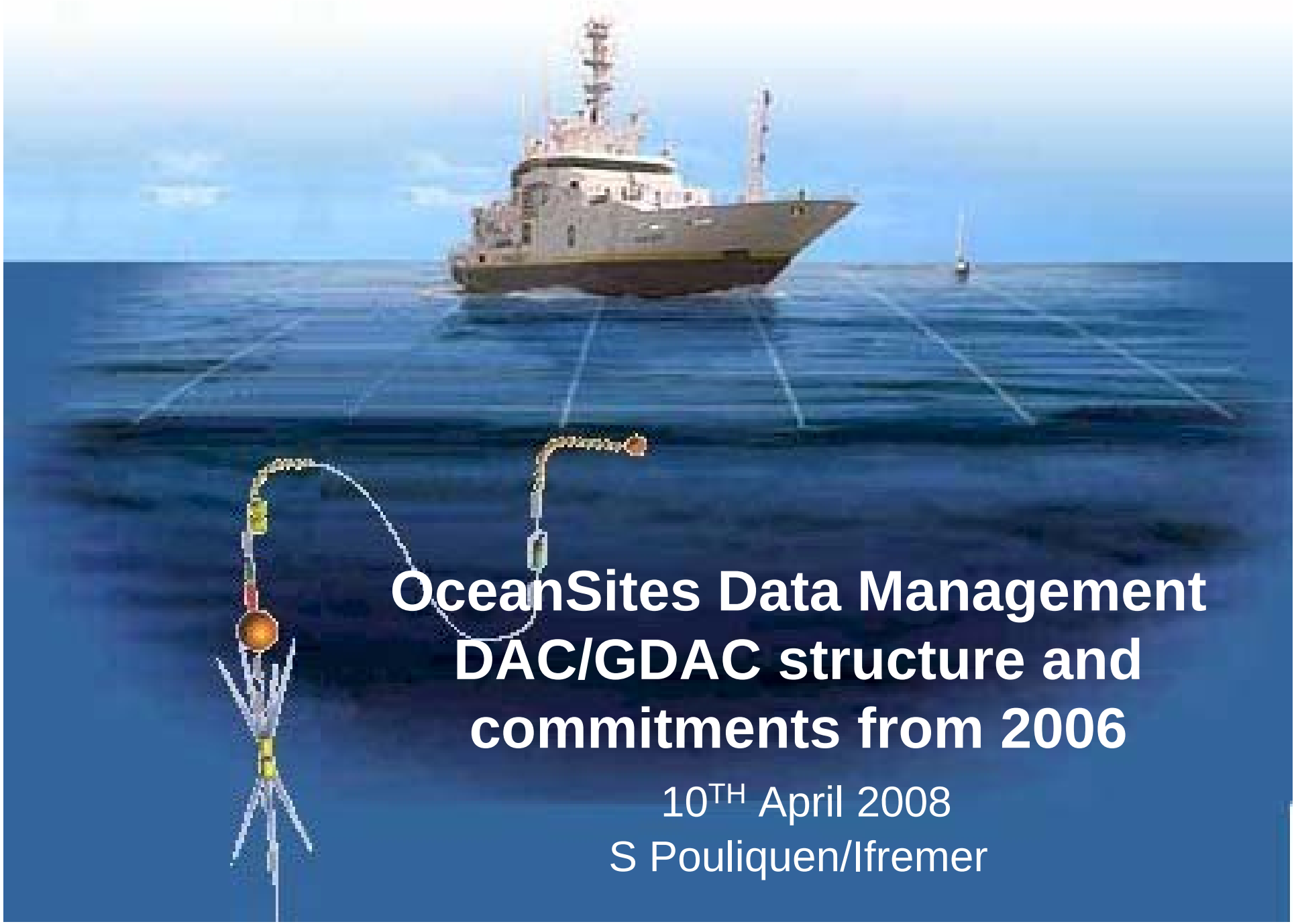


OceanSites Data Management DAC/GDAC structure and commitments from 2006

10TH April 2008

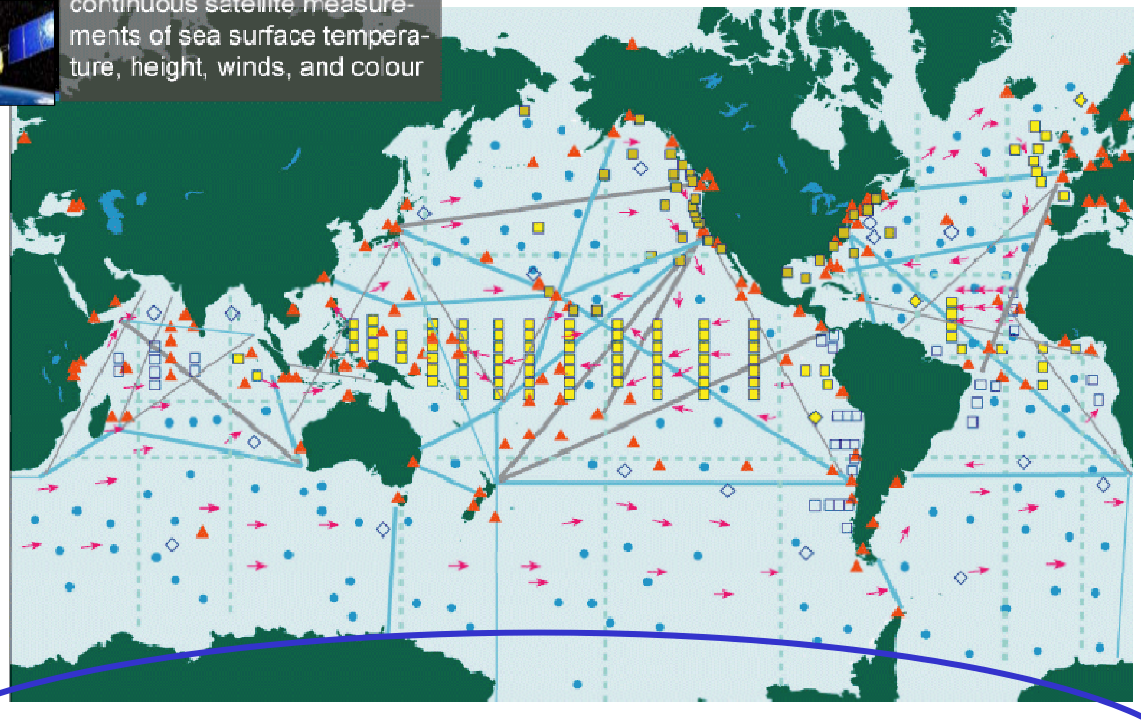
S Pouliquen/Ifremer

Initial Global Ocean Observing System for Climate

Status against the GCOS Implementation Plan and JCOMM targets

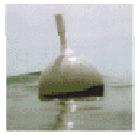
Total *in situ* networks **60%** February 2008

continuous satellite measurements of sea surface temperature, height, winds, and colour



87% **Surface measurements** from volunteer ships (VOSclim)

200 ships in pilot project



100% Global **drifting surface buoy** array

5° resolution array: 1250 floats

62% **Tide gauge** network (GCOS subset of GLOSS core network)

170 real-time reporting gauges

81% **XBT** sub-surface temperature section network

51 lines occupied

100% **Profiling float** network (Argo)

3° resolution array: 3000 floats

43% Repeat **hydrography** and **carbon inventory**

Full ocean survey in 10 years

Reference time series 24%

58 sites

48% Global **reference mooring** network

29 moorings planned

79% Global tropical **moored buoy** network

119 moorings planned

GCOS Ocean Observations

GOOS

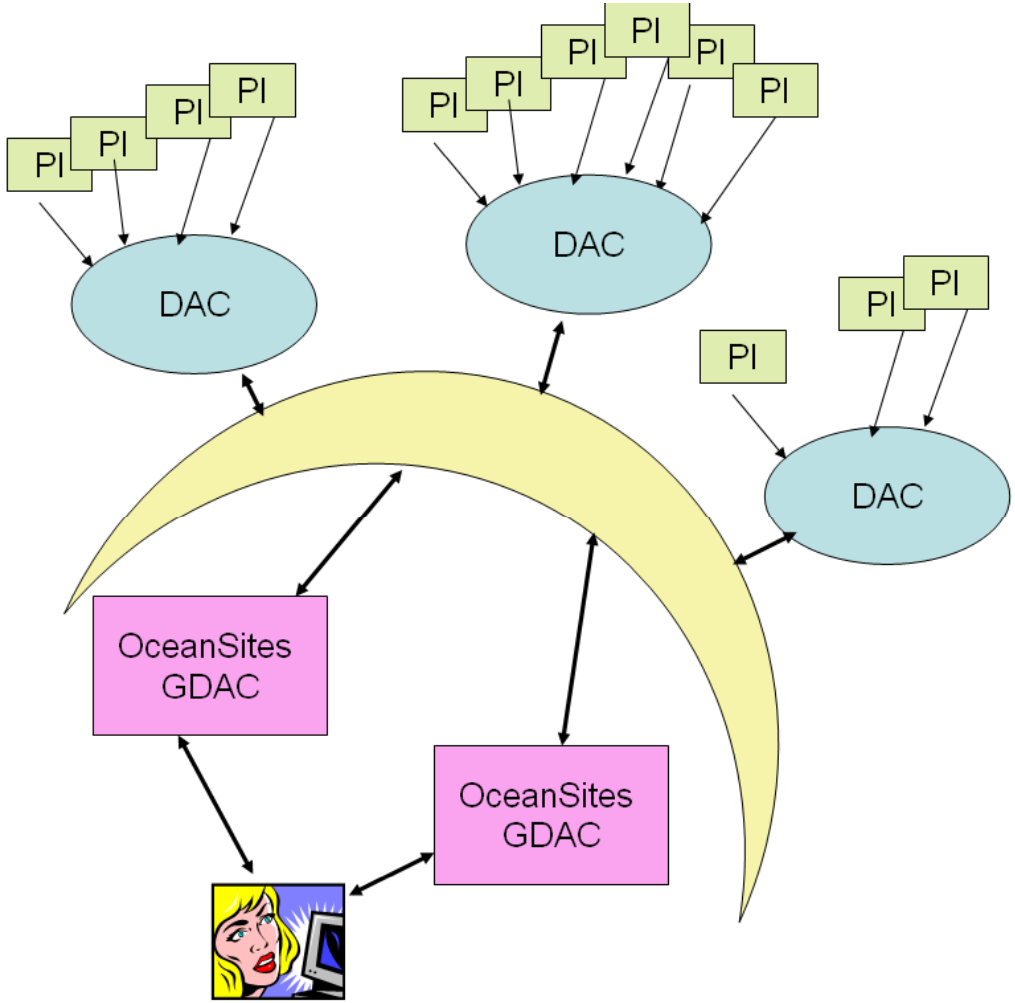
jcomm OCEANOGRAPHY & MARINE METEOROLOGY

What are the motivations to work on Data mgt

- There are potential users both in research community and operational oceanography fields for both forecast and reanalysis
- Data access is improving within different domains that are addressing the same communities(Argo, Gosud, Carbon, Clivar…)
- National and International programs are emerging (Orion/OOI in USA, GMES in Europe, GEO at International level…)

⇒ There is a need to build for OceanSites a data management organization that is coherent with these programs without reinventing the wheel

An architecture have been proposed



The role of the 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 an Ocean 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, data transfer and update with the Pis he is working with.

It 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 role of the different actors

- **Gdac** (Global data access center): a Gdac is in charge of
  - providing a virtual access to the data that are served by the Dacs .
  - maintaining the OceanSITES catalogue
  - synchronizing his catalogue with the second GDAC.
  - In Future implementing viewing services for the OceanSITES dataset.

Two Gdac for security reason

We should study open source solutions available for discrete data that may have been developed by other groups for such distributed datasets

Data Format

- Unique ID:
  - Unique Site ID by maintaining a Catalogue at GDAC Level. Introduction of a new site submitted to OTS science team
  - UNIQUE station id: Site ID+ serial number assigned by the GDAC
    - If GTS transmission : WMO number is needed (unique Id provided by WMO **A84nn** where A represents the WMO region where mooring is deployed, and nn represents the mooring ID in that region. We therefore assume that there won't be more than 99 IDs required in every WMO region.
    - The project name should be recorded somehow in metadata
- At GDAC level only the best version of the data is available.
- There may be additional products (interpolated data ,…) provided by PI/DACS to Gdacs
- Original data + additional data should be archived either at National level of via an OTS archive facility if there is a candidate
- A dataset can contain data retrieved in RT and data recovered from the instrument : Data-mode should be at measurement level
- Recommended that both good and bad data should be kept in files transmitted to GDAC and marked with appropriate QC flags
- There was an agreement on using the Flag proposed in the manual but we have to refine the comments to set these flags in a uniform way through PIs and DACS

Data network

- Agreement on the role of PI and GDAC
- Agreement that as the DACs are not funded specifically via OceanSITES we should not put too much requirement on them.
- Minimum requirement for a DAC: put/push data available in OceanSITES format on an FTP/to GDAC FTP site. Then GDAC will bring the data on a centralized server according to the GDAC design
- Future possible evolution :
  - Add an OpenDap /Dapper server on GDAC ftp to be able to provide remotely viewing/access services to OceanSITES data (Like IPRC already does for other datasets)
  - Some Dacs to implement OpenDap/Dapper servers linked together to GDACs

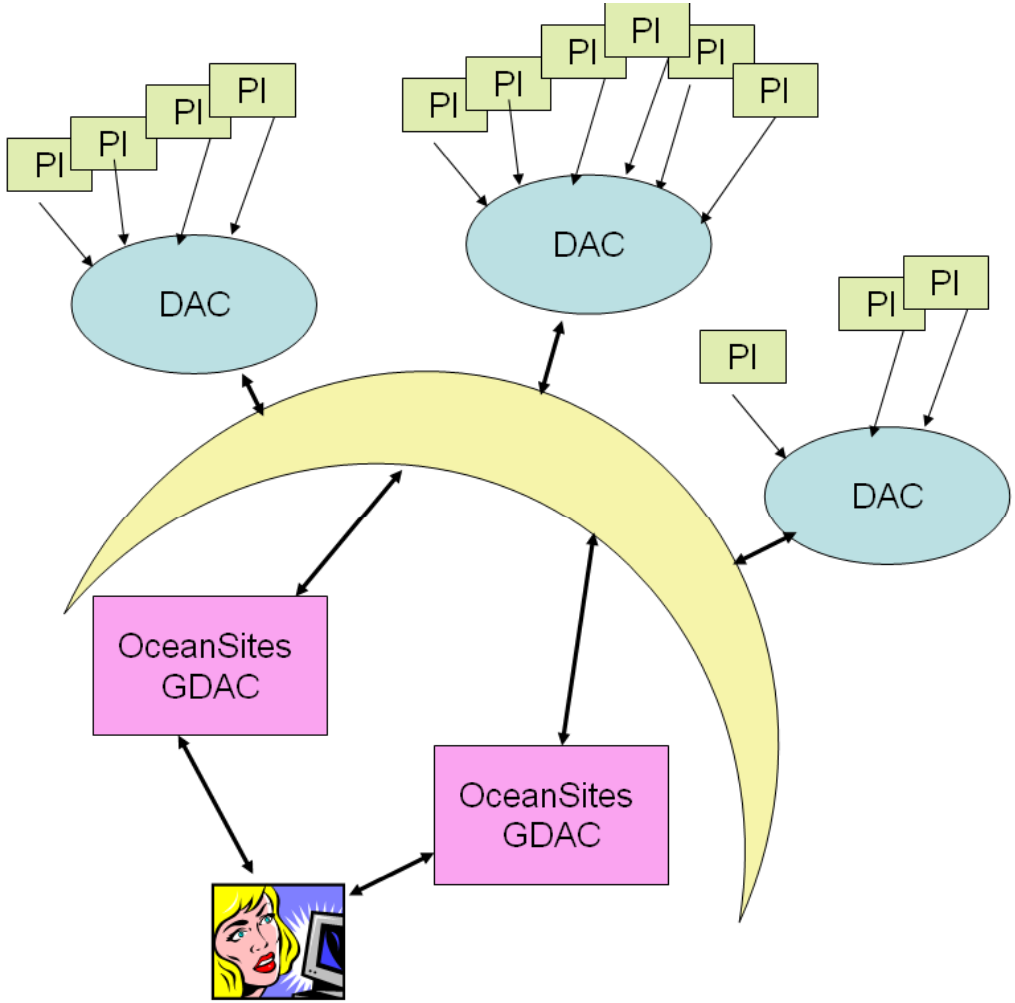
Data Network Which TOOLs

- Need to make some trial test before deciding if we can use “operationally” Dapper and MAS (IPRC volunteered)
- MMI can help us as a clearinghouse on metadata expertise
- Need to be in contact with BODC/GF3 and CF teams to inform them on how we use their standards and get the evolution we need..
- When Opendap type of server is implemented we may should be able to develop visualization and downloading facilities on this network a bit more sophisticated that a “get“ from an FTP…

Proposal of Commitments

- PMEL : ok to provide Keo & then TAO/Pirata data in oceansites format
- MBARI could volunteer to be one of the GDACs… OK to provide some of the MBARI data in OceanSites
- 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 OCANSITES format
- CCHDO: could provide ship occupied CTD repeated stations. Will provide tool to convert in 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… Will continue the way they started with Coriolis
- IPRC:: Can Act like a DAC for some orphan OCEANSites DAC .IPRC could pursue work on DAPPER and LAS to give visibility of the network..
- HAWAII: HOTS: CTD data , WHOTS with WHOI (RT and DM)
- WHOI : has already started and will continue
- BATS : CTD data like for HOTS
- Coriolis volunteers to host one of the GDAC
- NOAA is interested in hosting one GDAC in 2007

An architecture have been proposed



Present status

- DAC:
  - Coriolis with NOC: DAC for European Sites
  - NDBC : DAC for some USA sites : WHOI + NDBC + ???
  - PMEL: DAC for TAO/TRITON/PIRATA & Keo
  - CDIA: DAC for CO2 data
  - CCHDO : ????
  - IPRC : ???
- GDAC
  - Coriolis
  - NDBC

Future Steps

- Present status:
  - GDACS retrieve data from the DACS and make it available on an FTP site.
  - Presently once a day could be up to 4 times a day
  - Coriolis plan to provide an additional access via OPeNDAP
  - ➢ This will always stay as it serve the users who don't have RT needs

- How to serve users who want to access data within minutes after is available ? :
  - In this case data have to stay at the DAC
  - The DAC has to set up an OpenDap /Dapper server
  - GDAC has to set up an entry point to link these OPeNDAP servers to provide the portal to users
  - Data on the DAC must be in the OceanSITES format