

N 9 3 - 2 4 9 3 2

# **Data Availability and Data Archeology from the Former Soviet Union**

---

***Yuri Sychev, Nickolai Mikhailov***

## **1 Introduction**

Acquisition of data on the ocean is believed to start in 1872, when the Royal Navy ship "Challenger" performed oceanographic stations its round-world voyage (1872-1876). First oceanographic studies of the World Ocean refer to the 80s second half of the XIX century. During its round-world expedition "Vityaz" (1886-1889) headed by S.O.Markov, performed hydrological measurements in the Baltic Sea, Atlantic and Pacific Oceans. According to information available the regular expedition observations (prototype of future complex international program on the ocean research) started in the second half of 80s last century under the auspice of Kiev commission for exploration of German Seas. Systematic hydrological observations were organized by Hydrographic Department of Russia in 1876-1879 according to the program similar to the Kiev one and observations were regularly made by ships of custom service over the Russian area of the Baltic Sea.

The increasing demands in oceanographic data contributed to considerable progress in exploration of the World ocean during current century whole tendency to increase and become more significant has been observed for the last 30-40 years. Most probably various expeditions which were carried out during International Geophysical Year in different regions of the World Ocean are to be reference point in performing intensive oceanographic observations of Marine environment.

In the former USSR oceanographic observations are made by research and hydrographic vessels, commercial and fishery ships as well as and oil production platforms, coastal hydrometeorological station and other observing platforms. Oceanographic observations data, available from main sources of information on the ocean-research vessels - are also considered in the report.

According to RIHMI-WDC NODC information above 1348 national RVs have been functioning in the World Ocean in the course of all historic period, most part of which (above 950 ships) refer to RVs with small displacement having performed observations in former USSR seas: White Sea, Barents Sea, Baltic Sea, Black Sea, Caspian Sea, Sea of Japan, Sea of Okhotsk, Bering Sea and Arctic Seas. In the last decade general number of RVs carrying out expeditions research in the oceans and seas amounted to 250-300 vessels, small displacement included.

Expedition research efforts have been mainly realized under five former USSR agencies: State Committee for Hydrometeorology, Ministry of Fisheries, Academy

of Sciences, Main Department of Fleet Navigation and Oceanography and State Committee for People's Education. The averaged number of expeditions per year amounted to 600-650 with amount of information obtained about 30,0 Mbyte.

Observation data have been concentrated at agency centres of data acquisition and also have been sent via mail and telegraph channels to NODC with a view to creating the state oceanographic observations data holdings. The way of data transmission was regulated by some resolution of the former USSR Government and the ship owners were obliged to pass the materials obtained in expeditions to the National Oceanographic Data Centre, which was operating as a department of Research Institute of Hydrometeorological Information - World Data Centre (RIHMI-WDC).

Distribution of major functions on data processing and dissemination as well as data exchange problems between RIHMI-WDC NODC and the Centres have been coordinated at the level of the former USSR marine agencies. On the whole oceanographic data processing was rather study at the state level, being provided by interaction between data suppliers, Centres and RIHMI-WDC NODC for implementation of the scheme considered with respect to organization, methodology, software, technology and information aspects.

## **2 Data Collection and Accumulation**

Functioning of the multilevel system of oceanographic data processing provided for a high enough level of data collection completeness and preservation of data holdings of oceanographic observations. Oceanographic data collection in this form started in the second half of the sixties after RIHMI-WDC NODC had been established. Since 1969-1970, data for the expedition investigations, which make up not less than 85-90% for the oceans and foreign seas and about 70-80% of data for the former USSR/CIS seas, were systematically accumulated at the NODC. The major part of materials is submitted to the NODC in a standard form in line with the "Scheme of the scientific and technical report on r/v cruise" which includes the regulation for the presentation and transfer of data for the expedition investigations. In 1980-1982 a certain amount of data which are obtained from computers on board the r/v, became available at the NODC on magnetic tapes. So, the last version of the "Scheme of the scientific and technical report on r/v cruise" includes the descriptions of oceanographic data record formats.

To provide the completeness of data acquisition NODC performed early control of observation materials available from the ship-owners and centres and ones in 5-7 years checked availability of the archival historical data in RIHMI-WDC with reference to materials available with suppliers and agency data centres. Data of research cruises carried out outside the country, are submitted to RIHMI-WDC via the IOC/IODE system using WDC changed and are also resulted from operation of data acquisition centres (TROPEX, MEDALPEX) under different projects and bilateral exchange of observation data.

At the present time NODC data are collected, which were submitted by 954 Soviet and 1325 foreign ships whose general number equals to 1837.9 thousand of oceanographic stations and 395 thousand of bathythermograph soundings.

National data for oceans and foreign seas have been submitted by 6.9 thousand of RVs cruises for the period 1902-1990 and amount to about 40 % (with reference to oceanographic stations number) from the total number of data sets collected. Significant volume of oceanographic data has been collected over the former USSR sea areas. In such a case the main part of marine data has been obtained at fixed observation sites - cruise sites, stations of standard and secular sections. Participation in international cooperation on oceanographic data acquisition and processing made it possible to significantly enlarge oceanographic data holdings.

The composition of oceanographic observation types and parameters, which are included into data holding, is varied and depends on the time of observations, the vessel's equipment, the investigation program and other reasons. Before the sixties, the r/v carried out, mainly, hydrological and hydrochemical observations, bathythermagraphical observations, current observations and simultaneous meteorological observations. During the last 20 years the range of the observable parameters became substantially wider and currently it includes eleven types of observations. It is necessary to note, that hydrological and hydrochemical observations were carried out during the greater percent of cruises, collected at RIHMI-WDC NODC.

### **3 Archival Data Set Formation**

#### **3.1 Primary Data Processing and Archiving**

Archival data sets creation procedures which are realized, with software in OC EC computer and MS DOC environment provide for data assimilation, which are received from different sources and on different media (manuscripts, magnetic types, diskettes) , and for the preparation of a series of typical archival data sets on magnetic types which are oriented on long data storage.

Data quality control is the most laborious elements of this procedures. Data QC which are used to prepare archival data sets of the deep-sea hydrological and hydrochemical observations, are similar to QC GTSP procedures, which are the following:

- Impossible time
- mpossible coordinates
- Land - sea
- Impossible depth
- Impossible instrument code
- Impossible parameter values
- Profile 10° square profile envelope( for the Atlantic and Pacific ocean to the

- equator)
- Regional profile envelope (for seas of the former USSR)
- Increasing depth
- Constant profile
- Spike (top, within the profile, bottom)
- Kink
- Density Inversion

Archival data set creation provided the basis for the technical availability of accumulated historical observations data. In 1983-1985 the time lag from data production to their inclusion into archival data sets amounted to 1.0-1.5 years. During the last years the time lag increased to 2.0-2.5 years due to the functioning difficulties of oceanographic data collection and processing system. The final production of the scheme block 2 is the archival sets on magnetic tape and documentation with the archive description, which includes information about magnetic tape content, quality procedures, data record structure and other characteristics of archival set.

#### **4 Integrated Data Set Creation**

Typical archival sets are oriented on the long data storage and, thus, are characterized by a certain redundancy of record formats and by some inconvenience in their practical use. To increase data technical availability NODC carries out the project on the creation of the Integrated Global data set of the deep-sea observations set GLOBAL, which fulfills the current requirements for data availability and data handling convenience. The GLOBAL data set has the next composition:

- bathymetric data ( Nansen bottle or similar instruments)
- temperature, salinity, PH, O<sub>2</sub>, alkalinity, nitrites, nitrates and other hydrochemical characteristics;
- bathythermograph data (MBT,XBT)
- temperature;
- sounding sets (of CTD and STD type)
- temperature and salinity.

There are four main levels in the GLOBAL data set:

- level I/basic level (hydrological and hydrochemical data arranged by Marsden squares);
- level II (data interpolated for the standard depths);
- level III (climatic characteristics in Marsden squares);
- level IV (climatic characteristics in regular grid-point form).

At present the preparation of GLOBAL level I data is being finalized. I,III,IV level sets for the sea water temperature and salinity have been prepared for individual regions.

The current version of the GLOBAL set of the basic level includes the observation data, registered in the archival sets as of January, 1, 1991 (data from the beginning of observations to 1988 or to 1990 for individual regions). To obtain the characteristics of the presence and time-space distribution of oceanographic observation data, available at RIHMI-WDC NODC, metadata from the GLOBAL set of the basic level were processed for the Soviet cruise data and the USA. Time diagram (Fig. ) show data distribution over years of observations for the Atlantic Pacific and Indian Ocean, individually and together. It is seen from the diagram, that the main part of data falls at the last 30 years. In this case, the largest number of observations for this period was carried out for the Atlantic Ocean.

## 5 Data Archaeology

Retrieval of archival data available in manuscripts and not entered on technical media (data archaeology) is one of the main ways of increasing historical data quantity. Since entering of observational data on technical media is a rather time and money consuming work, it is no wonder that in some cases large quantities of these data can be found. In 1991 the NODC of RIHMI-WDC made great efforts to search and catalog these data. There are two main data sources: WDC archives with a large amount of information being held as hard copies and archives of national marine institutions not received by RIHMI-WDC for various reasons.

### 5.1 WDC-B Data

At present the WDC-B holds data from almost 3,000 cruises of 54 foreign countries containing about 192,000 hydrological stations. General distribution of the number of cruises for decades is shown in Fig. \_\_. It shows that half of the data obtained is referred to the last two decades. When moving from the current time to the beginning of the century the data amount decreases but remains significant since these historical data are of great importance. 74 cruises are referred to the first half of the century and 9 ones to the first two decades. Only 13 countries have shares exceeding 1% of the whole data amount. Their total share is about 85%, more than 60% being obtained from 4 countries: Japan - 27%, USA - 17.5%, Canada - 9.3%, Germany - 8%.

As it was expected the observations were mostly made in the North Atlantic (30%) and the North Pacific (39.6%). The South Atlantic and South Pacific account for 5.1% and 7.3% respectively, the Indian Ocean - about 4% and the Arctic Ocean and adjacent seas - about 2.5% of the observations made. Peculiarities of the time-space distribution of the WDC data not recorded on technical media are shown on the Fig. \_\_.

### 5.2 National Data

Six month retrieval made it possible for the NODC to obtain and catalog the data held in manuscripts from nearly 1,000 cruises (about 56,000 hydrological

stations) of Soviet research vessels. Over 60% of data were obtained from cruises of the 1985-1990 period which reveals a complicated final stage of the Soviet state started in 1985. The remaining 30% are time distributed in a natural fashion (Fig.\_\_). 35.5% of cruises available as archaeological data before 1985 apply to the seas adjacent to the CIS (Commonwealth of Independent States). Cruises in foreign seas and oceans display stronger attraction to the Northern Hemisphere than average one for the WDC data and that is quite natural (Fig.\_\_). As expert estimates show the further retrieval of national data similar to the one made will make it possible to use widely the data from some 500 - 800 cruises performed by marine institutions of the former USSR.

The large amount of the data held in manuscripts - about 56,000 national and 192,000 foreign hydrological stations - will require great efforts for their recording. It will take approximately 2,000-2,500 man\*month with a total cost of about 2.5 million roubles in late 1991 prices.

### **5.3 Data Availability Policy**

The current status of the former USSR states give rise to difficulties in the information exchange, oceanographic one in particular. At present the process of establishing of state structures and signing of vital interstate agreements is going on. Until the process is over it is too early to speak of the mutual commitments of the CIS members in the field of the World Ocean research and therefore in the field of data exchange and collection. Nevertheless Russia is ready to meet all the commitments of the former USSR in the World Ocean research. The appearance of the official documents confirming this readiness is a matter of time. Thus, observational data being a national archive as of December 1991 are considered as a national Russian archive. The policy of openness implemented in the recent time made it possible to involve some of the classified data in the international exchange. This applies to the data of the Navy. Now the Navy allowed the data from 350,000 MBT soundings for international exchange. Beginning at mid 1992 the Navy is going to pass 5,000 MBT data a year with a delay of a month. The possibility of making deep-sea observations unclassified is being discussed.

In the nearest future Russia is likely to reduce significantly a number of new observations for well known regions. So the retrieval of data from the observations already made, recording them on technical media and creating high quality data sets is one of the main tasks for the Russia data managers.