4th OceanSITES Data Management Team Meeting

Date: 25-26th March 2010
Location: Paris,France
Authors: H. Viola, T. Carval, B. Burnett, S. Woodruff and D. Snowden
Meeting information: http://www.jcomm.info/oceansites-dmt-2010

Table of Contents :

Attendees:

Taco de Bruin, Thierry Carval, Steve Diggs, Nan Galbraith, Matthias Lankhorst, Mike McCann, Maureen Pagnani, Cecile Robin, Derrick Snowden, Hester Viola, Robert Weller, Scott Woodruff, Patrick Gorringe, Kate Roberts, Jing Zhou, Susan Becker

Apologies from:

B Burnett, S Pouliquen, K Ando, P Cornillon, Y Hanafusa, P Freitag, P Rao, E Schulz, U Send, A Kozyr, A Dickson.
For more detail about attendees, see Appendix I.

Welcome, Introductions, Review of Agenda

Bob Weller welcomed everybody to the meeting and reminded the Data Management Team (DMT) that since the last meeting, OceanObs09 had raised the visibility of OceanSITES data and emphasized the willingness of new communities to share data, especially in real time. He then asked participants to briefly introduce themselves.

The morning session was to be chaired by Hester Viola in Bill Burnett's absence. The group began by reviewing the agenda and first covering the Site Approval .

As the agenda was quite tight, time wise, the team agreed to create a "Parking Lot" for items that required more discussion than was possible during the meeting. The group collated the discussion points and summarized them in Section 5.1. They will also become topics for future teleconferences (i.e. WebEX).

1. Review of OceanSITES Data Holdings and 2009 Meeting

1.1. Update on Data Management for 2010

On behalf of Bill Burnett, Hester Viola presented the current set up of the data system and reviewed responsibilities for PIs, DACs and GDACs (Figure 1). The responsibilities of each participant are documented fully in the User Manual and the Data Management Handbook.

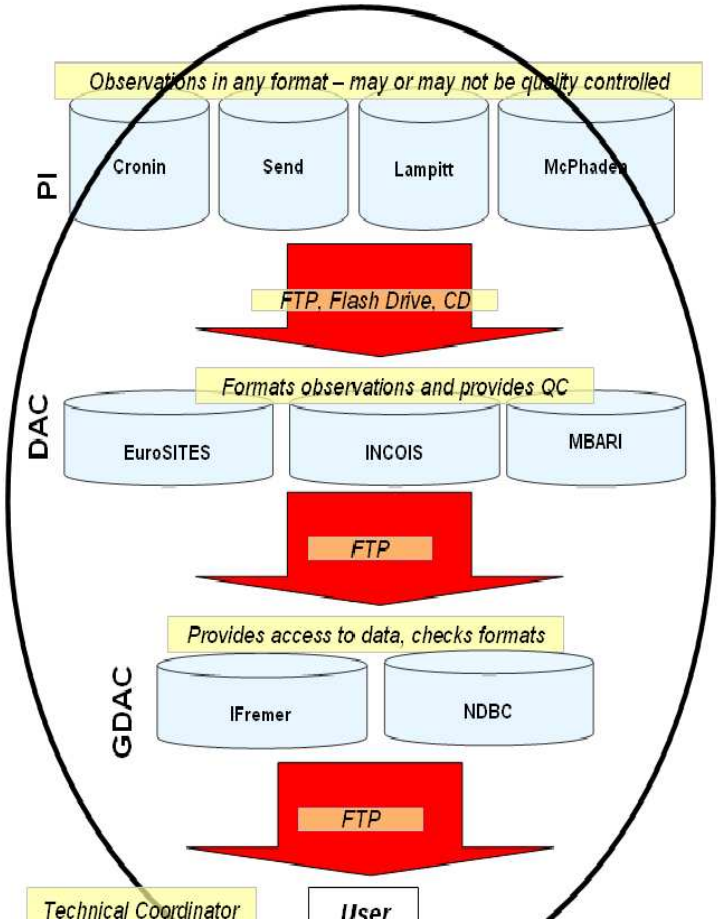

Figure 1: Current Architecture of GDAC, DAC and Data Providers. The PI and DAC organizations listed are only a few examples from a larger set of PIs and DACs. The Technical Coordinator supports the entire operation.

1.1.1 GDAC Resources

OceanSITES data is available online via ftp and OPeNDAP, from the Global Data Assembly Centers:

- IFREMER Coriolis (FTP): ftp://ftp.ifremer.fr/ifremer/oceansites/
- US NDBC (FTP): ftp://data.ndbc.noaa.gov/data/oceansites/
- US NDBC (OPeNDAP): http://dods.ndbc.noaa.gov/thredds/catalog.html
- IFREMER OPeNDAP: (under development)

1.1.3 Virtual meetings

Since the previous meeting in Venice, in September 2009, Bill Burnett had arranged virtual (WebEX)

meetings each month, to discuss specific technical issues. These meetings were held:

- December 7 , 2009
- January 11, 2010
- February 1, 2010
- March 8, 2010

and very well attended. This method proved to be successful during the "off-months" and will continue to be used in the future (starting again in June 2010) to coordinate data management requirements and decision making.

1.1.2 File naming convention

The group confirmed the following file name convention, as finalized during one of the virtual meetings. However the group modified the definition to ensure any data file without a PARTX will not end with an underscore (i.e. the final section of the filename definition was changed from _<PARTX>.nc to <_PARTX>.nc).

Therefore the filename definition will be adopted as follows:

OS_XXX_YYY_T<_PARTX>.nc

where

OS - OceanSITES
XXX - Platform code from the OceanSITES catalogue
YYY - Deployment code (unique code for deployment - date or number)
T - Data Mode (R: Real-Time, P: Provisional, D: Delayed Mode, M: Mixed)
<_PARTX> - User defined field for identification of data (Parameters or nature of data)

- Previous definitions of parameter names (TSVU etc) are now made optional and placed in the PARTX descriptor at the PI/DAC discretion.

It was noted that the GDACs will a have a script to include a full listing of all sensor data available in the files, from fields located in the index file. The GDACs will develop a catalog structure from the index file to provide data users with the ability to query file contents without having to look in the files, therefore the types of data do not have to appear in the file name **(Action from previous meeting: GDAC).**

It is important to note that the OceanSITES ftp directory relies on the definitions and hierarchy defined by OceanSITES, so the definitions of SITE, PLATFORM and DEPLOYMENT were drafted since the last face-to-face meeting. The definitions drafted during the virtual meetings were reviewed by the team and modified slightly. These are to be included in the User Manual.

1.1.4 ftp Directory Structure

Several options for the directory structure were discussed in the virtual meetings. Although the option of including a top-level directory grouping of Geographic Area (ocean basin and sub-basin) had been suggested as usable, there was subsequent discussion via e-mail that arose concerns about that technique – therefore it was thought necessary to re-visit the discussion in the DMT meeting.

In this "face-to-face" meeting, the team agreed that, the level of Geographic Area should be removed from the ftp directory structure and presented as one way for users to view the data system. The concept of Area could therefore be generated later as part of user interfaces other than the ftp directories.

Overall, the team was keen that the ftp be organized in a way that is most convenient for the GDACs

and DACs.

There was then a short discussion about what the upper level directory should be, the decision being that there should be a new directory called DATA at the top level, the directories for each SITE should be the next level and that there should not be a directory for each platform. Therefore for all sites the OceanSITES NetCDF would be directly accessible in a single directory named according to the SITE name. Metadata files can be either placed directly within the SITE directory with the NetCDF files or in a directory called METADATA situated next to the NetCDF files (i.e. under the SITE directory).

Once implemented, these decisions were predicted to allow the GDACs to begin synchronizing their contents within a month.

It was noted that the DACs should have a way to view just their own file structure (using a symbolic link or similar method); this view could either just list all files, or group the files into SITE directories. (**Action: GDACs, DACs**). In order to manage deletion of files from the directory, a DAC manager should replace a file to be replaced with an empty file with the same name and the GDAC will periodically remove any file with size zero and automatically report the deletion to the DAC. (**Action: GDAC**)

1.1.5 Definitions of terms and grouping of sites

The definitions of terms as drafted following the WebEXmeetings were presented for review by the team. The team felt that several of them needed modification. The definitions below are the modified versions:

1.1.5.1 Definitions of sites: platform and deployment

Site – Used as the ftp Parent Directory for OceanSITES GDACs

- A defined geographic location where sustained oceanographic, meteorological or other types of observations are made. A site should be thought of as a point in space, i.e. a nominal position, with a small areal extent around it, such that successive observations from anywhere within this area reasonably represent conditions at the nominal position for the major scientific questions that the observations address.
    - Example: CIS - an OceanSITES site in Central Irminger Sea.

Platform – Used in the File Name for OceanSITES GDAC – this must be unique throughout OceanSITES

- An independently deployable package of instruments and sensors forming part of a site. It may be fixed to the ocean floor, may float or may be self-propelled
    - Example: CIS1 - a mooring in the Central Irminger Sea,.

Deployment

- An instrumented platform performing observations for a period of time. Changes to the instrumentation or to the spatial characteristics of the platform or its instruments constitute the end of the deployment
    - Example: The CTD data for CIS-1 deployment performed in May 2009 (200905) and distributed as file OS_CIS-1_200905_R_CTD.nc.
    - A change to the sampling scheme or a recalibration of existing instruments or sensors will not be considered a new deployment.

The team recalled that during the WebEX meetings it was decided that it was not necessary for OceanSITES to have its own definition of Instrument and Sensor.

1.1.5.2 Groupings of sites: networks and arrays

Sites will be given a Network and/or Array name as metadata where it is necessary to represent relationships between sites either in space or logistically/scientifically.

A single site may be part of one or more network and/or arrays, defined as follows. These are virtual groupings, not impacting on data storage or file contents, and they may be used in the future – along with the geographic area grouping – as a way to present data sets in more advanced user interfaces.

Network

- A grouping of sites based on common shore-based logistics or infrastructure. A single site may belong (or may apply) to one or more networks, but does not have to be allocated to a network. Documenting the network is recommended only if it identifies structures beyond a single project or a single operating institution
    - Example: EuroSITES is a group of 9 European sites in the Atlantic and Mediterranean) where the link is a logistical one with shared organizational resources.

Array

- A grouping of sites based on a common and identified scientific question, or on a common geographic location. It is valid for a single site to belong to one or multiple arrays – or none. Documenting the array is recommended only if it identifies commonalities beyond a single project or a single operating institution.
    - Example: An IRMINGERSEA array would identify the sites CIS, LOCO-IRMINGERSEA, and OOI-IRMINGERSEA as sharing a common scientific interest and/or geographic location.

The Network/Array currently has no place in the metadata or in global attributes of the files. There needs to be some mechanism for storing this information within the metadata or data. Alternatively, it will exist only in the JCOMMOPS database and therefore only be available via the Site Catalog text file and Station List. (**Action: DMT in discussions via WebEX/telephone**)

Review of Meeting Goals

The meetings goals outlined by Bill Burnett were then presented, discussed and revised by the team. The final goals agreed upon were as follows:

1. Define timeframes for beginning synchronization of the GDACS.
2. Ensure that all DACs are aware of their responsibility to collect the information necessary for metadata sheets.
    - And the suggested ways to document the metadata - in SensorML, word documents or text outputs from NetCDF file dumps (assuming that the NetCDF file does indeed include all of the metadata necessary).
3. Discuss how to provide users with ability to peruse and obtain OceanSITES data via a web portal without worrying about the ftp directory structure. Work on the requirements for users and on defining the goals of the data management system.
4. Begin movement toward web services (for data and metadata) :
    - Add an OPeNDAP/Dapper server on both GDACs to be able to provide remotely viewing/access services to OceanSITES data.
    - GDACs (and some DACs) might implement SOS or similar web services to link together.
5. Finalize decisions for user manual updates and appropriate time lines.

1.2. Review of 2009 Action Items

Action items from the last meeting (and before) are included in Appendix II with status filled in for those who provided details on their progress.

1.3. OceanSITES Data Management Project Office Update

Hester Viola presented a short update on the project office activities during the previous 6 months and progress on action items. She outlined the actions she had completed or that were underway relating to data management from previous meetings. Then she discussed some information products and documents which had been created or updated:

- New sites - working with site operators of candidate sites and convincing them to become part of the OceanSITES system and provide necessary documentation etc.
    - Compiling list of new sites for review by the Steering team and encouraging full documentation to be provided by PIs and metadata by data managers.
- Merged the Station Data Spreadsheet with the Site Catalog. Current file at http://www.jcommops.org/FTPRoot/oceansites/documents/oceansites_station_data_2009.xls
    - Began developing JCOMMOPS metadata loading routines to check status of GDACs and automatically update database.
    - Added some extra sites.
    - Noting that names of SITES and PLATFORMS will be needed based on discussions at the meeting.
- Made some changes to the **Data Policy** and **Data Access** documents based on a decision by the Executive Board to require all raw data to be shared ASAP from telemetered sites. Sent out the new documents for review by the Steering and Data Management Teams.

- Wrote a document on the process for adopting new sites in OceanSITES "Site Approval Process" – see meeting website for draft.

- Created a reference to all Quality Control documents about biogeochemical data available from OceanSITES participants and others. See http://www.jcommops.org/dbcp/community/standards#QC

Hester then explained that she had followed up on action items and with DACs, updating the list of DAC status and data flows, and tracked progress of data provision. Within JCOMMOPS, Hester began working on the necessary technical routines to create a monthly map showing GDAC data flows and the database design to store all OceanSITES metadata and deployment information in the JCOMMOPS database. Also added to the web mapping tools at JCOMMOPS for all OceanSITES layers (synchronized with Google Earth) see http://w4.jcommops.org/website/JCOMM/viewer.htm

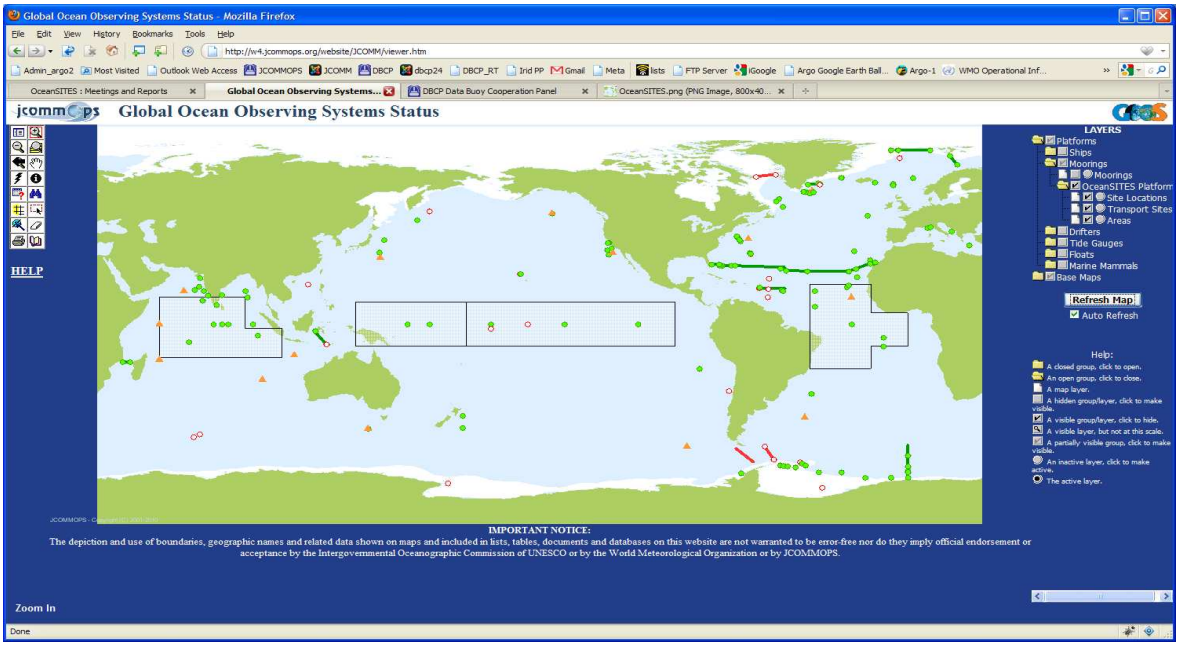

Figure 2. Interactive web map view with OceanSITES layers as an alternative view to Google Earth (showing same data as Google earth but in a flat-earth view and with other JCOMM network layers for overlay)

Hester then presented some future plans relating to data management and explained that further information about the role and activities undertaken in 2009-10 were available in the report prepared for the meeting "Update from the Project Office."

The team then discussed the Global Telecommunication System (GTS) distribution and the following issues related to the upcoming transition to the Binary Universal Form for the Representation of meteorological data (BUFR) format which WMO has decided will be used instead of the older traditional alphanumeric codes (TAC), starting in 2012.

The OceanSITES format for GTS will, in the future, be similar or the same as other moored buoys. There will be no need in future to use other formats such as the Argo Float message format (TESAC) or the XBT format (BATHY) as is done today.

- While presently TACs such as the Argo Float message format (TESAC) and the XBT format (BATHY) are required to transmit some types of OceanSITES data, in the future a BUFR format for moored buoys is anticipated to handle these requirements as well.
- The moored buoy BUFR template is currently under review and Hester Viola and Bill Burnett are actively involved.
- OceanSITES should have representation, other than Hester, Bill, and Scott Woodruff on the JCOMM Task Team on Table Driven Codes (TT-TDC). Maureen Pagnani therefore volunteered to be involved on TT-TDC in BUFR issues for OceanSITES.

Finally, as part of the discussion of the project office activities, it was agreed that citation information as it appears in the User Manual should be put into the data policy document. (**Action: Project Office**)

1.4. Updates from Data Providers – Contacts and Data Flow from DACs

1.4.1 DAC contact points

The following list outlines the eleven (11) DACs "validated" as operational DACs during the meeting with the appropriate points of contact. Points of contact should be reviewed or confirmed.

DAC	Point of Contact
NDBC	B Burnett and J Zhou. Scripps (M Lankhorst and K Baker)
PMEL	P Freitag
JAMSTEC	Y Hanafusa
INCOIS	P Rama Rao
CDIAC	A Kozyr
EuroSITES	NOC (M Pagnani) and IFREMER(T Carval)
MBARI	M McCann
WHOI	N Galbraith
CCHDO	S Diggs
NIOZ	T De Bruin
IMOS	P Gorringe, ~~E Schulz~~ (pending agreement from IMOS directors)

The GDACs can be contacted at the general email address gdac@oceansites.org. It was also noted that there is a need for sharing of tools and programs to assist DACs in processing the OceanSITES format. The User Manual could list some references to tools and programs available to assist new or existing DACs in creating or converting files or metadata. Eventually the OceanSITES website could also include a page which offers of assistance and useful tools for new DACs or data managers.

For each DAC a report is presented, either based on inputs before the meeting or during the meeting directly from the Data Manager responsible for the OceanSITES data.

Reports are also included below from other potential DACs including Norway and Hawaii.

1.4.2 NDBC

- As recommended at the last meeting NDBC is in the process of including all TAO data in OceanSITES and preparing appropriate metadata sheets. Recovery of historical metadata needs to be planned out.
- SIO – Matthias Lankhorst provides NetCDF files directly to NDBC so they are providing some of the DAC processing.
    - MOVE has been processed up until the last deployment and some reprocessing is underway for velocities and bottom pressure data
    - CORC is deployed and needs to have its SITE and PLATFORM names confirmed before sending data to the GDAC. Metadata and site description documents are also not yet completed.
    - CCE processing is underway. Metadata files are not yet provided to NDBC. Site description has been provided to the Project office.
    - The OOI site off Argentina will be the first OOI site to be tackled. Site description documents have not been compiled yet.

- CALCOFI data is being processed by Karen Baker at Scripps. Metadata is currently being compiled. Data processing is a challenge as CalCOFI is providing data to many different 'clients'.

1.4.3 PMEL

- While NDBC is including all TAO data in OceanSITES and preparing appropriate metadata sheets (as discussed in section 1.4.2) metadata sheets and observations from the PIRATA and RAMA arrays will need to be provided by PMEL. Different deployments need to be identified in the file names.
- PMEL is working diligently to process the CO2 data from a number of OceanSITES moorings. Within the next few months, this data will be submitted to CDIAC and available to the OceanSITES GDACs.
- The team noted that there had been no indication of whether PMEL could process biogeochemical data in real-time, so this should be checked. PMEL should be contacted to follow up on real-time access to Carbon and Biogeochemical data (**Action: NDBC, Project Office, Co-chairs**)
- KEO data is missing Metadata files

1.4.4 JAMSTEC

- JAMSTEC has previously committed to providing Triton and JKEO data but very little progress has been made so far in making data available for OceanSITES. Have not been able to keep up with the new OceanSITES data format and ongoing discussion for the next format and metadata standard to date. However, still going to develop a DAC for JAMSTEC OceanSITES stations. Have started re-structuring OceanSITES data management in JAMSTEC and expect some progress beginning from April. Committed to having someone at the next meeting.

1.4.5 INCOIS

- Since the last meeting the data and metadata for the Equatorial Current Meter Mooring Arrays in the Indian Ocean viz., EQCM1-93E, EQCM2-83, EQCM3A-77, and EQCM3-76 have been converted to OceanSITES compliant format and uploaded to OceanSITES. The Sensor ML file for EQCM1-93E has been prepared according to conventions specified in *OceanSITES_sml_manual_proposition* document. Their SensorML file will be validated by the OceanSITES Data Management Team. The Sensor ML files for the other arrays will be prepared once it is approved.
- The 3 sites previously provided have not been updated since 2009

1.4.6 CDIAC

- Willing and able to provide the DAC services for the data from the Moorings and Time-series carbon measurements in OceanSITES format. Not long ago CDIAC opened the web site for these measurements at: http://cdiac.ornl.gov/oceans/Moorings/moorings.html. This site provides information on all moorings and time series that are suppose to measure CO2 system. Unfortunately, the data reporting to CDIAC from these sites is very slow. As of today data is only available from a few sites including TAO moorings and some coastal stations. About a year ago we submitted the carbon data from TAO moorings to OceanSITES in NetCDF format. Since then no feedback on the data file was received.

1.4.7 EUROSITES

M Pagnani reported that 61 datasets (in Version 1.1) had been uploaded so far. Mainly they include only temperature and salinity. The post-deployment data will be put up for CIS (chlorophyll and Oxygen). One challenge is trying to recreate legacy metadata. This needs additional flags to indicate validity.

Naming of files and version modifications have impacted file production. The EuroSITES Project will run until March 2011, but no guaranteed continuation. Thierry Carval reported that the DYFAMED sites had lost funding for regular CTD measurements, but that ongoing funding might be provided from the Mediterranean Ocean Observing System (MOOS). Matthias also mentioned that the German site, K276, did not have sufficient resources to provide data due to retirements. He volunteered to find a way for Tom Mueller (IFM-Geomar) K276 to provide OceanSITES format data onto the GDACs.

1.4.8 MBARI

Three sites are currently pushing subsurface temperature and salinity data and meteorology data to the NDBC GDAC in real time. We will soon add subsurface velocity profile data is to processing stream. Long-term MBARI funding is available for only 2 moorings, so next year we expect to push data from only two sites.

As a result of discussions at the Paris meeting the file naming convention was changed to make the platform_code more specific and to use the real time data code, i.e. OS_M1_20091020_TS.nc has become OS_MBARI-M1_20091020_R_TS.nc. Additional non-standard global attributes are being added from MBARI's Shore Side Data System to give detailed provenance information within the NetCDF file. Efforts are underway to map this information to ISO 19115/19139 and SensorML. These formats will provide provenance information in a standard structured ways, but will require putting the information in separate files. MBARI will continue to include the provenance information as a global attribute with the understanding that it is not meant to be machine "understandable."

1.4.9 WHOI

WHOI-UOP is managing three sites; NTAS, Stratus, and WHOTS. Near-real-time surface meteorological data is being submitted for all 3 sites; also receiving and submitting some subsurface data from NTAS in near-real-time.

Delayed mode met data has been submitted for NTAS through 2006, for Stratus through 2008, and for WHOTS through 2008. The newer surface meteorological data sets are still going through final processing and quality control procedures internally. Submitted delayed mode subsurface temperature, salinity or velocity data.

- Stratus subsurface data, up to about 2005, is nearly ready, and the plan is to translate this in the near future; just requires availability of staff.
- NTAS subsurface data, which consists almost entirely of temperature and salinity, has not been processed enough to be submitted. NTAS funding is, or was originally, primarily for surface data, but there is a plan to share some useful subsurface data, when it becomes available. ADCP QC flags are also not present yet in CF, or OceanSITES.
- WHOTS subsurface data is being handled by University of Hawaii; so it is not clear if they will submit this data, or receive help getting it into OceanSITES NetCDF.
- Wave and CO2 data have to be managed by other DACs.
- SensorML metadata files have been created, but only for the real-time data sets; the NetCDF CDL files should suffice for other data types. These will be placed in a metadata directory for each site.

1.4.10 CCHDO

The CLIVAR & Carbon Hydrographic Data Office (CCHDO) at Scripps is undertaking data processing for all CTD and Bottle data from HOT and BATS.
Some progress has been made, but no files placed on the GDAC yet.

- HOT data is usable, but converting to the OceanSITES format is time consuming. Some data is done, but needs to be put through the format checker etc…
    - The parameters available is vast, so some prioritization will need to be done by the GDACs.

- BATS data is not so easy to process, so a visit is required (dependent on external funding) to work through the issues. The Biogeochemical data is a particular issue as many of the variables do not yet exist in the OceanSITES file format.
    - The team suggested that whenever a new data type was encountered, which was not in the OceanSITES format, the CF conventions and BODCs vocabularies should be consulted and an email sent to the data management team email list to discuss the requirements.
- Line P is also going to be processed in the near future. S Diggs will visit to secure the data in May 2010. Sites will need to be identified to form a nominal location to include in OceanSITES (i.e. the end points of the line only or another regularly visited location/area). Data should flow onto OceanSITES GDACs by mid-late 2010 for CTDs and bottles.

Steve Diggs requested advice from the group about whether the processing should be done in chronological order or if the GDAC should just be given whichever files are available. The group suggested that there are no rules about that and that the path of least resistance should be followed, allowing as many files to be uploaded to the GDAC in the shortest time possible. It was noted that metadata was going to be much easier to populate for current/recent data sets, indicating that would be faster.

Steve also asked if everybody was reprocessing files every time a format changed or file names were redefined and suggested that some kind of web-based format converter could be useful for those converting a large number of files from one version to another, or changing one variable or attribute value.

The values put on "uncertainty" for global attributes seem to vary from DAC to DAC, so a clear instruction is needed from OceanSITES i.e. is it acceptable to leave them blank or write "none" and what functionality it actually provides to data users integrating data.

1.4.11 NIOZ

Taco De Bruin reported that he had completed an inventory of all OceanSITES data as part of the scope of the work. The LOCO sites, Mozambique Channel and Irminger Sea were priorities.

The processing of data from 2000-08 had been completed with Quality Control and 2009 data is currently being processed into the database. NIOZ does not yet have the tools to convert to OceanSITES format.

He committed to providing metadata files to the GDACs as a priority (**Action: T De Bruin**) and then soon after to putting data onto the GDACs.

1.4.12 IMOS

Australia has one deep sea mooring (SOTS) which has been deployed. There have been delays in making the data flow truly operational, but 3 months of data has been received and will be processed. There is also the long-running (30-year) reference network of moorings which may be appropriate as OceanSITES sites.

The Australia Ocean Data Network (AODN) will improve the sharing of data from all platforms under Australia's responsibility. Pending agreement from other IMOS team members Patrick Gorringe and/or Kate Roberts would act as the DAC focal point from now on and will be added onto the DMT list.

1.4.13 Norway (Potential DAC) :

- 3 sites are identified but only one is in EuroSITES. No extra information.

1.4.14 Hawaii (Potential DAC)::

Roger Lukas reported that U Hawaii is very grateful for the great help that Steve Diggs and Matt Shen have provided in converting our HOT physical oceanographic data from WOCE to OceanSITES format and then transferring those data to appropriate online data archives. Data already goes to NODC and to a large community of users through the web site (ww.soest.hawaii.edu/HOT_WOCE). Could certainly take over the data conversion and transmission tasks for our HOT physical oceanographic data from Steve if that is desirable, however unlikely to be able to offer DAC services for other data at this stage.

Matthew Church is now the lead PI for the HOT program.

2. NetCDF Data Files and User Manual

Matthias Lankhorst chaired a session discussing the NetCDF file format and structure issues. Including items such as Coordinate Variables, Data Variables and Attributes, and Quality Control.

1.

2.

2.1. File naming convention and Directory Structure

Matthias Lankhorst presented the file naming convention in order to have confirmation that the team could approve it. The team agreed. Therefore the new convention was adopted and can be included in the user manual version 1.2.

He then quickly discussed the directory structure as agreed to earlier in the meeting and asked for confirmation from the group, so that the decision could be documented.
The directories will appear on the GDACs as follows:
/oceansites_index.txt
/oceansites/data/SiteName/os_PlatformCode_DeploymentCode_DataMode_<_PARTX>.nc

And for example the metadata could appear as follows:

/oceansites/data/SiteName/metadata/os_PlatformCode_DeploymentCode_DataMode_<_PARTX>.xml
The team agreed to this structure, this could therefore be included in the user manual version 1.2 and GDACs can begin the restructure and synchronization as soon as possible.

2.2. Redundant data streams i.e. real time data replacing delayed mode data and archiving

One issue outstanding from the last meeting was how to properly store real-time and delayed-mode versions of the same data on the GDAC servers when the existing data streams are replaced on the GDAC with what was judged the “best copy.” At the 2009 meeting it was agreed that GDACs will continue to hold only the best copy but at that time it was also felt that more discussions with the Steering Team were required to understand perspective on the overall requirements and what should be the GDAC capabilities.

While the need was recognized to find a way for real-time data to be replaced by delayed-mode data (or in the event of corrections of the delayed-mode data) so as to avoid unnecessary redundancy and make readily available the best quality data, many people in the team expressed the possible need to have the original data still available to users. Requirements are increasingly recognized by the scientific community for more precise traceability of the data inputs used for climate studies and other applications (e.g., so that the results can be independently reproduced), thus arguing to the extent practical for preservation of any OceanSITES data versions that have been supplied to users. This archival function would not necessarily be supplied by the GDACs, but somewhere that OceanSITES has to agreed to an official source of the original data. A source which is easy to find and access, whether that be the GDAC website or an archiving center.

It was suggested as one possibility that NOAA/NODC and SeaDataNet each take a copy of the GDAC files monthly to preserve the real-time (or then available) data while the GDACs continue to retain only the best copy in accordance with current policy. The following additional issues were also aired,

but it was agreed that they should be better articulated and made the topic of a future WebEX (Action: DMT members in discussion via WebEX):

- Mechanisms are needed to archive the dataset and govern its orderly replacement. Could just be a regular yearly archival process, or PI-defined deprecating or the DAC moving files to an alternative location.
- Duplicates from different data streams can also introduce archival complications. If the same variables need to be reported from different DACs or different instruments (e.g., with T/S Carbon from CDIAC), then different short/long/standard names are needed.

2.3. Coordinate Variables – Geoid Reference data, Depth and Lat/Long

WGS84 is in the manual and is the agreed reference.

Depth = sea_level, mean_sea_level, mean_lower_low_water, wgs_84 lookup in reference attribute for depth. If missing then sea_level is default. Verify that this is CF compliant. (**Action: Derrick Snowden**) a

Mandatory nature of Depth as a variable. Requirement: We need to accommodate the specification of multiple sensor heights as variables. Measurements reduced to levels for models are derived products and should not be a duplicated field in the data file. Need to have sensor height for each instrument. Clarify how to use the depth parameter as a variable to include different heights with different measures. An ancillary variable called sensor_height or sensor_depth should be required for any data where it would provide useful information and where the coordinate variable 'depth' does not already contain this information (i.e. because of different instruments being at different heights).
(**Action: T Carval, B Burnett**)

Readers should refer to:
http://www.unidata.ucar.edu/software/netcdf/docs/BestPractices.html

Lat/Lon should be able to be a time series or a one dimensional field. OceanSITES needs to be able to capture the time-series as a trajectory and can keep the station data type (with nominal lat/lon reported as a coordinate) and have a GPS lat/lon position as a variable. Lat/Lon variables are allowed to have two dimensions. OceanSITES needs to check with CF/Unidata CDM to determine how to save the lat/lon/pressures.

(**Action: work out how to implement that. Present Different options**)

1. MMcann could stick to station data model type, but if there is a need to save lat/lon measured then should be variables for GPS posi (GPS_latitude etc.)
2. Implement a trajectory ‘type’ to assume that the coordinate values are changeable. Example in the user manual of 5 different CTDs measuring over time.

(**Action: make sure this attribute is consistent with Unidata (updates) and update the manual to say any of the types can be used including “trajectory”.**)

The team confirmed that empty/fill values are not allowed in coordinate variables.

2.4. Raw Data discussion

As discussed in Section 2.2, the preliminary decision at the last DMT meeting in Venice was to only maintain the best-copy data and replace (delete) older versions.

- DMT had another long discussion about the philosophy of keeping data. Who will

accept the responsibility to interact with the Archives? PI or GDAC? It makes a difference in how we determine what to do with raw data.
- This topic impacts all of OceanSITES – not just Data Management. Will need OceanSITES Steering Team involvement to determine best course of action.

2.5. Bottom Pressure vs. in-line Pressure

The group agreed to add an attribute to the PARA variable called sensor_mount with defined values., while noting that it should keep up with the ongoing discussions on the CF mailing lists.

2.6. Sensor attributes

Sensor_name, sensor_SN, Attributes on the parameter level. (or ancillary variable)

char <PARAM>_sensor_name(DEPTH, STRING128)
char <PARAM>_sensor_serial_number(DEPTH, STRING128)

char parameter(PARAMETER, STRING128)
char sensor_name(PARAMETER,DEPTH,STRING128)
char sensor_serial_number(PARAMETER,DEPTH,STRING128)

- General agreement that this should be ancillary variable
- Should be labeled as "highly desirable" but not mandatory.

It would be useful to have a vocabulary to capture the sensor name, and have attributes to report start and end dates for the deployment of new sensors.
(**Action: update to user manual, create list of sensor names**)

2.7. Optional/mandatory attributes

These 3 attributes are mandatory: standard_name, units and _FillValue. These 11 attributes are highly desirables: QC_indicator, QC_procedure, valid_min, valid_max, sensor_name, uncertainty sensor_serial_number accuracy, precision, resolution, DM_indicator. The other attributes are optional.

We should "require" more of the attributes in the user manual.
- Need a tool to report on the quality of the attributes in the file so that we can evaluate and improve the situation.

This may be partially achieved by the semi-automated metadata creation tool to be presented by D Snowden in agenda item four.

Attributes of data variables: these should be listed as required for most.
- Can the format checker see if the attributes that are highly desirable are missing and send feedback to the DAC?

(**Action:** GDACs - Need to weight the importance of different attributes).

2.8. Undocumented attributes

The team agreed that these can be included, but that they'll just be ignored. The question was then raised, as to what to do about undocumented variables. The team again decided that these should also be ignored. Not necessary to have OS-specific attributes. Need a process to suggest that PIs work with the DMT to incorporate ASAP and to ensure that the names don't exist already elsewhere. Append to the variable name that it is PI-defined for local use and that OceanSITES can change it later. This could require a vocabulary.

2.9. Temperature and Salinity Scales

OceanSITES needs to specify scales in "Units" attributes based on the new calculations. The team implemented a variable attribute – called reference scale however they are not really units..
See Appendix IV for more information about Salinity Scales

2.10. Updated Parameter Lists

Can't be taken out of the manual, but needs to be more dynamic than the manual. The latest version is online – but need to clarify responsibilities and update mechanism
(**Action: T Carval and M Lankhorst**)

2.11. Carbon Vocabulary

Matthias Lankhorst reported that he met with Andrew Dickson about carbon partial pressure of CO2 in sea water. A surface one already exists so the name could be renamed or a new one made. Need to specify that the humidity is 100%, ambient temperature and pressure Units: Pa (or microatm)

- In general we are diverging from CF. These differences should be reconciled with the list.

(**ACTION: Assigned to Lankhorst/Kozyr/Dickson**)

2.12. Other Parameter Names

Wave properties – don't have any CF. Need to know the scope of the variables and what exists. Coordinate with US/IOOS to see what they use. Also ask Val Swail what other types of instruments measure waves and parameters considered vital from JCOMM Pilot Project (WET). (**Action: Derrick and Nan - Need a list of short/standard/long names by next meeting**).

- Chlorophyll from fluorescence. Can verify against OOI list of variable names (**Action: M Lankhorst**)
- Tubidity (**Action: M Pagnani**)
- Nutrients
  - Nitrate (NO3) and Nitrite and the sum of the two. M McCann will send CF names relating to Nitrate to Matthias Lankhorst (**Action McCann**)
- Others from Repeat Hydrography. Send CLIVAR parameter names to Matthias (**Action: S Diggs**)
- Sediment traps, work out extent of data to be captured in OceanSITES format. (**Action: M Pagnani, R Lampitt**)

Lower priority parameters to address in future:

- Phytoplankton, Zooplankton, biomass concentration, tagged species, spectral irradiance, nutrients, clarity (secchi depth), CDOM, multi-spectral radiometers, pictures and videos.

2.13. QC Variables and Attributes - Procedures regarding bad data

Bad data should be published with Flag 4, but the decision to make such data available is up to the discretion of the PI or DAC. The group decided to revisit the code tables for data quality indicators and that there should be some documentation in the User Manual.

2.14. Review and Discussion of implications on Synchronisation of GDACs

The participants approved the ftp organization. It will be proposed to the steering team and the GDACs will move forward with automating the synchronization.
Metadata directory location - should be specified in the user manual (**Action: GDACs**). Do we need an index file of metadata files?

3. SensorML Metadata Files

The chair of the morning session, Thierry Carval, provided an introduction on which metadata is most important and how it needs to be managed. He stressed that SensorML is a good format for recording instrument metadata, there is a need to think of the content more than the format.

a. Description of site/deployment/instrument/data QC
b. Description of data files content
c. Description for data access

At the moment the SensorML examples can describe provisional and real-time data. Proposals for delayed mode will be discussed later.

3.

3.1. Metadata formats and status and Aligning metadata content within various files

Cecile Robin presented on the use of SensorML for describing instrument metadata in OceanSITES.

IFREMER implementing an SOS server . ISONET from EuroSITES network focused on sensor registry – catalog to hold instrument metadata to offer services for discovery of sensors.

The key points in understanding SensorML -
4 types of classes – 2 are physical and 2 non-physical (process chain and model_
Sensor ML logic – containers (system and process chains) can contain any other class

Profiles are used to constrain the content and templates used for examples.
OceanSITES profiles. The site is modeled and has a nominal location defined as a BBOX. If the site is a single point, the top left and bottom right can have the same location. The platform and deployment are modeled as a Deployment System with a nominal position and deployment characteristics (dates etc).

Next step: provide sample files of deployment description for example CIS (**Action : Cécile and Nan**). Other participants will scrutinize, comment and write their own metadata files. Check if outputs from one instrument can be documented as going to multiple NetCDF output.
(Action: C Robin)

3.2. Generating Metadata documents

Derrick Snowden (NOAA) made a presentation on behalf of Ted Habermann from NGDC focusing on how to document data sets and ensure their long term preservation and usability.

OceanSITES is a complex collection of information at many levels. We can break the levels of documentation and the relationships down to five levels.

a. Program
b. Project
c. Network
d. Platform
e. Deployment

Derrick explained that NOAA wanted to show how to document, search/discovery and implement a

THREDDs catalog at deployment level (e.g. NetCDF files as the source). For each NetCDF file, the system creates 3 views (XML/NCML plus also an ISO 19115 and HTML report) of the available metadata and displays it as a catalog. ISO XML has capacity to describe anything necessary.

THREDDs offers an easy-to-use view of the metadata – as converted from NetCDF. The NcML xml file looks a lot like the NetCDF file therefore it could be used as a User Interface to the data system:

Derrick explained the process of the Spiral Tracker used to slowly improve the content of metadata in consultation with data providers Defining requirements for metadata and iterative approach in populating metadata. It can also be used to verify and improve on metadata content in NetCDF and where it could be improved or more compliant with requirements and provide automated feedback to DACs and PIs on missing metadata or non-compliance (e.g. where attributes are "highly desirable"). ISO has very good capacity to store information about data quality and lineage.

It checks on 43 elements to see if the file contains standard naming which could be an issue affecting the current score. Using the Spiral checker a quick assessment of some OceanSITES, their data was completed as an example, CIS has a score of 16/43 which is rather low Minor alterations to NetCDF file format and naming would yield a significant payoff

This is all using the content of the DODs server on the NDBC GDAC.

NetCDF manual suggests certain names for terms (Unidata data discovery conventions) and OceanSITES has used slightly different terms in some cases. The DMT could consider changing to adhere to the exact names, if appropriate. This would allow for registration at community data portals, search capability, possibly assist in archive submission. It will also improve the experience with the information rich ISO standard which fits in with SensorML approach to documenting sensors and instruments.

Eventually the goal would be to generate complete ISO files (with links to SensorML) and use those to populate the NetCDF which avoids the constraints of the nc contents in storing metadata.

Repository available on the NGDC site with "Decision support system" FAQs and Geonetwork is used as a user interface (but can also be used to author metadata files). Once the ISO metadata is available on the web accessible folder – all of the metadata and therefore links to data can be harvested directly for national/international catalogs. Automatically harvests from url. Note: There is a big overlap between SensorML, ISO 19115-2 and then CF sits within the intersection.
http://www.demo.ngdc.noaa.gov/geonetwork

3.3. Metadata Provenance

Mike McCann gave a presentation on systems in place in MBARI to track lineage for data sets. Provenance can be represented in NetCDF, however representing the information in ISO or SensorML would be more complete and rigorous, as an approach, or a future plan.

3.4. MMI ontology presentation

Nan Galbraith presented an update on the work of the MMI group and some progress she had made in developing ontology (for instruments) on behalf of OceanSITES. Any decisions to be made by OceanSITES at this stage may be premature, but it is time to start thinking about potential approaches. How much can OceanSITES get involved in creating our own vocabularies –as well as using other external ones.? What terms need to be defined and constrained?

Ontologies provide a way to

a. Define terms and concepts
b. Structured relationships between terms – subsets and hierarchies
c. Machine-readable information
d. Stable, reusable information

Ontologies can bring together several different vocabularies and different standards. URI links could be used to go directly to the definitions. Ontology also allows versioning of the definitions. MMI provides a registry and repository. There are many ontologies available – 4 vocabularies in particular are useful tests Q2O, plus OceanSITES own vocabulary. QA/QC, Provenance, Instruments and Site Terms – which Nan had started doing (OceanSITES DM thesaurus). The DMT should consider how this will impact the OceanSITES program and if we do need more than the BODC etc. that are already in use. Nan suggests developing a thesaurus as a first step.
(**Action: Nan Galbraith as lead, DMT**)

4. Various Topics

4.

4.1. Synchronization of Global DACS and redundant data streams

Jing Zhou presented the items in the index file. It will contain parameters as the final elements of the index file –with standard names delimited by a space. He presented the process for synchronizing which deals with DACs upload areas and then goes to each other to retrieve the missing info.

On the GDAC ftp server, each DAC has an individual ftp connection to upload its data and metadata files. Incoming files are checked by the GDAC, rejected if they fail the format checker. Valid files are then copied from the DAC which will just be a list of files, i.e a flat directory structure) to the appropriate GDAC ftp directory under the correct SITE. Once a day, both GDACs will perform a synchronization.

Jing will finalize the synchronization scripts and provide them to IFREMER The metadata files will need to be synchronized by some similar mechanism. (**Action: GDACs**).

IFREMER would like to be able to check the Global Attribute in case the file is not correct – as well as checking the file names and the location expected. This process should have error trapping and exception reports to go back to the DAC. Format checker must be implemented in the same way for both GDACs (**Action: GDACs**) This format and file check must be integrated into the synchronization process The GDACs need access to a list of contact details and site and platform codes available for each DAC (**Action: GDACs, Project office**)

Jing talked about TAO metadata and discussed the nature of the data set which included averaged data in real time and raw data in delayed mode. The team thought there should be a query to the Steering Team about whether we want to just report original resolution. i.e. can we have the real time (low res) version for an older period even if something more recent has been delayed mode QC'd and replaced a period (**Action: Burnett**)

4.2. Site approval process

Hester Viola described the approval process as drafted after the previous meeting in 2009, for information and so the team understood the responsibilities of the DACs in seeking and providing metadata from PIs and agreeing to the names of sites and platforms, with the project office.
The process is available in the meeting report from the 7th Steering Team Meeting.

4.3. Delayed Mode Data Management

Thierry Carval described the global attributes that are used to indicate the mode of data processing and historical information to track the changes and processing performed on the data.
The user manual will need to be updated to give more examples of history use (**Action: Carval**)
The level of quality control and values in the code table may need to be revisited, or updated to allow several values. Develop a recommendation of best practice for managing the reporting of QC for the whole file (**Action: Lankhorst, Carval, Project Office**)

The reference to the QC_manual used should be a URL reference to a document that is accessible online, rather than a name of a document. Additionally perhaps it would be good policy to include

something like a date version rather than just a URL, the problem being that if the URL is not persistent over the long run perhaps the document could still be tracked down

Also need to add the Resolution and accuracy and qc_indicator .

4.4. Data Management Handbook and Quality Control Manuals

Hester Viola presented on behalf of Bill Burnett about the Quality Control documents available from other JCOMM programs, especially the DBCP. She explained that there are two working groups set up for Quality Control of 1. Physical (Chair B Burnett) and 2. Biogeochemical Parameters (Chair: A Dickson).

For Physical Parameters, the Data Management Handbook, which will include advice on QC, was drafted and sent out by Bill Burnett in August 2009. An update was provided for this meeting and is available on the meeting website. http://www.jcommops.org/dbcp/community/standards#QC

QC of Physical parameters (OceanSITES Data handbook)
B Burnett/NDBC (lead), Yasunori Hanafusa/JAMSTEC, N Galbraith/WHOI, M Pagnani/NODC, T Carval/Coriolis, Pattabhi Rao/INCOIS and P Frietag/PMEL

The project office compiled all of the existing documentation on QC of Biogeochemical Parameters on behalf of the Working group on Biogeochemical QC in late 2009 so that it could summary and synthesize these into a document for OceanSITES.

QC of Biogeochemical parameters (focused on moored autonomous instruments)
A Dickson (lead), R Lampitt/NOC, F Chavez/MBARI, D Wallace/IFM-Geomar, T Trull/CSIRO, M Church/Hawaii, R Johnson/Bermuda

She explained that the QARTOD had made progress in documenting QC mechanisms for meteorological data, and that Bill Burnett planned to ensure all of the information available from QARTOD is put into the Data Management Handbook. (Action: B Burnett, Working Group Members)

5. Conclusion and review against meeting goals

The action list prepared by T Carval and H Viola was reviewed and modified based on feedback from the group. The list is attached in Appendix III.

Goal	Success in achieving
1. Define timeframes for beginning synchronization of the GDACS	**Achieved, aiming for end of May 2010**
2. Ensure that all DACs are aware of their responsibility to collect the information necessary for metadata sheets o And the suggested ways to document the metadata - in SensorML, word documents or text outputs from NetCDF file dumps (assuming that the NetCDF file does indeed include all of the metadata necessary)	**Requirements were communicated and DACs present acknowledged their responsibility.**
3. Discuss how to provide users with ability to peruse and obtain OceanSITES data via web portal without worrying about the ftp directory structure. Work on the requirements for users and what the goals of the data management system are.	**Need guiding principles from the Steering Team on the overall vision and goals of the data system.** **NOAA's metadata generating tools, to be compiled as part of the spiral tracker project should provide one user-friendly view.**
4. Begin movement toward web services (for data and metadata) : o Add an OpenDap /Dapper server on both GDACs on ftp to be able to provide remotely viewing/access services to OceanSITES data. o GDACs (and some DACs) to implement SOS or similar web services to link together	**Not covered in detail.**
5. Finalise decisions for user manual updates and appropriate time lines.	**Achieved**

5.

5.1. Topics from the "Parking Lot"

The following topics were identified as needing further discussion during WebEX/telephone meetings so that the whole data management team has an opportunity to participate. They are presented here in no particular order.

A. Archiving and different data streams

How do we identify redundant data streams? Need to find a way for real time to be replaced by delayed-mode version and avoid redundancy. It is proposed that NODC could take a copy of the GDAC files monthly to preserve the real time while the GDAC retains the best copy.

- Mechanism to archive the dataset and replace. Need a process. Could just be a yearly archive or PI defined deprecating or moved to somewhere else by the DAC.
- Duplicates from different data streams. If same variables need to be reported e.g. with T/S Carbon from CDIAC then need different short/long/standard name.

(**Action: DMT members in discussion via Webex**)

Make a process proposal for the Data Management Plan – new items for archiving and alternate data streams – and DAC has a deprecated directory, Use history to write the location and date. Of the file that superseded another. Put info on the location to redirect to the current copy of the file as it was.
Do we still only want to conserve the best-copy of the data?

B. Handling Lat/Lon time-series as a trajectory or station type
Lat/Lon should be able to be a time series or a 1d.
Decision – need to be able to capture the time-series as a trajectory.
Can keep the station data type (with nominal lat/lon reported as a coordinate) and have a GPS lat/lon position as a variable.
Lat/Lon variables are allowed to have two dimensions. We need to check with CF/Unidata CDM to determine how to save the lat/lon/pressure and come up with options on how to implement this in future (**Action: M McCann**)

C. Mandatory nature of Depth as a variable.
Manual says that for Moorings, depth is mandatory. For other platforms not. Measurements reduced to levels for models are derived products and should not be a duplicated field in the data file. Need to have sensor height for each instrument. One file for each one coordinate axis is the question
Therefore revist one variable per axis requirement.
Clarify how to use the depth parameter as a variable to include different height with different measures

D. Creating a User Interface to provide a "Geographic Area" or "Ocean Basin" via on the data and Capturing Network/Array details

- The geographic area concept should be available to data users via some sort of user interface.
- Network/Array definitions and where this information should be stored. Do we need it in the data or metadata

E. Metadata preservation or archaeology
Preserving (or re-creating) historical metadata files.
Example TAO metadata will be needed for older files)

F. Citing data effectively
How do we best ensure proper citation of data? Do we need to use Digital Object Identifiers (DOI)? Scenario: Users not taking the metadata file with the data file and providing the (undocumented) data to a third party.

G. How to make Repeat Hydrographic data look like "sites"
If a repeat hydrography program does not do its casts at the same locations each cruise, how can we make the data look like an OceanSITES site?

The topics identified should be prioritized and form the discussion in future WebEX/telephone meetings.
(Action: **B Burnett, S Pouliquen**)

6. Next meeting:

Regarding face-to-face meetings, Taco De Bruin offered to host the next meeting, whether it is in late 2010 or in 2011. Australia also expressed interest in hosting a future meeting, perhaps in 2011.

7. APPENDIX I: Attendees

Proxy co-chairs :
Mr Derrick SNOWDEN
Program Manager
Climate Observation Division
1100 Wayne Ave, Suite 1202
Silver Spring MD 20910
United States
Tel: +1 301-427-2464
Fax: +1 301-427-0033
Email: Derrick.Snowden@noaa.gov

Thierry CARVAL
Global Data Manager
IFREMER (French Institute for Sea Research and Exploitation) Centre de Brest, BP70 29280 Plouzané, France
France
Email: Thierry.Carval@ifremer.fr

Steering Team Representative :
Dr. Robert WELLER
Senior Scientist, WHOI
Clark 204a MS 29
Woods Hole MA 02543
United States
Tel: +1 508 289-2508
Fax: +1 508 457-2163
Email: rweller@whoi.edu

Project Office Represaentative :
Ms Hester VIOLA
OceanSITES Project Office, JCOMMOPS
8-10 rue Hermès
Parc Technologique du Canal
31520 Ramonville St Agne
France
Tel: +33 5 61 39 47 82
Fax: +33 5 61 75 10 14
Email: viola@jcommops.org

Mr Taco F. DE BRUIN
NIOZ Royal Netherlands Institute for Sea Research
Landsdiep 4 t'Horntje
Postbus 59
1790 AB Den Burg, Texel
Noord Holland
Netherlands
Tel: +31 (0)222-369479
Fax: +31 (0)222-319674
Email: taco.de.bruin@nioz.nl

Dr. Stephen DIGGS
Data Manager, CLIVAR Hydrography
UCSD/SIO 9500 Gilman Drive MailCode 0214
La Jolla CA 92093-0214
United States
Tel: +1-858-534-1108
Fax: +1-801-650-8623
Email: sdiggs@ucsd.edu

Ms. Nan GALBRAITH
Information Systems Specialist
Woods Hole Oceanographic Institution
Woods Hole MA 02543
United States
Tel: 01-508-289-2444
Email: ngalbraith@whoi.edu

Dr. Matthias LANKHORST
Scripps Institution of Oceanography
9500 Gilman Drive
Mail Code 0230
La Jolla CA 92093-0230
United States
Tel: +1 858 822 5013
Email: mlankhorst@ucsd.edu

Dr. Mike MCCANN
Software Engineer
Monterey Bay Aquarium Research Institute
7700 Sandholdt Road
Moss Landing CA 95039
United States
Tel: (408) 775-1769
Fax: (408) 775-1646
Email: mccann@mbari.org

Ms. Maureen PAGNANI
National Oceanography Centre, Southampton
European Way
Southampton
SO14 3ZH
United Kingdom
Tel: +44 (0)2380 596255
Email: mred@noc.soton.ac.uk

Ms. Cecile ROBIN
Service IDM/ISI
French Institute for the Exploitation of the Sea, IFREMER Centre de Brest
Z.I. Pointe du Diable
B.P. 70
29280 Plouzané
France
Email: Cecile.Robin@ifremer.fr

Mr Scott WOODRUFF
IT Specialist
(R/PSD3), 325 Broadway
Boulder CO 80305
United States
Tel: +1 303 497-6747
Fax: +1 303 497-7013
Email: scott.d.woodruff@noaa.gov

Jing ZHOU
Data Systems Architect/Developer
NOAA National Data Buoy Center
1100 Balch Blvd.
Stennis Space Center MS 39529
United States
Tel: 1-985-710-2009
Email: Jing.Zhou@noaa.gov

Guests :

Susan BECKER
Scripps Institution of Oceanography University of California San Diego 9500 Gilman Dr
La Jolla CA 92093
United States
Email: sbecker@ucsd.edu

Kate ROBERTS
Assistant Director Information Systems
University of Tasmania (eMarine Information Infrastructure)
University of Tasmania
Private Bag 21
Hobart TAS 7001
Australia
Email: Kate.Roberts@utas.edu.au

Dr Patrick GORRINGE
Deputy Director Science Engagement
University of Tasmania
eMII Project
Private Bag 21
Hobart Tasmania 7001
Australia
Tel: +61 3 6226 8558
Fax: +61 3 6226 2997
Email: Patrick.Gorringe@utas.edu.au

8. APPENDIX II: Action List from 2009 and earlier

Items which have been completed are not shown

Meeting at which item was added	Agenda item	Action	By Who	Assisted by who	by when	Status	Comments
OS-DM-2	B-1.3	Commitments to providing data: All operators of sites and representatives of DACs present at the meeting agreed to make their data available in the new format, follow the agreed procedures, and to make the data flow, both in real-time and in post-recovery mode.	All DMT participants	GDACs, Project Office	next meeting	Underway	Also OS-SC-6 Agenda Item 2. Data flowing from MBARI, PMEL, CCHDO/SIO (Bermuda, Hawaii), EUROSITES, WHOI, NDBC, INCOIS
OS-DM-2	B-1	Update the data flow documentation and provide guidelines for DACs on how to submit data to be included in the GDAC data Handbook	S Pouliquen	B Burnett	next meeting	Underway	Included in the Draft OceanSITES Data Management Plan (handbook)
OS-DM-2	B-3.1.1	Propose Quality Control best practice for physical and met parameters	B Burnett	Working group on Physical QC	next meeting	Underway	Also OS-SC-6 Agenda Item 2. Draft went out to Working group in August as part of the "OceanSITES Data Management Plan (handbook)"
OS-DM-2	B-3.1.2	Propose Quality Control best practice for Bio-Geochemical parameters	A Dickson	Working group on BCG QC	next meeting	No Progress	Also OS-SC-6 Agenda Item 2

Meeting at which item was added	Agenda item	Action	By Who	Assisted by who	by when	Status	Comments
OS-DM-3	B-1	Cochairs to request all delayed mode data be made available for PIRATA and RAMA buoys, as they are considered part of OceanSITES?	U Send, R Weller	M McPhaden, P Freitag		No Progress	
OS-DM-3	B-1	SIO (on behalf of NDBC as DAC) to provide data for MOVE to GDACs and metadata files	M Lankhorst			Underway	ML: T&S complete, velocities and bottom pressure are being processed. Metadata files need to be created (NetCDF dump?)
OS-DM-3	B-1	Clarify which data will be made available via CCHDO/SIO and generate metadata descriptions	S Diggs			Underway	Progress documented in meeting report from Paris 2010
OS-DM-3	B-1	Check which EuroSITES data will be available. EuroSITES to populate archive (ANIMATE, MERSEA - ADCP and BioGeoChemical data) and develop metadata for all sites	M Pagnani	R Lampitt, J Karstensten		Underway	61 data sets currently available. Metadata is a challenge.
OS-DM-3	B-1	GDACs to match all directories and file names to Site Catalog (or request updates). E.g INCOIS and WHOI should be separated based on SITE code.	GDACs	Project Office		Underway	Synchronization will commence by May 2010, with SITE as the top-level directory
OS-DM-3	B-1	Track presence of Historical data sets and encourage DACs to provide older data to the GDACs (charts or google earth to show presence over time)	Project Office	GDACs		No Progress	DACs can assist by providing a list (excel) of outstanding datasets and those being processed.

OS-DM-3	B-1	Merge the station data spreadsheet with the site catalog spreadsheet. Until the site catalog can be automaticallly generated from the JCOMMOPS database (with GDAC index files inputted each day)	Project Office	GDACs		Underway	Depends on GDAC Synchronization
OS-DM-3	B-3	DMT members to check that the current SensorML file format adequately captures the required metadata	DMT members	Cecile Robin, Maureen Pagnani		Ongoing	
OS-DM-3	B-3	For elements which are not included in the current schemas (e.g. Data mode (DM/RT), etc) there is a need to register the entries in the MMI vocabularies	C Robin	N Galbraith		Underway	NGalbraith: Some ontologies have been created
OS-DM-3	B-3	Working group members (6) to comment on the Data Management Plan document and then the DMT members should review the document.	DMT members	B Burnett	Dec-09	Underway	BBurnett: Document still in development. Hope to provide second draft to members by the data management meeting.
OS-DM-3	B-3	For variables that have not been previously documented, Working group can feed back gaps and Co-Chairs and Project Office will find expertise to provide input and expand on document. Review by DMT (the document will then be included in the JCOMM Best Practices Catalog)	B Burnett	Cochairs, Project Office	Jan-09		
OS-DM-3	Action 2	Does IMR want to serve as a DAC or rely on the EuroSITES DAC to distribute their data to OceanSITES	IMR?			No Progress	
OS-DM-3	Action 6	Post citation requests for users using the citation provided into the NetCDF files on the OceanSITES website.	Steering Team members			No Progress	
OS-DM-3	Action 7	Cecile Robin to work with Nan Galbraith and Mike McMann to describe more complicated site than ESTOC.	C Robin	N Galbraith, M McCann		Underway	

OS-DM-3	Action 8	IFREMER and NDBC will compare their implementation of SensorML and work for a proposal for OceanSITES in 2010.	GDACs			Underway	
OS-DM-3	Action 9	Approach US NODC "world data center" and perhaps SeaDataNet, to archive all raw data as GDACs will store only Best Copy Data. Request information from the NODCs to see if they would serve as a long-term archive repository for OceanSITES.	B Burnett, S Pouliquen			Underway	BB: NODC and NDBC will meet for a Quality Control Workshop in May. Discussions regarding OceanSITES data holdings will be held at that time.
OS-DM-3	Action 14	In the user manual there is a Parameter Name and Standard Name table. The Team should develop a mechanism to enable the DACS to update this list when required. Action – Data management team.	B Burnett			Ongoing	
OS-DM-3	Action 17	GDACs to produce regular statistics on data access e.g. yearly summary of downloads (charts about origin of user, regularity etc)	GDACs	Pouliquen and Burnett		Underway	BBurnett: Once the GDACs are synchronized, NDBC will work with Ifremer to install tools to track visitors and data downloads
OS-DM-3	Action 18	GDACs implement a RSS mechanism to inform users on data updates	GDACs	Pouliquen and Burnett		Ongoing	BBurnett: Work ongoing at NDBC, should be installed end of March
OS-DM-3	Action 19	GDACs implement OpenDAP access on top of the FTP server	GDACs	Pouliquen and Burnett		Underway	BBurnett: Installed at NDBC - will work to install similar system at Ifremer.
OS-DM-3	B-3	Update the Users Manual to reflect the format changes agreed upon at the meeting	T Carval	GDACs	After every meeting	Ongoing	We need to document the update/approval procedure.

9. APPENDIX III: Action List from 2010

Meeting at which item was added	Agenda item	Action	By Who	Assisted by who	by when	Status	Comments
OS-DM-4	Action 1	GTS - Regarding BUFR Templates, Maureen Pagnani could represent OceanSITES and assist in reviewing/developing requirements for BUFR templates appropriate to OceanSITES. (Action: M Pagnani)	M Pagnani				
OS-DM-4	Action 2	Each DAC manager present during the meeting should provide a 1-2 paragraph Status report for the Meeting report	DACs			Underway	
OS-DM-4	Action 3	Arrange FTP directories in a way to present SITE as the top level (rather than a Geographic Area) and set up synchronization (data and metadata) The DACs should have a way to view just their own sites/files in a structure dedicated to their own area (using a symbolic link or similar).	GDACs, DACs		May-10		
OS-DM-4	Action 4	The GDACs will periodically remove any file with size zero and automatically report the deletion to the DAC.	GDACs				(In order to manage deletion of files from the directory, a DAC manager should replace an old file with an empty file with the same name)

OS-DM-4	Action 5	User Manual Update (v 1.2.1): The Network/Array currently has no place in the metadata or in global attributes of the files. There needs to be some mechanism for storing this information within the metadata or data	DMT members				
OS-DM-4	Action 6	The project office should provide advice to NDBC on naming of sites within the TAO array.	Project office			Underway	
OS-DM-4	Action 7	The example SensorML file generated by India (INCOIS) should be distributed to the team.	Project office			DONE	Placed on the meeting website
OS-DM-4	Action 8	PMEL should be contacted to follow up on real-time access to Carbon and Biogeochemical data	Project office, B. Burnett	R Weller			
OS-DM-4	Action 9	Directory structure: The Data Management team should develop a proposal for the Steering Team or Executive committee on decisions made regarding the directory structure and netcdf file format changes.	DMT members				
OS-DM-4	Action 10	Archiving and different data streams: How do we identify and manage redundant data streams? Whether due to two DACs processing data from one site, or real-time data being replaced by delayed mode, a process should be determined and suggestions made to the Steering Team. Action: Need to discuss in a webex meeting.	DMT members	B Burnett	Next Webex		
OS-DM-4	Action 11	Based on outcomes of future webex meetings, Make a proposal for a process to be documented in the the Data Management Plan on how to deprecate data that should not appear anymore on the GDAC directory.	GDACs	DMT members			

OS-DM-4	Action 12	User Manual Update (v 1.2): WGS84 is in the manual. Depth = sea_level, mean_sea_level, mean_lower_low_water, wgs_84 lookup in reference attribute for depth. If missing then sea_level is default. Verify that this is CF compliant.	D Snowden				
OS-DM-4	Action 13	User Manual Update (v 1.2.1): Clarify how to use the depth parameter as a variable to include different heights with different measures. An ancillary variable called sensor_height or sensor_depth should be required for any data where it would provide useful information and where the coordinate variable 'depth' does not already contain this information (i.e. because of different instruments being at different heights). Action: May need to discuss in a web ex meeting.	T Carval, B Burnett	DMT Members	Next Webex ?		
OS-DM-4	Action 14	User Manual Update (v 1.2): Lat/Lon should be able to be a time series or a 1d. How do we capture the time-series as a trajectory? work out how to implement that. Present Different options. CDM_station_type make sure this attribute is consistent with Unidata (updates) and update the manual to say any of the types can be used including “trajectory”. List pros and cons for each.	M Lankhorst, M McCann				
OS-DM-4	Action 15	Proposal to add extra parameter-level attributes in NetCDF file (!) to contain calibration info. Also consider creating a vocabulary for the possibilities.	T Carval, M Lankhorst	M Pagnani			

OS-DM-4	Action 16	Bottom Pressure vs. in-line Pressure: Sensor attributes, Sensor_name, sensor_SN, Attributes on the parameter level. (or ancillary variable – yes preferably) – should be mandatory (put unknown). Need a vocabulary to capture the sensor name. If they are variables then could have attributes to report start and end dates for the deployment of new sensors. update to user manual, create list of sensor names	T Carval	DMT Members			
OS-DM-4	Action 17	Sensor orientation, clock times and sampling scheme: Optional/mandatory attributes - Attributes of data variables: these should be listed as required for most. Derrick Snowden offered to explain how this can be done with the tools developed at NGDC, but can the format checker see if the attributes that are highly desirable are missing and send feedback to the DAC?	GDACs				
OS-DM-4	Action 18	Undocumented variables: Also should ignore undocumented variables. Need a process to suggest that PIs work with the DMT to incorporate asap and to ensure that the names don't exist already elsewhere. Append to the variable name that it is PI=defined for local use and that OceanSITES can change it later. make paragraph to describe the issue and the suggested approaches.	Steve, Nan				
OS-DM-4	Action 19	Updated Parameter Lists: Can't be taken out of the manual, but needs to be more dynamic than the manual. Latest version is online – but need to clarify	M Lankhorst, T Carval			Underway	

		responsibilities and update mechanism.					
OS-DM-4	Action 20	Carbon Vocabulary : Check with Alex Kozyr and Andrew Dickson on how to deal with CF to propose new names (PCO2, ZCO2DRYAIR, PH_TOT)	M Lankhorst	A Kozyr, A Dickson			
OS-DM-4	Action 21	Wave properties – don't have any CF. Need to know the scope of the variables and what exists. Coordinate with US/IOOS to see what they use. Also ask Val Swail what other types of instruments measure waves and parameters considered vital from JCOMM Pilot Project (WET). Need a list of short/standard/long names by next meeting.	D Snowden	N Galbraith, V Swail, IOOS			
OS-DM-4	Action 22	Investigate ways to capture the following biogeochemical measurements in the CF names: - Chlorophyll from fluorescence (Action: M Lankhorst) - Tubidity (Action: M Pagnani) - Nitrate (NO3) and Nitrite and the sum of the two. (Action: M McCann) - Others from Repeat Hydrography (Action: S Diggs) - sediments. (Action: M Pagnani, R Lampitt) - Later we should consider - Phytoplankton, Zooplankton, biomass concentration, tagged species, spectral irradiance, other nutrients, clarity (secchi depth), CDOM, multi-spectral radiometers, pictures and videos.	M Lankhorst, M Pagnani, M McCann, S Diggs, R Lampitt	DMT Members			

OS-DM-4	Action 23	Philosophy of what we are trying to provide needs to go back to the Steering Team. What are the real goals of putting data on the GDACs and aims with regard to data quality.	DMT members				
OS-DM-4	Action 24	Set up an OpenDAP server	T Carval				
OS-DM-4	Action 25	User Manual Update (v 1.2): Review network and array definition	M Lankhorst			DONE	
OS-DM-4	Action 26	User Manual Update (v 1.2): Add information page on data converters, already developed tools to create data files.	T Carval, J Zhou				
OS-DM-4	Action 27	Plan how best to manage (source?) historical metadata for TAO	J Zhou				
OS-DM-4	Action 28	Organise, plan and transition to data format version 1.2. Research how to create a converter from format 1.1 to 1.2.. (May be distributed and run by the DACs)	GDACs				
OS-DM-4	Action 29	How do we best ensure proper citation of data? Do we need to use Digital Object Identifiers (DOI)? Scenario: Users not taking the metadata file with the data file and providing the (undocumented) data to a third party. Citation details must be in the data policy document Action: Topic for future Webex meeting	B Burnett		Next Webex		
OS-DM-4	Action 30	The german site, K276, does not have sufficient resources to provide data due to retirements. Find a way for Tom Mueller (IFM-Geomar) K276 to provide OceanSITES format data onto the GDACs.	M Lankhorst	SeaDataNet Contact?			
OS-DM-4	Action 31	Create a SensorML repository linking to metadata from sites.	GDACs				
OS-DM-4	Action 32	Propose samples SensorML deployment description (example: CIS)	C Robin, N Galbraith				

OS-DM-4	Action 33	Give feedback to spiral checker on OceanSITES NetCDF format	D Snowden				
OS-DM-4	Action 34	User Manual Update (v 1.2): Add min and max depth in index file, from netdcf global attributes	T Carval				
OS-DM-4	Action 35	Ask the Steering Team its opinion on the products derived from OceanSITES and how to deal with low resolution data once high resolution data is made available (example: daily averages from TAO)	B Burnett	Project Office			
OS-DM-4	Action 36	Add a 'readme' file at the top level of the GDAC, referring to the data policy. Provide information about citations to the Project Office to put it into the data policy.	GDACs				
OS-DM-4	Action 37	Use the format checker for incoming files on GDAC (ideally the same checker available to users)	GDACs				
OS-DM-4	Action 38	Develop a recommendation for how to approach the global data quality information - do we modify the code table or change to free text with a suggested ontology	T Carval, M Lankhorst, Project Office				
OS-DM-4	Action 39	User Manual Update (v 1.2): QC_manual example should give and example of a URL to refer to, instead of a document name.	T Carval				
OS-DM-4	Action 40	User Manual Update (v 1.2): Merge the comment fields in QC-indicator code table. And update them to clarify	T Carval				
OS-DM-4	Action 41	User Manual Update (v 1.2.1): Nan's request for a depth dimensioned use of QC_indicator	T Carval				
OS-DM-4	Action 42	User Manual Update (v 1.2.1): How to manage calibration details? Just free text PARAM_calibration string? Need some examples. Matthias provides an example (the GDAC will	M Lankhorst, N Galbraith				

		accept these files even if not yet documented in user's manual)					
OS-DM-4	Action 43	DACs to provide a spreadsheet to the Project Office listing data that is being processed and not yet available	DACs				
OS-DM-4	Action 44	Who will be in charge of long term archive? Ask the NODCs (US-NODC, SeaDataNet). Relates to Action 10.	B Burnett				Also from last meeting
OS-DM-4	Action 45	User Manual Update (v 1.2): No empty values allowed for coordinate variables.	T Carval				
OS-DM-4	Action 46	Provides links to the ODMTT thesaurus (Ocean Data Management Task Team).	N Galbraith				
OS-DM-4	Action 47	Assess all topics for further discussion (meeting report section 5.1) and priorities for future webex meetings	B Burnett, S Pouliquen				Agenda item 5.1 from Meeting Report

10. APPENDIX IV: Information about Salinity Scales

5.2.7 Thermodynamics and Equation of State of Seawater - TEOS-10

257 This Agenda Item was introduced by Mr Keith Alverson (Head, IOC Ocean Observation and Services Section) referring to **Document IOC/IODE-XX/14.5 (*Thermodynamics and Equation of State of Seawater. TEOS-10*).**

258 Mr Alverson informed the Committee that there are three very good reasons for continuing to store Practical Salinity rather than Absolute Salinity in National Oceanographic Data Centres. First, Practical Salinity is an (almost) directly measured quantity whereas Absolute Salinity (the mass fraction of sea salt in seawater) is generally a derived quantity. That is, we calculate Practical Salinity from measurements of conductivity, temperature and pressure, whereas to date we derive Absolute Salinity from a combination of these measurements plus other measurements and correlations that are not yet well established. Calculated Practical Salinity is preferred over the actually measured in-situ conductivity value because of its conservative nature with respect to changes of temperature or pressure. Second, it is imperative that confusion is not created in national databases where a change in the reporting of salinity may be mishandled at some stage and later be misinterpreted as a real increase in the ocean's salinity. This second point argues strongly for no change in present practice in the storage of Practical Salinity SP in national databases of oceanographic data. Thirdly, the algorithm for determining the "best" estimate of Absolute Salinity is immature and will undoubtedly change in the future so we cannot recommend storing Absolute Salinity in national databases. Storage of a more robust intermediate value, the Reference Salinity, would also introduce the possibility of misuse of salinity data without providing any real advantage over storing Practical Salinity so we also avoid this possibility.

259 **The Committee adopted** Recommendation IODE-XX.4

Recommendation IODE-XX.4

THERMODYNAMICS AND EQUATION OF STATE OF SEAWATER. TEOS-10

The IOC Committee on International Oceanographic Data and Information Exchange, **Noting** the report from SCOR WG127 that describes the significant work done on defining a new procedure to calculate the thermophysical properties of seawater;

Further noting that instrumentation that derives salinity from a measure of conductivity converts first to practical salinity and then using ancillary material and another algorithm, practical salinity can be converted to absolute salinity;

Considering that WG127 remarked that the ancillary material and the conversion algorithm from practical salinity to absolute salinity is still immature and likely to change as more data accumulate;

Further considering that the recommendation from WG127 is that:

i. principal investigators should prepare and submit salinity data in the practical salinity scale to archive centres;

ii. archive centres should continue to archive practical salinity to ensure the long term integrity of the archive;

Notes and agrees with the scientific importance of defining new thermophysical properties of seawater and expresses its interest in the outcome of considerations by the 25th IOC Assembly for its adoption;

Recommends

I. that data centres continue to archive practical salinity, not absolute or reference salinity in accordance with recommendations from WG127;

II. that WG127 widely document these practices and their rationale, similar to what was provided to IODE-XX in appropriate journals (e.g. Journal of Atmospheric and Oceanic Technology) and newsletters (e.g. EOS);

III. that in communications describing the new procedures, WG127 clearly indicate that salinity be reported to data centres in practical salinity;

IV. that the scientific community be asked to provide to national and international data centres all necessary metadata, including manufacturer and model of instrumentation used to measure salinity;

V. that data centres ensure that instrument information be stored along with the practical salinity measurements;

VI. that data centres that are compelled to accept absolute salinities request from the submitters information about the versions of the algorithm and ancillary material used to convert from practical to absolute salinity.