Data Management Activities

Bob Keeley

OOPC- 9

Southampton, Jun, 2004

OIT Project Issues

These include:

- effective telecommunication
- needs for agreed standards and protocols
- provision for innovative data inquiries
- interoperability independent of individual platforms

Summary

- ISO-19115
- Metadata
- Data dictionary
- QC standards
- Metrics
- Unique tagging
- xml
- Surface salinity
- Time series moorings
- Argo
- CLIVAR
- DMACS / Russia
- New sampling strategies

ISO-19115

- ISO-19115 is a standard for describing collections of data
- There is a joint activity undertaken by JCOMM / IODE ETDMP and ICES / IODE SGXML groups
- They are comparing various metadata models including FGDC, EDMED, CSR, Blue Pages to ISO-19115
- The work ties in with a common terminology for data and metadata
- Results will be presented to ICES, IODE and JCOMM in 2005

Metadata in real-time

Taking the OOPC requirement stated to GSC for JCOMM to “develop and implement … the real-time transmission … of all metadata relevant … for SST and subsurface temperature profiles…”

- JCOMM OPA proposes BUFR as the means to accomplish part of this but recognizes more thought needed.

- A first draft of the scope and specifications of a pilot to accomplish this Is just completed. It divides the problem into 4 categories
    a) real-time,
    b) operational/pulled,
    c) delayed,
    d) historical

    And proposes a pilot to deal with categories 1 and 2.

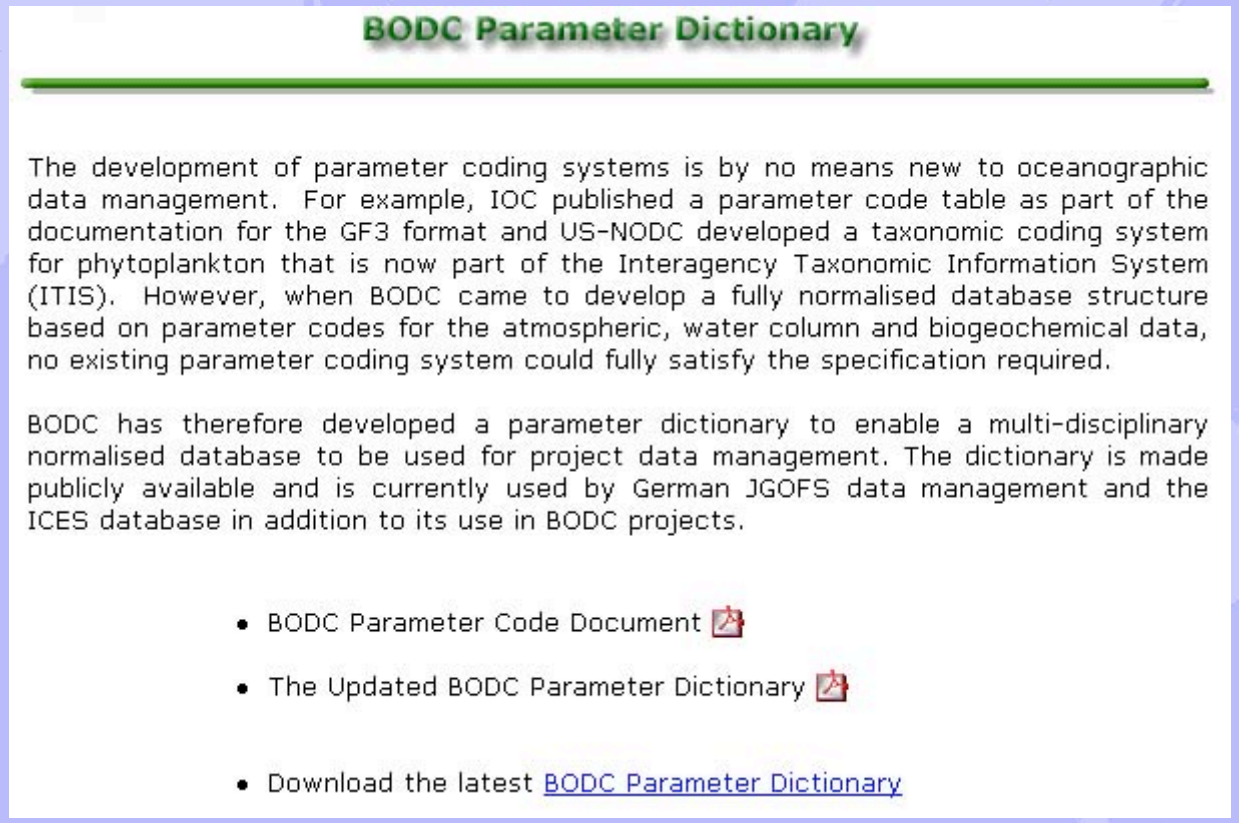
Data Dictionary

Objective: To standardize the naming of variables and metadata

- BODC has mapping of variables from UK, France and working on Netherlands and Sweden
- more than 14,000 entries
- documentation and search capability on-line

- MEDS is working with institutes in Canada and with NODC and intends to work with BODC
- allows on-line searching, editing and new entries
- first step in converging to a common terminology

BODC on-line



BODC Parameter Dictionary

The development of parameter coding systems is by no means new to oceanographic data management. For example, IOC published a parameter code table as part of the documentation for the GF3 format and US-NODC developed a taxonomic coding system for phytoplankton that is now part of the Interagency Taxonomic Information System (ITIS). However, when BODC came to develop a fully normalised database structure based on parameter codes for the atmospheric, water column and biogeochemical data, no existing parameter coding system could fully satisfy the specification required.

BODC has therefore developed a parameter dictionary to enable a multi-disciplinary normalised database to be used for project data management. The dictionary is made publicly available and is currently used by German JGOFS data management and the ICES database in addition to its use in BODC projects.

- BODC Parameter Code Document
- The Updated BODC Parameter Dictionary
- Download the latest BODC Parameter Dictionary

http://www.bodc.ac.uk – follow links to documents

BODC Sample

ICES-IOC SGXML Parameter Dictionary : BODC Dictionary

Free form text search (leave this blank to retrieve all data) : silicate

This page is a single word or phrase search of the Parameter Dictionary which utilises the dictionary term, abbreviation, and methodology fields.

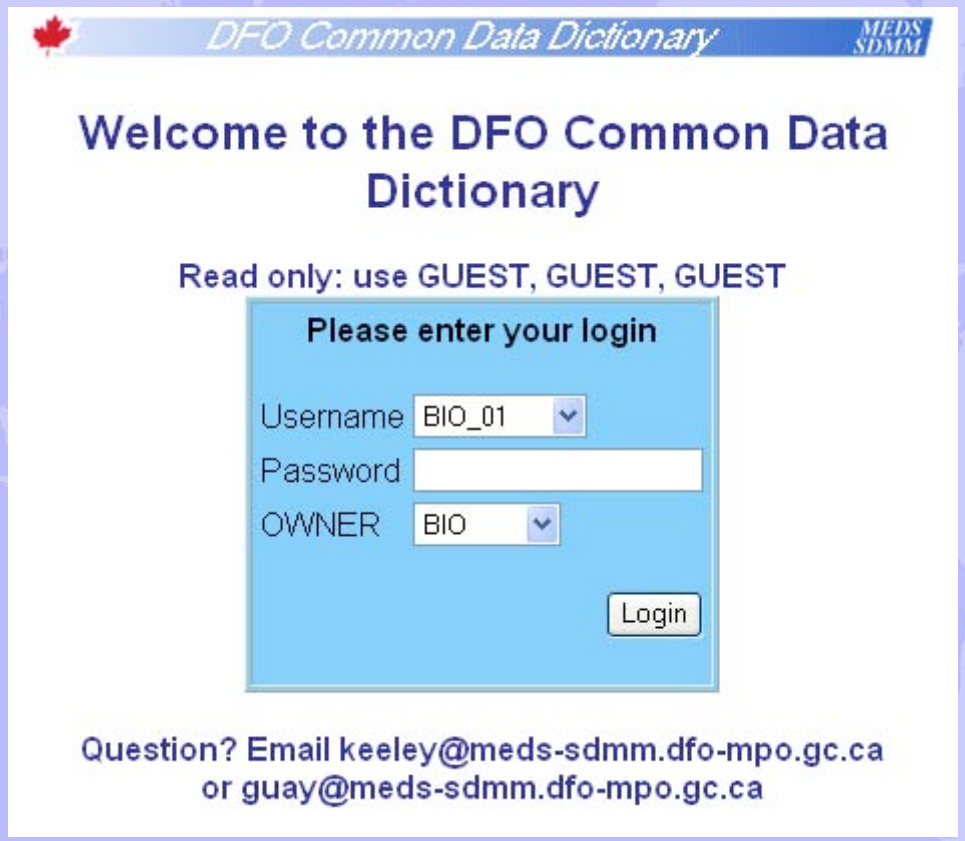
Display the Parameter Dictionary Sorted by :

Dictionary Term | Abbreviation | Parameter Code | Methodology | Unit of Measurement

This output has been sorted by Dictionary Term

dictionary_term	short_name	code	methodology	unit_of_measurement	min_value	max_value	null_representation	accuracy	creation_date
Dissolved silicate	Silicate	SLCAAADZ	Colorometric autoanalysis (unspecified filter type)	micromol/l	0.0	500.0	-1.0	3.2	19940916
Dissolved silicate	Silicate	SLCAAAD5	Colorometric autoanalysis (0.2 um pore filtered)	micromol/l	0.0	500.0	-1.0	2.2	19970508
Dissolved silicate	Silicate	SLCAAAD1	Colorometric autoanalysis (GF/F filtered)	micromol/l	0.0	500.0	-1.0	3.2	19940916

MEDS on-line



http://www.meds-sdmm.dfo-mpo.gc.ca/meds/About_meds/standards/login_e.asp

MEDS Sample

Find a code:

Find a meaning: silicate

Search Results:

Code	Meaning	Units	Status	Date Entered	Owner	Category	Edit
DSO$	The date as YYYYMMDD of analysis of Silicate	date	Active	2/24/2004 11:54:48 AM	MEDS	Date and time within day	
SLCA	Silicate (SIO4-SI) Content	mmol/m**3	Active	3/4/2004 10:31:19 AM	BIO	Chemistry	
SLCA	Silicate (SIO4-SI) Content	mmol/m**3	Active	3/4/2004 10:31:19 AM	IML	Chemistry	
SLCA	Silicate (SiO4-Si) content	mmol/m**3	Active	2/24/2004 11:54:53 AM	MEDS	Chemistry	
Silcate	Silicate	umol/L	Active	3/4/2004 10:31:19 AM	IOS	Chemistry	

QC Standards

• Argo has agreed on automated real-time procedures and more recently on delayed mode QC practices for salinity. Manuals describing both have been prepared and is available from the GDACs.

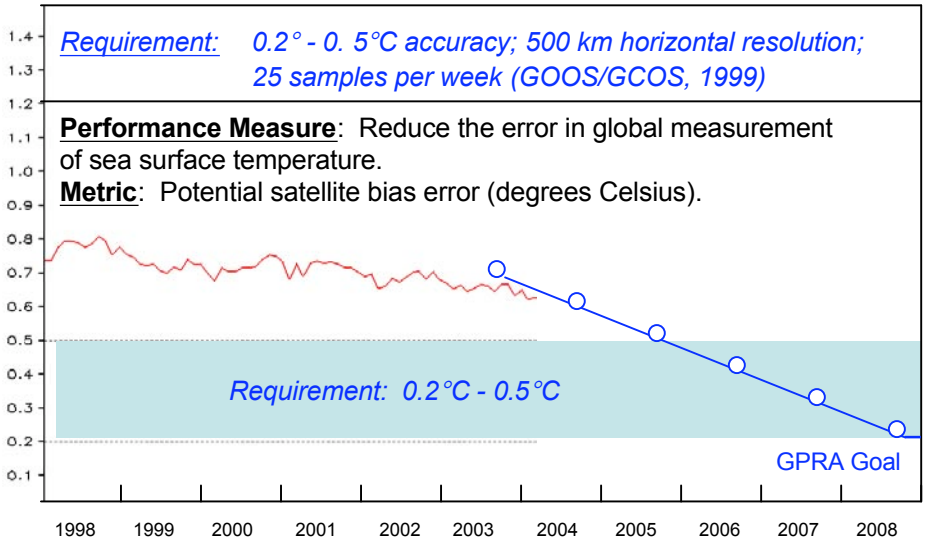
• Argo GDAC (IFREMER) and GTSPP real-time data centre (MEDS) starting a comparison of procedures with GODAE (Jim Cummings).

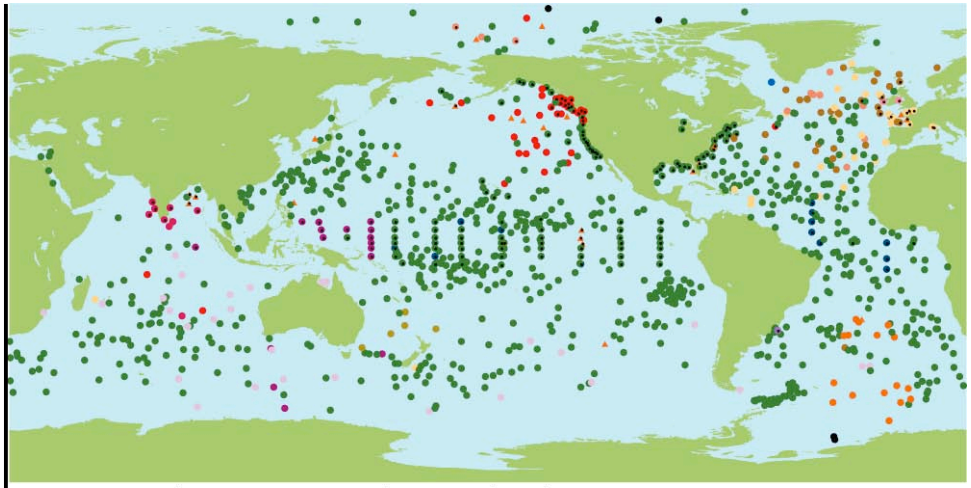
• JCOMM / ETDMP has a project to standardize procedures and the common link is through Keeley.

Metrics

- Collaboration between Mike Johnson of OPA and Bob Keeley of MEDS with possible contribution by Etienne Charpentier of JCOMMOPS
- Using OOPC defined climate observing requirements to show the progress of observing system to meet targets
- harmonizing presentations between draft versions
- first joint presentation for Q4 of 2004 available early in 2005
- Accessible from links at JCOMMOPS
- should be able to address SSS soon

Observing System Status: 2004, Q1.
Sea Surface Temperature

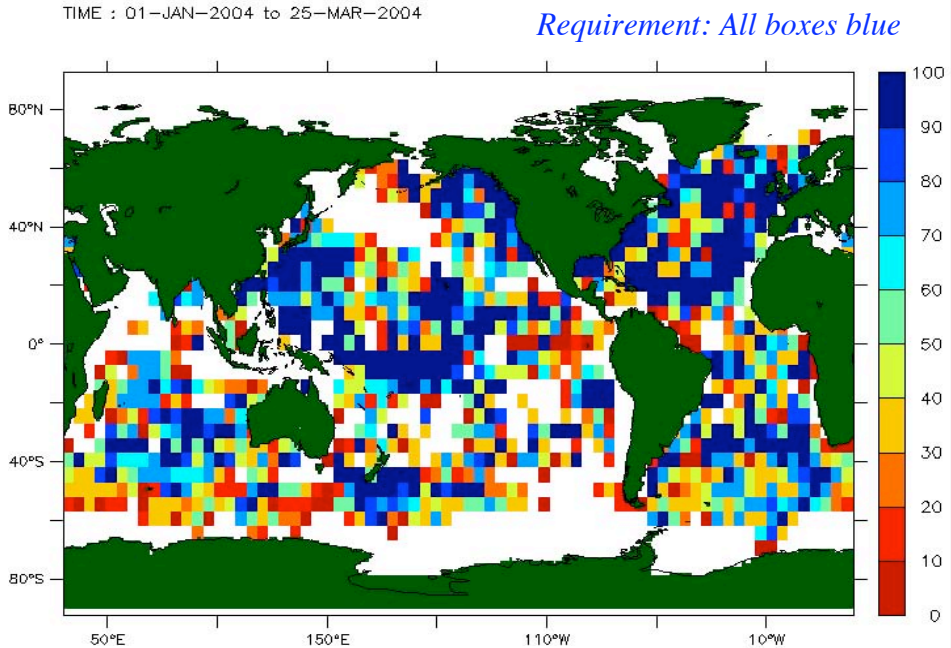




DBCP status by country, March 2004 (data buoys reporting on GTS)

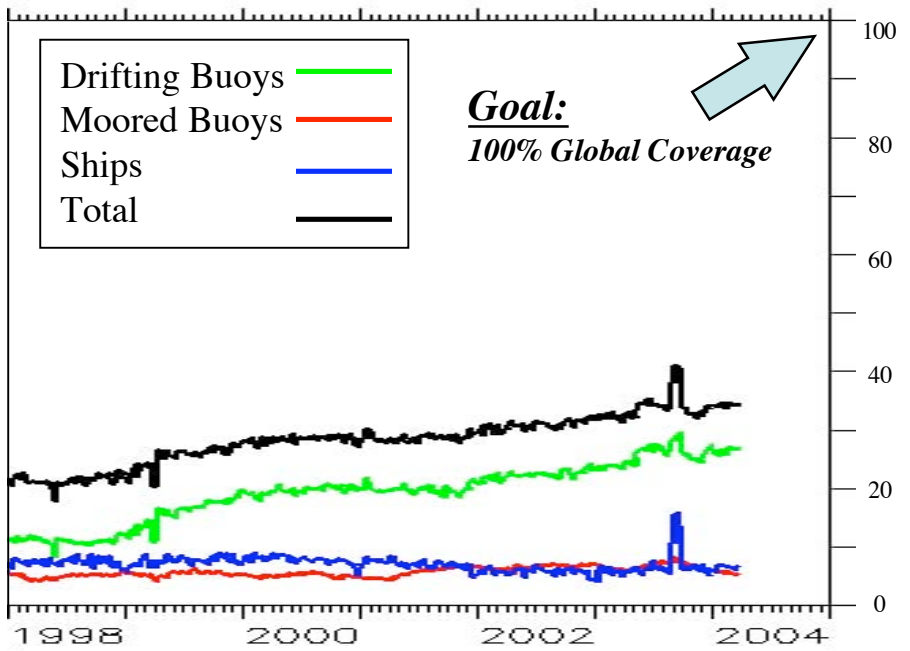
Drifting buoys: 909 Moored buoys: 174 Ship observations not shown

- AUSTRALIA (29)
- BRAZIL (1)
- BRAZIL/FRANCE/USA (13)
- CANADA (23, 16)
- FRANCE (23, 8)
- GERMANY (4)
- INDIA (3, 6)
- IRELAND (2)
- JAPAN (4, 12)
- NETHERLANDS (1)
- NEW ZEALAND (5)
- NORWAY (12)
- SOUTH AFRICA (16)
- UNITED KINGDOM (27, 7)
- UNITED STATES (761, 110)
- MOORINGS
- UNKNOWN



Drifting Buoys + Moored Buoys + Weighted Ship Observations

Percent of Weeks with 25 Accuracy Weighted sst obs in 5x5 Box 2004

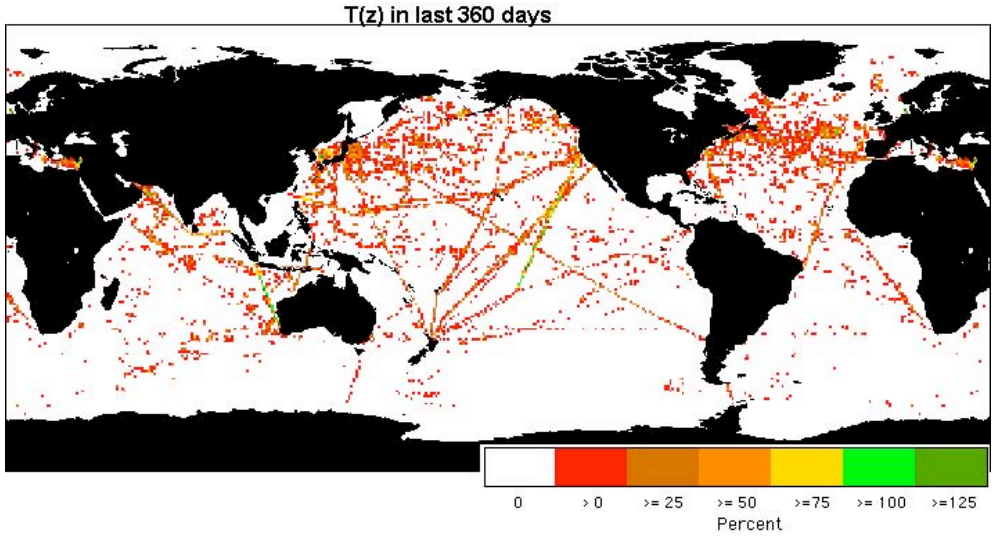


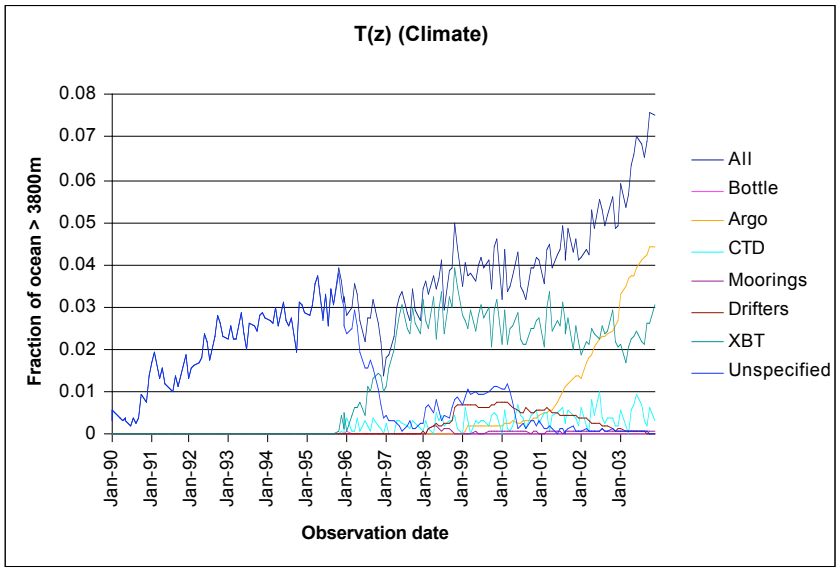
Percent of 5x5 Boxes with 25 Observations per Week

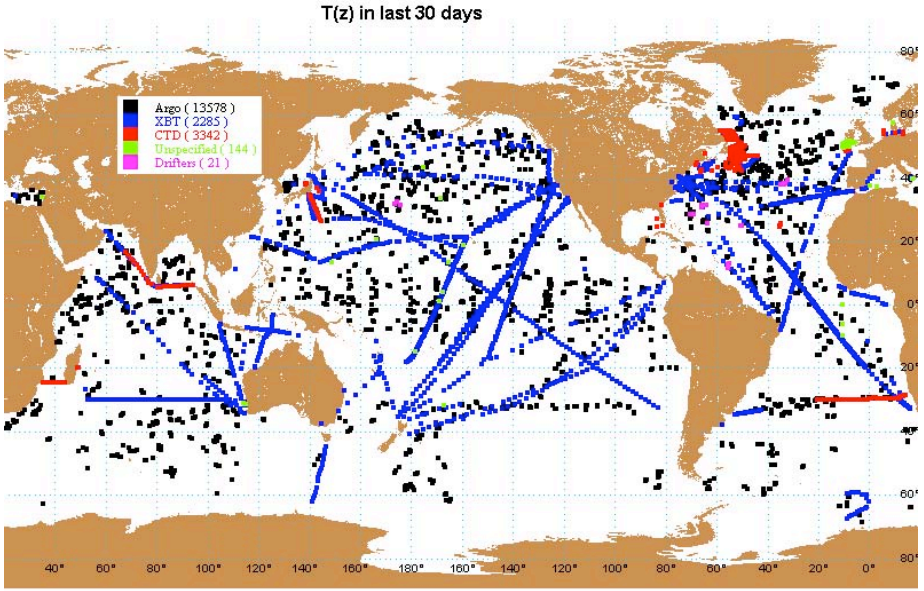
Climate Upper Ocean Temperature

We require

- Observations in deep water (measured as a fraction of the world with water depths > 3800m)
- Profiles deeper than 700m
- One profile every 30 days and every 1° latitude x 1° longitude
- Data delivery within 30 days of collection.
- Accuracy of ~0.1° C







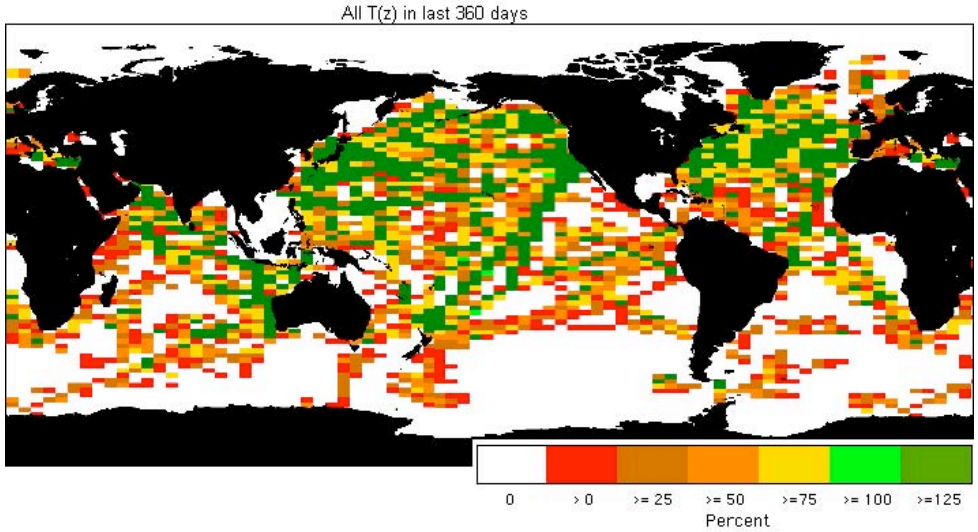
MEDS – SDMM
Jun, 2004

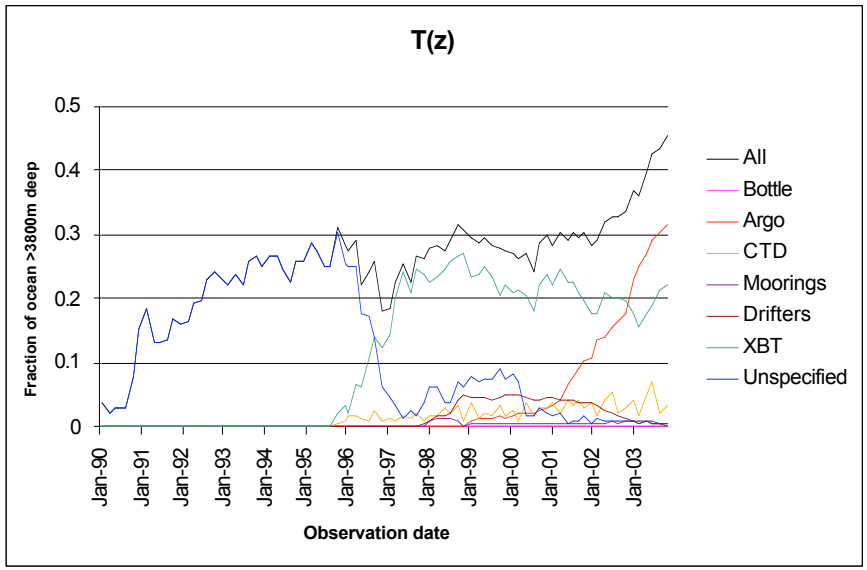
OOPC, Jun 2004

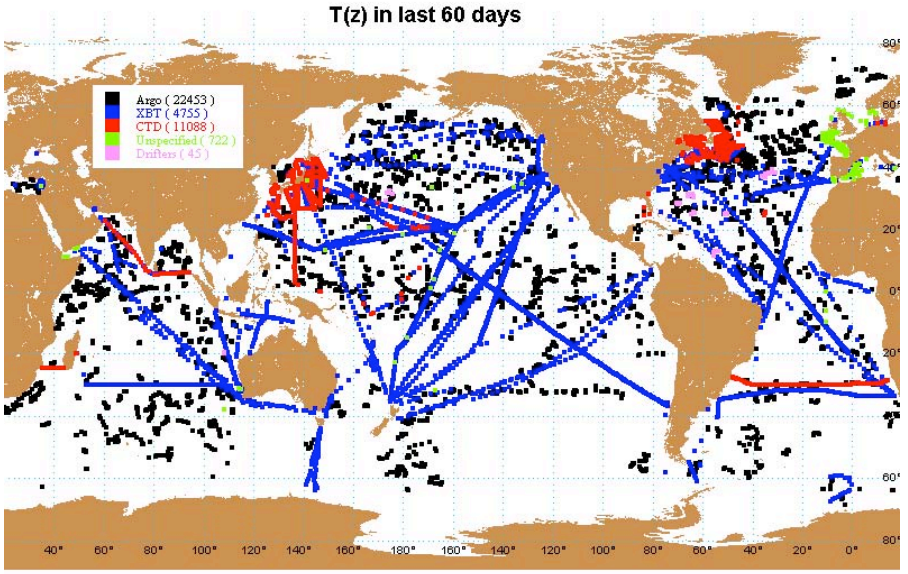
Large Scale Upper Ocean Temperature

We require

- Observations in deep water (measured as a fraction of the world with water depths > 3800m)
- Profiles deeper than 700m
- One profile every 60 days and every 2º latitude x 5º longitude
- Data delivery within 30 days of collection.
- Accuracy of ~0.5º C







MEDS - SDMM
Jun, 2004

OOPC, Jun 2004

Unique Tags

- Allows for unequivocal match of real-time to original profile arriving in delayed mode
- Collaboration between MEDS, SEAS, Australia for profile data (not Argo)
- Started in April, 2004
- Initial analysis of effectiveness to be completed by Sep, 2004
- Extensible to other kinds of real-time data
- ETDMP looking at how to generalize the idea for all kinds of data with a report for March, 2005

SEAS and CRC32

"Best" data now

Full resolution on disk

SEAS process

Create BATHY to calculate CRC

BATHY

GTS

Full resolution with CRC

Calculate CRC and add to data

MEDS

NODC

BATHY with CRC

Match CRC of full resolution to BATHY

XML

The SGXML WG of ICES / IODE:

- Japanese working on scheme to accommodate data from Tokyo Bay
- Sample mappings of profiles, current meter records, water levels, data derived from net tows to “xml bricks”
- Carried out comparisons of xml bricks to GML and identification of commonalities
- Results will be placed on Marine XML website
  http://www.marinexml.net/
- EUMarineXML project appears to be concentrating on electronic charts and developing a registry of schemas and documentation

Surface Salinity

• Organized under the Global Ocean Surface Underway Data (GOSUD) Project

•France has built a global centre and is now offering data in a similar way to how Argo data are offered.
http://www.ifremer.fr/sismer/program/gosud/

• Site is collecting both real-time and recent data for download.

• A collaboration has been agreed with HRMM group to acquire data through their connections to US vessels. The details are to be worked out, but we will devise a straightforward way to link surface met and ocean data even though they reside in separate archives.

• Tools are being built to compare data received directly at the global centre and what are circulating on the GTS. Comparisons will be used to encourage real-time distribution of the data and greater contributions of data to the Project.

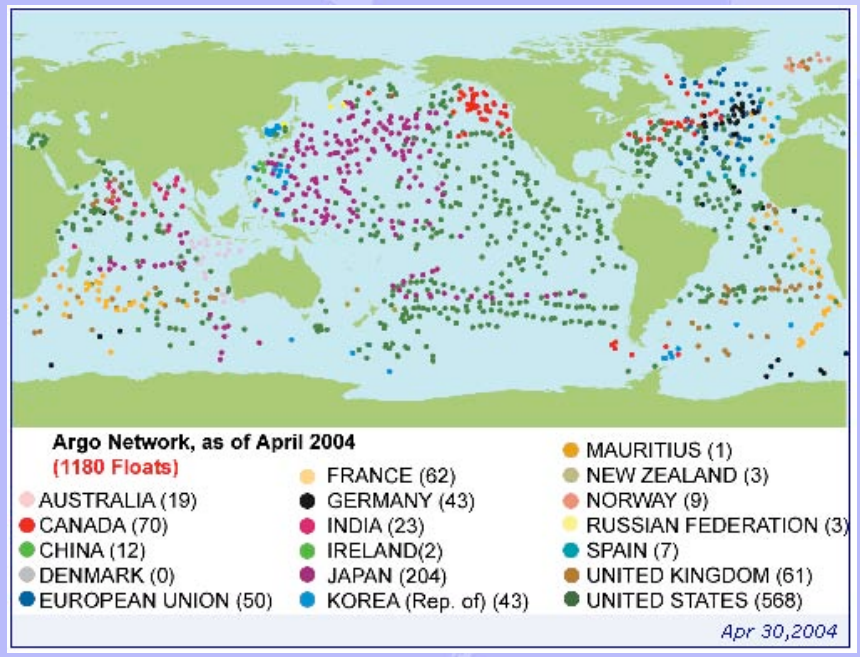
Time Series Moorings

- Reported on by Weller

Argo Development

~1200 floats operating

Oceans getting filled



Argo Data Management

- Agreement on delayed mode salinity procedures and “record keeping”
- Format change at start of 2004 to accommodate information on salinity QC
- First delayed mode data starting to arrive at GDACs
- First CDs will be issued in next 12 months.

- Upcoming challenges
  - floats keep evolving – now reporting oxygen from ~20 floats
  - moving towards real-time data distribution in BUFR
  - developing a global float performance picture
  - acquiring recent CTD data for salinity QC
  - correcting data from floats with sensor drift before GTS distribution

Performance Report (draft)

Report date (yyyymmdd): 20040604

- ID = WMO number
- Active = Y if float still operating and N otherwise
- Manufacturer = float manufacturer
- Instrument = details about the float and sensors
- Controller = details about the controller board
- Duty_Cycle = Time between cycles (days)
- Prf_press = profile depth (m)
- Start_Date = date of deployment (as yyyymmddhhmm)
- Start_Lat = deployment latitude (+ or - 90 +ve N
- Start_Long = deployment longitude (+ or - 180 +ve W
- Ldate = date of the last cycle received (as yyyymmddhhmm
- Exp_levels = number of expected levels
- Exp_cycles = (Ldate {or today}-Start_Date)/Duty_Cycle
- NFull = number of cycles missing fewer than 5 levels
- NPart = number of cycles missing 6 or more levels
- Missing = number of cycles missing

CLIVAR

- Global Synthesis and Observations Panel – chair: Dean Roemmich
- First (pre) meeting on data system: 24-26 Mar 2004, Scripps
- Some action items:
  - DACS to set up OPeNDAP servers
  - DACs to report new data on web pages monthly or quarterly
  - DACs to assist drafting of DM requirements document
  - Basin Panels to identify observations taken and where data are
  - Develop requirements for a “DIU”
  - DACs collaborate with GODAE on QC
  - SSS, Surface Met, ADCP DACs to examine integration
  - Re-analysis workshop to examine QC’ing historical data
  - WHPO Best practices manual to be revised
  - DACs to examine better integration of data
  - CLIVAR DACs to hold regular meetings

DMACS / ETDMP

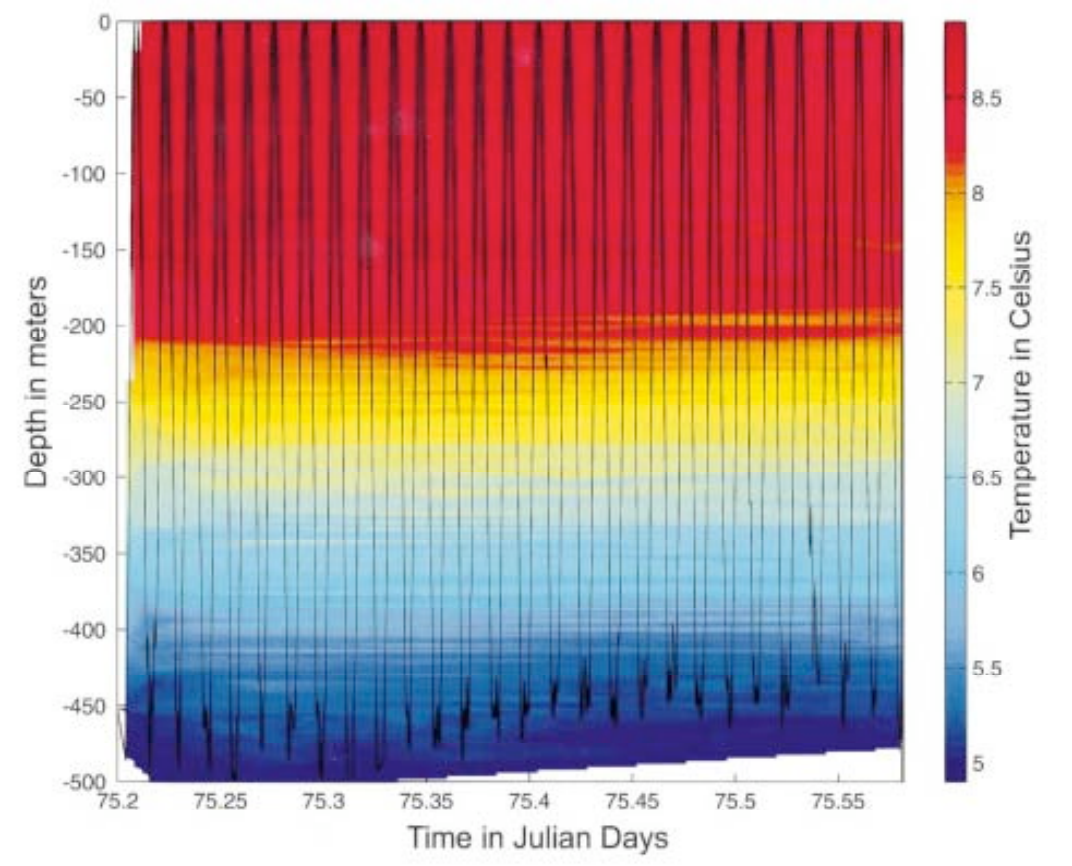
• JCOMM ETDMP had meeting last Sep attended by Steve Hankin of DMACs

• Working on a pilot with the Russian WDC to show how to fuse data from different sources.

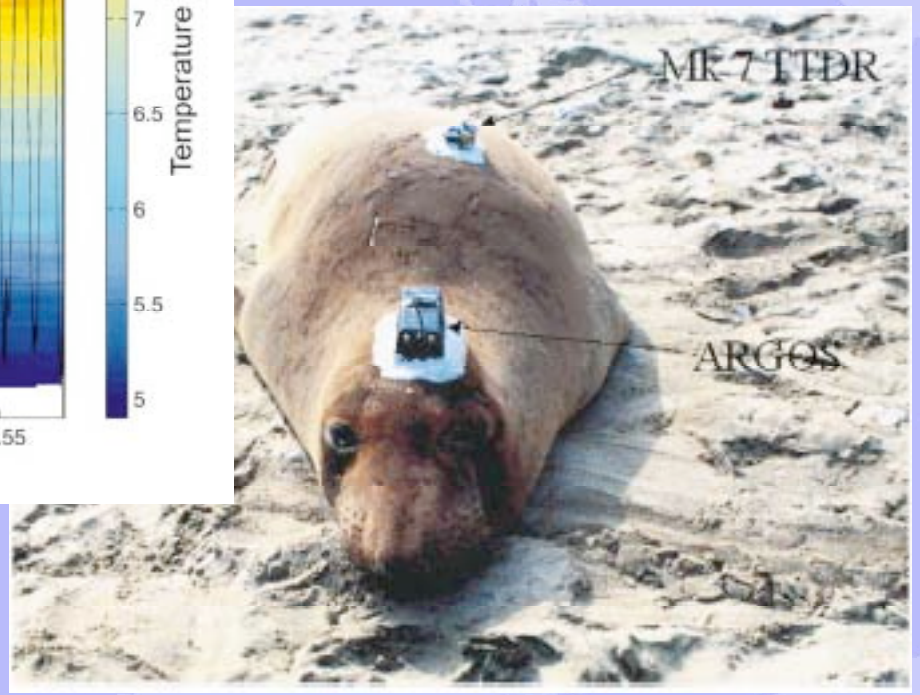
• Target for demonstration is JCOMM-2, Halifax, 2005

• JCOMM representative to attend DMACS meeting

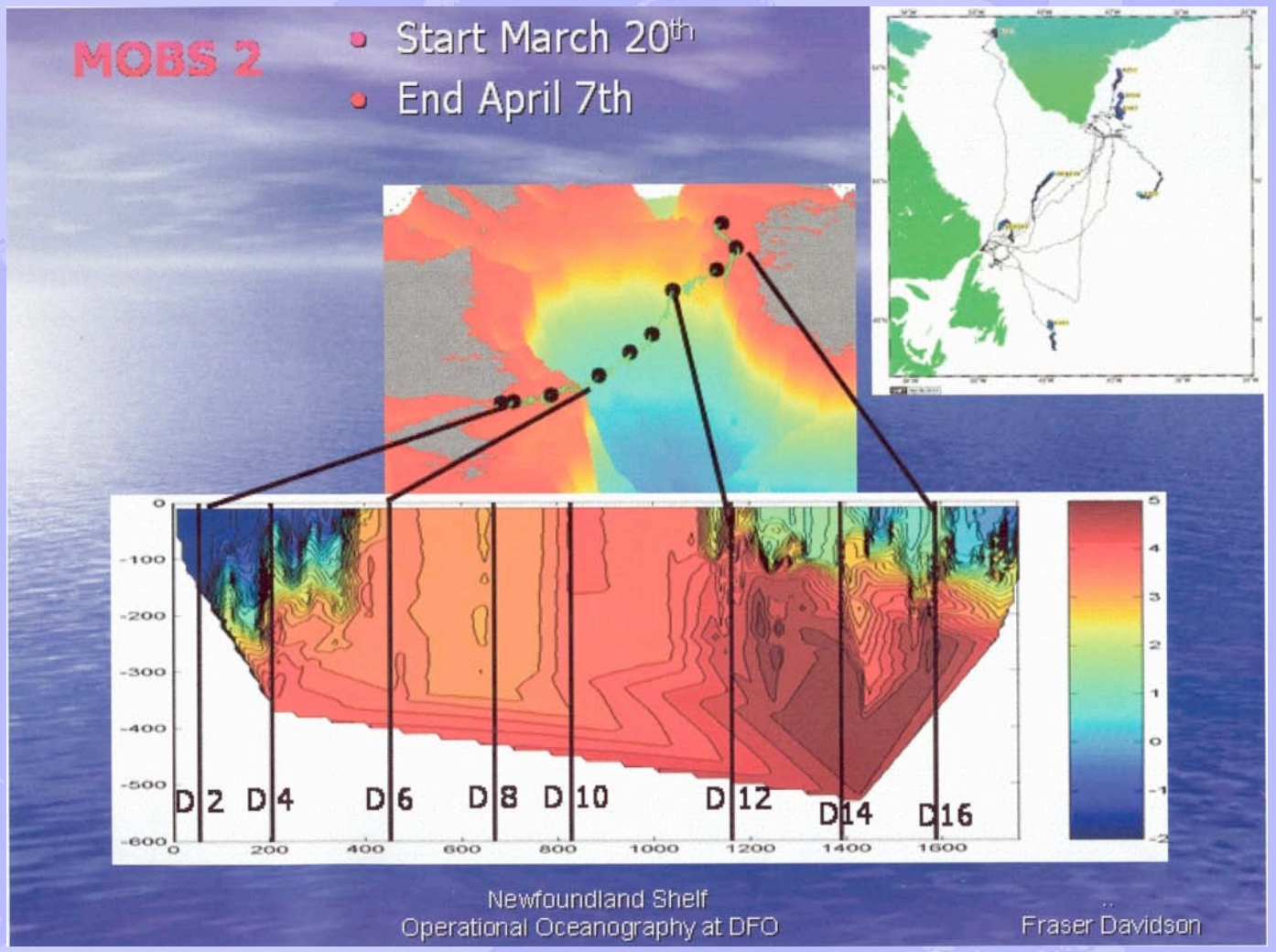
New Sampling



This is not very new but there are people seriously discussing real-time exchange of such data



Most recent example



Actions

1. Clearly defined targets for the data systems to meet are important. They send a message to implementation bodies that there are goals against which they can and will be measured.

2. OOPC needs to review its targets, to clarify those that are still fuzzy and periodically revisit existing ones to ensure they still make sense.

3. OOPC should make direct comments on the adequacy of the and suitability of steps taken by JCOMM and IODE to improve the data systems

4. OOPC should contribute to the review of IODE to influence its evolution.