1st OceanSITES Data Management meeting

Ref:OceanSITES-DataManagementMeetingReport-01.B

Date 16th -17th February 2006

Location Hawaii/USA

Authors: S Pouliquen

Affiliation : Ifremer

Table of Content

Document Change Record

| Author	Modification	Issue	Date
S Pouliquen	Initial Version	1A	23rd Feb 2006
S Pouliquen	Final version with comments from Uwe, Scott and John	1B	14th Mar 2006

1. CONTEXT AND OBJECTIVES

S Pouliquen introduced the meeting with a brief summary on the international context, the conclusion of previous OceanSITES Steering meetings and the objectives of this meeting.

At the last JCOMM meeting, Mike Johnson, head of the Observations Program Area, presented OceanSITES as one part of the Initial Global Observing System for Climate. Compared to the other elements of this observing system, even if a lot of sites are performing good measurements, their data availability to the community is far from being coordinated (except for the TAO/TRITON/Pirata arrays).
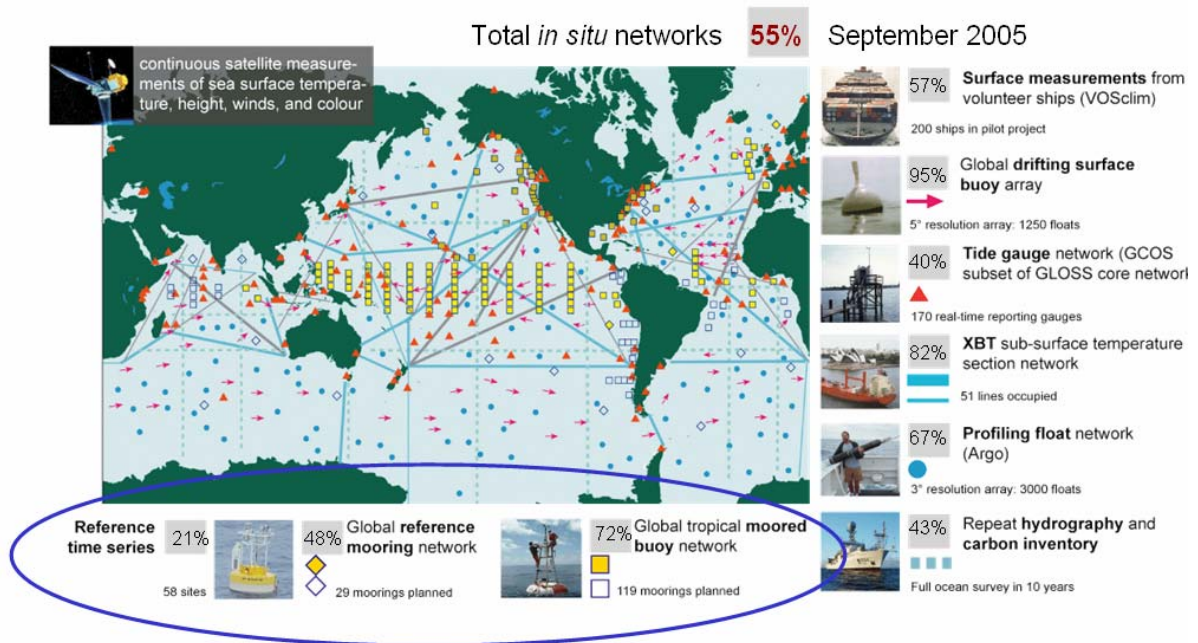
OceanSITES has a lot of potential users both in research community and operational oceanography fields for forecast and reanalysis purposes. This means that there is a need for access to data in near real time and in delayed mode. In particular, GODAE experiment , that will end in 3 years, is a potential important user for OceanSITES for ocean model validation that at present under-uses timeseries data.

It's no more a question of tools: data access has improved a lot in past years within different domains that are addressing the same communities(Argo, Gosud, Carbon, Clivar…). Moreover there are national and International programs that are emerging (Orion/OOI in USA, GMES in Europe, GEO at International level…) and push our community to move forward and build for OceanSites a data management organization that is coherent with these programs without reinventing the wheel
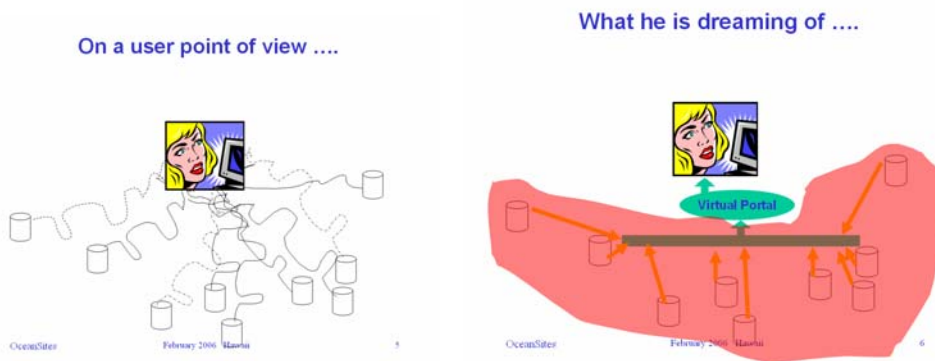
The goal is to move from the left hand side figure the right hand one

When the OceanSITES steering team first addressed the data management issues in 2003 and 2004, the following decisions were taken:

- Start with an easy to implement solution based on Argo model

- Set up a common generic format able to serve the core variables defined for OceanSites project

- Provide the capability to distribute both delayed mode data + some real-time data

- All data have to be allowed for public access

- Rely on a few global data centers to provide an easy centralized access to data by mirroring the national/laboratory servers.

- Start to set up the Gdacs with a few volunteer contributors: Coriolis volunteered to be one of the GDACs

Since last meeting in Puerto-Rico, a data management team has been set up with representatives of different institutions, projects involved in timeseries acquisition and distribution. AT present 43 persons are on the OceanSITES Data Management team mailing list (ots-dm@ifremer.fr). This group has worked by email and achieved the following results:

- A common format has been defined for OceanSITES

- First Data has been made available in this format

- A proposal for a data management network has been proposed and the role of the different actors defined

The purpose of this meeting is to revise what has been done and agree to a strategy and actions in order to take significant steps in 2006.

- for the experience gained by the different partners define the format for next 2 years

- Get commitment from centers to provide their data in this format

- Define the data management network strategy to progress in 2006

- Choose the technology we will rely on

- Initiate activities to progress on real-time quality control procedure for core parameters.

The question of funding associated with OceanSITES data management activities was raised and it was mentioned that, as for ARGO, the project by itself has no funding. We can only rely on national programs (like Coriolis, Orion, NCOF,…) or International ones. At European level there is some funding via MERSEA and CARBOOCEAN projects.

The critical importance of improving access to OceanSITES data within a year was also discussed. The European contributors said that they needed OceanSITES data for MERSEA to serve the ocean modeling community and that the question of sustaining observing system within GMES (Global Monitoring for Environment and Security) was underway. The risk is that only Argo is identified as critical in-situ observing system for GODAE. It seems that discussion between the Observing community and GODAE one is less advanced in other countries. For this community, a site by itself is not so useful, they will be ready to invest in this type of data when they have access to a network of timeseries data. Also in the context of NOAA (office of global programs) and JCOMM, OceanSITES has to deliver the data as a network in order to become a real part of the ocean observing system. Therefore we need a road map to define and provide the most important observations to be distributed by OceanSites.

2. EXAMINATION OF THE OCEANSITES DRAFT FORMAT

The objective of this first topic was to examine the Draft format described in the OceanSITES user manual V0.7 (Site definition, Mooring Metadata definition, Parameter definition, Need for technical information, Integration of Vessel Data) and to achieve to an agreed version of the common format.

It was agreed that an homogeneous data format was necessary to standardize the OceanSITES dataset that by essence is global, processed in different institutes, has to be consistent and continuously managed. The data format has to handle data but also OceanSites specific information about data that should be processed homogeneously. The global data set may be available through different distribution channels that would be set up progressively by different providers. The following distribution channels could be easily set up once data are available in a unique format (FTP servers, CDROM, OpenDAP servers, LAS servers, OpenGIS servers, Web services …)

The following key issues were identified and discussed:

- Unique ID

- Which data to provide to users

- Which QC flags

- Guidelines to create data files

Action T Carval to update the User Manual according to the conclusion below

2.1. Unique Id to avoid future confusion between different platforms.

There were a lot of discussions on Site and Station definition and identification. The first idea to use WMO number to identify a station was abandoned because there are stations that are ship occupied and the assigned WMO numbers can only be linked to mooring stations.

Therefore it was decided that GDACS will have to handle two catalogues:

- A catalogue of Sites. A Site is an area that is regularly sampled by a few stations. The introduction of a new site a submitted to the approval of the OceanSITES steering team or, when it will exist, Project Office. "Irminguer Sea" , "Hawaii", "Bermuda" are for example OceanSITES sites. A unique ID for each Site will be defined and attached to this name commonly used in the research community

- A catalogue of Stations: A site like Bermuda or Irminger Sea may have several stations, occupied by ships, moorings, etc. Each station is attached to a unique site. Its unique ID is "the Site ID+ serial number" assigned by the GDAC. For example the unique ID for the CIS mooring in Irminguer Sea could be 0001-0010 which means 10th station of the first Site. The link with the project name CIS will be kept in the metadata that will describe the station.

- If data from a mooring station is distributed on GTS than a WMO number is necessary. WMO, via JCOMMOPS, has proposed to guaranty to OceanSITES mooring a unique ID A84NN where A is the WMO region, 84 indicates that this is an OceanSITES site, NN number assigned to the mooring in that particular region. To get this type of ID? contact WMO (oca@wmo.int) . Moorings that have

already a number will keep it : i.e. PIRATA LAMBADA mooring will keep 13009 WMO ID;

- Finally it was asked to the OceanSITES steering team to more clearly define a site that is part of the OceanSITES network in order to send a clearer message to the outside of OceanSITES and help data manager to define the format needed. The following definition was provided by steering team on Sunday "Our network has to restrict itself to truly Eulerian observations, to be distinguishable from other networks. Transports can be provided as PRODUCTS"

2.2. Which data to provide to users

It was decided that only the best quality data available at any time will be distributed through the central portal. This portal will make available both original and processed data, at the discretion of the data provider. (Each provides features of value to potential users, and providing both maximizes the utility of the entire data set.). Original data may be improved by obtaining more frequent samples (as when an instrument is recovered from a mooring), by re-calibration, or by quality-control processes that flag -- but do not delete -- suspect data. Processed data may be improved by interpolation, re-calibration, quality control -- including the replacement of suspected bad data by interpolated values -- or other algorithms deemed suitable by the data provider. As improved data is received, it will replace any data set of the same type (original or processed) that it improves upon. Maintenance of previous versions, and of "raw" data sets that are not normalized into COARDS/CF units, is left to the discretion of the data provider.at least until a central OceanSITES archive centre is found.

Coriolis can easily retrieve from GTS other timeseries data outside OceanSITES referenced sites. We agreed to provide them on GDAC as long as they are clearly labeled as "other timeseries data" from outside OceanSITES.

2.3. What quality control?

It was recommended , for the original data (see definition above) to keep in files both good and bad data and to use Quality Control flags to indicate to the user whether or not a measurement is good, doubtful or bad. Where this is impossible to implement, data deemed bad may be removed. Therefore the QC-Flags are kept at the individual measurement level.

The following QC scale was agreed as a starting point for OceanSITES. They will be further improved and defined by the Working group that will work on real-time QC. Comments will be added in the user manual by the Working group that will work on real-time QC (see point 4) to help PIs and Dac set them in a uniform way.

Code	Meaning
0	No QC was performed
1	Good data
2	Probably good data
3	Bad data that are potentially correctable
4	Bad data
5	Value changed
6	Not used
7	Not used
8	Interpolated value
9	Missing value

2.4. Parameter Dictionary

The question of what dictionary to use is a bit tricky because we found no satisfactory on the shelf solution as OceanSITES handles both atmospheric and ocean data. We have decided to use codes based in part on GF3 family names (The notion of family name in the GF3 standard should be more explicitly defined within OceanSITES). This is recommended for several reasons:

1. The group identified that GF3's inclusion of units for some codes poses a problem because OceanSITES units would be indicated in attached metadata without changing the code..

2. There will be many new codes needed and it is not clear GF3 will keep up

3. There is not the intent to stay aligned with GF3, for the above two reasons; although there is the intent to keep communicating with the maintainers of the GF3 standard

As a summary we have decided to use a code based on GF3.. The "standard_name" will be based on COARDS/CF conventions, as described in http://www.cgd.ucar.edu/cms/eaton/cf-metadata/CF-1.0.html. It is this standardization that will make OceanSITES data viewable by a number of existing tools. Name doesn't incorporate units

Many OceanSITES variables will not have existing CF or GF3 family names, and

OceanSITES will work to develop appropriate additions to those standards and forward them to the appropriate standards bodies. As we would like while moving forward on the OceanSITES metadata terms, that GF3 and/or CF organizations feel involved and keep updating their standards we will inform BODC and CF on how to we use these codes within OceanSITES. *Action T Carval & N Galbraith*

2.5. How to create a file?

The format is flexible enough to put everything in the same file but may be full of padding. Therefore we should provide guidance on how to combine data in same files or how to split into files.

From a user point of view, if between two deployments the configuration of the mooring has not changed at all there is no reason to cut this timeserie in pieces. We then recommend to have one file per configuration but to indicate via the Data-Mode indicator whether each single measurement is real-time data retrieved via satellite link or full resolution data retrieved at sea from instrument (RT and DM) . This information will be complemented by a QC flag that will inform on quality control procedure applied.

We agreed not to mix original data and processed data in same file. Processed data mean filtered, interpolated, re-sampled, and mixing them would require different time and depth vectors for each variable. We don't consider calibrated data as processed data but as higher quality version of the same dataset (which would thus just the replace the previous version). On GDAC/ FTP site we should study the most appropriate organization to distribute these different datasets (full resolution or original, processed dataset like interpolated,..)

Also data from similar sensors in a mooring or ship station should be in one file, e.g. all microcats in a mooring, or all ADCP bins. Sensors with very different sampling schemes should be in separate files. Again, these are recommendations, to make the handling as uniform as possible for future users, but it is not binding.

We consider that it is important to record the processing steps of a dataset into an history record. We are not ready yet for an implementation of an history section but a proposal needs to be drafted (*Action J Graybeal*)

Only the metadata attached to the data themselves have been defined at that stage. We need much more metadata to describe a configuration of a station. We think that an XML description would allow a more exhaustive description. SensorML description seemed worth being investigated. *Action Maureen Edwards and John Graybeal to start the action..*

The following File Naming convention was accepted and we agreed to revisit it next year. When the GDAC FTP will be operational:

Data OS-XXX-YYY-ZZZ-WWW-Data.nc

- − Xxx: site code
- − YYY: mooring code, Unique id
- − ZZZ: configuration
- − WWW: data type

Metadata OS-XXX-YYY-ZZZ.xml

Examples :

OS_ATL001_001_200502_CTD.nc : This file contains the CTD data from CIS(001) mooring, from Irminguer (ATL001), for the deployment performed in February 2005. Note that the data start in February until a new configuration is re-deployment.

OS_ATL001_001_200502.xml: This file contains the meta-data of CIS(001) mooring, from Irminguer site (ATL001) for the deployment performed in February 2005. These meta-data are valid from February 2005 until a new configuration is re-deployment.

3. TIMESERIES DATA EXCHANGE

The objective of this second topic is to define a strategy for OceanSITES data exchange and ease access for users both in real time (operational ones) and in delayed mode (mainly research). We have to define the role of the different actors, the architecture of the servers and the link with existing projects.

3.1. Definition of overall strategy for time series distribution: role of PIs, DACs, GDACs
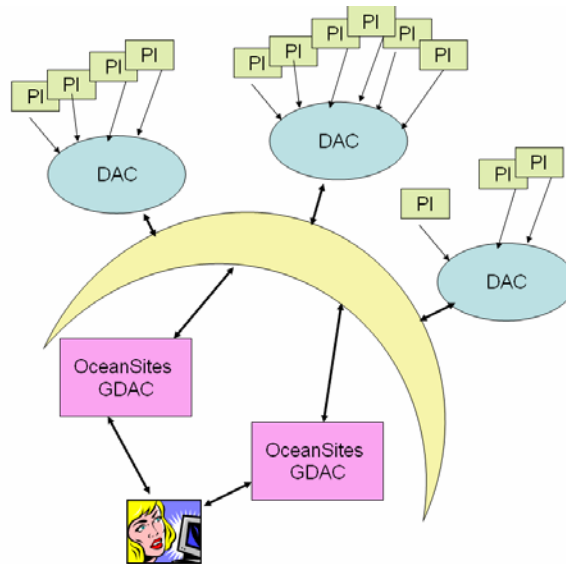
The proposal described in detail in annex 1 was presented. It is based on a distributed data system with three different actors:

- **Pi:** He is responsible of a specific Site. He assures the at sea activities, and provides to the Dac the data and metadata necessary for delivery at OceanSITES project.

- **Dac:** A Dac (Data Access Center) is responsible of setting up a site server, according to the specification approved by OceanSites data management group. He guarantees:
  - Data availability
  - Compliance to the agreed format
  - The Quality of the data according to OceanSites agreed procedures
  - Organization of data processing, formatting, data transfer and update with the Pis he is working with.

- **Gdac** (Global data access center): a Gdac is in charge of
  - Providing a virtual or centralized access to the data that are served by the Dacs .
  - Maintaining the OceanSITES catalogue
  - Synchronizing his catalogues with the second GDAC.

- In Future implementing viewing services for the OceanSITES dataset.

As all the components of this data system will probably be hosted by research institutes, which are not working 24H/7day, for security reason we propose to have two GDACS one in Europe and one in USA. The probability that the two central portals are down at the same time is low.

During the discussion, there has been an agreement on the role of the PI and GDAC. As OceanSITES is starting to set up its data management architecture without any additional funds, there was an agreement that a DAC should not be overburdened but at least provide the data he is in charge of in the OceanSITES format to the GDACs.

Considering the fact that we want an OceanSITES data portal up in 2006 , with a significant amount of data available, we agreed that as a starting point, the easiest solution was for the the GDACs to collect the data from the different DAC to a centralized server according to the GDAC design

Coriolis is willing to set up the European GDAC in 2006. NOAA has an interest in developing the USA GDAC. An actual commitment does not exist yet and the exact location would still need to be defined (e.g. PMEL, ICOADS,NDBC). The decision is expected in October 2006 for an implementation in 2007. Moreover a WDC within NOAA (e.g.NODC) would probably be one appropriate archive facility for OceanSITES. MBARI mentioned that they may be interested in setting up an interim GDAC.

It was also mentioned that Dacs should study the possibility to send some of the real-time data on GTS for operational users.

3.2. Can MMI help OceanSITES

J Graybeal presented the Marine Metada Interoperability Initiative (http://marinemetadata.org) , funded by NSF, that aims to provide tools and accumulate documentation and advice to people dealing with marine metadata. MMI can act as a clearinghouse on topics such as (meta)data management for OceanSITES;

3.3. What are the available technology for distributed access

Jim Potemra accepted to present on behalf of D Holloway/P Cornillon a presentation that S Pouliquen had received from Dan by email the night before.

Tools like OpenDap solve the problem of accessing to distributed data that are in different format and use them together as long as they are semantically compatible. The initiative to agree on a common format will help to use OpenDap technology in a more distributed way if necessary in the future.

Aggregation servers provide access to multi-files and aggregate in one dataset a timeseries of data. The Mooring Aggregation Server is built on Dapper but it's still under development.. We think that in 2006 the solution is not mature enough to be implemented but IPRC agreed to continue with MAS /Dapper teams the test for OceanSITES. If the results are good , we can envisage to set up an OpenDap /Dapper server upon the GDAC ftp to be able to provide remotely viewing/access services to

OceanSITES data (Like IPRC already does for other datasets). Later in 2007 some Dacs may even think to implement OpenDap/Dapper servers linked together to GDACs

3.4. OceanSITES from PMEL prospective

PMEL has set up an efficient data management system to deliver regularly TAO/TRITON/PIRATA data to its users. Maintaining data on more than one site takes more resources and coordination, but can be accomplished with a reasonable amount of effort. If OceanSITES netcdf format is within accepted standards PMEL should be able to comply and provide data in that format to the GDACs. PMEL is also handling the Kurushio Extension Observatory and is willing to provide these data to OceanSITES. Even if TAO operations is transferring to NDBC, as a research endeavor the flux measurements should remain at PMEL, which means that to get all the TAO/Pirata array OceanSITES will have to interact with both PMEL and NDBC. Moreover pCO2 data are independently processed and should be provided through CDIAC. In 2006, PMEL will first provide KEO data in OceanSITES format and then use those tools to implement a second processing chain to provide TAO/PIRATA data to OCEANSITES on FTP every day of daily average.

4. INITIATE REAL-TIME QC

4.1. Proposal for Temperature and Salinity derived from European Project Mersea

Maureen presented the RT QC that has been defined for MERSEA extending what was done for ARGO. Tests are done on platform id, data, global range check, drifting mooring, range checks based on regional climatology, spike, and stuck value.

We agreed to work on a way to fill uniformly the QC flags by providing guidelines to Real Time operators… There should be an OceanSITES real-time QC manual that will be initiated from the Mersea document. The following working group was set up on this subject (***Action PMEL, MBARI , Coriolis, NOC, Jamstec***)

It was decided that first step for delayed mode QC was for each operator to document their procedure and make them available on OceanSITES WWW site (***Action all PIs)***

If decided by steering team, the operator will be encouraged to provide an estimate of the accuracy of the data. In the format we have to add metadata to provide this information as well as reference to a QC manual. We need to provide precision in metadata for each sensor per mooring configuration, and accuracy for each sensor data set. The method should be documented in metadata. (***Action T Carval***)

*The Steering Team will provide guidelines for defining the accuracy information to be provided with metadata **(Action Uwe Send and B Weller)***

4.2. Presentation of the QARTOD effort

QARTOD is a multi agency effort to address the Quality Assurance and Quality Control for IOOS. It has held 3 meetings since 2003. They are focusing on some variables and all information is on their WWW site (http://www.qartod.org).

Their next meeting is 21-23 June at a place to be defined. We think that a representative of Oceansites data team should be involved in this meeting. Julie Thomas has been added to the Ots-dm mailing list.

5. ACTION FOR YEAR 2006

5.1. Dacs

- PMEL : ok to provide KEO & then TAO/Pirata data in OceanSITES format

- MBARI OK to provide some of the MBARI data in OceanSITES format

- CDIAC OK to provide CO_2 data in OceanSITES format

- UCSB ok to provide some of the data in OceanSITES format (BTM)

- LOCO : willing to provide some of the data in OceanSITES format

- CCHDO: could provide ship occupied CTD repeated stations. Will provide tool to convert WOCE format hydrographic data into OceanSITES format

- Jamstec: while their system is set up will provide Triton data in OceanSites format. in 2007

- NOC: need to include ADCP and current meter data plus biogeochemical data … Will continue to provide Animate data via Coriolis

- HAWAII: HOTS: CTD data , WHOTS mooring with/via WHOI (RT and DM)

- WHOI : has already started and will continue

- BATS : CTD data like for HOTS

- IPRC: Can Act like a DAC for some orphan OceanSITES locations. IPRC will pursue work on DAPPER and LAS to give visibility of the network..

Action OceanSITES steering team to identify orphan sites to be processed by *IPRC*

5.2. GDACs

- Coriolis volunteers to host one of the GDAC

- NOAA is interested in hosting one GDAC in 2007,decision to come in October. MBARI may study the possibility to host a GDAC

6. JOINT SESSION WITH DATA TEAM

S Pouliquen presented the summary of the data management meeting and the recommendations and decisions were endorsed by the steering team.

In order to prove a real network and also test the efficiency of the design data management structure, it was decided to focus of the following variables :
- Physical : T & S & current   (for the Godae community)
- Bio-chemical variables: pCO2, oxygen, nitrate, CHL-a
- Air-sea Fluxes (wind, relative humidity, air/sea Temperature, short/long wave radiation)

We agreed to address in priority data from 2000 to present to overlap with ARGO. Longer timeseries are welcome. This list is a minimum one that does not prevent teams to provide more…

The steering team endorsed the decisions proposed in paragraph 5 by the data management team.
It was emphasized that it should be better if some data were at GDACS before summer: it should be possible fo Mersea and WHOI.. Some BTM(UCSB) should be available .CCHDO proposed to provide tools to convert Ship based timeseries into OceanSITEs format that could be used for transferring HOTS and BATS dataset..

Finally then data management part of the WWWW site and the brochure were revised by the data tem. The WWW site Data part should contain:
- Links directly to data access and not to project
- Need section to documentation
- Need link to GDACS when they are up
- Need link to tools that people wants to share

Data management team agreed to meet once a year. Prior to the steering annual meeting. It would help to manage to get funding to help people attend again the data management meeting until travel support becomes available via projects on this topic. A co-chair of the Data management with the following functions (coordination of activities, chasing after people, documentation, organizing meeting) would be appreciated by Sylvie.

7. ACTION LIST

Number	Action	Who	When
1	Update the OCEANSITES User Manual	T Carval	01/03/2006
2	Define the Site Catalogue and assign unique Site codes to each of them	T Carval + Steering Team chairs	15/04/2006
3	Define the Station Catalogue and the process to assign a unique code to a new station	T Carval and S Pouliquen	15/04/2006
4	Refine real-Time QC procedure and Flag assignation	PMEL MBARI CORIOLI NOC JAMSTEC	15/06/2006
5	Complete the User Manual with codification for the parameter identified as first priority section 6	T Carval & N Galbraith	15/04/2006
6	Refine GDAC FTP organization	T Carval & S Pouliquen	15/06/2006
7	Keep OCEANSITES DM team inform of relevant information from MMI project	J Graybeal	Continuous
8	Propose a format for metadata (study SensorML)	M Edwards &J.Graybeal	30/06/2006
9	Test Dapper and MMA technology and advise OceanSITES DM team on possible utilization	IPRC	Next meeting
10	Dacs to provide their data in OceanSITES format according to commitment provided in section 5	ALL	30/06/2006
11	Implement First GDAC	Coriolis	30/06/2006

8. LIST OF PARTICIPANTS

Name	Institution	email
ANDO Kentaro	JAMSTEC	andouk@jamstec.go.edu
BODEN Tom	DOE/ORNL/CDIAC	bodenta@oml.gov
CARVAL Thierry	Ifremer	Thierry.Carval@ifremer.fr
CHAVEZ Francisco	MBARI	chfr@mbari.org
CORNILLON Peter	URI/OpenDAP	pcornillon@uri.gso.edu
DECARLO Sharon	UH/APDRC	decarlo@hawaii.edu
DIGGS Steve	UCSD/SIO	sdiggs@ucsd.edu
EDWARDS Maureen	NOCS	mred@noc.soton.ac.uk
FREITAG Paul	NOAA/PMEL	Paul.Freitag@noaa.gov
GRAYBEAL John	MBARI/MMI	graybeal@mbari.org
HACKER Peter	UH/APDRC	phacker@hawaii.edu
HANAFUSA Yasunori	JAMSTEC	hanafusay@jamstec.go.jp
JIANG Songnian	UCSB	songnian@opl.ucsb.edu
LETHABY Paul	UH	lethaby@hawaii.edu
LUKAS Roger	UH	rlukas@hawaii.edu
MARRA John	NOAA/DSC-IMEA center	John.Marra@noaa.gov
POTEMRA Jim	UH/IPRC	jimp@hawai.edu
POULIQUEN Sylvie	Ifremer	Sylvie.Pouliquen@ifremer.fr
SANTIAGO-MANDUJANO Fernando	UH	santiago@soest-hawaii.edu
SEND Uwe	UCSD/SIO	usend@ucsd.edu
SWIFT Jim	UCSD/SIO	jswift@ucsd.edu
VAN HAKEN Hendrik	Royal NIOZ (LOCO)	aken@nioz.nl
WOODRUFF Scott	NOAA/ESRL	Scott.D.Woodruff@noaa.gov

9. AGENDA

Thursday 16th Feb : **Morning :**

9:00 - 9:15 Why do we need Data management activities within OceanSITES and what are the objectives of this meeting: S Pouliquen

I- Examination of the OceanSITES draft format
The objective of this half day is to examine the Draft format described in the OceanSITES user manual (Site definition, Mooring Metadata definition, Parameter definition, Need for technical information, Integration of Vessel Data) and to achieve to an agreed version of the common format at the end of this half day.

9:15 - 9:45 Proposed Format description T Carval
9:45 - 10:45 Any Feedback from OceanSITES partners while fitting data into this format?
- Coriolis Feedback T Carval
- Animate Feedback M Edwards ?
- WHOI : T Carval on behalf of N Galbraith
- ????
  10:45 - 11-15: Coffee Break
11:50 – 13:00: Discussion and agreement on OceanSITES Format: Lead discussion T Carval
  13:00 - 14-00 Lunch Break

Thursday 16th Feb : **Afternoon :**
II- Timeseries data exchange
The objective of this half day is to define a strategy for OceanSITES data exchange and ease access for users both in real time(operational ones) and in delayed mode(mainly research)
We have to define the role of the different actors, the architecture of the servers and the link with existing projects.
14:00- 14:30 Definition of overall strategy for time series distribution: role of PIs, DACs, GDACs : S Pouliquen and S Woodruff
14:30 – 15:00 Can DMAC Help us : ??? on behalf of D Holloway
15:00 – 15:30 Can MMI help us: J Graybeal
15:00 – 16:00 Other Contributions ??????
  16:00 : 16:30 Coffee Break
16:30 – 18:00 Definition of an architecture (Gdacs , Dacs, Identification of the contributing DACs for 2006, Link with GTS data , elaboration of an initial time plan) Lead S Pouliquen

Friday 17th February Morning

III- Initiate Real-Time QC
9:00 – 9:30- Proposal for Temperature and Salinity derived from European Project Mersea M Edwards
9:30 – 10:00- Presentation of the QARTOD effort   S Diggs
10:00 – 10:30- Selection of the first candidate parameters for homogeneous real-time QC and definition of a time plan for implementation
  10:30 – 11:00 Coffee Break

IV- Wrap-up and Actions. S Pouliquen

__Friday 17th February Afternoon__

Joint session with Data Team

13:15-14:00   Data Team report   *Sylvie Pouliquen*

14:00   Discussion of items from data meeting that need input from Steering Team

16:00   Adoption of recommendations and decisions from Data Team

16:30   Implementation path for Data Plan, increased flow to the data archive.

17:00   The purpose, evolution, mandate of the OceanSITES Data Team
               Issues: Staff, staff support,
                       Active involvement, next meeting,
                       JCOMM involvement,
                       Goals, getting the message out,
                       Visibility/brochure role,
                       Tech report updates
18:00   adjourn

__Friday 17th February Afternoon__

10. ANNEX 1 : DATA MANAGEMENT NETWORK PROPOSAL

What could be the OceanSITES data distribution?
Draft written by S Pouliquen, U Send, S Woodruff
1 October 2005

Context:

1- A lot of OceanSITES data are available on internet, some are not, but in any cases their format are very heterogeneous.
2- These data would be more widely used if they were easily accessible especially by operational oceanography applications.
3- Only a few moorings/network of moorings transmit data on GTS and often limited to some parameters.
4- Some mooring sites are administrated by institutes who have computer support to install and maintain an operational delivery system of data, but some others rely on scientists who don't have much support on such activities and should rely on other institutions to make their datasets available.
5- OceanSites is willing to contribute to JCOMM (Joint commission for oceanography and marine meteorology WMO/IOC).

Proposal:

In past year the OceanSITES data management group has defined a draft common format base on Netcdf, and we agreed that most of the data should be distributed in this format. For some operational networks (like Tao/Triton/Pirata) for which data are already widely used by operational users, it may be not easy to change distribution format and probably some mapping techniques should be used to connect these sites to the OceanSITES network.

We propose to build the OceanSITES data distribution in two steps:

1- Provide an integrated access from Global Data Access Centers (GDACs) to datasets in OceanSITES format by using Ftp servers (+OpenDap technology if necessary)
2- Study mapping techniques (probably with MMI project led by Monterey) to connect the sites that won't be able to convert their data in OceanSITES format

In order to safeguard OceanSITES data for future generation, we should implement periodic archive of OceanSITES data in at least one place, and at least for delayed mode data.
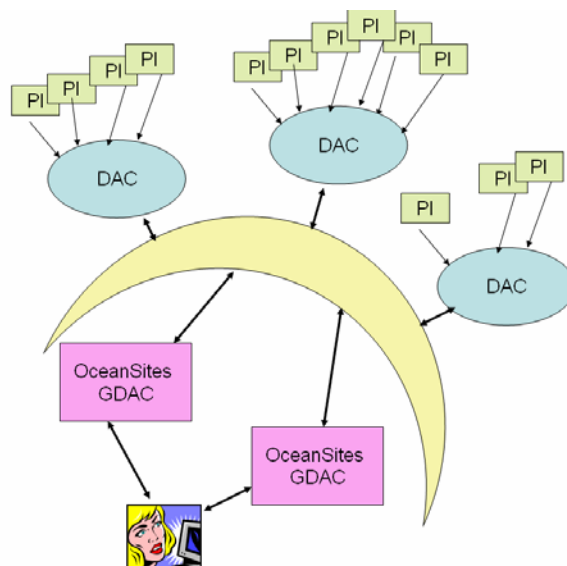
Definition of the roles of the different contributors to OceanSites

The OceanSITES network will be a collection of data servers registered in these two catalogues. Each data server resides in a single place and can serve one or more mooring sites coming from one or more institutions. Each mooring site described on OceanSites WWW site should reside only on one server.

The roles of the different contributors to OceanSITES network can be define the following way:

- **Pi:** He is responsible of a specific Site. He assures the at sea activities, and provides to the Dac the data and metadata necessary for delivery at OceanSITES project.
- **Dac:** A Dac (Data Access Center) is responsible of setting up an Ocean site server, according to the specification approved by OceanSites data management group. He guarantees that the data available on his server are compliant with the format and the Quality control procedure that have been define by OceanSites project. Each Dac has the responsibility to organize data processing, data transfer and update with the Pis he is working with. Each data server resides in a single place and can serve one or more mooring sites coming from one or more institutions. Each mooring site described on OceanSites WWW site should reside only on one server.
- **Gdac** (Global data access center): a Gdac is in charge of providing a vitual access to the data that are served by the Dacs . They are in charge of maintaining the same OceanSITES catalogue (ISO19115 suggested). These two portals should synchronize their catalogue periodically. Once the FTP servers will be on, in order to give more user friendly visibility to OceanSITES data, Gdac's portals should implement view services for the OceanSITES dataset. We should study open source solutions available for discrete data that may have been developed by other groups for such distributed datasets

A Dac has to have the capability to guarantee data availability and integrity on their server 95% of the time for at least 5 years.

A Gdac can also play the role of a Dac for his country or continent.
For security reasons we would like to have two GDACs for Ocean sites
We have to study the value added of putting an OpenDap protocol upon these FTP servers

To help operational users in retrieving automatically these data, on a regular basis, OceanSites project should minimize the number of different FTP servers. We suggest having one per continent

Structure of an FTP server for a Dac

1- Structure of an FTP Server:
   a. An index file containing a line per file describing principal metadata which would allow a user to build his own recovery program especially for latest updated datasets without using OpenDap technology A consolidated index file would be built at GDACs
   b. A directory per Site then per mooring then per Deployment
      <site name>/<mooring name>/ <deployment id>
   c. In each <deployment directory> we propose to have the following organization.
      /Metadata
      /1DTimeseries parameters (like microcats, CTD)
      /2D timeseries parameters (like ADCP)
      /Video
      / <other type of data> that don't fit into the categories above

      We need to have a metadata file per deployment because the mooring can be modified. Because of Netcdf organization, to optimize file space and not generate useless big files, it's should be better not to mix in a same file data data of different organization (To be confirmed)

      i. One metadata file per deployment that describes the mooring
      ii. In each data directory a file containing a copy of some metadata that describe the mooring, and the time series of the different parameters measured by this mooring.

---

iii. RealTime data and Delayed mode data will be stored in different files in the same directory. The naming convention will inform the user

Remarks:

1. The naming convention will be important to inform the user on the data that are contained in the file without reading it. A suggestion to be refined could be
<R or D><Site><Mooring><deployment ID><type of parameters>.nc

2nd step
Study solutions to connect OceanSITES servers that are not able to convert their data in OceanSITES format. We should encourage OceanSites contributors to only use this solution exceptionally. As MMI project is studying such techniques we should work with them to solve this problem if we need to solve it

Organization:
2 Gdac portals: Proposed: Associated with ICoads (NOAA/NCAR) and Coriolis
Dacs servers :
 • European moorings : Coriolis
 • ….To be completed

1or 2 archiving facility: TBD