NON-SPATIAL ENVIRONMENTAL DATA

K. Voigt
GSF - National Research Center for Environment and Health, Institute of Biomathematics and Biometry, Neuherberg, Germany

Keywords: non-spatial environmental data, environmental database, metadatabase, environmental information system, meta-information system.

Contents

Summary

The enormous increase in environmental information implies a rise in online databases, CD-ROMs and Internet resources in these fields. A clear management strategy is required to handle this variety of data as well as data-sources. In order to understand the subject of environmental data and databases, different ways of classification methods are introduced. Classification can be fulfilled according to storage media (online database, CD-ROM databases, Internet resources). Another classification method is reached by types of databases. For the given types examples in the medium Internet are given and explained. The DAIN—Metadatabase of Internet Resources for Environmental Chemicals—is introduced, and its facilities (user statistics, search form)

are explained. As a specialty, Environmental Information Systems (EIS) and Environmental Meta Information Systems are described. Approaches in Europe and in Germany are explained. The urgent need for applying data mining concepts to manage the growing data on the Internet with respect to environmental problems is stressed in the last section of the paper.

1. Proliferation of Environmental Information

In environmental sciences we meet an enormous output of and demand for data. Since there is no indication that the increase in information and data in the complex fields of environmental sciences will slow down within the foreseeable future, we shall have to cope with a growing flood of environmental information. A scientific approach is urgently needed to deal with this information abundance. The enormous increase in environmental information implies a rise in online databases, CD-ROMs and Internet resources in these fields. The Gale Directory of Databases is the largest collection of commercially available databases worldwide. It comprises in its latest issue 11 339 databases in all disciplines. The Gale Directory is divided into the following subject classes: business, health/life sciences, humanities, law, multidisciplinary, news/general, science/technology/engineering, and social sciences. Science and technology comprise 17% and health and life sciences 11%. This means that approximately 28% of the listed databases contain environmental information which corresponds to around 3200 commercial databases.

In contrast to the situation in the Seventies, Eighties and early Nineties, we now have the medium of the Internet available which allows us to get access to the environmental data and information. With the estimated number of 195 million Internet users in August 1999, many people have the tool to use these data-sources. The problem arises where to find the information wanted. Answering this question, a management strategy to handle this variety of data-sources is needed.

2. Classification of Environmental Databases

Several ways exist to describe environmental databases. First, databases are categorized by media type. These media types are commonly accepted as online databases, CD-ROMs and Internet resources. Second, databases belonging to a medium are grouped according to their subject into ‘general interest databases,’ ‘scientific and technical databases’ and ‘business and regulatory databases.’ The third way of classification is the division of environmental databases into database types. The mentioned ways are only the common and obvious ones. Due to the very diverse subject matter of environmental sciences, many other possibilities exist to classify these databases. For example, environmental databases can be specialized on a specific type of environmental information like ecotoxicity data on chemical substances, or concentration data in environmental media. On the other hand, several databases focus on a specific use of chemical substances, like databases on pesticides or on solvents. For further reading a paper entitled ‘Environmental Information Databases’ published in the *Encyclopedia of Computational Chemistry* is strongly recommended. A major textbook on the topic of environmental information systems was published by Günther in 1998.

In this paper the focus lies on non-spatial environmental information, databases and information systems. However, the overlap of spatial and non-spatial data frequently occurs especially in environmental information systems.

2.1. Definition of Environmental Data, Information, Database and Information System

Environmental data are technical, spatial and temporal data for the environmental media of air, water and soil. They pertain to questions of waste, noise, dangerous substances, fauna and flora, landscape, nature and species conservancy. With the help of analysis and interpretation of those data environmental information can be created.

An environmental database is a particular type of database that stores mainly environmental data. According to environmental informatics experts, a database can be called an 'environmental database' if the following three conditions are fulfilled:

- the majority of data are environmental data;
- a database system is used for the storage of these data;
- the database is established as the basis for environmental uses and inquiries.

Applying this definition to environmental information on chemical substances, the following conditions have to be included:

- the majority of data are chemical data and information;
- the database is established as the basis for chemical questions and uses.

Environmental Information Systems (EIS), used as a technological-organizational infrastructure to provide environmental information from special fields in different environmental databases, are often geographically localized. Therefore, Environmental Information Systems are sometimes considered to be extended Geographical Information Systems (GIS). However, EIS also hold thematic data (i.e. environmental facts such as measurement values on chemical substance attributes, environmental documents such as text data on literature, research projects, laws and regulation) or data with temporal reference (e.g. land use alterations of restoration areas, or seasonal fluctuations in dangerous substance measurements). GIS can neither cope with the problem of handling thematic data adequately nor of managing time series data.

These definitions are taken out of a textbook entitled *Umweltinformatik*— 'Environmental Informatics' written by Bernd Page and Lorenz Hilty.

2.2. Media for Environmental Databases

As mentioned above, the media: online, CD-ROM and Internet, are used.

2.2.1. Online Databases and CD-ROMs

The term 'online database' has been established for databases offered to the public by specialized database providers called hosts, using national and international public data

networks or the Internet. CD-ROMs (Compact Disk Read Only Memory) have been introduced approximately 15 years ago and are still, in our context, essentially a read-only medium.

2.2.2. Internet Resources

The medium 'Internet' is a medium created by hundreds of developments that mark our movement toward the new millennium. Every day the World Wide Web grows by roughly a million electronic pages adding to the thousands of millions already online. For the first time in history, millions of people have virtually instant access from their homes and offices to creative output of a significant—and growing—fraction of the planet's population. The question nowadays is not whether to exploit the electronic information superhighway, but how.

In the content described in this paper environmental databases on the Internet are resources which belong to the so-called 'free' Internet, i.e. they are available to the Internet user free of charge.

Nowadays, most databases—i.e. online databases as well as contents of CD-ROMs—can be retrieved via the medium Internet. These days we still distinguish among online databases, databases on CD-ROMs and databases on the (free) Internet.

2.3. Subject Classes for Environmental Databases

The topic of the classifying environmental databases according to their subject was treated in a special issue of the journal *Database*. This booklet was published under the title "Environment Online, the Complete Environmental Series from *Database* magazine." This issue includes three different articles which cover the following three different subject classes:

- general interest databases
- scientific and technical databases
- business and regulatory information.

2.3.1. Classification Types for Environmental Databases

This is the most comprehensive way to categorize environmental databases. In the specialized literature of information sciences, databases are divided into several types according to their database structure and their information types. On the first level fact-based, text-based and integrated databases are distinguished. The term 'fact-based databases' is not clearly defined. It is commonly accepted that this database type is comprised of facts. However, bibliographic references are often given in factual databases, although they are of secondary importance. The term 'fact' in this context is based on the concept of attributes and characteristics in database theory. Therefore, fact-based information is more structured than textual information. In text-based databases the information type 'text' plays the leading part. On the one hand, text is modeled or represented, on the other hand this is done by the text itself (or abstract). It has to be stressed that most databases are 'mixtures' of text-based and fact-based databases. They

are so-called heterogeneous databases. Integrated databases means that they are comprised of a combination of textual, factual, graphical, tabular, etcetera, information.

Additionally, the following classification of environmental databases can be made which is illustrated in Figure 1.

The different types of databases will be explained now, and some examples of free Internet resources will be given.

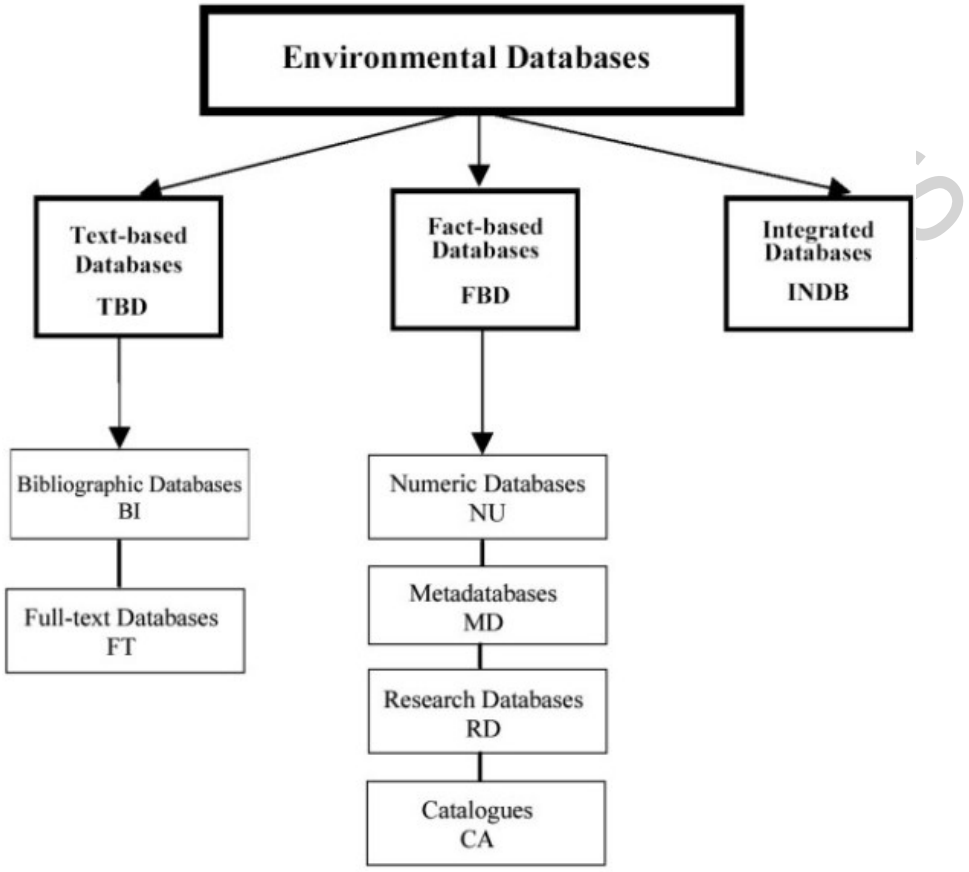

Figure 1. Types of environmental databases.

ChemExper	4000	Belgium. Search for chemical then click 'MSDS.' Lists suppliers as well. Should rise to 12 000 sheets soon.

In addition, general sites other categories like government and non-profit sites chemical manufactures and suppliers, pesticides, and miscellaneous sites are described.

This resource is an extremely helpful tool in finding information on chemical substances with respect to environmental and health issues. As demonstrated, it gives a structured list, but has no database structure. No search in specific data-fields is provided.

Other directories of environmental Internet resources are given in Table 2.

-
-
-

Bibliography

Günther O. (1998). *Environmental Information Systems*, 244 pp. Berlin: Springer Verlag. [This textbook treats the sub-topics data capture, data storage, data analysis and decision support, and metadata management.]

Kumar L. (1998). *Gale Directoy of Databases, Volume 1: Online Databases and Volume 2: CD-ROM, Diskette, Magnetic Tape, Handheld, and Batch Access Database Products*, Vol 1: 1–1022, Vol. 2: 1–1264, Detroit, MI: Gale Research. [This is the largest compilation of commercially available databases worldwide.]

Page B. and Hilty L. M. (1995). *Umwelt-Informatik*, 360 pp. München: R. Oldenbourg Verlag. [This book represents the basic research work in the field of environmental informatics in Germany.]

Ottahal A. (1996). *Der neue Umwelt-Datenbank-Führer*, 535 pp. Köln: Verlag TÜV-Rheinland. [In this directory of environmental databases 550 online databases and CD-ROMs are listed and explained.]

Stoyan D., Stoyan H., and Jansen U. (1997). *Umweltstatistik, Statistische Verarbeitung und Analyse von Umweltdaten*, 348 pp. Stuttgart: B.G. Teubner Verlagsgesellschaft. [The statistical analysis of environmental data is explained and demonstrated in examples.]

Einax J., Zwanziger H. W., and Geiss S. (1997). *Chemometrics in Environmental Analysis*. Weinheim: VCH. [Statistical analysis of environmental chemical data is discussed.]

Voigt K. (1998). Environmental Information Databases. *The Encyclopedia of Computational Chemistry* (eds. Schleyer P. V. R., Allinger N. L., Clark T., Gasteiger J., Kollman P. A., Schaefer III H. F., Schreiner P. R.), pp. 941–952. Chichester: John Wiley and Sons. [The topic of environmental and chemical databases is explained in detail.]

Voigt K., Welzl G., and Rediske G. (1999). Comparison of Environmental Databases: Online, CD-ROM, and Internet. *Proceedings of the 20th Annual National Online Meeting* (ed. Williams M.E.), pp. 487–497. Medford: Information Today Inc. [In this article a matrix of 50 environmentally-relevant descriptors is analyzed and compared for three meta-databases: online, CD-ROM and Internet.]

Zorn P., Emanol M., Marshall L., and Panek M. (1999). Mining meets the web: finding needles in the haystack. *Online* **23**(5), 17–28. [In this article the urgent need to apply data mining concepts to managing the Internet content is stressed.]

Biographical Sketch

Kristina Voigt graduated in food chemistry at the Technical University in Berlin in 1979. From 1980 to 1983 she worked in the Environmental Chemicals Department at the German Environmental Protection Agency (UBA) in Berlin. In 1983 she moved to the GSF National Research Center for Environment and Health in Munich. There her main task is testing and evaluation chemical-relevant and environmental-relevant data-sources (online databases, CD-ROMs, Internet resources). Since 1997 she is in the Institute for Biomathematics and Biometry. There her main task lies in the research field of environmetrics and information management. Since 1988 she has been head of the research project “Information System for Environmental Chemicals.” In 1997 she received her Ph.D. at the University of Erlangen-Nürnberg (Prof. Gasteiger, Computer-Chemistry-Centre). The title of her thesis was: Set-up of metadatabases for environmental chemicals and comparative evaluation of online databases and CD-ROMs. She has published extensively in books, reports, proceedings, and in environmental orientated journals. She holds

training courses and seminars for environmental- and chemical-relevant data sources in Germany. She is giving lectures in the faculty of environmental sciences at the University of Lueneburg. She is a member of the editorial board of Online and CD-ROM Review; a member in organizing committees and refereeing boards of international conferences like the International Online Meeting in London the ECO-INFORMA Conference and the Conferences eon Environmental Informatics; a member in the Management Committee of the Special Interest Group “Informatics in Environmental Protection” in the German Society for Informatics.