

DATA BASES AND MANAGEMENT IN OCEANOGRAPHY

R.M.S. Bhargava and J.S. Sarupria

The Indian National Oceanographic Data Centre has been functioning ever since the National Institute of Oceanography (NIO) came into existence in 1966. At the time when the establishment of NIO was approved, the Data Centre was recognised as a national facility to acquire, process, store and disseminate oceanographic data pertaining to the Arabian Sea, Bay of Bengal, Laccadive Sea, Andaman & Nicobar Sea and the Indian Ocean. Since then, the data centre has become the nucleus of oceanographic data and information in the country. Today the centre holds data for about 20,000 stations in the Indian Seas collected during IIOE (1960-65) by *INS Darshak* (1973-74); *R.V. Gaveshani* (1976-89); *ORV Sagar Kanya* (1983-89); *FORV Sagar Sampada* (1985-86) and *Samudra Manthan* (1983-86) during various cruises.

Our country has made great strides in oceanographic research in the last two decades and has firmly established itself in the international scene. India is also providing information and data to the neighbouring countries for their oceanographic research programmes. The country has collected large amount of oceanographic data in and around the Indian Seas on various parameters by several organizations, particularly the National Institute of Oceanography, Indian Navy, Indian Council of Agriculture Research, Geological Survey of India, Oil and Natural Gas Commission Survey of India, India Meteorological Department, Fisheries Survey of India and Merchant Naval ships.

The main objective of the centre is to satisfy, as far as possible, the needs of marine scientists and user community in respect of oceanographic data. As is known, the collection of data is not only arduous but costly too. It has to be managed, in a way it can be used by anyone since it is national property. This task of management is performed by the Data Centre. The Data Centre ensures that the user gets good quality and accurate data which can generate good information.

In order to fulfil this objective, the centre engages itself in the following activities:

Acquisition of data & information

Various research organisations are collecting data and undertake oceanographic cruises. The centre keeps contact with them, approaches them by post or personally and collects the cruise information and actual data.

Processing

The data and information, thus acquired, are reformatted, checked for its quality and accuracy wherever possible and stored in the computer-aided storage system.

Storage

For storage in a computer, specific formats for each type of data are prepared. An Integrated File System is developed for storage of data and information.

Dissemination

The data and information are stored so that on user's demand, the data can easily be retrieved for dissemination. An Integrated Inventory Information System (IIIS) has been developed for handling user's enquiries and for dissemination.

The summary of station coverage by major oceanographic research vessels and their operations, is as follows:

Sr. No.	Name of the ship, ship code & owner	Year of operation	Total No. of oceanographic stations covered
1.	R.V.Gaveshani (GV) (NIO, CSIR)	1976 (January)	6061 (till 1989)
2.	O.R.V. Sagar Kanya (SK) (DOD)	1983 (July)	2240 (till 1989)
3.	R.V.Samudra Manthan (ST) (GSI, Govt.Deptt)	1983 (June)	1861 (till 1987)
4.	R.V.Samudra Kaustakha (KS), (GSI, Govt,Deptt)	1984 (September)	1507 (till 1987)
5.	Sagar Sampada (SS) (DOD)	1985 (February)	1482 (Aug. 1988)

DATA TYPES

The oceanographic data have been classified in the following types:

1. (A) Station Data

1. Physical & meteorological parameters - Temperature, salinity, CTD, BT, wind, nansen cast, sound velocity and humidity
2. Chemical parameters - Nutrients, CO², N, Ammonia pH, trace metals and hydrocarbons.
3. Biological parameters - Primary production, chlorophyll, zooplankton, fisheries, benthos, and microbiology.
4. Geological parameters - Geochemistry of sediment, and heavy metals.
5. Pollution - Heavy metals, pesticides, and petroleum hydrocarbon

(B) Time series data (Fixed station data)

1. Physics - Wave, currents, wind, temperature, and sound velocity.

(C) Underway data

1. Physics - XBT, and thermosalinograph.
2. Geophysics - Magnetics, gravity, bathymetry, and seismic.
3. Biology - Chlorophyll and plankton

(D) Remote Sensing data

- Wind fields, SST, waves, chlorophyll and turbidity.

DATA CATEGORIES

The data can also be divided into two categories viz. sensitive and non-sensitive.

Sensitive data

The sensitive data include BT, acoustic, trace metal and radioactive elements followed by seismic data. However, the most sensitive area is

the Continental Shelf and margin for which data are not disseminated without the concurrence of concerned authority.

Non-sensitive data

As far as non-sensitive data are concerned it is generally made available for exchange with minor reservations. The area beyond the EEZ is the international waters and therefore data from this area should be made available for exchange without any restrictions.

Oceanography is a multidisciplinary science, embracing physics, chemistry, biology, geology and geophysics and their related branches. Data have been collected by a few research Institutes and Universities on one or more of these aspects. The INODC holds the data collected by these organisations and also those collected during the International Indian Ocean Expedition. In addition to this, the centre has also acquired data from Institutes and Data Centres of other countries, particularly the NODC of USA and the World Data Centre-A in Washington. The following is the type and quantity of data managed by the INODC:

Discipline	Parameter	No.of stations for which data is available
Physical Oceanography	(Nansen cast) temp. sal., O, STD/CTD, MBT, XBT/DBT wave, currents, surface meteorology.	23431
Chemical Oceanography	(Nutrients), PO ₄ -P, NO ₃ -N, NO ₂ -N, SiO ₄ , pH, major elements of sea salt, trace metals & toxic elements.	18220
Biology	Chlorophyll productivity rate, zooplankton and benthos.	19321
Geology	Grain size, sediment composition, sediment load, foraminifera and geochemical	2120
Geophysics	Bathymetry, gravity, magnetics and seismics.	125113 lkm

As stated earlier, the area of interest is the Indian Ocean, particularly its two arms viz. the Bay of Bengal and the Arabian Sea. Again, more emphasis is laid on the data of the Exclusive Economic Zone of India which covers 2,015 m sq.km. along India's coastline of 7517 km. including those of Laccadive and Andaman and Nicobar coasts. The above mentioned stations are therefore spread over in these areas in the Indian Ocean.

The following data bases have been created and are currently available at the centre.

Station Level

1.1. **Inventory and parameter data base:** This data base includes the details of stations such as platform identity, station no., cruise no., date, time, position of the station in terms of latitude and longitude and number of observations made, as also the indexing of the record with INODC reference number which is the key in each record of the data base. This base contains the information on about 6061 stations of *R.V. Gaveshani*, 2100 stations of *ORV Sagar Kanya*, 1482 stations of *FORV Sagar Sampada* and 900 stations of *Samudra Manthan*.

Physical

1.2. **MBT data base :** This data base contains data on temperature at five meter interval, maximum depth upto 275 m. The temperature data were digitized from continuous BT data profile at 5 meters interval. These data have been marked with temperature inversion and intracruise variability. This data base presently contains 1398 station's data.

1.3. **XBT data base:** This is also a temperature data base, it contains the temperature data at five meter interval. These data have been digitized from graphic charts. The maximum depth of this data base is upto 750 m. This data base contains the temperature data for 931 stations.

1.4. **Nansen cast data base:** This contains the data on temperature and salinity at different depths. These data have been corrected by standard method for depth and temperature. This data base has been linked with a FORTRAN programme and physical oceanographic functional library which linearly interpolates the temperature and salinity at standard depths and then compute the sigma-t (density), geopotential anomaly, steric anomaly, thermosteric anomaly and sound velocity. The data base contains data for more than 1000 stations.

1.5 Surface meteorological data base: This data base contains data on SST, wind speed and direction, barometric pressure, air temperature, weather code, cloud type and amount, visibility, sea state etc. In this data base, some of the parameters have been coded with WMO standard codes. This data base presently contains data for 2037 stations.

Biological

1.6 Primary productivity and chlorophyll data base: This contains data on primary productivity, chlorophyll, pigment, POC, ATP at different depth upto the euphotic zone. This data base is also linked with a FORTRAN programme, which computes the column values for primary production, chlorophyll and ATP. This base presently contains data on 316 stations.

1.7 Zooplankton data base: This base contains data on nets, haul types, stratum depth, duration, biomass etc, along with inventory level information of the stations. This data base is having 1642 records from more than 1500 stations.

1.8 Zoobenthos data base: This base is having data on the bottom sampling and details along with the type of bottom texture, density and biomass of zoobenthos. This base also contains a station inventory level information and is having data for more than 1100 stations.

1.9 Nutrients and oxygen data base: This base contains information for oxygen, nitrate, nitrite, pH, ammonia, phosphate and silicate at different depths. This base contains data of 1000 stations collected on board *R.V. Gaveshani* and *ORV Sagar Kanya*.

2. Underway data base

2.1 Navigation inventory: This base contains information on latitude and longitude, along with date, time and file numbers. This base is having the information on 25, 542 lkm pertaining to 13 cruises.

2.2 Geophysical data base: This contains data on magnetics, bathymetry, gravity and seismic field. This format is compatible to the MGD-77 forming system of geophysical data. This base has the data for navigational (46,600 nm); magnetics (34,557 nm) bathymetry (46,400 nm) gravity (21,298 nm) and seismic (12,550 nautical miles).

3. **Time series inventory:** This data base has recently been initiated and contains data on currents, wave, and some mooring or anchored and station inventory, where observations have been collected at fixed time intervals.

INODC has developed the following systems and are used by the Data Centre for oceanographic data base creation and management.

A. Integrated File System (IFS)

An Integrated File System (IFS) was designed under a Pilot project from IOC, for management of Oceanographic Data and Information. This system has three type of hierarchial data base structure. The system basically provides the storage facility for all categories of oceanographic data. Each category has been divided into several record types viz. Level 1 and 2 for inventory and documentation information, Level 3 for actual data and Level 4 is for further details in some special types of data. The IFS format is presently used for inventory of station, underway and time series data. Data level formats are developed for storage of Nansen cast, BT, CTD, chemical, primary productivity, zooplankton and zoobenthos data.

B. Integrated Inventory Information System (IIIS)

Integrated Inventory Information System (IIIS) has been developed for the management of inventory level information. It has the modules namely station, parameters and documentation. IIIS also shows how a ROSCOP based inventory type data base structure can be evolved to meet the end user's requirement and to provide generalised inventory information services. It is supported with a retrieval system to handle all type of enquiries. It also helps in the preparation of cruise and station catalogues.

C. Oceanographic Station Inventory Information System (SIIS)

Oceanographic inventory information of data, collected during *RV Gaveshani* and *ORV Sagar Kanya* cruises, have been organised and structured using popular dBase III PLUS data base package available on DCMPS/386 microcomputer and working under MS-DOS environment. Using dBASE command level programming language, software modules are developed additionally, to support built-in-capabilities of dBASE III, for developing a very user friendly online system.

The system developed is capable of carrying out major functions such as (1) systematic storage and management of data, (2) selection of inventory information as per user's demand and (3) preparation of catalogues and reports.

Query handling system searches and sorts the required information, depending upon users specification (Geographic area, depth, season etc.) and stores it in a new data base. Further, the system generates TEXT FILE with INODC cruise nos. and INODC station nos. of searched data. Using Text File data, corresponding to INODC nos., data can be easily retrieved from large data files available on mainframe Computer.

D. Software for geophysical record management

Using dBASE III Plus software package, a data base has been developed for management of Geophysical data records. It contains data record of 48 Geophysical cruises undertaken by *R.V. Gaveshani*. Record includes magnetic, bathymetric, seismic line kms., names of scientists who collected data, data storage form and facility, area, time, year etc. for each cruise. Data base can be updated as and when data are received.

Online Inventory Information system for Geophysical line coverage data management

Geophysical line coverage data management system developed on DCM PS/386 contains navigational data details of 13 cruises of *RV Gaveshani*. Navigational data details include latitude, longitude, position, date, time, kind of data collected (magnetic, bathymetric or seismic etc.) along with cruise, fix and line identification numbers. Data base is open for constant updating. Software modules are developed for 1) cruise and fixwise easy appending, editing and viewing of data 2) printing cruise catalogue and fixes catalogue and 3) retrieving selected information.

E. Quality control for MBT data

A three level procedure has been developed for checking the quality of BT data. The criteria used in the procedure are: temperature distribution, watermass property and a standard temperature-climatology profile derived from historical data. In each level the quality flags are assigned for the inversion station, abnormal intracruise variability and climatologically abnormal stations. This procedure has enabled

to verify the quality of MBT data. Three flags have been incorporated in the data base for :

- i) temperature inversion;
- ii) abnormal intracruise variability in coastal area; and
- iii) abnormal intracruise variability in open ocean -

F. System for microbiological data storage

A system based on numerical coding has been developed to store the microbiological data pertaining to physiological groups, bacterial activities, microbial population etc. Under this system, a two digit code for identity of a record and another two digit code for identity of field for microbiological activities have been used. The system also identifies the types of sample used for analysis purpose.

G. Underway and time series data bases

New data bases have been initiated for underway and time series data. IFS formats have been designed and developed for storage and exchange of current meter data, geochemistry (Sedimentology) and geophysical data, including navigation, magnetic and gravity and bathymetric data. Geophysical merge data format is similar to MGD 77 format.

The Data Centre is engaged in compilation of data and bringing out the data products for user's convenience. The following products have so far been brought out.

1. **Inventory reports (total 10 nos):** Stations' and cruises' inventories have been published in nine reports. These reports contain inventory information of *R.V. Gaveshani* (1976-88) and *ORV Sagar Kanya* (1983-86). One inventory report pertains to 13 geophysical cruises and 6480 fixes of *R.V. Gaveshani* for the year 1983 to 1988.

2. **Data reports:** Five data reports have so far been prepared as follows:

MBT data	-	2
Surface meteorological data	-	1
Nansen cast data	-	1 and
Chemical data	-	1

All these data, referred to in these reports, were collected on board *R.V. Gaveshani* and *O.R.V. Sagar Kanya*.

3. **Oceanographic atlas:** Two volumes of oceanographic atlases have been published, one for EEZ and the other one for mechanical Bathythermograph data. One atlas for EEZ of India incorporates the physical, chemical and biological data collected during 1976-80 and shows the general annual distribution at three depths viz. 0, 50 and 100 m. The other atlas is devoted to MBT data and depicts all the available profiles in one degree square.

A list of data products is appended.

The data will increase many folds in the coming years, particularly with the development of remote sensing methods and are required to be effectively managed. It is in the national context that INODC has been recognised as the National Data Centre. INODC has also been made as a focal point for remote sensing data. However, INODC should have support from all data generating agencies and other Data Centres.

In order to make data management more effective, it is useful to survey and establish the national sources of data generation, if necessary even by scouting. The survey might be based on the following aspects: types of data, formats, frequency of measurement, time interval, type of environment and geographical area covered. The vast data generated are also required to be properly documented which would help users in getting required data.

Therefore, keeping in view the work done, developments and requirements, the following line of action is proposed, so that the Data Centre may serve useful purpose:

1. Establishment of a Centralized and independent Data Centre.
2. Acquisition of data from all resources.
3. Development of proper formats for storage
4. Checking, quality control and processing of data
5. Development of data bases.
6. Preparation of inventories, catalogues , reports etc.
7. Development of computerized storage and retrieval systems for prompt and effective dissemination of data and information.

8. Development of "Online" Communication (including LAN and WAN between the Centres and user organisations).
9. Creation of dedicated and trained manpower who must be treated on par with other scientific cadres. It may not be out of place to suggest here, that data management may be included as a topic in oceanographic courses.
10. Development of infrastructure facilities like computer, reprography etc. supported by sufficient grants for their maintenance and operation.

DATA PRODUCTS

Inventories

1. Inventory of stations and cruises of *R.V.Gaveshani*. Vol.1, (1976-80), Ref.No. 1101, Vol.1. (August 1986).
2. Inventory of stations and cruises of *R.V.Gaveshani*, Vol.2, (1981-85), Ref.No. 1102 (August 1986).
3. Inventory of stations and cruises of *O.R.V. Sagar Kanya*, (1983-85), Vol.3 Ref.No. 1103, (August 1986).
4. Inventory of cruises and stations of *FORV Sagar Sampada*. 1-20 cruises, (February 1985-September 1986).
5. Inventory of Geological Samples collected by *R.V.Gaveshani* cruises during 1976-85, vol. 4, Ref.No.1104, December 1986.
6. Report on Inventory of the Geophysical cruises and fixes of *R.V.Gaveshani* (1980-88), February 1989.
7. Inventory of cruises and stations of *R.V.Gaveshani* and *O.R.V Sagar Kanya* Ref.No.1105, Vol.5, (1986), March 1989.
8. Inventory of cruises and stations of *R.V.Gaveshani* and *O.R.V. Sagar Kanya*, Ref.No.1106, Vol.6 (1987), April 1989.
9. Inventory of cruises and stations of *R.V.Gaveshani* and *O.R.V.Sagar Kanya*, Ref.No.1107, Vol.7 (1988), May 1989.

Data Reports

1. Mechanical Bathythermograph Data Report (*R.V.Gaveshani* cruise 1976-86), Ref.No.1302, Vol.1, January 1989.
2. Mechanical Bathythermograph Data Report (*ORV Sagar Kanya* 1983-86), Ref.No.1302, Vol.2, January 1989.
3. Surface Meteorological Data Report of *R.V.Gaveshani* (1976- 85), Ref.No.1303, September 1989.
4. Nansen Cast Data Report of *R.V.Gaveshani* (1976-80), Ref.No.1304, November 1989.

Atlases

1. Oceanographic Atlas of the EEZ, Ref.No. 1301, (January 1988).
2. National Guidelines for the Distribution and Exchange of Samples and Data (June 1986).
3. Station Coverage in the Exclusive Economic Zone of India (December, 1986).
4. Integrated Inventory Information System (IIIS), (Jan, 1988).
5. Integrated File System (IFS) for Oceanographic Data Management, (March 1988).
6. A guide to formats for Marine Geophysical data storage and exchange (July 1988).
7. A guide to formats for Biological data storage and exchange (July 1988).
8. A quality control procedure for temperature data, Technical Report, 1989.

