

PROBLEMS WITH RENDERING THE INFORMATION ACCESSIBLE AND PRESENTING THE INFORMATION IN PUBLICLY ACCESSIBLE DATABASES

MARCIN WICHOROWSKI, JAN MARCIN WĘSŁAWSKI
AND JACEK WYRWIŃSKI

*Institute of Oceanology,
Polish Academy of Sciences,
Powstancow Warszawy 55, 81–712 Sopot, Poland
wichor@iopan.gda.pl*

(Received 14 September 1999)

Abstract: Based on the experience acquired from working on the Regional Oceanographic Database (ROD) at the IOPAS in Sopot and from cooperation with similar scientific facilities, a general review of problems with the security of data ownership, copyrights, rendering the scientific information accessible, propagating it and differences in the access to data in European Union and Poland was presented. A system for accessing processed data, educational data and source information was proposed.

Keywords: data acquisition, databases, data collections, data storage, inventories, legislation, oceanographic data

1. Types of data in environmental databases

Institutions carrying out environmental research base progressively on archived data. It results from the requirements set for the quality of research (much statistically significant information) as well as from the technological progress allowing to gather a lot of information in a short period of time. In the past it was possible to publish scientific articles based on single measurements (for instance, series of manual measurements of salinity with Nansen cylinders). Though today, the minimum requirement is to present a continuous series of thousands of measurements obtained from electronic probes. That is why a typical phenomenon for numerous scientific institutions is an excess of gathered data, as well as delays in processing and making it accessible. Gathering, storing and giving access to data is becoming an important element of the scientific policy of many institutions. Because the technological leap occurred in Poland only recently, most of the significant data over ten years old was lost beyond recall or its retrieval would be too expensive.

This problem concerns especially the countries of the former Soviet Union, where many international “rescue programs” were established in order to retrieve and archive in a modern way oceanographical and meteorological data which are important for the global models of climate.

Putting in, storing and using data in bases one might encounter legal restrictions which are considerable. Since we focus on the public environmental database, we will not encounter legal difficulties of personal or industrial databases, etc. Categories of information which are usually introduced to the environmental databases are as follows:

— *raw results of measurements* — the owner of such data is an institution which provided means to gather it. The people who are carrying out measurements, putting them into bases and conducting the quality control of material do not have rights to such set for such data do not possess “author intellectual contribution”. Usually, very few specialists who are employed by the institution running databases are allowed to use such data. The way to gather such data should make them durable. The principle should be to use several conveyors for data storage (for example, server, server replication and CD archives);

— *basic processed data* — the owner of data is still the institution, though the copyrights are now specified. They are possessed by a person, who edited the material and thanks to his competence changed the character of data from the raw sequences of numbers gained from the analyzer to certain environmental quantities, i.e. the concentration of chlorophyll in water, the level of storm accumulation etc. The sets of processed data undergo calibration as well as control of data and they carry the responsibility for the integrity of a result. The person who processes data is also responsible for the quality of information given by him and to some level for the results of its use. The processed data may not be used by outer users unless the institution–owner does let do so. In this case, it is necessary to provide information about the author. Unpublished processed data is the most valuable material of every single database, for it is the scientific information ready to be used, which can be introduced to the original publication. The protection of this part of data set must be very clear and effective. A typical example of a scientific piracy is publishing an article based on the original data from the base of the institution, which [the institution] didn’t permit to use stored information in such a way. Editor’s offices of magazines, which publish original works, are not able to state whether the author of a work is indeed the disposer of data and they usually require to sign a declaration by the author which brings on him the entire responsibility. In this way, claims of the institution, which was robbed of data, must be pointed against a certain individual and practically they are hard to be carried out;

— *archived data (from earlier published works, publications etc.)* — in the scientific environment, such data is considered to be commonly accessible, which means that it practically does not have the owner, who might limit the access to the data. It is necessary though to quote the author along with giving the name of the institution, which is the owner of the data. The introduction of earlier published data to the base is, however, restricted by the laws of the very first publisher, for

instance, most of the magazines reserve all rights to publish pictures, tables, photographs all by themselves and their use requires a written permission of the certain publisher;

— *data of compilation character* — synthesized from many sources, are placed in databases as the commonly accessible material, the copyrights are reserved for the author of the compilation, and he is also responsible for providing the sources which he has been using. Considering the compilations, which usually are represented by diagrams, it is worth to remember that the source data of over 50 years does not require to recognize copyrights, they can be copied utterly. Because of good manners though, one should provide such a source. Newer source materials might have various sorts of restrictions, though after processing they become the original work of the author of compilation. In case of managing a scientific–educational database there is no basis for legal claims, however this problem arises if the data contained in the base is sold or used for commercial purposes (i.e. in marketing);

— *metainformation* — directories, tables of contents or the so-called database maps; they are considered to be objects without an individual intellectual contribution (features of an authorship work) and can be used without restrictions when located in a publicly accessible database.

Programs and their applications used in databases can also come under the law. According to the definition of a computer program contained in the Polish Norm, it is “a set of instructions designated to be used in a computer” and software is a broader concept including documentation, conveyors and additional materials (i.e. illustrations). According to the 3rd article of the Copyright Law (1994) databases are protected by copyright, even if they contain unprotected materials, yet their selection, arrangement or comparison have a creative character. The 12th article of the Law states “unless it is concerned by the work contract, the employer, whose employee created the object as a result of his duties in work, gains the authorship and property laws”. The 14th article states “... a scientific institution has the right to be the first to publish the work of its employee, who created the object as a result of his duties in work, and the author is entitled to a compensation”. Therefore, the work created by the employee is regulated by the law and any use of the work by the employer should be additionally compensated. According to the same article the employer may use the scientific content of the work without any farther compensation and may give access to the work to third parties. The legislation of the countries belonging to the European Union contains a Directive no. 96/9 describing the legal protection of databases. The main point of the directive states: “the use of some elements of a database is an unsubstantial exploitation in contrast to the use of a major part of the database. The base can be accessed in three instances:

1. for private purposes — in case of non–electronic databases;
2. for non–profit purposes, such as education or research; the indication of the source is necessary;
3. for public safety or if any legal proceedings require it.

2. The ways of rendering the data accessible

Because of a wide variety of data stored in environmental databases (in respect of type and size), it is characteristic of them to vary in access and presentation of data, as well as to have a wide range of potential recipients. The typical users of environmental databases are: scientific and educational institutions, metainformation banks, industrial companies, consulting firms, administration, public services and other institutions.

In the process of rendering the data accessible it is necessary to point out the producer (a person, who gained and elaborated the data; it is probable that the producer holds the copyrights to the data, if the raw data had to be processed in a creative way), owner (a person, who catalogued the data, elaborated the access, storage and distribution of data; the way of drawing up and putting together the data is copyright), disposer (a person, who supervises the data in everyday use) and the user (a person using the data).

The way of rendering the data accessible is not only a technical problem. During the process of giving the access to data one must consider the copyrights of producers and owners of data as well as his. One should also take under consideration not only uncontrolled use of our data, but also unauthorized access to data for which one is responsible; in other words, one has to secure data entrusted to us against unauthorized access.

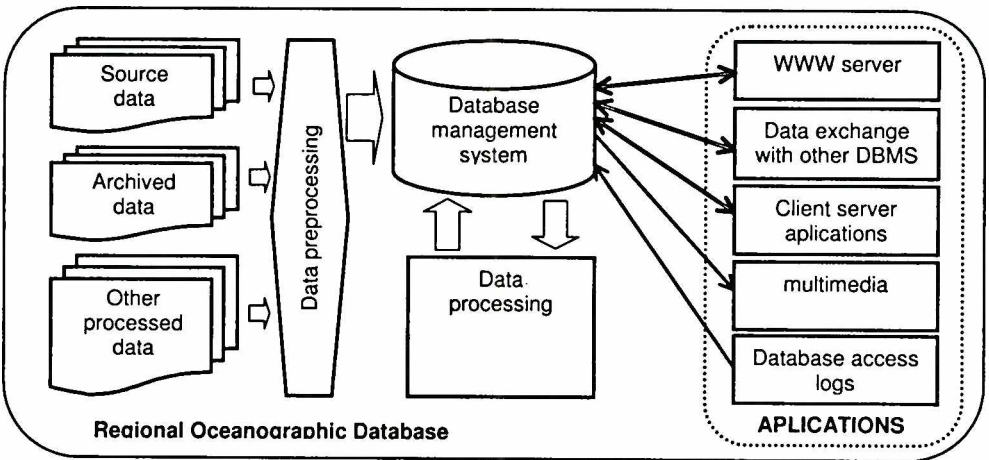
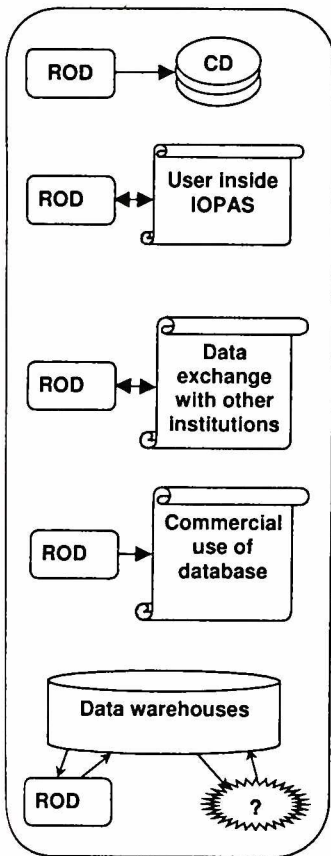


Figure 1. Data flow in ROD

Figure 1 presents a general diagram of information flow in ROD, which also considers the process of preparing data gained from measuring instruments, from archives and other system providing data (i.e. database management systems at other institutions, data banks). Only the group of employees entering data into ROD using the client-server application is granted the direct access to the database. The rest of the employees and outer users access the database through a WWW server. The data is also provided to ROD by a system of access registration. A part of the base is an archive of data on CDs. Some of them are accessible on-line.

The end users gain information from the database through different types of interfaces (HTTP, ISAPI, CGI, JAVA, FTP, SMTP, systems of files) which can be accessed by common applications (i.e. WWW browsers) or by applications created for ROD. The access to data contained in the base is free. Large volumes of data (acoustic data, CTD) require preparation of specific sets of information from archived raw measured data and they are accessible after consulting the database operator. The users are allowed to access an index of data. Such a solution allows the control over the usage of data. Furthermore, the base contains graphic data (pictures, maps) which are copyright. The copyright also covers all types of multimedia publications that base on sets of data contained in ROD.

The diagram below shows the ways of rendering information in a database accessible, and also shows law conditions concerning these ways.



Giving access to information contained in the base on CDs and other multimedia conveyors requires to place all ©. Such a publication is protected by copyrights.

Users inside the institution can use the data at all times.

The exchange of data with other institutions gives these institutions unrestricted access to the data within a certain constricted range, excluding the access to data gained from other institutions, which cannot be rendered accessible without a license.

Rendering data accessible to remaining institutions, including commercial institutions, requires estimating the value of data and billing. Sale of data gained from other institutions requires obtaining certain licenses.

A specific example of giving access to data to a user about whom we have incomplete information. It usually concerns data, which is accessible through the medium of third parties. It is impossible to control the proper use of data obtained by him.

3. Examples of rendering data accessible in EU

SMHI — Swedish Meteorological and Hydrological Institute.

SMHI is the disposer of hydrological data of the Baltex program (BHDC) as well as of Swedish Ocean Archive (SHARK) and National Oceanographic Data Center databases. In both bases there are gathered data obtained by SMHI and other institutions participating in monitoring programs and Baltex program. Data

index is commonly accessible. SMHI data can be obtained after declaring a form specifying the range of data. Additionally, BHDC data is accessible after obtaining a form from the member of Baltex program, registration of a user in BHDC and signing a license contract. In every case a fee, the way of preparing data and restrictions, which concern the control of data, are negotiated.

EDMED — European Directory of Marine Environmental Datasets.

The base of metainformation established in 1991 under the auspices of the MAST program. The aim of the database is a stimulation of efficient use of existing data by scientists, administration and other potential users. The base is a tool for the identification of useful sets of data located in the European laboratories, institutes, government agencies, data banks and private firms. At present the base contains 2180 sets of data from 453 institutions. The disposer of the base is the British Oceanographic Data Center. The database can be accessed in the Internet at <http://www.nbi.ac.uk/bodc/edmed.html>. The disposer warns that the resources are catalogued independently of their copyrights and their use has to be arranged between the user and owner of the data.

4. Conclusions

1. Metainformation bases are the most common way of giving access to data by institutions. The user can be granted access to useful information only after his identification. It secures the institutions from uncontrolled use of their data.
2. Processed and unpublished data are the most valuable material of every base. The protection of this part of the base has to be very clear and efficacious.
3. There is no uniform standard of giving access and managing a database in Poland or in Europe. It means that the user has to recognize the principles of using the resources of each database.
4. The databases bought as a whole, usually containing processed data, are accessible without restrictions to selected groups or users with the requirement of stating the source of used data (the requirement of the producer and owner).
5. The economic activity of a unit allows it to profit from the law barring unfair competition.

References

- [1] J. Barta and R. Markiewicz, *Database protection in the copyright system — the present day situation and perspectives*, Mat. of the Conf. INFOBAZY'97 — Databases for science, 23-25 June 1997, Gdansk, 1997
- [2] Z. Okoń, *Computer programme and software*, <http://prawo.fnet.pl/art/prkomp/003/indexdwn.html>
- [3] A. Karpowicz, *Copyright guide*, PWN, Warsaw 1994
- [4] *Advertising without risk*, T. Drozdowska (ed.), Verlag Dashofer, Warsaw 1999
- [5] Dziennik Ustaw RP no. 24 (1994) *Copyright Law*