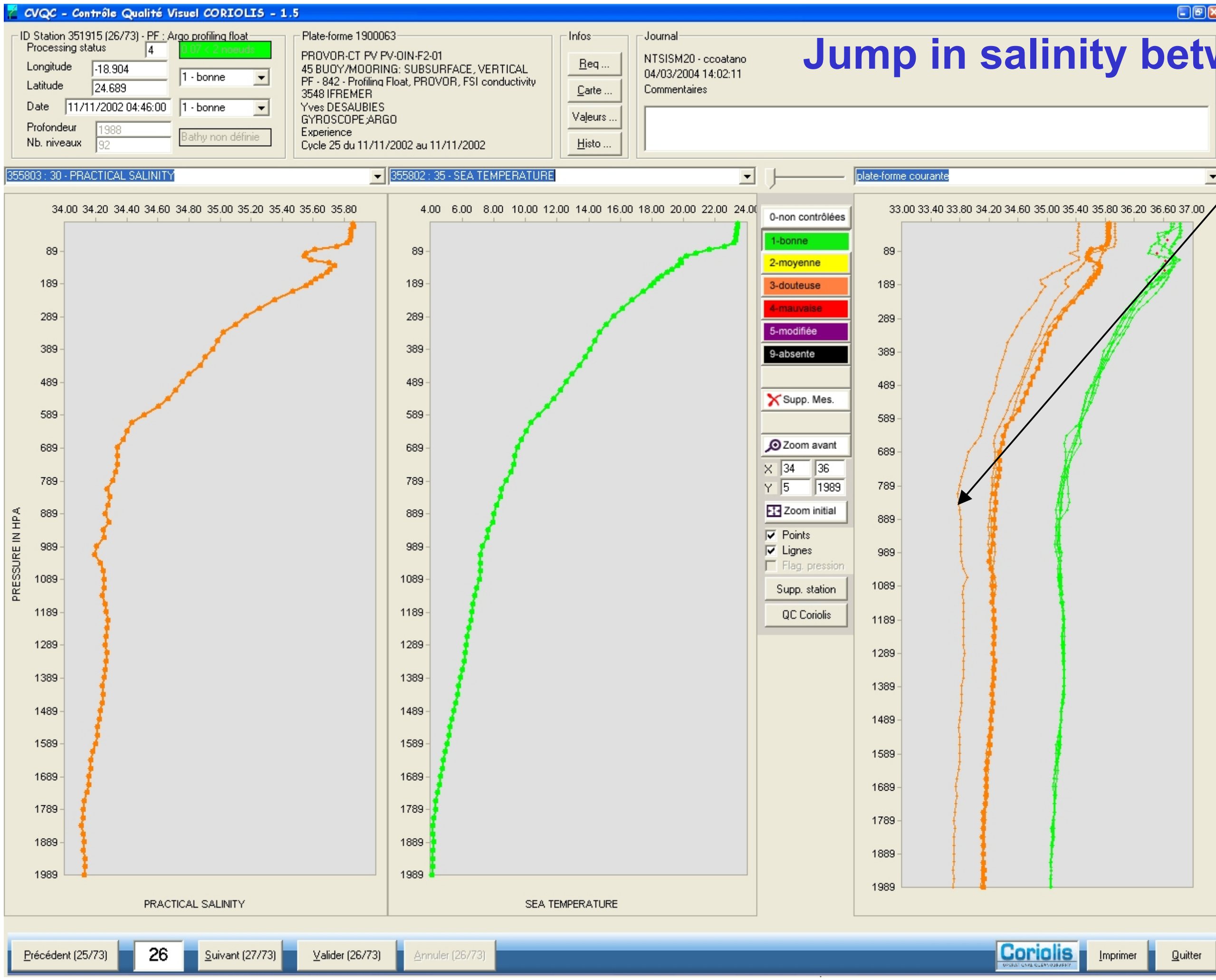
⇒ **2 steps: Real time and delayed mode QC**

Real Time QC for ARGO Available within 24h

16 automatic tests have been defined . They are applied automatically by all data centers . Rejected data are then checked by an operator. All good data are sent on GTS. All data together with QC flags are sent to Gdacs

- **Gross error tests: date, position, float speed at drift, temperature, Salinity**
- **Profile coherence: decrease of the pressure, spike detection, excess gradient between two points, density inversion, constant value or overflow for T or S**
- **Coherence between profiles: jump or big drift in temperature or salinity between two cycles**
- **Grey List: For the float in this list, all profiles must be checked by an operator because their behavior is "strange"**

Real Time QC for ARGO



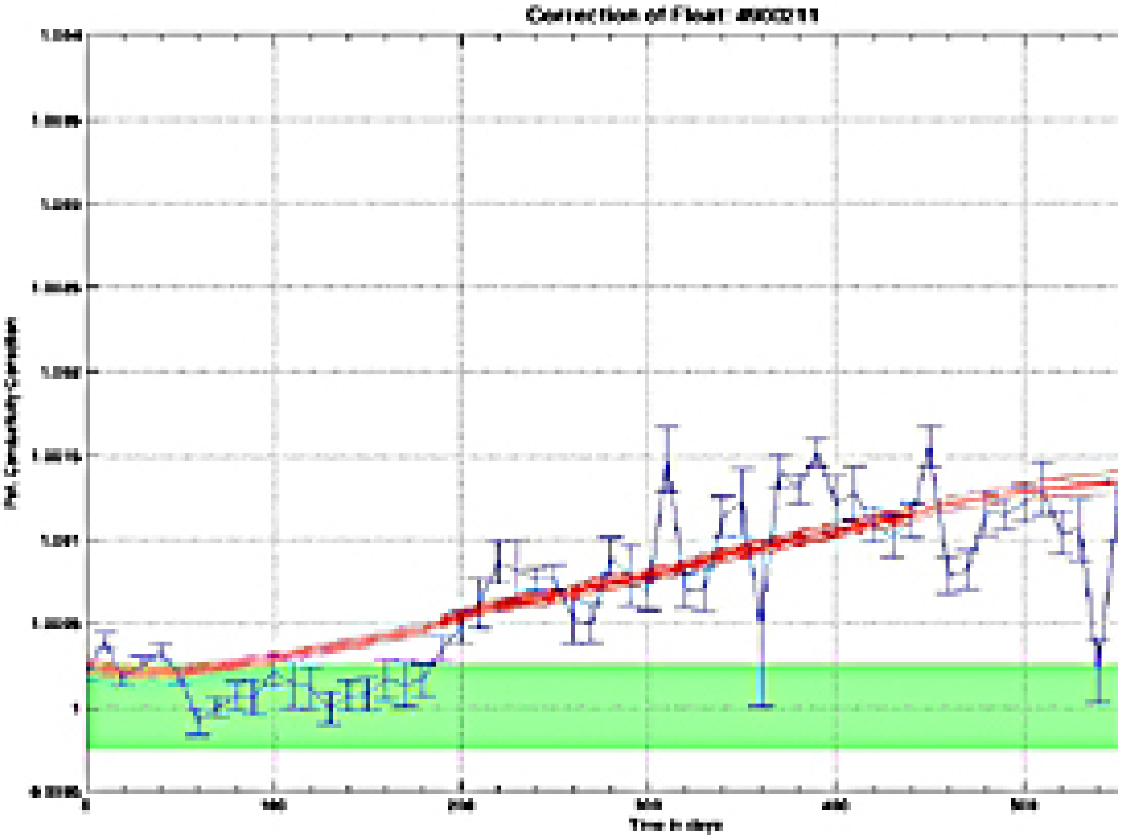
Jump in salinity betwenn two cycles

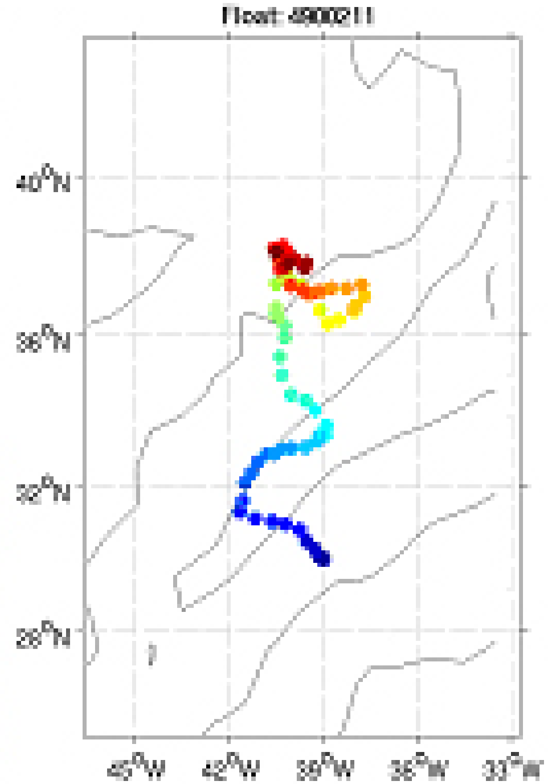
Temperature in green, Salinty in orange

Delayed mode QC for ARGO Available within one year

- **The free-moving nature of profiling floats means that most float measurements are without accompanying *in situ* "ground truth" values for absolute calibration (No CTD).**
- **Using 2-stage objective mapping methods, salinity data mapped from a historical database of existing profiles can be compared to float measurements**
- **A weighted average in the vertical (giving more weight to stable water masses) results in a single salinity offset for each float profile, as compared with the mapped data.**
- **Looking at the trend of these residuals allows detecting an sensor offset or a drift and quantifying this within error bars.**

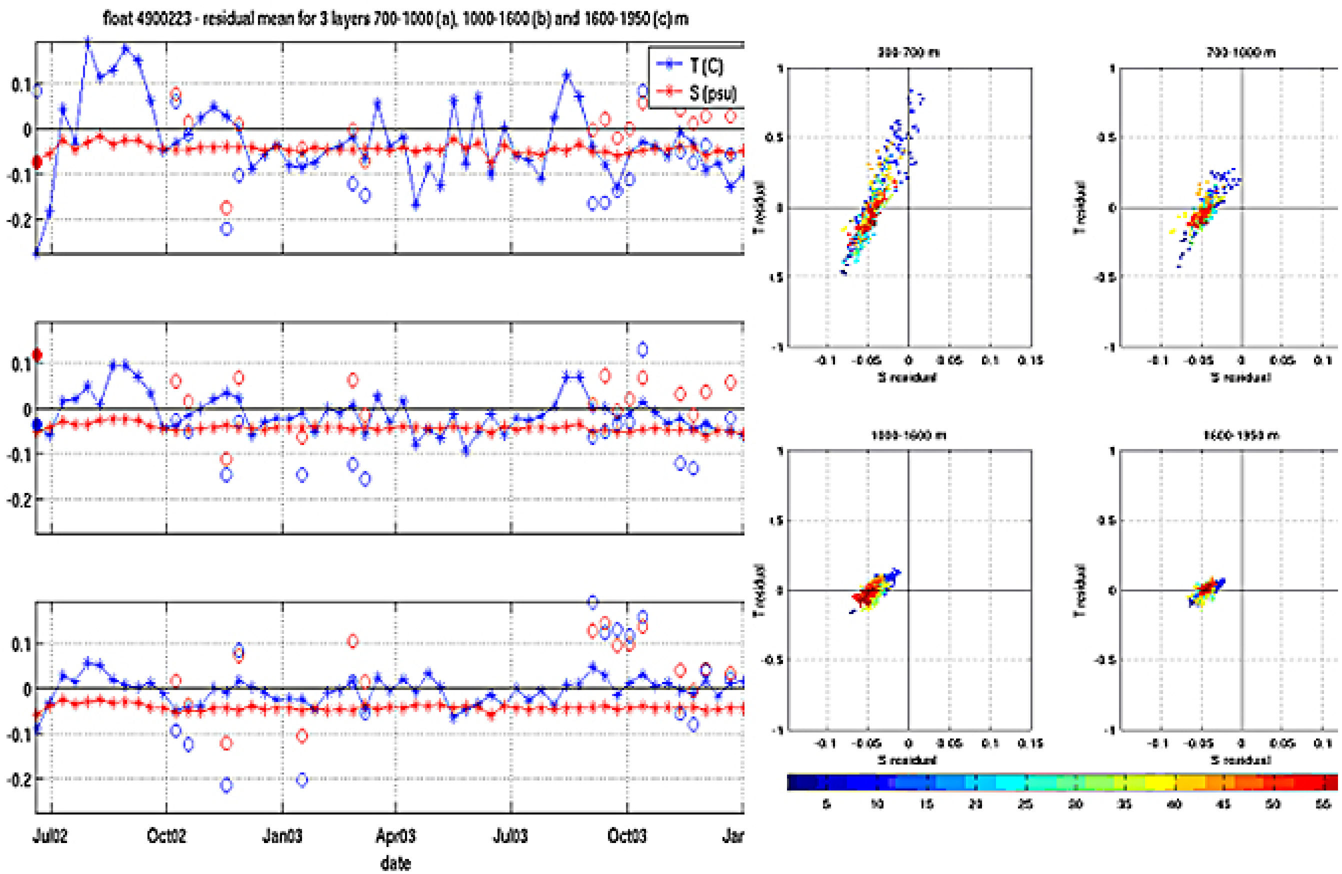
Delayed mode QC for ARGO





In blue individual cycle calculated corrections,
In red the proposed correction calculated by linear fit on a 6 month sliding window.
When correction stays within green limits (+/- 0.01PSU) no correction is applied

Delayed mode QC for ARGO



Another statistical method also used to estimate sensor drift consist of calculating weekly objective analysis with all the available QC'd profiles coming from CTD, moorings, floats and in monitoring the error residual for each float over time both in temperature and salinity by averaging these residuals on a number of levels.

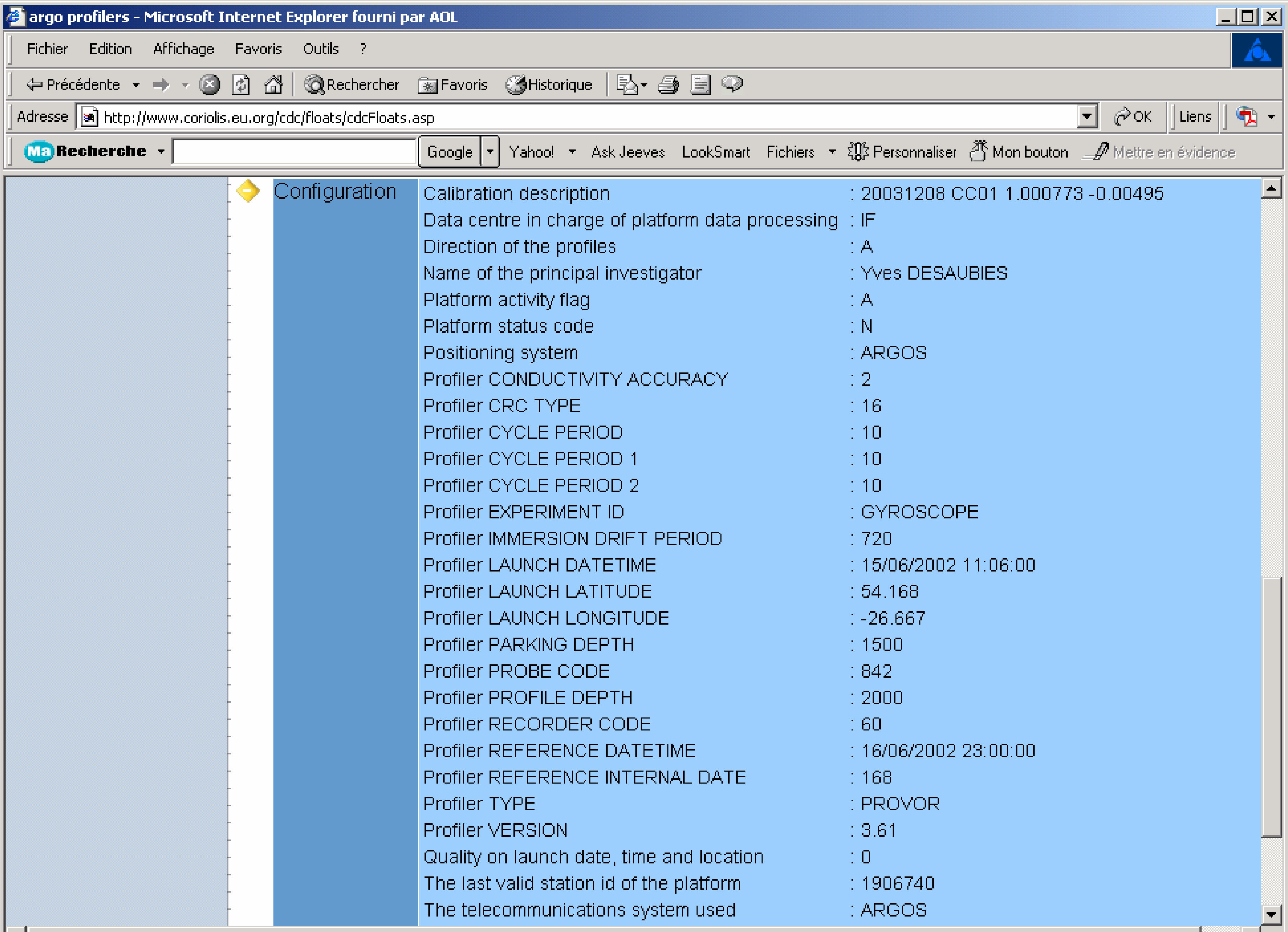
Data Format

- **Data format has always been a nightmare both for users and data managers and they are both dreaming of the "Esperanto" of data format.**
- **Historically:**
  - ⇒ ASCII format (easy to use by human eyes but not for softwares),
  - ⇒ binary format (more compact, easy for software but not shareable among platforms (Windows, Unix,…)),
  - ⇒ self-descriptive, multiplatform formats (Netcdf, Hdf…) that allow more flexibility in sharing data among a network and are read by all softwares that are commonly used by scientists.
- **BUT it's not enough…..**

Data Format

- **We need to agree on a common language**
  - ⇒ How can a user know that “subsurface temperature” is called TEMP in ARGO, Temperature in TAO, and is different from Temperature in GHRSST
  - ⇒ This is the purpose of metadata normalization handled by groups such as MarineXML or ISO19115
- **We need more than the measurements themselves:**
  - ⇒ Metadata that record the context of data acquisition (sensor, experiment, data center, PI,…)
  - ⇒ Raw data and corrected data to enable future reprocessing
  - ⇒ Quality flags that characterize the data
  - ⇒ History of what has been done on the data , what are the processing steps they have gone through,Calibration information if any.

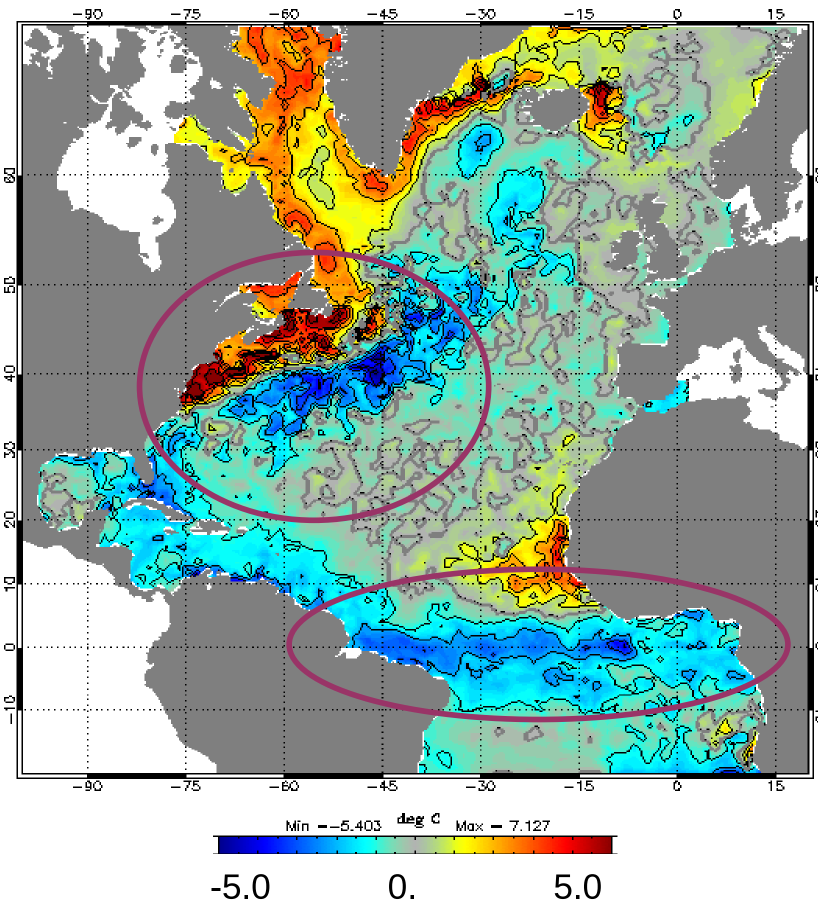
Data Format: metadata



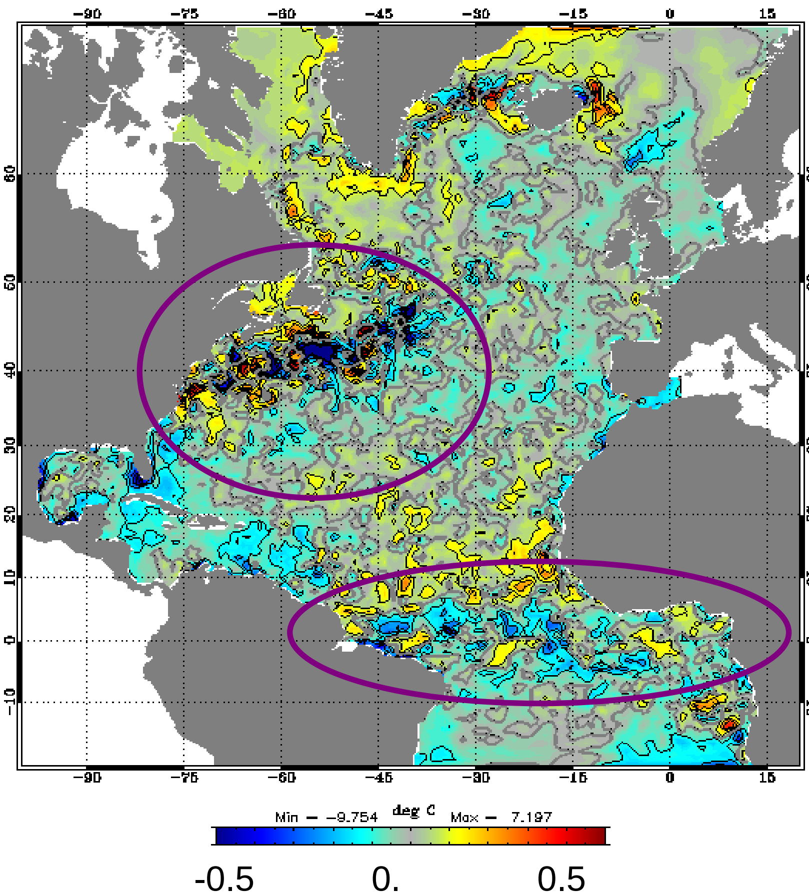
Calibration description	: 20031208 CC01 1.000773 -0.00495
Data centre in charge of platform data processing	: IF
Direction of the profiles	: A
Name of the principal investigator	: Yves DESAUBIES
Platform activity flag	: A
Platform status code	: N
Positioning system	: ARGOS
Profiler CONDUCTIVITY ACCURACY	: 2
Profiler CRC TYPE	: 16
Profiler CYCLE PERIOD	: 10
Profiler CYCLE PERIOD 1	: 10
Profiler CYCLE PERIOD 2	: 10
Profiler EXPERIMENT ID	: GYROSCOPE
Profiler IMMERSION DRIFT PERIOD	: 720
Profiler LAUNCH DATETIME	: 15/06/2002 11:06:00
Profiler LAUNCH LATITUDE	: 54.168
Profiler LAUNCH LONGITUDE	: -26.667
Profiler PARKING DEPTH	: 1500
Profiler PROBE CODE	: 842
Profiler PROFILE DEPTH	: 2000
Profiler RECORDER CODE	: 60
Profiler REFERENCE DATETIME	: 16/06/2002 23:00:00
Profiler REFERENCE INTERNAL DATE	: 168
Profiler TYPE	: PROVOR
Profiler VERSION	: 3.61
Quality on launch date, time and location	: 0
The last valid station id of the platform	: 1906740
The telecommunications system used	: ARGOS

Difference SST(model) –SST(Reynolds)12/06/2002

Without In-Situ assimilation



With In-Situ assimilation



Analysis of temperature and salinity over the Atlantic

Version 3:

Operational since December 2002

Method

- **Optimal interpolation (Bretherton et al.,1975)**

Data

- **Temperature and salinity profiles from Argo profilers, XBT, XCTD, CTD, buoys**
- **Time series (Pirata moorings, ..)**

Configuration

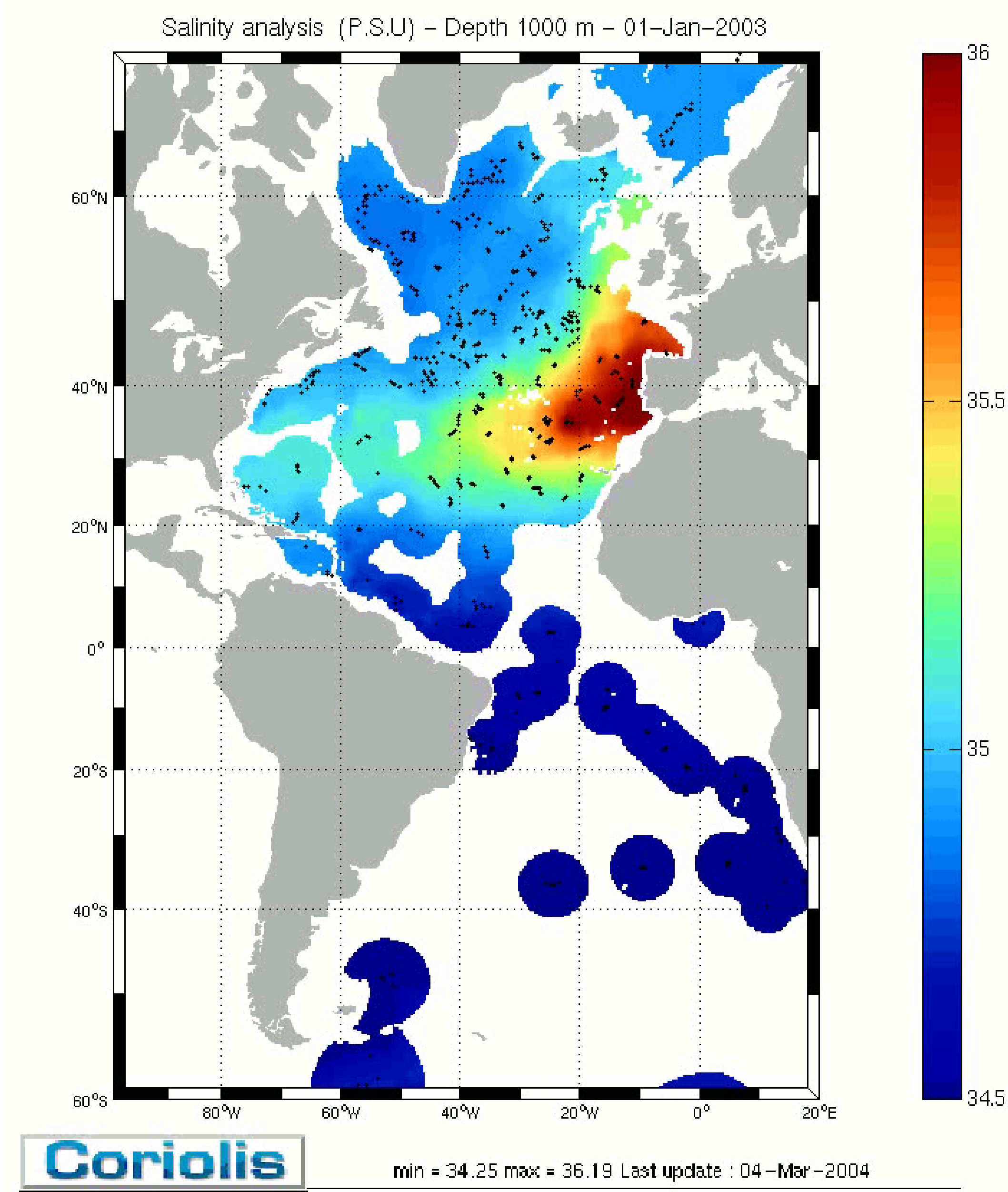
- grid with 1/3° resolution
- 59 levels from 0 to 2000 m

Output:

- T & S fields
- Analysis residuals for each observation

Foreseen:

- Extension to Global in 2004



Conclusions

- **Collaboration is mandatory to reach the Godae goals**
  - ⇒ Implementation level
  - ⇒ Data management level

- **Computer techniques won't solve all problems: we have to first agree on a common language before integrating apple and pears in the same basket….**

- **Modelers have to provide clearer guidance of what they need and the consequences induced by the non-sustainability of some networks…. It's your job to do these impact studies**

Some Challenges for the future

- **Agreement on efficient common quality control procedures both in real time and delayed mode to provide coherent datasets independent of what platform sampled it….**

- **Agreement on common languages to be able to use « Data Mining » tools that can work on semantic questions like:**

  I want a cloud free SST image over the Bay of Byscay in March 2004 together with the surface SST from drifters acquired in same area at same time