

## Update On JGOFS Data And Information Management

by Bernard Avril

Over the last year, the JGOFS Data Management Task Team (DMTT) has undertaken a number of activities with the assistance of the International Project Office (IPO). Some of the most important are highlighted in this update on the task of locating, organizing and preserving the vast array of data collected by JGOFS investigators from many nations over more than a decade.

### New Web Site Launched

Last July, the IPO launched a new JGOFS Internet site (<http://www.uib.no/jgoofs>) with three objectives in mind. The first is to improve the overall organisation of the site and to make it more "user friendly." The second is to preserve accumulated JGOFS knowledge with minimal support until the programme ends in December 2003. And the third is to facilitate the handover of information to future ocean biogeochemistry and ecology programmes as part of the JGOFS legacy. The IPO welcomes suggestions for improving the quality and long-term value of the web site. Specific efforts have been made to exhibit JGOFS achievements in data management and to increase the visibility of proper data management in the marine sciences (<http://www.uib.no/jgoofs/DMTT/dmtt.html>).

### Data Rescue Meeting

One of the goals of the DMTT and the IPO is to identify, document and collect all data and associated metadata from JGOFS studies. Accordingly the IPO invited representatives of countries that participated in JGOFS but are not represented on the DMTT to a "data rescue" meeting, held in Ispra, Italy, in June 2002 to learn about their JGOFS activities, to encourage them to contribute their data to the proposed Interna-

tional JGOFS Master Data Set, and to discuss with them the technical details of transferring, integrating, distributing and archiving their data sets and metadata.

National representatives were invited to attend from Belgium, Italy, The Netherlands, Norway and Spain. Participants in the meeting were DMTT chairwoman Margarita Conkright of the U.S. National Oceanographic Data Center (NODC); JGOFS Global Synthesis Working Group (GSWG) chairman Reiner Schlitzer; JGOFS/GAIM Task Team chairman Patrick Monfray; Maaïke Vancauwenberghe of the Federal Office for Scientific, Technical and Cultural Affairs, Belgium; Cesare Corselli of the Milano-Bicocca University, Italy; Taco de Bruin of The Netherlands Institute for Sea Research; Jan-Roger Andersen of the Norsk Marint Datasenter, Norway; Aida Fernández Ríos of the Instituto de Investigaciones Marinas, Spain; Michael Diepenbroek of the World Data Centre for Marine Environmental Sciences (WDC-MARE), Germany; Charles Trees of the U.S. National Aeronautics and Space Administration; Nicolas Hoepfner of the European Union Joint Research Centre, Italy, and Bernard Avril of the JGOFS IPO.

The national representatives benefited from a clear overview of international data management collaboration, from the expertise of the DMTT and the World Data Centres and from the development of tools dedicated to data management and visualization. They also gained insights into ocean modelling, thanks to the presentations and comments by the modellers. On the other hand, the national presentations provided valuable information about relevant projects, cruises and data sets.

Scientists were encouraged to collaborate with the WDC system for future studies as well as for JGOFS

data sets. Conkright pointed out that one central, unique place for data sets is better for scientists, and this should be the WDC system.

She noted, further, that original quality control (QC) flags are not removed in data sets deposited in the WDC system; new QC flags are added instead. Different versions are labelled with unique date stamps and digital object identifiers. When QC flags are not consistent or when errors are found by the WDC, the data originator is contacted for comments. The final products identify the scientists associated with each variable in the data set and their institutions, in addition to the data submitters and their institutions.

Since that meeting, the various countries represented have made significant progress with regard to data management matters related to JGOFS science. For example, the Dutch JGOFS data sets are currently being transferred into the WDC system, several CD-ROMs with specific data sets from various countries are being completed, and a new national oceanographic data centre has been organised in Italy. More information about the Ispra meeting is available via the JGOFS home page ([http://www.uib.no/jgoofs/Publications/other\\_pub/Ispra\\_Data\\_June2002.pdf](http://www.uib.no/jgoofs/Publications/other_pub/Ispra_Data_June2002.pdf)).

### JGOFS Core Parameters

With help from the IPO, DMTT members have updated the list of JGOFS core parameters, along with the most common units and metadata. Last fall, the members of the JGOFS Scientific Steering Committee (SSC), the DMTT and the IPO reached a consensus on a minimum list of JGOFS parameters that participating national programs should focus on in their submissions to the JGOFS master data set.

The core parameters list was published in JGOFS Report no. 37 in

August 2002 in almost final form. It is available via the JGOFS web site ([http://www.uib.no/jgoifs/Publications/Report\\_Series/JGOFS\\_37.pdf](http://www.uib.no/jgoifs/Publications/Report_Series/JGOFS_37.pdf)).

### Final Data Products

The ultimate purpose of the DMTT is to provide scientists with a comprehensive biogeochemical data set in a common file and format for use not only in current modelling and synthesis activities, but also in future global change studies. The task team is currently finalizing the compilation of major JGOFS data contributions.

The Pangea information system initiative for JGOFS at the World Data Centre for Marine Environmental Sciences (WDC-MARE) is continuing (see description in *U.S. JGOFS News* 11:4, pp. 18-19). Its goal is full accessibility and long-term archiving of all JGOFS data and meta-data through the WDC system. Yet the pace has been slow because of the lack of basic funding. Members of the Pangea team met in January in Bremen, Germany, with JGOFS representatives to assess the Pangea initiative with regard to production of the JGOFS master data set (see [http://www.uib.no/jgoifs/Publications/other\\_pub/JGOFS-PANGAEA\\_Bremen2003.pdf](http://www.uib.no/jgoifs/Publications/other_pub/JGOFS-PANGAEA_Bremen2003.pdf)).

The work plan developed in Bremen was discussed further at a DMTT meeting in mid March, held at the British Oceanographic Data Centre (BODC) in Bidston, UK. Attending the meeting were DMTT chairwoman Margarita Conkright, Cynthia Chandler of the U.S. JGOFS Data Management Office, Jaswant Sarupria of the Indian National Oceanographic Data Centre, Marie-Paule Torre of the JGOFS-France/Processus Biogéochimiques dans l'Océan et Flux (PROOF) Data Management Office, Joachim Herrmann of the German JGOFS Data Management Office, Roy Lowry of the BOFS Data Management Office, BODC, GSWG chairman Reiner Schlitzer, and Bernard Avril of the JGOFS IPO.

DMTT members agreed on an overall plan for final JGOFS data products with the following components:

- Each JGOFS contributing country with DMTT representation will finalize its own data products before March 31, preferably as CD-ROMs or on an ftp site.
- The DMTT, with help from the JGOFS IPO, will organize the production of an international data DVD, to be titled "International JGOFS Data Collection, 1988-2000. Volume 1: discrete data sets." This DVD will be distributed during the JGOFS Open Science Conference, to be held in Washington, D.C., in May.
- This DVD will include an integral copy of all JGOFS CD-ROMs and data products available by March 31. For already published products, reproduction agreements have been secured or are being sought. The DMTT will hold the copyright for this DVD product, which will be structured by countries, projects and cruises as much as possible. The whole product will include 10 contributions never previously released. The U.S. NODC has agreed to organize the production of the master copy of the DVD and 1,000 copies, with financial help from the JGOFS IPO.
- The DMTT supports the plan of the Pangea team to publish a second volume in 2004 or 2005, to be titled "International JGOFS Data Collection. Volume 2: integrated data sets." All data sets in this volume should be accessible in a common file and data format through the Pangea user and visualization interface and be organised by project, cruise and data set for each parameter. The DMTT and IPO will continue to work with the Pangea team until JGOFS comes to an end in December.