

INTERNATIONAL DATA MANAGEMENT

J. S. Sarupria

National Institute of Oceanography, Dona Paula, Goa

Introduction

Advances in oceanographic/marine science and technology depend to a significant degree upon the effective flow of information and data from the collectors to various types of users. The principle purpose of a World Data Centre is to provide on a continuing basis information/data in a useable form to the so called "Secondary users" community i.e. to individuals or organisations which have or will in the future have need for the oceanographic information. This is done when the primary purpose for which the data were collected, has been satisfied.

The international centres are a focal points for the accessing and dissemination of oceanographic data and information. Efficient acquisition, processing, archieving and distribution procedures are needed to provide service not only to the scientific community but beyond it to the larger community of users in public and private agencies concerned with marine policy and economic development.

History and Objectives

During the International Conference on Oceanographic Research, 1960 at which the setting up on Intergovernmental Oceanographic Commission was recommended, certain initial tasks for the new commsission were outlined and amongst these was included the requirement for a structure to co-ordinate Intern tional Oceanographic Data Exchange. The first formal establishment of a Working Group on exchange of oceanographic data was made by IOC Resolution I - G adopted during October 1961, which prescribed that :

The mission of the Working Group shall be to facilitate the exchange of oceanographic data, standardization of forms for reporting and coding data, the encouragement for the preparation of data catalogues and the assistance in the development of a National Oceanographic Data Centres (NODC).

At the Sixth Session (September 1967) the Commission amended the term of reference of the Working Group by adding the following :

1. To review and re-appraise the whole International Oceanographic Data Exchange System.
2. To continue and encourage the present work leading towards format standardization for use with automated techniques for data input, storage, retrieval, dissemination and exchange.
3. To continue the development of means whereby geological, geophysical, biological, air-sea interaction data and data form continously recording' sensors are incorporated into the International Oceanographic Data Exchange System.

Later on in the Fourteenth Session of the IOC (Paris 1987) the name of the Working Committee on International Oceanographic Data Exchange was changed to the Technical Committee on Oceanographic Data and Information Exchange, with acronym IODE.

IODE system consist of National Oceanographic Data Centres (NODCs), Responsible National Oceanographic Data Centres (RNODCs) and two World Data Centres WDC-A (Washington, USA) and WDC-B (Moscow, USSR).

The principle users of oceanographic data archived in Data Centres of the IODE system are; International agencies, planning and development groups at the regional level, coastal and offshore industries such as fishing, shipping, petroleum, coastal protection, national and international scientific programmes. Many thousands of inquiries are answered by the NODCs within the IODE system each year.

Data holdings

The international marine data base now contains data from more than 2,000,000 observations. Included in the data received by the World Data Centres (Oceanography) are data from more than 80,000 oceanographic stations. The number of oceanographic stations for which data are held by the centres now exceeds 900,000 and 400,000 bathythermographic profiles are being archived. Other data holdings in the centres include biological observations, series of current observations and pollution data. A large number of geological and geophysical measurements are held in the World Data Centre in Boulder, Colorado.

At present there are 42 NODCs/DNAs and 13 RNODCs operating under the IODE umbrella. Though it means that only less than 50% of all IOC member states established NODCs. These centres are located in the member states which posses more than 80% of all marine data globally collected. The last few years were marked with an increase of interest among IOC member states in data collection and management - 7 new data centres have been established, 11 member states are considering the establishment of RNODCs and few of them received support from the IOC.

Monitoring mechanism

The international network continues to have big gaps in Africa and Asia. Many NODC's are understaffed and have limited facilities, which do not permit them to copy effectively with increasing demands in Oceanographic Data Management from National and Internationall Communities. In fact, only about 15 NODC's are in a position to take RNODC's responsibilities and out of these 7 carry out RNODC's activities. They are Argentina, Canada, Japan, U.K., USA and USSR.

Exchange of oceanographic data within the international system is monitored through the following mechanism:

- i) The announcement of National Oceanographic Programmes (NOPs) informing planned research activities.
- ii) The "Report of Observations/Samples collected by Oceanographic Programmes" (ROSCOP) which contains information about measurements taken in completed research activities.
- iii) Data inventories published by WDCs, Oceanography, some NODCs and RNODCs.
- iv) The Marine Environmental Data Information Referral System Catalogue (MEDI Catalogue). The next MEDI directory data sets will contain descriptions of holdings in national centres and in organisations with data sets holding such as e.g. TOGA and WOCE analysis centres.

Data formatting systems

IFS - Integrated File System (IFS) has been designed for the efficient management of oceanographic data and information at Indian National Oceanographic Data Centre (INODC). This system provides an easy way in answering the queries for facility for oceanographic data in a classified manner. It identifies four categories of data collection principally, each category has been divided into three or four levels and each level is further divided into several record type. Level 1 and 2 are for inventory and documentation information, Level 3 and 4 are for actual data storage purpose.

General Format-3 (GF3): This is a system originates for formatting geo-scientific data series into sequential files on digital storage device. It is not a fixed format in the conventional sense but is a generalised system which allows to use in a number of ways of organizing his data whichever way he choose, the data fit into GF3. Provision is made in GF3 for the user to describe, in the same files that carry the data, the exact format he has selected and all codes he has used, as well as ample space for plain language documentation. GF3 is a character oriented format and may be used on any digital storage device supporting sequential files. Although designed originally for use with other digital storage media such as Winchester disks, Floppy disks, Optical disks and tape cassette mass storage devices. The underlying structures of the GF3 system is built around logical records of a fixed length of 1920 bytes and all relevant data and information are stored in these records in character coded format.

Although GF3 developed originally as a formatting standard for data exchange purposes, it is equally well suited for use in the archiving of data, It is highly versatile system capable of accomodating virtually any type of digital oceanographic data including physical, chemical, biological, geological, geophysical and meteorological measurements. The main requirement for data series to be included in GF3 is that they are digital and are referenced in a space-time frame work based on geographic coordinates.

The routine collection of geo-scientific data is often geared towards rather specific types of measurements. The occasions will therefore arise when the full flexibility of the GF3 system is not required and where the user may prefer a format tailored specifically to a particular type of data. GF3 is

well suited for this purpose in that it provided a framework within which data specific standard formats can be created. Such formats can be considered as subsets of the GF3 formatting system. In this manner, GF3 may tailor specially to standard type of data as and when required.

Data can be stored and retrieved in GF3 with simple rather short programmes. The GF3 system includes a Fortran-based software package, GF3-Proc, to facilitate the reading and writing of data in GF3. The package is designed as a suite of Fortran sub-routines and provides a complete software interface to GF3 enabling the user to handle GF3 in an automated fashion. It is also designed to be portable between different computer systems so as to be readily available to scientists and data centres alike.