

# DATABASES FOR MARINE BIOLOGISTS AND BIOTECHNOLOGISTS: THE STATE-OF-THE-ART AND PROSPECTS

*Vishwas S. Chavan*

*Biological Oceanography Division, National Institute of Oceanography, Dona Paula, Goa 403 004, India*

**Abstract:** Only 1% of the presently available 5000 database titles are relevant to marine biology and biotechnology. Nearly 60% of these are bibliographic in nature. There are almost no textual and numeric databases, which are the prime need of researchers. This paper presents a review of the major databases related to marine biology and biotechnology, with a discussion on the scope for textual and numeric databases.

bases suggests that only 1% (i.e. about 50) are related to marine biology and biotechnology. Out of these 50 databases, less than 10 are totally dedicated to the subject; other databases, particularly bibliographic ones, cover various aspects. Of these 50 databases more than 60% (about 30) are bibliographic databases. Out of the remaining 40%, textual and numeric databases, nearly 60% (about 12) are dedicated to genetics, encompassing protein sequences, nucleic acid sequences etc. (Table 2 in Appendix). This clearly shows a lack of subject-specific databases.

## 1. Introduction

Computer-based information storage, retrieval and scanning in any developed country have proved to be of great importance in handling the enormous amount of data being generated every day. R&D units in developing countries are catching up in computer and telecommunications-based information storage, processing, retrieval and transmission. Scientific organisations are no exception to this. There are more than 5000 databases worldwide on various subjects. Databases related to scientific topics have dominated the database manufacturing houses as it is essential to keep abreast of the latest information for initialising and conducting scientific research studies.

The present paper provides a review of the databases available internationally and in India for marine biologists and biotechnologists; it discusses their shortcomings and the prospects for database manufacturers and subject experts in creating more informative, user-friendly databases.

## 2. State-of-the-art

Of the more than 5000 databases and databanks presently available, very few are dedicated to marine science. Those that are include ASFA (Aquatic Sciences and Fisheries Abstracts), Aquaculture, Aqualine, Oceanic Abstracts, Fishbase etc. (Tables 1 and 2 in Appendix). An appraisal of such data-

### 2.1. Databases and the Third World

All the databases listed in Tables 1 and 2 are from developed nations. The contribution from Third World countries towards globally circulated databases is practically nil, with a few exceptions such as Fishbase from the Philippines.

However, recently, in the developing countries, particularly in India, databases related to marine biology and biotechnology have been created (Table 3 in Appendix), but dissemination of the stored knowledge is very poor due to lack of sophisticated communication facilities. Databases can either be retrieved from floppy disk or printouts, which create an obstacle in transferring information.

## 3. Facts

Most of the databases cover a broad range of topics and hence subject specificity is lost. Because of an insignificant contribution from developing nations and problems with dissemination as well as retrieval of information, marine biologists and biotechnologists are deprived of recent advances in the field of interest.

Most of the bibliographic databases cover the information of the past 15 to 20 years, i.e. 1970 onwards. Thus searching older literature by traditional ways takes a lot of time.

There is no uniformity in the structure, storage and operating and retrieval systems of different databases. Presently,

producers evolve their own formats and schemes without reference to the ones already available on the subject, increasing the chances of data loss during format conversion and reconditioning. This is irrespective of the fact that the need for a national and international format for exchange of bibliographic information has long been felt. Such problems create confusion for the end-user.

It has been observed that there is repetition of information in many of the bibliographic databases. Many of the existing bibliographic databases do not use a standard thesaurus. In many cases, thesauri to go by are not available. Command languages have not been standardised and hence end-users have to learn more than one common language.

Though more descriptive textual and numeric databases are needed today, very few have been developed, and these have been neglected by subject experts too.

The world's oceans contain nearly one million biotic species of which only 45 000 have so far been documented by the National Oceanographic Data Centre (NODC), USA. Oceans are a renewable source of food; such a large gap in information is an impediment to research, management and conservation activities in the marine environment.

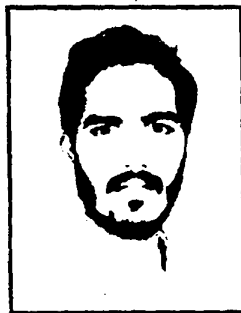
#### 4. Prospects

There is great scope for subject-specific and global bibliographies on particular aspects of marine biology and biotechnology. In this respect subject experts have to play a major role in maintaining the quality of the product through critical scrutiny and selection of the records.

Literature published prior to 1970 needs to be computerised to save time, as does literature from Third World countries. International database manufacturers should collaborate with the national libraries and information departments of these nations in order to collate the literature on the topic of concern.

Coordination at both national and international level is imperative. Organisations like Unesco should take more initiative in keeping check over duplication of records in different databases and the development of textual and numeric databases. A code for the standardisation of databases could be redrafted by such an agency.

#### The author



Subject experts and different organisations should collaborate to create more textual and numeric databases which would prove helpful in sustainable resource utilisation, strengthening taxonomy and research, and in identifying endangered species, extinct forms, the extent and effect of pollution and the status of marine biodiversity and environment.

Decentralised delivery options such as floppy disks, CDROM and magnetic tapes are of significance, especially in developing countries like India. These would prevent users from using unreliable telecom systems and thus expenditure on online communication could be minimised. By this means of dissemination, database producers could exercise greater control over the quality of the product. As well as this, decentralisation could offer different types of products such as a full database.

#### 5. Conclusion

Only 1% of the total databases available are related to marine biology and biotechnology. Over 60% of these are bibliographic, the majority of which covers information published after 1970. It has been felt that standardisation of databases and computation of information published before 1970 is imperative. More textual and numeric databases describing marine biota in detail need to be developed. Decentralised delivery options need to be adopted for better dissemination of knowledge in developing countries.

#### Acknowledgement

The author records his sincere thanks to Carol Tenopir, Associate Professor, School of Library and Information Studies, University of Hawaii at Manoa, Honolulu for reviewing this article and providing valuable suggestions to improve its quality. I am thankful to my colleague Dr T.G. Jagtap for his help in improving the quality of the text. Thanks are due to my senior colleagues Dr B.N. Desai, Dr D. Chandramohan and Dr A.H. Parulekar for their encouragement.

#### Vishwas S. Chavan

Vishwas S. Chavan holds a Masters Degree in Environmental Sciences from the University of Poona, India. He is presently working at the National Institute of Oceanography, Goa, India as an Information Officer on the Biotechnology Information System (BTIS) established by the Indian government's Department of Biotechnology. He is involved in and responsible for developing NIODMLI (NIO's Database on Marine Life of India), identified by Unesco's Intergovernmental Oceanographic Commission (IOC) as an inventory database in the field of marine biology. He is engaged in developing a computer-based taxonomy for marine flora and fauna.

## Appendix

Table 1: Major international bibliographic databases related to marine biology and biotechnology.

Name/title of database	Mode	Provider	Coverage	Features
AGRIS	Online CDROM Floppy	Food and Agricultural Association, Paris	1986-present Quarterly update	Contains multilingual information from over 110 countries and agricultural institutes.
AGRIS INTERNATIONAL	Online CDROM Floppy	US National Agricultural Laboratory, USA	1975-present Monthly update	Contains 1 291 330 records on various aspects of agriculture.
Aquaculture	Online CDROM Floppy	National Oceanographic & Atmospheric Administration (NOAA), USA	1970-present Closed file	10 400 records related to diseases, economics, engineering, food, nutrition, growth requirements, legal aspects of marine, brackish, freshwater organisms.
Aqualine	Online CDROM Floppy	Water Research Centre, USA	1960-present Monthly update	138 000 citations related to water quality, management, development and pollution.
Aquatic Sciences and Fisheries Abstracts (ASFA)	Online CDROM	Cambridge Scientific Abstracts, USA	1978-present Monthly update	363 341 citations related to life sciences of seas, inland waters as well as legal, political and social aspects related to aquatic life.
BEASTCD	CDROM	SilverPlatter Inc., USA	1973-present Annual update	Breeding, feeding, management and biotechnology of animals valuable in agriculture and aquaculture.
Biological and Agricultural Index (BAI)	Online CDROM Tape	H.W. Wilson Co.	1964-present Monthly update	Index 225 periodicals related to biology and agricultural science.
BIOSIS PREVIEWS (Biology Abstracts) formerly known as Bioresearch Index	Online CDROM Floppy	Biosciences Information, USA	1969-present Weekly update	7 713 784 records from 9000 journals, symposia, reviews, reports, research communications related to biosciences and medical research.
Biotechnology Abstracts (Derwent Biotechnology Abstracts)	Online CDROM Floppy	Derwent Publications, UK	1982-present	123 404 records related to biotechnology from major journals and patent literature.
CA Search	Online CDROM Floppy	Chemical Abstracts Serv., USA	1967-present Biweekly update	10 323 927 records covered by Chemical Abstracts Services.
CAB Abstracts	Online CDROM Floppy	Commonwealth Agricultural Bureau, UK	1972-present Monthly update	2 878 913 records from 26 abstracting journals related to agriculture and biology.
Current Biotechnology Abstracts (CBA)	Online CDROM Floppy	Royal Society of Chemistry, UK	1983-present Monthly update	Covers 200 core journals in the field of biotechnology. Contains 42 923 records.
Current Contents Search	Online Floppy	Institute of Scientific Information, Inc., USA	1990-present Weekly update	Covers leading journals in life sciences, agriculture, biology, environmental sciences with other branches of science and technology. Contains 1 477 528 records.
EMBASE (Excerpta Medica)	Online CDROM Floppy	Elsevier Science Publishers, The Netherlands	June 1974-present Weekly update	Covers 4000 journals and periodicals related to biomedical sciences with 4 714 889 records.
ENVIROBIB (Environmental Bibliography)	Online CDROM Floppy	Environmental Studies Institute, USA	1973-present Bimonthly update	426 075 records related to ecology, atmosphere, resource management and health.
ENVIROLINE	Online CDROM Floppy	Environmental Information Center, USA	1971-present	150 741 citations from 5000 national and international publications related to environment.

**Table 1 cont:** Major international bibliographic databases related to marine biology and biotechnology.

<i>Name/title of database</i>	<i>Mode</i>	<i>Provider</i>	<i>Coverage</i>	<i>Features</i>
Food, Science and Technology Abstracts (FSTA)	Online CDROM Floppy	International Food Information Services, UK	1969–present Annual update	420 999 abstracts from 1800 journals on food sciences and technology.
General Science Index (GDS)	Online CDROM Tape	H.W. Wilson Co.	1978–present Monthly update	Index 140 periodicals related to different aspects of science including biology, botany, zoology, genetics, environment and oceanography.
HSDB	Online CDROM Floppy			
Life Sciences Collection	Online CDROM Floppy	Cambridge Scientific Abstracts, USA	1978–present Monthly update	Covers 17 abstracting journals in life sciences. Contains 1 392 381 records.
MEDLINE	Online CDROM Floppy	National Library of Medicine, USA	1966–present Weekly update	Covers 3200 journals from 70 countries related to biomedical sciences and biotechnology. Contains 6 856 539 records.
MERCK INDEX	Online	Merck & Co. Inc., USA	Late 19th Century to present Semiannual update	Online version of printed Merck Index Vol. 1 encyclopedia.
Oceanic Abstracts	Online CDROM Floppy	Cambridge Scientific Abstracts, USA	1964–present Bimonthly update	223 368 records related to marine sciences from 3500 worldwide sources.
PASCAL	Online CDROM Floppy	CNRS, Paris	1973–present	3 866 245 records related to biology, biotechnology, life sciences, medicine, chemistry, pollution, food and agricultural sciences.
Pollution Abstracts	Online CDROM Floppy	Cambridge Scientific Abstracts, USA	1970–present Bimonthly update	175 051 records related to various aspects of environmental pollution.
SCISEARCH (Science Citation Index)	Online CDROM Floppy	Institute for Scientific Information, USA	1974–present Biweekly update	Covers world's 90% scientific and technical literature with 10 706 877 records.
SSIE Current Research	Online CDROM Floppy	National Technical Information Services (NTIS), USA.	1978–Feb. 1982 Closed file	439 000 records related to reports of Basic and Applied Scientific research projects.
SUPERTECH	Online CDROM Floppy	R.R. Bowker Inc., USA	1973–present Monthly update	Covers technical and industrial aspects of biotechnology. Contains 86 104 records.
Toxline	Online CDROM Floppy	National Library of Medicine, USA	1981–present	Citations related to toxicity. Contains 1 647 025 records.
Waternet	Online CDROM Floppy	American Water Works Assoc., USA	1971–present Bimonthly update	Index to the publications of AWWA Research Foundation with 23 197 records.
Water Resources Abstracts (SWRA)	Online CDROM Floppy	US Department of Interior, USA	1968–present Monthly update	243 000 records from journals, monographs, reports, patents, articles conferences related to water resources.
Wildlife & Fish Worldwide	CDROM	National Information Services Corporation	1971–present semiannual update	200 000 records on mammals, reptiles and fishes.
ZOOLOGICAL RECORD ONLINE	Online CDROM Floppy	Biosciences Information, USA	1978–present	1 308 826 records from 6000 journals related to zoology.

**Table 2:** *Major international databanks. Textual and numeric databases related to marine biology and biotechnology.*

<i>Name/title of database</i>	<i>Mode</i>	<i>Provider</i>	<i>Coverage</i>	<i>Features</i>
BIOTEC	CDROM	Cuban Academy of Sciences and MultiConsult S.C.		Information on research centres, publishers, associations, universities and organisations from 70 countries working on biotechnology. Contains 5765 entries.
Biotechnology Research Faculty Profile	CDROM Floppy	North Carolina Biotechnology Centre, USA		Contains curriculum vitae of biotechnologists from USA.
BROOKHAVENS	Tape			Useful for application in protein engineering.
Cambridge Structural Database (CSD)	Tape CDROM	Cambridge Crystallographic Data Centre, UK	1969–present	Contains numeric and bibliographic information related to organocarbon crystal structure.
Complex Carbohydrate Structure Database (CCSD)	CDROM	University of Georgia USA		Contains carbohydrate structures
Dictionary of Natural Products	CDROM	Chapman and Hall	1992	Information about 8000 natural products.
EMBL data library	CDROM Tape Floppy	European Molecular Biology Laboratory		Contains nucleotide sequences.
FISHBASE ver. 1.0	Floppy	International Centre for Living Aquatic Resources Management (ICLARM), Manila, the Philippines		Full colour pictures, distribution maps, information on habitats, population aquaculture and diseases of 2000 tropical marine animals with facility to compare species groups or geographical areas.
GENEBANK	CDROM	National Institute of General Medical Sciences	1982–present	Contains genetic sequences
HYBRIDOMA	Tape	CODATA and International Immunological Society		Databank of cloned cell line and their immunoreactive product.
Linnaeus Protist Software ver 1.0	CDROM	Expert-Centre for Taxonomic Identification (ETI) Amsterdam, The Netherlands		Full description of ecology, reproduction, economic importance, distribution, biochemistry, identification characters, taxonomy, synonyms, photographs, drawings, sounds and films of 360 marine protistan species from northern European coastal waters.
Natural Resources Database	CDROM	National Information Services Corporation	Annual update	Information on wetland, Pacific island ecosystems, national parks, resource management, endangered species with bibliography.
NBRF-PIR (NUC ACID)	Tape	National Biomedical Research Foundation		Contains 3205 sequences and 7 151 795 nucleotides.
NBRF-PIR (Protein)	Tape	National Biomedical Research Foundation		Contains 6330 sequences and 1 850 665 residues used for protein identification.
NODC Code list	Print	National Oceanographic Data Centre, USA		Coding list of plant and animal species.
PSD-Kyoto		T. Ooi Chemical Research Institute Kyoto, Japan		Contains protein sequences.
Zoological Records	Online CDROM Floppy	Biosciences Information Services, USA and Zoological Society of London, UK	1971–1981 Monthly update	286 500 records related to systematics and taxonomic information of various groups of organisms along with bibliographic information.

**Table 3:** Major Indian bibliographic, numeric and textual databases related to marine biology and biotechnology.

<i>Name/title of database</i>	<i>Mode</i>	<i>Provider</i>	<i>Coverage</i>	<i>Features</i>
Abstracts of Current Literature in Toxicology (ACLIT)	Print	Indian Toxicological Research Institute, Lucknow	1984–present	Covers 250 journals related to environment and toxicology. Growth rate 1000 records/year.
Bibliography of Indian Agriculture (BIA)	Print	Indian Agricultural Research Institute, New Delhi.	1944–present	17 500 records from 1200 Indian and foreign publications related to various aspects of Indian agriculture.
Bibliography of Indian Zoology (BIBLPHY INDIAN ZOOL)	Print	Zoological Survey of India, Calcutta	1958–present	55 000 records related to literature on fauna of Indian region. Growth rate 2000 records/year.
BTIS-BOSE	Print	Distributed Information Centre, Bose Institute, Calcutta	1989–present	10 000 records related to genetic engineering.
Current Literature on the Science of Sciences (CLOSS)	Print	National Institute of Sci. Technology and Development Studies, New Delhi		10 000 records related to science and technological developments.
Current Research in Plant Sciences (CRIPS)	Floppy Print	National Botanical Research Institute, Lucknow	1990–present Monthly update	21 000 records from 600 national and international periodicals.
ENVIS (Paryavaran Abstracts)	Floppy Print	Ministry of Environment and Forest, New Delhi	1984–present Quarterly update	Contains 4000 records from journals, periodicals related to environment. Growth rate 1000 records/year.
Food Science and Technology Abstracts (FSTA)	Print	CFTRI, Mysore	1987–present Monthly update	34 000 records from major Indian publications related to food science technology.
Food Technology Abstracts	Print	CFTRI, Mysore	1987–present Monthly update	6200 records from major Indian publications related to food science technology.
Indian National Bibliography	Print	Central Reference Library, Calcutta		Covers printed material from 14 Indian languages. Contains 15 000 records.
Indian Science Abstracts	Floppy Print	INSDOC, New Delhi	Biweekly update	Covers 750 journals and periodicals.
National Union Catalogue of Scientific Serials in India	Print	INSDOC, New Delhi		35 000 titles related to science and technology.
OCEANDRU ALERT (NICDAP)	Print	Central Drug Research Institute, Lucknow	1988–present	Sources of drugs manufacture and scientific data on drugs including drugs from the sea.
OCEANLINE	Print	National Institute of Oceanography, Goa	1989–present Biweekly update	1500 records with growth rate of 1000 records/year related to oceanography.
REFERA 1.1	Floppy Print	Distributed Information Centre, MKU, Madurai	1696 to 1990	6500 records on rotifera group of organisms from national and international sources.
<i>Numeric and textual databases</i>				
Databank on Toxic Chemicals (DABTOC)	Print	ITRC, Lucknow	1986 onwards	Multilingual information on 228 chemicals.