Report of the Argo Data Management Meeting, Ottawa, 18-20 Sep, 2002

The meeting began at 0900 on 18 September.

The welcome was given by Dr. Savi Narayanan, Director of the Marine Environmental Data Service (MEDS) on behalf of Dr. W. Watson-Wright, Assistant Deputy Minister of Science in the Department of Fisheries and Oceans, DFO, Canada. Dr. Narayanan noted that there were over a dozen countries and agencies represented at the meeting and this was always a good indication of the international interest and support for a project. She explained that the DFO was the primary governmental agency in Canada for carrying out ocean research. She remarked that DFO's mandate to understand Canada's ocean and aquatic resources, protection of the marine and freshwater environment, management and protection of fisheries resources, and maritime safety, commerce and ocean development. She noted that Data are a valuable resource that can be used and re-used to help in answering questions relating to DFO's mandate and that the department is committed to managing the data and information returned from its scientific research. She then reviewed the organization of Argo in Canada. Two PIs, Allyn Clarke at the Bedford Institute of Oceanography and Howard Freeland and the Institute of Ocean Sciences are responsible for the data. The MEDS is responsible for the data management activities acting as the Data Assembly Centre, DAC, for Canada. They are tracking about 50 floats at present in 3 oceans. Canada is committed to buying, deploying and tracking additional floats. She noted that Argo is an extremely important program as it permits sustained, in-situ observations to define the state of the ocean on its broadest scales. Argo complements conventional ship based, line mode sampling and will only succeed through international cooperation. She expressed her wish for success of the meeting, knowing that there was still much to be done to make Argo an operational program.

Bob Keeley, the local host and co-chair, thanked Dr. Narayanan for her presentation. He informed the meeting of the working hours and other local arrangements. He noted that he and his co-chair, Sylvie Pouliquen, would be sharing chairing duties for the different agenda items. Arrangements were made to provide a tour of MEDS to interested participants. Participants were provided an opportunity to modify the agenda and it is shown in annex 1. Finally, participants introduced themselves. A complete list is given in annex 2.

1. Review of National system development and milestone updates

Before the meeting, national reports were received from a number of countries. Additional written reports were invited for inclusion in the meeting report. Oral presentations of these reports were not made. Instead participants were invited to add remarks to supplement their reports. These reports are included in annex 3.

France noted that they provide DAC services to others, and consequently report included information on floats deployed and operated by Germany, the U.K., Denmark and the European Union.

Because both Chile and Peru were attending for the first time, they were invited to present information about the Argo programme in their countries. A summary of their presentations is included in annex 3 with the other national reports.

2. Review GDAC operations

a. ftp & www implementation

Thierry Carval presented a brief review of the operations of the Argo data system because there were so many new participants to the meeting. He then went on to describe the operations of the

French GDAC. Details of their operations can be found in Annex 4. He noted that presently the ftp site was updated every 24 hours, but the www site was updated more quickly. He described their intentions to include a new "geo" directory for antarctic floats and that this would need to be coordinated with the US GDAC.

He explained the differences between the "geo" and "latest data" directories. In the geo directory, the day subdirectories are by observation day. in the latest data directory, the subdirectories is by day of processing. He explained that this latter use of day of processing was consistent with providing clients with the most recently received data. Since delayed mode data will be replacing real-time versions, the day of processing accurately indicates when the data were received.

At the last meeting, it was stated that it would be desirable to have all of the data from a single float to be built into a single file rather than the four file types presently used (profile, trajectory, technical and metadata). Limitations in the structure of netCDF make this difficult without greatly expanding file sizes (see discussions on ease of use later for more details). If there is a strong request from users for this to be done, GDACs will undertake the study. Users are encouraged to contact the GDACs.

In response to questions, Thierry noted that the data present in the "geo" and "latest data" directories were copies made from the files uploaded by the DACs to the "dac" directory. He also noted that the GDACs have software to ensure what appears in the "geo" and "latest data" directories are exact copies of what was received. He reminded the meeting that DACs are responsible for the individual profile files, and the float trajectory, technical and metadata files only. The files appearing in the other directories are the GDAC responsibility.

At present, trajectory data only reside in the "dac" directories. Uwe Send asked how a client could get access to float data from a geographic area like the Atlantic Ocean in a way similar to profiles. Thierry remarked that at present there was no way to do this. He also noted that this would be very difficult to do since trajectory files contained all of the surface drift information for a float., and a float may well cross from one geographic area to another. However, he suggested that such a request for data may be better met through the use of the www server rather than the ftp server. IFREMER expected this server to be providing profile data by October of this year and trajectory information a few months after.

Mark Ignaszewski explained that the US GDAC is online and operating in fully automated mode. The site is compliant with “US GODAE/IFREMER Data Servers”, Version 2.2 and the data files are compliant with version of the file specification as documented in the “Argo Data Users Manual”, Version 1.0. There are three DACs currently online: Canada, Japan, and the United States. The French DAC is expected online soon.

A Web interface has been added to site that allows the user to retrieve data based on position and time.

New addresses have been established for the US GODAE Server and, therefore, for the US GDAC. The web address is www.usgodae.org/argo/argo.html; the ftp address is ftp://usgodae.usgodae.org/pub/outgoing/argo (or ftp://usgodae1.usgodae.org/pub/outgoing/argo/if coming from behind a firewall). (The old addresses still work for the moment.)

He went on to provide additional details about the US GDAC operations (again details can be found in annex 4). He noted that his ftp server is updated hourly and has the same structure as found at IFREMER. He noted that there is still no synchronization between the two GDACs but that this was expected by the end of 2002. He also noted that he was still carrying out manual notifications of problems found in files uploaded from the DACs to the GDAC, simply to check his notification process. The files on his server conform to version 2 of the Argo netCDF format.

The www interface was just installed and may still have a few problems. He expected to have a Live Access Server V6 up and running perhaps by the end of 2002.

The discussion noted that a number of DACs had recently started to send real-time data to the GDACs. These versions of the data are more complete than the data derived from the GTS and found in the "gts" directory on the ftp servers. The meeting asked that the GDACs remove the GTS versions of the data whenever the real-time data was also received directly from the DACs. This will remove any confusion over versions of the data and ensure that the clients get the best available version.

Before the next AST meeting , a status report of the data from floats that GDACs receive only through the GTS will be made. Considering this status, a decision will be made to keep or not this directory on the GDACs ftp servers.

Some DACs noted that they were having problems building netCDF files and asked if it would be possible to share software, or receive some other kind of assistance from other DACs. No simple solution was found to this, since each DAC had their own computing environment to operate in. The variation in environment meant that sharing software was not likely, although advice was always available. Both GDACs also have software that checks the structure of the netCDF files to ensure that files sent to them conform to the latest version of the Argo format. All available software written either by GDACs or DACs will be provided on the ftp site as example for people to start with netCDF files.

Other participants noted that some countries or agencies may not be able to support DAC functions and asked what should be done. In this case, we already have an example in the French DAC providing services to others. Any country or agency wishing to find out what services they might use should contact either co-chair of the Data Management Team (Sylvie Pouliquen or Bob Keeley).

In designing the GDAC operations, it was thought that there should be a delay of up to one hour between the time files were uploaded to the GDACs and when they appeared on the servers. The meeting decided this was not necessary and that removing this delay would allow for automatic processes to function more efficiently. Appropriate changes will need to be made in the documentation.

Mark noted that in most profile files, the first element in the pressure and data arrays were measurements made at the surface. However, he had found some files that recorded temperature and salinity measurements with the first element of the pressure array being the deepest pressure. The format specifications document makes no mention of ordering of the data. The meeting decided that profiles should be ordered from surface to bottom and requested that all DACs take this into consideration.

Bob Keeley noted that the netCDF format records pressures. However, the TESAC code form used on the GTS requires data be reported by depth. It is important that each DAC submitting data to the GTS make the conversion from pressure to depth using the standard UNESCO algorithm (UNESCO Technical Papers in Marine Science No. 44, Algorithms for computation of fundamental properties of seawater, 1983).

b. Tools to ease data use

Howard Freeland presented some comments on using the data obtained from GDAC ftp servers. His premise was to use the ftp server to acquire profile data on a regular basis for a relatively small area of an ocean. He noted a few problems that he encountered including the following.

- There is a single inventory file at present and as the volume of data grows, this file will become substantial in size.

- The daily files in the "geo" directory contain data from whole oceans, and without and efficient inventory, clients may choose to download daily files only to find out no data exists in their area of interest.
- Even if some data do exist, the multi-profile daily files have large amounts of wasted space consumed by null values.
- To new users, netCDF files are "user hostile" and something needs to be done to help them.

In discussions, a number of points were brought up. It was noted that the ftp site is perhaps better configured for those clients that will be downloading larger volumes of data and have the necessary tools and experience with netCDF. In contrast, the www site when implemented likely will be better suited to provide data from smaller areas or restricted time periods. The multi-profile files do exhibit large numbers of null values, but this is a consequence of the netCDF structure and data being presented in daily files.

The GDACs also noted that they intend to produce an ASCII version of the data and that this should help users inexperienced in netCDF. It was remarked that desirable properties on an ASCII format would be that it easily moves into Matlab and spreadsheet applications.

It was suggested that one strategy to help users wishing smaller amounts of data but on a regular basis, would be for the GDACs to offer a subscription service. A client might provide their area of interest, type of data desired, a desired format and perhaps other criteria and the GDACs would provide regular uploads of data meeting their criteria. There are different ways to implement such a service (e.g. the UCAR LDM) and problems of unsubscribing that would need to be resolved. The GDACs were requested to consider if such a service is possible and how it might be implemented.

The meeting concluded that the first priority for the GDACs is to implement their www servers since this would provide an alternative to the basic ftp servers currently operating.

c. Float Ids

Mark noted that the inclusion of the letter Q in a float identifier has undesirable consequences.
- The Q is only added to data placed on the GTS, but it does not form part of the float identifier used by the WMO.
- DACs keep or remove the Q when sending data to the GDAC and this inconsistency is not good.
- There is confusion when looking for data from a particular float whether the Q designator is to be used or not.
- When GTS dissemination uses BUFR instead of TESAC, there is no need to use the Q designator.

The AIC strongly supported these views and urged the meeting to eliminate the use of the Q in any references to float identifiers. The meeting agreed. All DACs are instructed to remove the Q in float identifiers in netCDF files. MEDS is requested to remove the Q when it forwards GTS data to the GDACs. Appropriate changes in Argo GDAC and DAC documentation must be made.

3. Products

a. Classes of products

This discussion classified products into three categories
- Data related - such as maps of the temporal and spatial coverage of floats, timeliness reports, etc.

- Network related - such as an assessment of the adequacy of the coverage of floats
- Science related - such as maps of temperature or salinity fields, etc.

The purpose of the products in the first two categories is to measure how well Argo is meeting its goals, whereas the last category is for research purposes. It was agreed that science products are appropriately left to the scientific community to generate, but the other two classes should be produced by the data system. It was also noted that the data system should be careful about generating products that appeared to be science related because this would likely be one of the earlier such products available from Argo and may give the impression that participants to Argo have first access to the data. This is not the message to be presented to the international community.

It was remarked that a number of data and network related products currently are being generated by DACs and GDACs. It was agreed that all these should be evaluated to evaluate which are the most effective and which are still not being generated by anyone. This is a task handed to the products working group. Membership of the group is shown in annex 6.

b. Data CDs for groups with poor internet access

Charles Sun of the US NODC presented a proposal of what the CDs for Argo could look like. Some of the attributes he suggested included inclusion of the Ocean Data View software that is included on the WOCE Data Resource DVDs, a number of other data visualization tools, data files, etc.

In the discussion certain attributes were decided. These included the following.
- The medium used should be CD rather than DVD.
- Internet links to Argo sites, such as the AIC, would be useful since not all of the users would have limited access to the Internet.
- The CD should have the same layout and tools as the GDAC servers. This will familiarize users with the GDAC structures and so make it easier for them when they are able to access the GDACs.

It was somewhat difficult to judge exactly what such users would want. Belbeoch had already attended a meeting of such users approximately one year ago, and Keeley was shortly to be attending a meeting to discuss "Potential Application of Ocean Observations for Pacific Island Nations". Their experience should help to guide what would be the more appropriate contents and they agreed to contribute their experience. The group emphasized that this CD production needed to keep in mind that the target audience were those with poor Internet access.

Questions were raised about what might be the cost of such a CD, how often it should be produced and what updated versions might contain. It was agreed that to useful the update frequency needed to be more often than yearly. It was not clear how many such CDs would be required, how they would be distributed or what might be the costs. It was suggested that rather than having one site create the updates, it may be better to create a template of the CD. This template would be distributed to any group wishing to create CDs for local or regional use.

It was agreed that a working group (see annex 6) should prepare a draft version of the CD for evaluation by the group and the AST. The draft should be circulated to members of the data management Team by December of 2002 with the target of having the CDs produced by September 2003. It was agreed that the time period of the data coverage should start 1 Jan 1999. This working group should also address the questions raised and resolve what should be done at least to begin.

4. Data format issues

a. ASCII version for groups who cannot use netCDF

At a SEREAD meeting, it was remarked that the netCDF format was too complex for some Argo data users. It was suggested that a simpler ASCII form might be used. The committee agreed that it was important to define an ASCII format available on GDAC www sites when users ask for data. It was decided that the Format Working Group would look at this issue trying to use, if possible, an existing format or at least adapt one. In the meantime the MEDATLAS format will be available at the Coriolis GDAC.

b. Metadata file format

The point was made to standardize the metadata files available at GDACs. Some DACs provide metadata in netCDF, others in ASCII (mainly ncdump of the netCDF file). It was agreed that all DACs will send metadata in the netCDF format described in the "User Manual". However, GDACs will define an XML view of this netCDF file that will facilitate the user access to these metadata files on the ftp sites.

c. BUFR

Keeley noted that little time had been spent on this issue. He had contacted the WMO to find out their current plans for assigning data in BUFR to different bulletins on the GTS. This is important since it is the bulletin headers that help a user to narrow down what data to decode. He noted that this was still being sorted out.

Keeley noted that writing data into BUFR is at least as complicated as writing into netCDF. There is no international library of BUFR routines, and so anyone needing to do so would have a significant amount of software development to undertake. He suggested that one consideration may be to have the conversion of Argo data into BUFR be centralized, or at least limited to a few sites. Woodward informed the meeting that Service Argos was building a subsystem in their software to put data into BUFR in preparation for the conversion of drifting buoy data distribution to BUFR in about 2 years time.

At other meetings, Keeley had informally polled some Argo users and heard that there was not a large demand at present to put data into BUFR. This was the general view of the meeting as well. The meeting agreed that encoding into BUFR would permit GTS distribution of all of the Argo data, not just profiles, but that this was not high on the list of priorities to be done at this time. Keeley was asked to continue to work at this, but on a lower priority basis.

5. Real-time QC

a. Review of effectiveness

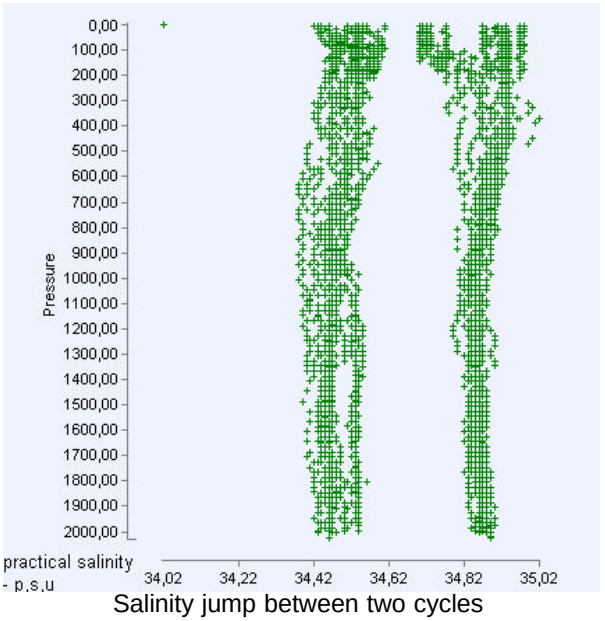
Tran and Keeley presented results of their review of the effectiveness of the real-time QC procedures. They looked at the data that had arrived at MEDS from the GTS and passed these profiles through the automatic QC procedures agreed to at the last meeting. It was expected that some of the profiles would fail since it was known that not all data were passing through the agreed procedures yet. They examined approximately one year of data. They found that there were a number of instances where the Top and Bottom Spike Test was flagging good data as bad. This was exclusively at the top of the profile and was a result of too stringent rules being applied. The group debated the desirability of automatic tests having criteria set to be more effective at catching bad data at the expense of flagging good data or vice versa. In the end it was agreed that this test should be removed immediately from the automated procedures. It was also agreed that a working group should address this problem to come up with better criteria for this test and other tests that would be effective at detecting bad data.

b. Climatology test

A study conducted by Claudia Schmid, using a climatology test based on WOA98 (plus standard deviations provided by Levitus), and the NCEP analysis, concluded that 3% of good temperature profiles and 6% of good salinity profiles had questionable values that could delay their being sent to the GTS within 24 hours of observation. Problems occur, for example, in regions with strong and variable fronts and eddies (e.g. Gulf Stream) and regions with large interannual variability (e.g. eastern tropical Pacific). In addition, the salinity maximum of the subtropical underwater in the float data examined is often larger than in the climatology (while looking perfectly fine). At the present time most regions have adequate climatology for temperature. However, there is such a scarcity of historical salinity that much more work is required to establish a reliable climatology.

The conclusion was that it was better to pass some bad data rather than hold back good data to make the 24-hour target. Moreover, data assimilators have their own QC before assimilation so they prefer to have access to more data even if some are bad. Considering the above mentioned failure rates of good profiles, it does not seem appropriate to let a climatology test delay the transmission of data to the GTS. However, the results of this test can be made available to the community in the netCDF files posted to the GDACs.

However the main difficulty is to detect a sensor drift in real-time because, especially if the drift is evolving slowly, most assimilation centers would not reject these data. Greg Johnson and Annie Wong agreed to study the possibility to detect a drift in real-time data and, to provide the DAC with an adjustment that may be applied to data before sending to the GTS. The result of this study will be presented at the next Data Management meeting in 2003.



Salinity jump between two cycles

In the meantime we have to provide the information to users by inserting comments about suspicious floats in the metadata file. Another way of notifying users of suspect floats may be to use the DBCP Buoy QC mailing list. A proposal will be drafted by Keeley and Pouliquen and sent to DACs to get their views.

c. Additional tests

No additional tests for profiles were proposed.

There was discussion about developing tests to quality control trajectory data. There were a number of unresolved questions, such as how to determine the locations where a float surfaces as compared to where it first reports, where it sinks, etc. There are no agreed procedures for making these determinations and so the co-chairs will raise the question with the AST and seek their advice on how to proceed.

d. CLS implementation

Bill Woodward presented the actions being undertaken by Service Argos in support of Argo. He noted that at the US offices in Largo, Maryland, AOML installed its software to process, automatically quality control and distribute the data to the GTS in real-time. Much of the data from US floats pass through this system. He noted, however, that at CLS in Toulouse, France, they were wanting a more streamlined solution and were contracting to have software developed to do this in the future.

The meeting agreed that both in the USA and in France, CLS had to apply the Real-time QC tests before inserting data onto the GTS. CLS/France should apply the AOML real time QC procedure while they are developing a more integrated solution. DACs who are using CLS services to put data on the GTS were encouraged to send data in netCDF rapidly to both GDACS.

The meeting recommended that Argo should also have a better representation at DBCP so that recommendations coming from them that affect Argo are properly vetted to meet Argo needs.

6. Delayed Mode QC

Several presentations were made on different aspects of the scientific (delayed-mode) quality control process. A discussion of the present status, areas of consensus, and plans for implementation and improvement followed these. Conclusions were:

- The scientific QC process should include at least two (or three) steps – one is a salinity recalibration process using a standard climatological database. Following salinity recalibration, another step is examination of individual profiles by a P.I. or equivalent salinity expert. These steps should be carried out in an equivalent manner everywhere.
- The salinity recalibration process will be based on the system developed by PMEL (Wong et al, 2002), with appropriate regional adaptations.
- The climatological database to be used initially for salinity recalibration should be discussed and agreed by participating regional centres for each ocean. For example, a suggestion was presented by JAMSTEC for constructing a Pacific database using data from WOD98 and Hydrobase. WOD98 contains more recent data but the internal QC of Hydrobase is more rigorous.
- For updating the climatological database with recent data, Argo should seek help from CLIVAR and its hydrographic program office.
- Regional adaptations and improvements to the recalibration system should be discussed and agreed by participating regional centres. Those regional centres will be responsible for making the recalibration software and climatological dataset widely available (distribution might be via the AIC).
- The roles and responsibilities of regional centres need to be described and widely available. This needs to include descriptions of how regional data will be compiled, how data from floats moving into and out of regions are handled, standardizing on climatology.
- An effective process for feeding information on salinity recalibration back into the real-time data stream should be studied. It was noted that real-time recalibration will be sub-optimal, but in principle it's possible to make significant improvement to salinity estimates from many floats (see item 5b).

- In cases where the P.I. or salinity expert rejects or changes recalibration information, a reason should be provided in the data file. A set of guidelines should be developed to assist salinity experts and improve the uniformity of this procedure across the data system. Profile examination by salinity experts requires a substantial commitment of resources, but is necessary in Argo.
- The details of the scientific QC process needs to be sorted out before delayed mode data start moving into the GDACs.
- The technique for documenting the results of the QC process needs to be clarified. A small working group consisting of Schmid, Wong, Keeley, Gronell, and Carval will undertake this).
- Specific instructions and examples are needed to show how to insert information from scientific quality control into Argo netCDF data files (corrected salinity, salinity error, flags, comments). The first step to be implemented will be transfer of salinity recalibration information between regional centres and DACs. It was noted that for floats needing correction to pressure data, that correction should be done prior to the salinity correction.

There are a number of issues raised here and there is no one solution for them all, nor is there enough experience to be able to carry out some of the tasks mentioned. Some of the work will proceed through the WG dealing on QC Process Documentation, and the Formats WG. For those issues for which there is no immediate solution, the co-chairs will try to organize email discussions, meetings or seek advice from the AST.

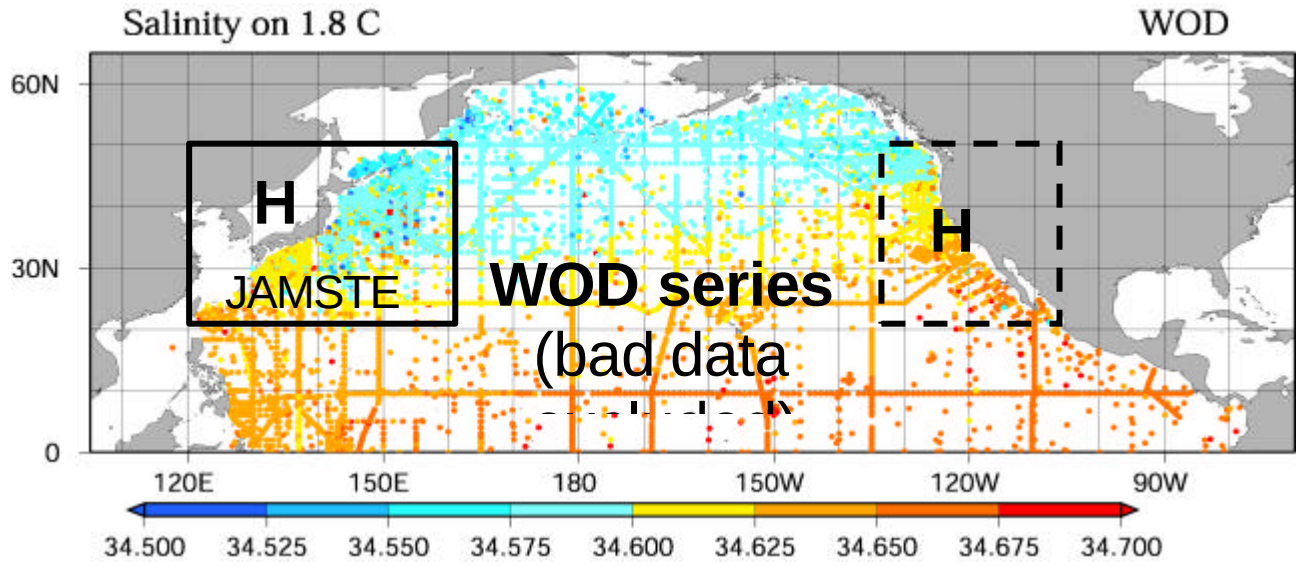


Figure 1: Schematic of JAMSTEC suggestion for merging WOD and Hydrobase climatological databases. For each ocean basin, regional data centres involved in salinity recalibration should reach agreement on the best climatological database.

a. Adoption of Wong procedures

The PMEL salinity correction system for Argo floats was first presented at AST-3 in March 2001 and again at the Data Management Meeting at Brest in November 2001. At AST-4 in March 2002, it was recommended that international Argo partners test the system. This system is presented in detail in Wong et al. (Wong, A. P. S., G. C. Johnson, and W. B. Owens (2003)

Delayed-mode calibration of autonomous CTD profiling float salinity data by theta-S climatology. Journal of Atmospheric and Oceanic Technology,). A preprint of this article and diagnostic plots from the system as applied to U.S. floats are available at http://floats.pmel.noaa.gov/argo. At this meeting, we focused on how the system provides both a salinity corrections and an error estimate for the corrected salinity based on the theta-S climatology statistic, and how the system takes temporal variability of theta-S relations into account. We showed an example where the system works well, and where it does not work well, and demonstrated how tuning some of the system parameters could improve the system performance. We also indicated, based on French, Japanese, and UK experience, how improving the theta-S climatology can also greatly improve system performance. Planned improvements to the current system include making the length of the calibration time-series variable, allowing for discontinuities in the float calibration, tuning mapping spatial scales and the calibration time-series variable, adding screening of near-surface points, and incorporating improved climatological data sets as available. Outstanding issues after these improvements are made include how to incorporate sub optimal salinity correction projections into the real-time QC process and how to update the climatological database with new CTD data. An even more distant issue is how to build a system for intercomparing float salinity data to make an internally consistent database, more of a reanalysis issue than a delayed-mode quality control issue.

b. JAMSTEC experience

S. Minato and T Kobayashi made a presentation. They showed the results of their work testing the Wong procedures using float data from the North and Subtropical Pacific. For climatology, they used a combination of World Ocean Database 98 (WODB98) and Hydrobase. They made a number of interesting comments including the following.

- They requested a step by step standardization of the process for carrying out scientific quality control
- They found that there were a significant number of occurrences where the CRC result was in error.
- They requested a clear description of how to include test results and information into the History structure in the netCDF file.
- They noted that in he subtropical Pacific, the 1500 db level was too shallow for calibrations.

c. Gould experience

Gould reported application of WONG et al DM QC process to floats from Indian Ocean deployed on 32S.

Motivation was
a) to get experience of using the system
b) to apply the system to a “typical” ocean area (i.e. one with a sparse climatology)
c) UK interest in the area is that it is one in which decadal scale changes from anthropogenic impacts might be first detected.

Results are typified by float 09315. It migrated from an area with a recent 1990s climatology to one where the climatology is over 30 years old. In that time the ocean has changed and this means that the application of Wong et al forced the calibration away from the truer uncalibrated values.

Adding more recent (from deployment cruise and from WOCE to the WOD’98 climatology improved matters.

Conclusions reached were

- Using recent CTD data, the detection limit for temporal changes is 0.01 (based on spatial variability) i.e. over 3 months can’t distinguish between calibration offsets and ocean

changes.

- Software is geared to handle slow drifts. Profiles with sudden offsets present problems. (e.g on first few profiles - settling in)
- Recent, high quality climatology on WOCE scales is essential to make the "Wong et al" method work
- This is recognised by Wong et al but tools to ingest recent CTD data are needed
- DM QC centres need speedy access to ship-based CTD data.
- Even sparse data are helpful
- Unless salinity sensors improve greatly we will need to keep gathering CTD data on WOCE scales to enable Argo to look at spatial scales of decadal change.
- Data runs of longer than 3 months are needed to increase confidence in corrections. This has implications for DM QC timescales.

Discussion suggested that a role for regional date centres would be to update climatologies and to ingest recent CTD data. This requires partnerships with research programmes (CLIVAR) and its data system (probably continuing to use WOCE Hydrographic Programme Office)

d. British Oceanographic Data Centre (BODC)

Representatives from the BODC made a number of general comments. Included in them were the following points.

- If regional data centres construct their own climatology there will be chaos.
- The description of the Data State Indicators is confusing. It seems overly complex and will require some clarification in order to be effective.

7. AIC web site

a. List of national contact points

Participants were urged to be sure that the appropriate national contact points for the data system were known by the AIC.

b. Revised web site

The AIC was established in early 2001 to serve the Argo community, under the auspices of IOC & WMO, assisting as appropriate in the implementation of a global network. The web site provides the formal mechanism for informing designated contact points in Member States about float deployments, shows how to track float positions, and gives access to float data, in compliance with the IOC Resolution XX-6. All material included on the website is free.
Administration, development, and maintenance of this web based information system are done by the Technical Coordinator (TC). CLS, which is hosting a dedicated server, provides operational Internet connection, mail server, firewall and backups. Some extra costs (regarding the basic logistic contract with CLS), including hardware and software have been covered through the integration within JCOMMOPS.

The Argo TC presented the AIC website developments. The committee was asked for feedback about possible problems of access, page loading times, etc., and to suggest future upgrades or developments.

Beyond the real-time Monitoring System (interactive & dynamic maps), the website provides:

- General information on Argo
- Tools for national and international coordination
- Tools for assisting network implementation
- Tools for assisting Data Management
- JCOMMOPS integrated tools (Argo, DBCP & SOOP)

- Materials for communication
- Documentation
- Links
- Homepage http://argo.jcommops.org
- Forum http://forum.jcommops.org
- FTP ftp://ftp.jcommops.org
- General Mailing List argo@jcommops.org
- Technical Mailing List argo-tech@jcommops.org

The Argo TC proposed to gather different tools (pieces of code, software, etc.) to convert data formats, and to design a dedicated web page for their distribution. This should assist new Argo participants in their data processing. National Argo websites host some pages with products (maps, graphs, etc.) for their platforms. The Argo TC advised that the URLs to access these pages should be based only on the WMO ID and not on a mixture of Argos, WMO and internal Ids.

The meeting noted that a lot of improvements have taken place on the AIC web site and most of the information is now available. A number of additional comments were made as follows.

- Users asked to have the floats currently operating listed in order of most recent first.
- The web site has still the problem of the maps cut in the middle of Pacific Ocean that has to be solved.
- In the monitoring pages users asked for weekly updates to the map presently done monthly and would appreciate time history of floats operating on a monthly basis.
- In the notification form the TC was requested to add an "are you sure button", before sending the email all around the world.
- For the documents referenced on the AIC site and not maintained by AIC it was asked to point to the master sites rather than duplicate the file on AIC.
- AIC proposed to set up a tool to help on finding deployment opportunities. The committee encouraged AIC to continue on this way.
- AIC issues a letter once a month in which there is place for country reports. Users/PIs/DACs are invited to send their contribution to the TC. The Argo Science Team should study how to send this letter to a wider community than the Argo one.

c. Milestones

The meeting noted that the milestone application on AIC is a good tool to see how the network and data system is becoming operational. DACs are encouraged to update their page regularly.

8. Long Term Archive plan

Charles Sun presented a proposal for the operations of the long-term archive for Argo. He described functions being designed at the US NODC to build what they are calling the Global Argo Data Repository, GADR. He described a number of functions including the following.

- a web site with links to documents, DACs, GDACs.
- a web site providing on-line access to the data with sub setting and data visualization tools, multiple format generation capabilities
- an archive maintained in Oracle.
- download capabilities and schedules to capture data from the GDACs.
- quality checking software for data acquired from the GDACs.
- capabilities to integrate data from other instruments with Argo data.

A number of questions and comments were raised in the discussion. These included

- The proposed web site shows a high degree of duplication with AIC (documents, list of DACs or GDACs)

- The Argo ASCII format will be decided by formats working group
- Are QC procedures needed at all at GADR?
- What distinguishes the data on the GDACs from the data at the GADR?
- Where does someone get oldest to newest data?
- Is there a clear boundary between the functions performed by the GDACs and the functions performed by the GADR?

The meeting came to following conclusions about the operations of the GADR.

- If problems are detected in the data downloaded to the GADR, they will undertake no changes. They need to contact the originator, either through the DAC or the PI directly, to inform them of the problem found and let them resolve the issue.
- The GDACs should hold all of the Argo data and they are the source. The GDACs said that they will provide access to all the archive as long as they are funded and they have no plans to remove data from their ftp and www sites.
- The GADR should safeguard versions of the Argo data and information found at the GDACs.
- The Argo data at the GADR should be up-to-date with the data on the GDACs.
- The GADR should provide users with data sets on demand that integrates Argo data with other types of data collected.
- The suite of tools to permit sub setting and visualization of data on the GADR should be identical or similar to those found at the GDACs.
- CD or other hard copy generation of Argo data should be done in cooperation with the GDACs

The meeting decided that the functions of the GADR needed to be further refined and formed a working group to do this. (see annex 5)

There was some discussion about what can or should be done about other groups outside of the Argo data system setting up servers for Argo data. The hazard is that these servers may be out of date with the versions of data offered and this could cause confusion. It was accepted that such a situation may very well arise but that there was no mechanism for the data system to control such actions. The best that we can do is ensure users of the Argo data system are consistently pointed to the GDACs as the master source of data, and to request servers outside the data system to do the same.

9. Review of documents

The meeting was asked to comment on all of the documents that were already available. These include

- The Data Management Handbook
- Users manual for formats
- Format descriptions - metadata, profile, trajectory
- GDAC operations
- Real-time QC procedures

It was noted that it will be necessary for authors / editors of these manuals to make certain updates as a result of decisions taken at this meeting. Additional comments were as follows.

- The Data Management Handbook needs a clearer description of the data flow in the Argo data system, and more polished diagrams illustrating the data flow
- The Users manual for formats did not have a properly descriptive name. Since this manual included not only the format description but also the user manual it should be called the Argo Data Format and User Manual.
- The Format descriptions manual be removed from distribution since all of its information was contained now in the Users Manual just discussed.

- Text in the Format descriptions manual had some minor errors that needed corrections. Those noting the errors should provide them to Carval.

Additional documentation would be needed for parts of the data system being developed. So, we would need documentation on the recommended scientific quality control procedures, and a document describing operations of the GADR. It was also suggested that a Frequently Asked Questions, FAQ, be including at the GDACs to help guide users on the best way to access data, either on the ftp or www servers.

Finally, it was noted that the current suite of documentation exists only in English but that Argo includes a number of non-English speaking countries. It was agreed that the co-chairs should investigate what might be done to have the manuals translated into other languages.

The format of the metadata file was raised once again. At the moment, these files are in netCDF but some participants thought this was inconvenient. The GDACs will provide an XML version of the netCDF metadata files on the ftp sites.

Charles Sun had asked PMEL to review the Argo formats and he informed the meeting of their comments. They noted that

- A lot of attribute type information is embedded in variables as character data. Most programs that read netCDF files (with the exception of ncBrowse) do not deal well with character variables.
- Because there are no dimension variables for position or time, it is necessary to manually associate a variable (that defines an axis, for example, depth) with temperature. ncBrowse can do this association, but it is necessary to use the "New Map..." button to create a mapping.
- The Argo netCDF files do not specify a "CONVENTIONS" attribute. This attribute can be used by software, for example, EPIC, to determine how to read the file.

Other comments related to the format were as follows.

- We need to be sure the text in the user manual is very clear about what is placed in fields. For example there is some additional clarity needing to describe what is included in the field for the position and time of the profile.
- The QC flags should be written as arrays but under the current format specification these are character strings.
- The software written by the GDACs only carried out checks of the structure of profile files. It was noted that some checks would be necessary to ensure the content also made sense.

10. Other business

A few additional points were discussed:

- Paul Hill of Seimac asked to attend. He would like a forum where his company can show off what capabilities they have. This is not the only industry representative who would like to present technical development to Argo community. Perhaps the AST could organize a technical meeting at the occasion of a conference where many Argo actors will be present. This may also help new countries starting with floats to have an easy access to technical information.
- Schmid pointed out that a profile that fails a QC test because of a salinity offset will be flagged but we want scientific QC to look at these and generally they would not look at data flagged wrong. Carval, Gronell, and Wong will propose a way set up a correct data flow for scientific QC (see item 6).
- Roemmich emphasized that the next challenge for the Argo data system is to set up the delayed mode loop. For next AST meeting, he wants examples of floats whose data passed through scientific semi-automated and expert QC and results in standard Argo format in at

least one Argo basin. PMEL will do it over Pacific, CORIOLIS proposed to do it over Atlantic. Data system members should help in getting the data to the experts and storing the information they produce.

11. Time and Place of Next Meeting

Mark Ignasewski offered to host the next Data Management Team meeting in Monterey sometime in the October to December time frame of 2003. This was accepted.

Keeley proposed that the Data Management meeting to be held after Monterey should overlap a day with the AST meeting. This would allow for direct contacts between the two teams to see how each was working. This was taken under consideration.

The meeting adjourned at 1400 on Sep 20. A complete list of action items is given in annex 5. Working groups and members are given in annex 6.