REGIONAL OCEANOGRAPHIC DATABASE AS A MODERN SEA RESEARCH TOOL

JACEK PISKOZUB, KRZYSZTOF PORAZIŃSKI, SŁAWOMIR SAGAN, WALDEMAR WALCZOWSKI, MARCIN WICHOROWSKI AND JACEK WYRWIŃSKI[1]

Institute of Oceanology,
Polish Academy of Sciences,
Powstancow Warszawy 55, 81-712 Sopot, Poland
[1]wyrwa@iopan.gda.pl

(Received 10 April 1999)

Abstract: Based on the experience acquired from working on the Regional Oceanographic Database at the IOPAS in Sopot a general view of database as a tool for supporting scientific research was presented. Article describes general structure, data flow and techniques of using the ROD.

Keywords: database, oceanography, tool, scientific research, BLOB

1. Introduction

This article describes a database created and developed at the Institute of Oceanology belonging to the Polish Academy of Science (PAS) in Sopot. The Regional Oceanographic Database (ROD) was developed as a tool aimed at aiding research.

2. ROD concepts and the range of application

The IOPAS has systematically stored measured data in the form of graphs and tables since the 1970's. Most of the source data collected during this period of time was not preserved. However, not much data was gathered, because it had to be processed manually.

When in the 1980's the digital data storing media (disks and magnetic tapes) were introduced, larger amounts of data were stored. At first, simple 8 bit personal computers were used. By using this type of storage for an increasing number of

tasks and by introducing better and more efficient measuring equipment, the quantity of collected data has increased.

Usually, the researcher had to store the data and was responsible for processing, control and its availability to the interested scientists.

After a while it became obvious, that it is necessary to create a unified archivization structure, which should provide easy access to information.

That is why the ROD was created. Its goal is common to all databases: quick and unrestricted access to all archives and to thematically connected data.
The database was developed gradually in the following stages:

—people assigned to control different groups of data were chosen;

— all archives, stored in different places and by different people, were catalogued;

— the storage media were modernized where needed;

— methods and equipment, used to register the archive data, were verified;

— quality of the stored data was estimated.

A database of information about the archives was created. It was a group of linked HTML documents accessible through the Internet. As a by–product, a database containing information about the measuring equipment used in the past and present was formed, as well as another database of methodology used in measuring the sea properties.

The system in its present form does not contain the actual data, but instead, it directs the interested person to the researcher, who conducts specific type of research and who is responsible for the data. However, a system was created that forces a certain procedure of handling the newly collected data and keeps the information up to date.

Further plans include the creation of access to the actual data using the present structure. It is to be achieved by merging with the SQL database, which operates on the archives.

3. Functional description of the database

The concept of the Regional Oceanographic Database is based on the information resources model accessible to the public in the WAN net. Communication with the user is provided through a graphic interface based on a HTTP protocol and on a group of HTML documents. The Internet is used as a transmission medium. This is the most common electronic information exchange system, furthermore it does not require any professional software. The configuration options of the WWW server allow the owner to choose the recipient of data and to manage access to the database.

The Regional Oceanographic Database was divided into five modules because of the different types of format of the data and different ways of its collecting and accessing:

Hydrologic database — contains the STD data gathered during research, and basic information used to analyze different measurements. Because of the large

volume of data in the database, it contains only an index and a way to contact a scientist managing the data. The database is based on a MS SQL 6.5 server.

Sea ecology database — contains complete information on the European Arctic fauna. The database was created as a group of HTML documents. The numerical data is available in the form of tables through the MS SQL 6.5server.

Physical database — contains physical data gathered during research. For bigger volumes of data it contains only an index and a way to contact a scientist managing the data. The database was created as a group of HTML documents.

Chemical database — contains geochemical, biochemical, radiochemical data and data describing probes of different components of the environment (organisms, residues, radionuclids) developed by the Chemical Department of the Institute of Oceanology PAS. The database was created as a group of HTML documents. The numerical data is available in the form of tables through the MS SQL 6.5 server.

Laboratory and measuring equipment database — contains information on sets of laboratory and measuring equipment. The database contains information on all parameters of the measuring equipment, which helps to determine the quality of measurements. The database is based on a MS SQL 6.5 server.

The access to all modules is identical and can be found on the main page of the Regional Oceanographic Database at *http://www.iopan.gda.pl/rbdo*.

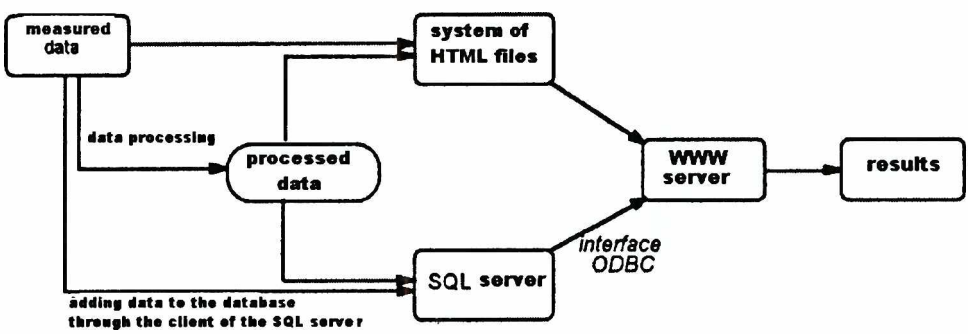


Figure 1. The data flow chart of RBDO

The Institute cannot provide the access to some large volumes data due to the lack of storage media. However, the information is stored using mass memory devices, and can be accessed through the person, who manages the data. The access to the index of information is unrestricted.

Because the data is supplied to the database and verified by several people, it was necessary to create procedures describing different stages of entering information into the database. The flow of information is presented in Figure 1.

The client receives data through the WWW server. At the present two servers are used: CERN – contains a group of HTML documents and MS IIS v.2.0 – cooperates through the ODBC interface with the database server MS SQL 6.5. The access to the database through the ODBC is as follows:

1. user's HTTP request is received by the Internet Information Server;
2. the Internet Information Server launches HTTPODBC.DLL (ISAPI) and delivers information contained in the HTTP request;
3. HTTPODBC.DLL reads information from the IDC (Internet Database Connector) file;
4. HTTPODBC.DLL executes the script contained in the IDC file and generates a request to the SQL server through the ODBC;

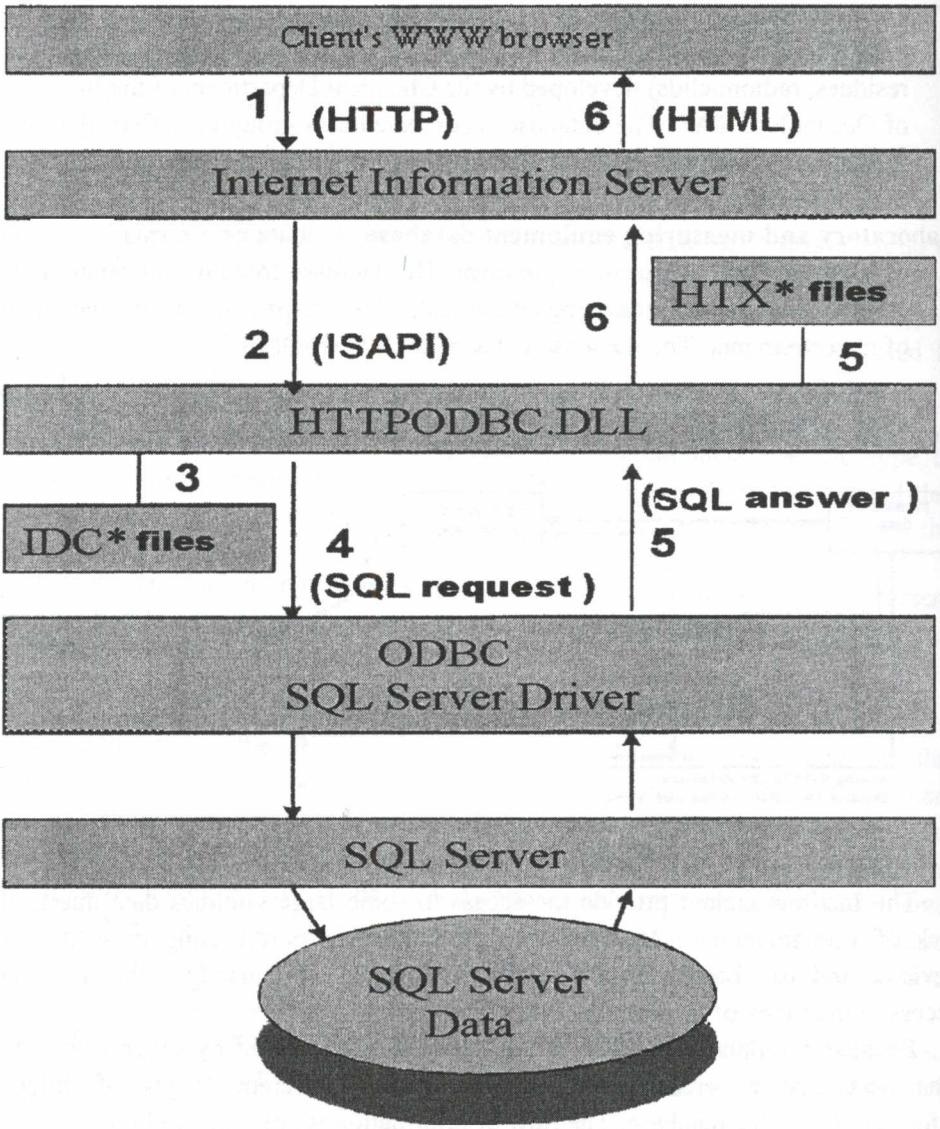


Figure 2. *The scheme of carrying out a query using ODBC*

5. the answer of the server, enhanced by a document scheme contained in the HTX (HTML extension) file, is sent to the Internet Information Server;
6. the User receives a HTML document containing the requested data from the database.

4. Principles of data access

The Institute of Oceanology PAS conducts basic oceanographic research. It means that the collected data should be accessible to people and institutions interested in the use of data for research. It does not mean, however, that the Institute and researchers, who are the authors of the data in the database give up their copyrights. The copyrights are published in the Polish Copyright Law (especially article 14) from the 4th of February 1994, as well as in the other international agreements.

In the usage of the database, the following rules concerning access to the data measured by IOPAS were formed:

— part of the database, which is accessible through the Internet, contains information on measuring devices used by the Institute of Oceanology; research trips and experiments during which data was collected; places where research was conducted (names and positions of geographic measuring stations) for each parameter and Internet addresses of the people responsible for the data;
— the rest of the database (raw and processed data) can be accessed by the Institute's employees and is accessible to researchers upon special request. Because the Institute of Oceanology and its employees reserve copyright to the collected data, even a marginal usage should contain the IOPAS in its references. Also, the research papers based mainly on IOPAS data should be co–written by its employees;
— part of the data can be placed in the unrestricted Internet database to exemplify collected data.

5. Using the Database as a research tool

The hydrological sea research consists mainly of the temperature and salinity measurements as a function of depth (pressure). The classic method, based on probing the water with batometers, measuring the salinity using the chemical method and measuring the temperature with reversible thermometers supplied a couple hundreds of location measurements from each expedition, where temperature and salinity were measured. Based on that data, the density of the sea water was calculated, vertical profiles (transsection) and horizontal distribution of salinity and temperature were graphed.

The last decade brought a radical increase of data volume collected, because of electronic probing devices. The CTD probes work at a frequency of 8 Hz (Guildline) or 24 Hz (Seabird) and they measure temperature, pressure and conductivity of sea water. The next stage is the usage of microstructural probes, which measure hydrophysical fields and transmit the information at the frequency of 1000 Hz.

Our computers can store and process data from one research expedition. For example one arctic expedition brings about 110 MB of raw data from vertical probing. During the expedition all source data is stored on hard disks and backed up on floppy disks. A file contains not only the CTD data, but also the date, the time, the number of expedition, the name of the station and the ship's position transferred automatically from the GPS. We plan automatic storage of meteorological data (the direction and strength of wind, the air temperature, the atmospheric pressure) in every file. At the Institute the source data is transferred to optical disks and stored in the Database. Even during the expedition data is processed and averaged (every 0.5, 1 or 5 meters) and the necessary oceanographic parameters are calculated.

The next step in the sea research is a transition from the CTD measurements to measurements using a hauled scanning probe. During the movement of the ship, the probe enclosed in a special container is raised and lowered from the surface to the bottom of the sea (Baltic Sea) or to the depth of 300 meters (Arctic). The whole cycle is repeated every 200–500 m. The data is stored and the whole profile is shown on a monitor in real time. It gives us an entirely new picture of the sea, a chance to discover new structures and processes, and also a huge increase in the amount of data which has to be collected, processed and interpreted. One hour of probing results in about 1MB of raw data, each short Baltic expedition brings dozens of hours of measurements.

All raw and processed hydrological data is stored in the Institute's Database. There is also a bulletin published containing the expedition maps and tables with the probing places. Apart from the hydrological data importance on its own, it is also a basis for all other oceanographic measurements when combined with meteorological data.

The data collected and processed in this manner is in a raw form, because it consists of unprocessed single probings. Therefore an attempt is made to group all data by topic. A CD containing all hydrological data concerning the Arctic Front was being systematized throughout many years and has been recently published. It presently contains:

— raw data;
— probing data averaged every 0.5, 1 and 5 m containing pressure, potential temperature, salinity, potential density, anomalies in the specific volume of the sea water, stability and frequency of the Visl–Brundt vibrations;
— horizontal distribution of temperature, salinity, density of water at some levels, calculated speed of geostrophical currents;
— vertical distribution — profiles of temperature, salinity, density of sea water, geostrophical component of currents.

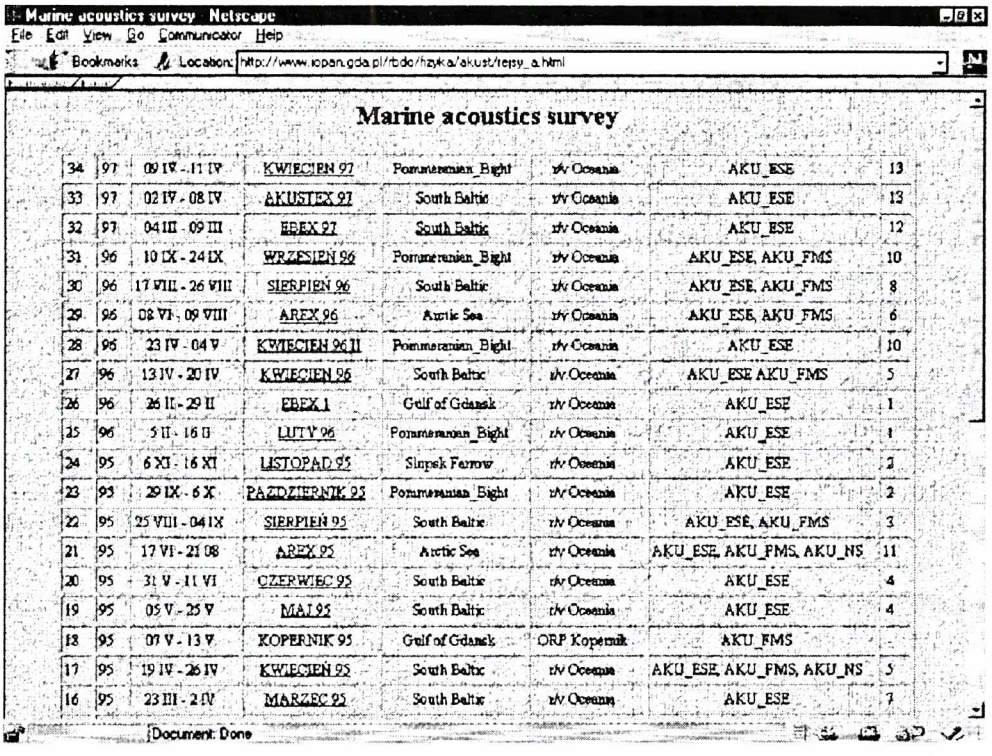
The processed profiles and maps are stored in a graphical form and in the surfer files (*.srf) format, together with all elements necessary for editing (*.grd, *.dat, *.bln files). The user may edit and modify the files. The publishing of next CDs, concerning the measurements in the Confluence Area of the Norwegian Sea,

Barents Sea (the Institute has a great collection of data from that area) and measurements with a hauled and stationary probe in the Slupsk Furrow, is planned. The CD will also contain meteorological data and it will be available in the Institute's network.

6. Problems with giving access to large volume files

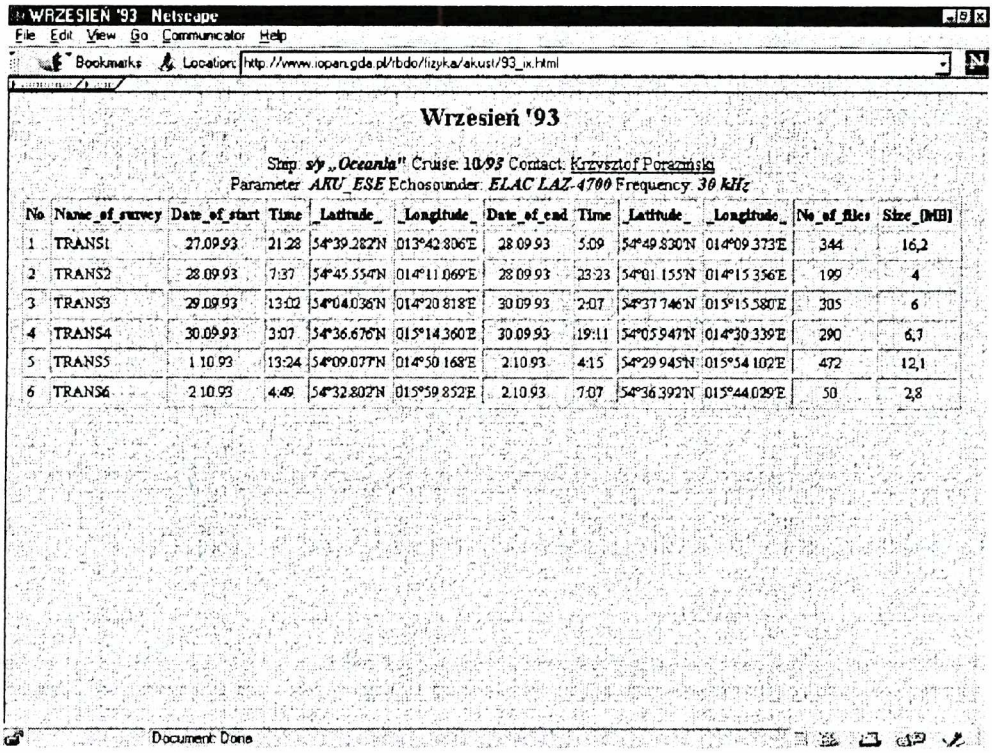
During expeditions, the Sea Acoustics Laboratory of the Institute of Oceanology PAS studies the migrations of sea organisms. The uniqueness of this measurement is that the acoustic signal coming from the sea has to be continually registered. The signal from the echo–probe is sampled in an analog–digital converter with a frequency of 3–5 kHz, and takes up 50–100 kB in a single file on a magnetic media. The amount of data depends on the depth of the measurement. The signals are stored in 30 s intervals, therefore a few miles of measuring results in dozens MB of data. There are two problems: storing the data and making them accessible in the network. The first one can be solved by storing the data on compact discs. At present we have 14 CD's with the data from the years 1986 to 1997.

The second problem — making that data accessible — is much more complicated. The data stored on the CD's is raw and can be used in order to get other results. That data is usually not useful for people outside of the Acoustics Laboratory, because they do not have any calibration data, descriptions of measurements, etc. These measurements are done for the particular research,



34	97	09 IV - 11 IV	KWIECIEN 97	Pomeranian_Bight	r/v Oceania	AKU_ESE	13
33	97	02 IV - 08 IV	AKUSTEX 97	South Baltic	r/v Oceania	AKU_ESE	13
32	97	04 III - 09 III	EBEX 97	South Baltic	r/v Oceania	AKU_ESE	12
31	96	10 IX - 24 IX	WRZESIEN 96	Pomeranian_Bight	r/v Oceania	AKU_ESE, AKU_FMS	10
30	96	17 VIII - 26 VIII	SIERPIEN 96	South Baltic	r/v Oceania	AKU_ESE, AKU_FMS	8
29	96	02 VI - 09 VIII	AREX 96	Arctic Sea	r/v Oceania	AKU_ESE, AKU_FMS	6
28	96	23 IV - 04 V	KWIECIEN 96 II	Pomeranian_Bight	r/v Oceania	AKU_ESE	10
27	96	13 IV - 20 IV	KWIECIEN 96	South Baltic	r/v Oceania	AKU_ESE AKU_FMS	5
26	96	26 II - 29 II	EBEX 1	Gulf of Gdansk	r/v Oceania	AKU_ESE	1
25	96	5 II - 16 II	LUTY 96	Pomeranian_Bight	r/v Oceania	AKU_ESE	1
24	95	6 XI - 16 XI	LISTOPAD 95	Slupsk Farrow	r/v Oceania	AKU_ESE	2
23	95	29 IX - 6 X	PAZDZIERNIK 95	Pomeranian_Bight	r/v Oceania	AKU_ESE	2
22	95	25 VIII - 04 IX	SIERPIEN 95	South Baltic	r/v Oceania	AKU_ESE, AKU_FMS	3
21	95	17 VI - 21 08	AREX 95	Arctic Sea	r/v Oceania	AKU_ESE, AKU_FMS, AKU_NS	11
20	95	31 V - 11 VI	CZERWIEC 95	South Baltic	r/v Oceania	AKU_ESE	4
19	95	05 V - 25 V	MAJ 95	South Baltic	r/v Oceania	AKU_ESE	4
18	95	07 V - 13 V	KOPERNIK 95	Gulf of Gdansk	ORP Kopernik	AKU_FMS	
17	95	19 IV - 26 IV	KWIECIEN 95	South Baltic	r/v Oceania	AKU_ESE, AKU_FMS, AKU_NS	5
16	95	23 III - 2 IV	MARZEC 95	South Baltic	r/v Oceania	AKU_ESE	7

Figure 3. *An example of a screen containing a list of expeditions*



WRZESIEŃ '93 Netscape

File Edit View Go Communicator Help

Bookmarks Location: http://www.iopan.gda.pl/tbdo/lizyka/akust/93_ix.html

Wrzesień '93

Ship: *s/y „Oceania"* Cruise: *10/93* Contact: Krzysztof Poraziński
Parameter: *AKU_ESE* Echosounder: *ELAC LAZ 4700* Frequency: *30 kHz*

No	Name_of_survey	Date_of_start	Time	_Latitude_	_Longitude_	Date_of_end	Time	_Latitude_	_Longitude_	No_of_files	Size_[MB]
1	TRANS1	27.09.93	21:28	54°39.282'N	013°42.806'E	28.09.93	5:09	54°49.830'N	014°09.373'E	344	16,2
2	TRANS2	28.09.93	7:37	54°45.554'N	014°11.069'E	28.09.93	23:23	54°01.155'N	014°15.356'E	199	4
3	TRANS3	29.09.93	13:02	54°04.036'N	014°20.818'E	30.09.93	2:07	54°37.746'N	015°15.580'E	305	6
4	TRANS4	30.09.93	3:07	54°36.676'N	015°14.360'E	30.09.93	19:11	54°05.947'N	014°30.339'E	290	6,7
5	TRANS5	1.10.93	13:24	54°09.077'N	014°50.168'E	2.10.93	4:15	54°29.945'N	015°54.102'E	472	12,1
6	TRANS6	2.10.93	4:49	54°32.802'N	015°59.852'E	2.10.93	7:07	54°36.392'N	015°44.029'E	50	2,8

Document: Done

Figure 4. *An example of a screen describing one of the expeditions*

therefore converting data to standard forms (for example calculating the reverse dispersion component — Sv) is unnecessary and time–consuming from our point of view. The data is left in its raw form and converted only in order to get necessary results. A separate problem which arises is how to make all the data accessible at once (a couple of GB) and the network's capacity.

Because of these two problems we can only give access to the metainformation — catalogs of gathered data. There is a list of all expeditions, the name of the ship and information on which CD the data is stored. The data from each expedition is divided into separate measurements. Every particular information concerning the research contains the time (date, time — GMT) and the place of the measurement (latitude and longitude).

There are also additional information such as: the type of the equipment used, the frequency of the signal, the type of the A/C converter used (this describes the format of data in the file), other information connected with the subject, and also the personal information (esp. e–mail) of the person supervising the data. This type of access seems to be more reasonable when questions about the data are individual. They can be forwarded to competent people, who supervise or work on the specific kind of data.

At present we are working on presenting the data in a graphic form and on showing the measurement places on maps.

Our concept has already brought contacts with some foreign research centers.

7. Conclusions

Due to the present technology we can easily create scientific databases and place them in large networks. These databases can be used as an aid in research. The access to the database using common WWW browsers makes it easy to use, popularizes science and helps exchange information between the different fields of science. The costs of maintaining and developing the scientific databases are partly recompensed by the systematization of collected data. It helps to raise the quality of published research papers, and the importance of Polish science in the world. The development of the Regional Oceanographic Database was possible thanks to grants and the KBN policy concerning financial aid for the Center of High Power Computers, the Academic Computer Network TASK, the LAN Network of the Institute of Oceanology PAS and grants for developing the ROD.

References

[1] MS Internet Information Server v.3.0 documentation files