

[date]

[address]

Ref. JGOFS/...

Subject: Data Management for future ocean biogeochemistry/climate programmes: Lessons learned from JGOFS

Dear Dr. ...,

This letter concerns the issue of data management for ocean biogeochemistry in new research programmes being planned under the auspices of IGBP and SCOR. We believe that this matter is highly relevant for you, but if you know of a colleague who is more directly responsible with support and funding as well as management of ocean data in your institution, we would be grateful if you could also pass our letter to that person.

The JGOFS Project has been highly successful in providing new insights into global biogeochemical cycling in the oceans through a multi-national effort. A considerable effort was invested in new measurements of ocean properties during JGOFS. Yet a much smaller stress was directed toward ensuring the accessibility and ultimate stewardship of these expensive and irreplaceable data through a coordinated international data management effort. As new programmes are being designed and implemented, we must learn from the JGOFS data management experience. The following list is a set of recommendations for new programmes to consider:

- Establishment and support of a centralized International Project Data Centre responsible for: acquisition of data, data access, and distribution of data to the World Data Centres system to ensure its long-term archival;
- Establishment and support of experienced/full-time national data coordinators who will: identify cruises and Principal Investigators (PIs) associated with the national programmes, work with PIs to ensure data and metadata are complete and in common file and data formats and submit data to the International Project Data Centre;
- Setting of a time limit before which all countries participating in the programme must submit their data to the Data Centre;
- Establishment of a coherent set of standards for reporting data and metadata;
- Working with funding agencies to ensure compliance with programme data policies.

These recommendations are designed to ensure the rapid dissemination of data and its long-term preservation and accessibility.

During the JGOFS Project, key biological and chemical variables were sampled by over 20 countries at the regional scale (process studies in the North Atlantic, Arabian Sea, Equatorial

Pacific, Southern Ocean and North Pacific), global scale (carbon survey) and from long-term measurements at key ocean sites. As we proceed with the final synthesis and modelling phases of JGOFS, it is likely that JGOFS may fall short in providing a satisfactory data legacy for future generations.

The JGOFS data management plan was set up so each nation had a data coordinator responsible for that nation's data. Data are either managed by a national JGOFS data manager (e.g., Australia, Canada, France, Germany, India, Japan, U.K. and U.S.), or reside with individual PIs. A Data Management Task Team (DMTT) was formed to coordinate the data management efforts, but in effect, the DMTT does not represent all nations involved in JGOFS activities, and does not have the manpower or financial resources to go far beyond coordination and cataloguing of data collections carried out nationally. The lack of a centralized International Project Data Centre severely hampers the use of JGOFS data for synthesis and model validation, now and in the future.

Additional problems identified with the current JGOFS Data Management are:

- Ambiguity in many countries as to what constitutes a JGOFS cruise;
- No time limit, and in most countries, no requirement for delivery of data to a Data Centre from where it can be properly archived and disseminated;
- Reluctance by PIs to share data;
- Data in diverse formats with incomplete documentation or missing key core JGOFS parameters.

Efforts to acquire funds to compile all JGOFS international data into a common file and data format, to be distributed internationally, have thus far failed in the U.S. Other national funding agencies have also been approached, in vain. These efforts failed mainly because such supporting activities, although extremely important to make optimal use of the scientific data, have fared poorly in competition with proposals to initiate new science projects. This is very unfortunate, since the acquisition of data is very costly compared to the small cost of effective data management.

JGOFS was proactive in setting a data policy and establishing an international data management committee very early in the Project, and the JGOFS system became a model for other programmes (e.g., GLOBEC). Yet, in spite of this effort and because of the complexity of interdisciplinary data and of JGOFS itself, much remains to be done to secure the accessibility of all data collected in the Project. We offer these recommendations to the planners of the next generation of ocean programmes, to help all move toward a new and better, integrated data management system in the next few years.

We are very interested in your views on the data management issue facing future programmes and invite you to share your thoughts with us. We will be passing on a more comprehensive view of the JGOFS experience to planners of the next generation of ocean research programmes and look forward to your input.

Sincerely,

[signature]

Hugh Ducklow  
(Chair, JGOFS SSC)

[signature]

Margarita Conkright  
(Chair, JGOFS DMTT)

[signature]

Bernard Avril  
(Asst. Executive Officer, JGOFS IPO)