

2nd ARGO DATA MANAGEMENT MEETING

12-14 November 2001

IFREMER/France

1 INTRODUCTION

The second Argo Data Management meeting was held in Brest the 12-14 November 2001. About 30 participants from 8 different countries attending the meeting were greeted by Claude Marchelot of IFREMER. . He presented Ifremer institute as well as the "Informatique et Data Management" department activities. Coriolis data center is part of the department. The purposes of this meeting were first to present the achievements obtained during the past year, second to overview the status of the different Argo data centers and last to set actions for the coming months.

This Argo Data Management Committee was officially created at the last Argo Science Team Meeting in Victoria in March, 2001. It is co-chaired by Sylvie Pouliquen/Ifremer/France and Robert Keeley/MEDS/Canada

The meeting objectives on which the participants agreed were :

- 1. To finalize Netcdf formats. This action was considered by everybody as a main priority because it is needed by the national centers to start transmitting their data to the Global Data Centers (GDACs).
- 2. To finalize GDAC profile, track and metadata data handling either on FTP or WWW servers.
- 3. To develop a strategy for writing the Argo Data Management Handbook. The purpose of this handbook is to provide all the elements necessary for a new Argo Data Center to begin operations in harmony with the Argo data system
- 4. To settle the question of standardization of delayed mode quality control process and the definition of Regional Centers in charge of regional delayed mode QC.
- 5. To develop a strategy for inter-meeting work on specific tasks.
- 6. To settle a strategy for migration to BUFR for data transmission on the GTS.
- 7. To develop a strategy for countries to report status of their national data center implementation to provide visibility to the Argo program managers.
- 8. To start development of the long-term archive of the Argo data.

Most of these objectives were met during this meeting and you will find in this document a summary of the discussions as well as the list of actions issued at the end of the meeting.

The agenda is found in Annex 1, the list of participants in Annex 2, and all the relevant documents presented at the meeting on the meeting WWW site at the following address http://www.meds-sdmm.dfo-mpo.gc.ca/meds/About_MEDS/Meetings/ArgoDM/ArgoDMTHome_e.htm

2 REPORT FROM HYDERABAD MEETING

S Pouliquen presented a rapid report on the Argo Indian Ocean meeting that was held in Hyderabad, India last July. To meet Argo coverage, 600 floats need to be deployed. The deployments will mainly start in 2002-2003 and the main contributors will be India, Australia, USA, UK and Japan. Even if most of the countries around the Indian ocean are not able to buy floats, they are really interested by this program and willing to help providing for example vessel facilities. They may need help to put in place data centers and ocean modeling facilities to use efficiently these data.

3 DATA HANDBOOK

R. Keeley presented the draft form of the Handbook that contained a table of contents, and some preliminary text for the various sections. He also presented a series of questions about the contents and what should or should not be included in the Handbook.

In general the Handbook will be built as a series of linked web pages. The Handbook itself will not contain detailed documents, such as the description of the construction of a GDAC site. This

information will be referenced by the Handbook and contained in a separate document, appearing as a linked annex from the Handbook. A link to the handbook will be placed on the Argo Information Centre pages.

It was decided that a draft version should be prepared and ready in time for the upcoming Argo Science Team meeting scheduled for 12 Mar, 2002 in Hobart. Sylvie Pouliquen, Bob Keeley and Lesley Rickards agreed to do this, with Sylvie taking the lead.

In discussion, it was decided that detailed descriptions of DAC operations in the different countries would not be part of the Handbook. It was determined that there was a need for such descriptions but this would be provided through the building of appropriate links on the AIC pages to documents that each DAC should prepare to describe their operations and to be found on their web pages.

During the course of the meeting, documents for annexes were identified and the individuals who would provide these were noted. In this part of the meeting, it was recognized that an annex dealing with data formats was required, including not only the format structure, but also a description of what the contents would be for each field in the format. Thierry Carval agreed to provide this.

It was decided that general descriptions of data monitoring products, or tools would be included in the Handbook. These will be required by the data system, but are not something that would be described in detail. Such details will be found at the AIC.

Another annex required was one that extends what has already been written to describe the operations of the GDAC. This annex needs to provide the details about what steps a DAC must take to make a connection to the GDAC, and the responsibilities of each, DAC and GDAC, in the data transfer process.

Though not part of the Handbook, it was suggested that information be gathered from the various float manufacturers of the models that have been deployed by different participants. This information would be referenced as another annex. The AIC was requested to accumulate this information.

4 DATA FORMAT

At the 1st Argo data Management meeting it was decided that a unique format will be used on the Internet for data distribution to users, and for data exchange between national data centers (DACs) and Global data centers (GDACs).

It was also decided that not only profile data but also metadata, trajectories and technical data will be included in this standardization. A small working group, lead by T. Carval started 6 months ago by email. It had been decided that defining a format based on netCDF was a good option because;

- It is a widely accepted data-format by the user community.
- It is a self-describing format for which a lot of tools are widely available.
- It is a reliable and efficient format for data exchange.

There was some concern expressed about dividing the data into different pieces (i.e. profile, trajectory and metadata files) and the possibility that in so doing, some data may be lost. While keeping everything together in a single file structure is expedient for those who want data on a float by float basis, it makes provision of data by area and time very difficult. As both needs where mentioned by the users, it was stated in the end, that the format should be able to handle either a file per profile or a file for all the float profiles. The trajectories, metadata and technical will be provided in one file per float and per type. The data system was cautioned to take the necessary steps to ensure data were not lost.

The proposal made by the group was discussed during the meeting and some modifications were agreed to. A new version, Version 2.0, was produced and is included in annex 3.

Profile files will be either single or multi-profile to cope with all user needs.

This version will be used for the coming year for data exchange between centers and data distribution to users. A review will be made at the 3rd Argo data management meeting, and the format may be updated if necessary.

If remer will write the User Manual that describes these formats in detail. It will be available for the next Argo Science Meeting in March 2002.

5 QUALITY CONTROL (REAL-TIME AND DELAYED)

The discussions were handled in two parts with R. Keeley providing information about the real-time procedures that had been decided by email, and D. Roemmich leading the discussion on delayed mode QC procedures.

5.1 Real-time

There was a general discussion about the strategy of using less stringent tests and thereby permitting bad data into the real-time data distribution system versus more stringent tests and excluding more data from real-time distribution. The meeting agreed to use less stringent tests expecting that many users of the data in real-time will be modelers and will have automated assimilation procedures to remove profiles that fall outside expected values.

C. Schmidt presented some statistics on the performance of the automated real-time procedures (details are in the US National Report). R. Keeley offered to examine the performance in greater detail to get statistics on profiles that pass and fail the procedures, but also to look at how many floats fail that should not as well as floats that pass but have problems in their profiles.

D. Roemmich remarked that the real-time procedures do not exploit the fact that a profile from today should be quite similar to the previous reported profile. He noted that he had an algorithm that should be tested for inclusion in the real-time procedures. The test would be better in the T-S plane rather than T-P or S-P. He also noted that A. Wong could contribute to this discussion having developed QC procedures to be applied in delayed mode. The meeting agreed to create a working group of Bob, D. Roemmich, A. Wong, R. Molinari and R. Keeley to develop the algorithm.

The U.K. reported that they are having Service Argos decode data from the U.K. floats, and turn the data into TESACs for distribution on the GTS. The U.S. developed software to implement the agreed QC tests for the Service Argos office in the U.S. It was not certain if the French offices of Service Argos were also using the procedures for the U.K. floats. The U.K. will pursue this with Service Argos and S. Pouliquen offered to assist.

The U.S. noted in its national report that it was using a Levitus climatology test and a test against a first guess field generated from an NCEP model to screen data for the GTS. After discussion, they agreed to stop using the tests to determine if data should go to the GTS, but to continue to use the tests as a way to see how effective such tests would be if everyone applied them. They will provide the results of this at the next data management meeting.

5.2 Delayed Mode

The discussions on delayed mode quality control began with a presentation by A. Wong. She described the technique that she has developed to check for salinity drifts in float data. To be effective, she requires about 10 profiles in order to build up the needed statistics. Even then, her procedure cannot distinguish between seasonal variability and sensor drift. In the end, the PI needs to make the decision about what reflects the ocean properties and what is caused by sensor problems.

D. Roemmich remarked that it was the responsibility of nations as part of their contribution to Argo to undertake delayed mode QC. He also remarked that how a country does this was its own concern. However, it was critical that all countries agreed to do the same thing so that users of delayed mode Argo data would have confidence in the uniformity of quality. He suggested that all PIs should look at Wong's procedures and either agree to use them, or propose an alternative. He agreed to contact members of the AST to ask that this debate take place at the next meeting in March of next year.

Since 10 profiles take approximately 100 days to accumulate, Wong noted that if her procedures were to be adopted by everyone, it was unrealistic to expect delayed mode data to be returned from PIs in the 90 day period proposed by the AST. Instead, a 150 day period was suggested as more realistic.

There was some discussion about what to do with the error correction information produced by Wong's procedures. It was decided that there must be a place in the data format to accommodate these. Consequently, the netCDF format for profiles was modified.

There was discussion about the role of regional centres in carrying out quality assessment. S. Pouliquen pointed out that these regional centers were important because they will perform delayed mode QC made on Basin level, proposing correction on all the float in this basin, whatever nation had deployed it. We must be aware that most of the time basin will be populated by different nations. R.Molinari remarked that some centres would want to look at all of the data (floats and others) collected in particular regions. In looking at the composites, it is likely that questions would be raised about some data. Roemmich replied that where there was a PI for a float (and not all deployed floats would have a PI) they had the last word on what was considered correct data for their floats. If questions were raised, they would be directed to the PI who would make the final decision and forward any modified data files back to their DAC who would pass the data to the GDACs.

It was not clear what centres were interested in doing these regional analyses. However, for floats with no PIs, it would be these regional analyses that had the final word on data quality. It was suggested that an action to incite countries to set up such centers should be taken by the AST.

6 GLOBAL DATA CENTERS: FTP AND WWW SERVERS

At the last meeting we agreed on the following points:

- Users will have access to all the Argo data at two places, called the Global Data Centers (GDACs) located at Coriolis/IFREMER/France and USGODAE/FNMOC/USA
- These GDACs will provide both FTP and WWW interfaces to users.
- The FTP sites will be the same and the GDACs will propose an architecture at the 2nd Argo data management meeting together with data exchange and synchronization procedures.
- Concerning WWW sites, GDACs will provide similar services but no requirement was put on them to have the same interface. A review of the two sites will be made at the 3rd Argo data management meeting.

6.1 FTP server

In the proposal made by IFREMER and FNMOC the same ftp architecture was used for data exchange and distribution to users with no data duplication on the sites. The ftp site is organized in 3 main directories :

- by Data center and Float
- by Ocean(Atlantic, Indian, Pacific)
- by processing dates

This organization seems to meet scientists' needs. Nevertheless, at the lowest level of these directories, some scientists want to have multi-profile files instead of one file per profile as proposed originally. From the GDACs both a file for all the profiles of a float and a file per profile will be available because both needs exist. This implies data duplication on the servers but GDAC managers agreed to implement these functions.

The data transfer between national data centers and global data centers, as well as synchronization process between global data centers were accepted as described in the document "US GODAE/IFREMER Data Servers as part of the Argo data distribution network " version 1.2, March 2001.

Coriolis/IFREMER	 ✓ FTP site set up the 4th July 2001, automatically updated daily since July 2001 : incoming ftp directories for AOML, MEDS and US-GODAE were created. Accounts for other national centers will be added as needed. ✓ November 2001 : synchronization process once a day between Coriolis center and US Godae server. ✓ December 2001 automatic processing of the national centers data who will be able to create Netcdf Argo float files.
-------------------------	---

Information on implementation at each global data center was provided by both GDAC managers:

6.2 WWW server

The Global Data Centers WWW server will have to serve a large community of users:

- PIs who have deployed floats
- Scientists preparing float deployment
- Scientists who want to study an area
- Ocean modelers who need data for their model
- Argo managers who report on Argo program
- Public (media, schools, museums, ...)

To fulfill their needs it was accepted that the WWW server will have to provide visualization and sub-setting tools. For each set of tools we have defined the main functionalities:

	•Access to any Temperature and Salinity data in an
	area
	\Rightarrow Individual profiles
	\Rightarrow Multi-profiles graphics
	⇒ Time-series parameter evolution (depth- time contour plot)
	\Rightarrow Temperature-Salinity diagrams
Visualization tools	•With each T,S display provide a map with float
v isualization tools	locations for orientation; associate line colors with trajectories (if many floats) or trajectory segments (if one float)
	•Access to float trajectories (maps with trajectories or
	surface/depth displacements or profile positions, option to color-code deep arrows with deep T or S)
	•Access to manipulation tools for graphics: Choose
	subsets/groups, only alive floats or all, set axes/zoom,

	 control contour/color map, merge "historical" data as background, color coding for trajectories (e.g. by PI/country/project, by date, by drift depth,) Statistics on data coverage and its evolution, float/performance statistics,
Sub-setting Tools	 •Temporal, geographical and drift-level criteria •Full resolution profiles or interpolated on standard levels • Submerged or surface or total displacements • Parameters (T, S, positions, deep/surface velocities,) • [Experiment/Cruise, PI, accuracy/error level] • Format ⇒ASCII, Netcdf ⇒One file per profile, float, day, week, month, ocean, ⇒

A subset of these functionalities will be available on one of the GDACs WWW site (Coriolis) and presented in March 2002 at the Argo Science team.

7 AIC STATUS AND ROLE

M. Belbeoch presented a report of activities that have taken place at the AIC. He remarked that an Argo label has been designed and will soon be ready to provide to manufacturers to put on their floats. It is printed in the 6 official languages of the U.N. and provides the telephone number of the AIC. Discussions with the float manufacturers were carried out to ensure application of the label would not have any detrimental effects on float performance. The cost is approximately 1 USD per label.

One of the main tasks of the AIC is to provide information about float deployments and when floats are likely to enter the EEZs of nations. The AIC provides a web-based form for float deployers to fill out each time a float is deployed. The information automatically is routed to every IOC representative via an email message. The AIC web pages also support a mapping tool that shows the locations of floats that have reported. By clicking on a float location, the tool takes the user to the web page on the original web site where the information is located.

Discussions decided that it was very effective to have the map display presented by the AIC. The display, however, should have 30 degrees E as the map edges rather than 180 degrees E as it was now. The meeting asked Belbeoch to explore what tools would permit this.

The meeting noted that the individual pages linked from a float position on the AIC map often allow for data downloads. This has been useful to date to provide access to data before the GDAC servers come on-line. However, the servers will soon be operating and in order to protect the integrity of the GDACs to serve the master version of data, the present links from the AIC map to individual web pages is to be turned off. Instead, the AIC should point clients who want data to go to the

GDACs. It is still important for the AIC to point to national web pages concerning Argo, but this should be done from other pages.

There was also some discussion to decide a consistent way to identify floats. In the TESAC code form, a letter Q precedes the float identifier. Some DACs were using the QC as part of the identifier and some were not. The meeting agreed that the Q would be used and asked the AIC to make the necessary modifications to its pages to reflect this.

The meeting completed this discussion with the creation of a working group to guide the developments of services on the AIC pages. The present membership would be S. Pouliquen, R. Keeley and M. Belbeoch. Another member would be sought to help.

8 NATIONAL REPORTS

National reports were presented by Canada, France, Japan, Korea, Russia, U.K., and U.S.A. A written report was prepared as a PowerPoint presentation by Australia. These reports (with updates) are in an annex to this report.

No report was received from China. However, the U.S. reported that someone from China was going to be visiting AOML to look at their processing system. This visit was being supported through the POGO program. A visit was also going to be made to Steve Riser's operations as well.

It was noted that some DACs or agencies have easier connections to the GTS than others. It was suggested that agencies having difficulties should contact either of the co-chairs of the Data Management Team who would then talk to other centres to see what help can be provided.

9 ARGO SCIENCE TEAM REQUIREMENTS

Unlike the first Argo Data Management meeting, members of the Argo Science Team had been invited to participate at this 2nd Argo Data Management meeting. The purpose of this invitation, was for the group, to be sure that, at least, the main Science Team requirements were known by the committee, and that the implementation the Argo Data Management group was proposing, was in agreement with these requirements.

The main messages the Argo Science Team had passed on the committee were:

- Even if the Data Centers use the experience acquired on "Upper Ocean "data handling, the DACs mustn't forget the characteristics of profiling floats and especially the stability in time of the instruments.
- At the level of the Argo network we will have to achieve a quality that would guarantee that any Argo profile is equivalent to another wherever processed, manufactured, deployed, etc. This implies that the next main action is to work on Delayed Mode Quality Control and on the best way to implement the Regional Data Centers in charge of delayed mode quality control at the level of basins/oceans.

In the future the Argo Science Team and Argo Data Management Team will have to work closely together in order to achieve these ambitious goals.

10 TRANSITION TO BUFR

A brief discussion took place regarding the transition to using BUFR to transmit data over the GTS. No specific actions were taken at the meeting. Keeley and Pouliquen will seek members to participate on a working group to start this work.

11 LONG TERM ARCHIVE

The last meeting decided that the Argo program would rely on the U.S. NODC to ensure the long term archive of the data as it is already doing for other international programs like WOCE. C. Sun made a proposal to be reviewed for two main reasons:

- The Long time Archive Server must be complementary and not duplicate the GDAC's services.
- The Argo Science Team must specify more clearly their needs concerning this archive.

The proposal made by NODC had two main functionalities:

- Recover and archive from GDACs the different profile versions that are available on GDAC FTP servers.
- Provide a WWW interface to Argo data.

Concerning the first point, Argo Science Team members who were at the meeting, specified that they only wanted to archive the original and best value of each profile. So NODC only needs to act as a backup of GDACs for real time data (no access to data for users, only for GDACs if necessary) and only keep one version of each profile in the archive.

Concerning the second point it was remarked that NODC mustn't duplicate what was already done at the GDACs. Because of GTSPP, WOCE and other archives NODC is handling, what NODC can provide as a complement to GDAC is an access to Argo data and also to other in-situ measurements. NODC was advised to modify his proposal in this way. R Keeley, S Pouliquen and C Sun will work together to update the NODC proposal.

12 ARGO SYSTEM PRODUCTS

This item was introduced by R. Molinari. He noted that there are at least three classes of products. The first are those that evaluate the data, the second evaluate the network and the third are scientific products. Evaluation of data includes such things as comparing float profiles to profiles from other instrumentation such as XBTs or CTDs. It is important to be sure that changing observation instrumentation does not introduce an offset in measurements that later on may be confused as a climate signal.

Evaluation of the network includes products that demonstrate how quickly data are being distributed both in real-time and delayed mode. It could also include statistics on typical float lifetimes, displays of the distribution of floats or their trajectories.

Scientific products would include comparisons to climatology, estimates of how well the array improves characterization of the ocean, comparisons to models, presentation of available trajectory information, etc.

Discussions noted that Argo needs to work with groups such as SEAREAD to provide products that promote the Argo program. The products generated should be specific to the Argo program, though. It was suggested that the AIC should examine existing web sites of Argo participants looking for sample products that represent the full program rather than just those of a particular nation. Such products could be listed on the AIC web pages.

R. Molinari agreed to organize a working group to define the desirable products. M. Belbeoch agreed to carry out the survey suggested and to provide this to him and his team. The group would present results to the AST meeting next March.

It was also agreed that there would be a discussion of products available at individual web sites at the next DMT meeting.

13 DATA VERSION CONTROL AND DOCUMENTATION

R. Keeley presented the idea of Data State Indicators (DSI) was a product of discussions he had with Neville Smith at an OOPC meeting. The DSI is analogous to the scheme employed by satellite data managers to tell a user the general state of processing of the data. Because of the nature of oceanographic data, a DSI was applied at a finer granularity of data, but the concept was the same.

The meeting agreed that this appeared to be a useful thing to do. Instructions were needed to ensure the data system set these uniformly among participants. This information would appear either in the descriptions of the formats, or in another annex for the Data Management Handbook. Keeley was asked to prepare the necessary information.

There was also a discussion on how to report the state of completion of the data system. Keeley presented the results of discussions that had taken place. A set of milestones was proposed against which each DAC and GDAC would measure their progress. This information would be set up and maintained on the AIC. M. Belbeoch agreed to implement what had been discussed. The system to handle the information at the AIC would be completed and DACs and GDACs requested to report their progress before the upcoming AST meeting.

14 INTEGRATING OTHERS KINDS OF DATA

A discussion was initiated on how to provide access to temperature and salinity data acquired other than by Argo floats but compatible with the formats of Argo.

A presentation was made from Coriolis/France experience whose goals are to provide to French operational ocean modelers (Mercator, French Navy), access to all the temperature and salinity measurements available whatever instrument had acquired them (profilers, XBT, CTD, thermosalinographs, moored or drifting surface buoys).

The advantages of providing such data access are:

- To improve significantly data coverage (both in space and time) mainly in the upper layers.
- To facilitate the sensors inter-comparison, sensor drift estimation, mainly for expandable devices (profilers, XBT) which needs high-resolution measurements(CTD) for calibration.
- To minimize the operations related on data management using homogeneous techniques and relying on the same operation teams.

The main difficulties encountered by the Coriolis team were:

- How to define the vertical reference that is usually meters for XBT, thermistor chains and pressure for Profilers and CTD? This conversion is not so easy when salinity is not available?
- How to acquire metadata information which are not standardized globally (Cruise references (RV, Ship of Opportunity), Call Sign, WMO number, ..)?

• How to deal with data that have different sensor resolution and accuracy?

Integrated data management is feasible from the technical, network, and data policy point of view. A lot of work had been done for various programs (GTSPP, WOCE, etc.) and now for Argo, which may be put into concrete form through homogeneous data access, data format, real time and delay mode quality control procedures.

15 INTER-MEETING WORK

R. Keeley proposed that meeting to take place between sessions should be carried out by working groups formed to prepare the basic proposals which would then be placed before the larger group. This idea had already been employed in preparation for this meeting. This was accepted and some new working groups were formed resulting from this meeting. The list of groups, members, responsibilities and report schedule is listed in annex 4.

16 NEXT MEETING

The Marine Environmental Data Service located in Ottawa, Canada will host the next session of the Data Management Team. Tentative dates are September 18 to 20, 2002.

17 ACTIONS LIST

N	Action	Date	Participants
1	Write the Argo User manual describing the Netcdf Argo formats and field descriptions	March 2002	T. Carval
2	WG: Write the preliminary version of Argo data management Handbook	March 2002	S. Pouliquen, R. Keeley, L Rickards
3	Update GDAC document according to the recommendations provided by the Data Management Committee	End 2001	S. Pouliquen, M Ignaszewski
4	Evaluation if the efficiency of real time Quality Control tests	September 2002	R Keeley, C Schmidt, L. Petit de la Villéon
5	Report on the effectiveness of climatology and comparisons to first guess meteorologiccal fields for real-time QC	September, 2002	C. Schmidt
6	French GDAC to provide a subset of visualization and sub-setting tools on www servers	March, 2002	S. Pouliquen,
7	Elaborate the Bufr format for Argo floats transmission on GTS to be reviewed at next meeting	September 2002	R. Keeley
8	Update the AIC WWW site according to the recommendations provided by the Data Management Committee	March 2002	M. Belbéoch
9	Include milestones interface on AIC WWW server	February 2002	M. Belbéoch
10	Fill in the milestones by DACS and GDACS manager	March 2002	DAC, GDAC Managers
11	Action on CLS France for implementation of the same real time Quality Control procedure as in CLS USA	End 2001	S. Pouliquen, L. Rickards
12	Evaluation of Annie Wong/PMEL delayed mode Quality Control procedures by the Argo Science Team members	March 2002	D. Roemmich

13	Revise the Long Term Archive proposal according to the recommendations provided by the Data Management Committee	End 2001	C. Sun, S. Pouliquen, R. Keeley
14	Create working group for products elaboration from Argo data and make a proposal at the next IAST meeting	Report for March 2002	R. Molinari
15	Elaborate the list of DACs contact points	February 2002	R.Keeley, S. Pouliquen
16	Argo data CDROM: NODC will propose a CD design or a CDROM prototype at the next Argo Data Management meeting.	September 2002	C. Sun, S. Pouliquen, M. Ignaszewski
17	Create the inter session working groups	End 2001	R Keeley
18	Provide DACs with software to check Netcdf files format	End 2001	T. Carval
19	WG: Develop algorithms that should be included in real-time QC tests to exploit characteristics of floats	March, 2002	D. Roemmich,A. Wong,B. R. Molinari,C. R. KeeleyD. C. Sun
20	Next Argo Data Management Meeting at Ottawa	Sep18-20, 2002	R Keeley, S.Pouliquen