Oceanographic Data Management

IGBP Working Group B2

prepared by Roy Lowry (BODC, UK)
and Bernard Avril (JGOFS IPO)

Objective

- To provide recommendations and guidelines on data, metadata and information management issues to IGBP and SCOR for their oceanographic programmes

Strategy

- Examine information management practices and legacy of JGOFS, GLOBEC and LOICZ
- Look at lessons learned
- Compare JGOFS and WOCE information management strategies
- Look at the relevance of non-IGBP international oceanographic data, metadata and information management infrastructure

Strategy

- Discuss a range of generic oceanographic management issues
- Discuss the working draft of the IGBP/SCOR Data Management Policy
- Draw together our conclusions for the WG report

JGOFS

- Practices
  - National programmes sometimes supported by national data management
  - National data management (where it existed) co-ordinated by DMTT reporting to the SSC
  - Part-time post at the IPO supporting DMTT through web site, metadata management and liaison with JGOFS contributions not covered by the DMTT
- Legacy
  - Breaking the biological and chemical data management “taboo”
  - Enhanced data management infrastructure
  - Discovery metadata in GCMD
  - DVD of discrete data sets

Other IGBP/SCOR Projects

- Contributions welcome!

JGOFS Lessons Learned

- **Basics can be surprisingly difficult**
  - What is the scope of a project in terms of data sets, cruises or even national activities?
  - Data sharing culture is not universal/homogeneous even in a project with a data sharing policy!
- **Using existing infrastructure leads to problems**
  - Heterogeneous data sets
  - Heterogeneous data management coverage (including no coverage which inevitably means no data delivery)

JGOFS Lessons Learned

- Centralised information management required from the outset
- Coherent data management framework required from the outset
- Getting adequate data management resources is almost impossible

JGOFS Lessons Learned

- Agreeing the rules for data sharing within the programme at the outset pays dividends
- Obtaining support for these rules from funding bodies pays dividends
- Involvement of the scientific community in data management pays dividends
- An attractive carrot is worth ten big sticks

Other Projects / Programmes

- Other lessons learned?

JGOFS / WOCE Comparison

- **Two areas where WOCE scored better than JGOFS**
  - Project scope definition
    - Tightly managed as cruises, sections and data sets
  - Nature of the final product
    - Data organised into type collections (CTD, ADCP, etc.)
    - Consistent structure and format

JGOFS / WOCE Comparison

- **Why?**
  - WOCE established a centralised information management unit (DIU) even before the project started. JGOFS equivalent (IPO) was not fully functional until 1996 - half way through the project.
  - International JGOFS data management was a federation of national initiatives whereas WOCE obtained national resources for the benefit of the international programme. Argo is following the WOCE strategy with equal success.
  - WOCE was mostly concerned with physical oceanography and modelling.

IGBP–Relevant International Activities

Possible Examples:

- JCOMM
- GOOS
- ICES
- IOC/IODE
- WDC network
- Others?

Generic Issues

- Data policy (politics and principles of data sharing)
- Data and metadata management functional specification
- Data and metadata management infrastructure models
- Implementation and resourcing
- Roles of scientists and data managers
- Data management technology
- Scope of current oceanographic data management

Data Policy

- **Data moratoria for inter-project and intra-project data sharing**
  - Who decides them?
  - Who polices them?
- **Data usage accreditation**
  - Who decides acknowledgement or co-authorship?
  - What if data used and not attributed to the originator?
- **Data quality standards**
  - Who determines what is acceptable?
  - Who applies these standards?
- **Data interoperability**
- **Long term data accessibility and security**

Functional Specification

- **Objective is to present all project data and metadata as an integrated entity**
- **Different PI "scenarios"**
  - Unsupported individual
  - National programme with national data management infrastructure
  - Project with "integral" data management
- **Can we produce a route map from these PI sources to integrated data sets?**

Infrastructure Models

- Centralised or distributed?
- Granularity of distribution
- Incorporation of existing infrastructure
- National “agenda” and/or international “focus”?

Roles of Scientists and Data Managers

- Scientific input to data management is essential
  - Does anybody dispute this?
  - How can this best be achieved in practice?
- Data managers should
  - Help scientists do better science by:
    - Accelerating data access
    - Raising confidence in data by assuring quality
    - Providing effective data handling and visualisation tools
    - Providing access to comprehensive baseline data
  - Assure the future legacy of the project

Technology

- What do we want from technology?
  - Interoperability
  - Increased value from existing accessible data
  - Cost-effective data management
  - Access to ‘hidden’ data sets
  - Capability to use data in operational (near real-time) mode
  - ??

Science Not Addressed by Current Data Management?

- What are the implications of increasing socio-economic focus of science on oceanographic data management?
- What are the implications of bio-informatics?
- What else should be we looking out for?

IGBP/SCOR Oceanographic Data Management Policy

- **Some issues for discussion**
  - Time frame, scope and distribution of cruise reports
  - Is a Publications Committee necessary?
  - What is the value of elevating data sets to citable status?
  - Should project data management get involved in data rescue and archaeology?

IGBP/SCOR Oceanographic Data Management Policy

- Peer review of all project data sets is an important component of the policy
  - Is this a practical proposition?
  - Cruise or discipline based?
  - Comparison with WOCE DAC model
  - What is the added value?
- Should the vehicle for the metadata portal be incorporated in policy?
- Role of existing NODC infrastructure
- Other issues?

IGBP/SCOR Oceanographic Data Management Policy

- Should data and information in IGBP/SCOR oceanographic projects be:
  - Attributed to the responsible data producer
  - Established as accurate through quality control and analysis
  - Interoperable to facilitate global integration and synthesis
  - Identifiable so as to be citable
  - Accompanied by complete and accurate metadata conforming to appropriate international standards
  - Accessible as and when required by project scientific activities
  - Released into the public domain in a timely manner
  - Properly secured and preserved for the long term
  - Anything else?