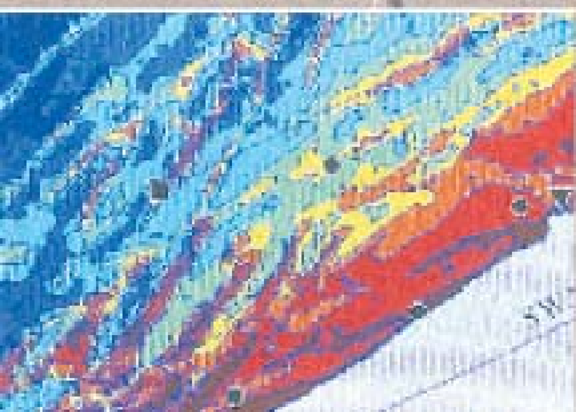
PROCEEDINGS
THE COLOUR OF OCEAN DATA

International symposium on oceanographic data and information management with special attention to biological data
Brussels, Belgium, November 25-27, 2002

Edited by
Edward Vanden Berghe,
Murray Brown,
Mark J. Costello,
Carlo Heip,
Sydney Levitus
and Peter Pissierssens



IOC Workshop Report No.188
VLIZ Special Publication No.16

PROCEEDINGS

THE COLOUR OF OCEAN DATA

International symposium on oceanographic data and information management with special attention to biological data

Brussels, Belgium
November 25-27, 2002

Edited by

***Edward Vanden Berghe*[1], *Murray Brown*[2], *Mark J. Costello*[3], *Carlo Heip*[4], *Sydney Levitus*[5] *and Peter Pissierssens*[6]**

[1]Flanders Marine Institute (VLIZ), Vismijn Pakhuizen 45-52, B-8400 Oostende, Belgium
[2]834, Elysian Fields Av., New Orleans, Louisiana 70117, USA
[3]Huntsman Marine Science Center, 1 Lower Campus Road, St Andrews, New Brunswick, E5B 2L7, Canada
[4]Netherlands Institute of Ecology, Centre for Estuarine and Marine Ecology (NIOO-CEME), Korringaweg 7, Postbus 140, NL-4400 AC Yerseke, Netherlands
[5]National Oceanic and Atmospheric Administration, National Oceanographic Data Center, SSMC3, 4th Floor, 1315 East-West Highway, Silver Spring, MD 20910, USA
[6]Intergovernmental Oceanographic Commission (UNESCO), 1 rue Miollis, F-75732 Paris Cedex 15, France

This symposium was jointly organized by the Flanders Marine Institute (VLIZ), the Belgian Science Policy, the Intergovernmental Oceanographic Commission / International Oceanographic Data and Information Exchange (IOC/IODE), the Ocean Biogeographic Information System / Census of Marine Life (OBIS/CoML)

VLIZ

Belgian Science Policy

CENSUS OF MARINE LIFE

IOC Workshop Report No.188
VLIZ Special Publication No.16

Disclaimer

Bibliographic reference

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus, P. Pissierssens (Eds). Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25–27 November 2002. Paris, UNESCO/IOC, VLIZ, 2004. x + 308 pp. (IOC Workshop Report, 188) (VLIZ Special Publication, 16)

Pictures cover

A. Craasmann
M. Decleer

Co-published in 2004 by

United Nations Educational,
Scientific and Cultural Organization
Intergovernmental Oceanographic Commission
1 rue Miollis
F-75732 Paris Cedex 15
France

Vlaams Instituut voor de Zee vzw
Flanders Marine Institute
Vismijn, Pakhuizen 45-52
B-8400 Oostende
Belgium

Printed in UNESCO's workshops

Printed in France

(SC-2004/WS/1)
ISSN 1377-0950

Preface

More and more, ocean data management has to play a crucial role in global as well as local matters. Vast amounts of physical, chemical and biological oceanographic data are collected in all seas and oceans of the world. International networks are created, to standardize data formats and facilitate data exchange. Global databases of existing data have been compiled, global programmes are operational and new ones are developed, data are routinely exchanged. The Intergovernmental Oceanographic Commission, with the network of National Oceanographic Data Centres, and the International Council of Scientific Unions, with the World Data Centres, have played a major catalysing role in establishing the existing ocean data management practices.

No one can think of data management without thinking of information technology. New developments in computer hard- and software force us to continually rethink the way we manage ocean data. One of the major challenges in this is to try and close the gap between the haves and the have-nots, and to assist scientists in less fortunate countries to manage oceanographic data flows in a suitable and timely fashion.

So far major emphasis has been on the standardization and exchange of physical oceanographic data in open ocean conditions. But the colour of the ocean data is changing. The 'blue' ocean sciences get increasingly interested in including geological, chemical and biological data (*e.g.* the newly created IODE Group of Experts on Biological and Chemical Data Management and Exchange Practices – GE-BCDMEP). Moreover the shallow sea areas get more and more attention as highly productive biological areas that need to be seen in close association with the deep seas. How to fill in the gap of widely accepted standards for data structures that can serve the deep 'blue' and the shallow 'green' biological data management is a major issue that has to be addressed.

And there is more: data has to be turned into information. In the context of ocean data management, scientists, data managers and decision makers are all very much dependent on each other. Decision makers will stimulate research topics with policy priority and hence guide researchers. Scientists need to provide data managers with reliable and first quality controlled data in such a way that the latter can translate and make them available for the decision makers. But do they speak the same 'language'? Are they happy with the access they have to the data? And if not, can they learn from each other's expectations and experience?

The objectives of this symposium were to harmonize ocean colours and languages and create a forum for data managers, scientists and decision makers with a major interest in oceanography, and open to everyone interested in ocean data management.

Edward Vanden Berghe
on behalf of the Scientific Committee

The symposium

The ‘Colour of Ocean Data’ symposium was organized from 25 to 27 November 2002, in the Palais des Congres in the centre of Brussels, by the Flanders Marine Institute, the Intergovernmental Oceanographic Commission of UNESCO, the Belgian Science Policy, and the Census of Marine Life. Nearly 200 participants were registered; there were 44 oral presentations, 40 poster presentations and eight demonstrations.

In a series of five sessions, various aspects of data management were discussed. The main aim was to allow the different communities to learn about developments on related fields and to learn from each others’ experiences. For each session of oral presentations, there was a corresponding session with poster presentations and demos. The wide variety of topics that were discussed was indicative of the breadth of the field; the ensuing discussions clearly demonstrated the timeliness of the symposium. All powerpoint files of the oral presentations, and all abstracts for both oral and poster presentations are available through the VLIZ web site (http://www.vliz.be/En/Activ/Events/Cod/cod.htm). The chairs of these sections kindly agreed to act as editors for the papers that were presented in their session.

The two last hours of the symposium were devoted to a short panel discussion. Two representatives each of international organizations, of the data management community and of the scientific community were given the opportunity to expand on their views on oceanographic data management, their views on the role of data centres, and expectations from user communities. The conclusions from this panel discussion are included in these proceedings.

Symposium sessions

- Opening session
    - Belgium: F. Demeyere, Deputy Head of Cabinet, Government Commissioner, attached to the Minister for Scientific Research, Mr. Yvan Ylieff
    - Flanders: R. Herman, representative of the Government of Flanders (Afdeling Technologie en Innovatie, Ministerie van de Vlaamse Gemeenschap; Administratie Wetenschap en Innovatie)
    - UNESCO: P. Bernal, Assistant Director-General UNESCO and Executive Secretary IOC·
    - M. Altalo (Corporate Vice President, Energy Solutions, USA)
      Social relevance of ocean data management – importance of products following from data management
    - P. Mathy (Head of Unit, European Commission; Directorate RTD I: Preserving the Ecosystem; Unit RTD I.4: Marine ecosystems. Infrastructure)
      Perspectives for marine sciences and technologies in Framework Programme VI
- Marine capacity building in global programmes; Chair: Murray Brown, Phoenix Training Consultants, USA
- Biodiversity data; Chair: Mark Costello, Director, Ecoserve, Ireland

- Ecological and community data; Chair: Carlo Heip, Director, Centre for Estuarine and Marine Ecology, Netherlands Institute of Ecology, The Netherlands
- New internet developments; Chair: Peter Pissierssens, Head, Ocean Services IOC (IOC/OCS), France;
- Case studies; Chair: Sydney Levitus, Director, World Data Centre for Oceanography – Silver Spring, USA
- Panel discussion; Chair: Savi Narayanan, Marine Environmental Data System, Canada

Scientific Committee

- Efstathios Balopoulos (HNODC, Greece; IOC/IODE)
- Murray Brown (Phoenix Training Consultants, USA)
- Mark Costello (CoML/OBIS; Ecoserve, Ireland; Hunstsman Marine Science Centre, Canada)
- Carlo Heip (NIOO/CEME, The Netherlands)
- Syd Levitus (NODC Washington/WDC-A, USA)
- Jan Mees (VLIZ, Belgium)
- Nick Mikhaelov (RIHMI-WDC, Russia)
- Savi Narayanan (MEDS, Canada)
- Peter Pissierssens (IOC/IODE)
- Peter Schalk (ETI, The Netherlands)
- Ben Searle (AODC, Australia; IOC/IODE)
- Edward Vanden Berghe (VLIZ, Belgium)
- Magda Vincx (University Ghent, Belgium)
- Ron Wilson (University of Delaware, USA)

Organizing Committee

- Edward Vanden Berghe (VLIZ, Belgium): chairman
- Rudy Herman (AWI, Belgium)
- Jan Mees (VLIZ, Belgium)
- Frank Monteny (The Belgian Science Policy, Belgium)
- Peter Pissierssens (IOC/IODE)
- Greg Reed (IOC/IODE)
- Serge Scory (MUMM, Belgium)
- Jan Seys (VLIZ, Belgium)

Acknowledgements

Our sincere thanks go to our sponsors, who made this conference possible, and to the scientific community that made organizing this symposium such a pleasure.

Thanks to Ingrid Dobbelaere and Kristin de Lichtervelde for the layout of the proceedings and the help with the practical organization of the symposium itself.

Table of contents

SESSION: CAPACITY BUILDING

SESSION: BIODIVERSITY AND BIOGEOGRAPHY

SESSION: ECOLOGICAL AND COMMUNITY DATA

SESSION: NEW DEVELOPMENTS

SESSION: CASE STUDIES

PANEL DISCUSSION

SESSION: CAPACITY BUILDING

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p.3-14
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

The Global Mangrove Database and Information System (GLOMIS): present status and future trends

Shigeyuki Baba[1,2], Chris Gordon[3,4], Mami Kainuma[2], Jesse S. Ayivor[3,4] and Farid Dahdouh-Guebas[5]

[1]Faculty of Agriculture, University of the Ryukyus
Okinawa, 903-0129, Japan
E-mail: mangrove@ryukyu.ne.jp

[2]International Society for Mangrove Ecosystems (ISME), c/o Faculty of Agriculture, University of the Ryukyus, Okinawa 903-0129, Japan

[3]Volta Basin Research Project, University of Ghana
PO Box LG 209, Legon, Accra, Ghana

[4]Centre for African Wetlands, University of Ghana
c/o PO Box LG 67, Legon, Accra, Ghana

[5]Laboratory of General Botany and Nature Management (APNA), Mangrove Management Group, Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium

Abstract

The Global Mangrove Database and Information System (GLOMIS) is an activity implemented by the International Society for Mangrove Ecosystems (ISME). ISME took over the research, training activities and studies formerly commissioned by the UNDP/UNESCO projects on the management of mangrove ecosystems. GLOMIS is funded by the International Tropical Timber Organisation. GLOMIS started in 1997 and aims at constructing and operating a global mangrove database for the compilation of mangrove-related information scattered around the world. This information is being collected by four GLOMIS Regional Centres (Brazil, Fiji, Ghana and India) and disseminated worldwide from GLOMIS Headquarters (Okinawa, Japan). The collected data is available on the GLOMIS homepage (http://www.glomis.com) as well as CD-ROM and a five volume soft-cover publication. Over a period that currently extends from 1997 to 2003, GLOMIS has built a database on a variety of published and unpublished mangrove data (*e.g.* distribution, wood and fish productivity, environmental, ecological and socio-economic value), mangrove projects, and institutions and people working on mangroves. In the future, GLOMIS plans to integrate the data in a geographic information system for use by researchers, planners, policy- and decision-makers, and coastal zone managers. GLOMIS faces the challenges of combining terrestrial and marine data from the mangrove ecosystem that is strategically located at the interface between land and sea. This paper reports on the relevance, preparatory stages, operation, outputs and future of GLOMIS.

Keywords: Mangrove database; GIS; Internet.

Introduction

Mangroves are plants distributed in intertidal areas of the tropics and sub-tropics, and comprise the only flowering trees (actual forests) that thrive under seawater conditions. The term mangrove ecosystem is also applied to this particular floral assemblage, and its fauna, which is recognised as one of the world's most remarkable ecosystems, supporting a unique biological diversity – partly terrestrial, partly marine, and partly intertidal in origin. Because of this, mangroves receive a particular label within coastal and marine ecology on one hand, and within terrestrial ecology (forestry) on the other, and the original combination of both marine and forestry data is inherent to the research associated with mangrove ecosystems.

Mangroves are functionally and structurally diverse. They occur in saline, waterlogged soils mostly along protected coastal shores where tidal fluxes alternately inundate and expose the muddy sandy soils and sediments. The ecosystem is dominated by halophytic woody plants, and includes the substrates on which they grow, micro-flora and a wide array of terrestrial, aquatic and benthic fauna. Mangrove ecosystems receive large inputs of matter and energy of both marine and terrestrial origin due to their strategic location at the interface between land and sea, which partially explains their high productivity. The mangrove forest itself provides many economically useful materials like firewood, wood for charcoal production and raw materials for paper and chipboard. The forest also provides fodder for livestock, and materials for medicines, and dyes. It also has many ecosystem functions: mangroves serve as nursery grounds and as habitat for breeding, hatching and spawning of many marine and terrestrial organisms. Mangroves are also important for their role as coastal stabilizers and as biological filters and sinks for several pollutants. One recent intriguing finding is that the below ground carbon content of mangrove forests is 4-18 times higher than the carbon content of tropical rain forests (Fujimoto, in press). This indicates that positive action in the area of mangrove conservation and rehabilitation would contribute immensely to the sequestration of CO_2.

When managed sustainably, the mangrove forest can yield goods of high economic value and earn tremendous foreign exchange for the producing country without compromising its ecological and environmental integrity. Rönnbäck (2001) reported that 1ha of mangrove in developing countries generates 1 to 11.8 tons of fisheries catch with a market value of € 900-12,400, and worldwide tidal marshes and mangrove areas are worth € 1,600,000,000,000 (Costanza *et al.*, 1997). Unfortunately, conservation of the mangrove has not been the practice as over-exploitation, mismanagement and conversion to other uses have resulted in a substantial loss of mangrove ecosystems worldwide. Recent studies have shown that in the African, American and Indo-West Pacific regions where mangroves are most extensively distributed, these valuable yet fragile ecosystems are faced with both natural and human threats. The natural threats include storms or hurricanes, tidal waves, herbivory, pests and diseases, sediment movement, river floods and coastal erosion. Human threats on the other hand include mangrove felling, mining, salt winning, over-fishing, as well as water pollution, oil contamination, and solid waste disposal. Mangroves are also threatened by global warming as a result of heavy industrial activities and the increasing emission of CO_2 into the atmosphere. (Dahdouh-Guebas, 2002; Diop *et al.*, 2002; Lacerda *et al.*, 2002; Vannucci, 2002).

In general, mangrove ecosystems are delicate due to the fine balance that exists between the marine and terrestrial components, and once destroyed, they cannot be restored easily. It is essentially this recognition that prompted the International Society for Mangrove Ecosystems (ISME) and the International Tropical Timber Organisation (ITTO) to establish the Global

Mangrove Database and Information System (GLOMIS) for the conservation, rational use and sustainable management of mangroves as consistent with the ITTO objectives.

The International Society for Mangrove Ecosystems

The International Society for Mangrove Ecosystems (ISME) is an international non-profit, non-governmental scientific society, established in August 1990. ISME took over the research, training activities and studies formerly commissioned by the UNDP/UNESCO projects on the management of mangrove ecosystems. The need to establish an international mangrove society arose from the alarming rate of mangrove forest degradation occurring worldwide. The ISME's Mangrove Charter states that "the Society shall collect, evaluate and disseminate information on mangrove ecosystems" and "promote international cooperation". ISME is affiliated with the International Council of Scientific Unions (ICSU) and is part of the International Union of Biological Sciences (IUBS) through the International Association of Biological Oceanography (IABO). ISME is an NGO associated with the United Nations Department of Public Information with a Roster status and is included in the list of NGOs in consultative status with the United Nations Economic and Social Council (ECOSOC). The Headquarters of ISME is located at the University of the Ryukyus (Okinawa, Japan) and the executive body consists of a President, three Vice-Presidents, a Treasurer and an Executive Secretary who are drawn from a wide range of countries. ISME membership counts over 40 institutional and 823 individual members from 80 countries and regions as of December 2002.

ISME aims at creating a well-organized network of people interested in all aspects of mangrove studies by promoting research and surveys for conservation, rational management and sustainable utilization of mangroves. It also intends to serve as an international databank on mangrove ecosystems. Current major activities of ISME are as follows:

- International Tropical Timber Organization (ITTO)/ISME project 'Global Mangrove Database and Information System';
- Japan International Cooperation Agency (JICA)/ISME annual mangrove management training course 'Sustainable Management of Mangrove Ecosystems';
- Restoration of degraded mangroves worldwide; and
- Organizing international workshops and symposiums.

ISME also has a number of publications, such as ISME Newsletter, series of scientific journals (technical reports, occasional papers and proceedings), world mangrove distribution maps [World Mangrove Atlas (Spalding *et al.*, 1997)], technical manuals for mangrove restoration (Field, 1996), non-technical mangrove reading (Field, 1995) and slides and videos for public awareness.

In addition to ITTO, an intergovernmental body dedicated to the sustainable development and conservation of tropical forests, ISME collaborates with the United Nations Educational, Scientific and Cultural Organisation (UNESCO), United Nations University (UNU), the Food and Agriculture Organisation (FAO), World Bank and with many international NGOs such as IUCN-Pakistan and local NGOs to implement various mangrove projects.

The Global Mangrove Database and Information System

Background

The Global Mangrove Database and Information System (GLOMIS) was started in 1996 under the sponsorship of ITTO [PD 14/97 Rev. 1 (F)]. GLOMIS was designed to develop a global mangrove database and information system that essentially promotes information sharing and co-operation among scientists, professionals, decision makers and coastal dwelling people of the tropical and sub-tropical zone that depend on mangroves for economic livelihood. Mangrove-related information scattered around the world is compiled to avoid unnecessary duplication of work. GLOMIS puts emphasis on exchange of information and cooperation among scientists, governments and the people for the conservation, rational use and sustainable management of the world's mangroves. GLOMIS is based at the ISME Secretariat in Okinawa, Japan, and is supported by four Regional Centres located in Brazil, Fiji, Ghana and India.

The GLOMIS project initiative stemmed from the realization of the problem that mangroves all over the world continue to suffer massive destruction on a daily basis for various reasons despite the immense benefits the society derives from them, and the need to adopt a coordinated scientific approach to readdress the problem.

Objectives of GLOMIS

To help solve this problem, therefore, the project sets out to achieve a development objective of ensuring that mangrove ecosystems worldwide are managed and utilized rationally for use of timber, fodder, charcoal and fuelwood materials, and many other forest products as well as maintenance of coastal water quality and stability to ensure the sustainable production of fisheries.

Specifically, GLOMIS seeks to:

1) Build a database of all published and unpublished available data on the distribution and productivity of mangrove species and major mangrove forest types worldwide; their environmental condition and site specificity; their structure, growth rates and commercial timber yield; their utilization and management for timber production, fisheries and other purposes; and their environmental, ecological and economic value. Data on institutions, people and projects related to all aspects of mangroves are also entered in the database.

2) Implement the functional stage of a geographic information system (GIS) within GLOMIS that would provide planners, policy and decision makers and coastal zone managers with tools – including dollar value – to evaluate management options and implement rational plans for sustainable utilization of mangroves for forestry, fisheries and other uses.

Operation and outputs of GLOMIS

GLOMIS is a searchable database of scientific literature relating to mangroves, institutions and scientists working on any aspect of mangroves, as well as regional projects and programmes related to mangroves. The Project GLOMIS mainframe consists of the GLOMIS Headquarters (GLOMIS HQ) in Okinawa, Japan at ISME Secretariat, and four Regional Centres; Brazil

(Universidade Federal Fluminense UFF in Fortaleza covering Latin America and Caribbean), Fiji (The National Trust of Fiji in Suva covering Australia, New Zealand and Oceania), Ghana (Center for African Wetlands, University of Ghana in Accra covering Africa and Middle East), and India (M.S. Swaminathan Research Foundation in Chennai covering Continental Asia and S.E. Asia). The GLOMIS HQ, on its part, collects mangrove information from Europe, North America and Japan. Mangroves are distributed in about 100 countries worldwide and the Regional Centers with the GLOMIS HQ are responsible for collecting necessary information from mangrove-occurring countries within their geographical areas of operation. The project is carried out by the Project Coordinator, Assistant Coordinator, leaders and staff members of Regional Centres, Webmaster, ISME Secretariat and GLOMIS Board members who supervise the operation of GLOMIS. Four categories of mangrove related references, people, projects and institutions are compiled at each Regional Centre and are transferred to GLOMIS HQ (Fig. 1). For system components Linux is used as the Operation System. For security purposes, the system is protected by a firewall. Since no firewall is perfect, a back up server system has been prepared.

During the first phase (Phase I) of GLOMIS (1997-2001), the foundation of the GLOMIS database was established through the development and construction of a standardized format. To enter data in the standardized format, a user-friendly software data entry system was developed and distributed to each Regional Center. The main server of GLOMIS is located at GLOMIS HQ, Okinawa, Japan, where data collected at the Regional Centres are sent for editing to ensure the dissemination of reliable data. When it becomes necessary, data are sent back to the respective Regional Centre for correction.

In March 2001, the GLOMIS website (http://www.glomis.com) was launched, ushering in Phase II of the GLOMIS project. One of the most important functions of GLOMIS is its Internet-accessible information system and database that provides necessary information on mangroves to people anywhere in the world. In addition to the searchable database, the GLOMIS homepage provides access to the ISME/GLOMIS Electronic Journal (EJ) designed to inform mangrove scientists of new developments with regard to mangrove ecosystems. The homepage also contains other sources of information such as GLOMIS News to provide reports from each Regional Centre, conference information, and news on GLOMIS and mangrove ecosystems. A five volume soft-covered book and a CD-ROM with GLOMIS data as of September 2001 were published as outputs of Phase I to reach out to non-Internet users (ITTO/ISME, 2001a, 2001b). The CD-ROM contains additional information including photographs and mangrove distribution maps which were not put on the Internet, due to the size of the files, given that a number of countries use slower modems (9.6 - 28.8 Kbps) for Internet access, and thus have difficulty in retrieving large amounts of data. The database was therefore designed to be as light as possible for easy data accessibility. Currently, GLOMIS data are being distributed free of charge. As of December 2002, the GLOMIS database had received 7,135 inquiries (Fig. 2).

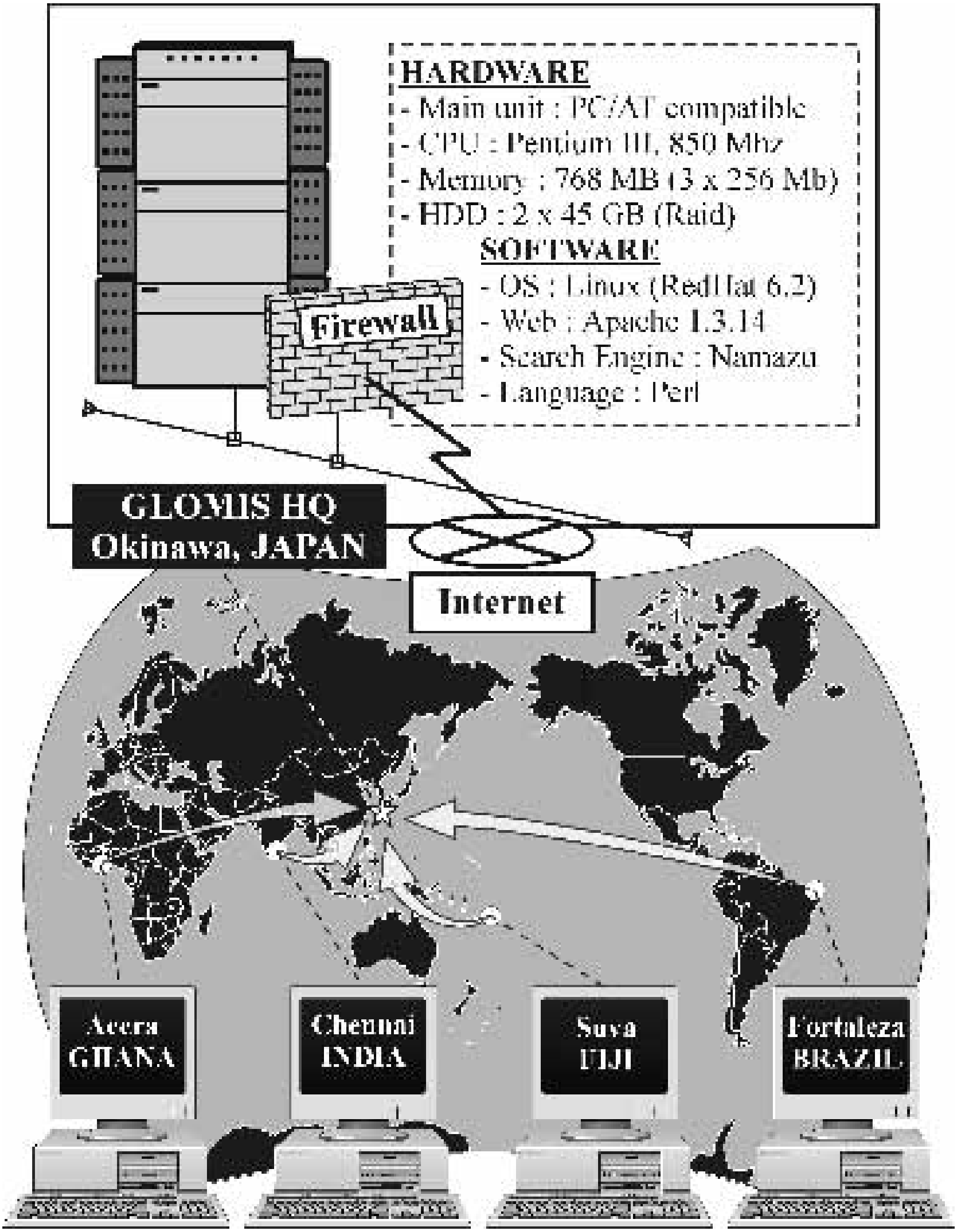


Fig. 1. Data transfer and system components of GLOMIS. The four Regional Centres forward the data to the GLOMIS headquarters in Japan (GLOMIS HQ) where it is put on the server through the system components given on the top of the figure.

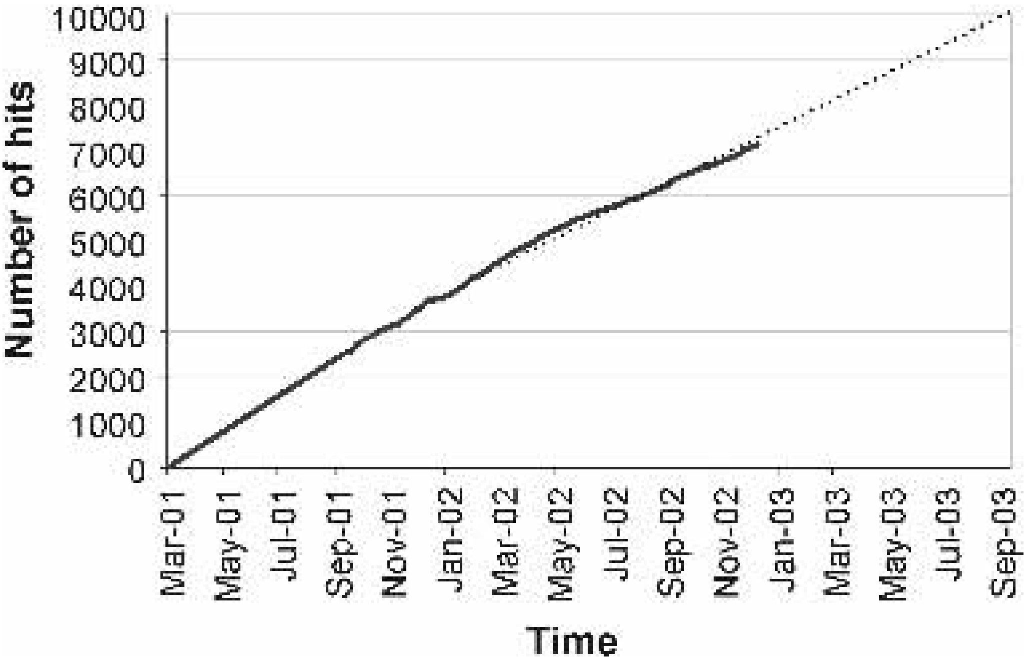


Fig. 2. GLOMIS (cumulative) hit numbers in function of the time. The dotted line represents the estimated month to 10,000 hits.

The database provides a catalogue list of four categories, namely mangrove-related references, people, projects and institutions, as earlier mentioned. Each datum is named as follows: CTCC0000000.txt where CT (Category type) is either IN (Institution), PE (People), PR (Project) or RE (Reference), and CC (Country code) is BR (Brazil), FJ (Fiji), GH (Ghana), IN (India) or JP (Japan) depending on where the data is entered. To date, a total of 5,525 data are compiled and disseminated, out of which 4,800 are reference information. The database can be searched either by keywords, categories or free words. GLOMIS has over 245 predetermined keywords concerning mangrove ecosystems in the area of biology, management, research methods and geography. A category search can be done from four categories and free word search can be done by entering the word of interest (Fig. 3). An example on the result of the database search is shown in Fig. 4.

The second phase (Phase II) of GLOMIS (2001-2003) focused on updating and maintenance of the database while establishing quality control of the data. Currently only published sources are entered in the database as Reference. Entries on People are only made at the consent of the individual concerned or where the individual is already on a public database. Some duplication of data has been observed in the database apparently due to the overlapping entries sent from different Regional Centres. Software to automatically pick up these duplicated data for correction is now being developed by the GLOMIS technical support partner Fujitsu-Okinawa.

Since GLOMIS currently does not have a library function, and due to copyright issues, hard copies of references in the database are not provided to the end users, however with the combination of the reference and people database, end users can make contact with the authors of articles to make direct requests for copies of publications. One of the major difficulties the GLOMIS faces is collecting scattered information from all over the world and keeping track of the latest information in the field. The staffs of the Regional Centres are physically travelling to the respective countries in their regions to collect information in addition to contacting mangrove-related people through traditional mail, e-mail and the Internet. Efforts are being made to publicize GLOMIS at as many as possible of the various mangrove and coastal ecosystem related conferences so as to have more inputs into the database. The strengthening of Regional Centres as well as active support from end users is indispensable for the success of GLOMIS.

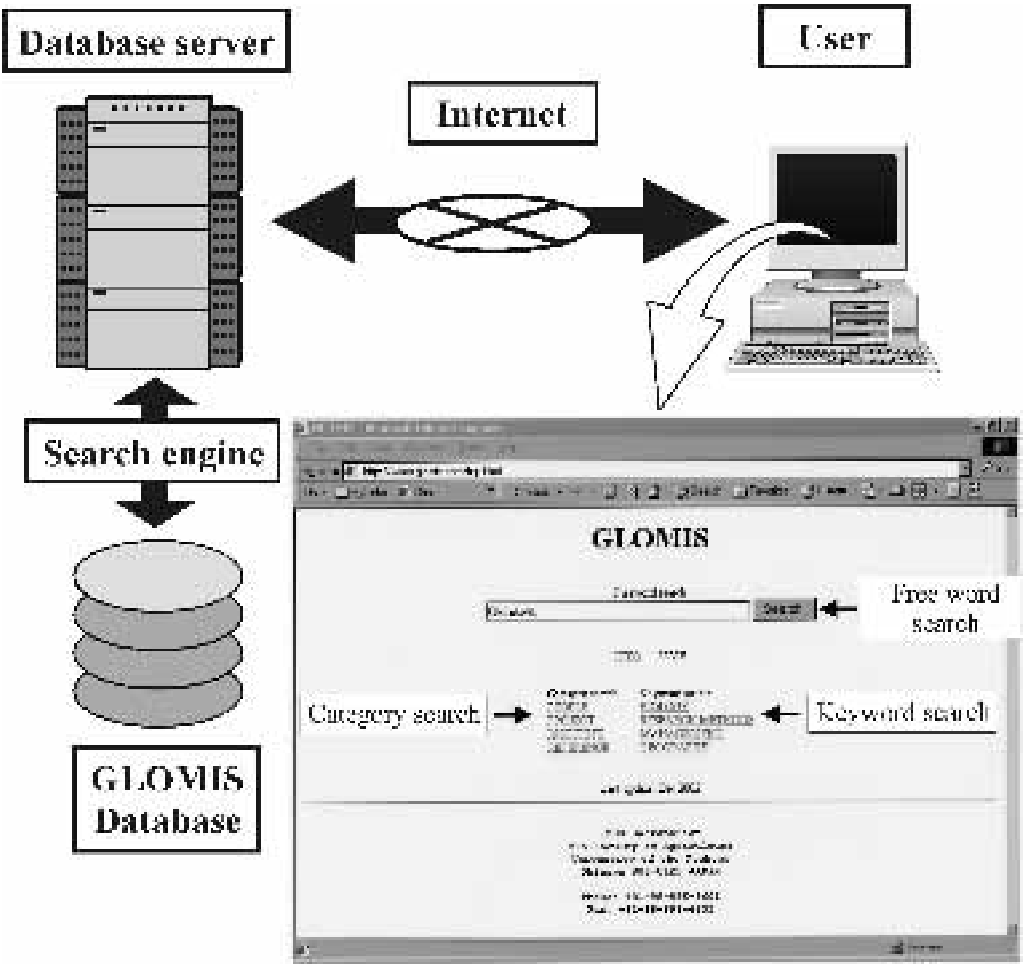


Fig. 3. The GLOMIS data retrieval components and the three methods (bottom right computer screen).

The GIS within GLOMIS is currently limited to the world-wide distribution of mangroves based on the data in ISME's World Mangrove Atlas (Spalding *et al.*, 1997). For South and Southeast

Asia, Australasia, the Americas, West Africa, and East Africa and the Middle East, a total of 25 mangrove distribution maps are provided. These do not include the distribution of each single species, the latter of which is given per country as species lists. However, currently, GLOMIS offers the option of inserting the location and identity of mangrove species worldwide.

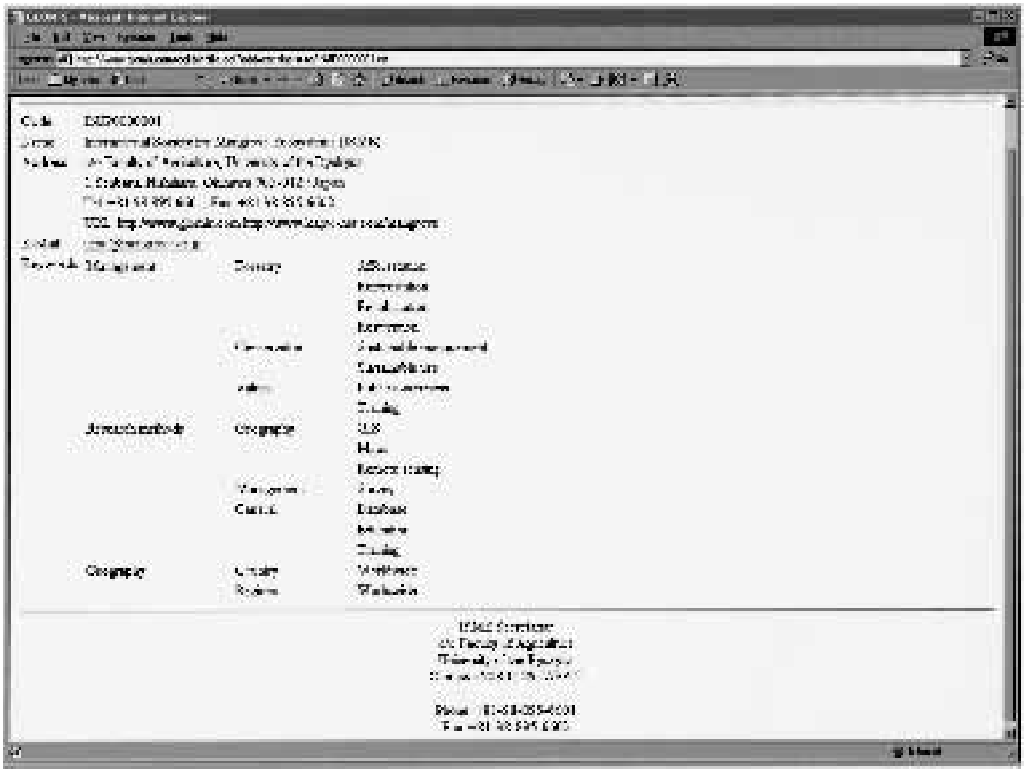


Fig. 4. Retrieved data from a category search on 'Institution'.

Conclusion and recommendations

The objectives of the first two phases of GLOMIS were broadly achieved. One major handicap was the inability of the poorly-resourced Regional Centres to cope with the quantity of work assigned to them, leading to frequent staff changes and its associated problems. It was also thought that the project would have been more beneficial to end users if it had indulged in assembling literature on all issues related to mangroves (i.e. establishing regional reference libraries) and not merely constructing a database of references, which in some cases could not be traced. Capacity development and information exchange have also not been the focus of the first two phases of GLOMIS even though these elements are very essential in the current developmental process. ISME therefore intends to further enhance GLOMIS by introducing the next phase that would be better financed and that would seek to broaden the outlook of the project to include other aspects that would make it more development-oriented. To achieve this, the proposed coming phase of GLOMIS, which will cover a period of two years, will focus on research, establishment of reference libraries, capacity development for stakeholders and end users, and information exchange through more effective networking.

The GLOMIS database is a valuable resource that needs to be continued and be kept up-to-date. The maintenance and updating of the server and management costs for regional centres and GLOMIS HQ are expensive. For the survival of the GLOMIS database, charging private companies and users who live in developed countries may need to be considered to pay for maintenance of the database. Although the original idea of GLOMIS was to establish a GIS-based system, this goal has not yet been fully met; it should be achieved, to the extent possible, in order to establish a truly stand-alone system. This may be done by providing digital and visual GIS-data such as mangrove distribution, productivity, and destruction in the world.

In addition, presently there is a tendency to view the database solely for the purpose of natural science study; it will be also necessary to strengthen the database by adding more data on local socio-economics and anthropological aspects of mangrove ecosystems. To achieve a real conservation and sustainable utilization of mangrove ecosystems, local community participation is indispensable. Compiling and disseminating recent trends of socio-economic and anthropological data will contribute to the capacity of local governments and communities to conduct and make their own mangrove management plans.

It is anticipated that at the end of the next two-year period the following outputs have been achieved:

- GLOMIS has been enhanced and made more available and valuable to end users especially policy makers, researchers and post-graduate students. This would be done by adding more current references to the GLOMIS references component and digital and visual data in a new component.
- Each GLOMIS Regional Centre plays a role as a reference point or a library facility stocked with available published and unpublished data on the distribution and productivity of mangrove species and major mangrove forest types worldwide.
- Targeted individuals, groups, organizations, institutions and societies in mangrove growing areas within the four regions where GLOMIS operates have increased their abilities to perform core functions, solve problems, and define objectives in relation to mangrove ecosystems conservation, sustainable management and rehabilitation.
- The targeted entities have been assisted to understand and deal with their development needs in a broad context and in a sustainable manner.
- Information exchange among stakeholders and effective networking in respect of mangrove conservation, sustainable management and rehabilitation have been enhanced within the regions where GLOMIS operates.

Finally, it is known that mangroves, seagrasses and coral reefs are integrated and interdependent coastal ecosystems. Conservation of one of these ecosystems requires the conservation of them all. For successful conservation, restoration and sustainable utilization of mangroves and their adjacent ecosystems, continuous networking and information sharing of researchers, policy makers, local communities and people involved with other coastal ecosystems will be important. Also the inclusion and sharing of data of both terrestrial and marine origin in the mangrove ecotone is a challenge.

Acknowledgements

We are grateful to the International Tropical Timber Organization (ITTO), for its continuous support and guidance to implement the GLOMIS project. We are indebted to the Government of Japan for its financial support to GLOMIS. We would also like to thank the Okinawa Prefectural Government for their support to the GLOMIS project. The GLOMIS project could not have been accomplished without the support from the following members: The coordinators of GLOMIS Phase I – Dr M. Vannucci and Dr L.D. Lacerda (also leader of the Brazil Regional Centre); GLOMIS Board members – Prof. K. Ogino, Dr F. Blasco, and Prof. E.S. Diop; leaders and staff of Fiji and India GLOMIS Regional Centre in Phase I and Phase II – Fiji, Mr M. Jaffar (leader of Phase I), Mr V. Amato-Ali and Ms E. Erasito (laders of Phase II), Ms A. Deo and Ms A. Ravuvu (staff), and India, Dr V. Balaji (leader of Phase I) and Dr P.C. Kesavan (leader of Phase II), Mr Senthil Kumaran (staff); previous staff of the Ghana Regional Centre include, Ms L. Oduro-Koranteng and Ms S.K. Yawson. Support has also been received from the webmaster – Ms S. Watanabe; ISME Executive Committee members; and ISME Secretariat – Ms N. Oshiro, Ms K. Yamashiro, Ms K. Horikiri and Dr L. Vidhanaarachchi. Although we cannot list all of the names, we are grateful to the generous personal contributors who have encouraged and provided us with their information and publications. Shigeyuki Baba is the Executive Secretary of ISME. Chris Gordon and Mami Kainuma are coordinator and assistant coordinator of the GLOMIS Phase II, respectively. Chris Gordon and Jesse S. Ayivor are also leader and current staff of the Ghana Regional Centre, respectively. Farid Dahdouh-Guebas is a Postdoctoral Researcher from the Fund for Scientific Research (FWO-Vlaanderen) in Belgium.

References

Costanza R., R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, S. Naeem, K. Limburg, J. Paruelo, R.V. O'Neill, R. Raskin, P. Sutton and M. van den Belt. 1997. The value of the world's ecosystem services and natural capital. Nature 387:253-260.

Dahdouh-Guebas F. 2002. Remote sensing and GIS in the sustainable management of tropical coastal ecosystems, Environment, Development and Sustainability Special Issue, Kluwer Academic Publishers, Dordrecht, The Netherlands. 145p.

Diop E.S., C. Gordon, A.K. Semesi, A. Soumare, N. Diallo, A. Guisse and J.S. Ayivor. 2002. Mangroves of Africa. p.63-121. In : Mangrove ecosystems: function and management. Lacerda L.D. (Ed.). Springer-Verlag. Berlin. Germany. 292p.

Fujimoto K. (in press). Below-ground carbon sequestration of mangrove forests in the Asian-Pacific region. In: Mangrove management and conservation – present and future. Vannucci M. (Ed.). United Nations University (UNU), Tokyo, Japan.

Field C. 1995. Journey amongst mangroves. The International Society for Mangrove Ecosystems (ISME), Okinawa, Japan. 135p.

Field C. (Ed). 1996. Restoration of mangrove ecosystems. The International Society for Mangrove Ecosystems (ISME), Okinawa, Japan. 250p.

ITTO/ISME. 2001a. Introduction to global mangrove database and information system (GLOMIS) Vol. I-VI. The International Society for Mangrove Ecosystems (ISME). Okinawa. Japan. 1060p.

ITTO/ISME. 2001b. Introduction to Global Mangrove Database and Information System (GLOMIS). The International Society for Mangrove Ecosystems (ISME). Okinawa. Japan. CD-ROM.

Lacerda L.D., J.E. Conde, B. Kjerfve, R. Alvarez-Leon, C. Alarcon, and J. Polania. 2002. American mangroves. p.1-62. In: Mangrove ecosystems: function and management. Lacerda L.D. (Ed.). Springer-Verlag. Berlin. Germany. 292p.

Rönnbäck P. 2001. Mangroves and seafood production. The ecological economics of sustainability. Akademitryck AB, Edsbruk, Sweden.

Spalding M., F. Blasco and C. Field. Eds. 1997. World mangrove atlas. The International Society for Mangrove Ecosystems (ISME), Okinawa, Japan. 178p.

Vannucci M. 2002. Indo-West Pacific Mangroves. p.123-215. In: Mangrove ecosystems: function and management. Lacerda L.D. (Ed.). Springer-Verlag. Berlin. Germany. 292p.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 15-24
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Integration of environmental datasets, formats and software in the IODE Resource Kit

Murray Brown

Phoenix Training Consultants
834 Elysian Fields Ave., New Orleans, Louisiana 70117, USA
E-mail: murraybr@bellsouth.net

Abstract

The Intergovernmental Oceanographic Commission's (IOC) International Oceanographic Data and Information Exchange Program (IODE) has developed an extensive suite of Web-based training materials called the IODE Resource Kit. A major section, entitled 'Data Analysis and Products,' has been designed to introduce data management trainees to global and regional environmental datasets, the principal formats used for their storage and distribution, and public-domain software for marine data quality-control and analysis. A principal aim of this section has been to achieve integration between the major datasets selected for IODE training, using existing pathways between them afforded by specific format compatibilities. It is now possible to identify a suite of (mainly) public-domain software programs that not only provide bridging functions between various databases, but also perform many analysis and quality-control functions. These interconnections are functionally illustrated by three functional schematics used in the IODE training curriculum. Every dataset used in the IODE workshops can be located on the diagrams, and easy paths can be traced to any desired software program. A set of 50 illustrated 'Roadmap Tutorials' has been developed to give step-by-step directions on the processes involved, including raw data entry, spreadsheet and relational database manipulations, grid-and-contour methods, and finally multi-parameter synthesis in Geographic Information System (GIS) applications. At the present time, most major formats and datasets of principal interest to hydrographers and chemical oceanographers are included in the schematics and in the IODE curriculum.

Keywords: Marine data management; Training.

Introduction

Since the late 1990s the Intergovernmental Oceanographic Commission has sponsored an international training program in marine database management (funded principally by Flanders and Belgium), later expanded to include marine information management. Focused primarily on developing countries, the program has based its training curriculum on the IODE Resource Kit (Reed, 2004), which contains extensive fundamental documentation and very practical workshop Tutorials. This paper discusses the central core paradigm for training in marine data analysis and data products, one of the three main components of the data portion of the Kit.

Earlier experience with the OceanPC Program in the early 1990s had indicated to the IOC trainers that the mere availability of data and software resources – even when a modest degree of integration between the units was available – did not lead to development of national data collections or of especially noteworthy product development among the former trainees. It was apparent that a very robust practical and theoretical structure of relationships among the various datasets, formats and software programs must be developed, and that the training to be based on this structure must vividly demonstrate the possible connections and processes through practical demonstrations of them in ‘real-world’ situations.

Toward this end, the IOC trainers have canvassed the Internet to catalog the best ocean data sources, their formats, and the freeware that can be used to edit, display and analyze them. Special attention has been paid to finding or writing format conversion programs that accomplish connections between important resources. Using a list of principal datasets, formats and software, the trainers have developed integration diagrams that show explicitly the connections between them, and have written a complete suite of training Tutorials that show how to move among and between the resources. This suite is actually a step-by-step manual for the creation of a national ocean data collection for Namibia, and for the creation of important data products from that collection.

OceanPC Program

The predecessor program to the Data Analysis and Products section of the Kit was the OceanPC training activity sponsored by the IOC in the early 1990s. Based heavily on the many modular programs for data management and quality control written by Harry Dooly at ICES, the system of software could accommodate the World Ocean Database files, and contained very good quality-control utilities. It lacked sophisticated data analysis capabilities, and there were some issues of user-friendliness that caused concern, particularly in the area of data editing. Basically, it consisted of a host of small programs that ‘communicated’ with each other through two common formats: the ICES standard profile format, and the ICES spreadsheet format. The only other major ‘external’ program that the system supported was SURFER, through the export of simple XYZ flat files for gridding. Although the ICES programs could achieve some degree of data sub-selection, by space and time coordinates, this process was not straightforward and somewhat inflexible. It was felt then that the time had come to consider full relational database management technology, which was not at the heart of the ICES programs.

Planning the ‘Next Generation’

In early 1995, the IOC convened a meeting of the OceanPC trainers to discuss possible future directions for marine data management training, with a special view toward the software problems described above. Two diverging philosophies emerged, named by their proponents The Central Engine (CE), and the Daisy-Chain (DC) Models.

Central Engine Model

The main feature of the CE model is that it requires a massive, all-encompassing software program (the ‘Engine’) that is compatible with all principal formats and can perform all desired QC and analysis work, in addition to exporting data and products in all import formats and other necessary formats (e.g. GIS images and shapefiles, spreadsheets). This model, shown in Fig. 1, suffers from the problems associated with planning and supporting the creation of the major software required, and the prospect that the main program would probably be obsolete as soon as it was published. Quite frankly, it was discarded as impractical.

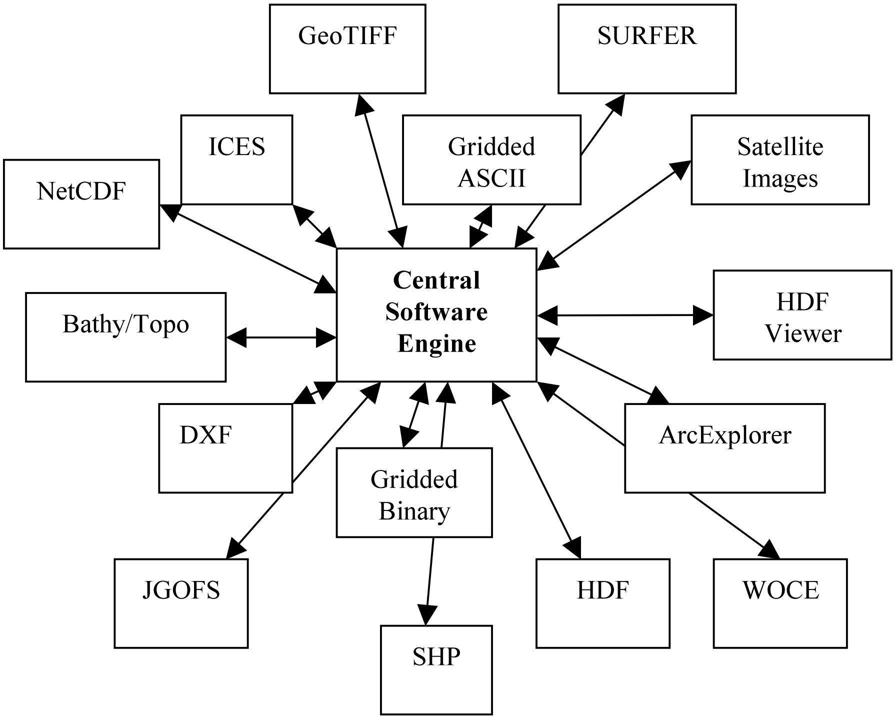


Fig. 1. Schematic representation of the Central Engine Model. The set of formats, databases and programs shown here is just representational, intended to show the profusion of connections required; it is not a faithful catalog of the resources considered at the time.

Daisy-Chain Model

The main feature of the Daisy-Chain Model is that it does not require any specific new software development, but instead relies on connections between existing application programs and on the future evolution of format compatibilities between these programs and major formats used in data publications. Its advantages lie in the fact that software development (except in the case of small format converter programs) is usually handled by other parties. Its disadvantages lie in the need for constant monitoring of the available software and formats in current use, in order to remain abreast of current developments.

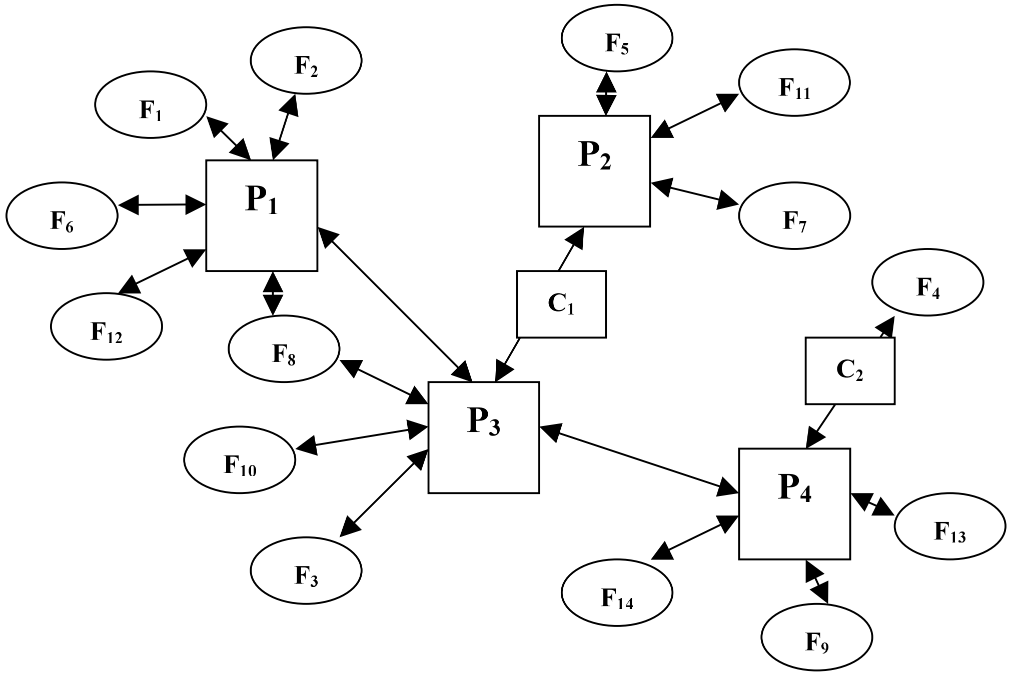


Fig. 2. Schematic representation of the Daisy-Chain Model. The 'system' is composed of various software programs (P), each with their own format (F) compatibilities, sometimes aided by special format converter programs (C).

Technology advances

Both because there was no budget at hand to continue the OceanPC development project, and because extremely rapid developments in Internet technology and the availability of new software seemed to require a strategic delay, nothing was done about system upgrade immediately after the 1995 meeting. A veritable explosion in the databases and software programs available to the ocean community occurred in the brief period after the introduction of Internet browsers (1994-6). Instead of searching high and low for resources, the IOC trainers (one of them an original OceanPC instructor) had the unexpected pleasure of wading through dozens of candidates.

New software included Ocean Data View (http://www.awi-bremerhaven.de/GEO/ODV/), an ocean database synthesis, quality-control and analysis program published by Reiner Schlitzer; Java OceanAtlas (http://odf.ucsd.edu/joa/jsindex.html), a similar program published by John Osborne; HDF Browser (http://www.intersys21.com/html1/hdf.html), a program for creating and viewing HDF files; HDFView (http://hdf.ncsa.uiuc.edu/hdf-java-html/hdfview/index.html), a program for viewing and exporting data matrices from HDF files; ArcExplorer (http://www.esri.com/company/free.html), a freeware GIS browser; and various upgrades of the popular SURFER gridding and contouring program (http://www.goldensoftware.com/demo.shtml). In addition a major portion of the ICES DOS-based software was upgraded to Windows format.

New database publications available either on-line or in CD-ROM format included the World Ocean Database 2001 (http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html); World Ocean Atlas 2001 (http://www.nodc.noaa.gov/OC5/WOA01/pr_woa01.html); satellite sea-surface temperature analyses from US NASA (e.g. POET [http://seablade.jpl.nasa.gov/poet/]) in global and regionally-subset form; the comprehensive WOCE dataset (http://www.nodc.noaa.gov/woce_v3/); numerous global relief datasets (gridded binary and ASCII) from the US NGDC (http://www.ngdc.noaa.gov); and very large sets of climate-related datasets from US NASA (http://redhook.gsfc.nasa.gov/~imswww/pub/imswelcome/plain.html).

It is difficult to list similarly any specific developments in the area of formats during this period, other than the usual committee-based development of new grand formats for international use by everybody, which activity is usually on-going, but has no practical impact. The mid- to late-1990s were marked, however, by a decrease in the publication of ad-hoc, new formats invented by CD publishers (e.g. the TOGA CD-ROM's ASCII chunk files and the WOCE Version 2 winds in binary grids), in favor of existing standard formats. HDF and NetCDF emerged as the formats of choice for satellite images and gridded climatological data. The ICES standard profile and spreadsheet formats dimmed in interest as users struggled with their eccentricities. The wild popularity of the World Ocean Database publications, however, did not focus interest on their format due to the ability of Ocean Data View to read the files directly, thus sparing the user the need to learn the ins and outs of stratified formats.

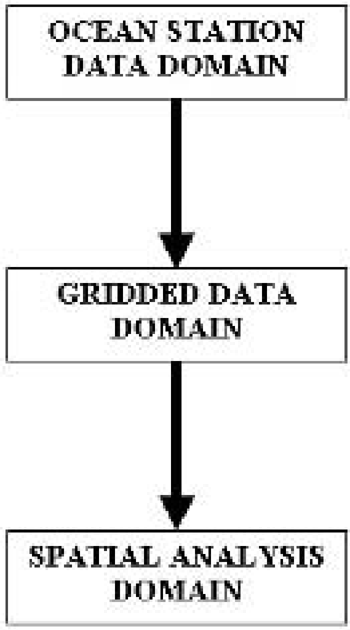
The bottom line to the technology explosion of the late 1990s was that the rapid availability of all these resources, amid the steady trend toward a smaller family of formats, simply presented itself as the solution to the problem. A ready-made Daisy-Chain Model of resources integration just appeared on the horizon, and it was greeted, adopted and fostered by the IOC trainers.

Emergence of the Resource Kit

The earliest use of a set of integrated resources was the IOC-supported Black Sea CD, used for marine data training in Sofia, Bulgaria, in 1998. It consisted of a simple one-page HTML interface to a set of programs, datasets and format descriptions. An integration diagram that showed the relationships between the formats was presented, but it was not included on the CD-ROM.

This first appearance of the integrated approach was followed by the use of CD-ROMs of increasing sophistications during the ODINEA and IOCINCWIO Marine Data Management Training programs for eastern African countries in the 1998-2000 period. As the pedagogic skills of the trainers increased, the total number of software programs decreased (from a high of over 40 to about 20 today), list of emphasized formats decreased (from 22 to 8), but the principal datasets used for training have remained constant at about 20. The early integration diagram has evolved into a set of three 'domain diagrams' in a separate section on Data Integration in the Data Analysis and Products part of the Kit. Beginning in 2001, manuals for each workshop have been prepared, primarily in view of the extreme size of the Kit (currently over 10,000 files). As well, tutorials have been included in the Kit, showing students how to navigate through the integration diagrams.

Integration diagrams



Although the Resource Kit's basic model for Data Analysis and Products is the Daisy-Chain Model, as described above, the many possible combinations and connections has made it necessary to break up the model into three 'domains,' as shown on the right. There are obvious areas of overlap, requiring a few programs or formats to appear on more than one diagram, but the division seems to successfully separate quite different areas of work into more-easily understood relationships. The domains portray the data analysis process as a sequential progression from 'data,' through gridding and contouring, to spatial analysis (GIS applications).

The Ocean Station Data Domain consists of the datasets, formats and software concerned with ocean survey data. It emphasizes the methods to create a national data collection from published and unpublished sources, and the methods to create 'data products' for analysis by other means (e.g. gridding and contouring in the Gridded Data Domain).

The Gridded Data Domain consists of the resources concerned with gridded data, and with Level 3 satellite images, which are – in a sense – visual representations of data grids on the earth surface. This domain also includes the mechanisms for creating special formats which can be used in the next domain.

The Spatial Analysis Domain consists of resources concerned with GIS analysis.

Ocean Station Data Domain

Two main 'routes' are emphasized in this domain: the pathway from raw data to Ocean Data View (or digital data capture), and the pathway from global archives to ODV (Fig. 3). The latter is easily navigated, due to ODVs built-in compatibility with principal formats. The former requires careful tutelage, due to the many different data reporting systems found in old hard-copy data or spreadsheets. Particular attention is paid to the 'units problem.' The role if ACCESS, as surrogate for all relational database management systems (RDMS), is entirely optional, because ODV currently performs such a wide range of data management functions that we have not found it necessary to use RDMS technology for our work. We do, however, provide training on methods to migrate data between ODV and a typical RDMS program if that is needed.

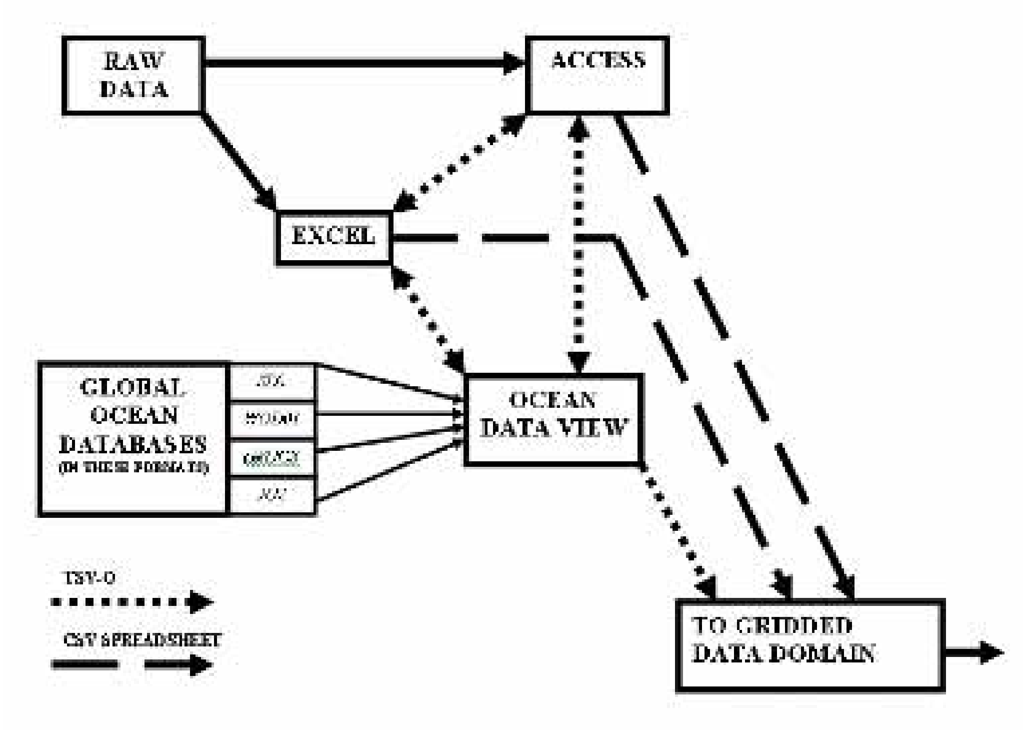


Fig. 3. Schematic diagram of the Ocean Station Data Domain. Training tutorials in the Kit emphasize the route from Raw Data to EXCEL to Ocean Data View, as well as direct data import to ODV from Global Databases and export from ODV of spreadsheet data for use in the Gridded Data Domain.

Gridded Data Domain

The Gridded Data Domain includes both data matrices obtained by gridding ocean station data and satellite images mapped to the earth's surface. Due to the limitation that the entire IOC training curriculum is focused on the ultimate use of ArcExplorer, which does not support image projection changes, the image mapping must be Cartesian (also called equirectangular or equatorial cylindrical equidistant). This Domain includes the enormous archives of HDF and NetCDF images and climatological means available from numerous online archives. Quite frankly, however, one of the most valuable formats in this Domain is the simple XYZ format used for transfer between programs. Data in XYZ format are ubiquitous on the Internet.

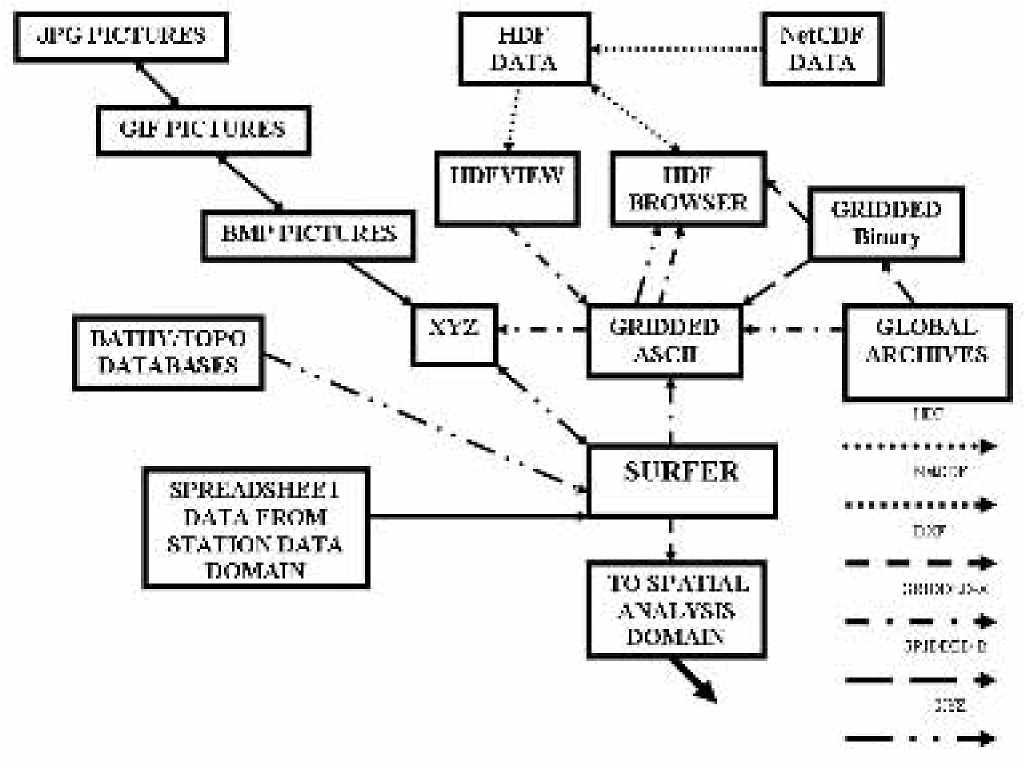


Fig. 4. Schematic diagram of the Gridded Data Domain. Tutorials in the Kit emphasize numerous interconnections between the many formats in this Domain. Some small format conversion programs are not shown here to simplify the diagram. SURFER has found extensive use as a format converter, in addition to its powerful grid-and-contour functions.

Spatial Analysis Domain

This Domain is concerned primarily with the synthesis of 'overlay' maps that combine data products and images from the Gridded Domain and from many global data resources. Students who have access to complete GIS systems can increase the scope of this Domain to include the use of images and vector files in other projections or coordinate systems. Students with only the resources provided in the Kit are limited to Cartesian display (e.g. X and Y in degrees, without projection). A perhaps surprising feature is the extensive use of the DXF format, an older workhorse interchange format in GIS work that is still widely supported.

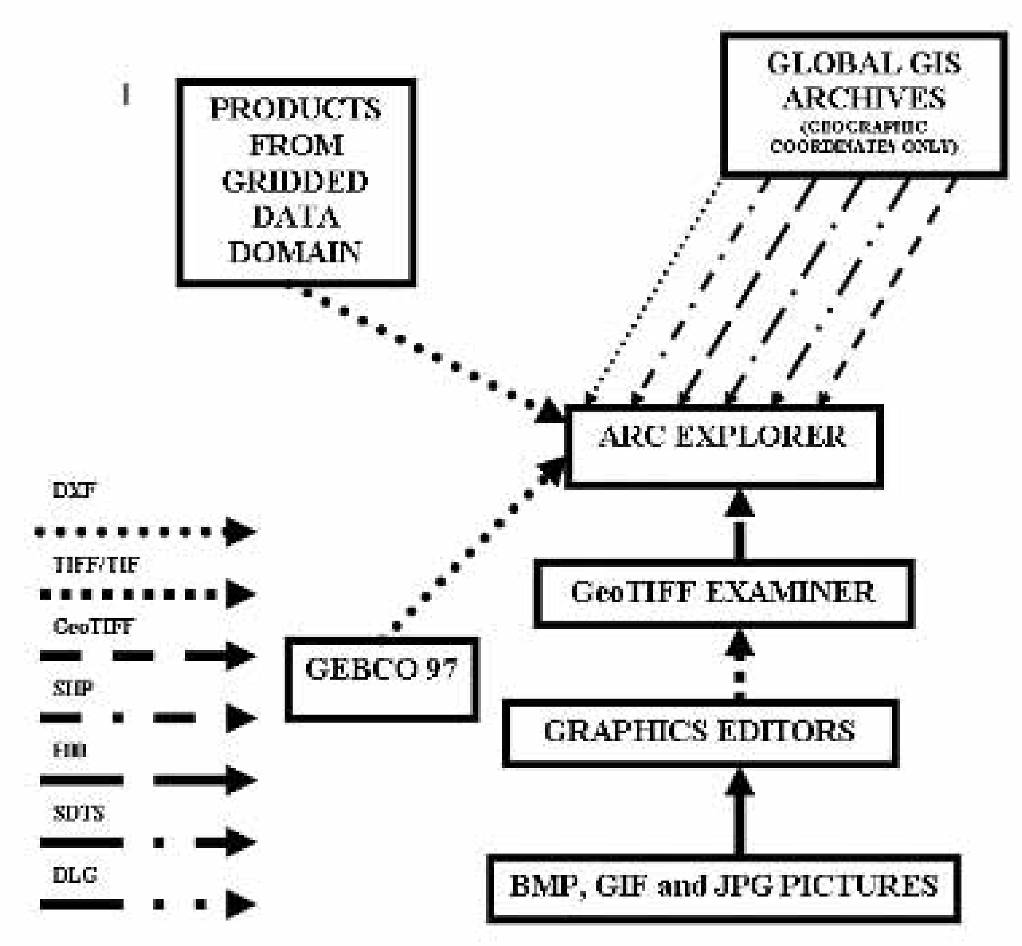


Fig. 5. Schematic diagram of the Spatial Analysis Domain. At this level the data products developed in the previous two Domains are combined with many different types of global resource data files in appropriate formats. Of special interest is the incorporation of GeoTIF images which often allows the use of actual photographic images for large-scale work (e.g. local analyses at ICZM levels of resolution).

Conclusions

The IODE Resource Kit has developed a methodology for the integration of a wide range of ocean data, data formats and applications software, using a set of three 'Domain' diagrams to illustrate the relationships. The marine data management training program of the IODE relies heavily on these diagrams to point students to documentation for the elements of the integration scheme, and to provide the big picture for the Roadmap Tutorials contained in the Kit.

Acknowledgements

This work of developing the IODE Resource Kit has been supported by the Intergovernmental Oceanographic Commission, the governments of Flanders and Sweden, and by details of employees from the governments of Australia and the United States. Special thanks are given to Mr. Peter Pissierssens who knows how to lead a team by letting them work in peace. For the development of the integration scheme, much is owed to the original OceanPC team, Mr. John Withrow, Dr Douglas McLain and Dr Harry Dooley.

References

IODE Resource Kit, n.d. http://www.oceanteacher.org. Intergovernmental Oceanographic Commission of UNESCO. Paris.

Reed G. 2004. OceanTeacher: building capacity in oceanographic data and information management. p. 39-44. In: Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002. Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). IOC Workshop Reports 188 (UNESCO, Paris). x + 308 pp. – also published as VLIZ Special Publication 16.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 25-38
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

The UNESCO Bilko Project: developing training capability for coastal and marine remote sensing

Craig Donlon[1], Ian Robinson[2], David Blackburn[3], Alistair Edwards[4], Malcolm Dobson[5], Tjeerd Hobma[6], and Jesus Morales[7]

[1]European Commission Joint Research Centre, Institute for Environment and Sustainability
Via Enrico Fermi, I-21020, Ispra, Italy
Present address: International GODAE SST Pilot Project Office, Hadley Centre, Met Office, Fitzroy Road, Exeter, EX1 3PB, United Kingdom
E-mail: craig.donlon@metoffice.com

[2]University of Southampton School of Ocean and Earth Science
European Way, Southampton, United Kingdom

[3]Buis les Baronnies, France

[4]Centre for Tropical Coastal Management, University of Newcastle
United Kingdom

[5]Scotland On-Line, Technology Park
Gemini Crescent, Dundee DD2 1SW, United Kingdom

[6]ITC, Department of Water Resources
Hengelosestraat 99, PO Box 6, NL-7500 AA Enschede, The Netherlands

Abstract

This presentation reviews the UNESCO Bilko project that commenced in 1987 and continues today. The primary aim of the project is to make remote sensing training materials accessible to those without specialist resources at their disposal and to promote good teaching practice by tapping the diverse skills and expertise of an expert community. Considerable resources have been generated by the project including a DOS- and Windows-based image processing software package. Pedagogical materials include a wealth of short self-study lessons focused on a particular remote sensing technique, oceanographic phenomenon or sensor that students can work through in their own time. Collectively, the Bilko project provides a remarkably diverse but comprehensive resource for teaching coastal and marine remote sensing. Recently, the project has adopted a thematic framework in order to deliver more focused material and to keep pace with rapidly evolving remote sensing sensors, platforms and algorithms. The project currently serves some 1900 users located in over 70 countries and has supported several international workshops and training courses with both teaching materials and expertise. Several networks have been developed that are monitored by the Bilko steering committee including a network dedicated to Bilko lesson authors and a network for Bilko users. All material is available from the Bilko project website (http://www.bilko.org) or from the Bilko Project Office, ITC, Enschede, The Netherlands.

Keywords: Coastal and marine remote sensing; Training; Capacity building; Image processing; Oceanography.

Introduction

The considerable international investment in remote sensing technology over the past 20 years has significantly increased our capacity to measure, predict and manage changes in the marine environment. Remote sensing measurements constitute an unparalleled data resource providing a unique perspective in both time and space. A variety of sensors on satellites and aircrafts can measure the sea surface temperature (which can also show current patterns), ocean colour (from which the chlorophyll concentration can be derived, and hence the phytoplankton distribution), variations in sea level (from which ocean currents can be estimated) and surface roughness (giving surface waves and derived winds). Arguably, remote sensing has resulted in a paradigm shift in the way that governments and scientists view the marine environment.

The application of remote sensing to address real-world problems is of paramount importance. In turn, improvements in strategic decision-making and the development of wider national and international marine public policy in the coastal and marine environment can be expected. But to achieve this vision, the benefits and possibilities offered by remote sensing must be familiar to a broad community including, for example, ocean scientists and engineers, marine, freshwater, and coastal planners, natural resource managers, policy makers, industry leaders, local, national and international agencies, non-governmental organizations, educators and students. For many, remote sensing is no longer a specialist technology pursued by a few, but is simply another, albeit extremely powerful, observational tool. But, in some areas of the world this is not the case and significant barriers still exist to the effective use of remote sensing in the marine environment. If programs, such as the Global Ocean Observing System (GOOS), that rely heavily on remote sensing in their regional implementation are to realize their full potential, shortfalls in both human and material remote sensing capacity need to be addressed.

The UNESCO Bilko project was initiated in 1987 under the Marine Sciences Training and Education Programme (TREDMAR) of UNESCO to develop training capability in coastal and marine remote sensing through a series of computer-based learning (CBL) modules. The project has provided eight modules of computer-based lessons to over 500 marine science laboratories and educational establishments and over 1900 individual users in over 70 countries around the world. The Bilko project continues to innovate and now through its user network, provides a virtual global faculty for coastal and marine remote sensing.

This paper provides an overview of the UNESCO Bilko project, its approach, client base, scope, accomplishments, and its strategic vision for the sustainable development of marine and coastal remote sensing capacity.

The UNESCO Bilko approach to building capacity in coastal and marine remote sensing

Marine and coastal remote sensing is a particularly complex subject requiring knowledge of physics, biology, oceanography and, computer science, amongst other fields. The remote sensing laboratory resides in the virtual world of a computer, and the benefits of remote sensing can only be explored by those having at least a basic grasp of all the above disciplines. This is true for both students and teachers. Students can be deterred by the prospect of challenging and detailed background study. Teachers, not familiar with remote sensing or having limited access to appropriate resources or infrastructure, may choose not to provide instruction in remote sensing.

When UNESCO considered these issues in the mid 80s, the interpretation of satellite images could be undertaken only on large computers using expensive commercial image processing software that typically had long learning curves. Satellite and aircraft images were often seen as having military importance, so access to the equipment and technologies for teaching image interpretation was privileged and even restricted. The original aim of the Bilko project was to facilitate 'hands-on' training in coastal and marine remote sensing for those traditionally excluded from such training. Three recurrent themes focused the project:

(i) The interpretation of satellite images is a skill of great international, strategic and economic importance which should be developed widely and distributed globally;
(ii) UNESCO should take a role in the development of new media of instruction, specifically the use of small computers to teach topics normally denied to those having no access to large computers or local teaching facilities; and
(iii) UNESCO should seek to improve the quality of teaching in selected fields by demonstrating what can be done by distance-education methods.

Rather than follow the conventions of the day, UNESCO chose to back a project based on personal computers (PCs) which were then beginning to emerge in large numbers. A specification for the target recipients was derived using a 'lowest common denominator' approach. It was assumed, for example, that recipients would be relatively isolated from others working in the field but have access to low-performance, standalone PC computers. Only marginal teaching support and access to reference materials would be available. Four primary actions were agreed for the project:

(i) A software tool for the display and interpretation of remotely sensed data that allows the topic to be taught using low cost computer terminals rather than specialized equipment would be developed and made freely available.
(ii) As the software on its own would have been of marginal benefit, it should be supported by lessons that exemplify its power and teach the fundamentals of image analysis and interpretation by allowing users themselves through 'hands on' practical exercises to explore different remote sensing data.
(iii) A distribution system would be developed to bring the software and accompanying lessons to anyone with an interest in coastal and marine remote sensing. Copyright of all project material was vested in UNESCO and agreement was reached such that recipients were encouraged to copy Bilko material for students and colleagues – the sole restraint being that it should not be sold.
(iv) A means for teachers and students to exchange, assess, criticize and enhance Bilko software and lessons would be developed and maintained to provide feedback to lead appropriate innovation and development.

The substance of these original actions is still valid today and each is maintained within the Bilko project at all levels.

The Bilko image processing software

Due to the severe constraints imposed by the typical PC computer systems that used the DOS operating system, Bilko functionality was kept deliberately modest. The Bilko software was originally developed for the 4-bit (16 colour), 640 x 350 pixel format of the Enhanced Graphics Adapter (EGA) display system. Rapidly, this was superseded by more powerful Video Graphics Array (VGA) display systems providing 256 greyscale/colour levels and a full screen 512 x 512 pixel display. Bilko was rewritten to embrace this enhanced functionality that provided considerable power. The DOS-based Bilko software that emerged was a fast, menu driven image processing toolkit that allowed the display and analysis of remote sensing data at a pixel resolution. The toolkit had inbuilt context-sensitive help screens to guide students and provide help for all of the functions provided with the software. Considerable attention was given to making the software extremely user-friendly so that students with no previous experience of computers would not be intimidated. Indeed, emphasis was placed on having an impressive image on the computer screen as soon as possible within every session in order to maintain student confidence and interest. In this way, attention was focused on the image data rather than on the software techniques that had placed it there (Robinson *et al.*, 1993). At that moment in time, the Bilko toolkit was a small revolution, in the sense that satellite and airborne digital images could now be analysed on a desktop computer by anyone.

The original DOS-based Bilko image processing toolkit has now been developed from a modest toolkit into a professional image processing environment called Bilko for Windows running under the Microsoft Windows operating system. The enhanced capabilities of the Windows operating system have allowed the Bilko software team to provide considerable functionality comparable to many commercial image-processing systems including tri-band colour composite, geometric rectification, a powerful image calculator and the ability to handle many image data formats and image data sets all at once. Bilko for Windows supports various common data file formats including HDF, 8-, 16- and 24-bit flat binary files, and calibrated GIF files. Although the Bilko team deliberately target 8-bit compressed file formats to ease the burden of data storage and transfer, the team also recognized that a number of standard data file formats are operationally used to supply data to users that should be supported (e.g. HDF).

The migration of Bilko from DOS to Windows has not compromised the original philosophy of the project, and the new functionality offered by the Windows system were harnessed and an innovative document interface was developed in which the navigation model, menus, icons, layout and colours are all intuitive and supportive of the user's goals. The document interface concept allows users to conceptualise operations that can be performed on a particular image data set. For example, an 'image document type' groups image handling operations together; a 'stretch document type' groups different contrast enhancement operations together. Table I describes the main Bilko document types used within the Bilko for Windows system.

Fig. 1 presents a demonstration Bilko for Windows session screen showing the use of traditional menu and toolbars characteristic of the Windows interface. Users are instantly presented with an interface that is familiar and follows the style of other Windows applications so that those with some Windows experience are confident and secure in the environment.

Bilko allows operations to be performed on multiple images that are connected (associated) together within the software. Each image is assigned a variable number which is used to refer to the image when writing mathematical formulas. This exceptionally powerful feature allows complex multi-image algorithms to be applied and even tested. It also provides the fundamentals for basic Geographical Information System (GIS) operations. In this way, the Bilko software can serve the needs of many people already familiar with remote sensing and perhaps working at an advanced level either in research or in the applications sector without having to invest in expensive image processing systems. This is an important feature as it allows Bilko to penetrate beyond basic remote sensing training needs contributing to material capacity at a local level.

Table I. Main document types and functions offered by the Bilko 2000 software

Bilko document types	Associated functions
Image	Filters, rectifications, re-sampling create image associations, image zoom functions
Stretch	Edit and apply a number of different lookup table stretches
Palette	Generate, edit and apply colour lookup tables
Filter	Specify and apply up to a 15x15 kernel digital filter
Formula	Use a text editor to script and apply complex mathematical operations to image data
Table	Generate ground control point files for remapping images to geographic projections, shows matrix results of covariance and correlation
Scatter	Generate a scatter plot derived from two connected (associated) images
Histogram	Produce a histogram distribution of image data
Set	Allows users to manage a collection of images as a ‘stack’ (allowing ‘data coring’ operations such as time series analyses, Hovmuller plots) or as a ‘tile’ allowing combination analyses

The basic functionality of the software system is supported by a specially written introductory tutorial that guides students through the various functions of the Bilko software. This is in effect the Bilko ‘user manual’ but is not written in the conventional style of a manual. Instead it follows a particular pedagogical style that has been the hallmark of all Bilko lesson material and, arguably, the root of the project success.

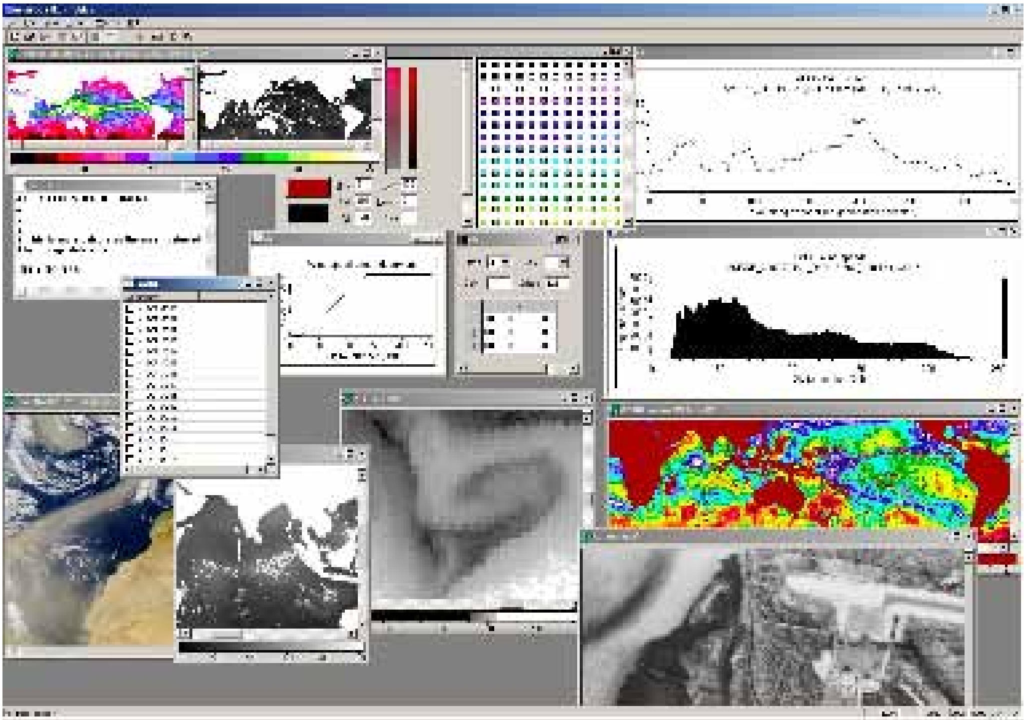


Fig. 1. The Bilko for Windows screen demonstrating many of the functions offered by the software.

Bilko lessons and modular structure

The Bilko project distinguishes between lessons and modules. Lessons constitute ‘A focused learning experience dealing with a single topic which can be completed within 1-3 hours.’ Initially, in the DOS-based era, Bilko lessons were provided in paper copy format and students would use the paper lessons together with the computer software. The Windows environment has allowed lessons to be written in the hypertext format providing an altogether different medium to present lesson material. Pop-up windows, hypertext jumps, interactive menus and questions are all possible. Furthermore, images, video clips, sounds and any other multi-media object can in principle be included within the lesson as required. However, experience has shown that writing in a hypertext environment should be treated with care and attention because many students still prefer to print a copy of the lesson and preserve computer screen ‘real-estate’ for using Bilko for Windows. In this sense, we have discovered that the use of Adobe Portable Document Format (PDF) lessons are as useful a medium as hypertext, and many media objects can still be included in the lesson but the lesson material can also be printed in a useful format.

Bilko lesson format

It is worth exploring the structure and format of Bilko lessons which, over the years, have proved exceptionally powerful in terms of generating material, editing for quality and, delivering models of good teaching practice. It is the consistent high quality and pedagogical style of Bilko lessons that set this project apart from other similar image processing toolkits and systems. Bilko lessons are designed for students and teachers working on their own with limited technical support or backup. Distance education demands that everything required to complete a lesson must be self-contained within a lesson or a module. Consequently lessons must be complete (as far as practically possible), written in a direct but clear style. They should foresee the questions that will be asked by students and provide the answers in the correct format and place. Feedback loops must be included to remove weaknesses of presentation and to focus student attention. Reasoning should be repeated which, although redundant, considerably helps understanding - a technique we refer to as 'constructive redundancy' (Blackburn and Edwards, 2002). In the absence of an instructor to question performance, lessons should be written as 'active text' to promote active learning by interspersing questions and other activities throughout the lesson. All questions should be provided with model answers and an indication of how they were obtained. Lessons call for considerable effort on the part of authors, and the Bilko development team has written a number of reference documents to facilitate their production (Blackburn and Edwards, 2002).

Lessons are edited by a Bilko editorial team to a common standard and put into a common format that is considered essential. One of the few inflexible rules of the Bilko system is that every lesson must begin with a statement of aims, objectives and a brief description of the lesson. This is followed by the lesson itself and then by a summary of the lesson. The repetition is deliberate to consolidate the knowledge gained during the course of the lesson.

The lesson 'aim' provides a statement of the aspirations that the lesson author holds for students working with the lesson. For example:

The aim of this lesson is to introduce the basic concepts of digital image analysis.

The aim sets out in global terms the desired outcome of the lesson and reveals what is educationally distinctive about one lesson in comparison to others. It allows a student or teacher to select the most appropriate lessons for a given purpose.

The lesson 'objectives' specifies tasks that students should be able to undertake as a result of studying the lesson. For example:

After completing the lessons in this introduction, you should have learned to:
1) Run the Bilko for Windows software and load digital images for analysis;
2) Select and apply the image analysis and image modification routines of the Bilko for Windows system;
3) Call for Help files when in doubt about any image handling procedure.

The lesson 'objectives' allows students to discover if they have satisfied the requirements set out under the heading of aims. Together with a short description of the lesson, the Bilko lesson 'aims and objectives' presents the student and teacher with a succinct appraisal of the lesson.

This allows teachers especially to review rapidly and to select appropriate Bilko material for inclusion within their own course structures.

Bilko modules

A Bilko module is simply a collection of Bilko lessons. A thematic module is a collection of lessons focused around one theme such as a subject area or a region.

Table II. UNESCO Bilko computer based learning training modules in applications of satellite and airborne remote sensing.

Module	Title	Reference
1 (DOS)	Some marine applications of satellite and airborne remote sensing. A computer-based learning module	MARINF/70. Unesco, Paris. 1989. 90p.
2 (DOS)	Applications of marine image data. Second computer-based learning module	MARINF/81. Unesco, Paris. 1991. 85p.
3 (DOS)	Applications of marine and coastal image data from satellite, airborne and in-situ sensors. Third computer-based learning module	MARINF/83. Unesco, Paris. 1992. 101p.
4 (DOS)	Applications of marine and coastal image data from satellite, airborne and in-situ sensors. Fourth computer-based learning module	MARINF/90. Unesco, Paris. 1993. 102p. (also available in Russian)
5 (DOS	Aplicaciones de datos de imagen costeros y marinos provenientes de satelites, aviones y sensores in situ. Quinto modulo de aprendizaje sobre base informatica	MARINF/96. Unesco, Paris. 1994. 141p.
5 (DOS)	Applications of marine and coastal image data from satellite, airborne and in-situ sensors. Fifth computer-based learning module	MARINF/96. Unesco, Paris. 1996. 110p.
6 (WIN)	The first Bilko for Windows: an experimental module	CD-ROM available together with all other modules from the UNESCO Bilko project office on request
7 (WIN)	Applications of satellite and airborne image data to coastal management	CD-ROM available together with all other modules from the UNESCO Bilko project office on request
8 (WIN)	Applications of remote sensing to fisheries management	CD-ROM available together with all other modules from the UNESCO Bilko project office on request

Table II. (cont.)

9 (WIN)	Applications of remote sensing to coastal erosion	In preparation expected mid 2003. Contact Dr Tjeerd Hobma. (hobma@itc.nl)
10 (WIN)	Vulnerability through the eye of a satellite	In preparation expected mid 2003. Contact Dr Craig Donlon (craig.donlon@jrc.it)
11 (WIN)	Applications of remote sensing to coastal management in India	In preparation. Contact Dr R. Sudarshana. (rsudarshana@hotmail.com)
12 (WIN)	Applications of remote sensing to natural hazards	In preparation. Contact: Dr Cees van Westen. (westen@itc.nl)

Each module is a self-contained package of:

(i) Image processing software (Bilko);
(ii) Introductory tutorial on how to use the software;
(iii) Lessons on the application of remote sensing to oceanography or coastal management; and
(iv) Satellite and airborne remotely sensed images to accompany the lessons.

Eight Bilko modules, containing over 60 individual lessons, have now been produced and distributed by the Bilko project in the period 1989-2002 (described in Table II). Several more thematic modules are now in preparation.

The concept of modules is particularly powerful in terms of building capacity at a regional level because local knowledge, resources and expertise can be used to train local students on specific issues. The development of a Bilko Module is a considerable task in itself, but the benefits in terms of establishing an active community and network of users that become engaged in developing the module, together with the support of the UNESCO Bilko, are immense. For example, the Global Ocean Observing System (GOOS) is considering the development of regional GOOS Bilko modules to build capacity in this way.

Project management

The UNESCO Bilko project has been steered from its inception (1987) by Dr Dirk G. Troost of UNESCO, assisted by a small team operating at different international institutions and convening on an ad-hoc basis. In 1998, a formal Bilko Steering committee was established and a Secretariat based at the Division of Applied Geomorphological Survey at ITC in The Netherlands, which now co-ordinates the day-to-day running of the UNESCO-Bilko Project. The international secretariat and editorial board maintain pedagogical standards, coordinate module development, production and distribution of Bilko material, manage client feedback, and strategic development. In particular, the secretariat maintains a comprehensive database of statistics relating to the Bilko project as well as collating all user feedback and project reporting.

Dissemination of UNESCO Bilko material

The first Bilko module was released with a series of 5 ¼' floppy disks and was a considerable success. Several distribution media types were considered before the floppy disk was chosen, as reviewed by Troost *et al.* (1991). The floppy disk was successful then because every PC computer system had the capability to read and write to this medium, they were cheap to duplicate in large numbers, and they allowed for modifications and updates to the material as required. The major limitation was storage space. As the copyright of all Bilko material is vested with UNESCO, all lesson material is open to modification, improvement and extension by teachers for their own purposes. Each lesson is provided in electronic format both as text and as a word processing document to facilitate this process. Such an 'open-source' approach to the use and development of Bilko lesson material is considered an important element of the project as it provides a fundamental mechanism for building confidence and capacity at a local level (Troost *et al.*, 1991). Today, Bilko modules are provided as a CD-ROM disk that provides a large amount of storage capacity, so much so that all of the Bilko modules are included on Bilko CD-ROM distributions.

At the time Bilko migrated from DOS to the Windows environment, the Internet was expanding exponentially. The development of web-browser technology initially appeared to provide a new form of dissemination and teaching that Bilko could harness. The Bilko team invested some time and energy towards understanding what could be achieved using a combination of Internet-based lessons while the Bilko software was operated locally. An advanced lesson, following the pedagogical style of Bilko, was written in the hypertext mark-up language (HTML) so that the web browsers could be used to display Bilko lessons. Many of the new features offered by the Bilko for Windows software were employed in the lessons, and student answers could be submitted to the lesson authors via the web browser. The authors could then assess the students' performance and return a marked appraisal of the students' performance by an e-mail. In theory, image data sets, lesson material and even the Bilko software itself could be obtained via internet transfer. The new Internet-based Bilko system appeared to be ideal. However, despite widespread promotion of the new Internet lesson within the Bilko community, the new lesson was used only by a small number of students having high bandwidth Internet connections and resident in the United States.

Consequently, the Bilko project decided that until Internet connectivity was better established, all Bilko material would be made available in CD-ROM format. As Internet connectivity grows throughout the world, the Bilko project will initiate further steps to harness and make full use of this new capability. However, at this present time, extensive use of the Internet to provide distance-learning modules is, in our opinion, premature. A virtual teaching laboratory based on the use of the Internet alone cannot provide training to a large number of countries where Internet access is unavailable or at best, sufficient only for small data transfer (e.g. e-mail).

Feedback: the currency of success

Feedback is essential to the Bilko project. Because all Bilko material is provided free of charge, there is no traditional measure of success in terms of sales or number of modules distributed. Feedback provides the main evidence that the project goals are being achieved, justifying the financial and human investments made by many who have contributed to the project free of charge. Feedback format is varied. In many cases it is an e-mail or a letter thanking the team for their endeavours whereas in other cases it is a reference to the Bilko project in a report or journal paper. Soliciting and quantifying the value of feedback on any basis is not easy.

The approach adopted by the Bilko project requires that each recipient of Bilko material returns a feedback questionnaire provided within the distribution. On return, this information is entered into a database that is periodically reviewed and updated by the Bilko secretariat. Originally, a traditional letter was sent to each member identified in the database. This has now been largely replaced by the use of e-mail where possible. It should be noted that where e-mail communication is not available, communications are still sent via personal letter. If a recipient does not reply, it is assumed that either they have moved on or are no longer in need of the Bilko project. Suggestions and requests made by users of Bilko are always treated seriously and are forwarded to the appropriate member of the Bilko Steering committee for consideration. Software requests and suggestions for improvements are reviewed and considered against the general philosophy of the project which is to provide new functionality only if the existing functions cannot be used in some combination to achieve the same result. Bilko is about teaching the tools, techniques and methods used in remote sensing, it is about exposing their details and not about providing closed 'black-box' solutions. There are many other image processing systems that are eloquently able to provide this latter type of functionality.

As the Bilko project has matured, a global network of users has developed in which many of these users themselves become authors of lessons in subsequent modules through interactions and feedback loops external to the 'formal' Bilko project. The wealth of accumulated knowledge and information, together with the generation and exchange of educational material within the Bilko project user network is viewed as a considerable achievement.

A virtual global facility for marine and coastal remote sensing: the UNESCO Bilko project

Today, the UNESCO Bilko project provides a framework in which individuals, teachers, projects and institutions can interact and work together creatively to build and develop capacity for the application of remote sensing in the coastal and marine environment. Bilko has grown from an experiment into a network of students, teachers and scientists connected by a common interest in the use and application of remote sensing images. For some, it has been the ability of Bilko to reveal the beauty and fascination of digital imagery of the ocean and coastal zone that ties them to the project. For others it is a means to an end; a way to teach remote sensing, a source of course material, part of a university degree, a dedicated training workshop or a daily tool that is used to analyse and interpret image data for research or decision making. For whatever reason people are involved with the Bilko project, it is true to say that collectively, they represent a global network.

Viewed from a different perspective, Bilko now constitutes a virtual global faculty for marine and coastal remote sensing. It has built confidence and knowledge in the use of remote sensing, abolished the view that remote sensing is the domain of wealthy countries and exposed the strategic benefits of remote sensing in the marine and coastal environment. As no charge is made for any material, the Bilko Project sees itself as an 'honest broker' of capacity building initiatives without vested interest. This self-perception is justified considering that the project has largely developed on a volunteer basis, both within the secretariat and in relation to the many lesson authors, translators, software programmers and teachers that have contributed to its material free of charge.

The ultimate target of the Bilko program is to place the information, knowledge and the power of remote sensing into the hands of the people who actually live and work in the areas so often imaged by satellites and aircraft systems. In this model, capacity can be built at three related levels:

(i) At the individual level providing standalone tools and data for teaching of remote sensing and its application in the marine and coastal environment,
(ii) At the entity level (e.g. a university department, government laboratory or training workshop) where large numbers of students may be reached and teachers themselves can be trained (training of trainers) using Bilko materials,
(iii) At the system or regional level (e.g. national and international projects) where thematic material can be produced at a local level within the overall framework of Bilko. Here, the creation of self sustaining, user-producer networks that build capacity from within but focused on specific national issues may flourish. This is the so called 'user-producer' network that, armed with the substantial experience and resources provided by the Bilko project, has become a reality in several cases (Robinson *et al.*, 1994).

At the individual level, the Bilko project database holds records for over 1900 subscribers actively engaged in the project (as indicated by a positive response) in over 70 countries. The UNESCO Bilko website server (http://www.bilko.org) currently has a minimum of 5 downloads of Bilko software and lesson material per week. However, because all Bilko project material is freely available and the project encourages its duplication and dissemination external to the project itself, there are probably many more users of Bilko than we report here.

At the entity level, Bilko has been used to give hands-on experience of image processing and to teach practical aspects of remote sensing to both bachelors and masters students at universities for over ten years. At the University of Newcastle, Module 7 is used with students reading for a MSc in Tropical Coastal Management whilst oceanographic lessons are used for students reading for a BSc in Marine Biology. At the University of Southampton School of Ocean Sciences, Bilko is used in both MSc and BSc courses. Typically, dedicated lesson material is prepared for particular courses, yielding many valuable lessons produced by groups of teachers and students. Other courses known to have made use of DOS Bilko modules include those held in Nairobi, Kenya; La Réunion; Caracas, Venezuela; Ensenada, Mexico; Enschede, The Netherlands; Lepe, Spain and Rio Grande, Brazil. In the case of the course held in Lepe, Spain, this course led to the preparation of Bilko Module 5.

More recently, several successful remote sensing short-courses, one in Zanzibar and one in Thailand (see Fig. 2), have utilized Bilko for Windows to provide targeted training in specific areas. The first short-course was an IOC-UNEP Regional Training Course on Seagrass Mapping using Remote Sensing which was held at the Institute of Marine Sciences, Zanzibar in November 1997. Participants from East Africa and international organisations in the region attended the course. The workshop concentrated on the application of remote sensing to coastal management with a special focus on the mapping of seagrass areas and seagrass standing crop. The second course formed part of an International Ocean Color Coordinating Group (IOCCG) Training course on Applications of Marine Remote Sensing which was held at the Asian Institute of Technology (AIT), Klong Luang, Thailand in November 1999. This was the third

IOCCG training course set up to increase the user base of ocean-colour data throughout the world. Some 30 participants from 12 countries in the SE Asian region attended the course. Several Bilko lessons focussed on the Indian Ocean using thermal infrared, colour and altimeter radar data of sea surface temperature, chlorophyll and, sea surface height were specially written for this course.

Fig. 2. Students at work using Bilko for Windows. (a) Course participants from mainland Tanzania, Zanzibar and Kenya work on Bilko Module 7 practical lessons at the Institute of Marine Sciences, Zanzibar (November 1997). (b) Course participants work on dedicated lessons designed for the IOCCG training course at the Asian Institute of Technology, Bangkok, Thailand (November 1999).

The Bilko project is entering a new phase in terms of developing capacity at the system or regional level. International organizations such as the Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM), Global Ocean Observing System (GOOS) and Global Climate Observing System (GCOS) all have dedicated working groups to develop appropriate and sustainable capacity in order to implement their programmes successfully. Marine and coastal remote sensing has a fundamental role within globally oriented programmes and an identified priority action is for training in coastal and marine remote sensing. The UNESCO Bilko project provides an ideal framework in which coastal and marine remote sensing capacity can be built in a sustainable manner at this level, allowing for the training of individuals and teachers and, through the modular framework providing a means for regional groups to develop their own targeted teaching material that can also be used in other areas of the world. It is hoped that the Bilko project will play a mutually beneficial role in the development of marine remote sensing by entering into close relationships with international organizations.

Summary

The UNESCO Bilko project has developed and matured, since its conception in 1987, within the UNESCO TREDMAR programme. It now provides dedicated computer-based learning materials for the teaching and study of coastal and marine remote sensing to over 1900 registered users distributed globally within 70 countries. It has provided a means for students and teachers to approach and work within the complex subject of marine and coastal remote sensing, generating confidence and knowledge that ultimately results in better analysis and decision-making for the marine and coastal environment. The Bilko project will continue to grow and, in the new era of 'operational' oceanography, integrated coastal management and Internet connectivity, will continue to evolve in order to provide foundations necessary to build a sustainable capacity for marine and coastal remote sensing.

References

Blackburn D. and A.J. Edwards. 2002. Technical notes for Bilko authors, available from the Bilko project office. ITC, Netherlands.11p.

Blackburn D. and A.J. Edwards. 2002. Guidelines for Bilko authors, available from the Bilko project office. ITC, Netherlands. 10p.

Callison R.D., I.S. Robinson, A.P. Cracknell and D. Cummings. 1989. Some marine applications of satellite and airborne remote sensing – a computer based learning module. MARINF/70 UNESCO (Mar. Sci. Div.), Paris, 90p.

Troost D., D. Blackburn, I. Robinson and R. Callison. 1991. A PC-based textbook for the marine sciences. Case study in global collaboration. Impact of Science on Society 164:339-351.

Robinson I.S. 1988. Education and training for marine science applications of remote sensing. Int. J. Rem. Sens. 9:1847-1857.

Robinson I.S., D. Troost and D. Blackburn. 1994. Developing training capability in marine remote sensing through computer based modules serving a global network. Proc. WMO/IOC technical conference on space-based observations, WMO/TD-No 649:280-286.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 39-44
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

OceanTeacher: building capacity in oceanographic data and information management

Greg Reed

Intergovernmental Oceanographic Commission (IOC) of UNESCO
1, rue Miollis, F-75732 Paris Cedex 15, France
E-mail: g.reed@unesco.org

Abstract

Within the framework of the International Oceanographic Data and Information Exchange (IODE) program, the IOC has developed a capacity building program to train data and information managers in developing countries with the objective of establishing and strengthening National Oceanographic Data and Information Centres. As part of these capacity building activities, IODE has developed OceanTeacher, a comprehensive self-training and resource tool for oceanographic data and information management.

OceanTeacher is a browser-driven system designed to provide training tools used primarily during IODE capacity building courses but also for self-training and continuous professional development. The OceanTeacher system comprises two components: the IODE Resource Kit and the Resource Kit Training Manual. The Resource Kit contains a range of marine data-management and information-management material, including software, quality control and analysis strategies, training manuals, and relevant IOC documents. The Resource Kit Manual is a collection of outlines, notes, examples, and miscellaneous class work documents used in conjunction with the Resource Kit to organize a training program in marine data and information management.

This paper details the OceanTeacher system and describes the approach being taken by IODE in providing capacity building to assist developing countries including regional group training courses and workshops to instruct data centre managers in the different aspects of oceanographic data and information management.

Keywords: Capacity building; Distance learning; Oceanographic data management.

Introduction

The International Oceanographic Data and Information Exchange (IODE) program of the IOC has developed the OceanTeacher system, a comprehensive training and self-study tool for ocean data and information management to support its capacity building strategy.

OceanTeacher contains a range of marine data management and information management material, including software, quality control and analysis strategies, training manuals, and relevant IOC documents. It is an extensive self-training and resource tool for newly established Oceanographic Data Centres, designed to assist managers and staff members to acquire the skills to set up and run new IODE centres. It provides a broad spectrum of background information on global data and information archiving activities, specifications for data storage in standard formats, and the software tools needed to perform many quality control, sub-setting, and analyses techniques. In addition, datasets and information relevant to specific geographical regions are provided as a plug-in 'custom pack'. While aimed at developing countries, OceanTeacher is of significant value to developed countries and their marine science agencies.

IODE data and information management capacity building

Capacity building has always been an important component of the IODE program since its inception in 1961. Assisting member states to acquire and build the necessary resources for ocean data and information management is an essential and critical part of all IODE activities. Training courses, mostly hosted by National Oceanographic Data Centres, have been held annually with resource persons from the IODE community providing instruction.

Although these training courses have been successful in training data centre staff, the IODE program had never reached the stage where an agreement was made on a 'standard curriculum'. This meant that courses and their content could vary from course to course and that the course programs were not necessarily in line with the requirements of the participants. As these training courses were mostly one-off activities, the long-term impact was often disappointing, as the participating countries did not have the necessary infrastructure to implement the acquired knowledge. There was little or no follow-up to the courses, so the impact of the course on the participants' day-to-day work could not be monitored.

Regional networks as a structure for IODE capacity building

IOC activities in developing countries are mostly organised using a regional approach. The IOC regions include IOCINCWIO (North and Central Western Indian Ocean - 'Eastern Africa'), IOCEA (Central Eastern Atlantic - 'Western Africa'), IOCARIBE (Caribbean and adjacent region), IOCINDIO (Central Indian Ocean), WESTPAC (Western Pacific region), Black Sea and Southern Oceans regions. The IODE program has recognized the importance of building networks based on this regional structure. This approach has many advantages such as:

- Promotion of communication between members of the network;
- Promotion of south-south cooperation in training and institutional capacity building;
- Facilitating collaboration with other programs and projects;
- Facilitating follow-up for training activities;
- Facilitating exchange of data and information using compatible technologies and formats.

Ocean data and information networks

Since 1998 the IOC has been developing and enhancing the new ODIN (Ocean Data and Information Network) capacity building strategy that links training to infrastructure and operations. These networks promote regional cooperation, data exchange and creation of regional data products. The first project to be implemented under this new strategy was the ODINEA project for seven East African countries in the IOCINCWIO region. In order to support this new capacity building strategy, IODE began the development of a CD-ROM based 'NODC-In-A-Box' product. This product proved to be a useful capacity building tool and provided both the tools and instruction on how to manage and manipulate oceanographic data as well as providing basic regional datasets. In 1998 it was decided to implement a Pilot Project for the development of a computer-based training tool based on the ODINEA CD-ROM, to be called the ‘IODE Resource Kit’.

The success of the ODINEA project led to the development, and approval by a donor, of the ODINAFRICA II project involving 20 African member states. This project has a duration of three years and a budget of US$4 million. The objectives of the ODINAFRICA-II project are:

- Providing assistance in the development and operation of National Oceanographic Data (and Information) Centres and establish their networking in Africa;
- Providing training opportunities in marine data and information management applying standard formats and methodologies as defined by the IODE;
- Assist in the development and maintenance of national, regional and Pan-African marine metadata, information and data holding databases;
- Assist in the development and dissemination of marine and coastal data and information products responding to the needs of a wide variety of user groups using national and regional networks.

The first ODINAFRICA II capacity building workshop on ocean data management was held in April 2001 and the existing IODE Resource Kit was expanded to include a Training Manual and was renamed OceanTeacher.

Ocean teacher

OceanTeacher is a browser-based training and self-study tool for ocean data and information management. It is used in IODE capacity building programs and also for self-training and continuous professional development. OceanTeacher comprises two components: i) the IODE Resource Kit, and ii) the Resource Kit Training Manual. The Resource Kit contains a range of marine data-management and information-management material, including software, quality control and analysis strategies, training manuals, and relevant IOC documents. The Resource Kit Training Manual is a collection of outlines, notes, examples, and miscellaneous class work documents used in conjunction with the Resource Kit to organise a training program in marine data and information management. OceanTeacher can be viewed with a web browser either on-line or off-line. A CD-ROM version can be prepared at any time for use in capacity building workshops or for distribution to interested organisations or individuals. It is not essential that users have internet connection to use the system.

IODE Resource Kit

The IODE Resource Kit is a comprehensive self-training and resource tool for newly established Oceanographic Data Centres, designed to assist managers and staff members to acquire the skills to set up and run IODE data centres. It contains a range of marine data-management and information-management materials, including software, quality control and analysis strategies, training manuals, and relevant IOC documents. The Resource Kit provides a broad spectrum of background information on global data and information archiving activities, specifications for data storage in standard formats, and the software tools to perform many quality-control, subsetting, and analyses techniques.

The Resource Kit is modular in design and contains three basic modules:

- Module 1. IODE Data Centre System
- Module 2. Data Management Systems
- Module 3: Data Analysis and Products

Module 1 discusses the roles and responsibilities of an oceanographic data centre and describes the IODE global network system of data centres. It further describes data and information management within a science program and how the data manager can provide valuable data and information sources to managers and project scientists during a science program. A comprehensive reference library containing relevant IOC manuals and guides, online tutorials and standard reference material is also included.

Module 2 describes some of the skills essential for an ocean data manager including computer systems, database technology, metadata and information management, data observation and collection instructions, data quality control, the use of the internet for data and information exchange, and an introduction to geographical information systems.

Module 3 describes in detail a number of data formats and the source of collateral data. It also includes a data classroom and software toolbox. The data classroom provides a training curriculum in the use of selected software to quality control and analyse ocean station data, using software tools such as the Ocean Data View program, and standard spreadsheet and relational database programs. The data classroom emphasises the connections between available software and global databases, based on the use of common formats. The software toolbox provides a number of useful software tools that can be immediately installed and run. Manuals and test datasets are included. These software packages are freeware and shareware applications.

The modular approach taken with the development of the Resource Kit enables i) selected experts to contribute and regularly update the content; and ii) course programs to be designed based on individual, national or regional priorities using material from each module.

Data management training manual

The Data Management Training Manual is a collection of outlines, notes, examples, and miscellaneous class work documents that can be used in combination with the IODE Resource Kit. The aim of OceanTeacher has been to organise the original source documents and reference materials into the Resource Kit itself, while saving the instructional materials that point to these documents for the Training Manual. The Manual and its inherent course outline have been developed over a number of years, principally during IODE training workshops for the ODINEA project. The long-term exercises, referred to as ‘Intersessional Goals,’ are those pioneered by the ODINEA participants. It was felt that the best way to learn the material in the Resource Kit was to undertake real-world projects to find, quality control, analyse, synthesise, and publish marine data and information.

A typical ocean data management workshop would cover:

- Basic computer skills
- The importance of marine data in general, and within the national and regional environment in particular
- How to set up an oceanographic data centre within the IODE System
- The infrastructure requirements, including hardware and software tools
- How to manipulate and analyse the principal types and formats of marine data
- How to produce ocean data products and to disseminate these products, both over the internet and by traditional methods

The topics covered are based on material from the Resource Kit, although not necessarily all during the same workshop session. Some topics are covered only once, while others are taught more than once, but at increasing levels of difficulty, during the 3-year training cycle.

Information management training manual

The Marine Information Management Training Manual provides a foundation for the professional education essential for the modern information worker. It is currently available in two volumes:

- Course 1. Basic introduction to marine information management, information concepts; information software and technology and the organisation of the collection using a defined integrated library management system (ILMS).
- Course 2. Building on Course 1. Advanced applications of ILMS; creation of research support services and information seeking and retrieval particularly in the electronic environment.

As well as specific information handling techniques, particularly in the electronic environment, students examine the political, economic and social context of the formation of an information centre. Teaching is underpinned by relevant research and utilises web-based learning modules as well as practical exercises. Topics covered include:

- Evaluate and assess the information requirements of the organisation
- Project manage the creation of an information centre in the organisation
- Understand the software and hardware required to underpin information management
- Apply efficient techniques to information seeking exercises
- Build, organise and document a library collection both paper and electronic
- Use the defined integrated library management system to support all management activities
- Set up and maintain research support services
- Identify the major electronic resources in marine science and organise access
- Oversee the production of internal publications and advise on e-publishing
- Introduce the concept of knowledge management to the organisation
- Network within the activities of marine science information associations
- Identify opportunities for continued professional education
- Provide information skills training on marine science resources to Information Centre users

Regional data collections

The Ocean Data Management Training Manual is supplemented by regional data collections containing marine and coastal data for use in environmental analyses. The underlying goal of the collections is to provide the user with data that can be synthesized, particularly in Geographic Information System (GIS).

Regional datasets have been prepared for various training courses, including:

- IOCINCWIO Regional Data Collection (coverage from 15o E to 80o E, 0o to 40o S)
- IOCEA Regional Data Collection (coverage from 50o W to 20o E, 0o to 40o N)
- Red Sea Data Collection (coverage from 10o N to 30o N, 30o E to 50o E)
- ODINCARSA Regional Data Collection (coverage from 33o N to 65o S, 120o W to 50o W)
- Africa Update (WOD01) Data Collection

Typically, these regional data collections contain the following specific data types:

- Base Mapping (Bathymetry, Coastlines)
- Chemistry and Hydrography (Chlorophyll, Nitrate, Nutrients, Phosphate, Ocean Station Data, Oxygen, Salinity, Sea Temperature)
- Geology and Soils (Geophysics, Soils)
- Meteorology (Air Temperature, Precipitation, Winds)
- Physical Oceanography (Currents, Waves)
- Terrestrial Biology (Land Cover, Vegetation)

All regional data collections are available on CD-ROM and on-line at the OceanTeacher site.

Conclusion

The Ocean Data and Information Network capacity building strategy, developed by IODE programme, links training to infrastructure and data centre operations. This model has proved to be successful within the African and South American framework and will be expanded to include other regions. The OceanTeacher system is an important component of this strategy – providing a training and self-study tool for ocean data and information management for use in IODE capacity building programs and also for self-training and continuous professional development. The aim of OceanTeacher is not just to train individual scientists, but also to build new training capacity in the participating institutions. Toward this end, OceanTeacher is designed to let former students become new local trainers.

The IODE program of IOC will continue to develop and maintain the OceanTeacher system as both an on-line and off-line resource to support its capacity building strategy. The on-line version of it is available at OceanTeacher.org. A CD-ROM version, for off-line viewing, can be ordered from IOC.

SESSION: BIODIVERSITY AND BIOGEOGRAPHY

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 47-55
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Twelve years of FishBase: lessons learned

Gert Boden and Guy G. Teugels[†1]

Ichthyology, Africa Museum
Leuvensesteenweg 13, B-3080 Tervuren, Belgium
E-mail: boden@africamuseum.be

Abstract

FishBase is an extensive database with information on fishes that was started in 1990 and that is available on the internet since 1998. The number of hits increased from a few 100.000 in 1999 to more than 4 million in August 2002. Most users are individuals searching for information by common name. Researchers are also intensively using FishBase since they visit the specialized topics. Due to their limited number, it is obvious that their user sessions are relatively less numerous compared to those of the individuals. FishBase is trying not to duplicate information that is already available on the internet. It rather gives species-level links to these particular pages, mainly other databases and distribution maps. FishBase has a 'science first' attitude, *i.e.* information is based on cited scientific publications.

Keywords: FishBase; Biodiversity; Internet.

Introduction

FishBase started in 1990 as an initiative of the European Commission. The database fitted in with the whole program between the European Union and the developing countries. Since most of the available information and knowledge on research on fishes is only available in the developed countries, FishBase was seen as a tool for the transfer of this information and knowledge to the developing countries. The development of the FishBase-concept was given to ICLARM (International Center for Living Aquatic Resources Management, now the World Fish Center), member of the CGIAR-group. In 2000, the FishBase-Consortium was founded in order to maintain the database permanently. The seven members of this Consortium are all more or less complementary in their specializations: FAO (Food and Agriculture Organization, United Nations), ICLARM (World Fish Center), IfM (Institut für Meereskunde an der Universität Kiel), MNHN (Muséum National d'Histoire Naturelle, Paris), MRAC (Musée Royal de l'Afrique Centrale, Tervuren), NRM (Naturhistoriska Riksmuseet, Stockholm) and UBC-FC (University of British Columbia, Fisheries Centre, Vancouver).

What is FishBase?

A CD-ROM version of FishBase was released every year since 1996. Since 1998 FishBase is also available on the internet, the main site being www.fishbase.org. Two mirror-sites are

[1] Since this paper was presented, Guy Teugels passed away. We will deeply miss him.

available since 2001: www.fishbase.de at Kiel, Germany, and ichtyonb1.mnhn.fr at Paris, France. FishBase on the internet opens with the ‘Search FishBase’ page. This page shows some numerical overview data and links to important tools. In October 2002 FishBase included 26.945 valid species, 75.240 synonyms, 135.700 common names, 32.655 pictures and 27.175 references. Eschmeyer (1998) counted the number of valid species at about 23.250, but estimated that the total number of valid species could reach 30.000 or 35.000. FishBase already contains a great part of the valid fish species and newly described fish species are regularly added. The second part of the first page gives the different possibilities to search for information. This can be by common name, scientific name, family, country, reference or other topics. Within a few clicks one is able to see the ‘Species Summary’ page for the species requested, with general data, links to more detailed data and links to other web pages for that particular species when available. These include information on taxonomy, morphology, distribution, ecology, reproduction and many other topics.

FishBase also offers a Fish Forum, which is a platform where people can ask questions and eventually provide answers. Other subjects include a Fish Quiz, a ‘Best Photos’ page and Fish Watcher.

FishBase hits and user sessions

Web statistics for FishBase on the internet are available since 1999. The evolution of the number of hits and user sessions per month is given in Fig. 1. In 1999 it reached only a few 100.000 hits each month, but the number increased to reach a peak of more than 4 million hits per month in August 2002. The number of user sessions per month follows the same growth, ending at about 200.000 in August 2002. Many of these users visit FishBase more than once.

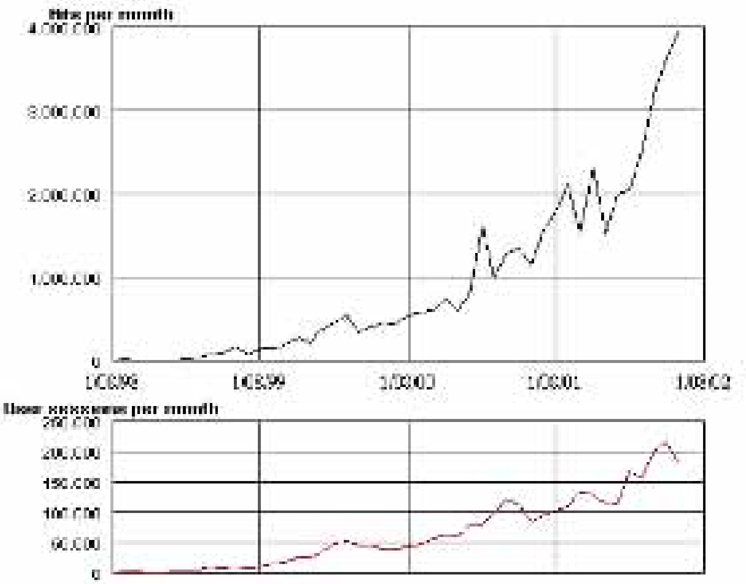


Fig. 1. Number of hits per month and user sessions per month in the period from 1 August 1998 to 1 August 2002.

While the number of unique users on the internet doubled in the period between January 2000 and May 2002, FishBase had more than 5 times more unique users in 2002 compared to 2000 (Table I). As more people are exploring the internet, a lot more are discovering FishBase as a standard database to search for information on fishes. This can be demonstrated by the fact that more unique users are entering the 'Search Page' and keep coming back after having seen the important amount of information and tools available in FishBase.

Table I. The number of internet users worldwide versus the number of FishBase users, during the period January 2000-May 2002 (source: NUA Internet Services)

	Global internet usage	FishBase usage
January 2000	254,29 million	20948
January 2001	455,55 million	43860
January 2002	562,47 million	62481
May 2002	580,78 million	114385

FishBase was running on only one server from the start in 1999. Soon, this server was not powerful enough to handle the rapid growth of FishBase users on the internet. In order to have a higher capacity, two mirror-sites were added at the end of 2001. With these mirror-sites FishBase was able to deal with a lot more visitors. This explains the enormous growth of FishBase users in the period after December 2001 (Fig. 1), while the number of unique users even doubled in the short period between January 2002 and May 2002 (Table I).

Who uses FishBase?

Based on the FishBase guestbook entries in June 2002 (Fig. 2), the main part of FishBase users are individuals. They represent nearly half of the total visitors. Together with visitors from private organizations and universities, they account for approximately 80% of the FishBase users. The international research centers and museums provide in terms of percentage less users. However, due to the limited number of researchers in the world, it is still an impressive percentage.

Since FishBase is available on the internet, it can be used worldwide (Fig. 3). Most of the visitors, about 60%, originate from North America or Europe. Asia is represented by 20% of the FishBase users, leaving another 20% for South America, Africa, Australia and Oceania. We can conclude that FishBase on the internet is less used in the developing countries, although it was primarily meant for them. The main reason is the connection with the internet. At many places in the developing countries there is simply no connection with the internet, and if there is, the connection is costly and not always reliable. If this will improve in the future, the percentage of users in the developing countries will increase. But until then most researchers in the developing countries are using the CD-ROM versions of FishBase rather than the internet version.

If we take a closer look at the FishBase usage by topic (Fig. 4), the most used topics are common names, species summaries and photos. This corresponds with the high number of individuals working with FishBase. Most of them only know the fishes by their common names and are interested in general information, preferably with a picture or photo of it. These three topics are important for individuals, because the fish can be recognized or identified. However, scientific names, used by scientists and specialists, also represent a relatively great part of the topics used. This is clearly an indication that FishBase is used by many researchers, who find it a useful source for information on their investigation. In general, we can see an exponentional decline in usage with increasing specialisation. The more a topic is specialized, the less people will use it. Specialized topics like fish collections, trophic ecology, reproduction and physiology are mainly interesting for people who are familiar with the specific terms used. Individuals not familiar with all these terms will not use these topics. Therefore we can conclude that FishBase is visited mainly by individuals, but also researchers are intensively using FishBase.

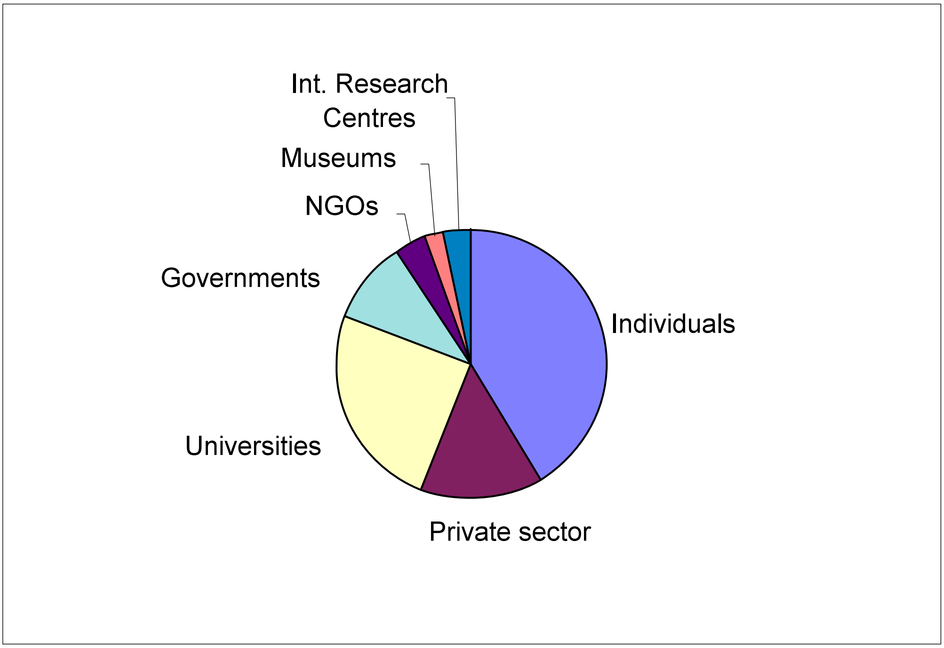


Fig. 2. FishBase users by activity, based on the FishBase guestbook entries in June 2002.

New items in FishBase

In Fig. 5 we compare the percentage of FishBase usage with the percentage of internet usage by country. A great part of the internet and FishBase users live in the USA, and as a result the point lies close to the idealistic line, somewhat to its left. But for FishBase, the interesting points are on the right side of the idealistic line. For these points, the FishBase usage is lower than suggested by general internet usage in these countries. Cases are China, South Korea, Japan, Taiwan, Russia and India. In order to improve the FishBase usage in these countries we have to look for reasons why people in these countries are not using FishBase as expected. A closer look at the countries reveals one thing in common: the scripts they use for their languages are non-Latin scripts. Therefore, FishBase now includes common names of fishes in a variety of non-

Latin scripts. Currently Greek, Chinese, Thai and Arabic are supported and more will follow in the near future.

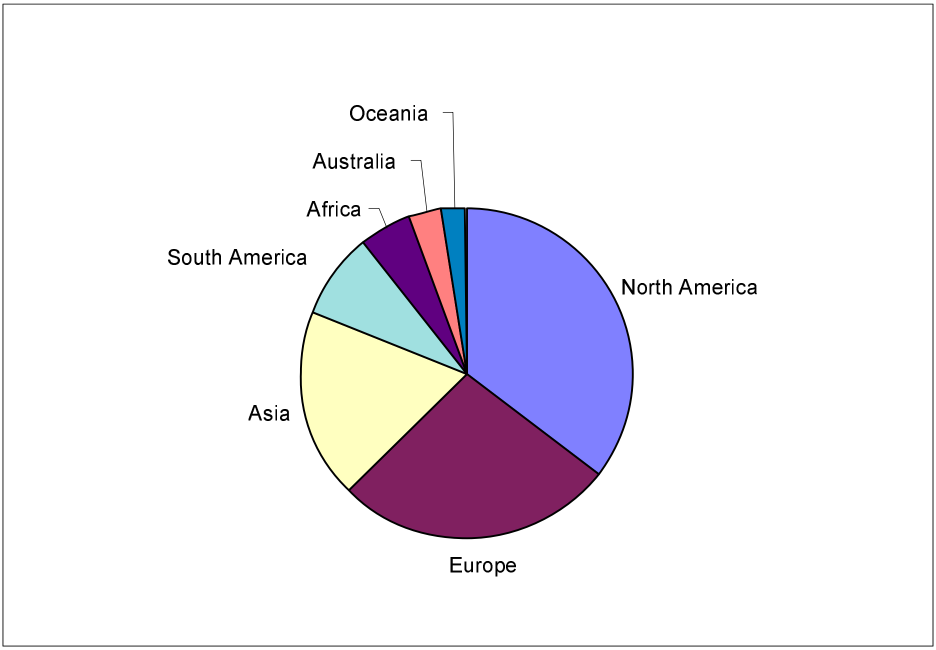


Fig. 3. FishBase users by continent, based on the FishBase guestbook entries in June 2002.

Recently, also other features were added to FishBase. It is now possible to make your own field guide for an area of interest, such as countries, ecosystems, or smaller areas. Technical terms used in FishBase are linked to the respective definitions in the glossary, available in four languages: English, French, Spanish and Portuguese. More pictures are added in FishBase every day and for some species these pictures are divided into different sections: general pictures, pictures uploaded by FishWatchers, stamps, pictures of diseases and pictures of eggs and larvae.

No duplication of effort

The internet is an enormous source and therefore a lot of information on fishes is already available through other databases or pages. Rather than to duplicate it, FishBase makes species-level links. Some of the databases already linked with FishBase are Eschmeyer's Catalog of Fishes, IUCN's Red List data, Museum collection databases, the FAO databases (SIDP, FIGIS, Catch and Aquaculture), ECOTOX and Genbank. For more regional information, FishBase has some links with national databases, *e.g.* the Fish Database of Taiwan and Checklist of Marine Fishes of Turkey. Extra references can be given by some linked bibliographies, like the Zoological Record, Fish and Fisheries and Scirus.

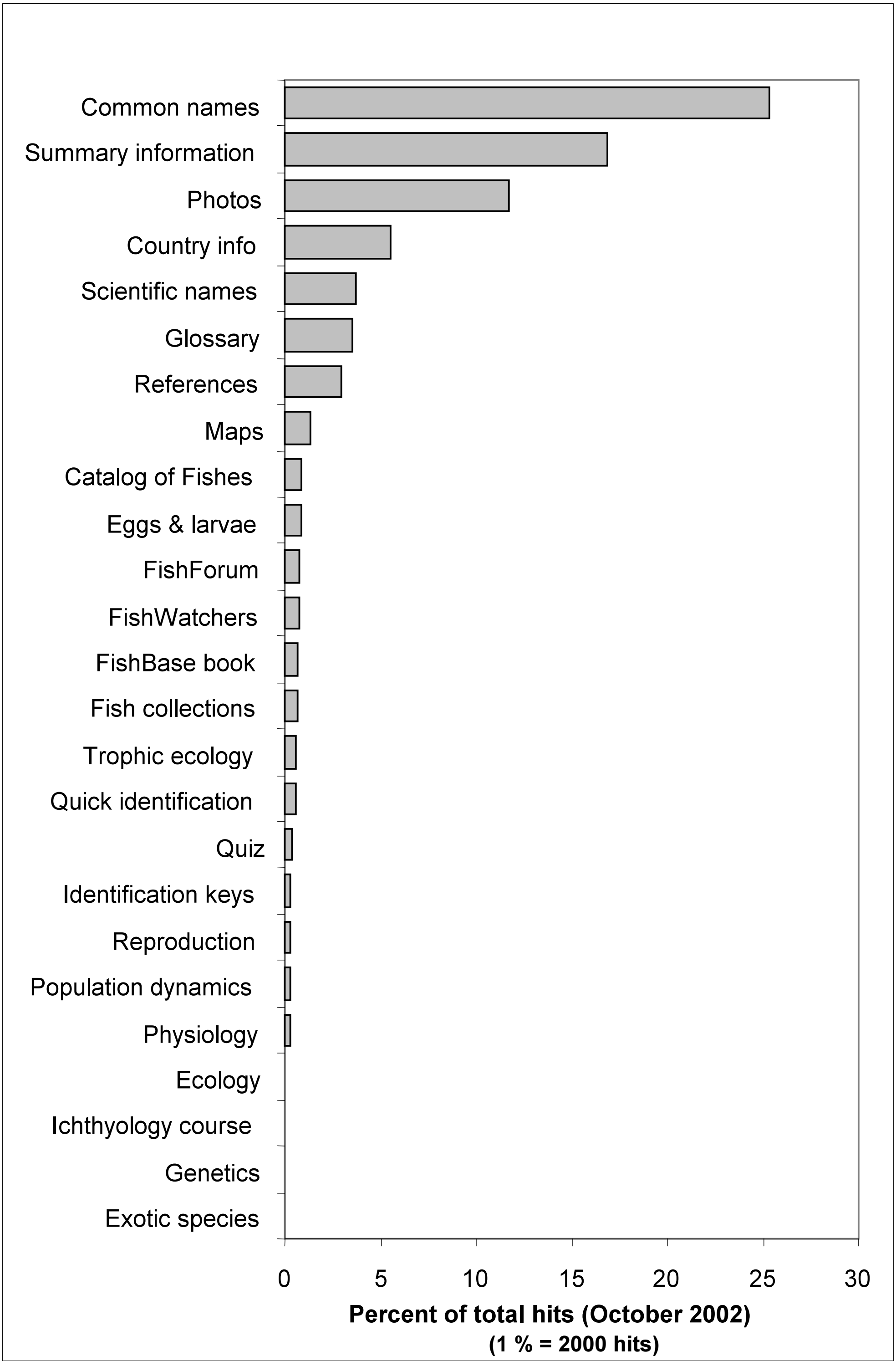


Fig. 4. Frequency of FishBase hits by topic.

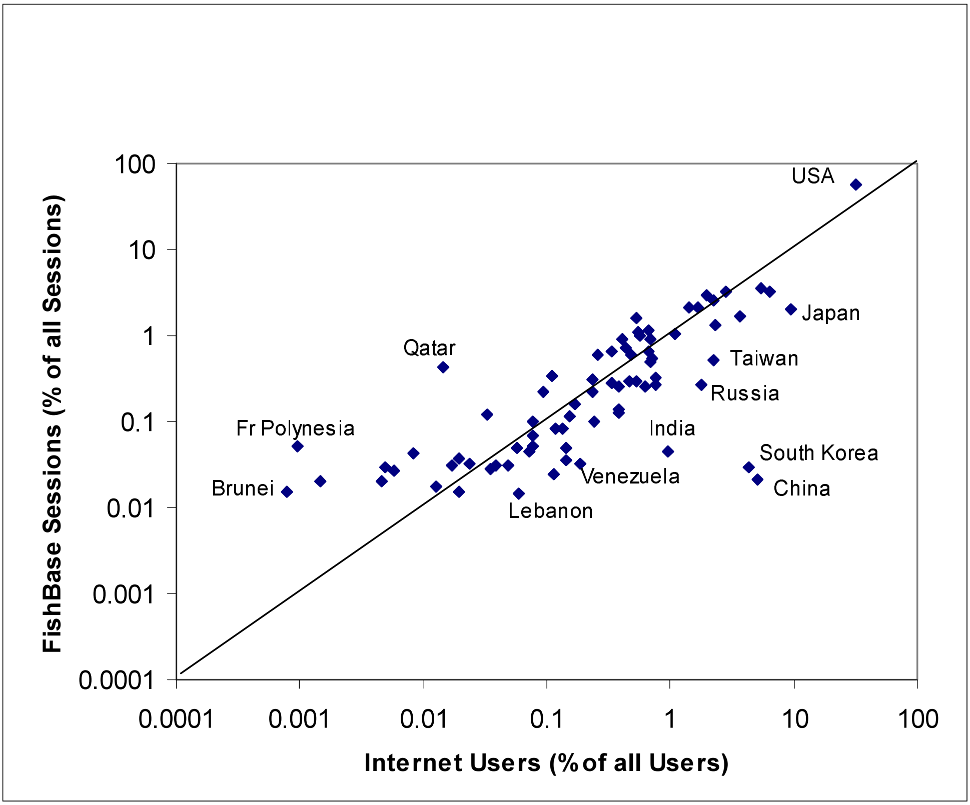


Fig. 5. FishBase usage compared with internet usage by country, in 2001.

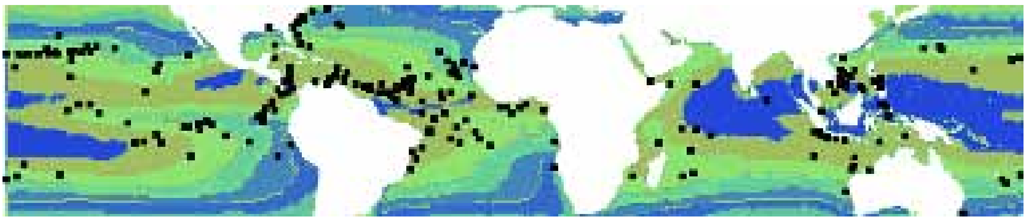


Fig. 6. Example of an 'active' distribution map: the distribution of Exocoetus volitans *Linnaeus, 1758 in relation to the minimum monthly sea surface temperature with the WhyWhere map.*

An important tool for FishBase are 'active' distribution maps for each species, constructed on-demand from occurrence data. Especially for the marine environment one can work with the OBIS map or the WhyWhere map. On these maps it is possible to compare the distribution of a certain species with different marine environmental parameters. For example the distribution of the species *Exocoetus volitans* can be seen in relation to the minimum monthly sea surface temperature (Fig. 6).

Conclusions

The conclusions of Froese (2001) published after ten years of FishBase remain the same today and can be summarized as follows:

- A ‘science first’ attitude. The information in FishBase is based on scientific publications and can be used for scientific research. Therefore FishBase strictly follows the scientific standards for the presentation of it.
- A ‘Yes’ attitude towards people. People who approach and want to contribute with FishBase are not neglected.
- Invite, accept and act on criticism. FishBase welcomes criticism as a way to improve its quality. Even if criticism is unjustified, we try to accommodate it and we interact with the user.
- Data quality and quantity first. FishBase is in the first place a database where people, especially researchers, can find information on fishes. Therefore the quality and the quantity of the information have to be good. This is more important than a good-looking website.
- Keep the design of the database simple. Software is always changing and new programmers have to be able to take over.
- Invest in people. A well-trained staff can do more than student assistants who are working for a short period.
- Give more credit than expected to contributors. It is always a pleasure to know that you are appreciated for the work you have done.
- What is not used is useless. The criterion has to be the actual use of information. FishBase gives priority to the well-used and successful topics.

Some additional conclusions can be given now, mainly about the FishBase usage. FishBase is visited by a lot of individuals, mainly people who are interested in general data about fishes. Mostly, the scientific name of the fish is not known by them, and they rather use the common name as a source to find information about it. With the pictures or photos, they are able to recognize the fish. Therefore, the topics used by them are the common names, the general species summary page and pictures. As a result, FishBase will have to focus on these topics and priority will be given to them. For the common names, some new non-Latin scripts are already added to FishBase and more will follow in the future. In this way, people can search for information on fishes using their language and script, which will increase the number of FishBase visitors, especially visitors from the developing countries. FishBase has already species summary pages for 26.945 valid species. It is already a great part of all known valid species, but not complete. Harmonizing with Eschmeyer’s Catalog of Fishes will be necessary to include all valid fish species. The most important key information like distribution and diagnosis, as well as pictures and photos, has to be entered for every species in the near future. With this information in FishBase, one should be able to identify or recognize every valid fish species.

Other topics are mainly used by scientists and specialists, a relative low number compared with the number of individuals. But it's still an impressive number as scientists are just a little part of the entire population of internet-users. Proved by the fact that scientific names are well used in FishBase, this little group is intensively using the information in FishBase for their research. Therefore, FishBase cannot overlook this part of the users and also has to enter this information. After all, the publications of scientists are used as the basis for the information available in FishBase.

All these lessons are not neglected and are the basis of the success of FishBase. FishBase has now grown to one of the most reliable databases on the internet. As a matter of fact, it is mentioned in different scientific publications as a tool for research. Recent articles are recommending FishBase as an excellent example of an online database (Knapp *et al.*, 2002).

Acknowledgements

We are grateful to Dr Rainer Froese, FishBase coordinator (IfM, Kiel), for his assistance and comments. The contribution of GB forms part of the FishBase project at MRAC, financed by the Belgian Cooperation (DGIC).

References

Eschmeyer W.N. 1998. Catalog of fishes. California Academy of Sciences, San Francisco, USA. 2905p.

Froese R. 2001. Ten years of FishBase. EC Fisheries Cooperation Bulletin 14:13.

Knapp S., R.M. Bateman, N.R. Chalmers, C.J. Humphries, P.S. Rainbow, A.B. Smith, P.D. Taylor, R.I. Vane-Wright and M. Wilkinson. 2002. Taxonomy needs evolution, not revolution. Some changes are clearly necessary, but science cannot be replaced by informatics. Nature 419: 559.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 57-63
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

NeMys: a multidisciplinary biological information system

Tim Deprez[1], Edward Vanden Berghe[2] and Magda Vincx[1]

[1]Marine Biology Section, Ghent University
Krijgslaan 281/S8 ,B-9000 Ghent, Belgium
E-mail: tim.deprez@rug.ac.be

[2]Flanders Marine Data and Information Centre, Flanders Marine Institute
Vismijn Pakhuizen 45-52, B-8400 Ostend, Belgium

Abstract

Several systems to store taxonomic information exist, and more are under development. None of these satisfied the needs identified for a taxonomic information system as was needed by the different research groups within the Marine Biology Section of Ghent University. As a result, a taxonomic database system was developed in house, based on experiences with previous taxonomic databases, and implementing desirable features of existing systems. The result is NeMys, a system capable of capturing the basic taxonomic information (name, synonymy, classification) linked to literature (including full text of the publications), distribution records, morphological description, molecular genetics, and pictures. The whole structure is generic, and can easily accommodate widely differing taxonomic groups, as is demonstrated by the nematode and mysid data that now form the content of the database.

Keywords: Taxonomy; Database; Mysida; Nematoda; Generic; Web application.

Introduction

During the last decade the concept of taxonomic databases has been established within the world of classical taxonomy (ITIS, Species2000, Catalogue of Life, Tree of Life). Relatively more effort was paid to well-known or commercially important groups like fishes (*e.g.* FishBase), algae (*e.g.* Algaebase), birds (*e.g.* Birds of Europe - ETI), Cephalopoda (Cephbase)… These efforts can be seen as very valuable in terms of raising appreciation for this new approach within the field of taxonomy.

The decreasing number of taxonomists working on certain taxa urges the need of storing biological data on these taxa in a fully digital way.

NeMys was developed after a series of experiences with other digital information systems mainly on free-living marine Nematoda (NemasLan, TaxonLan) and Mysida (MysidLan, TaxonLan) (Deprez *et al.*, 2001(a); Deprez *et al.*, 2001(b); Deprez, 2001). These database applications – the starting point in entering taxonomic, biogeographic and morphologic information – had a suit of disadvantages also visible in many other biological database systems:

1. lack of standards do not allow exchange of data between taxa; 2. there is in many cases an unbreakable link between the program code of a user interface and the datasets; 3. datasets are not available online; 4. there is a limitation in data types that can be entered.

The main philosophy behind the new NeMys project is that a biological information system must be able to handle all kinds of information and that this information must be accessible by anyone, anywhere. At any time a link between the data and the data source must remain (*e.g.* a literature source will be linked in a digital format to a data record).

This paper presents the current state of the NeMys data structure, the datasets available and some future possibilities.

Overview – data structure

In NeMys all data are stored in a relational database hosted on an SQL server: all information is stored only once and all sets of basic information (*e.g.* literature, taxonomy, geography) can easily be linked to each other. The user interface making use of recent web technologies (ASP and SVG) combines all the separated datasets in a user-friendly format. An overview of the basic database structure described below is displayed in Fig. 1.

The 'taxonomic structure' visualized by a taxonomical tree can be seen as the backbone of the system. For the structure of the table in the database holding this taxonomic information ('tu' – taxonomic units-, see Fig. 1), a variation on the structure as used by ITIS, and identical to the taxonomic datasets hosted at VLIZ (*e.g.* APHIA; http://www.vliz.be) is chosen. This structure makes use of a parent-children relation to store the hierarchical structure. Of all taxa the following basic information is stored: taxon name, taxon authority (who described the taxon and when), taxon rank (*e.g.* species, genus, family, …), taxon parent (the taxon under which the current taxon is ordered in the hierarchy), accepted taxon (if the taxon synonymies with another taxon, then a link to the accepted taxon is created) and taxon data source (in which literature source was the existence of the taxon found).

To this taxonomic tree al kinds of information can be linked directly or indirectly.

Directly linked information

'Literature information' linked to the taxonomic tree is of basic importance within the NeMys concept. Therefore a distinct dataset holding the literature reference ('tbldocuments' in Fig. 1) and a link to the possibly scanned document in PDF format (http://www.acrobat.com) is used. This literature dataset is reused in the system many times to act as data source for data entered in the system (*e.g.* taxon data source in the taxonomic structure, see above). The link between taxon and literature ('doc_link' in Fig. 1) is enriched with additional information like an indication of the type of document (*e.g.* original description, ecological review, …) and a remark (*e.g.* the page in the literature source where the taxon occurs, comments on pictures in the article, …)

A second type of information immediately linked to a taxon is 'pictorial information' ('tblpicture' in Fig. 1). Any type of pictorial information can be entered accompanied with basic

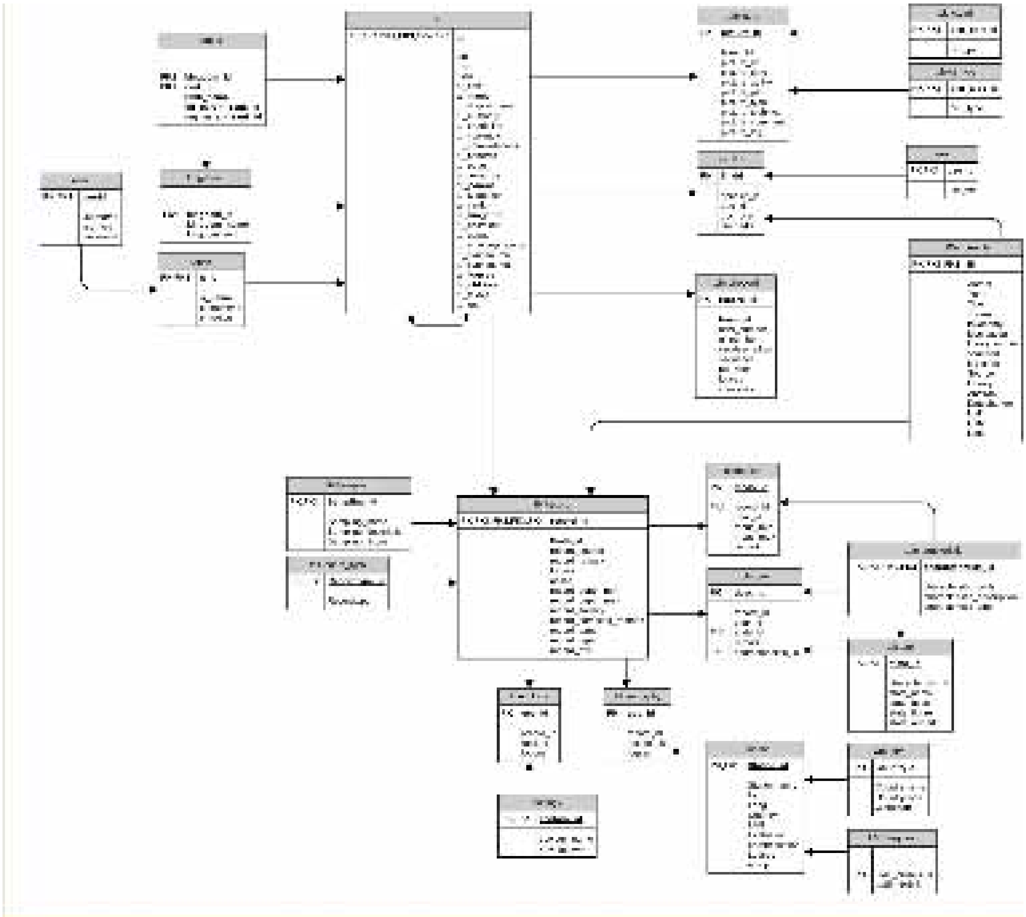


Fig. 1. Relational database diagram of the NeMys database system.

additional information like picture type (*e.g.* scanned drawing, scanning electron microscope picture, …), body part displayed on the picture, picture author, date at which the picture was taken, technical comments on the picture, observations about the picture and the picture file name.

A third type of direct linked information, are ‘notes’ which can hold any possible information not immediately fitting in the structured information (‘notes’ in Fig. 1). As datasource for a note a literature source is always needed. An example of a note can be the misspelling of a taxon name in a certain literature source.

The fourth type of directly linked information is molecular records (‘molrecord’ in Fig. 1). For these records the standard data fields are used as used by the NCBI-GenBank, EMBL-EBI and DDBJ. A link to these online datasets is automatically created.

Indirectly linked information

Morphological, ecological and geographical information can be seen as indirectly linked information to the dataset. These data are not directly linked to a taxon but to a record that is linked to a taxon; this is considered as indirectly linked information. Each of these records are always holding information derived from one literature data source.

'Morphological information' is split up in two types: morphological information ('tblmorfo' in Fig. 1) defined through characteristics ('tblcharacteristic' in Fig. 1) and states ('tblstate' in Fig. 1), and measurement information ('tblmeasur' in Fig. 1).
Morphological characteristics and their states are in most cases defined by a taxonomist when starting up and specifying a certain dataset. These states can be illustrated more in detail by a picture (which is useful for identification purposes) and by a textual description. Adding or changing these characteristics and states can be done at any time but includes risks on data loss.
Measurement information is also divided in different types of measurements also defined separately for the taxon of interest. Measurement data (*e.g.* length of male, number of spines,...) are always stored in two values: one minimum value and one maximum value.

A second important set of indirectly linked information is 'geographic information' ('tblgeography' in Fig. 1). Geographic information is stored by linking an exact defined location from a geographical gazetteer ('location' in Fig. 1) to a record. This geographic gazetteer is a list of locations with exact coordinates and additional geographical information. Biogeographic information can be used in creating maps, regional checklists and as part of an identification procedure.

Next to geographic and morphologic information there is the possibility to add an unstructured remark to a record, or ecologic information.

Overview – user interface

A user interface to consult the data structure was created making use of the ASP technology and can be fully consulted through an internet browser. This technology is widely known to be fast and relatively easy in use, and one of the main advantages is that all database activity is performed server-side. For displaying scanned literature, Acrobat Reader freely available for any platform (Windows, Linux, MacOs) or another package which is able to open files in pdf-format, is used. Maps are displayed after a free available small plugin, capable of reading the xml based svg-format, is installed.

The user interface for all datasets is exactly the same which enlarges possibilities of easy and quick creation of new datasets and interlinking the different datasets.

In general the user interface is split up in two main sections: one to consult data (available for everybody) and one to enter data (available for users having a login). The data consultation section allows users consulting all data available through searches starting from taxonomy, geography, literature or morphology. Maps, pictures and scanned literature can be consulted online.

Datasets

At this moment two datasets (both consultable through http://intramar.rug.ac.be/NeMys) are available within the system: one on Mysida and one on free-living marine Nematoda. The dataset on Mysida already uses all features available while for the Nematoda some types of data are now started to be entered.

Within the Mysida database about 1150 species names are entered of which about 180 are synonyms. Morphologic, biogeographic, bibliographic and pictorial data is available for all 129 European species available in the datasets. Beside the European data, full biogeographical and bibliographical data is also available for the genera Anchialina and Gastrosaccus (both globally occurring). In total more than 3000 references about Mysida are available of which 800 are already scanned (being more than 9000 pages). In the current state of the dataset, 930 pictures are linked to taxa (all European taxa have at least one illustration), about 950 exact locations (mainly from European marine waters) are available in the geographic gazetteer and about 600 records holding morphological and/or geographical data extracted from literature are available. In total there are now 2850 links between literature and (mainly European) taxa. This high number of links reflects the bibliographic value of this dataset for the European region.

The Nematoda dataset has in its current state mainly data available for Antarctic and North Atlantic Marine Nematoda. In total 3055 species, 3843 literature references (of which 1600 are scanned), 170 locations in the geographical gazetteer and 2861 morphological and/or geographical records are available for this dataset. This dataset has the potential to grow into a much larger dataset than the Mysida as within this group a much higher number of species are known (about 4000 cf. Warwick *et al.* 1998).

Discussion: current and future possibilities

As described, NeMys already has a large amount of biologically important data types that can be linked to the database. The generic structure enables adding possible future new data types.

One of the key principles also used in the ancestors of NeMys (NemasLan, MysidLan) is the storage of the data source next to the data. This enables future data control and correction of possible misinterpreted data in the literature source. In most other similar database systems only a reference is stored while in NeMys the reference and a digital copy of the data source are stored.

Data available in a database only exposes its value once this data can be accessed in a user friendly environment. Originally the datasets were consulted through an MSAccess user interface. The main disadvantage of these types of user interfaces is that they are platform-related, they are only available in a local network, they need to be installed and they ask relatively more activity from the server and thus are relatively slower.

One solution which solves most of these problems is a user interface making use of the user's web browser and where most activity is performed server-side. The user interface created for consultation of the datasets does is relatively easy in use, fast and does not expect the user to install software except for displaying maps.

The generic structure used for storage of morphological and geographical information makes it also possible to use this data for preliminary identification purposes. A cyclic procedure based on delimiting choices made by the user (morphological, geographical or taxonomical) offers the possibility of identifying specimens much quicker than making use of classic dichotomous keys (if available). This whole process is completely dependant on the data entered in the system. Therefore it is of main importance to keep in mind what data is available and what data entry strategy is used (taxonomically oriented, geographically oriented). The data structure offers possibilities to create dichotomous identification keys which can also be used in a paper version or when no internet connection is available.

The use of a well-documented geographical gazetteer linked to taxa combined with an automatic mapping facility offers large biogeographical data analysis possibilities. The svg-technology used for biogeographical mapping has the advantage that it is a very quick way for displaying geographically referenced data and that this within purely biogeographical data can be easily combined with maps on biological and non-biological data (*e.g.* salinity, temperature, coral reef distribution, …). The requirement to install a plugin at the user side is still a disadvantage.
The use of the xml-based svg-technology for mapping purposes already opens possibilities in mapping data available on distributed databases (not yet available). Once distributed datasets are set up, it will thus be possible to generate biogeographical maps displaying data on different taxa originating from different databases hosted on different systems (*e.g.* Fishbase, Algaebase, NeMys, …). Before all this will be possible a limited set of common standards (through XML??) will have to be formulated.

In the near future most attention will be paid to the datasets: completing datasets and starting up new datasets on new taxa (*e.g.* Marine Copepoda). The visibility of the datasets will be enlarged by publishing the datasets through a portal accessible through the website of VLIZ. Through this portal it will be possible to perform simple taxonomical and geographical search actions. More detailed analyses of the dataset will be performed on the web interface hosted at the Marine Biology Section (Ghent University).

In conclusion it can be stated that NeMys can be seen as a very broad biological information system. Next to the pure taxonomical aspect of the system it offers possibilities in data storage for a much broader range of end users.

Acknowledgements

Many people helped in getting the data structure and datasets at the current level. Data input in the Mysida dataset was done by Tim Deprez and Bjorn Possé. Data input in the Nematoda dataset was done by Guy De Smet, Sandra Vanhove, Ann Vanreusel, Magda Vincx, Jan Vanaverbeke, Maaike Steyaert and Jan Wittoeck.

Discussions on the data structure were mainly held by Tim Deprez, Magda Vincx, Guy De Smet and Edward Vanden Berghe.

List of acronyms with correlated URLs

Algaebase	-	http://www.algaebase.org/
APHIA	-	Marine species register for the North Sea: http://www.vliz.be/vmdcdata/aphia/index.htm
ASP	-	Active Server Pages: http://www.asp.net/
Catalogue of Life	-	http://www.biodiversity.reading.ac.uk/catalogue-of-life/
Cephbase	-	http://www.cephbase.utmb.edu/
DDBJ	-	DNA Data Bank of Japan: http://www.ddbj.nig.ac.jp/
EBI	-	European Bioinformatics Institute: http://www.ebi.ac.uk/embl/index.html
ETI	-	Expert Center for Taxonomic Identification: http:// http://www.eti.uva.nl/
FishBase	-	http://www.fishbase.org
ITIS	-	Integrated Taxonomic Information System: http://www.itis.usda.gov
NCBI	-	National Center For Biotechnology Information: http://www.ncbi.nlm.nih.gov/
PDF	-	Portable Document Format: http://www.acrobat.com/
Species2000	-	http://www.sp2000.org
SVG	-	Scalable Vector Graphics : http://www.w3.org/TR/SVG/
Tree of Life	-	http://tolweb.org/tree/phylogeny.html
VLIZ	-	Vlaams Instituut voor de Zee, Flanders Marine Institute: http://www.vliz.be
XML	-	Extensible Markup Language: http://www.w3.org/XML/

References

Deprez T. et al. 2001a. MYSIDLAN 4.0 Taxonomy and biodiversity of shallow coastal Mysidacea of the Western Indian Ocean [CD-ROM]. Magda Vincx, Marine Biology Section: Gent, Belgium. 1 CD-ROM.

Deprez T. et al. 2001b. Taxonlan, a fully generic digital archive of species for biodiversity, morphology and taxonomy research. In: Proceedings of the Eleventh international Meiofauna conference, July 15-20, 2001. Boston USA, p.70.

Deprez T. 2001. Taxonlan, a generic digital archive of species in terms of biodiversity, morphology and taxonomy. Bionet-International NEWS 7:5-6.

Warwick R.M.; H.M. Platt and P.J. Somerfield. 1998. Free-living Marine Nematodes. Part III. Monohysterids. Synopsis of the British Fauna. The Linnean Society of London and The Estuarine and Coastal Sciences Association.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 65-70
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

The Marine Species Database for Eastern Africa (MASDEA)

Esther Nazi Fondo[1], Melckzedeck Knight Osore[1] and Edward Vanden Berghe[2]

[1]Kenya Marine and Fisheries Research Institute
PO Box 81651, Mombasa, Kenya
E-mail: efondo@kmfri.co.ke

[2]Flanders Marine Data and Information Centre, Flanders Marine Institute
Vismijn Pakhuizen 45-52, B-8400 Ostend, Belgium

Abstract

The Marine Species Database for Eastern Africa (MASDEA) is a taxonomic/biogeographic database on the species of marine plants and animals found in the Eastern Africa region. After the Rio Conference, and increased awareness on biodiversity and need for integrated coastal zones management and planning, the gap in knowledge of the marine species of the region became very obvious. In spite of the existence of a number of taxonomic databases, none of these covered the needs for our region. Some global databases with adequate cover for Eastern Africa exist, but only incorporate information on a limited number of taxonomic groups.

The goal of MASDEA is to fill in the gap of knowledge on the taxonomy and biogeography of marine species in the region. Its objectives are as follows:
1. collect all available literature on the marine species of the region;
2. enter taxonomic information and distribution data on the species into the database;
3. seek the support of taxonomic experts of different groups of the region, to assist in quality control;
4. identify gaps, and locate the relevant literature to be entered in the database;
5. make the database available to scientists in the region and beyond.

Keywords: Biogeography; Database; Eastern Africa.

Introduction

The earth has a vast array of living organisms that possess certain consistent features by which they can be identified and sorted into distinct groups (taxa). There is currently no consensus on the total number of species: estimates range from 4 million to 100 million. In the last 250 years roughly 1.7 million organisms have been named and formally described (Gewin, 2002). Taxonomy is important in that it gives unique names to organisms and these names can be used all over the world despite different languages (Jefferey, 1989). Taxonomy provides the reference system for all biology, so organizing, updating and streamlining the process of providing taxonomic information in the most easily accessible way is a priority.

At least three kinds of taxonomic databases are recognized:

1. Taxon databases, where the information contained relates directly to species (or other taxa);

2. Accession, specimen or collection databases, which contain information on: individual specimens (in a museum collection), accessions (cultivated in a botanical garden), cultures (maintained in a microbial culture collecting), or observations (recorded in a mapping project);

3. Name databases or 'nomeclators'
   Several global initiatives exist to inventory all published taxonomic names (e.g. Species 2000, All species etc.). Some databasing efforts are directed specifically at gathering information on marine organisms [e.g. Veron's global list of Corals (Veron and Stafford-Smith, 2000), Sheppard's Corals of the Indian Ocean (Sheppard, 1998), FishBase (Froese and Pauly, 1997), WDC-A's NODC Taxonomic Codes, later succeeded by the ITIS database).

These databases however are limited on a number of taxonomic groups, or are specific to certain regions. In the Western Indian Ocean region a lot of marine species have been described by visiting taxonomists and a lot of literature has been published in different journals and monographs. The MASDEA database is aimed at collecting all this information from the region, compiling it and making it accessible to the scientists of the region. Marine organisms are not sensitive to national boundaries and the database includes information from 21 countries or regions from the Red Sea to South Africa, and eastwards to the Western Islands.

Objectives

After the Rio Conference on sustainable development, the gap in information regarding marine species became obvious. At the RECOSCIX-WIO (Regional Cooperation of Information exchange of the Western Indian Ocean) which is located at Kenya Marine and Fisheries Research Institute, Mombasa, Kenya, a lot of literature was available. In 1996, Dr Edward Vanden Berghe who was then the coordinator of the RECOSCIX-WIO Project designed and developed the MASDEA database.

The objectives of the database were to:
- collect all available literature on the marine species of the region;
- enter taxonomic information and distribution data on the species into the database;
- seek the support of taxonomic experts of different groups of the region, to assist in quality control;
- identify gaps, and locate the relevant literature to be entered in the database;
- make the database available to scientists in the region and beyond.

The database has several potential uses:

The data are useful for conservation, as a source of knowledge on the diversity of species in the region, and assists in making decisions on planning of marine protected areas. It provides historical information: one can keep track of extinct species and also keep track of old information and literature. Current and valid names as well as synonyms of species are clearly indicated as well as the authority (author who described the species) therefore avoiding confusion that might result from consulting the literature without knowledge of how to interpret it. Accounts for newly discovered species are entered into the database as information becomes available, thus avoiding losing information in future.

Guiding principles for development

The software for the database is developed in Microsoft Access and a number of guiding principles were observed while developing the structure and the user interface (Vanden Berghe, 1997):

- Importance of keeping wrong and outdated information, this includes name changes, misidentification and information on synonymy and current valid names. Biogeographic records and scientific findings in general tend to live their own life once they are published. Corrections to outdated or wrong information does not always reach all those who read the original publication. Because of this, the database might look 'incomplete' to some, and supplemented with information that is already known (but not by all) to be wrong.
- Importance of leaving an 'audit trail', where all the sources of information is documented in the database. Thus the database is a reference tool that serves as a roadmap to the literature available on biogeography of the Western Indian Ocean.
- Not forcing an extra layer of codes on the users so as to avoid confusion. Taxonomic nomenclature is a coding system in its own right codes, necessary for databases technical reasons, linking records in the relational structure of the database are completely hidden from the user.
- Keeping it simple by having only those fields that are essential to capture the biogeographical information.

Structure of the database

The database consists of a synonymized list, with distribution records referring to the taxon name with which the distribution information was originally published. Each distribution record (in principle present in one of the countries of the region) is referenced to the literature. Several flags makei t possible to keep track of records that are known to be false or doubtful. This feature, together with the way synonyms are treated, ensures that false records with invalid names are only entered once, and appropriately flagged.

In the database, four tables form the core of the structure, Countries, Records, Literature and Taxonomy tables. The records table links country and taxonomy, and a reference to the literature, making it possible to keep track of sources of information on the level of occurrence of taxa in a country. A separate table gives information on the higher taxonomy; book-keeping of the flow of information into the system is possible through a table with session information.

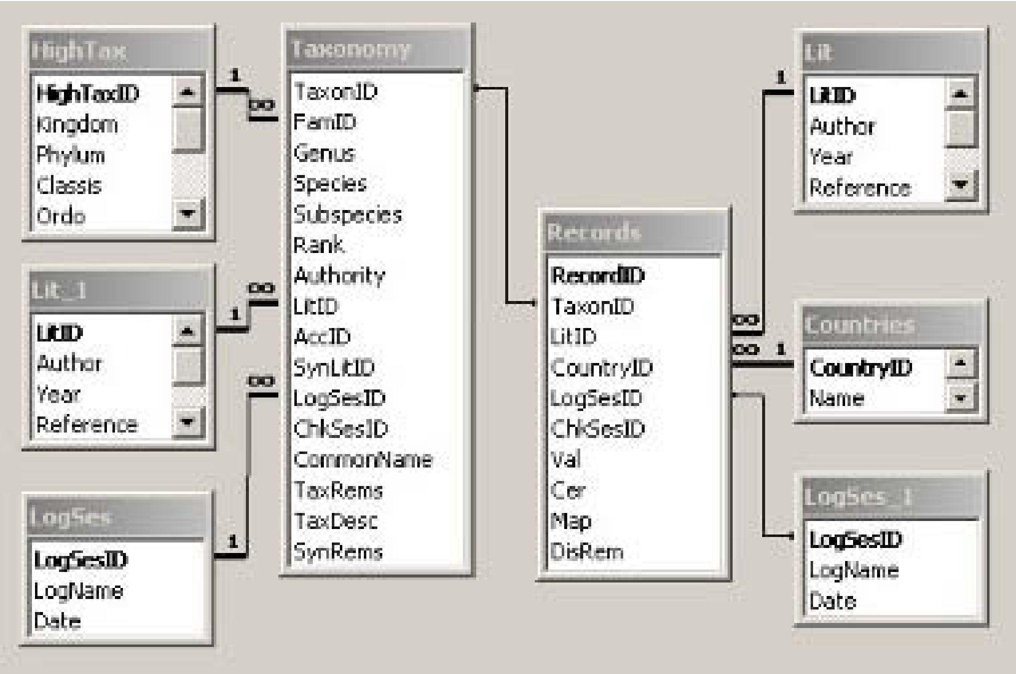


Fig. 1. Relational structure of the MASDEA database, implemented in Access.

Current status

Data entry began in December 1996 and approximately 30,000 species and 15,000 distribution records had been entered at the time of writing (August 2002). The database is not completed and the status of some of the groups from different sources is as follows:

- Sea turtles from the Food and Agriculture Organization of the United Nations (FAO) Species Catalogue and The World Conservation Union (IUCN) Manual for Marine Mammals
- Fish information from Fish Base, and FAO species catalogues
- Sheppard's work on reef-building corals (Sheppard, 1998)
- Decapods (from different sources)
- Echinoderms from Clark and Rowe (1971)
- Molluscs from several sources
- Algae (from Silva *et al*., 1996)
- Several works on general natural history e.g. Vine (1988), Richmond (1997)

The database is now accessible online at: http//www.vliz.be/vmdcdata/Masdea/about.htm
Online, the database can be consulted in three ways:

- Browse: to navigate through the taxonomic tree by clicking on one of the list of taxon names;
- Search: to go directly to the taxon of interest by entering the (partial) name of the taxon in a search form;
- List: to create checklist for a geographical area.

List of countries/regions included in the database

Aldabra
Chagos
Comores
East Africa
East Africa and Madagascar
Eritrea
Kenya
Madagascar
Mascarene Islands
Mauritius
Mozambique
Red Sea
Reunion
Rodriguez
Seychelles
Somalia
Tanzania
Western Islands

Partners

The host of the RECOSCIX-WIO Project, responsible for the data input is the Kenya Marine and Fisheries Research Institute (KMFRI). The Flanders Marine Institute (VLIZ) has been responsible for technical developments and maintenance of the database.

Acknowledgements

We wish to acknowledge the following people who deserve our thanks: Ms. Eunice Onyango of KMFRI who assisted in data entry; Dr David Obura of CORDIO, for validation of the corals; Professor Marco Vannini of the University of Florence and Dr Renison Ruwa of KMFRI for information on Decapods; Yves Samyn and Maria Garcia of VUB and Dr Claude Massin of the Royal Belgian Institute for Natural History for providing information on Echinoderms and also Ward Appeltans for the logging of records. We are grateful to the Kenya Marine and Fisheries Institute and the Flanders Marine Institute (VLIZ).

References

Clark A.M. and F.W.E. Rowe. 1971. Monograph of the shallow water Indo-west Pacific Echinoderms. Trustees of the British Museum (Natural History) London. Publication number 690. 238p., 31 plates, 100 text figures.

Froese R. and D. Pauly. 1997. FishBase 97: Concepts design and data sources. ICLARM: Manila, Philippines. 256p.

Gewin V. 2002. All living things, online. Nature 418: 362-363.

Jeffrey C. 1986. Biological Nomenclature. 3rd Edition. E. Arnold, London, UK. 86p.

Richmond M. 1997. A guide to the sea shores of Eastern Africa and the Western Indian Ocean. SIDA. Sweden.

Sheppard C. 1998. Corals of the Indian Ocean. Western Indian Ocean Marine Science Association. Zanzibar, Tanzania. CD.

Silva C., P. Basson and R. Moe. 1996. Catalogue of the benthic marine algae of the Indian Ocean 79. University of California Publications in Botany.

Vanden Berghe E. 1997. MASDEA- Marine Species database for Eastern Africa. WINDOW Newsletter 8(2): 5-8.

Veron J.E.N. and M. Stafford-Smith. 2000. Corals of the World. Australian Institute of Marine Science. Townsville, Australia. 3 volumes.

Vine P. 1988. Red Sea invertebrates. Immel Publishing London. 224p.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 71-76
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Development of an Atlantic Canada marine species information system based on a museum collection: a case study

Angela Martin, Lou Van Guelpen, Gerhard Pohle and Mark J. Costello

Atlantic Reference Centre, The Huntsman Marine Science Centre
1 Lower Campus Road, St. Andrews, N.B., Canada, E5B 2L7
E-mail: arc@dfo-mpo.gc.ca

Abstract

The Atlantic Reference Centre (ARC), established in 1984 as an amalgamation of regional collections, is a research museum of Canadian Atlantic marine biota and a provider of regional marine biodiversity information. Computerization of all catalogued lots has been accomplished through museum catalogue database restructuring and various funding programs. Because of its value as an archive of marine biodiversity data, concurrent ARC goals have been to place the database on-line and to use it for developing a species information system (SIS) for Canadian Atlantic marine biota. The SIS is an evolving entity of products, often collaborative, and funded by various biodiversity information initiatives. The ARC SIS is currently comprised of specimens and catalogues; on-line sites with querying, mapping and analysis capability for each museum taxon or displaying images, information, and a distribution map for each Canadian Atlantic fish species; a comprehensive list of all species inhabiting the Bay of Fundy, supplemented with substantial species information; a list and classification of all fish species in Canadian Atlantic waters; an on-line site integrating the biodiversity databases of the ARC and other providers with DFO physical and biological databases, into a biodiversity atlas service; and funding to 1) place ARC fish records on Species Analyst Canada, 2) develop a comprehensive Canadian Atlantic marine biodiversity register, 3) develop a web site for the Bay of Fundy species information system, and 4) perform quality assurance and quality control of the ARC museum database, thus improving data quality in the SIS. Future plans include collaboration with LarvalBase and FishBase in 2003-04, and expansion of the Canadian Atlantic marine biodiversity register. The ARC SIS will aid students, educators, researchers and managers in studying, protecting, and promoting the sustainable use of Atlantic Canada's natural resources.

Keywords: Biodiversity; Information; Marine; Atlantic Canada.

Introduction

The Atlantic Reference Centre (ARC) is a research museum of Canadian Atlantic marine biota containing an archive of biodiversity information in the form of preserved specimens, and paper and computerized museum catalogues. This paper describes how, through a wide variety of short-term projects, the ARC is developing these resources into a Species Information System (SIS) for waters of Atlantic Canada, comprised chronologically of:

- ARC museum specimens, and paper and electronic catalogues,
- on-line sites with querying, mapping and analysis capability for each museum taxon,

- an on-line site displaying images and information for each Canadian Atlantic fish species,
- a comprehensive list of plant, invertebrate, and vertebrate species inhabiting the Bay of Fundy, supplemented with substantial species information,
- a list and classification of all fish species in Canadian Atlantic waters,
- an on-line site displaying a distribution map for each Canadian Atlantic fish species,
- an on-line site integrating the biodiversity databases of the ARC and other providers with DFO (Department of Fisheries and Oceans) physical and biological databases, into a biodiversity atlas service,
- 2003 funding to 1) place ARC fish records on Species Analyst Canada, 2) develop a comprehensive Canadian Atlantic marine biodiversity register, 3) develop a web site for the Bay of Fundy species information system, and 4) perform quality assurance and quality control of the ARC museum database, thus improving data quality in the SIS.

The museum collections

The ARC, a joint operation of the Huntsman Marine Science Centre (HMSC) and the DFO, is located in St. Andrews, New Brunswick, Canada on the Bay of Fundy shore. Established in 1984, the ARC united two organizations: the HMSC Ichthyoplankton Laboratory, established in 1977 and archiving a valuable collection of Canadian Atlantic ichthyoplankton and adult fishes, and the DFO Identification Centre, initiated in the 1960s and with excellent holdings of local invertebrates and fishes, some dating back to the early 1900s. Collection growth has continued to the present such that holdings now represent all Canadian Atlantic waters.

A major benefit of the collecting history is that, for selected DFO research surveys, especially ichthyoplankton surveys, all material collected was archived at the ARC and is of great value for ecological, population, and biogeographic, as well as systematic studies. ARC holdings by December 2002 comprised 126,403 catalogued lots (a lot contains all specimens of a species from one collecting event, such as a bottom trawl) (Table I). The ichthyoplankton collection is the largest in the United States and Canada (Poss and Collette 1995). The ARC also houses type specimens and osteological preparations.

Table I. ARC collection holdings, December 2002

Collection	Catalogued lots	Representing
Invertebrates	13,475	24 phyla
Fishes	24,318	204 families
Larval fishes	88,610	105 cruises

In-house taxonomic expertise has contributed to major collaborative research undertakings, such as the US GLOBEC (Global Ocean Ecosystems Dynamics) program on Georges Bank, and DFO's Scotian Shelf Ichthyoplankton Program on Nova Scotia's offshore banks. This expertise also includes marine zooplankton, nektonic and benthic invertebrates, and freshwater insects.

The computerized catalogue

Since its foundation the ARC has had a policy of cataloging its holdings in computerized databases. However, changing technological developments resulted in a long and complex process of catalogue database restructuring over time involving dBase, MINISIS, Revelation, Advanced Revelation, MUSE, and finally Specify. This database restructuring was a learning process that wasted resources. From this experience we learned that when in-house database expertise is lacking, selection of a dedicated museum database management system is advisable and should be undertaken with care.

In 2002 the ARC adopted the museum database Specify (http://usobi.org/specify), for which technical support is freely available. Specify is the successor to MUSE (see MUSE at http://www.usobi.org/specify/DevaSearch/search.html). In recent years the manual addition of records into MUSE and Specify has enabled computerization of all ARC catalogued lots, completed in 2002. Also completed in 2002 was a long-term project to electronically catalogue the ARC's extensive ichthyoplankton holdings, utilizing DFO databases. These tasks were accomplished through data management and biodiversity funding programs in Canada and the USA, together with various government internships and summer youth employment programs (see Acknowledgements).

The culmination of ARC database development is a functional museum database into which newly catalogued specimens are regularly added. Since this database is an archive of marine biodiversity data, concurrent ARC goals are to make the database available on-line, and to use it as a basis for developing a SIS for Canadian Atlantic marine biota.

Data on-line and the Species Information System

The ARC SIS is an evolving entity comprising products funded by various, usually unrelated, biodiversity information initiatives, and sometimes done in collaboration with other institutions (see Acknowledgements). In 2000 the ARC museum database was an integral component of a program to develop a system for integration, visualization, and analysis of distributed biogeographic and oceanographic information for marine populations using the internet and built-in tools for analysis (http://netviewer.usc.edu/web/index2.html). Another major outcome of this project was a web site for interactive mapping and listing, on a taxon basis, of the ARC's georeferenced holdings (http://gmbis.marinebiodiversity.ca/aconw95/aconscripts/ /ARCSpecimenMap4.html). This site was last updated in November 2002, and contains all ARC georeferenced records (approximately 91% of catalogued lots).

In 2001, production of the 'Distribution of rare, endangered and keystone marine vertebrate species in Bay of Fundy seascapes' (Bredin *et al.*, 2001) required the development of a preliminary list of fish species inhabiting Canadian Atlantic waters.

The next ARC SIS building block, in 2001, was an internet product entitled 'Fishes of Atlantic Canada: a photographic compendium' (http://collections.ic.gc.ca/compendium). The web site provides images, taxonomy and names, and important biological and ecological information for each fish species living in Canadian Atlantic waters. The fish list was developed from that described above.

Also in 2001, the ARC began the 'Development of a species information system for the Bay of Fundy'. The result was a database detailing taxonomy, general distribution and abundance, basic biological and ecological traits, conservation, economic, or other importance, and pertinent literature for annelids, crustaceans, echinoderms, and fishes inhabiting the Bay of Fundy. In 2002 this project continued as the 'Completion of a marine species information system for the Bay of Fundy', with the same details for phytoplankton, seaweeds, molluscs, remaining arthropods, birds, and mammals. During these projects, to facilitate future biodiversity data interoperability, standardized taxonomy was adopted using Sears (1998) for plants, the Integrated Taxonomic Information System (ITIS) (http://www.itis.usda.gov) for invertebrates, and FishBase (www.fishbase.org) for fishes.

Late in 2001, the ARC developed a framework for a registry of Canadian Atlantic marine biodiversity datasets (http://www.marinebiodiversity.ca/mbw/index.html). The DFO Maritimes Science Data Inventory, a compilation of regional departmental datasets, was modified to incorporate those from other federal and provincial government departments, universities, museums, nongovernmental organizations, industry, and the literature. This was a non-standards-based framework containing metadata for each electronic and non-electronic dataset describing ownership, contacts, content in detail, computerization, and availability. It is anticipated that this framework will contribute to the basis for a national plan for marine biodiversity (Zwanenburg *et al.* 2003), as well as to the ARC SIS.

In 2002 the ARC produced 'An atlas of distributions of Canadian Atlantic fishes' (http://collections.ic.gc.ca/FishAtlas), in which a modified distribution map was created for each ARC georeferenced species and placed in an electronic atlas. Each species is linked to the same species in the photographic compendium above. The species list for the atlas was much larger than that for the compendium, since the very large ARC ichthyoplankton collection, and thus many more species, had been added to the museum database. FishBase standardized fish taxonomy was adopted for all species. From this project the ARC SIS evolved, with the addition of the new fish species, standardized taxonomy, and inclusion of synonyms previously used by the ARC.

The most recent SIS development was a grant to integrate on-line the biodiversity databases of the ARC and other providers with DFO physical and biological databases, into a biodiversity atlas service. This atlas service will feature collection-based and integrated browsing and mapping of data, with links to services such as ITIS, FishBase, CephBase (http://cephbase.utmb.edu/), Hexacorals (http://www.kgs.ukans.edu/Hexacoral/), and internet images. Query results will be exportable in various formats, including input for on-line biodiversity analytical applications.

After presentation of this paper at the Colour of Ocean Data symposium, the ARC obtained funding in 2003 to 1) provide ARC museum database records for fishes in Darwin Core Version 2.0 format to Species Analyst Canada (http://www.cbif.gc.ca/speciesanalyst/fish/speciesquery_e.php), enabling on-line searching of ARC fishes; 2) develop a Canadian Atlantic marine biodiversity register containing a list and classification (based on Sears (1998) for plants, ITIS for invertebrates, and FishBase for fishes) of marine species in all major phyla inhabiting the Canadian Atlantic, and the range and concentrations of each species. This will become an on-line resource; 3) develop a web site for the Bay of Fundy species information system described above; and 4) perform comprehensive quality assurance and quality control of the ARC museum database, thus improving data quality in all components of the SIS.

Future plans involve growth of the ARC SIS. In 2003, ARC larval fish museum database records will be placed on LarvalBase (www.larvalbase.org), and on FishBase in 2003 or 2004. The Canadian Atlantic marine biodiversity register will be expanded to include taxonomy, identification, images, distribution and abundance, conservation status, basic biological and ecological traits, and commercial, scientific, educational, and/or social importance.

Discussion

Museum collections provide verifiable records of species distributions. Despite changes in knowledge about species identity over time, museum material can always be re-examined for confirmation or correction.

Collections with multiple samples in space and time are of even greater research value. For example, the material can be studied for population structure (size, growth rate, age, maturity), genetic variation, diet and parasites. The ARC collections are thus of value not only to taxonomists and systematists, but to researchers of genetics, population biology, fisheries, parasites, and ecology.

It is unfortunate that many marine laboratories and institutes have not curated their collections as a resource for future generations. Those with collections have difficulty keeping their catalogues up to date. A survey in Europe, as part of the European Register of Marine Species project (Costello 2000), found 20% of collections had no catalogues at all, and only 36% had all of their collections catalogued on paper. The rate of computerized catalogues was less, with only 10% with full coverage of their collection, and 36% with no computerized catalogues.

Little use can be made of museum collections if basic information on their content, in terms of species, numbers of individuals, and geographic origins, are not known. Ideally, this information is not just catalogued on paper, but in a computerized database searchable from the internet. Such on-line publication of the collection information will increase awareness and hence usage of the collections.

As a provider of biodiversity information the ARC goal is to complete the SIS and become a primary source of data for species inhabiting Canadian Atlantic waters. We will continue to develop biodiversity products like those detailed above, and others, such as electronic species atlases. This paper shows how it is possible to develop an integrated SIS using a series of short-term projects and variety of funding sources. This comprehensive and easily accessible assemblage of marine biodiversity information will aid students, educators, researchers and managers in studying, protecting, and promoting the sustainable use of Atlantic Canada's natural resources.

Acknowledgements

We would like to acknowledge the funding agencies that contributed to the ARC SIS: Fisheries and Oceans Canada, the Centre for Marine Biodiversity, the USA National Ocean Partnership Program, the Alfred P. Sloan Foundation, the Census of Marine Life, Environment Canada's Ecological Monitoring and Assessment Network, the Gulf of Maine Council on the Marine Environment, Industry Canada, the New Brunswick Environmental Trust Fund, Natural Resources Canada's GeoConnections program, Environment Canada's Canadian Information System for the Environment, the Global Biodiversity Information Facility, Human Resources and Development Canada, Canadian Heritage Young Canada Works, and the New Brunswick Student Employment and Experience Development Program. Collaborators in the SIS component projects were: Fisheries and Oceans Canada, the Specify Project, and the Atlantic Canada Conservation Data Centre. We especially would like to thank Fisheries and Oceans Canada for its support through the years for the ARC and for our biodiversity information collaborations and initiatives. The paper was reviewed and enhanced by an anonymous reviewer. Attendance at the Colour of Ocean Data conference was made possible by a grant to the Ocean Biogeographic Information System from the Sloan Foundation through Rutgers University to MJC.

References

Bredin K.A., S.H. Gerriets and L. Van Guelpen. 2001. Distribution of rare, endangered and keystone marine vertebrate species in Bay of Fundy seascapes. Prepared for the Gulf of Maine Council on the Marine Environment by the Atlantic Canada Conservation Data Centre. v + 73p.

Costello M.J. 2000. Developing species information systems: the European Register of Marine Species. Oceanography 13(3):48-55.

Poss S.G. and B.B. Collette. 1995. Second survey of fish collections in the United States and Canada. Copeia 1995(1):48-70.

Sears J.R. (Ed.). 1998. NEAS keys to the benthic marine algae of the northeastern coast of North America from Long Island Sound to the Strait of Belle Isle. A publication of the Northeast Algal Society, xi + 163p.

Zwanenburg K., K. Querbach, E. Kenchington and K. Frank. 2003. Development of a science plan for marine biodiversity in Canada. Canadian Technical Report of Fisheries and Aquatic Sciences 2432, viii + 72p.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 77-89
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Seamountsonline, an online information system for seamount biology

Karen I. Stocks

San Diego Supercomputer Center and Scripps Institution of Oceanography
9500 Gilman Drive, MC 0505, La Jolla, CA 92093-0505, USA
E-mail: kstocks@sdsc.edu

Abstract

Seamounts (undersea peaks in the ocean's floor) support unique biological communities that are of interest both as natural laboratories for studying marine biodiversity and as commercially-fished habitats requiring careful management. SeamountsOnline (http://seamounts.sdsc.edu) is a web resource through which researchers and managers can access global information on species recorded from seamounts. Users of the system can access species lists from seamounts of interest, distribution information on species of interest, sampling effort and method data, and a searchable bibliography of seamount literature. The system is based on a relational database that holds information in the following main tables: taxonomic names, seamounts, sampling events, and observations (the recording of a particular species from a particular sampling event). Features of the system include mechanisms for crediting data providers, versioning with archival of previous versions, referencing of sources at the data-element level, open source schema and scripts, and participation in the Ocean Biogeographic Information System (www.iobis.org). A brief example shows how the system has been applied to describing spatial patterns in fish endemism on the Hawaiian and Emperor seamount chains. The article is aimed at both potential users of the system interested in seamount data and researchers interested in database and online information system development for ecological data.

Keywords: Seamount; Ecological information system; Endemism; Island biogeography theory; Database design, Hawaiian seamounts.

Introduction

SeamountsOnline (http://seamounts.sdsc.edu) is a project aggregating existing data on seamount biota and making them freely available through an integrated online interface for research, management, education, and other non-commercial uses. This article describes the data content, functionality, and system design of SeamountsOnline.

Seamounts are undersea peaks in the ocean's floor – submerged mountains created by volcanism or tectonic uplift. Though seamounts are strictly defined as rising at least 1000m above the surrounding seafloor (International Hydrographic Organization, 2001), this definition is not particularly meaningful biologically, and SeamountsOnline includes data from smaller peaks. While the true number of seamounts is unknown, they have been found in every ocean basin and it is estimated that at least several tens of thousands exist (Smith, 1985).

Seamounts are of interest to ecology and biogeography because many support unique biological communities and high numbers of endemics (species found on only one seamount or restricted seamount chain and nowhere else in the oceans to date). Recent large studies have found rates of endemism of 30 to 40% and more on seamounts off Tasmania, New Caledonia, and Chile (Parin *et al.*, 1997; de Forges *et al.*, 2000; Koslow *et al.*, 2001), though other seamounts have lower rates (*e.g.* Fock *et al.*, 2002). It has long been proposed that seamounts may act as centers of speciation, refugia for relict populations, and/or stepping-stones for trans-oceanic dispersal (Hubbs, 1959). Because they vary in their community structure and levels of biodiversity, and also in physical factors thought to be important to biological communities (depth, latitude, distance from like habitat, primary productivity, dissolved oxygen, etc.), seamounts represent case studies for understanding some of the fundamental processes that create and promote biological diversity in the oceans. Synthesizing data across many seamounts is necessary to approach these questions. But while over 200 seamounts have been sampled biologically, rarely have data from more than one seamount chain been compared. SeamountsOnline is designed to provide integrated access to this heterogeneous and distributed body of information to facilitate synthetic studies.

At the applied level, seamount resource management is a pressing concern. Many seamounts can support high concentrations of commercial fish, particularly orange roughy (*Hoplosthesus atlanticus*), hoki (*Macruronus novaezelandiae*), oreos (Oreosomatidae), amourhead (*Pseudopentaceros wheeleri*) and rockfish (*Sebastes*) (Probert *et al.*, 1997; WWF/IUCN/WCPA, 2001; Dower and Perry, 2001). Many of these have not been managed sustainably (Koslow *et al.*, 2000, Stone *et al.*, 2003). Furthermore, bottom trawling has been demonstrated to be highly destructive to the epifaunal communities of deep corals, sponges, crinoids, hydroids, etc. found on seamounts (Koslow *et al.*, 2001). In some cases, corals may be over 100 years old (Rogers, 1999), indicating that the recovery time from disturbance may be extremely long. Countries including the United States, Australia, and New Zealand have begun siting marine protected areas on seamounts, and the United Nations Informal Consultative Process on the Ocean and the Law of the Sea is considering several proposals to create a policy mechanism for protecting seamounts in high-seas waters. Decisions about managing and protecting seamounts can be improved by access to the best available scientific information on seamounts.

SeamountsOnline (http://seamounts.sdsc.edu) began in 2000 with funding from the US National Science Foundation as a way to meet the information needs of seamount researchers and managers. It is a work in progress and it is continually expanding. While it was created specifically for seamount information, the overall system design could be applied to a variety of habitat-specific, species-level distribution datasets. Today, the number of groups seeking to share ecological information through the web is increasing, as is the time spent designing and implementing new systems. A description of the SeamountsOnline system is given here, along with a discussion of some of the main design decisions, in the hopes that it may prove useful to others setting up similar projects.

System description and funtionality

Data content

SeamountsOnline is designed to hold species-level observations from seamounts. The core data within the system consists of a record that a particular species was observed or collected on a particular seamount. This will be called an ‘observation’ throughout this paper. The database can hold observations where only a genus- or family- level identification is given, but does not hold information for observations identified to higher taxonomic levels or to other biological categories (*e.g.* ‘zooplankton’). All metazoan taxa are being included. Supporting each observation are data on how, where, and when the observation was made. Both presence/absence data and quantitative data can be accommodated.

SeamountsOnline is a continually-growing system, and the data content is by no means complete. At the time of writing (5/2003) the database held ~8500 observations on ~2000 species and ~110 seamounts (Fig. 1). Coverage is most comprehensive for fishes and crustacea on the Hawaiian/Emperor/Mid-Pacific seamounts, fishes and invertebrates from the Nasca/Sala-y-Gomez chain and the Norfolk Ridge seamounts, and fishes from the Great Meteor Seamount.

The majority of the data in the system have come from literature publications that have been hand-entered by the SeamountsOnline staff. For quality control, only data that have been published in peer-reviewed literature or by recognized government or research organizations are used. Additional valuable holdings have been provided as electronic datasets by seamount researchers and institutions. SeamountsOnline thanks those who have already stepped forward to contribute their electronic datasets: Bertrand Richer de Forges and collaborators for the ORSTOM data from the Norfolk Ridge (summarized in de Forges *et al.*, 2000), Heino Fock and collaborators for the Great Meteor Seamount fish data (Fock *et al.*, 2002), George Boehlert for ichthyoplankton data from Hancock Seamount (Boehlert and Mundy, 1992), and the Scripps Institution of Oceanography's Vertebrate and Benthic Invertebrate Collections (http://collections.ucsd.edu).

Functionality – using the system

The SeamountsOnline web site has three avenues through which users can retrieve data, as described below.

Search for species. Through this avenue, the user can specify 1) a species or genus of interest and/or 2) either a particular seamount by name or a geographic area by latitude and longitude bounds. By searching on a species or genus name the user will retrieve all seamount locations where that taxon has been found. By searching on a seamount name or region, the user will retrieve all species that have been recorded from that place. The default returned data table contains the taxonomic name (genus, species, subspecies [where applicable], and authority), seamount name, latitude and longitude with precision estimate, and the author(s) and publication year of the original data source. Optional fields that can be included are family and phylum name, depth of capture, date collected, number collected, and the full bibliographic citation for the source data. Each observation includes a sample number that is live-linked to the full sample information (see below). The results from a search can be downloaded either as a tab-separated text table or as an Excel file. Users can choose to download the default core of information, or can select additional optional fields to include. They can also choose to download the full sample information associated with each observation, either as one merged table or as separate species-observation and sample tables.

Search for samples. This avenue is designed to give information about how thoroughly a particular seamount has been sampled. This information is important for evaluating whether species lists are complete and representative, and whether data are comparable between seamounts. For example, if a seamount of interest has only been sampled with a bottom trawl, then the absence of a particular pelagic species from the observations does not necessarily mean that the species is not present there. To search for samples, the user selects a seamount of interest and retrieves a list of all the samples known from that seamount. The returned data for each sample includes, where available, the date(s) taken, the latitude and longitude location with precision, the depth, the depth zone (*i.e.* benthic or midwater), whether the sample was quantitative, the station or sample name/number given in the original data source, the gear used, the taxonomic groups recorded (*e.g.* 'only fish were counted'), the cruise and vessel, and the individual or institution taking the sample (Fig. 2). The full set of results from a given seamount can be downloaded as either a tab-separated text file or as an Excel file.

Search References. In the process of building SeamountsOnline, over 1000 bibliographic citations relating to seamounts have been collected and entered into a commercial reference manager database. Users can search this bibliographic database for authors, seamount names, or other terms of interest. The result of a search is a text list of references with the author, year, title and source given for each reference. This can be downloaded by copying from the screen and pasting to a local application. There is also a feature for downloading the entire bibliography as a text file in the same format as the screen return. The coverage is strongest for biological aspects of seamounts but also includes some references about seamount geology, hydrology, etc.

System design

SeamountsOnline is based on a relational database designed for this project. Two versions of the database exist: data are initially entered into a desktop version in Microsoft Access and then periodically ported over to an Oracle version for serving. The Access and Oracle versions differ in that the Access version is designed for maximum ease and quality control of data entry and the Oracle version (in which no data entries or edits are done) is designed to maximize query speed and performance. The query interface was programmed in Perl/CGI. All scripts are available upon request to the author (for non-commercial uses); the database schema can be viewed from the 'Database Design' menu button on the SeamountsOnline home page.

The primary tables of the relational database are described below. In addition, each table contains the following fields for use in data management: the source of the data (as a reference number that links to the bibliographic database), the person who entered the data, the date of entry, the person who checked the data for correctness, the date of checking, the date of last edit, in-house comments about data entry, and a field for general notes.

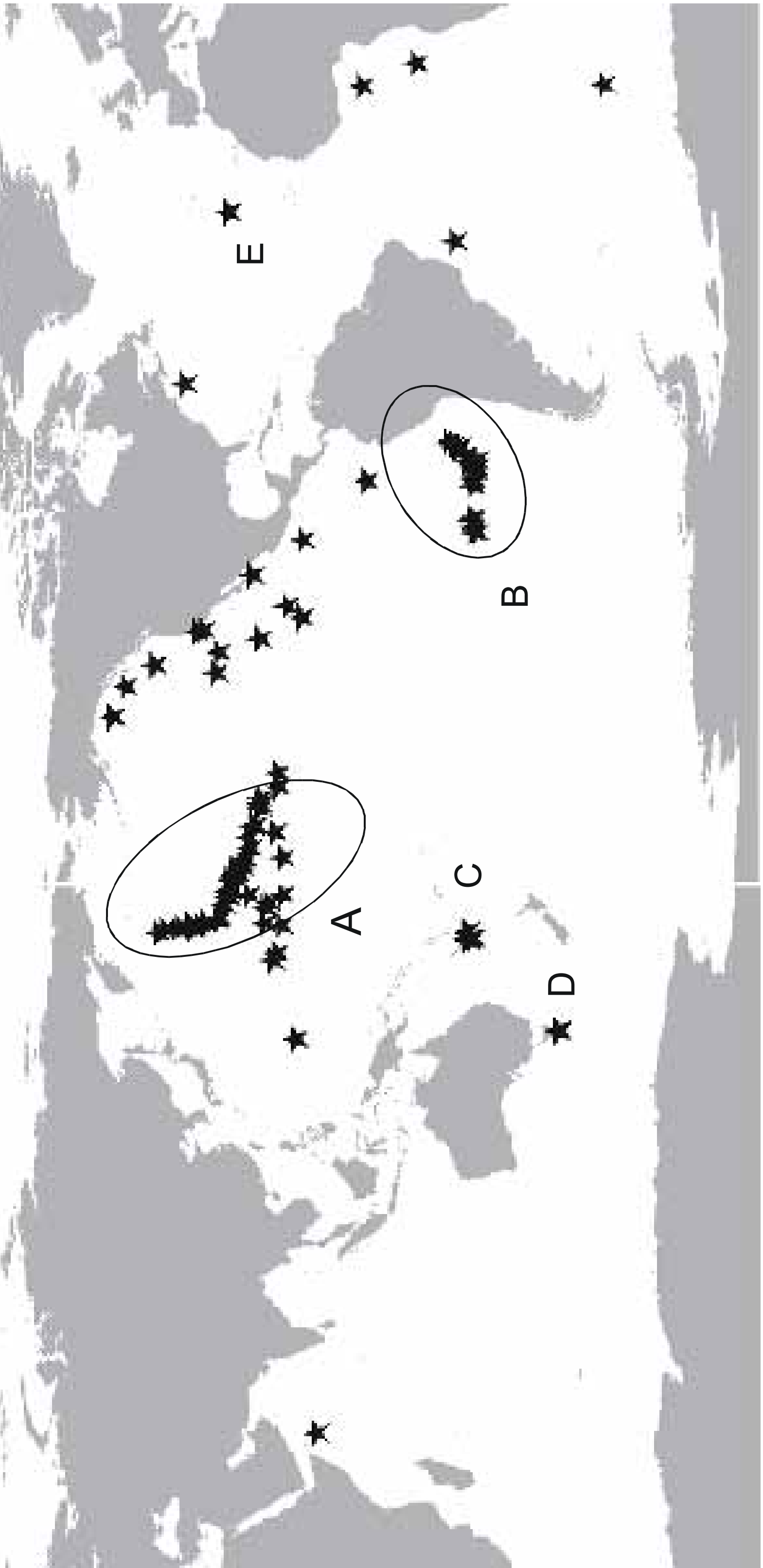


Fig. 1. Seamounts for which SeamountsOnline contains data. Seamount areas mentioned in the text: A = Hawaiian/Emperor chains, B = Nasca/Sala-y-Gomez chains, C = Norfolk Ridge seamounts, D = Tasmanian seamounts. E = Great Meteor Seamount.

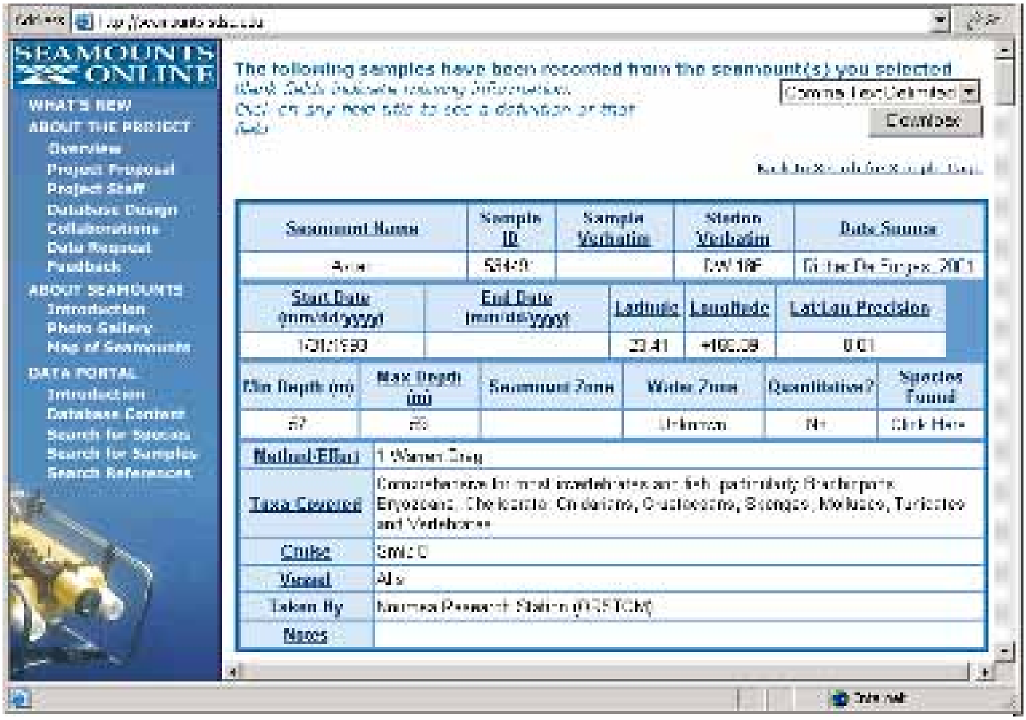


Fig. 2. Example of the data returned from a 'Search for Samples' (information from just one of multiple samples is shown).

Name. This table holds information about a taxonomic name including genus, species, subspecies (if applicable), and authority with year. In the Access version, the name is linked hierarchically to Family, Order, and Phylum tables. In the Oracle version, the higher taxonomic levels are incorporated into the Name table. All taxonomic names are entered as given in the original data source. The author is currently working on linking names to current valid names using the Catalog of Life (www.sp2000.org) and the taxonomic literature. Information on the geographic range of each species and whether it is a seamount endemic is also under development.

Seamount. This holds information about a particular seamount including name(s), seamount chain and region it is in, the central latitude and longitude with precision estimates, and the minimum depth. In development are fields for substrate type, geological age, seamount shape (*i.e.* plateau, pinnacle, etc.), surface area of the summit, and number of summits.

Sample. Information about each sampling event is held in this table. Sampling events can include both physical collections (net hauls, trap deployments, benthic cores, fisheries records, etc.) and visual observations (direct observation, video, and still photography). The data elements from this table are the same as those described in the 'Search for Samples' part of the Functionality section, above. This table has the flexibility to hold both point samples and line samples (*i.e.* net hauls with a start and end latitude and longitude). In development is a 'seamount zone' feature that characterizes whether the sample was taken from the top, sides, or base of a seamount. Each sample is linked to a seamount in the Seamount table.

Observation. This is the core of the database. It links a species from the Name table to a sample in the Sample table to record when and where a particular species was found. Supporting this are data, when available, on the person who made the identification, the identification guide/key used, the museum and accession number of any specimens that were preserved, the number found, and whether adults, juveniles, larvae, gravid females and/or eggs were recorded.

Supporting these tables are a few smaller tables including Expedition, which records information on major cruises; Institution, which gives the location and contact information for places where specimens are held; and Person, which gives the institution and address of people acting as identifiers of specimens or as chief scientists on cruises.

Design considerations

With recent improvements in internet connectivity and database technologies, efforts to integrate, organize, and distribute digital scientific data are increasing greatly. These efforts include large-scale projects like the Global Biodiversity Information Facility, but also many small-scale systems developed by individual researchers and small groups who may not have formal training in information system development. How these systems are structured and handled may greatly impact their utility and the support they receive from the research and management communities. Below are listed some of the system decisions that the SeamountsOnline team felt were critical to creating a robust system, and the rationale behind each decision. They are offered in the hope that they may assist others in developing their systems.

Credit for data providers. For large-scale or multi-taxa datasets it is rare that all the pertinent data are collected by a single research effort. By nature, therefore, database compilation often requires the integration of data from multiple sources. The data providers often have little incentive to share their data. While SeamountsOnline cannot compensate data providers, it does seek to give visible credit to the original providers of data. All data served through SeamountsOnline have a tag citing the original data source. Users of SeamountsOnline must agree to cite the original data source before they can search the system. While SeamountsOnline has no means to enforce this agreement, it at least assures data providers that the system is not seeking to take credit for their work.

Versioning. One of the primary purposes of SeamountsOnline is to act as a data resource for scientific research on seamount ecology. It is important for publishing researchers to be able to cite the data from which they drew their results, and for later researchers, perhaps trying to reproduce results, to be able to retrieve exactly those data. For this reason, data updates in SeamountsOnline are done periodically, each update is given a version number, and all previous versions are permanently archived at the San Diego Supercomputer Center and are available on request. Operationally, this is done by having a desktop version of the database which is continually updated with new data entries, and a separate version on the server that is updated intermittently from the desktop version.

Sampling effort information. In order to understand the distribution of a species it is important to know not just where it is found but also where it is not found. By tying observations to information, where available, on how the sample was taken and what groups were recorded (*e.g.* 'all cephalopods' or 'commercial fish only') SeamountsOnline assists users in determining whether the absence of a particular species is due to a lack of sampling or to a potentially true biological absence. In addition, it allows the users to re-create the species lists from particular samples to look at, for example, the rate of new species accumulation with increased sampling.

Habitat focus. SeamountsOnline is organized along the theme of a particular habitat type. By linking observations to an identified feature in the Seamount table, it allows information to be held about that feature, such as geological age, bottom type, etc. This same approach may be useful for other databases that focus on spatially-discrete geographical locations, such as islands, hydrothermal vents, trenches, seagrass beds, etc.

Data source referencing. All data items within SeamountsOnline are supported by a reference for the data source. This gives the flexibility to link multiple elements within one record of a table to different references (*i.e.* the source for the minimum depth of a seamount may be different from the source of the latitude and longitude of that seamount). References of many kinds, including unpublished datasets and web pages, can be recorded in the system.

Open data/Open source. SeamountsOnline has made a commitment to be an open-source system. All code and system designs are freely shared (for non-commercial uses only), and can be requested from the author. Most people developing ecological or biogeographic datasets are interested in ecology and not in information science and programming. While documenting scripts and database schema appropriately for an outside person to use takes time, the discipline overall will be best served by leveraging from past work.

Compatibility. As the ability and inclination of small groups to post their data on the internet grow, organizing elements will be necessary to create pathways through the large and heterogeneous volumes of data that emerge. The world of biological information is quickly organizing itself through new data warehouses and large 'umbrella' projects. Groups like the Ocean Biogeographic Information System (www.iobis.org), the Global Biodiversity Information Facility (www.gbif.org), and the US National Oceanographic Data Center (www.nodc.noaa.gov) are working to facilitate access to biogeographic data. Linking in or registering with these umbrella sites allows smaller datasets to be found and used more easily. To this end, SeamountsOnline has become a participant in the Ocean Biogeographic Information System, a globally-distributed network of marine systematic and ecological data resources (Zhang and Grassle, 2002). All data within SeamountsOnline can be searched and viewed through the main OBIS data portal at www.iobis.org. For information on taking part in OBIS, please visit the OBIS web site.

Example application: Hawaiian fish endemism

The data within SeamountsOnline are currently being used to look at patterns of endemism on the Hawaiian and Emperor Seamount chains (Fig. 1). Data from 17 published studies on these seamounts and one online data resource were entered into SeamountsOnline (Barsukov, 1973; JAMARC, 1973; Katayama, 1975; Chen, 1980; Gooding, 1980; Nakaya *et al.*, 1980; Kanayama, 1981; Dolganov, 1982; Nakabo *et al.*, 1983; Parin and Mikhailin, 1983; Yabe, 1983; Humphreys *et al.*, 1984; Uchida and Tagami, 1984; Randall and Chen, 1985; Borets, 1986; Uchida and Uchiyama, 1986; Chave and Mundy, 1994; Froese and Pauly, 2002). Then, the 'search for species' option was used to sequentially select seamounts of this chain and return all species observations associated with that seamount. For each seamount for which more than 10 fish species were recorded, the table of observations was downloaded and imported into Excel. From this a species list was created for each seamount.

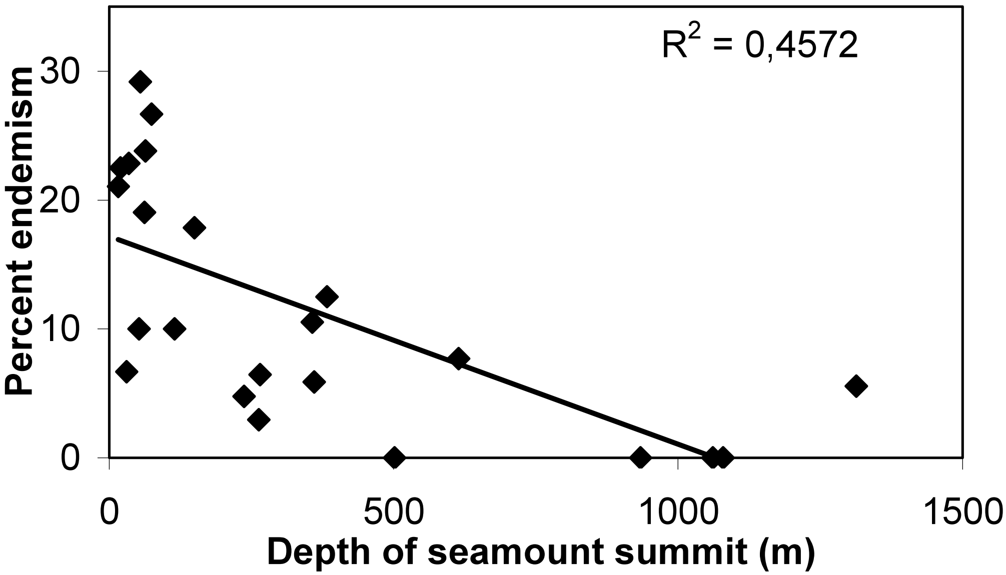
Then, each of the 213 fish species found was designated as a potential endemic or not based on the known distribution as found in the above sources plus: Jordan and Evermann (1973), Tinker (1978), Boehlert and Sasaki (1988), Blum (1989), Masuda *et al.* (1992), Pequeño (1997), Randall (1998), Castle and Smith (1999), Randall (1999), Randall and Lim (2000), Shinohara *et al.* (2001). Endemics were separated into two categories: those found only on seamounts of the Hawaiian/Emperor chains, and those found on these seamounts and also around the Hawaiian Islands. (Endemicity information is currently being added to the SeamountsOnline database, but is not yet available through the web site.) From these data, the total percent endemic species was calculated for each seamount. Absolute numbers of endemics were not evaluated because the sampling effort varied greatly across seamounts and in some cases was not well documented. Percent endemic species was regressed against both the depth of the seamount summit and the distance from the main island of Hawaii as measured in degrees of latitude/longitude. Regressions were run using Statistix 7 analytical software (www.statistix.com).

Overall, 5% of the fish species found on these seamounts are known only from the seamounts of the Hawaiian/Emperor Chains and an additional 12% are known only from these seamounts or the Hawaiian Islands, for a total of ~17% endemic to the area. For those species endemic to both the islands and seamounts of Hawaii there is a significant relationship with distance away from the main island of Hawaii and with depth of the seamount (Fig. 3) ($P<0.01$ for both). Note that depth and distance are correlated on this seamount chain. The relationship with distance is what would be predicted from a simple application of Island Biogeography Theory (MacArthur and Wilson, 1967), assuming that the islands of Hawaii, which have much more habitat area than the small seamount peaks, represent the main pool of species. While a causative factor cannot be determined from this work, it is an example of how a clear spatial pattern can emerge from the compilation of datasets that are each too small to show any pattern alone.

The future of SeamountsOnline

SeamountsOnline is a continually-growing project designed as a permanent archive and distributor of data from seamounts. The San Diego Supercomputer Center has committed to providing permanent server access to this database, and the Ocean Biogeographic Information System will perpetuate data maintenance in the event the author no longer can.

However, the growth of the SeamountsOnline dataset, and thus its utility to the community, will rely greatly on the willingness of researchers to share their datasets. While data can be entered by hand from literature publications, literature data are often summarized and lose resolution. Furthermore, manual entry of data from the literature is extremely time consuming. If researchers are willing to provide electronic datasets (in any format) to SeamountsOnline, the system will be able to incorporate them efficiently and continue to grow in the coming years. SeamountsOnline is grateful to those researchers, named in the Data Contents section above, who have already given data to the system. *If you have species-level sampling data from seamounts and are interested in providing them to SeamountsOnline, please contact the author.*



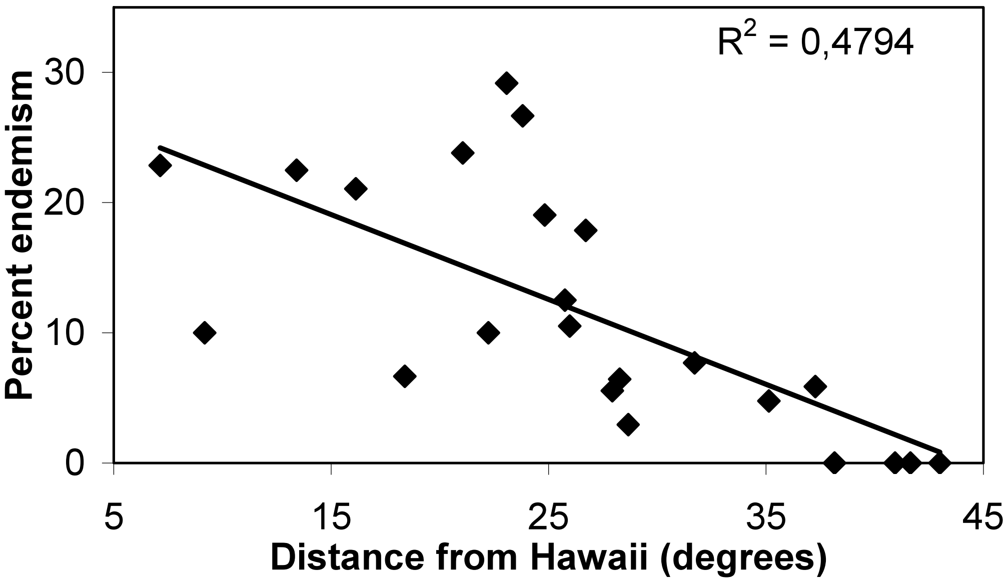


Fig. 3. Percent endemism in fish species from seamounts of the Hawaiian and Emperor chains compared to A) depth of the seamount summit and B) distance from mainland Hawaii in degrees. R^2 values are for linear regressions. $P<0.01$ for both regressions.

Acknowledgements

SeamountsOnline was funded by National Science Foundation Division of Biological Infrastructure grant DBI-0074498 and server support and other infrastructure is provided by the San Diego Supercomputer Center. Lisa Levin and David Stockwell at the University of California at San Diego have given valuable assistance as have Phoebe Zhang and Fred Grassle from the OBIS secretariat. Daphne Fautin was kind enough to provide a schema of her ‘Biogeoinformatics of Hexacorals’ database (www.kgs.ukans.edu/Hexacoral/) to assist in the design of SeamountsOnline. John Lin lent his talents to both web design and Perl/CGI programming, and Megan Williams and Elisheba Spiller contributed to database content development.

The project also gratefully acknowledges the researchers who have freely provided their data to SeamountsOnline: Bertrand Richer de Forges and collaborators for the ORSTOM data from the Norfolk Ridge, Heino Fock and collaborators for Great Meteor Seamount fish data, George Boehlert for ichthyoplankton data from Hancock Seamount, and the Scripps Institution of Oceanography's Vertebrate and Benthic Invertebrate Collections.

References

Barsukov V.V. 1973. The species composition of the genus *Helicolenus* (Sebastinae, Scorpaenidae, Pisces) and a description of a new species. Journal of Ichthyology 13:161-167.

Blum S.D. 1989. Biogeography of the Chaetodontidae: an analysis of allopatry among closely related species. Environmental Biology of Fishes 25: 9-31.

Boehlert G.W. and B.C. Mundy. 1992. Distribution of ichthyoplankton around Southeast Hancock Seamount, central North Pacific, in summer 1984 and winter 1985: data report. NOAA Technical Memorandum NMFS-SWFSC-176. 109p.

Boehlert G.W. and T. Sasaki. 1988. Biogeography of the armorhead, *Pseudopentaceros wheeleri*, and recruitment to isolated seamounts in the North Pacific Ocean. Fishery Bulletin 86:453-465.

Borets L.A. 1986. Ichthyofauna of the Northwestern and Hawaiian ridges. Journal of Ichthyology 26:1-13.

Castle P.H.J. and D.G. Smith. 1999. A reassessment of the eels of the genus *Bathycongrus* in the Indo-west Pacific. Journal of Fish Biology 54: 973-995.

Chave E.H. and B.C. Mundy. 1994. Deep-sea benthic fish of the Hawaiian archipelago, Cross Seamount, and Johnston Atoll. Pacific Science 48:367-409.

Chen C.H. 1980. Primary studies of demersal fish resources investigations on trawl grounds at Kanmu Seamount. Bulletin of Taiwan Fisheries Research Institute 32: 317-338.

de Forges B., J.A. Koslow and G.C.B. Poore. 2000. Diversity and endemism of the benthic seamount macrofauna in the southwest Pacific. Nature 405:944-947.

Dolganov V.N. 1982. Discovery of a rare rabbitfish species, *Chimaera owstoni* (Holocephali, Chimaeridae), in the region of the Northwestern underwater ridge. Journal of Ichthyology 22:145-146.

Dower J.F. and R.I. Perry. 2001. High abundance of larval rockfish over Cobb Seamount, an isolated seamount in the Northeast Pacific. Fisheries Oceanography 10:268-374.

Fock H., F. Uiblein, F. Köster and H. von Westernhagen. 2002. Biodiversity and species-environment relationships of the demersal fish assemblage at the Great Meteor Seamount (subtropical NE Atlantic), sampled by different trawls. Marine Biology 141:185-199.

Froese R. and D. Pauly (Eds). 2002. FishBase. World Wide Web electronic publication. www.fishbase.org. 12 December 2002.

Gooding R.M. 1980. Exploratory fishing on the Hancock Seamounts by the Townsend Cromwell, 1976-79. p.241-250. In: Proceedings of the Symposium on Status of Resource Investigations in the Northwestern Hawaiian Islands, April 24-25, 1980. Grigg R.W. and R.T. Pfund (Eds). Honolulu, Hawaii.

Hubbs C.L. 1959. Initial discoveries of fish faunas on seamounts and offshore banks in the Eastern Pacific. Pacific Science 13:311-316.

Humphreys R.L. Jr., D.T. Tagami and M.P. Seki. 1984. Seamount fishery resources within the southern Emperor - northern Hawaiian Ridge. p.283-327. In: Second Symposium on Resources Investigations in the Northwestern Hawaiian Islands, 25-27 May 1983. Grigg R.W. and K.Y. Tanoue (Eds). Honolulu, Hawaii.

International Hydrographic Organization. 2001. Standardization of undersea feature names. Bathymetric Publication No. 6, 3rd Edition. International Hydrographic Bureau.

JAMARC (Japan Marine Fishery Resource Research Center). 1973. Surveys of trawling grounds in the north central Pacific, 1972 season. JAMARC Report No. 7.

Jordan D.S. and B.W. Evermann. 1973. The shore fishes of Hawaii. Charles E. Tuttle Company, Rutland, Vermont. 392p.

Kanayama T. 1981. Scorpaenid fishes from the Emperor Seamount Chain. Research Institute of North Pacific Fisheries, Special Volume 56:119-129.

Katayama M. 1975. *Caprodon unicolor*, a New Anthiine Fish from the North Pacific Ocean. Japanese Journal of Ichthyology 22:13-15.

Koslow J.A., G.W. Boehlert, J.D.M. Gordon, R.L. Haedrich, P. Lorance, and N. Parin. 2000. Continental slope and deep-sea fisheries: implications for a fragile ecosystem. ICES Journal of Marine Science 57: 548-557.

Koslow J.A., K. Gowlett-Holmes, J.K. Lowry, G.C.B. Poore, and A. Williams. 2001. The seamount benthic macrofauna off southern Tasmania: community structure and impacts of trawling. Marine Ecology Progress Series 213:111-125.

MacArthur R.H. and E.O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, NJ. 203p.

Masuda H., K. Amaoka, C. Araga, T. Uyeno and T. Yoshino. 1992. The Fishes of the Japanese Archipelago, English Edition ed. Tokai University Press, Tokyo. 456p.

Nakabo T., T. Yamamoto and C.H. Chen. 1983. Two new species of the genus *Foetorepus* (Callionymidae) from the Emperor Seamounts, north-central Pacific. Japanese Journal of Ichthyology 29:349-354.

Nakaya K., K. Amakoa and K. Abe. 1980. A review of the genus *Lepidion* (Gadiformes, Moridae) from the Northwestern Pacific. Japanese Journal of Ichthyology 27:41-47.

Parin N.V. and S.V. Mikhailin. 1983. *Lepidopus calcar*, a new trichiurid fish from the Hawaiian Underwater Ridge. Japanese Journal of Ichthyology 29:27-30.

Parin N.V., A.N. Mironov and K.N. Nesis. 1997. Biology of the Nazca and Sala y Gomez submarine ridges, an outpost of the Indo-West Pacific fauna in the Eastern Pacific Ocean: composition and distribution of the fauna, its communities and history. Advances in Marine Biology 32:145-242.

Pequeño G. 1997. Peces de Chile. Lista sistemática revisada y comentada: addendum. Revista de Biologia Marina y Oceanografia 32:77-94.

Probert P.K., D.G. McKnight and S.L. Grove. 1997. Benthic invertebrate bycatch from a deep-water trawl fishery, Chatam Rise, New Zealand. Aquatic Conservation: Marine and Freshwater Ecosystems 7:27-40.

Randall J.E. 1998. Shore Fishes of Hawaii. University of Hawai'i Press, Honolulu, Hawaii. 216p.

Randall J.E. 1999. Review of the Dragonets (Pisces: Callionymidae) of the Hawaiian Islands, with descriptions of two new species. Pacific Science 53:185-207.

Randall J.E. and C.H. Chen. 1985. First record of the labrid fish *Bodianus cylindriatus* (Tanaka) from the Hawaiian Islands. Pacific Science 39:291-293.

Randall J.E. and K.K.P. Lim. 2000. A checklist of the fishes of the South China Sea. Raffles Bulletin of Zoology 8 Suppl.:569-667.

Rogers A.D. 1999. The biology of *Lophelia pertusa* and other deep-water reef-forming corals and impacts from human activities. International Review of Hydrobiology 84:315-406.

Shinohara G., H. Endo, K. Matsuura, Y. Machida and H. Honda. 2001. Annotated checklist of the deepwater fishes from Tosa Bay, Japan. p.283-343. In: Deep-sea fauna and pollutants in Tosa Bay. Fujita T., H. Saito and M. Takeda (Eds). National Science Museum, Tokyo. 380p.

Smith D.K. 1985. The statistics of seamount populations in the Pacific Ocean. PhD thesis, University of California San Diego, San Diego, California. 212p.

Stone G., L. Madin, K. Stocks, G. Hovermale, P. Hoagland, M. Schumacher, C. Steve and H. Tausig. 2003. Seamount biodiversity, exploitation and conservation. Presentation, proceedings. Defying Ocean's End Conference, Cabo San Lucas, Mexico, May 29 - June 3, 2003.

Tinker S.W. 1978. Fishes of Hawaii. Hawaiian Service, Inc., Honolulu, Hawaii. 532p.

Uchida R.N. and D.T. Tagami. 1984. Groundfish fisheries and research in the vicinity of seamounts in the North Pacific Ocean. Marine Fisheries Review 46:1-17.

Uchida R.N. and J.H. Uchiyama. 1986. Fishery atlas of the Northwestern Hawaiian Islands. NOAA Technical Report 38:1-142.

WWF/IUCN/WCPA. 2001. The status of natural resources on the high-seas. WWF/IUCN, Gland, Switzerland. 93p.

Yabe M. 1983. A new cottoid fish of the family Ereuniidae, *Marukawichthys pacificus*, from the central North Pacific. Japanese Journal of Ichthyology 30:18-26.

Zhang Y.Q. and J.F. Grassle. 2002. A portal for the ocean biogeographic information system. Oceanologica Acta 25:193-197.

SESSION: ECOLOGICAL AND COMMUNITY DATA

2004. p93-104
25-27 2002
88 x+ 308 p
16

An overview of the Fisheries Resource Information System and Tools (FiRST) version 2001: a database management system for storing and analyzing trawl survey data

Abstract

21,000

2001)

Introduction

eg. 1900s

1996).

301

1920s 1990s 70,000 1997).

In 1996, a [illegible]

A [illegible]

([illegible] 1997). A [illegible]

[illegible] *et al.*, 1997).

From 1998 to 2001, [illegible]

[illegible] 5766)'- [illegible] (1) [illegible] (2) [illegible] (3) [illegible]

[illegible] *et al.*, 2000). [illegible]

[illegible] ://[illegible]

[illegible] 2000), [illegible] *et al.*, 2000). [illegible]

[illegible] 2001) [illegible]

[illegible] *ie.* [illegible]

The Fisheries Resource Information System and Tools (FiRST)

[illegible] 2001) is a [illegible]

[illegible] *et al.* (2001).

[illegible]

- [illegible] 95 [illegible] 98;
- [illegible] 64 [illegible]
- [illegible] 1024 x 768 [illegible]
- [illegible] 5 [illegible] 12 [illegible]

[illegible] 1 [illegible] 2) [illegible] *eg.*, [illegible]

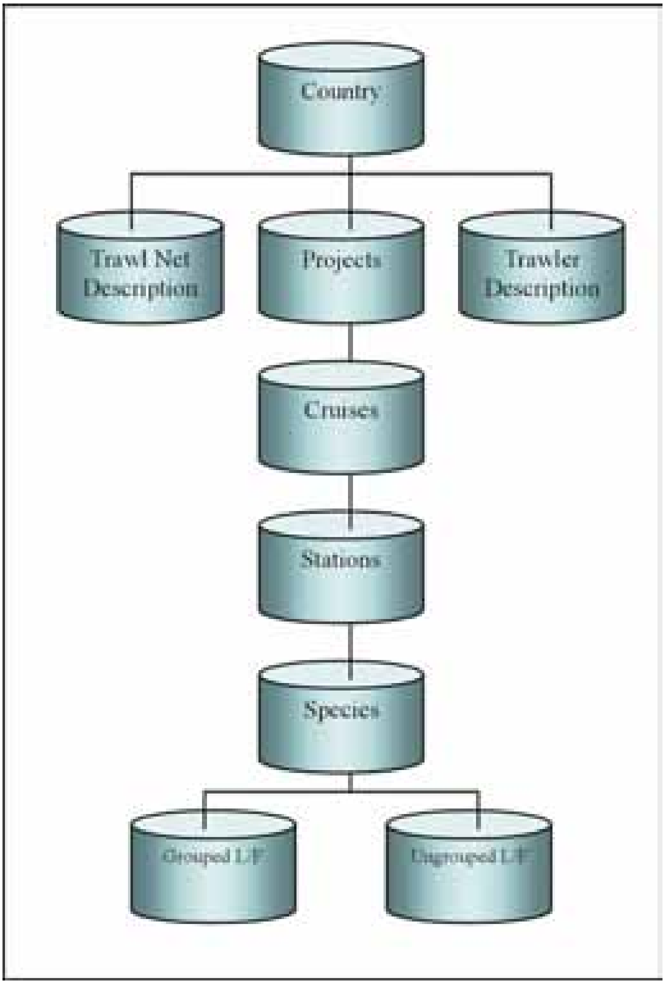


Fig. 1.
Schematic representation of the main tables in FiRST (ver. 2001) and their relationships. (Note: L/F means length frequency data from length based assessments).

Table I. Main tables of FiRST (ver. 2001), their contents and functions (adopted from Gayanilo *et al.*, 2001)

[illegible]	[illegible]
[illegible]	[illegible] 99. [illegible] *ie.* [illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible] (*eg.* [illegible])
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible]
[illegible]	[illegible] *Species* [illegible]
[illegible]	[illegible]

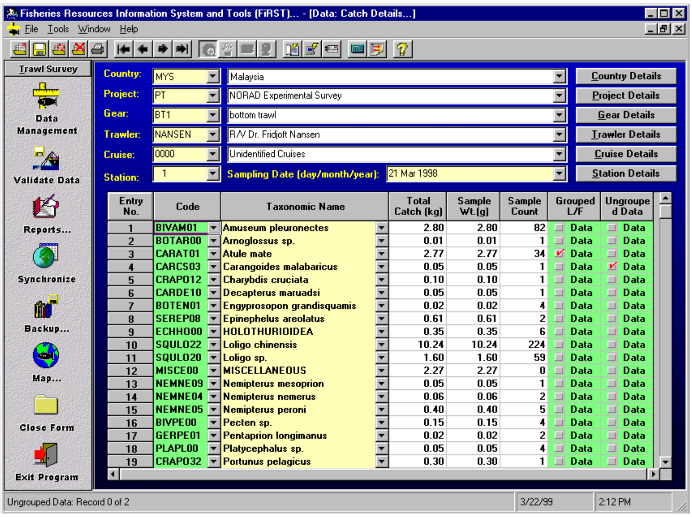


Fig. 2.
The catch form of FiRST (ver. 2001). Note that the last two columns indicate presence of related data.

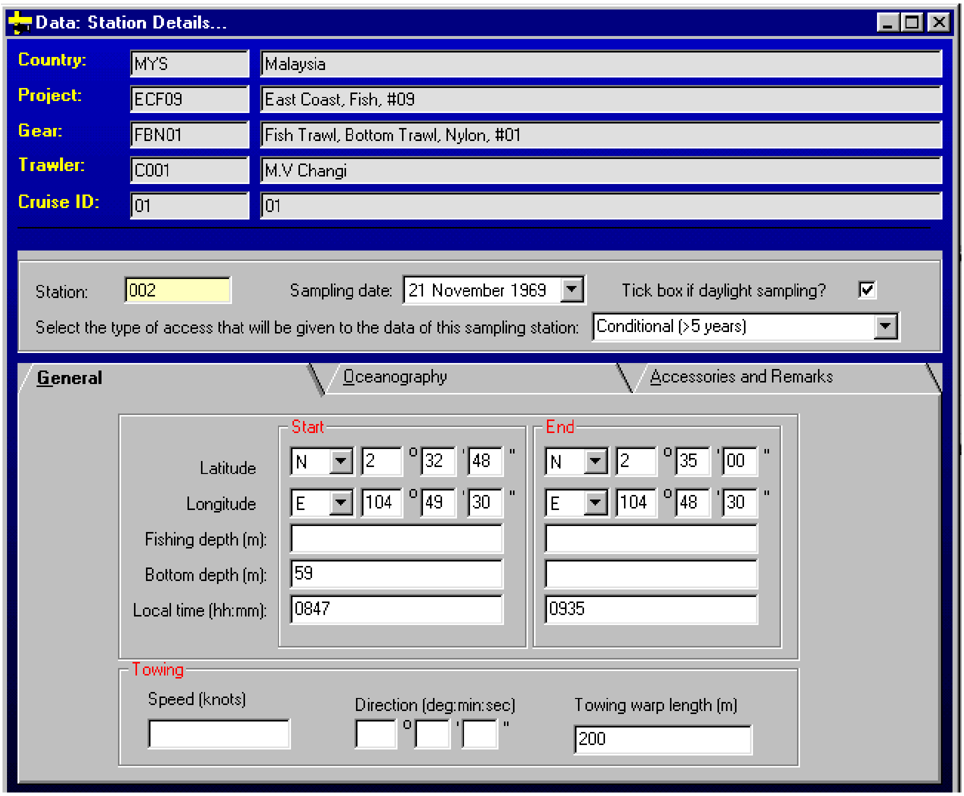


Fig. 3. The station form of FiRST (ver. 2001) to record station-related information.

[illegible]

[illegible]

[illegible]

et al., 2001): (1) [illegible] (2) [illegible] 5 [illegible] (3) [illegible] e- [illegible] (p/ [illegible]

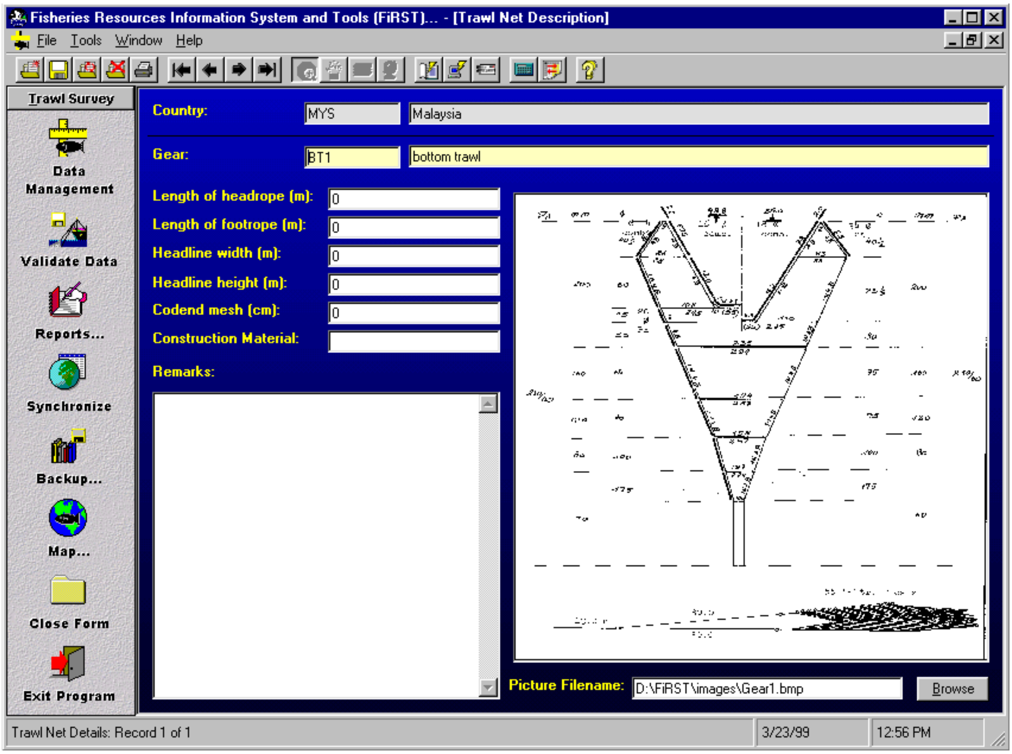


Fig.4. The gear details form in FiRST (ver. 2001). Note that a scanned picture of the gear used for the trawl survey can be stored.

[illegible] 2001). [illegible] *et al.*, 1996) [illegible] *et al.*, 2000) [illegible]

[illegible] 20,620 [illegible] 5 [illegible] 2001). [illegible] 2001 [illegible]

[illegible]

Table II. Available analytical modules in FiRST (version 2001)

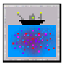
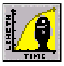
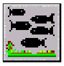
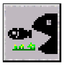
[illegible]	[illegible]	
	[illegible]	
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]

Fig. 5. Geographical coverage of trawl survey data in FiRST (ver. 2001).

Table III. List of trawl surveys contained in the database system (FiRST ver. 2001)

[illegible]	[illegible]	[illegible] [illegible]	[illegible] [illegible]	[illegible]
[illegible]				
[illegible] [illegible] [illegible]	, #01	13	341	1926
[illegible] [illegible] [illegible]	, #02	9	262	1927
[illegible] [illegible] [illegible]	, #04	4	153	1967
[illegible] [illegible] [illegible]	, #07	6	148	1972
[illegible] [illegible] [illegible]	, #09	37	822	1969-1973
[illegible] [illegible] [illegible]	, #11	4	85	1981-1985
[illegible] [illegible] [illegible]	, #13	6	130	1984
[illegible] [illegible] [illegible]	, #16	5	114	1986
[illegible] [illegible] [illegible]	, #19	5	186	1991
[illegible] [illegible] [illegible]	, #01	3	64	1926
[illegible] [illegible] [illegible]	, #02	13	303	1927
[illegible] [illegible] [illegible]	, #05	4	121	1971
[illegible] [illegible] [illegible]	, #06	4	79	1972
[illegible] [illegible] [illegible]	, #07	4	105	1973
[illegible] [illegible] [illegible]	, #08	4	93	1974
[illegible] [illegible] [illegible]	, #11	4	103	1980
[illegible] [illegible] [illegible]	, #12	4	192	1981
[illegible] [illegible] [illegible]	, #13	3	44	1984
[illegible] [illegible] [illegible]	, #15	1	50	1987
[illegible] [illegible] [illegible]	, #16	4	52	1988
[illegible] [illegible] [illegible]	, #18	5	61	1990-1991
[illegible] [illegible] [illegible]	, #19	4	32	1992-1993
[illegible] – [illegible] [illegible]	, [illegible]#01	3	87 1927	
[illegible] – [illegible] [illegible]	, [illegible]#07	10	300 1972	
[illegible] – [illegible] [illegible]	, [illegible]#11	6	134 1981	
[illegible] – [illegible] - [illegible]	, #12	3	141	1986
[illegible] – [illegible] - [illegible]	, #14	1	17	1986
[illegible] -	[illegible]#15 12	96	1989-1993	
	[illegible] [illegible] [illegible]	177	4,418	
[illegible]				
[illegible] - [illegible]	14 60 1975, 1979			
[illegible] p [illegible] - [illegible]	[illegible]	24	157 1947-1949	
[illegible] r	- [illegible] y	11	300	1980
[illegible] [illegible] y	- [illegible] [illegible] y	22	64	1992-1993
[illegible] gay [illegible] - [illegible]	gay [illegible]	1	62 1994-1995	
[illegible] y	- [illegible] y	17	158	1994, 1995
[illegible] - [illegible]	36	37 1992, 1993,		1995, 1996
	[illegible] 125 838			
[illegible] : [illegible]	- [illegible] [illegible]	106	5,890	1968 –1976 1985 –1986 1991, 1993, 1995
[illegible] : [illegible] [illegible]	- [illegible] 42	925 1969 –1973		
[illegible] : [illegible]	- [illegible] 2	1,376		1974 –1979
[illegible] : [illegible]	- [illegible] [illegible]	4	375	1979 –1980
	- [illegible] 32 1		,021	1984 –1987

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible] : [illegible]	- [illegible]	12	324	1988
	- [illegible]	7	90	1980
[illegible] : [illegible]	- [illegible]	4	15	
	[illegible]	55	1,450	
[illegible] : [illegible]	- [illegible]	12	613	1994 – 1995
	- [illegible]	13	393	1920, 1921, 1923
[illegible] - [illegible]	[illegible]	3	225	1978 – 1980
	[illegible] 16 618			
[illegible] : [illegible]	- [illegible]	78	3894	1979 – 1982, 1988
[illegible] - [illegible]	[illegible]	6	127	1993 – 1995
	[illegible] 84 4,021			
[illegible] : [illegible]	- [illegible]	5	96	1976
	[illegible]		[illegible]	

Table IV. Summary of the standard workbooks and worksheets as provided by FiRST (version 2001) to store socioeconomic an related information

[illegible] [illegible]	[illegible]	[illegible]
[illegible]	A1	[illegible]
	A2.1	[illegible]
	A2.2	[illegible]
	A2.3	[illegible]
	A2.4a	[illegible]
	A2.4b	[illegible]
	A2.5	[illegible]
Fleet Operation Dynamics	B1a	[illegible]
	B1b	[illegible]
	B2a	[illegible]
	B2b	[illegible]
	B3.1	[illegible]
	B3.2a	[illegible]
	B3.2b	[illegible]
	B3.3	[illegible]
	B3.4	[illegible]

[illegible]	[illegible]	[illegible]
[illegible]	B3.5	[illegible]
	B4	[illegible]
	B5a	[illegible]
	B5b	[illegible]
[illegible]	C1.	[illegible]
	C2.	[illegible]

Some illustrative examples of the results of the analysis

[illegible] 5 to 30% [illegible] (*ie.* 0.30-0.5) [illegible]

Table V. Some estimates of the declines in demersal biomass from trawl surveys in Asian countries (adopted from Garces *et al.*, 2001)

[illegible]	[illegible]	[illegible] (kg/km^2)	[illegible] (%)	[illegible]
[illegible]				
[illegible]	1947	10.60	100.0	[illegible] (1950)
	1980-81	2.13	20.1	[illegible] (1982)
	1992-93	1.96	18.5	[illegible] *et al.* (1995)
[illegible]	1949-52	4.61	100.0	[illegible] (1950)
	1992-93	0.47	10.2	[illegible] (1995)
[illegible]				
[illegible]	1977	3.72	100.0	[illegible] (2001)
	1998	2.20	59.1	
[illegible]				
[illegible]	1971/72	2.31	100.0	[illegible] *et al.* (2003, [illegible])
	1997	0.36	15.6	
[illegible]	1972	5.09	100.0	[illegible] *et al.* (2003, [illegible])
	1998	0.20	3.9	

[illegible]	[illegible]	[illegible]	[illegible] (km^2)	[illegible] (%)	[illegible]
[illegible]	1972	3.90	100.0	[illegible] *et al.* (2003, [illegible])	
	1998	1.11	28.5		
[illegible]	1986	1.52	100.0	[illegible] *et al.* (2003, [illegible])	
	1998	0.87	57.2		
[illegible]					
[illegible]	1961	0.70*	100.0	[illegible] *et al.* (2003, [illegible])	
	1991	0.10*	14.2		

* [illegible] x10[6].

[illegible] ([illegible]) [illegible] y [illegible] (*ie.* [illegible].

[illegible] (*eg.* [illegible] 50m [illegible] 100m [illegible]

Conclusion

[illegible] 21,000 [illegible] A [illegible] (*ie.* 5 to 30% [illegible]. [illegible] A [illegible] a [illegible] to [illegible]

[illegible] ([illegible] 2001) [illegible] V [illegible] C [illegible] A [illegible] R [illegible] T [illegible]

[illegible]

References

[illegible], A. M[illegible] S [illegible] 2003. ([illegible]). [illegible] A [illegible] C [illegible]

[illegible]
[illegible] 2000. [illegible] 2000 [illegible]
[illegible]
[illegible] 130p
[illegible] 1995. [illegible]
[illegible] (12992-1993). [illegible]
[illegible] 47 [illegible]
[illegible] 2000. F [illegible] 2000: [illegible]
[illegible] 2001. [illegible]
[illegible] p 15-23. [illegible]
[illegible]
[illegible] 1: [illegible]
[illegible]
[illegible] 1996. [illegible]
[illegible] 8. 126p
[illegible] 1997. [illegible]
[illegible] p 166-132. [illegible]
[illegible] 53. 208p
[illegible] 2001. [illegible]
[illegible] 7. 126p
[illegible] 2003. [illegible]
[illegible]
[illegible]
[illegible]
[illegible] 1995. [illegible]
[illegible]
[illegible] 2001. [illegible]
[illegible]
[illegible] 1980. [illegible]
234. 54p
[illegible] 1996. [illegible]
[illegible] p 1-16. [illegible]
[illegible]
[illegible] 3. 312p
[illegible] 1997. [illegible]
[illegible] p 1-7. [illegible]
[illegible] 53. 208p
[illegible] 2000. [illegible]
[illegible] 3(1):45-47.
[illegible] 2002. [illegible]
[illegible] (1989-1990). [illegible]
[illegible] 1982. [illegible] p 65-94. [illegible]
[illegible] 7. 124p
[illegible] *et al.* 2000. [illegible] 2000: [illegible]
2000. [illegible]
[illegible] 1950. [illegible] 25. [illegible]
[illegible] 149p

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 105-113
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Analysis of demersal fish assemblages on the Senegalese continental shelf considering fishing impact over the decade 1986-1995

Didier Jouffre[1], Gilles Domalain[1], Alain Caveriviere[1] and Djiby Thiam[2]

[1]Centre de Recherche Halieutique Méditerranéenne et Tropicale
rue Jean Monnet, BP 171, F-34203, Sète Cédex, France
E-mail: jouffre@ird.fr

[2]Centre de Recherche Océanographique de Dakar-Thiaroye
BP 2241, Dakar, Senegal

Abstract

The assemblages of demersal fish and associated species of the Senegalese continental shelf are analyzed in order to characterize their diversity and evolution during the 1986-1995 period. The biological data come from an historical series of scientific trawling surveys. The analytical process (dominance curves), as well as the interpretation of the results, support an evaluation of the ecosystems effects of fishing, in a zone which has undergone an increasing pressure of commercial exploitation for several decades. The stability observed over the period could be interpreted as an adaptation of the studied community to a relatively strong fishing pressure.

Keywords: Demersal fish; Community structure; Fisheries impact; Dominance curves; West Africa.

Introduction

The question of the assessment of fishing impacts on ecosystems is nowadays a topic of increasing interest for fisheries management purposes (Pauly *et al.*, 1998, 2000; ICES, 2000). This new approach to the management of fishing fits in a more general context, namely the global evaluation of the human impact on the environment which has led to the obligation of preserving biodiversity (Convention of Biological Diversity, 1992). In this context, methods of measurement of biodiversity making it possible to establish a diagnosis of changes in biodiversity are required (Blanchard, 2000). Among those, dominance curves and derived Abundance Biomass Comparison (ABC) plots (Clarke and Warwick, 1994) are often used.

The present study proposes the application of these methods to the demersal communities of Senegal, which appear to be a convenient case for exploring and discussing its applicability in tropical environments. Indeed:
- The demersal communities of the Senegalese continental shelf are submitted to a strong fishing pressure since several decades (Chauveau, 1985; SIAP, 2002; Chavance, 2002).
- Several scientific trawling surveys have been carried out over several years (Gascuel and Guitton, 2001) on these grounds of great economic interest (La pêche maritime, 1973; Gascuel

and Ménard, 1997). The present study uses one of the most complete and homogeneous dataset collected in the zone (*i.e.* a sampling serie by the RV/Louis Sauger during the 1986-1995 decade).

Material and method

Biological data

The biological data were collected by bottom-trawl scientific surveys. These data are exhaustively stored in homogeneous format (software First-SIAP, Gascuel and Guitton, 2001) in the Senegalese database of the SIAP project (Ba, 1999), from which the information used here was extracted.

- This sampling concerns 13 surveys between 1986 and 1995 (Table I), each one covering the whole Senegalese continental shelf (about 100 stations/survey).
- The sampling strategy for each campaign followed a stratified-random plan.
- The analyses were made on the standardised abundances and biomasses index (catch in number and kg/30' haul) of the sampled species (see list in Table II).
- The majority of the taxa were determined at the specific level, for a few taxa the determination was generic (genus or group of several species).
- For several sampling stations and some species, only the biomass (kg) is available (the abundance (N) was not recorded). In those cases, the abundance is estimated using the value of the biomass (per 30' haul) divided by the 'mean individual weight' of the concerned species. For each species the mean individual weight was calculated as the total biomass divided by the total abundance, using all the stations where the two informations were simultaneously recorded.

(NB: The consequence of the two last points on the results will be taken into account in the discussion).

Table I. List of the sampling surveys and numbers of fishing trawls (stations)

Code	Date	Number of stations
1	November 1986	92
2	May 1987	95
3	October 1986	97
4	March-April 1988	97
5	April 1989	100
6	November 1989	101
7	March 1990	99
8	March-April 1991	93
9	April-May 1992	98
10	October 1992	102
11	April-May 1993	98
12	March 1994	96
13	May 1995	95

Table II. List of the species

1	*Acanthurus monroviae*	48	*Merluccius* spp.
2	*Alectis alexandrinus*	49	*Mustelus mustelus*
3	*Aluterus punctatus*	50	*Mycteroperca rubra*
4	*Antigonia capros*	51	*Octopus vulgaris*
5	*Argyrosoma regium*	52	*Pagellus bellottii*
6	*Arius heudeloti*	53	*Parapristipoma octolineatum*
7	*Arius latiscutatus*	54	*Penaeus notialis*
8	*Arnoglossus capensis*	55	*Plectorhynchus mediterraneus*
9	*Balistes capriscus*	56	*Pomadasys incisus*
10	*Balistes punctatus*	57	*Pomadasys jubelini+peroteti*
11	*Batrachoides* sp.	58	*Pontinus kuhlii*
12	*Bodianus* spp.	59	*Priacanthus arenatus*
13	*Boops boops*	60	*Pseudotolithus senegalensis*
14	*Bothus podas*	61	*Pseudupenaeus prayensis*
15	*Brachydeuterus auritus*	62	*Pteroscion peli*
16	*Branchiostegus semifasciatus*	63	*Pterothrissus belloci*
17	*Brotula barbata*	64	*Raja miraletus*
18	*Calappa* spp.	65	*Raja straeleni*
19	*Chaetodon hoefleri*	66	*Rhinobatos rhinobatos*
20	*Chelidonichthys gabonensis*	67	*Sardinella aurita*
21	*Chilomycterus antennatus*	68	*Sardinella maderensis*
22	*Chloroscombrus chrysurus*	69	*Scomber japonicus*
23	*Citharus macrolepidotus*	70	*Scorpaena angolensis*
24	*Cybium tritor*	71	*Scorpaena* spp.
25	*Cymbium* spp.	72	*Scorpaena stephanica*
26	*Cynoglossus canariensis*	73	*Scyacium micrurum*
27	*Cynoglossus senegalensis*	74	*Selene dorsalis*
28	*Cynoponticus ferox*	75	*Sepia* spp.
29	*Dactylopterus volitans*	76	*Soleidae*
30	*Dasyatis margarita*	77	*Sparus caeruleostictus*
31	*Decapterus* spp.	78	*Sphoeroides* spp.
32	*Dentex angolensis*	79	*Sphyraena guachancho*
33	*Dentex canariensis*	80	*Squids*
34	*Dentex gibbosus*	81	*Stromateus fiatola*
35	*Dentex macrophthalmus*	82	*Synaptura* spp.
36	*Drepane africana*	83	*Torpedo torpedo*
37	*Epinephelus aeneus*	84	*Ttrachinocephalus myops*
38	*Epinephelus goreensis*	85	*Trachinus armatus*
39	*Eucinostomus melanopterus*	86	*Trachurus trecae*
40	*Fistularia petimba*	87	*Trichiurus lepturus*
41	*Fistularia tabaccaria*	88	*Trigla* spp.
42	*Galeoides decadactylus*	89	*Umbrina canariensis*
43	*Grammoplites gruveli*	90	*Uranoscopus cadenati*
44	*Ilisha africana*	91	*Uranoscopus polli*
45	*Lagocephalus laevigatus*	92	*Zanobatus schoenleinii*
46	*Lepidotrigla cadmani*	93	*Zeus faber*
47	*Lepidotrigla carolae*		

Data analysis

The community structure was investigated using Abundance Biomass Comparison (ABC) plots (Clarke and Warwick, 1994) which is an extention of the k-dominance curves theory (Lambshead *et al.*, 1983). The method allows to draw up a diagnosis of the diversity status of the studied community, with a gradual pattern from unstressed to grossly stressed. The analysis is computed using Primer software (Clarke and Warwick, 1994).

Results

The fishing effort is high in the studied area since several decades (Figs 1-2).
According to theory, the diagnosis for the studied community (Fig. 3) is 'stressed' or 'moderately stressed'. This diagnosis remains the same for each of the 13 sampling surveys, suggesting no significant evolution in the diversity status of the Senegalese demersal fish community along the studied period. This result is to be related to the lack of a significant trend in the global cpue (Fig. 4) and also to the lack of drastic change in the species dominances (Fig. 5) between 1986 and 1995.

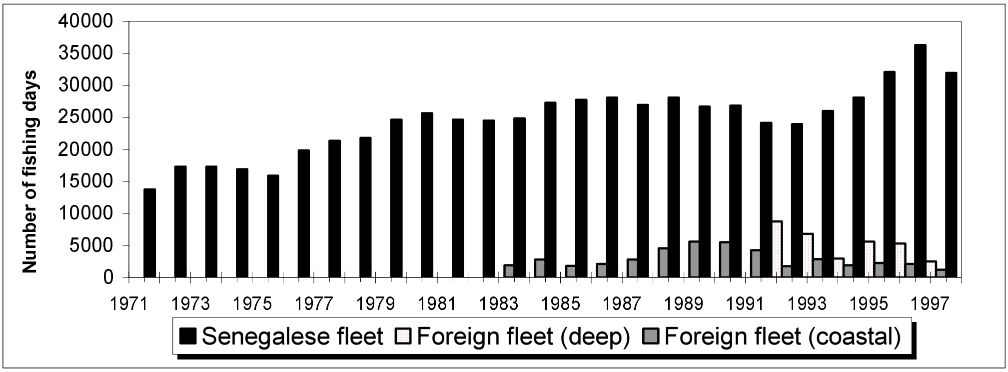


Fig. 1. Fishing effort of the industrial fishery during the last decades (from 1971 to 1998).

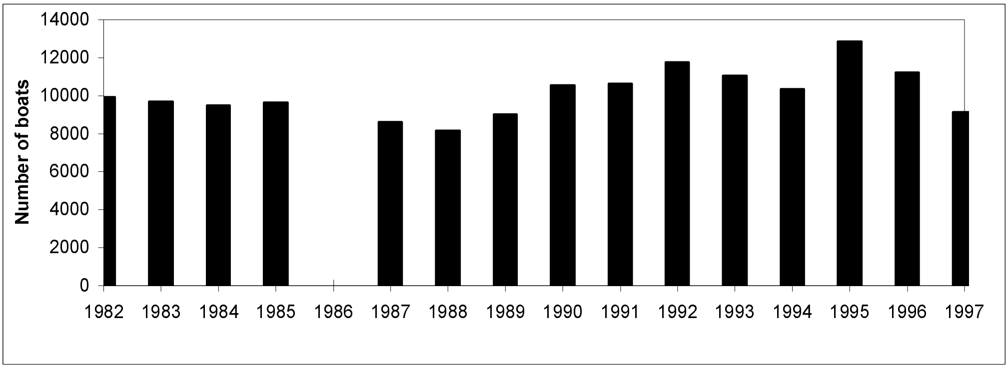


Fig. 2 : Fishing effort of the small scale fishery during the last decades (from 1982 to 1997).

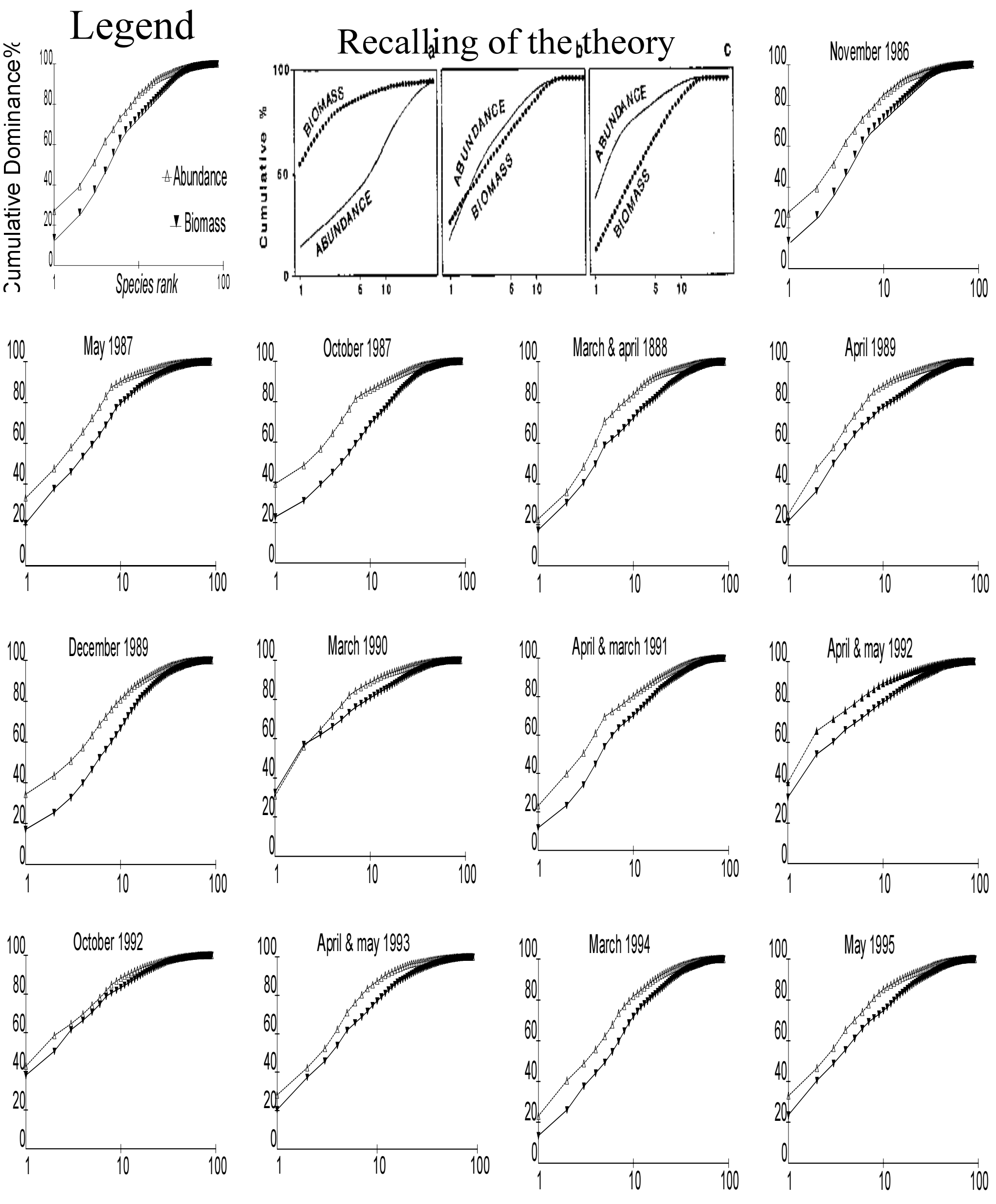


Fig. 3. ABC plot computed for the 13 surveys from 1986 to 1995.

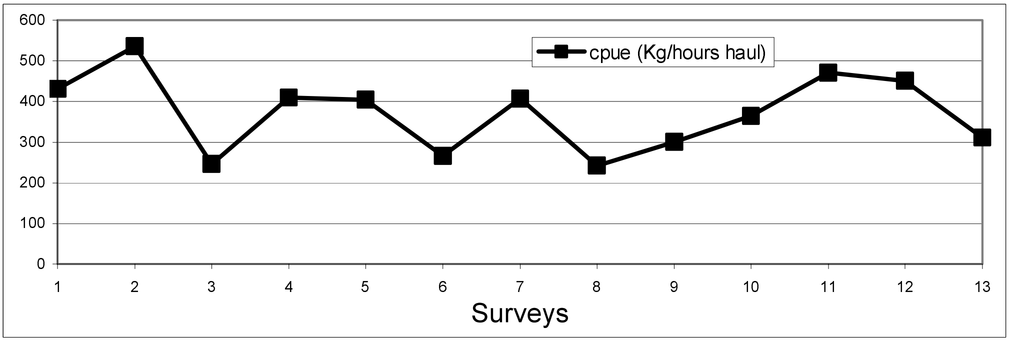


Fig. 4. Global abundance index (kg/hour haul).

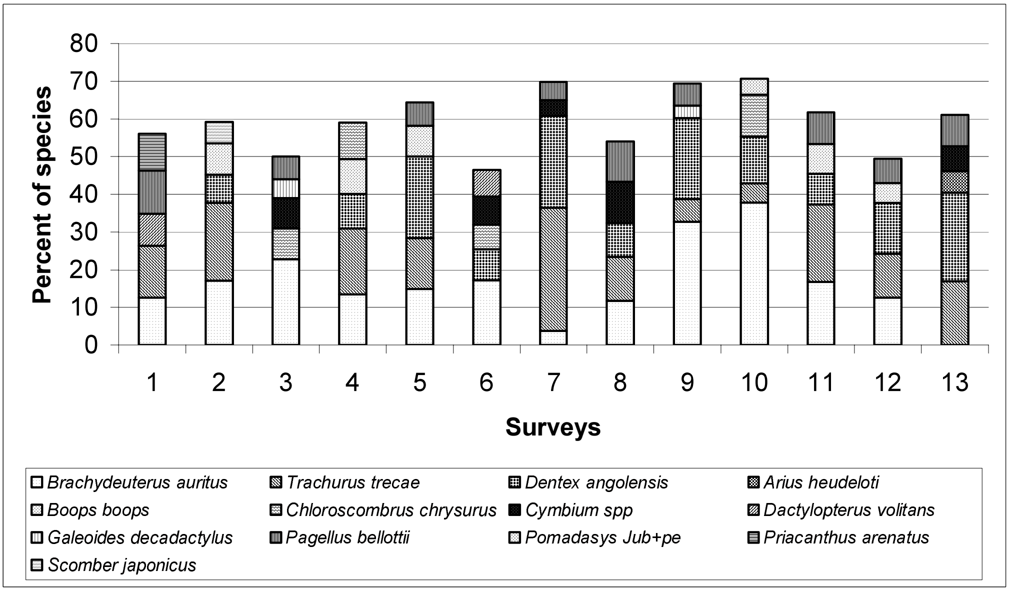


Fig. 5. Proportions of the dominant species.

Discussion

Concerning the adequacy of the data and the methodology two aspects must be pointed out:

- The first one concerns the lack of specific determination of the fauna. The theory of the method states that the determination level must be homogeneous for all the taxa. This is not the case here, some taxa representing a group of species. Fortunately this problem concerns only relatively rare taxa, that do not occur in the first rank of abundance or biomass. So, we can consider that these taxa have a minor influence on the shape of the curves and finally that they do not interfere on the related diagnosis. Considering this, there is an alternative with eliminating them from the analysis. We choose to keep them in, in order to fit the analysis to the biological information ‘as it is recorded’, without eliminating any taxa that could represent an halieutic interest.

- The second point is related to the estimation of abundance, when they have not been recorded (see methodology). This problem concerns only a few samples in the beginning of the sampling series but a greater proportion for the last surveys. Considering this problem, we choose to keep the related samples in the analysis, in order to compute the diagnosis on the most complete dataset possible. Doing this we consider the fact that the relative position of the biomass and the abundance curves (and consequently the related diagnosis) are more influenced by changes of the species occurring in the first rank than by changes of individual mean weight within these species.

The results can be summarized in the two folowing points: (1) the general diagnosis is ‘stressed’ (moderately or not) and (2) there is no significant evolution of this diagnosis (no apparent increase in this observed stress). We can notice that both of them are consistent with what was known or expected. Indeed:

- Concerning the first aspect, the result is not really surprising considering the fact that the community has been submitted to a strong fishing pressure for such a long time (see Figs 1 and 2). It is also coherent with many halieutic evaluation studies and models, even if they are generally based on monospecific studies (Caverivière and Thiam, 1992, 1994, 1996; Caverivière, 1994; Gascuel et Thiam, 1994).

- Concerning the second aspect, it is consistent with previous results on the same faunistic assemblages (Jouffre *et al.*, 1999) which suggests that the multipecies level of the biological structures doesn’t show the same sensitiveness to fishing impacts as the monospecific one. A study at the regional scale from Mauritania to Guinea (Jouffre *et al.*, submitted) has also suggested that it is essentially during the launching phase of the fisheries that the fishing impact on the composition of the demersal assemblage is the most perceptible.

Going further, the present results, compared to the general and theorical frame of the method, suggest that the targeted community is ‘moderately stressed’ (rather than ‘grossly stressed’). However, the question of the calibration of the method to refine the diagnosis based on this kind of data and for this objective remains largely open. Moreover, the data do not cover the initial state of the community (*i.e.* virgin state before fishing). That remains an essential question in order to reach a really operational diagnosis (in other words, to distinguish moderately stressed communities from strongly stressed ones).

Conclusion

On the practical level, the method of k-dominance curves/ABC plots seems applicable for a general assesment of multispecific halieutic resources in tropical areas, since one obtains a diagnosis which is coherent with expectations.

This method seems to be relatively robust, e.g. by tolerating deviation from theory, namely by tolerating to some extent taxa with levels of different determination provided they do not belong to the first ranks of abundance. However, for truly calibrating and validating the method considering this kind of data and objective (directed towards the halieutic impact), it would be necessary to apply it to a series of older surveys from Senegal, making it possible to include the launching phase of the fisheries. Applying this to similar regional data (such as available in Guinea and Mauritania) should also be interesting according to this objective.

Lastly, considering the ecological plan, the stability observed over the period (1986-1995) could be interpreted as an adaptation of the studied community to a relatively strong fishing pressure.

Acknowledgements

The authors wish to acknowledge all the contributors and partners of the SIAP project with special reference to CRODT and participants of the sampling surveys.

References

Ba M. 1999. Fisheries Information and Analysis System (FIAS) Project. ACP-EU Initiative.Bulletin 12(2-3 ; juin-septembre):28-30.

Blanchard F. 2000. Une approche de la dynamique de poissons démersaux exploités: analyse comparée de la diversité spécifique dans le golfe de Gascogne (océan Atlantique) et dans le golfe du Lion (mer Méditerranée). Aquatic Living Resources 14:29-40.

Caverivière A. 1994. Comparaison sur une période de 20 ans (1972-1992) des indices d'abondance obtenus sur le plateau continental sénégalais à partir des campagnes de chalutage de fond. p.163-177. In: L'évaluation des ressources exploitables par la pêche artisanale sénégalaise, tome 2. Barry-Gérard M., T. Diouf and A. Fonteneau (Eds). Orstom éditions, coll. colloques et séminaires, Paris. 424p.

Caverivière A. and M. Thiam. 1992. Indices d'abondance et niveaux d'exploitation des espèces démersales du plateau continental sénégalais. Estimations à partir des résultats des campagnes de chalutage stratifié (1986-1991). Centre Recherches Océanographiques Dakar-Thiaroye, Doc. Sci. 132:147p.

Caverivière A. and M. Thiam. 1994. Tailles moyennes et taux d'exploitation approchés des principales espèces de poissons débarquées par les chalutiers glaciers entre 1973 et 1989. p.365-384. In: L'évaluation des ressources exploitables par la pêche artisanale sénégalaise, tome 2. Barry-Gérard M., T. Diouf and A. Fonteneau (Eds). Orstom éditions, coll. colloques et séminaires, Paris. 424p.

Caverivière A. and M. Thiam. 1996. Évolution sur 20 ans (1969-1992) des tailles et des mortalités des principales espèces démersales du Sénégal (campagnes scientifiques de chalutage). Centre Recherches Océanographiques Dakar-Thiaroye, Doc. Sci. 144:17p.

Chauveau J.P. 1985. Histoire de la pêche maritime et politique de développement de la pêche au Sénégal. Représentation et pratique du dispositif de l'intervention moderniste. Anthropologie maritime 2:300-318.

Chavance P. 2002. Pêche en Afrique de l'Ouest. L'écosystème décapité. Science au Sud, le journal de l'IRD, 16, septembre/octobre 2002: p.16.

Clarke K.R. and R.M. Warwick. 1994. Change in marine communities: an approach to statistical analysis and interpretation. Plymouth Marine laboratory, Plymouth. 144p.

Gascuel D. and J. Guitton. 2001. Avancement du module trawlbase. Rapport de l'atelier de Mindelo le 09/10/2001. SIAP/FIAS Système d'information et d'analyse des pêches, Pr. 7 ACP RPR. 730, octobre 2001, document technique 4: 17p.

Gascuel D. and F. Ménard. 1997. Assessment of a multispecies fishery in Senegal, using production models and diversity indices. Aquatic Living Resources 10:281-288.

Gascuel D. and M. Thiam. 1994. Evolution de l'abondance des espèces démersales sénégalaises: estimation par modélisation linéaire des PUE. In: L'évaluation des ressources exploitables par la pêche artisanale sénégalaise, tome 2. p.191-213. Barry-Gérard M., T. Diouf and A. Fonteneau (Eds). Orstom éditions, coll. colloques et séminaires, Paris. 424p.

ICES 2000. Ecosystems effects of fishing. Proceedings of an ICES/SCOR Symposium held in Montpellier France, 16-19 March 1999. ICES Journal Marine Sciences 57:465-792.

Jouffre D., M. Thiam and A. Caverivière. 1999. Structural changes in the demersal fauna of Senegalese waters during the last decade: ecosystem effects of fishing? P.38. In: Book of abstracts of ICES/SCOR Symposium: Ecosystem effects of fishing, Montpellier 15-19 March 1999: 99p.

Jouffre D., G. Domalain, D. Thiam, S. Traoré, A. Caverivière, F. Domain and C. Inejih. (submitted). Schémas généraux de répartition et d'évolution des communautés de poissons démersaux d'Afrique de l'Ouest perçus par l'analyse des données de campagnes de chalutages scientifiques réalisées en Mauritanie Sénégal et Guinée entre 1987 et 1999. Actes du Symposium international: Pêcheries maritimes, écosystèmes et sociétés en Afrique de l'Ouest: un demi-siècle de changement. Dakar 22-24 juin 2002.

Lambshead P.J.D., H.M. Platt and K.M. Shaw. 1983. The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. Journal Natural History 17:859-874.

La pêche maritime. 1973. Un pays qui vit de la mer: le Sénégal. La pêche maritime 145:645-647.

Pauly D., V. Christensen, J. Dalsgaard, R. Froese and J.F. Torres. 1998. Fishing down marine food webs. Science 279:860-863.

Pauly D., V. Christensen and C. Walters. 2000. Ecopath, ecosim and ecospace as tools for evaluating ecosystems impact of fisheries. ICES Journal Marine Sciences 57:697-706.

SIAP. 2002. Symposium international: pêcheries maritimes, écosystèmes et sociétés en Afrique de l'Ouest: un demi-siècle de changement. Dakar 22-24 juin 2002. Conclusions adoptées par les participants au Symposium à l'issue de leurs travaux: http/www.ird.sn/activités/symposium.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 115-124
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Informational resources on aquatic alien species in Europe on the internet: present developments and future perspectives

Vadim Panov[1] and Stephan Gollasch[2]

[1]Zoological Institute, Russian Academy of Sciences
199034 St.Petersburg, Russia
E-mail: gaas@zin.ru

[2]GoConsult
Bahrenfelder Str. 73a, D-22765 Hamburg, Germany

Abstract

During the last two decades rates of invasions of European inland and coastal waters by alien species increased significantly. The development of an internet-based informational resources database on aquatic invasive species is considered as one of the most important mechanisms of information exchange within the scientific community. These resources may provide comprehensive information for management of aquatic invasive species in Europe, as well as for scientific and educational purposes. At present, the online informational resources on aquatic alien species from Europe are located in several online national, regional and global databases and information systems. However, at present the available online information is generally not sufficient for management purposes, such as prevention of new introductions, or control and eradication of the established invasive alien species. Future development of the internet resources on aquatic alien species in Europe should consider the integration in the developing global network of online interoperable databases and information systems. Priority should be given to the development of the interlinked regional information hubs, which should provide comprehensive information on aquatic alien species, including regional alien species directories with species-specific entries, and online access to the datasets of geo-referenced monitoring data. The development of species-specific entries for regional directories can be conducted by the European experts, including participants of the European Research Network of the Aquatic Invasive Species (ERNAIS).

Keywords: Invasive alien species; Aquatic ecosystems; Databases; Information systems; Internet resources.

Introduction

During the last decades the invasion rates of alien species[1] into European inland and coastal waters increased significantly. A number of alien species arrived in Europe from other parts of the world with ship ballast water and as hull fouling. The breaking of natural barriers between water basins by the construction of canals in the 19th and 20th centuries resulted in the spread of hundreds of alien species. Headlong intentional introductions also contributed to the number of invaders, including invasive and pathogen organisms. Currently several hundreds of species are considered as alien in the Mediterranean Sea (Galil, 2000; Galil and Zenetos, 2002). It is believed that about 100 aquatic invaders occur in the Baltic Sea, with 56 species first recorded during 1950-2002 (Baltic Alien Species Database, 2001; Leppäkoski *et al.*, 2002a). In the Black Sea, 59 species are considered as alien, including 49 species introduced during the last 50 years (Zaitsev and Öztürk, 2001). In many cases introductions of invasive species have caused significant losses in marine, estuarine and inland water biodiversity and economy in Europe (Leppäkoski *et al.*, 2002b).

The development of open databases and information systems on invasive alien species is essential for the effective international cooperation in data and expertise sharing, and provides support for management and control efforts. Internet-based information systems may serve as the main tool for the wide dissemination of information on taxonomy, biology, environmental impacts and possible control measures of invasive species. International law requires governments and other relevant organizations to support the creation and maintenance of the databases, information systems and interoperable distributed network of databases for the compilation and dissemination of information on alien species for the use in the context of any prevention, introduction, monitoring and mitigation activities. The 6th Conference of the Parties of the Convention on Biological Diversity agreed that a Clearing-house mechanism will be useful to facilitate scientific and technical cooperation on invasive species issues, and notified the Global Invasive Species Programme (GISP, 1999) as an international thematic focal point for alien species under the clearing-house mechanism, and calls on Parties, countries and relevant organizations to contribute to the creation and maintenance of the global information network, in particular to ensure effective international cooperation and expertise sharing (Decision VI/23, 2002).

Currently relevant international organizations and working groups are actively working on the development of open informational resources on aquatic alien species in Europe (GloBallast, 2001; Gollasch, 2002; Olenin *et al.*, 2002; Panov *et al.*, 2002). These resources may provide comprehensive information for the management of invasive alien species in European inland and coastal waters, as well as for scientific and educational purposes, and form a ‘European hub’ of the global invasive species informational network. The goal of the present paper is to overview the existing online informational resources on aquatic alien species in Europe, and to discuss perspectives of their further development.

---

[1] The following definitions are used: (i) ‘alien species’ refers to a species, subspecies or lower taxon, introduced outside its natural past or present distribution; includes any part, gametes, seeds, eggs, or propagules of such species that might survive and subsequently reproduce; (ii) ‘invasive alien species’ means an alien species whose introduction and/or spread threatens biological diversity (iii) ‘introduction’ refers to the movement by human agency, indirect or direct, of an alien species outside of its natural range (past or present) (Decision VI/23, 2002).

Global Informational Resources

ISSG Global Invasive Species Database

The Global Invasive Species Database was developed by the IUCN/SSC Invasive Species Specialist Group (ISSG) as part of the Global Invasive Species Programme (GISP, 1999). It provides global information on invasive alien species to agencies, resource managers, decision-makers, and interested individuals. The database focuses on invasive species that threaten biodiversity and covers all taxonomic groups from micro-organisms to animals and plants. Species information is supplied by expert contributors from around the world and includes: species' biology, ecology, native and alien range, references, contacts, links and images. Currently it includes searchable entries for 100 of 'World's Worst Invasive Alien Species', which consider both aquatic invasive species of European origin (fishhook waterflea *Cercopagis pengoi*, zebra mussel *Dreissena polymorpha*, green crab *Carcinus maenas*, common carp *Cyprinus carpio*, brown trout *Salmo trutta*), and harmful aquatic species, intentionally or unintentionally introduced into European inland and coastal waters from other regions (Pacific sea weed *Caulerpa taxifolia*, Japanese kelp *Undaria pinnatifida*, Atlantic ctenophore *Mnemiopsis leidyi*, Chinese mitten crab *Eriocheir sinensis*, Western mosquitofish *Gambusia affinis*, large-mouth bass *Micropterus salmoides*, common tilapia *Oreochromis mossambicus*). These informational resources are linked to the Baltic Sea Alien Species Database, Caspian Sea Biodiversity Database and the Regional Biological Invasions Center Information System (see the following chapter) (ISSG Global Invasive Species Database).

The FAO Database on Introductions of Aquatic Species (DIAS)

The FAO database on introductions of aquatic species currently contains about 3,150 records of introductions of freshwater and marine fishes, and other taxa, and can be queried through the Search Form. The database includes records of species introduced or transferred from one country to another. Coverage of accidental introductions of organisms (e.g. through ship ballast waters) is not complete and records on this topic have been generally entered only when important impacts on fisheries or on the environment have been caused (introduction of *Mnemiopsis leidyi* to the Black Sea, for instance).

Global Information System on Fishes (FishBase)

Global Information System on Fishes (FishBase) is one of the most comprehensive online informational resources on freshwater and marine fishes. FishBase was developed at the International Center for Living Aquatic Resources Management (ICLARM) in collaboration with the Food and Agriculture Organization of the United Nations (FAO) and with the support from the European Commission. FishBase is a relational database, which contains practically all fish species known to science (more than 25,000 species), including information on invasive European species (round goby *Neogobius melanostomus*, ruffe *Gymnocephalus cernuus* etc.), and fish species, alien for the European waters (Amur sleeper *Perccottus glenii*, stone morone *Pseudorasbora parva* and other). Species-specific entries include comprehensive information on fish species distribution, morphology, biology, references and links to other relevant internet sources of information (Froese and Pauly, 2002). However, detailed information on the distribution of invasive fish species in European waters, their invasion history and environmental impacts is not available in FishBase.

Global Ballast Water Management Programme (GloBallast) Information Resource

Comprehensive information of ballast water management is available at the Global Ballast Water Management Programme (GloBallast) website, which represents the global clearinghouse on ballast water management. The GEF/UNDP/IMO Global Ballast Water Management Programme (GloBallast) is assisting developing countries to reduce the transfer of harmful aquatic organisms and pathogens in ship ballast water, implement the IMO ballast water guidelines and prepare for the new IMO ballast water convention. The GloBallast website includes detailed information and links to other open sources of information on legislation, available ballast water treatment technologies, and provides reviews and links to the main online national, regional and national databases and information systems on aquatic alien species (GloBallast, 2002).

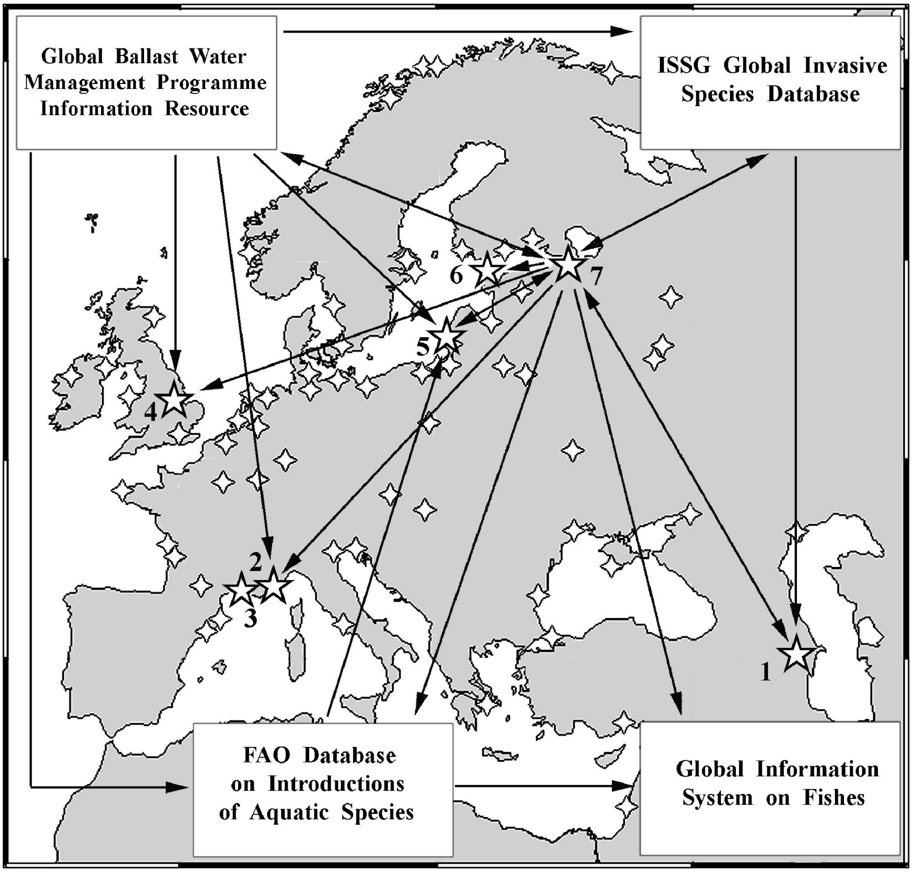


Fig. 1. Location of the main national and regional internet resources on aquatic alien species in Europe and links between them and with global online resources (1- Caspian Sea Biodiversity Database, 2- CIESM Atlas of Exotic Species, 3-Caulerpa taxifolia Database, 4- Directory of Non-native Marine Species in British Waters, 5- Baltic Sea Alien Species Database, 6- Marine Alien Species of Estonia Website, 7- Regional Biological Invasions Center Information System; small asterisks indicate ERNAIS sites).

National and Regional Informational Resources

Internet resources on aquatic alien species in the United Kingdom

Directory of Non-native Marine Species in British Waters

The Directory of Non-native Marine Species in British Waters, hosted on the website of the Joint Nature Conservation Committee (JNCC), includes entries on 53 species, alien for British coastal waters. The species-specific entries include data on species taxonomy, dates of introduction, origin, method of introduction, rates of spread, brief description of distribution, effects on the environment and control methods. The list of alien species includes such invasive organisms as the Japanese brown alga *Sargassum muticum*, the East Asian parasitic nematode *Anguillicola crassus*, the Atlantic polychaete *Marenzelleria viridis*, the New Zealand mud snail *Potamopyrgus antipodarum*, the Japanese oyster *Crassostrea gigas*, the North Atlantic clam *Mya arenaria*, the West Atlantic crab *Rhythropanopeus harrisii* and the Chinese mitten crab *Eriocheir sinensis* (Directory of Non-native Marine Species in British Waters).

The Chinese Mitten Crab homepage

The Chinese Mitten Crab homepage, hosted by the Natural History Museum, represents a useful online information source on the Chinese mitten crab *Eriocheir sinensis*. It includes detailed information on the taxonomy of this invasive species, the description of its life history with pictures of its life cycle, its distribution in North America, Europe and Great Britain (illustrated by maps), information on environmental problems, associated with this species introduction (The Natural History Museum, 2002).

Biological Records Centre (BRC) website

The Biological Records Centre (BRC) website, hosted by the Center for Ecology and Hydrology (CEH), is the national custodian of data on the distribution of wildlife in the British Isles, and provides online access to the data on the distribution of species via a sophisticated information system (The National Biodiversity Network Gateway). This system gives access to the National Biodiversity Network species dictionary and 10-km distribution maps. The pilot version includes data from the BRC database (including data on native and introduced crayfish) and several other data sets (including data on some marine invasive species) (The Biological Records Centre).

Marine Alien Species of Estonia website

The Marine Alien Species of Estonia website, hosted by the Estonian Marine Institute in Tallinn, includes standardized entries on 12 alien species, established in Estonian coastal waters, with information on species identification, natural range, invasion history in the Baltic Sea, distribution and population dynamics, ecological and economic impact, list of references. Alien species divided in three main subgroups by lifestyle: plankton (three species, including invasive

Ponto-Caspian fishhook waterflea *Cercopagis pengoi*), benthic invertebrates (six species, including invasive Ponto-Caspian zebra mussel *Dreissena polymorpha*, New Zealand mud snail *Potamopyrgus antipodarum*, Chinese mitten crab *Eriocheir sinensis*, Atlantic polychaete *Marenzelleria viridis*), and fishes (four species, including the invasive Ponto-Caspian round goby *Neogobius melanostomus*) (Marine Alien Species of Estonia).

The Caulerpa taxifolia *website*

The *Caulerpa taxifolia* website (information system) is located at the University of Nice-Sophia Antipolis (France), and provides distributional maps, information on environmental impacts and images of this extremely invasive Pacific macroalgae ('killer alga'), which possesses a very high potential to spread and replace native species.

CIESM Atlas of Exotic Species

The International Commission for the Scientific Exploration of the Mediterranean Sea (CIESM) Atlas of Exotic Species includes detailed description of 125 alien species of mollusks, 55 species of alien crustaceans and 91 alien fish species that invaded the Mediterranean Sea. The list of alien species includes several highly invasive species (the Pacific gastropod *Rapana venosa*, the West Atlantic crabs *Rhythropanopeus harrisii* and *Callinectes sapidus*, the Chinese mitten crab *Eriocheir sinensis*). Individual species entries include species images, their diagnostic features, biological information, references and distribution maps (CIESM, 2001).

Caspian Sea Biodiversity Database

The demonstration version of the Caspian Sea Biodiversity Database (CSBD) has been developed during 2001-2002 in the frameworks of the UNDP Caspian Environment Programme (CEP), and posted on the CEP website in June 2002. The CSBD consists of English and Russian versions, and currently includes entries on 36 aquatic species, both native (30 species) and alien (6 species) for the Caspian Sea ecosystem (Caspian Sea Biodiversity Database, 2002). Alien species in the CSBD are represented by three unintentionally introduced zooplankton organisms: the Mediterranean diatom alga *Rhizosolenia calcar-avis*, the Atlantic copepod *Acartia tonsa*, the Atlantic ctenophore *Mnemiopsis leidyi*, and three intentionally introduced organisms: the Atlantic-Mediterranean clam *Abra ovata*, and two fish species, *Liza aurata* (Atlantic-Mediterranean), and *Liza saliens* (Mediterranean). The CSBD also includes entries on two native species, considered as highly invasive outside the Ponto-Caspian Region, the cladoceran *Cercopagis pengoi* (fishhook waterflea) and fish *Neogobius melanostomus* (round goby). Entries in the database include information on species taxonomy, their distribution and biology, references and illustration of organisms. Entries on *Cercopagis* and *Mnemiopsis* include internet-links to entries on these species in the Regional Biological Invasions Center Information System (see below).

Baltic Sea Alien Species Database

An internet database on aquatic alien species in the Baltic Sea area was developed as an initiative of the Baltic Marine Biologists´ Working Group on Non-indigenous Estuarine and Marine Organisms in 1997; in 2000 a new concept of the online database appeared with support received from the Baltic Marine Environment Protection Commission (HELCOM). At present the database represents an interactive user-friendly tool, which includes several information retrieving options. The database species directory contains individual species entries. An entry includes the complete taxonomy of a species and available comments, complementing and specifying the Database features (year of introduction, ecological impact, etc.). The database search tool is a direct way to retrieve information according to the major features. It allows the retrieval of data by a single feature (i.e. by 'Taxon') or by combined features (i.e. 'Taxon' and 'Origin' and 'Ecological impact'), including multiple selections of items within any feature. A list of species, retrieved according to the selected criteria, is linked to relevant individual entries on species and references (Baltic Sea Alien Species Database, 2002; Olenin *et al.*, 2002), which include individual entries on such invasive Ponto-Caspian species as fishhook waterflea *Cercopagis pengoi* and the round goby *Neogobius melanostomus*. Some species-specific entries are hosted by the Regional Biological Invasions Center Information System (see below), the entry for the zebra mussel, *Dreissena polymorpha*, for instance.

Regional Biological Invasions Center Information System

The Regional Biological Invasions Center Information System (RBIC), hosted by the Zoological Institute of the Russian Academy of Sciences in St. Petersburg, is a new concept of the Group on Aquatic Alien Species (GAAS) website, which initially opened in 1999 (Panov, 1999). Currently RBIC is serving as the regional clearinghouse on invasive alien species (both aquatic and terrestrial), and as a web portal, providing access to the internet-based information resources on invasive species research and management in Europe and worldwide (Fig. 1) (Regional Biological Invasions Center, 2001a). The development of the Geographic Information System 'INVADER' as an international database on the internet is one of the RBIC priorities. Currently a demonstration version of GIS 'INVADER', with comprehensive geo-referenced information on the distribution of the Ponto-Caspian invasive cladoceran *Cercopagis pengoi* in Europe and North America, is available online (Regional Biological Invasions Center, 2001b).
Online geo-referenced distribution maps of selected aquatic invasive species, including the Atlantic ctenophore *Mnemiopsis leidyi*, the Ponto-Caspian mussel *Dreissena polymorpha* and the cladoceran *Cercopagis pengoi*, along with detailed descriptions of their taxonomy, invasion histories, biology, environmental impacts are available at the RBIC Illustrated Database of the Aquatic Invasive Species of Europe, interlinked with the Baltic Sea Alien Species Database, the Global Invasive Species Database and the Caspian Sea Biodiversity Database (Regional Biological Invasions Center, 2001c). Entry on *Mnemiopsis leidyi* provides an example of a comprehensive and user-friendly online information system on the aquatic invasive species, linked to other internet-based sources of information (Shiganova and Panov, 2002). Entry on *Mnemiopsis* in the RBIC Illustrated Database is already serving as open information system on *Mnemiopsis* for the Ponto-Caspian Region, updated on a regular basis.

The RBIC portal is also supporting web pages of some international working groups and networks, including the developing European Research Network on Aquatic Invasive Species, the Caspian Environment Program Regional Invasive Species Advisory Group and SIL Working

Group on Aquatic Invasive Species (Regional Biological Invasions Center, 2001d), and serving as a regional information hub for the developing global invasive species informational network (for more information on the global network see the online Report of the Joint Convention on Biological Diversity/Global Invasive Species Programme Informal Meeting on Formats, Protocols and Standards for Improved Exchange of Biodiversity-related Information 2002, and Report of the Workshop on Development of Nordic/Baltic Invasive Species Information Network 2002). RBIC is interlinked with the United States Geological Survey Nonindigenous Species Information Resource, which has been established as a central repository for accurate and spatially referenced biogeographic accounts of alien (nonindigenous) aquatic species, and represents American information hub of the global invasive species informational network.

Recommendations

At present, the online informational resources on aquatic alien species in Europe are located in several online national, regional and global databases and information systems, briefly described above. These informational resources are already linked within the World Wide Web, with the Regional Biological Invasions Center Information System (RBIC) serving as a regional web portal (Fig. 1). However, at present available online information is not sufficient for management purposes, such as prevention of new introductions, or control and eradication of the established invasive alien species. Probably only in the case of the invasive ctenophore *Mnemiopsis leidyi*, available information in the interlinked Global Invasive Species Database, Caspian Sea Biodiversity Database and RBIC is sufficient enough for the elaboration of adequate control measures in the Ponto-Caspian Region, as well as for undertaking preventive management options in the potential regions-recipients of this harmful species (in the Baltic Sea Region, for instance).

Future development of the internet resources on aquatic alien species in Europe should consider the integration in the developing global network of online interoperable databases and information systems (see Introduction). Priority should be given to the development of the interlinked regional information hubs, like those serving the Nordic/Baltic Region (interlinked Baltic Sea Alien Species Database and RBIC) and Mediterranean Sea (CIESM Atlas of Exotic Species). The creation of other regional information hubs is needed, specifically for the Ponto-Caspian Region. Its biodiversity and fisheries are threatened by alien species introductions. They represent an important donor area of aquatic invasive species for European inland and coastal waters and worldwide. These regional information hubs should provide comprehensive information on aquatic alien species, including regional alien species directories with species-specific entries, and online access to the datasets of geo-referenced monitoring data. Timely access to such open geo-referenced data, using already available internet and GIS technologies (like demonstrated at the RBIC and Biological Records Centre websites) may ensure service of the regional information hubs as effective management tools (early warning systems, for instance). The development of species-specific entries for regional directories can be conducted by the relevant European experts, whose contact information and area of expertise is available in the online database of the European Research Network of the Aquatic Invasive Species (ERNAIS) (Regional Biological Invasions Center, 2001). The timely incorporation of geo-referenced data from aquatic alien species monitoring into open national and regional databases is a challenging goal and the elaboration of effective mechanisms of achievement of this goal should be considered as one of the future priorities in the process of the development of online informational resources on aquatic alien species in Europe.

References

Baltic Sea Alien Species Database. 2002. Olenin S, E. Leppäkoski and D. Daunys (Eds). Alien Species Directory. http://www.ku.lt/nemo/mainnemo.htm

Caspian Environment Programme. 2002. *Mnemiopsis leidyi* in the Caspian Sea. http://caspianenvironment.org/mnemiopsis

Caspian Sea Biodiversity Database. 2002. Belyaeva V.N. (Ed.) http://www.caspianenvironment.org/biodb/eng/main.htm

Caulerpa taxifolia Database. http://www.caulerpa.org

CIESM Atlas of Exotic Species in the Mediterranean Sea. 2001. http://www.ciesm.org/atlas

Decision VI/23 COP6 of the Convention on Biological Diversity. 2002. http://www.biodiv.org/decisions

Directory of Non-native Marine Species in British Waters. Eno NC, Clark RA & Sanderson WG (Eds). http://www.jncc.gov.uk/marine/dns

FAO Database on Introductions of Aquatic Species (DIAS). 1998. http://www.fao.org/waicent/faoinfo/fishery/statist/fisoft/dias/index.htm

Froese R. and D. Pauly (Eds). 2002. FishBase. World Wide Web electronic publication. www.fishbase.org, 09 December 2002

Galil B.S. 2000. A sea under siege – alien species in the Mediterranean. Biological Invasions 2: 177-186.

Galil B.S. and A. Zenetos. 2002. A sea change – exotics in the eastern Mediterranean Sea. p.325-336. In: Leppäkoski E., S. Gollasch and S. Olenin (Eds). Invasive Aquatic Species of Europe. Distribution, Impacts and Management. Kluwer Academic Publishers, Dordrecht.

GISP. 1999. Global Invasive Species Programme. http://jasper.stanford.edu/gisp

GloBallast. 2002. Global Ballast Water Management Programme. http://globallast.imo.org

Gollasch S. 2002. Databases on aquatic alien species: North and Mediterranean Seas and non-European Initiatives. p.520-524. In: Leppäkoski E, S. Gollasch and S. Olenin (Eds). Invasive Aquatic Species of Europe. Distribution, Impacts and Management. Kluwer Academic Publishers, Dordrecht.

ISSG Global Invasive Species Database. http://www.issg.org/database

Leppäkoski E, S. Gollasch, P. Gruszka, H. Ojaveer, S. Olenin and V. Panov. 2002a. The Baltic – a sea of invaders. Can. J. Fish. Aquat. Sci. 59:1175-1188

Leppäkoski E, S. Gollasch and S. Olenin (Eds). 2002b. Invasive Aquatic Species of Europe. Distribution, Impacts and Management. Kluwer Academic Publishers, Dordrecht. 583 p.

Marine Alien Species of Estonia. http://www.sea.ee/Sektorid/merebioloogia/eesti/Marine_Alien_Species_of_Estonia.htm

Olenin S., E. Leppäkoski and D. Daunys. 2002. Internet database on aliens species in the Baltic Sea. p.525-528. In: Leppäkoski E, S. Gollasch and S. Olenin (Eds). 2002. Invasive Aquatic Species of Europe. Distribution, Impacts and Management. Kluwer Academic Publishers, Dordrecht.

Panov V.E. 1999. GAAS: Group on Aquatic Alien Species at the Zoological Institute in St.Petersburg, Russia. Biological Invasions 1:99-100.

Regional Biological Invasions Center. 2001a. Panov V, M. Dianov and A. Lobanov (Eds). http://www.zin.ru/projects/invasions

Regional Biological Invasions Center. 2001b. Panov V, M. Dianov and A. Lobanov (Eds). Interactive Geographical Information System for Documentation and Mapping of Alien Species Distribution. http://www.zin.ru/projects/invasions/gaas/invader/invader.htm

Regional Biological Invasions Center. 2001c. Panov V, M. Dianov and A. Lobanov (Eds). Aquatic Invasive Species of Europe. http://www.zin.ru/projects/invasions/gaas/aa_idb.htm

Regional Biological Invasions Center. 2001d. Panov V, M. Dianov and A. Lobanov (Eds). International Networking. http://www.zin.ru/projects/invasions/gaas/aa_netw.htm

Report of the Joint Convention on Biological Diversity/Global Invasive Species Programme Informal Meeting on Formats, Protocols and Standards for Improved Exchange of Biodiversity-related Information. 2002.

http://www.biodiv.org/doc/meetings/cop/cop-06/information/cop-06-inf-18-en.doc
Report of the Workshop on the Development of a Nordic/Baltic Invasive Species Informational Network. 2002. http://www.us-reo.dk/IS_Database.htm
Shiganova T.A. and V.E. Panov. 2002. *Mnemiopsis leidyi* A. Agassiz, 1865. Entry to the Illustrated Database, Regional Biological Invasions Center Information System. http://www.zin.ru/projects/invasions/gaas/mnelei.htm
The Biological Records Centre (BRC) website. http://www.brc.ac.uk
The Natural History Museum. 2002. The Chinese Mitten Crab. http://www.nhm.ac.uk/zoology/crab/
United States Geological Survey Nonindigenous Aquatic Species (USGS NAS) Information Resource. http://nas.er.usgs.gov
Zaitsev Yu. and B. Öztürk (Eds). 2001. Exotic species in the Aegean, Marmara, Black, Azov and Caspian Seas. Published by Turkish Marine Research Foundation, Istanbul, Turkey. 267p.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 125-131
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

History of the Continuous Plankton Recorder database

Darren Stevens and Philip C. Reid

Sir Alister Hardy Foundation for Ocean Science, The Laboratory
Citadel Hill, The Hoe, Plymouth PL1 2PB, United Kingdom
E-mail: dpst@mail.pml.ac.uk

Abstract

The Continuous Plankton Recorder survey has operated in the North Atlantic and North Sea since 1931, providing a unique multi-decadal dataset of plankton abundance. Over the period since 1931 technology has advanced and the system for storing the CPR data has developed considerably. From 1969 an electronic database was developed to store the results of CPR analysis. Since that time the database has undergone a number of changes due to performance related factors such as processor speed and disk capacity as well as economic factors such as the cost of software. These issues have been overcome and the system for storing and retrieving the data has become more user friendly at every development stage.

Keywords: Continuous Plankton Recorder survey; Database; Data storage; SAHFOS.

Introduction

The Continuous Plankton Recorder (CPR) is a high-speed plankton sampler designed to be towed behind commercially operated 'ships of opportunity'. The CPR is used to monitor plankton near the surface of the North Sea, the North Atlantic and in recent years the North Pacific when possible on a monthly basis. The resulting data are employed to look at seasonal cycles and year-to-year changes of species, their geographical distribution and the impact of climate change and fisheries on the plankton community. The first CPR was designed by Sir Alister Hardy and used during the Discovery Expedition to the Antarctic in 1925-1927 and a smaller version developed for use on merchant ships of opportunity by 1930. The CPR survey is currently operated by an international charity: the Sir Alister Hardy Foundation for Ocean Science (SAHFOS).

Since September 1931, the survey has towed nearly 5 million nautical miles throughout the Northern Hemisphere, with analysis completed on ~ 200000 samples, providing over 2 million positive plankton entries. During analysis ~ 500 different taxa (~ 200 phytoplankton, ~ 300 zooplankton) are identified (~ half to species) and counted. Data post 1946 are currently stored electronically in a relational database, developed using Microsoft Access; data prior to this are stored in paper format. This paper reviews the historic development of the CPR database, considering changes in computer hardware and software, through to the present system in operation.

Overview of the CPR survey and plankton analysis

CPRs are towed in the surface mixed layer (Hays and Warner, 1993). Water enters the machine through a 1.27cm square aperture, passes down a tunnel and through a silk filtering mesh. The silk moves at a constant rate (proportional to the speed of the ship) across the tunnel, powered by an impeller in the tail of the CPR (Fig. 1). The silk mesh is covered by a second layer of silk (covering silk) as it leaves the tunnel so the plankton is sandwiched between two layers of silk, before being stored in a tank of buffered formaldehyde solution. On returning to the laboratory the silk is processed in 10 nautical mile sections (samples). For a more detailed explanation of how a CPR works see Warner and Hays (1994). Using a microscope plankton analysis is divided into three stages: (i) phytoplankton (ii) zooplankton 'traverse' and (iii) zooplankton 'eye count'.

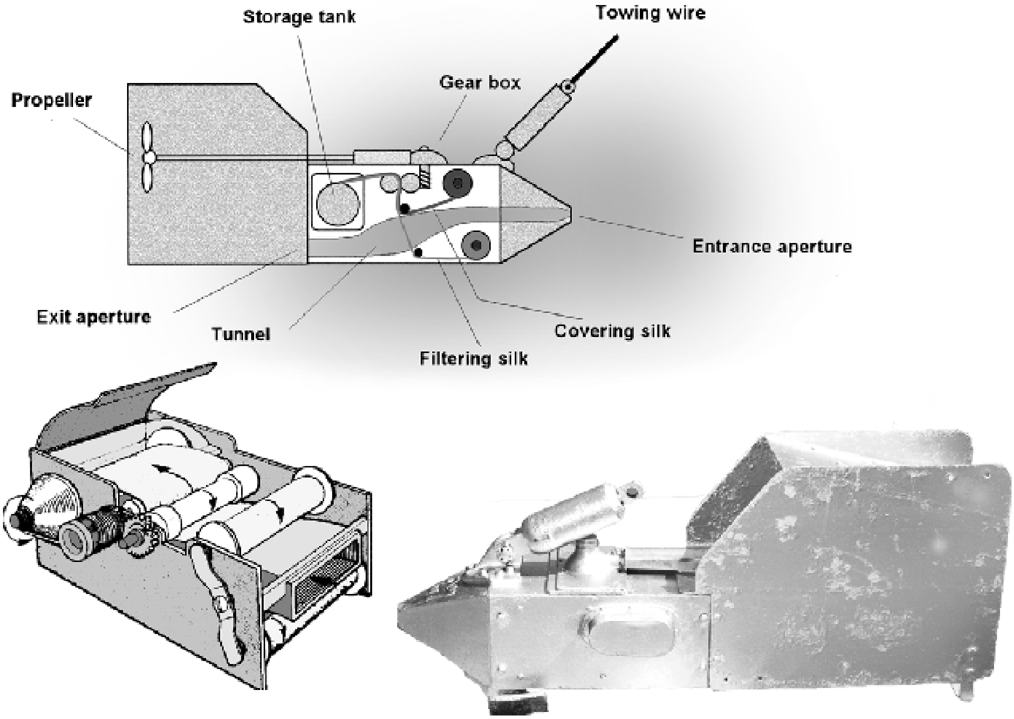


Fig. 1. Top: a cross section diagram of a continuous plankton recorder (CPR).
Bottom left: a diagram of the internal mechanism of a CPR.
Bottom right: a photograph of a CPR.

Phytoplankton analysis

The silk is viewed under x450 magnification and a traverse of the filtering silk is made in which 20 fields of view, each one centred on one mesh of the silk, are examined. The phytoplankton is identified and the number of fields of view in which each species is seen is then recorded. This method of analysis has not changed since 1958.

Zooplankton traverse

A traverse of both the filtering and covering silks at x54 magnification is conducted covering 1/40th of the silk. Zooplankton (< 2mm) is then identified and counted. This method of analysis has not changed since 1948.

Zooplankton eyecount

Larger zooplankton (> 2mm) are removed from the silk and identified and counted. This method of analysis has also not changed since 1948. For a more detailed description of the analysis process refer to Rae (1952) and Colebrook (1960).

Data storage

When the Survey first started data were plotted along routes. As data accumulated it soon became apparent that new ways of storing and processing the plankton results needed to be developed. Prior to 1969 all data were stored on cards and all calculations were completed by hand. Data were stored in paper format several times over to allow easy access. Initially the plankton data would be entered onto analysis sheets for each tow; these analysis sheets contained data for all species on all samples taken during that tow, this process is still in operation and paper records maintained. The data was then transposed onto cards for each species; these cards contained data relating to a particular species for all routes on which the species had occurred for one year. The plankton data was also log transformed and stored on large maps of the North Atlantic and North Sea in 1° of latitude by 2° of longitude rectangles for each species as arithmetic monthly means. At the end of the year this data would then be calculated, for a selected set of dominant species, into annual means for larger statistical areas called standard areas (Fig. 2). This exercise was labour intensive and a time consuming process and it is estimated that a team of more than 12 scientists/analysts spent a third of their time entering data for the Survey onto these charts. Mechanical calculators were used initially to assist data processing and progressively replaced by electronic equivalents of increasing sophistication. From the early 1960s computers were used to aid calculation and in the statistical analysis of data. However it was not until the late 1960s that plans to develop a computerised database were put in place.

The first CPR database was designed in 1969 on a KDF 9 based at the Edinburgh Regional Computing Centre (ERCC). The database was originally designed backwards using standard area data that had already been processed, rather than the raw data. This was due to the limitations of the computer equipment available at the time, and their inability to handle large amounts of data. The database was written in ALGOL (ALGOrithmic Language), a high level programming language. Both J.M. Colebrook and H.G. Hunt, researchers connected with the CPR survey, were involved with the early development of the database and continued to be involved until the late 1990s, providing a level of continuity not usually witnessed in computing. Programs were developed to help automate procedures such as calculating Cartesian co-ordinates of the midpoint of a sample, and calculating the 'cutting points' to divide a towed silk into 10 nautical mile samples. Originally this had taken some considerable time and effort but that was eased by the use of an Olivetti desktop programmable calculator during the late 1960s.

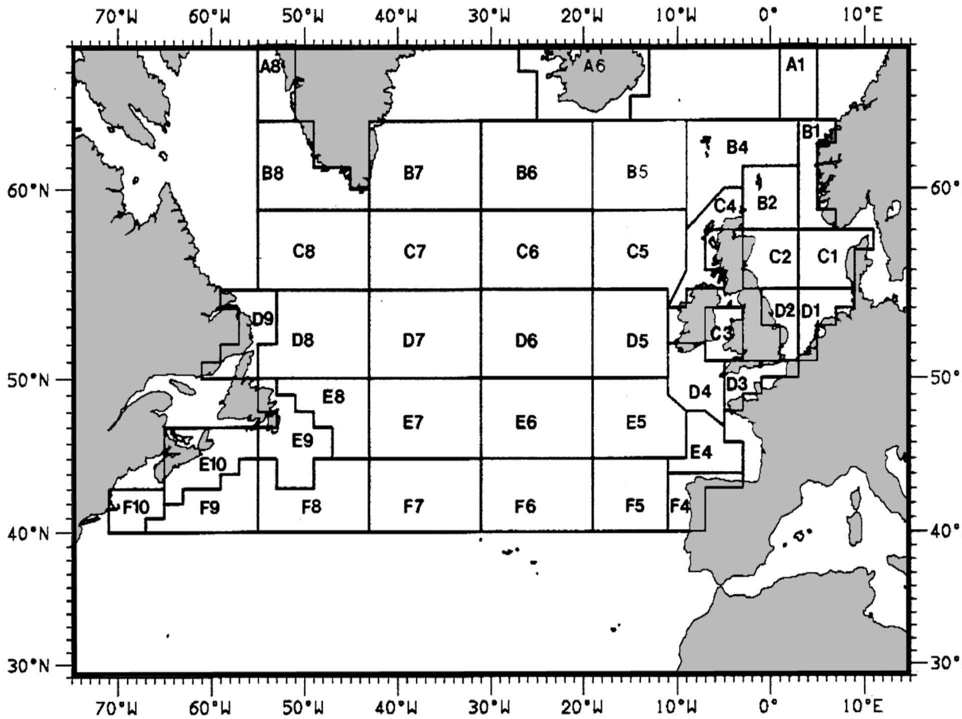


Fig. 2. Map of the CPR standard areas.

The ERCC replaced the KDF 9 in 1971 with an IBM 360/40 and support for ALGOL was withdrawn, the result was that all programs that had been developed for the CPR database had to be rewritten into a new programming language called IMP (Glover, 1970). Implementation language (IMP) was a programming language developed as part of the Edinburgh Multi-Access System (EMAS) and localised to Scotland (Stephens, 1974). A more detailed description of the early development of the CPR database can be found in Colebrook (1975).

During the early 1970's the CPR team took steps to ensure historic data was entered retrospectively using punched cards to form the first computerised CPR database containing raw data. The data was entered twice for each tow and the two sets of punched cards were compared to check for errors.

In 1977 the survey was relocated from the Oceanographic Laboratory Edinburgh, Scotland to the newly established Institute for Marine Environmental Research (IMER), later Plymouth

Marine Laboratory (PML), in Plymouth, England. This move again forced changes on the CPR database. Initially, data processing was still carried out at the ERCC, with the information being sent by teletype line. After the programs had been rewritten, this time into PASCAL and FORTRAN (IV) G, the CPR database moved home to a PDP 11 computer based at IMER. This was followed by a move to a GEC 41/90 computer in 1982 and then to an IBM 93/70 in 1989. During this period there was no need for any significant software rewrites, although the FORTRAN programming was now taking place in FORTRAN 77.

In 1991 the database moved to a RS 6000 computer, which experienced difficulties when communicating with a local area network, and was subsequently withdrawn from service resulting in another move for the database. By this point all programming was written using FORTRAN 77. In 1992 the database moved to a Sun Solaris 2.4 computer. This computer was used to store the data, but not to perform calculations on pre-processing data or any manipulation of the data; this was done by an IBM PC. At this point, researcher Harry Hunt had developed a library of ~100 program files to process and manipulate the data, including some programs to check for errors and to randomly distribute the CPR samples amongst the analysts. At that time the data stored electronically went back to 1946 with four files per year of data, creating nearly 200 raw data files. It was also necessary under the system to record zero values making the file sizes, in bytes, larger than necessary. As this system was a file-based database, managing the data proved difficult.

A decision was taken in ~1992 to move the data to a relational database. Harry Hunt developed programs to transpose the data from the current data files into a relational format and the data was stored using ORACLE, in 1993 (Warner and Hays, 1994). Due to the expense of maintaining ORACLE, this database was short lived and never fully adopted by the survey team as the pre-processing of samples was still taking place using the FORTRAN programs previously developed. In 1995 the CPR data was transferred to a Microsoft Access relational database developed by the Database Manager Andrew Warner. This was the first Windows based database designed for CPR data. For the first time data could be extracted from any area so researchers were no longer restricted to the CPR standard areas. Though all pre-processing of data was still performed using the FORTRAN programs, this was not a problem at this time as the programs to transpose the data into a relational format had already been written two years previously.

In 2000 the decision was taken to develop a new system to process the sample positions and ‘cutting points’, as well as some simple programs to check the data based on the FORTRAN system that was currently in use. The reason for this decision was due to the FORTRAN system being developed under the IBM OS/2 operating system which was no longer widely used, and worries about the future compatibility of the programs. A software developer Andrew Radmore, was employed to develop a new system using Microsoft Access to allow easier exchange of data between the data input and data output/storage systems. The input and storage systems were intentionally not fully integrated into a single database so that the data could be scrutinised carefully before being finally archived into the CPR database. This ensures the quality of the data is maintained.

Quality control and planning

When it comes to checking data quality, computers are not always the best solution. Simple error checking programs have been designed during the development of the CPR database but these are only used as guides for the senior analysts who check the data personally. There are many factors to take into consideration when examining the data, such as time of day and geographical position of the sample, as well as comparing the sample with others from the same tow. It would be foolish to attempt to replicate an experienced analyst's specialist knowledge using a computer, and would lower the quality of the data if such a system was implemented. The data is also cross-examined with information from the operations manager and the analysis team to ensure all archived data is correct. This ensures geographical and temporal data for successful tows are included in the database, excluding samples that are not analysed.

Planning for data storage is a major concern for all organisations involved with the curation of data. The CPR database has been moved onto new hardware systems using different software platforms during its life. It is important to ensure all data are stored on current media to secure access to the material in the future. Before a piece of equipment used to create data stores or backups is made redundant, all data contained on that media must be transferred to a new media. An organisation cannot be dependent on a single piece of equipment to read a certain type of media in case of failure, and the equipment may have become obsolete and non-replaceable.

Changes in software require planning; this is a costly and time-consuming exercise, though essential if data is to remain accessible. The need to keep up-to-date with software developments is twofold. Firstly you must ensure that your data is easily accessible by all users, as newer versions of software become standard. This does not mean upgrading software just because a new version is available. Secondly, as older software becomes obsolete so the number of professionals skilled in that area decreases, possibly creating recruitment problems in the future.

The future

The future development of the CPR database is likely to centre on the Internet and web development. The aim is to increase the ease of access to data from the CPR survey by allowing researchers to access the data themselves quickly via the World Wide Web. Steps have already been taken to achieve this with phytoplankton colour and *Calanus finmarchicus* data for the CPR standard areas available on the SAHFOS web site (www.sahfos.org). The inclusion of other environmental variables into the database is another development whereby preliminary steps have been taken and this work will be developed further.

As changes and developments in information technology continue to take place, so the CPR database will change accordingly, whilst maintaining the quality of the data and ensuring all data and programs are stored for future reference. The pace of change will be governed by the speed of hardware and software development to maintain compatibility with current systems, but all development will continue to ensure all systems are fully tested before implementation can occur.

As the development of new and larger capacity media on which to store data continues to evolve, it is necessary for all data managers to keep up-to-date with current developments in this area. All those involved with the archiving of data need to develop a strategy to ensure that historic data as well as recent data is transferred and stored on a current and popular data storage medium. If data are not stored on current media researchers will have difficulty accessing historical data or worse, the data could be lost completely.

The issue of software development must also be addressed, and a process for future development incorporated into the long-term aims and financial planning. Historically the CPR survey team has invested many hours into the development of the CPR database, which is a costly exercise. As processor speed continues to increase it will be necessary to upgrade software to utilise this increase in performance. Developers should budget accordingly for the necessary software changes.

The CPR database has a long history of changes and developments, providing us with a unique insight into these issues, helping us plan for the future.

Acknowledgements

We thank Tony John and Alistair Lindley for their insightful knowledge of the CPR survey. Mike Colebrook for providing an interesting history of the early development of the database. Sandra Chenery for information regarding the computing changes that happened at the West Hoe laboratory. Marion Smith for proof reading.

References

Colebrook J.M. 1960. Continuous Plankton Records: methods of analysis, 1950-1959. Bulletin of Marine Ecology 5:51-64.

Colebrook J.M. 1975. The Continuous Plankton Recorder Survey: automatic data processing methods. Bulletin of Marine Ecology 8:123-142.

Glover R.S. 1970. Laboratory Report: Oceanographic Laboratory, Edinburgh. Scottish Marine Biological Association. Annual Report 1969-1970:43-64.

Hays G.C. and A.J. Warner. 1993. Consistency of towing speed and sampling depth for the Continuous Plankton Recorder. Journal of the Marine Biological Association of the United Kingdom 73:967-970.

Rae K.M. 1952. Continuous Plankton Records: explanation and methods, 1946-1949. Hull Bulletins of Marine Ecology 3:135-155.

Stephens P.D. 1974. The IMP language and compiler. The Computer Journal 17(3):216-223.

Warner A.J. and G.C. Hays. 1994. Sampling by the Continuous Plankton Recorder survey. Progress in Oceanography 34:237-256.

SESSION: NEW DEVELOPMENTS

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 135-145
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Using XML technology for data and system metadata for the MBARI Ocean Observing System

Daniel Davis, Tom O'Reilly, Duane R. Edgington, Rich Schramm and Karen A. Salamy

Monterey Bay Aquarium Research Institute
7700 Sandholdt Rd., Moss Landing, CA 95039, USA
E-mail: dada@mbari.org

Abstract

The Monterey Bay Aquarium Research Institute (MBARI.ORG-WWW), a non profit, privately funded, research institute devoted to the development of oceanographic technology has been developing systems for long term environmental monitoring in Monterey Bay since 1987. The institute has initiated a project for expanding its ocean observing system capabilities through an expansion of existing moored data acquisition systems. The MBARI Ocean Observing System (MOOS) project is designed to develop highly flexible, configurable, and redeployable, ocean observing systems utilizing a wide variety of sensors, instruments and platforms, including moorings, ocean-bottom substations, and unmanned underwater vehicles (UUVs). The goal is to provide semi-continuous observations of important physical, biological, and chemical variables extended in space and time to support long term monitoring and event detection, such as the onset of an El Niño, as well as support for focused, intermediate-term, scientific process studies.
A primary challenge of this effort is the development of a software architecture to manage, control and access a variety of sensors, instruments, and platforms and manage the movement of data over a wide variety of communication links to users. The subsystem that performs these functions is called the MOOS Instrument Software Infrastructure (ISI). An associated challenge is to insure that all the ancillary data, the metadata, required to interpret both system and scientific data produced by the system are also properly defined, managed and made accessible to users along with the data streams. The approach taken by MBARI is to utilize XML technology to address this problem.
We provide an overview of the MOOS ISI system design and an overview of the scientific and system requirements that must be met by the metadata architecture. We describe a simple prototype process developed at MBARI and based on XML that can be used not only to define classes of metadata, but also to design user-friendly metadata forms to create, revise and access XML based metadata documents to satisfy a variety of requirements for the observing system.

Keywords: Ocean observing systems; XML; Metadata; Smart networks; UUV.

Introduction

There is a steadily increasing interest in the use of ocean observing systems for monitoring the health and conditions of the worlds' oceans. Both international and national efforts to design and implement such systems are well underway (Malone and Cole, 2000; Nowlin et al., 1996; GOOS-WWW; OCEAN.US-WWW). A variety of new technologies hold promise for resolving some of the difficult technical issues and challenge these types of systems pose. First and foremost is the use of web technology to provide broad user access to the observational products

of these systems. Less known, but also important, is the advent of 'smart networks', which incorporate the automatic ability to reconfigure an observing system with new platforms and sensors through the concept of 'plug and work'. All these technologies are based on modern notions of object-oriented software design and development for distributed systems.
A companion innovation to this revolution in software is new technology for managing information over distributed systems. The technology is an adaptation of general markup languages which have evolved from origins in document processing to extensible markup languages (XML) for managing information over distributed, highly configurable, network based systems (XML.COM-WWW). Just as these other web-based technologies are essential to the design and implementation of ocean observing systems, XML is gaining attention as an enabling technology for managing and accessing the information that observing systems produce (MEDI-WWW; MARINE.XML-WWW). A central feature of this technology is the ability to describe data independently of how the data is ultimately presented. Thus XML has the potential for being a universal language for describing data used over networks irrespective of how the data is ultimately used or displayed.

As ocean observing systems evolve, the types and form of the data they produce will evolve as well. This poses the challenging problem of utilizing and comparing similar or equivalent data over time in the presence of an ongoing evolution in sensor, instrument, and platform technology. A key to solving this problem is the early capture of metadata, along with the data, in a form that enables the scientific interpretation of the data over time.

Therefore a primary challenge of the metadata problem is providing highly flexible, easy to use solutions that can respond readily to changes in systems and the data they produce and are yet sufficiently simple that they have the potential of being widely used and disseminated. Describing specific metadata solutions for existing types of data is not the primary focus of this paper. The focus here is on a process that exploits existing XML technology to provide simple, easy to use methods to define, create and use metadata and to illustrate a prototype of this methodology for a system currently in development and planned for deployments in the immediate future.

Overview of the MBARI Ocean Observing System

A system diagram for MOOS is shown in Fig. 1. The system now under development will enable coordinated data acquisition from a diverse set of sensors, instruments and platforms. MOOS network software is already in the early stages of implementation at MBARI. The system requirements that are driving the design and implementation are the result of several years of analysis of the proposed scientific uses of the system, and discussion among scientists as well as ongoing deployments of prototypes dedicated to specific science experiments (MBARI.ORG-WWW). The required capabilities of MOOS present some very interesting engineering challenges. In this paper, we will look at these challenges from the software perspective, and review technologies both available and proposed, which may provide solutions. It is not the purpose of this paper to completely review the requirements and goals of the MOOS system but only to summarize the elements of the system required to provide a context for the subject of this paper.

A portable mooring system, with attached seafloor fiber optic network is the first element of MOOS; this platform will host surface, midwater, and benthic instruments, as well as a docking station for autonomous underwater vehicles (AUVs). The mooring will be but one component of a much broader ocean observing network, which incorporates ships, ROVs, AUVs, drifters, and other instrument platforms. These platforms (some already existing, some yet to be developed) will enable data collection over a broad geographic area and throughout the oceanic water column, from the sea surface mixed layer and euphotic zone, through the midwater, to the deep ocean seafloor. The instruments themselves will range from current-off-the-shelf (COTS) instruments such as CTDs and fluorometers, to novel custom-made devices developed at MBARI and elsewhere. Instrument command interfaces are very diverse, as currently there are no widely accepted interface standards; thus the MOOS network software architecture must accommodate a variety of protocols.

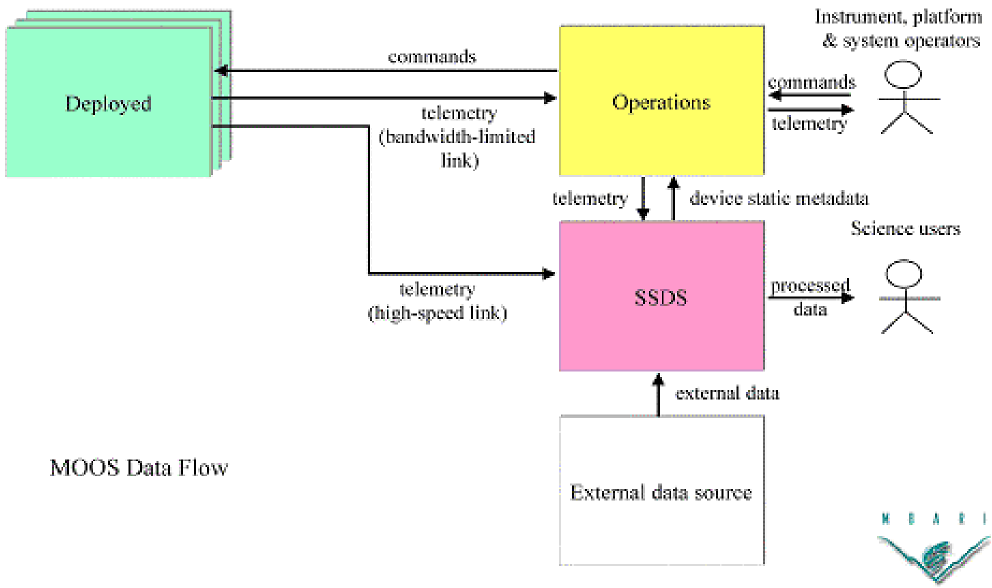


Fig. 1. System block diagram for MOOS.

The deployed subsystem

The deployed subsystem is the part of the system deployed at sea and can consist of moorings, ships, UUVs and a benthic network of substations connected to the surface by fiber optic links through a riser to a mooring, and then by radio links, or direct cable links, to shore. The subsystem nodes may also include a variety of UUVs in intermittent acoustic communication with local subsystem elements. UUVs may also 'dock' with subsystem elements and directly download data for transmission to shore through communication links. The deployed subsystem is also referred to as the 'wet side' of the system.

Shore Side Data Subsystem

A companion to the MOOS 'wet side' is a shore side data system (SSDS); also referred to as the 'dry side'. The SSDS is to be the repository for all the scientific data collected, and to provide users with access to this data. A primary requirement is to insure that all the metadata required for properly describing and interpreting the data from the system instruments is collected, catalogued, and accessible along with the data. In the context of a highly flexible and reconfigurable observing system such as MOOS, this problem is particularly challenging. The SSDS also provides a capability for accessing data sources external to the MOOS system.

Operational Subsystem

A third aspect of the MOOS system is the operational need to monitor, control, diagnose and recover from system failures in 'real time'. This places further demands on the MOOS architecture. The architecture must be capable of capturing the current status of system elements (sensors, instruments, platforms and communication links), as well as controlling and modifying them. An audit trail or 'history' of the system state is also required to support diagnostics and recovery and to fully identify the system state as an aspect of data interpretation. Additional system metadata will need to be defined and accessible to support the operational subsystem.

Smart network technology and the instrument software infrastructure

A fundamental feature of the MOOS effort is the utilization of recent advances in 'smart network' technology to design an instrument software infrastructure which provides real-time reconfiguration, remote device control, and automated event detection and response within a network with limited bandwidth links (radio-frequency and acoustic). Smart networks also provide a capability for 'plug and work' instruments, or automated device and service discovery, based on a distributed object-oriented software architecture. The flexibility for reconfigurability of the system created by the use of this technology must also be reflected in the approach to metadata issues. The software infrastructure being designed to support these features is called the MOOS Instrument Software Infrastructure or MOOS-ISI (O'Reilly et al., 2001).

The metadata problem

The conventional view of metadata is that it is ancillary information required to properly interpret a data set or data stream. It is generally discussed with reference to scientific data. Sometimes it is used in reference to catalogs of data used to search for and access datasets. An example is the Marine Environmental Data Inventory system developed and maintained by IOC (MEDI-WWW). Here we consider metadata only in an ocean observing system context, which not only includes metadata for the scientific data streams, but also metadata for the observing system as a whole. That is, any data required for interpreting the data from the system, or any data required to manage, control, or describe the state of the system at any time, will be viewed as metadata. Some of this 'system metadata' may be useful scientifically as well.

Any methodology used to address the metadata problem must provide methods to define, capture, access, and represent the metadata. It is commonly understood that the forms that metadata can take vary greatly. The fact that a certain person or laboratory performed the calibrations of an instrument may be important information when it comes to interpreting data. This is particularly the case in the presence of new sensor development. A description of the processing of raw sensor data to create data streams may be important metadata. The fact that a specific sensor came from a particular laboratory, or manufacturer at a certain time, may be critical. The fact that other data streams are available along with the data stream of interest may be another critical piece of information. Or even the status of the observing system, such as the results of automated event detection or system failures at critical times of observation, may be important information to properly interpret data. Thus the first requirement of any methodology used to define metadata is that it must be capable of describing information in a wide variety of forms. There is a need to represent a wide variety of data structures: numeric, string and character data, as well as lists, arrays, tables, and hierarchically structured and other structured forms of data.

Another critical challenge is to insure that important metadata is defined and entered as soon as a system element becomes operational. It must be easy to define and create important metadata for the system to support early metadata capture, or to easily modify metadata whenever the system is modified. There are numerous standards efforts dedicated to carefully describing what metadata may be required for various types of scientific data, but unless it is relatively easy to implement such standards, the metadata may not be created at all, or easily accessible to users of the system.

Another requirement is that even though users or systems may capture the same essential information in different forms, it must be relatively straightforward to convert from one form of metadata information to an equivalent, but different, form. As well there must be a capability to view and display data and metadata in a wide variety of ways. This means that the methods used to represent and store information may be quite separate from the means by which the information is displayed or viewed. In particular, it should be possible for users to view and interact with the metadata in a user-friendly manner, that is, a form natural to a particular user. Thus there should be simple interfaces to define what needs to be captured, for capturing it, and for accessing and displaying it after it is captured.
Given that ocean observing systems are evolving but may be deployed over long periods of time, or reconfigured for shorter deployments, it is necessary to have a technology that provides for incremental development and for readily modifying, without completely redoing, what has already been done.

Finally it is critical that the technology used to represent metadata does not constrain, imply, or define a particular science policy on the interpretation of data, but provides the widest variety of mechanisms for implementing policies on how metadata may be used to interpret data. In short, the technology should not constrain but instead facilitate what is most natural to its users.

A prototype process for using XML for metadata

The primary concept behind using XML for system metadata is that this technology provides a flexible and adaptable language and technology for defining, managing, rendering and displaying the data in a variety of ways. The features of XML technology provide opportunities to address a number of challenging problems in the metadata capture and management problem. There is widespread recognition that the technology meets the capabilities required by the metadata problem as outlined above. As well, there is already a rich tool-set of applications that can be utilized to address many of the challenging problems. A concern shared by some is that it would be unrealistic for users to learn the specific syntax and theory behind XML in order to obtain its benefits. A key to solving this problem is to have user friendly interfaces created by developers that eliminate any need for users to master the technical aspects of XML technology. In this way users can benefit from the features of XML without necessarily requiring them to know much about its details.

In the case of the MBARI Ocean Observing System we require mechanisms that allow scientific users to create, manage, and access metadata for scientific data. But we also need to provide operational users an equal ability to create information about sensors, instruments, platforms, communication links between sensors, instruments, or platforms and shore, and the status of system elements that meet system and operational needs as well.

We propose a prototype process that involves two basic steps. The first step is carried out by designers who define the types of metadata structures for the classes of objects that make up the system. Along with each object class, definition designers create user-friendly forms that can be used by the operational and scientific users of the system to create specific instances of these objects for the actual system. The goal of the first step is to specify what kinds of metadata are required by users for a class of sensors, a class of instruments, a class of platforms, or a class of communication links.

In summary the prototype process we have developed at MBARI for using XML in MOOS are:
1) A developer or designer uses an existing application like XML Spy to design a XML schema to describe a type of metadata needed for a particular class of system object, whether it is scientific or operational. The designer must take into account what information the user needs for objects in this class.
2) The developer or designer uses the schema to design a basic user interface class form for users to create or access a specific instance of the metadata described by the schema. The methodology illustrated below uses the visual interface design to create a JAVA application that displays a form that can be easily used to generate, view or revise, the XML metadata documents for each system object. Although the metadata may be stored in XML form, or even a compressed form, it is created, viewed and revised only through metadata object class forms.
3) System users, scientists, technicians and system operators use the forms for each class of system object to generate and view specific XML metadata without any requirement for the user to know or understand the XML language itself. Developers or designers can also develop other views, including partial views, of the metadata for specific purposes using the same tools. Designers can also design and develop converters using XSL (XML.COM-WWW) to create graphic displays or transform from one form of data to another.
4) The XML system metadata documents themselves are associated to the system objects, or their data streams, by the system infrastructure through a system of identifiers and timestamps. Data streams are associated with their source through its system identifier and are time stamped as well. The identifiers along with the timestamps provide a simple mechanism for obtaining any system metadata required by any particular data or object reference, at any time.

In the MOOS design approach there are four general classes of objects: sensors, instruments, platforms and communication links. Each of these object classes may have specialized subclasses of objects with their own metadata requirements. As stated above there may be separate metadata requirements for the science uses of the data, for system configuration and management, and for operational use. All system objects are required to possess a unique system ID. The software system infrastructure design associates to every system object metadata containers to hold that object's metadata.

In this approach, a series of individual metadata object class specifications are defined, which, when combined, form the complete set of specifications of the metadata required for all types of system objects. Specifications are defined using XML schema. The specific metadata for a specific system object are then instances of schema, i.e. XML documents, which conform to the specification schema. To illustrate the complete prototype process we will review these steps for a relatively simple class of system object called a sensor. The first step is to design a schema that describes the metadata required for the class of sensor. In this case we will do this for sensors that might occur, for example, on a CTD. Using an XML design tool, such as XML SpyTM, we could lay out the schema as shown in Fig. 2.

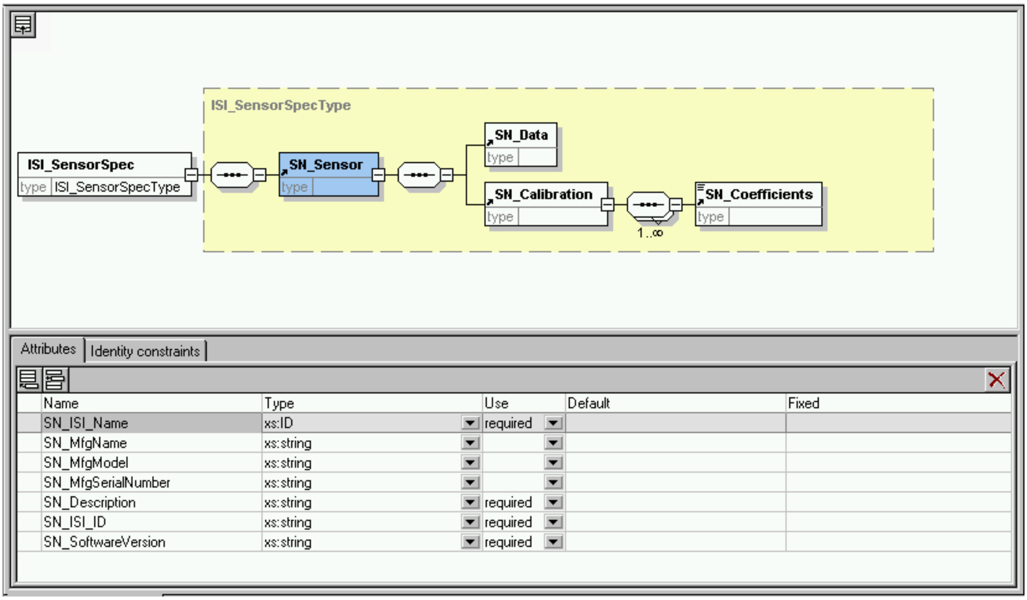


Fig. 2. XML sensor specification schema.

The root node ISI_SensorSpec is an ISI specification node. Its only attribute (not shown) is a locator field describing where it is filed or accessed over the web. This is followed by a SN_Sensor node, which has the following attributes as shown in Fig. 2.: SN_ ISI_Name , SN_MfgName, SN_MfgModel, SN_MfgSerialNumber, SN_Description, SN_ISI_ID, ISI_SoftwareVersion. This is followed by two nodes, SN_Data, and SN_Calibration. The calibration node is followed by a (possibly empty) list of SN_Coefficients nodes, consisting of name-value pairs. Each of the node elements has a list of attributes (shown only for SN_Sensor in the lower half of Fig. 2).

Once the schema design is established, the next step is to produce an entry form application that can create and access XML documents that conform to the schema. In this prototype methodology this is done using the IBM VisualAgeTM for JAVA IDE, and the EasyXML Bean Suite V4.0 available from the IBM Alpha Works developer site (IBM.ALPHA-WWW). This is primarily a matter of simply drawing a representation of the schema in tree form on a visual design surface, creating a user interface form using JAVA Swing or AWT bean elements for interacting with the XML beans representing the schema, and then 'wiring' the schema nodes and attributes to the interface elements. The visual composition process is shown in Fig. 3. The 'wires' describe how actions on the form bind to elements (nodes, or attributes) in the XML schema tree. The connections can be made two ways. If an XML document is retrieved, it is parsed by a SAX parser and the resulting events can cause information in the document to be used to fill in the form. Conversely events created by the user interacting with the form cause information in the form to be written to the nodes of the schema, and when saved, written out in syntactically correct XML according to the schema. As indicated, the tree diagram to the right of the visual form corresponds exactly to the nodes and attributes of the nodes defined in the schema in Fig. 2.

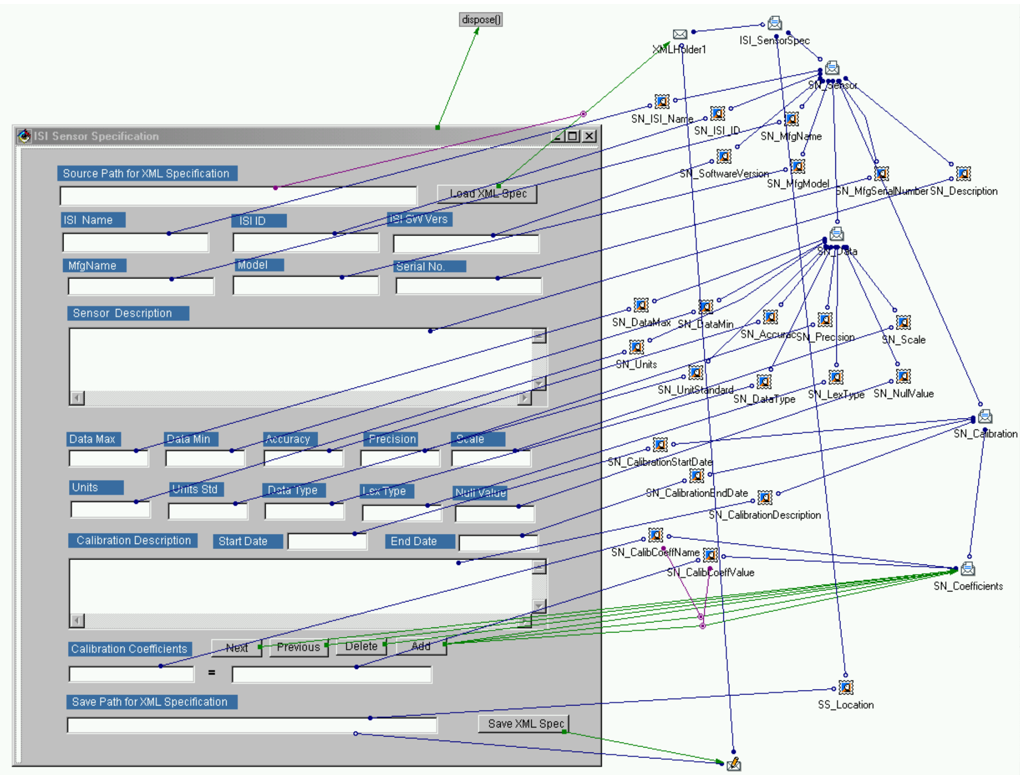


Fig. 3 . User interface form design, and wiring diagram, for XML sensor schema.

Once the visual composition process is complete, it is used to automatically generate JAVA code that can be executed on its own, or in an applet within a web application (Fig. 4). Using the generated application, the user fills in the fields on the form shown by the application for sensor specifications, and when complete, the application will create an XML metadata document based on the XML schema (Fig. 5). As indicated, because of the two-way nature of the connections in the composition, the same application that creates an XML metadata document for a system object can also be used to read or revise one already written. Literally no JAVA code needed to be manually written to create this application. The entire process is visual.

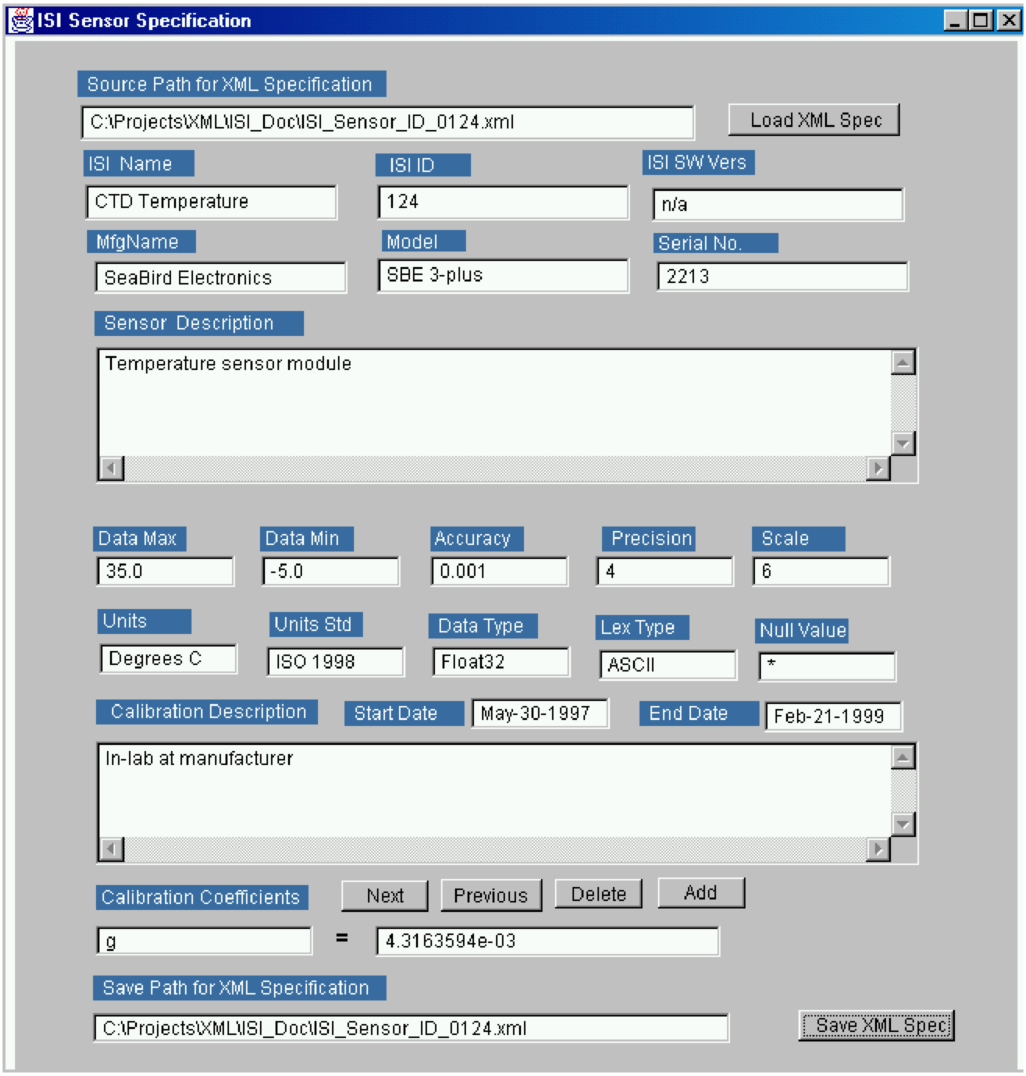


Fig. 4. User interface form to create, view, and revise XML sensor metadata.

When the application is executed, the form comes up with empty fields, unless default values are set in the form. The information that is required for each field is fairly self-explanatory to a user. Once the fields are filled out the user defines the locator reference in the 'Save path for

XML Specification' edit field and then selects the 'Save XML Spec' button. The XML document is saved in correct XML (Fig. 5). The application automatically enters the same locator reference in the 'Source path for XML Specification'. If the user wishes to review or revise the saved file, it can be loaded by selecting the 'Load XML Spec' button. In this way a collection of XML metadata documents can be created for all system elements. Although not shown here, a platform metadata schema and form has been prototyped that includes a list of the Sensor_ISI_ID, and file locators for the metadata documents for the sensors on that platform. In essentially what is an object-oriented methodology, the schema is a class definition, the XML metadata documents are instances of the class and the interface forms are create, view, or revise methods associated to the class. For this reason it is convenient to organize the information created and managed by these tools in an object-oriented directory structure. Each directory is associated to an object schema (a file of type 'xsd'), includes the creator/viewer/editor application corresponding to it, and all the XML metadata documents that conform to the schema and created by the application. This is just one simple way that a potentially large amount of metadata could be managed. The plan at MBARI is to create a catalog of metadata accessible within the database that manages the data, to provide access to the required metadata.


```
  <?xml version="1.0" encoding="UTF-8" ?>
- <ISI_SensorSpec SS_Location="C:\Projects\XML\ISI_Doc\ISI_Sensor_ID_0124.xml">
  - <SN_Sensor SN_Description="Temperature sensor module" SN_ISI_ID="124"
      SN_ISI_Name="CTD Temperature" SN_MfgModel="SBE 3-plus" SN_MfgName="SeaBird
      Electronics" SN_MfgSerialNumber="2213" SN_SoftwareVersion="n/a">
      <SN_Data SN_Accuracy="0.001" SN_DataMax="35.0" SN_DataMin="-5.0"
        SN_DataType="Float32" SN_LexType="ASCII" SN_NullValue="*" SN_Precision="4"
        SN_Scale="6" SN_UnitStandard="ISO 1998" SN_Units="Degrees C" />
    - <SN_Calibration SN_CalibrationDescription="In-lab at manufacturer"
        SN_CalibrationEndDate="Feb-21-1999" SN_CalibrationStartDate="May-30-1997">
        <SN_Coefficients SN_CalibCoeffName="g" SN_CalibCoeffValue="4.3163594e-03" />
        <SN_Coefficients SN_CalibCoeffName="h" SN_CalibCoeffValue="6.41530157e+04" />
        <SN_Coefficients SN_CalibCoeffName="i" SN_CalibCoeffValue="2.27235685e-05" />
        <SN_Coefficients SN_CalibCoeffName="j" SN_CalibCoeffValue="2.1715345e-06" />
      </SN_Calibration>
    </SN_Sensor>
  </ISI_SensorSpec>
```


Fig. 5. XML sensor metadata document created by interface form.

Conclusions

The main conclusion has already been widely recognized: XML technology is the appropriate choice for managing the metadata problem for ocean observing systems, as well as data exchange. We have also shown that it is relatively straightforward, using this technology, to design metadata specifications for classes of objects in an observing system and to design user friendly forms for creating, displaying, and revising metadata documents expressed in XML without requiring the user to read XML.

We have also illustrated a prototype process for developing metadata specifications, using XML schema, and show how users can create, access, view and revise metadata documents in XML.

We have also suggested that an approach to metadata design, use and management, using some object-oriented design principles for defining a class object hierarchy, with associated methods and instances, may be an appropriate metaphor. In this metaphor, metadata specifications become object class definitions, and are expressed using XML schemas. The forms for creating, viewing, revising metadata document for each metadata specification class become methods associated to the class, and are attached to the class. Specific metadata documents that describe specific sensors, instruments, platforms, and communication links are simply instances of the metadata specifications for these classes of system objects.

Acknowledgements

We wish to acknowledge our appreciation to the David and Lucille Packard Foundation for their generous support of research and development at MBARI, of which this effort is a part.

References

Malone T.C. and M. Cole. 2000. Towards a global scale coastal ocean observing system. Oceanography 13:7-11.

Nowlin W.D. et al. 1996. An observing system for climate. Bulletin of the American Meteorological Society 77:2243-2273.

O'Reilly Tom, Duane Edgington, Daniel Davis, Richard Henthorn, Michael P. McCann, Timothy Meese, Wayne Radochonski, Michael Risi, Brent Roman and Rich Schramm. 2001: 'Smart Network' Infrastructure for the MBARI Ocean Observing System. Oceans IEEE 2001 Conference Proceedings, Honolulu, Hawaii, Nov. 5-8.

Web sites

GOOS-WWW: http://ioc.unesco.org/goos/ (current November, 2002)
IBM.ALPHA-WWW: http://www.alphaworks.ibm.com/ab.nsf/bean/easyXML/ (current November, 2002)
MEDI WWW: http://ioc.unesco.org/medi/ (current November, 2002)
MARINE XML-WWW: http://ioc.unesco.org/marinexml/ (current November, 2002)
MBARI.ORG-WWW: http://www.mbari.org (current November, 2002)
OCEAN.US-WWW: http://www.ocean.us.net/home.jsp (current November, 2002)
XML.COM-WWW: http://www.xml.com/ (current November, 2002).

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 147-151
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Minimum requirements for reporting analytical data from marine environmental samples

Hans Egli[1], Manos Dassenakis[2], Hemda Garelick[3], René van Grieken[4], W.J.G.M. Peijnenburg[5], Leo Klasinc[6], Werner Kördel[7], Nick Priest[3] and Tanja Tavares[8]

[1] Syngenta Crop Protection AG, CH-4002 Basel, Switzerland

[2] University of Athens, Department of Chemistry, GR-15 771 Athens, Greece
E-mail: edasenak@cc.uoa.gr

[3] Middlesex University, School of Health and Social Sciences, London, N11 2NQ, United Kingdom

[4] Universitaire Instelling Antwerpen, Departement Scheikunde, B-2610 Antwerpen, Belgium

[5] RIVM, Laboratory for Ecotoxicology, NL-3720 BA Bilthoven, The Netherlands

[6] Rudjer Boskovic Institute, HR-10001 Zagreb, Croatia, 7 Fraunhofer-Institut für Umweltchemie und Ökotoxikologie, D-57377 Schmallenberg, Germany

[7] Fraunhofer Institute for Molecular Biology and Applied Ecology
D-57377 Schmallenberg, Germany

[8] Universidade Federal de Bahia, Instituto de Quimica
40.170-290 Salvador, Bahia, Brazil

Abstract

In view of the significance of marine environmental analytical data it is essential that the quality of both sampling and analysis be assured and that procedures used, as well as all relevant additional information are reported. There is a minimum level of information required in order to guarantee the fitness for use of the data. Emanating from discussions on the fundamental problems of the analysis of environmental samples for chemical contaminants, a general guidance is given regarding the minimum information that should be provided to adequately describe the methods of sampling, handling between sampling and analysis and the analytical methodology including calculation and validation procedures.

Keywords: Sea water; Sediments; Sampling; Analysis; Quality.

Introduction

Environmental analytical data are of major importance as they serve many purposes. They often attract high attention from the general press and the public and may even be used to support advisory or regulatory measures. Therefore it is essential that their quality is assured and supplementary information is published in addition to the analytical results. Only with sufficient additional information an appropriate use of the data is possible. The IUPAC Subcommittee on

Chemistry of Environmental Compartments has published a guidance document about the minimum information that should be provided if such analytical data are published (Egli *et al.,* 2003). This paper was prepared as a guide for the reporting of analytical data of marine environmental samples.

Each analysis concerning marine environment starts with taking samples from a much larger physical entity. All environmental compartments are considerably inhomogeneous, both spatially and temporally, and the inter-sample variability is usually large. Nevertheless, the sampling process has a crucial impact on the final analytical result. Between sampling and analysis, the analytes and other parameters can be subjected to alterations due to transport and storage conditions as well as the sample preparation procedures (*e.g.* filtering of water, sieving or drying of sediments) (Strickland and Parsons, 1968; Greeson *et al.*, 1990; UNEP/IAEA/IOC, 1991; USPHA, 1995; Grasshoff *et al.*, 1999).

The present recommendations list the information that should be provided about sample history and the factors that might have affected the results (data and metadata). The overall goal of the additional information to be provided is to enable the reader to judge the accuracy and representativeness of results, to compare results of different origin and to interpret them (Garner *et al.*, 1992; Ott, 1995).

Due to the variability in the aims and purposes of research projects that involve the analysis of marine environmental samples for chemical parameters, these recommendations cannot fully and unequivocally include all cases. They are rather meant to cover the most common project types. More detailed information would be needed if the analysis is part of a specific and comprehensive research project and less information would suffice in case of emergency situations, for instance in an oil spill.

Recommendations

Some general principles apply to all results, irrespective of the compartment of origin of the sample, the chemical or biological nature of the analyte and the method of analysis. They are grouped into two parts: (a) the sample description part, which includes information about the sample and its collection; (b) the sample analysis part, which includes information about all actions conducted on the samples between collection and generation of the final result of the analysis.

Sample description

The following information should be provided:

Adequate location of the sampling site and of the time of sampling
The clear location of sampling sites must be reported, preferably by indication of their longitude and latitude. A map may be very useful to visualize the sites. Hydrogeological, topographical, agricultural, industrial and other aspects that were used for the selection of sampling sites may also be of importance and should be reported. The sampling day must also be reported. The time of the day may be indicated depending on the goal of the study.

Description of the method of sampling
A description of the technique used for sampling is mandatory, including a description of the equipment and the type of samples (simple, replicate, composite etc.). The report of sample size is also useful as the larger the sample size, the lower is the sample-to-sample variability. In the case of non-homogeneous samples, such as coastal sediments, details of the method used for homogenization and taking sub-samples must be reported. In the case of on-line pre-concentration during sampling *e.g.*, an adsorbance onto a media surface, the efficiency and capacity of sampling device should be evaluated and reported (Keith, 1996; ISO, 2001).

Description of in situ *measurements*
In the case of *in situ* measurements with electrodes or sensors an adequate description of the experimental set-up is required together with the method of calibration and an estimation of the uncertainty of the final analytical result including both precision and accuracy (ISO, 1990 & 2001; Hunt and Wilson, 1986).

Additional information
Information about other site-specific factors that may affect the analytical data or be relevant for their interpretation should also be reported. This may include meteorological conditions prior to and during sampling, type and stage of tide in relation to yearly maximum and minimum or relevant anthropogenic activities adjacent to the sampling site (land use, agricultural practices, sewage systems, transport facilities, exploitation of resources, industries, waste treatment facilities, etc).

Sample analysis

The following information should be provided in the report:

Storage conditions between sampling and analysis
The time and conditions of sample storage must be described. This should include the material of the sample containers and preceding cleaning procedures. Moreover, evidence for sufficient stability of the analyte(s) and any pre-treatment (addition of preservatives, adsorption of the analyte onto cartridges, filtration or centrifugation,drying of the sample, etc.) must be reported. (Strickland and Parsons, 1968; Hunt and Wilson, 1986; UNEP/IAEA/IOC, 1991; Garner *et al.*, 1992; Grasshoff *et al.*, 1999).

Description of the analytical methodology
The entire analytical method must be described in details with the appropriate references. The description must include a sufficiently detailed description of the quantitative aspects and the performance of the method. The following information is necessary: a description of the calibration of the analytical system (concentration and chemical nature of internal and external standards, surrogate or isotopic tracers, number of calibration points, standard deviations); details of the mathematical procedures used to calculate and statistically interpret the results; number of replicate quantitations of the samples; the limit of detection of the detector and the limits of determination or quantitation of the entire analytical procedure, including the method of its calculation or estimation; an estimate of the precision and accuracy of the final analytical result (Hunt and Wilson, 1986; Grasshoff *et al.*, 1990; Greeson *et al.*, 1990; UNEP/IAEA/IOC, 1991; USPHA, 1995; Ellison *et al.*, 2000).

The requirements in this section are method depending and must be adapted accordingly.

In addition to the general requirements, the following specific information should be collected and reported (where applicable) in the cases of seawater and marine sediments:

Seawater
Depth; salinity; temperature; dissolved oxygen; separation method, if suspended material is separated (including filter type in case of filtration), or specification if no separation was made; other parameters if known or assumed to impact on the analyte to be determined such as suspended particles or transparency, dissolved organic carbon (DOC), pH, a statement on the level of biological activity, *e.g.* concentration of chlorophyll-A, position of main picnocline (thermocline or halocline) (UNEP/IAEA/IOC, 1991; Grasshoff *et al.*, 1999).

Sediments
'Horizon' sampled, sediment texture (particle size range or % clay, % silt, % sand); organic and inorganic carbon content; other parameters if known or assumed to impact on the analyte to be determined such as pH, redox potential, cation exchange capacity,biological activity,content in sorptive phases such as iron, aluminium or manganese, type of clay mineral. The results have to be presented as mg/g dry weight (UNEP/IAEA/IOC, 1991; MAP/UNEP, 1992).

The following rules should also be observed for the reporting of analytical data:

Units
SI-units, or internationally accepted units derived hereof, should be used and traditional units such as pounds, acres, inches avoided, even in countries where these units are still in common use. Concentrations should be expressed in units such as mol/m^3, ng/dm^3, ng/L, or µg/kg rather than in ratio numbers such as %, ppm, or ppb. These latter pseudo-units are ambiguous and may lead to misunderstandings. The expression of results in mmol/L (nmol or µmol) has to be preferred in the cases of nutrients and trace metals, whereas the mg/L (ng or µg) is better for organic compounds (MAP/UNEP, 1992; Ellison *et al.*, 2000).

Rounding
Figures should be rounded to a significant number of digits; no more than the last of the given digits should be uncertain due to the variability of the method.

Replicates
A clear indication whether the reported results are single measurements or mean averages of replicates is necessary, as well as a clarification of the type of replicates (i.e. replicate samples, replicate analysis, or replicate determination). Where mean results are reported, the range should be included and, where the reported result is based on at least three replicates, standard deviation should be given.

Conclusions

The analytical data about marine environment are of high significance as they serve many purposes and can be used to support advisory or regulatory measures. Therefore their appropriate use is possible only when their quality is assured and there is also sufficient additional information. This supplementary information may include detailed description of sampling and analytical procedures. Specific information is needed for various types of marine environmental samples like seawater or sediments. Significant attention is also needed in the presentation of the results concerning units, accuracy and precision. This paper was prepared as a guide for the proper way of reporting analytical data concerning marine environment. It is part of a wider guide that includes all types of environmental data and shall be published by International Union of Pure and Applied Chemistry (IUPAC). We think that the follow-up of the recommendations included in this paper shall be very useful for the better evaluation and use of environmental data and thus for a better environmental management.

References

Egli H., M. Dassenakis, H. Garelick, R. Van Grieken, W.J.G.M. Peijnenburg, L. Klasnic, W. Kordel, N. Priest and T. Tavares. 2003. Minimum requirements for reporting analytical data for environmental samples. IUPAC Technical Report. Pure and Applied Chemistry 75:1097-1106.

Ellison S.L.R., M. Rösslein and A. Williams. (Eds). 2000. EURACHEM/CITAC Guide. Quantifying uncertainty in analytical measurement. 2nd Ed.

Garner W.Y., M.S. Barge and J.P. Ussary (Eds). 1992. Good Laboratory Practice Standards. American Chemical Society Professional Reference Book, Washington DC.

Grasshoff K., K. Kremling and M. Ehrhardt (Eds). 1999. Methods of seawater analysis. 3rd Ed., Wiley-VCH, Verlag GmbH, D-69469 Weinheim.

Greeson P.E., T.A. Ehlke, G.A. Irwin, B.W. Llum and K.V. Slack. 1990. Methods for collection and analysis of aquatic biological and microbiological samples. Techniques of water-resources investigation of the USGS.

Hunt D.T.E. and A.L.Wilson. 1986. The chemical analysis of water, general principles and techniques. The Royal Society of Chemistry, Thomas Graham House, he Science Park, Cambridge.

International Organization for Standardization (ISO) Geneva. ISO 8466. 1990. Water quality. Calibration and evaluation of analytical methods and estimation of performance characteristics. Part 1. Statistical evaluation of the linear calibration function.

International Organization for Standardization (ISO) Geneva. ISO 8466. 2001. Water quality. Calibration and evaluation of analytical methods and estimation of performance characteristics. Part 2. Calibration strategy for non-linear second-order calibration functions.

International Organization for Standardization (ISO) Geneva. ISO 5667. 1980/2001. Water quality - Sampling - Parts 1-18, Guidance for various waters and sediment.

Keith L.H. (Ed.). 1996. Principles of environmental sampling. American Chemical Society Professional Reference Book, Washington DC.

Mediterranean Action Plan/United Nations Environment Programme. 1992. Computerization of MED POl data. Vol. 1, Revision 1.01.

Ott W.R. 1995. Environmental statistics and data analysis. CRC Press, Inc., NW Boca Raton, Florida.

Strickland J.D.H and T.R. Parsons. 1968. A practical handbook of seawater analysis. Fisheries Research Board of Canada, Ottawa.

UNEP/IAEA/IOC. 1991. Standard chemical methods for marine environmental monitoring. Reference Methods for Marine Pollution Studies 50. UNEP, Nairobi.

USPHA. 1995. Standard methods for the examination of water and wastewater. 19th Ed. Prepared and published jointly by: American Public Health Association, American Water Works Association, Water Environment Federation. Joint editorial board: Greenberg Arnold E., Andrew D. Eaton, Lenore S. Cleseri. Washington, D.C. American Public Health Association.

[illegible] 2004. p153-162
[illegible] 25-27 [illegible] 2002
[illegible] 88 [illegible] x + 308 p
[illegible] 16

BaltCom Datawarehouse – Online data mining using MS Analysis Services

[illegible]

[illegible]
[illegible] 64-66, 2800 [illegible]
[illegible]

Abstract

[illegible] 2.0 [illegible] (*Gadus morhua*), [illegible] (*Clupea harengus*), [illegible] (*Sprattus sprattus*) [illegible] (*Platichthys flesus*)] [illegible]

[illegible]

Introduction and requirement description

[illegible]

[illegible] 90's (2003), [illegible] (2001). A [illegible] (2003). [illegible]

[illegible]

About BaltCom

In 1995 [illegible]

[illegible] 1998 [illegible]

[illegible]

Input to fisheries assessment models and exploratory OLAP

[illegible]

- [illegible]
- [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Materials and methods

Selection of technology

[illegible]

[illegible]

Architecture and implementation

[illegible] (2003) [illegible]

[illegible] 90%. [illegible] *et al.* (2003).

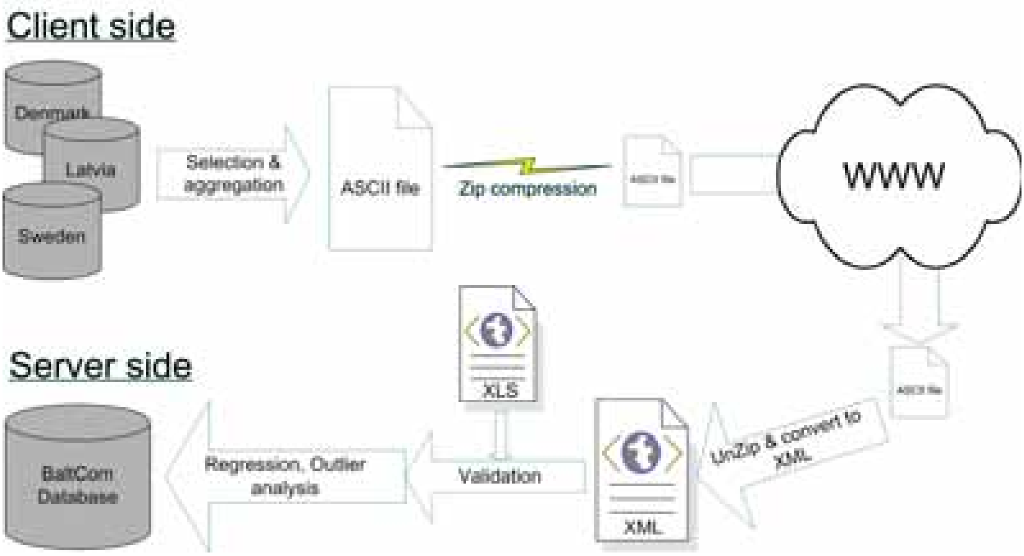


Fig. 1. Dataflow from user to database.

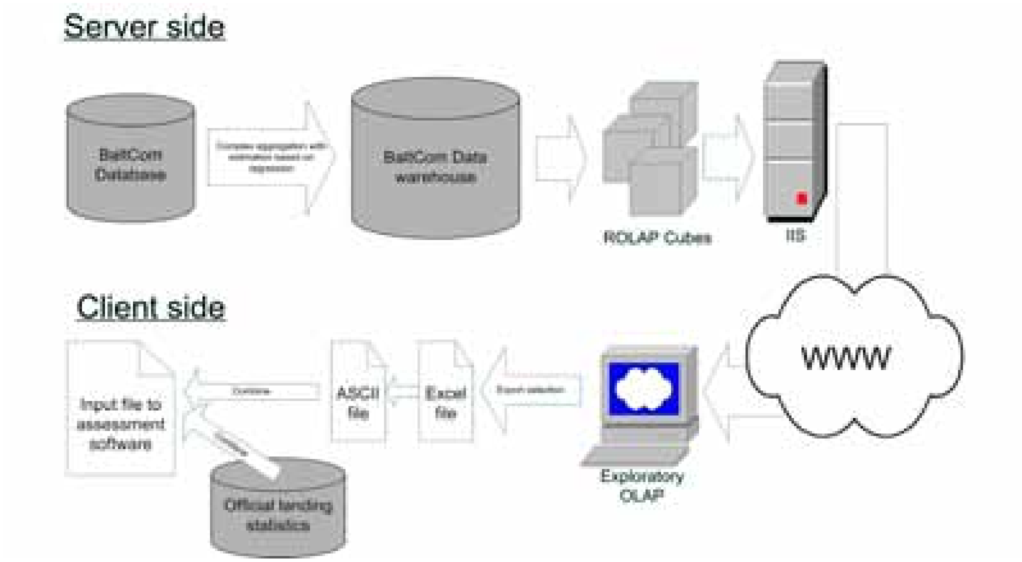


Fig. 2. Dataflow from database to user.

[illegible]

Results

[illegible]

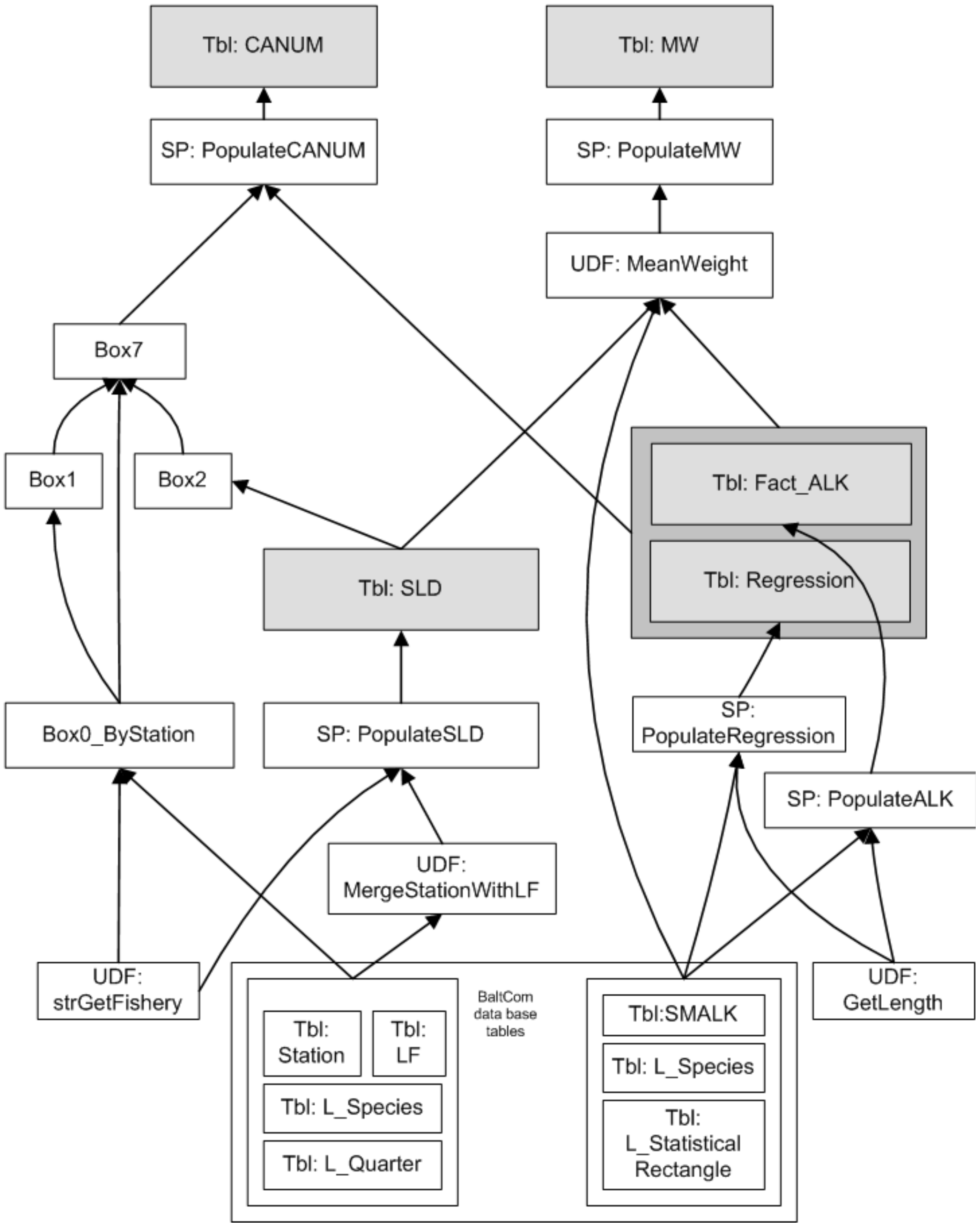
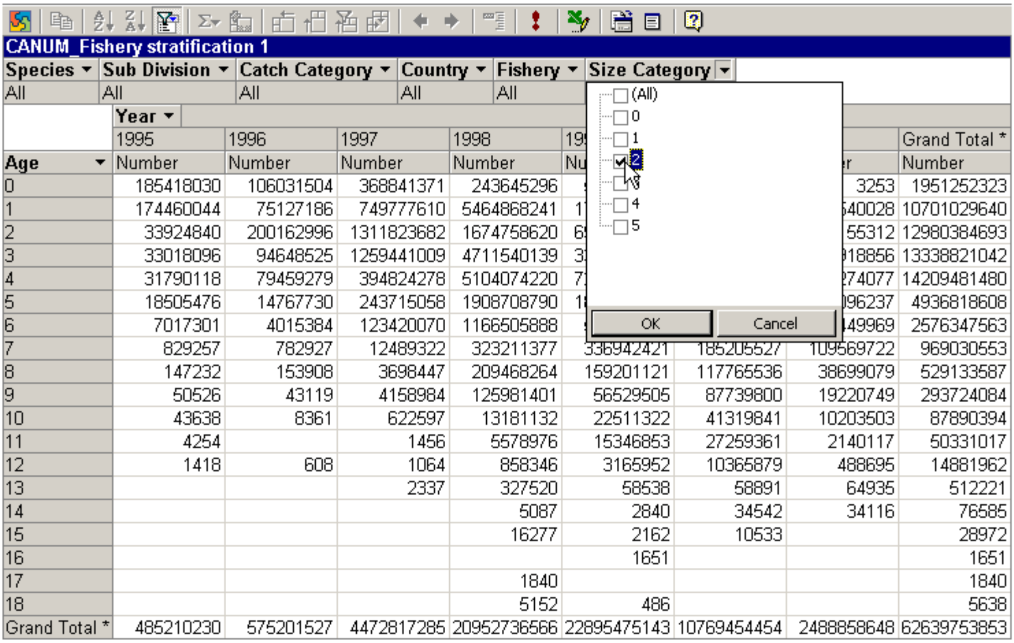


Fig. 3. SQL call stack populating datawarehouse tables. The SQL in each unit is documented in Degel and Jansen (2003). Tbl (in white box) = Database table, Tbl (in grey box) = Datawarehouse table, UDF = User Defined Function, SP = Stored Procedure.



Age	1995 Number	1996 Number	1997 Number	1998 Number	19[illegible] Nu[illegible]	[illegible]	[illegible]	Grand Total * Number
0	185418030	106031504	368841371	243645296	[illegible]	[illegible]	3253	1951252323
1	174460044	75127186	749777610	5464868241	[illegible]	[illegible]	[illegible]	10701029640
2	33924840	200162996	1311823682	1674758620	[illegible]	[illegible]	[illegible]	12980384693
3	33018096	94648525	1259441009	4711540139	[illegible]	[illegible]	[illegible]	13338821042
4	31790118	79459279	394824278	5104074220	[illegible]	[illegible]	[illegible]	14209481480
5	18505476	14767730	243715058	1908708790	[illegible]	[illegible]	[illegible]	4936818608
6	7017301	4015384	123420070	1166505888	[illegible]	[illegible]	[illegible]	2576347563
7	829257	782927	12489322	323211377	336942421	185205527	109569722	969030553
8	147232	153908	3698447	209468264	159201121	117765536	38699079	529133587
9	50526	43119	4158984	125981401	56529505	87739800	19220749	293724084
10	43638	8361	622597	13181132	22511322	41319841	10203503	87890394
11	4254		1456	5578976	15346853	27259361	2140117	50331017
12	1418	608	1064	858346	3165952	10365879	488695	14881962
13			2337	327520	58538	58891	64935	512221
14				5087	2840	34542	34116	76585
15				16277	2162	10533		28972
16					1651			1651
17				1840				1840
18				5152	486			5638
Grand Total *	485210230	575201527	4472817285	20952736566	22895475143	10769454454	2488858648	62639753853

Fig. 4. Pivot table with CANUM data for fisheries stratification 1.

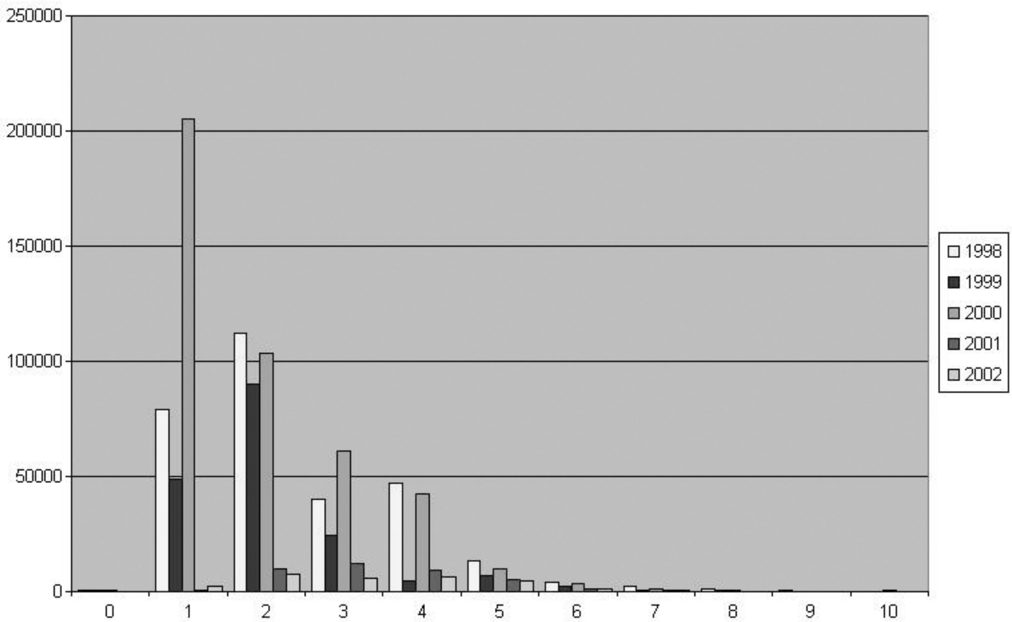


Fig. 5 Chart with CANUM data for fisheries stratification 1. Age in years on X-axis and number of fish per landed ton.

Discussion

Selection of technology

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Conclusions and recommendations

[illegible]

References

[illegible] 2003. [illegible]
[illegible] 2003. [illegible] 15p
[illegible] 2002. [illegible]
[illegible] 35. [illegible]
[illegible] 2001. [illegible] 550p
[illegible] 2002. [illegible]
[illegible]
96/002.
[illegible]
98/024.
[illegible] 2002. [illegible]
[illegible] 2002a [illegible]
[illegible] 2002c [illegible]
[illegible] 2003. [illegible] $$$
[illegible] 1998. [illegible]
[illegible] 306/1 [illegible] 2. [illegible] 2-5-103996-8. [illegible]
[illegible] 449 [illegible] 4490 [illegible]
[illegible] 2002. [illegible]
[illegible] 2002. [illegible]

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 163-175
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

MarineXML – using XML technology for marine data interoperability

Keiran Millard[1]

HR Wallingford Ltd
Howbery Park, Wallingford, UK OX10 8BA, Great Britain
E-mail: k.millard@hrwallingford.co.uk

Abstract

MarineXML is a project part-funded by the European Commission Fifth Framework Programme to demonstrate that eXtensible Mark-Up Language (XML) technology can be used to develop a framework that improves the interoperability of data in support of marine observing systems. With the advent of XML, the global oceanographic community has available an opportunity to create a truly universal marine data standard. There are already many XML based developments in the marine sector, however to date there has been little coordination. As a global standardisation initiative, XML can provide an evolutionary path for existing data formats and information systems while providing maximum benefits when used as a toolkit for the development of new systems.

By linking with related projects and initiatives, this project has the aim to develop a prototype of an XML-based Marine Mark-Up Language (MML) to show the integration between MML and data supported by other established standards. These include the International Hydrographic Organisation (IHO) S-57 standard, the OpenGIS Consortium's (OGC) Geographic Mark-Up Language (GML) standard and proprietary data formats such as those from marine instruments such as ADCP, expendable bathythermographs and ARGO floats, etc. It will specifically demonstrate how a MML approach supports data interoperability, widens data re-use and improves end-to-end data management in marine observing systems.

This paper presents the rationale behind the MarineXML project and indicates the tasks to be undertaken during the 24 months of the project beginning in February 2003.

Keywords: XML; Ocean data; Interoperability; European Commission.

Introduction

The huge diversity of data formats, proprietary data management systems, analysis packages, numerical models and visualization tools complicate the processing, management and accessibility of marine data. This diversity severely limits the multiple re-use of data and reduces our access to the data. Additionally the present complexity of accessing and integrating

---

[1] On behalf of the partners of the European Commission MarineXML project (www.marinexml.net)

data makes it tedious, often labour intensive to generate knowledge of marine processes and risks associated with marine ventures.

Today, the marine community is in a situation where large volumes of oceanographic data are collected by an increasing number of agencies and scientists for an increasing number of purposes. The desire to build regional or global databases, aggregating data from multiple sources is also increasing. The need for a common data framework to enable this integration has now become an essential component of building our knowledge of the marine environment.

Pressure is further being added with the rapid implementation of the Global Ocean Observing System (GOOS) and the various Committees and bodies overseeing its implementation, such as the IOC and WMO Joint Commission on Oceanography and Marine Meteorology (JCOMM). These bodies are building significant systems and capabilities to support near real-time users of marine data and information. The present lack of interoperability between data collection, processing and dissemination systems will add significant cost and complexity to these endeavours reducing potential economic, scientific and social benefits. The introduction of a well supported, unifying data framework will assist the marine community in general and the global observing programs in particular in overcoming these present difficulties and expensive restrictions.

Fortunately we are at a point in the development of information technologies where others have created an environment that can now be adapted to meet the interoperability needs of the ocean community. The potential of XML to support multiple portrayals/representations of a single data instance, provide automated processing and improve data exchange has been widely recognised by a number of industries, scientific disciplines, government agencies and the IT industry. For example the Geographic Mark-Up language (GML) developed by the Open GIS Consortium (OGC) for the use and exchange of geographic data or the Bioinfomatic Sequence Mark-Up Language (BSML) developed for the exchange of data related to the human genome project.

The success and acceptance of XML across other disciplines has prompted the International Oceanographic Data Exchange (IODE) Committee of the International Oceanographic Commission (IOC) to propose the development of an XML interoperability framework based on a 'Marine Mark-Up Language' (MML) specification (www.marineXML.net). The development of a MML specification will be a considerable benefit to the marine community, supporting the objectives of data interoperability and providing a framework on which:

- data providers can operate without changing their own (existing) data structures and systems
- application developers can build common programs and instruments that support multiple user formats and data standards and
- users can access and fuse heterogeneous data sets and access processing tools across the Internet or on local networks.

Similar to the way that GML enables the same geographic concepts to be re-used in a GIS, mobile phone or in-car navigation system, MML could enable marine data (e.g. temperature, wave height) from any source to be used in a mathematical model, navigation system or webpage. This also opens the possibility for data to be used more readily in mobile devices such as cell phones. The MML concept is shown below in Fig. 1.

This paper presents the rationale behind the MarineXML project. The following two chapters explain the issues to be considered in achieving marine data interoperability, the various initiatives and technologies that can be used as a basis for developing a marine mark-up language. Subsequent chapters detail the approach to be used within MarineXML during the twenty-four months of the project beginning in January 2003.

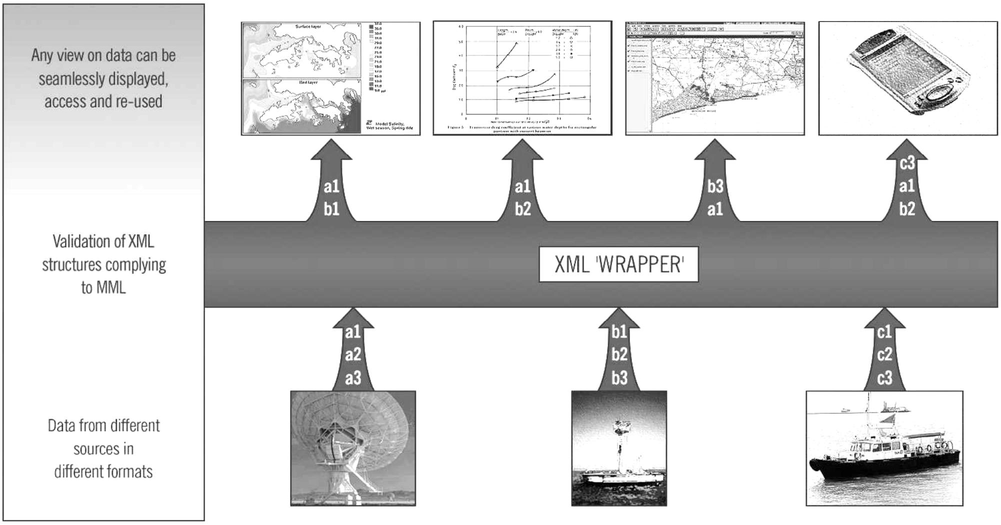


Fig. 1. MarineXML concept.

Issues in achieving data interoperability

Closed and open processing environments

Interoperability requires that relevant sources of data are accurately and easily identifiable, and messages exchanged between the sources and the requestor of information are understood. There are different situations in which interoperability is provided. These can be characterised by two dimensions related to types of processing environments and interoperability rules:

- predefined/closed versus open processing environment and
- closed/proprietary interoperability rules *versus* open interoperability standards.

The two dimensions are closely related. In a predefined/closed processing environment, all the known parties have a chance to agree on a set of closed/proprietary rules and build interoperability around them. On the other hand, in an open processing environment in which parties come and go, closed/proprietary rules are not appropriate because any agreement with an

initial set of players might not be suitable for those who join in later. Also, it could be expected that organisations would be reluctant to invest in a solution that has only limited applicability.
As a matter of fact, if one decided to adopt closed/proprietary interoperability rules for an open processing environment, that environment would eventually become a predefined/closed one as it would be restricted to those who decided to invest in implementation of the closed/proprietary interoperability rules.

Early integration dealt with predefined/closed processing environments and predetermined sets of data sources. So, there was no problem with identification of relevant sources. However, establishing efficient and effective data interchange was an important issue. At first, closed/proprietary integration rules were used and they sufficiently supported predefined/closed environments with a relatively small number of parties involved. In this situation introduction of any new player into the processing environment was costly and time consuming.
The first important open interoperability standard that was introduced independent of any particular application domain and IT (Information Technology) product or supplier was the character-coding standard ASCII. This standard ensured that characters in exchanged messages could be properly read and identified. A clear advantage of subscribing to this open interoperability standard was that no translation of character encoding was no longer necessary and character-handling software could be standardised, shared and reused. Indeed subscribing to the ASCII standard also enabled the orderly evolution to Unicode to meet the needs of the global community to handle non-Latin characters.

XML and data interoperability

Even with open interoperability at the character level, the content of the messages still has to be agreed on by all the parties involved in the data interchange. Publication of the XML standard by the World Wide Web Consortium (W3C, www.w3.org) introduced a standard grammar for creating data structures. With XML, standard parsers could easily decode the structure of any message. This still required an agreement on the data structure itself.

The agreements on structures are articulated in sets of DTDs (Document Type Definitions) or more recently in XML Schemas, which provide data models for data interchange. In recognition of the inevitability of the move into open environments many industries started working on common structures. This effort is co-ordinated by OASIS, the Organization for the Advancement of Structured Information Standards, which is an international body for coordinating international interoperability standards. Currently there are 21 standing OASIS committees working on industry specific standards and rules of conformance of structured information standards for interoperability.

Development of interoperability standards always involves selection of concepts (XML tags) which carry specific meaning (semantics). It is unfortunate that the underlying conceptualisation has to be implicit (or 'hidden') as there are no standard ways of representing semantics of XML tags. This is not a real issue when new interoperability standards are created for emerging applications like PICS for content selection or P3P for security and privacy, or well established applications with already commonly accepted data representation standards like MARC for library records. However, when there is a community of players who have already adopted and invested in different closed/proprietary interoperability rules, such as the oceanographic community, this approach is not sufficient. In this case there is a need for explicit semantics of interchange data models with well-defined mechanisms of mapping between the interchange models and closed/proprietary formats.

In oceanographic data interchange, there is a strong trend to move towards open interoperability rules based on XML technologies. But over many years most of the oceanographic data management organisations have accumulated data under closed/proprietary interoperability rules as they operated in many loosely connected predefined/closed processing environments. For these organisations moving away from closed/proprietary sets of interoperability rules is not an option as it would entail a massive conversion effort. Instead they should be provided with an opportunity to subscribe to open interoperability standards with well-defined explicit semantics with mapping mechanisms to the existing formats. It is therefore necessary to supplement the existing XML technologies with an open standard for defining data model semantics.

Adding semantics for interoperability

An open processing environment requires a mechanism for accurate identification of relevant data sources. Metadata or 'information about data' is commonly used for this purpose. XML based approaches to metadata specification are best exemplified by Dublin Core standard for general-purpose content definition. This small set of content descriptors gained popularity because of its relevance to many application domains and their simplicity. A more ambitious project for describing data sources is RDF (Resource Description Format). In this case resource definitions can unfold as a network of resource descriptors that can be either URIs (Universal Resource Identifier) or atomic values. But similarly to XML, both standards like Dublin Core and RDF lack explicit semantics that can be mapped against the existing metadata formats. Here again the marine community, in common with other thematic communities, is disadvantaged as different community members already have their own standards that cannot be abandoned.

In order to facilitate the interoperability of oceanographic data it is proposed that existing standards for data management be mapped against an ontology (or set of ontologies) which describes marine data at the *semantic level.* Ontologies are widely accepted as a method of knowledge sharing and seek to describe a conceptual model for the common terms and relationships involved in the domain. Only by reference to agreed common ontologies capturing the intended meaning of the data can effective interchange between data held by different independent agencies be made possible; by making ontologies explicit, formal and machine readable, a precise mapping can be developed and generic tools can be applied. Such common vocabulary and relationships can then be used as a basis for resolving queries on a semantic basis to provide more precision in search, and more trust in the results.

One of the main concerns of the Semantic Web activity (http://www.w3.org/2001/sw/) is to define ontologies in a Web accessible fashion, which enables decentralisation. Then each community can define their own ontological structures, but have sufficient commonality to communicate with others in an open and egalitarian manner. To this end, the W3C has developed RDF and OWL in order to represent semantic and ontological structures on top of the base interchange syntax provided by XML. Such an approach is well suited to the oceanographic domain that has to accommodate the legacy of a wide variety of existing formats.

Several approaches to interoperability have been proposed. A single ontology for the whole domain is perhaps the most straightforward; each legacy format will then be interpreted in this ontology. However, this is inflexible and likely to lead to a cumbersome ontology that has to accommodate all variations. An alternative approach would be allow each legacy format to provide its own ontology and allow a mapping to others; this is good for decentralisation, but without a reference ontology, leads to a proliferation of mappings and a serious problem of maintenance. A hybrid approach is to take some common components, either core fields (such as Dublin Core) and/or common vocabulary, and use this as a basis for common communication, with refinement of concepts following a match.

A logical approach for the marine community would be to take a hybrid approach that identifies:
- a common vocabulary for marine terms, within a glossary;
- a base ontology, with individual ontologies for different formats explicitly related to that base ontology;
- a common XML data format based on the common vocabulary and glossary for interchange between data sources, and for common interchange to the user community.

Mappings can then be produced to convert legacy formats into a marine XML; with the common vocabulary and domain ontologies in place, such a mapping should be straightforward.

Existing architectures supporting data interoperability

Whilst standardisation of a MML will take many years, developing a prototype can be achieved within the scope of a shorter project. Much of the basic research needed to develop a prototype MML already exists and through the integration of this existing work a prototype can be realised. These key areas of research are summarised in this chapter.

Grid technologies

The real-time monitoring of oceanographic observation systems requires the identification of a variety of monitoring resources (buoys, ships, shore stations, satellites) which can provide data for the environmental factors of interest. Such resources may not be administered by the same authority and may not provide the same data in the same manner. Nevertheless, they will need to be co-ordinated remotely and in real-time. Consider for example the case where a known instrument fails; the monitoring agent will need to seek out an alternative source and negotiate access to its data, including negotiating the meaning of the data, adapting the monitoring agent to the granularity and tolerances of the new source, all within real-time. Such a situation with access to large-scale data sources, and computing resources, with negotiated access to resources administered by other agencies, is akin to the Grid computing architecture, currently the subject of development programmes within the USA and Europe.

Ontological searching

There is on-going research into the use of ontologies for advanced data searching. For example, the EC Arion (www.arion-dl.org) project is currently developing an architecture to support the searching and retrieval of digital scientific collections that reside within research and consultancy organisations. This project is concentrating on providing generic tools to support the metadata-mediated search and retrieval of scientific data sets using ontologies. It is also using coastal zone management as a test application area. A further project which is relevant here is the Data Portal Project (http://www.escience.clrc.ac.uk/Activity/ACTIVITY=DataPortal) which is taking a similar approach to developing web and grid based approaches to accessing scientific data sets using a generic scientific ontology which can be specialised to specific domains. Existing systems have also been developed to support the use of controlled vocabularies within the Semantic Web framework. For example, the EC project Limber has been developing formats for representing multilingual thesauri using RDF. More generally, the

International Standards Organisation (ISO) standard on Topic Maps (www.topicmaps.org) provides a generic framework for representing vocabularies and relationships between them.

Geographic Mark-Up Language (GML)

The Geographic Markup Language (GML) is a development of the Open Geographic Information System Consortium, OGC (http://www.opengis.org). GML is the technology engine behind the geo-spatial Internet. The geo-spatial web is a global information infrastructure, integrating sources of geographic information using hyperlinks, and providing layered application services based on this infrastructure. The geo-spatial Web offers global integration of geographic information while supporting local data development and maintenance. Using GML, an XML 'dialect' developed for handling geographic data, you can deliver geographic information as distinct features, and then control how they are displayed. This approach is much more versatile and powerful than the conventional image-based methods using ‘flat’ data structures. GML is an emerging international standard; the current version (v3.0 ISO workitem 19136) of the specification formally bears the status of ‘adopted specification’ within the OpenGIS Consortium (OGC); this is the top rung of the OGC specification ladder. Basically the designation implies that GML is mature enough to be used in other implementation activities or incorporated into software products. GML 3.0 offers significant enhancements over GML 2.0 in a number of areas: topology, an enriched geometry model, support for coverages, temporal constructs, units of measure, and more.

eXtensible Mining Mark-Up Language (XMML)

The eXtensible Mining Mark-Up Language (XMML), developed to provide interoperability between data collected for mining drill (or bore) holes. The type of data suported by XMML is not marine, however, it does have very similar characteristics to ocean profile data and could be used to support ocean temperature and salinity (and other parameter) profile measurements.

Extensible Scientific Interchange Language (XSIL)

The Extensible Scientific Interchange Language (XSIL) is a flexible, hierarchical, extensible, transport language for scientific data objects, which has been developed by the Center for Advanced Computing Research at CalTech (www.cacr.caltech.edu/SDA/xsil/index.html). It is being used to support the Laser Interferometer Gravitational-Wave Observatory (LIGO), as large scale scientific programme in the USA. XSIL provides a set of simple elements for capturing generic scientific data sets and their parameters, including tables and arrays. It also allows a degree of metadata to describe the context of the data; this is extensible. XSIL also comes with a set of Java based tools for browsing and visualising XSIL data sets.

All XSIL objects may have Name and Type attributes; these are shown in the XSIL browser, and are used by the API (Application Programme Interface) to locate particular objects. For example: ‘Find the object of type Fruit with the name Apple’. There is a container object in XSIL, with the XSIL tag, that may have Name and Type attributes. The container may contain other containers, so that this object induces a hierarchy. The elementary objects of XSIL include Param, an association between a keyword and a value; an Array, similar to the array concept of Fortran or C; a Table, which is a set of column headings followed by a set of records.

Coastal Mark-Up Language (CML)

This is an on-going research project at Oregon State University, USA. The project looks to extend metadata in XML format that complies with the FGDC (Federal Geographic Data Committee) standard to construct interactive maps for data browsing and analysis. ESRI's ArcIMS product uses a form of XML called ArcXML which is an early XML implementation of the Geography Markup Language with some proprietary semantics. ArcXML can be used in the data publication process to specify legend and labelling conventions which will 'stick' to a data set such that it is only ever viewable the way that the original author intended. A further part of the CML initiative is the use of XML in the map images themselves. Current ArcIMS applications typically generate a map and then send a rasterized (GIF or JPEG) 'snapshot' of it to the client browser for viewing. Incorporating another XMLbased standard, Scalable Vector Graphics (SVG), graphic quality can be improved, but more importantly, it will increase the responsiveness of interactive maps to requests for zooming, panning and encapsulating animation.

Oregon State University proposes to develop a Coastal XML (CML) which will be the superset that incorporates the definitions for all of the coastal information types being housed by the IMS (Internet Map Server). This will include the type of FGDC compliant metadata for geospatial data as well as narrative texts and static or dynamic imagery relating to specific setting types and associated problem types for locations within the coastal zone. Although CML does not fully explore the complexity of semantic operability across multi-parameter datasets, the integration of metadata in XML with map data in XML is potentially useful to MarineXML

IHO S-57

S-57 describes the standard to be used for the exchange of digital hydrographic data between national Hydrographic Offices and for the distribution of digital data and products to manufacturers, mariners, and other data users. The most significant digital product being delivered in the S-57 format is the Electronic Navigational Chart (ENC). S-57 was adopted as an official IHO standard by the XIVth International Hydrographic Conference, Monaco, 4-15 May 1992 and is also specified in the IMO Performance Standards for Electronic Chart Display and Information Systems (ECDIS) (IMO Resolution A.817(19)as amended). S-57 Edition 3.1 was released in November 2000 following a familiarization period of one year. The previous edition (3.0), was published in November 1996, and was frozen for a period of 4 years in order to facilitate ENC production and to provide stability for ECDIS manufacturers.

The S-57 data model takes an object-oriented approach toward the modelling of hydrographic data. Each real-world entity is described by an object that belongs to a certain class. Each object has a set of attributes associated with it. There are two types of objects: feature objects and spatial objects. Feature objects are used to describe real-world phenomena such as beacons, buoys, land areas, sea areas,, etc. Spatial objects are used to represent the location on earth surface such as a point (called isolated nodes in the S-57 data model) with their latitude and longitude co-ordinates.

A feature object will use a spatial object to indicate its location. Both kinds of objects have attributes associated with them. For example, a beacon object will have attributes such as colour. A spatial object, say an isolated node, will have attributes like positional accuracy. These two kinds of objects are the conceptual building blocks of a dataset (which is called a cell in the S-57 standard). Currently there are four types of spatial objects. They are isolated nodes, connected nodes, edges and faces.

Cozdis

COZDIS (Coastal Zone Display and Information System) is an enhanced Electronic Chart Display and Information System (ECDIS) that is used today to provide ENC functions. ECDIS has been developed for navigational purposes and is grounded on an international data transfer standard S-57 which supports most of the MIS requirements including an ability to accept additional objects necessary for new subjects, and handling of the temporal variability of the dynamic objects.

COZDIS with XML capability extensions can facilitate the re-use and transformation of data. S-57 has been developed to cover the official electronic navigational charts, but can be easily extended to oceanographic and coastal applications. Given its well-defined structure, S-57 is open to structuring in an XML framework, increasing the usability of the data.

Development of a Marine Mark-up Language

Objectives of the project

The previous chapters have illustrated that not only is technology sufficiently mature to develop a Marine Mark-Up Language, there is also widespread application of this technology in both marine and non-marine areas. This provides the firm basis on which the MarineXML project is based and accordingly defines the four key project objectives below.

Objective 1. To produce a prototype marine data ontology framework for interoperability
This task will align a conceptual Marine Mark-Up Language (MML) specification with other XML and related standards. This will be achieved through research into a marine data ontological framework.

Objective 2. To produce working demonstrations of the data interoperability framework
The project will develop a working prototype test bed of the interoperability framework in Objective 1 that will test and demonstrate the functionality and interoperability of heterogeneous data sets from disparate agencies. It is intended that COZDIS will be the basic data management platform of these demonstration prototypes.

Objective 3. To develop a prototype MML specification
Using the outputs of the demonstration in Objective 2, the interoperability framework will be further developed as the basis of the MML specification. While the project will not develop the full MML specification, it will produce and document an intial framework.

Objective 4. To advance the standardisation of a Marine Mark-up Language
MarineXML will seek to ensure the standardisation of a MML by liasing with the SGXML during the project and provide support and advice to further the standardisation process beyond the end of the project.

Project task overview

To achieve these objectives the MarineXML project is executed as six tasks (work packets) as shown in Fig. 2. The main technical work will be centred around work packets 3 to 6 and the content of these task is outlined in the following sections.

Standardisation review (WP3)

MarineXML will review the range of existing and emerging standards that are in operation within the marine domain. This includes transfer standards such as S-57, discovery metadata standards such as MEDI and proprietry data formats used for instruments such as ADCPs. The requirement for interoperability with satellite data in the marine context will also be considered, but taking into account the volume of data present in many satellite data products, and the need for compatibility with existing standards such as HDF, it will be considered whether only metadata will be stored in XML in such cases.

WP3 will also investigate XML standards that exist in other domains to determine the relevance to marine data. This includes Geographic Mark-Up Language (GML), the eXtensible Mining Mark-Up Language (XMML), the Bioinformatic Sequence Mark-Up Language (BSML) and the eXstensible Scientific Interchange Language (XSIL). MarineXML will also link with the teams in the USA developing a Coastal Mark-Up Language (CML) and a Marine Information Mark-Up Language (MIML).

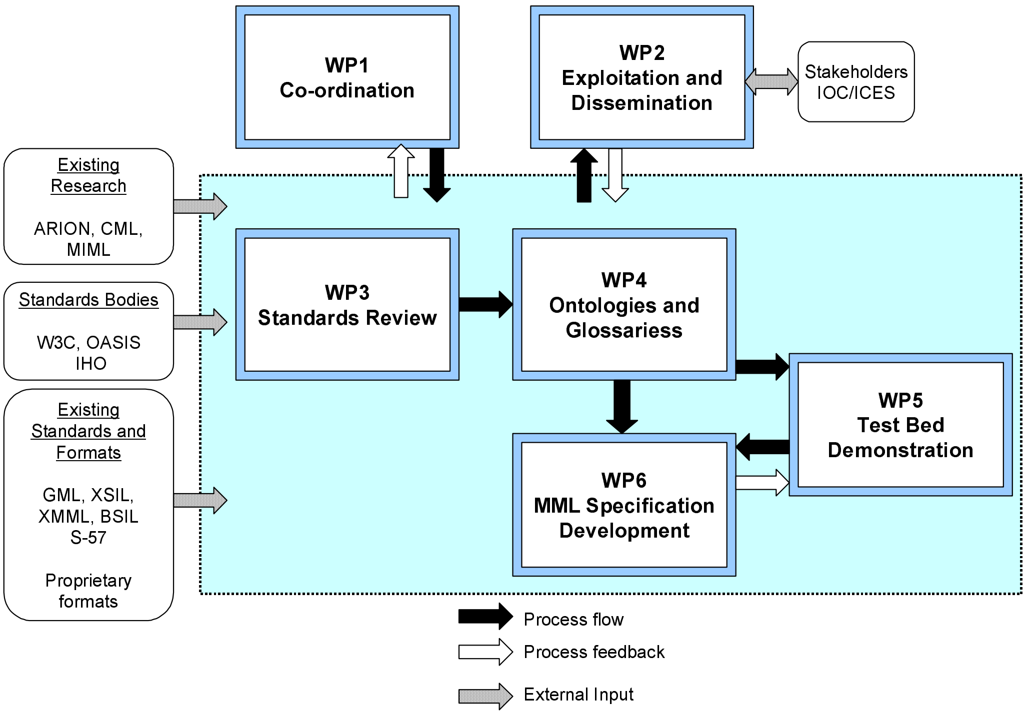


Fig. 2. Project PERT chart.

A non-formal descriptive analysis of interoperability between these standards will be conducted to form the basis for defining the ontologies in WP4. During the project, developments are regularly synchronised with related standards developments through a six month review with relevant standards bodies.

Ontologies and glossaries (WP4)
With these identified standards, the next step in the project is to explicitly document their relationships to each other and the mechanisms necessary to provide interoperability between them and the evolving MML framework. This will be accomplished by the introduction of ontological structures and significant efforts will be placed on developing appropriate and extensible structures within an ontological framework with associated glossary. Working with the EC ARION project (www.arion-dl.org), research will be conducted into the appropriate ontological structures suitable to address the complexity of data interoperability in the marine domain. Based on these ontological structures, a generic interoperability framework model will be created at a high level of abstraction to ensure broad applicability.

MarineXML will use existing glossaries from NASA (Global Change Master Directory), the British Oceanographic Data Centre's variable list, ASFA (Aquatic Sciences and Fisheries Abstract) glossary and GEMET (General Multilingual Environmental Thesaurus) used by the European Environment Agency. Other list of terms may also be identified. The aim is to determine an appropriate structure of glossary and level of detail to provide a semantic basis to a MML and the integration of the glossary into XML/RDF. This glossary will not be definitive, but sufficient to demonstrate the MML concept. It is expected that this glossary will provide detailed information structured hierachically on: Topic, Term, Variable, and Detailed Variable. Recommendations will be made as to how the glossary should be extended for future development of a MML.

Research will be undertaken to apply practically these theoretical ontological data structures to the real world data streams that will be used in the test bed applications. This will necessitate the creation of an initial MML ontological model based on the semantic structures arising from the marine glossary development. The glossary will be integrated into the generic ontological model to develop a skeleton version of the MML sufficiently complete to develop the test bed applications.

Marine Mark-up Language test-bed demonstration (WP5)
Once a generic framework has been created and it has been refined towards the marine requirements, the initial structure of the MML specification will be built. A focus will be made on specific structures needed to support the data to be used in two test bed applications. Additionally, as the test bed applications are being developed there will be a continuous feed back mechanism established to 'test' the validity of the initial MML specification in this real world situation.

The test bed applications will consist of a data flow comprising a mixture of physical, chemical and biological data coming from a range of in-situ and remote sensing sources. This will include such data collected by Argo floats, ADCP, expendable bathythermographs, drifters, etc. These scenarios will require a high level of data integration and fusion. Output data sets, and analysis products will be determined based on identified GOOS product needs.

The existing COZDIS capability will be modified to support XML based (MML) data structures. Based on the COZDIS platform, additional applications will be developed to support the automated input of data from the test-bed data streams. These applications will accept MML structured 'observations', provide basic quality assurance checking and validation (automated) and input the data into the MML enhanced COZDIS platform. Applications will also be developed to support the extraction of data in suitable formats for input into other (external to COZDIS) analytical, processing, display and modelling processes. Many of these processing applications already exist, but have very specific input data structure requirements. These export applications will use the XML based eXtensible Stylesheet Language Transformation (XSLT) to automatically provide the extraction (from COZDIS) and 'reformatting' process.

Marine Mark-Up Language specification development (WP6)
The ontological structure developed in WP 4 will be used as the basis of the Marine Mark-up Language. This WP will support the coding of the initial MML specification with a focus on the requirements of a specification to support the test bed scenario applications. The structures and application will be populated with identified content (as available in the test-bed), and an iterative testing and feedback process will allow the consortium to test and fine-tune the ontological framework. The prototype MML will be documented and published. This will include a formal description of the mark-up language and the test-bed applications

Conclusions

The political and technological climate is right for the development of global standards for marine data interoperability. These standards will realise large operational and scientific benefits to the marine community in terms of cost-savings and enhanced knowledge generation. MarineXML is a project part-funded by the European Commission Fifth Framework Programme to demonstrate that eXtensible Mark-Up Language (XML) technology can be used to develop a framework that improves the interoperability of data in support of marine information systems. The project runs for 24 months from February 2003.

The purpose of this paper has been to give a clear view of the objectives of MarineXML and an insight into how it proposes to achieve these objectives. As international interest grows in the application of XML to the marine environment duplication of effort and the development of competing camps becomes a real danger. By making explicit at an early stage what this project is setting out to achieve, it is hoped that links to related initiatives can be formed to prevent such dangers.

This paper has presented a comprehensive overview of the state of the art for XML in marine applications and much of this formed the basis of a project proposal to the European Commission. Since then, the advancement of the IOC/ICES SGXML has identified further work in this area. MarineXML will as a matter of principle work as closely as possible with SGXML towards the common aim of developing a Marine-Mark-Up Language. For this reason all work on developing a MML is communicated through the common 'MarineXML portal at www.MarineXML.net.

Acknowledgements

Much of the source text for this paper came from the proposal submitted for funding to the European Commission in April 2002 by the MarineXML team. Accordingly I would like to acknowledge all members of the MarineXML team for their inputs. The team is :

Keiran Millard	HR Wallingford
Jane Smallman	HR Wallingford
Jim Clipson	United Kingdom Marine Information Council
Dan Pillich	SevenCs
Torill Hamre	Nansen Environmental and Remote Sensing Centre
Stein Sandven	Nansen Environmental and Remote Sensing Centre
Juan Bicarregui	Central Laboratory of the Research Councils
Brian Matthews	Central Laboratory of the Research Councils
Pieter Harring	National Institute for Coastal and Marine Management
Edward Vanden Berghe	Flanders Marine Institute
Rob Atkinson	Social Change Online
Simon Cox	CSIRO
Hans Dahlin	EuroGOOS
Nic Flemming	EuroGOOS
Peter Pissierssens	International Oceanographic Data Exchange Committee
Greg Reed	International Oceanographic Data Exchange Committee

Special thanks go to Ben Searle who, whilst at AODC, contributed much enthusiasm and virtually all the text to the foundation drafts of this project proposal. MarineXML also acknowledges the funding provided by the European Commission under the Fifth Framework Programme to support this project.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 177-185
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

New internet developments: marine XML

Greg Reed and Peter Pissierssens

Intergovernmental Oceanographic Commission (IOC) of UNESCO
1 rue Miollis, F-75732 Paris Cedex 15, France
E-mail: g.reed@unesco.org

Abstract

With the emergence of XML as a data transfer protocol a mechanism to support available common standards across a range of marine data exchange situations is now available. The development of a Marine XML will support the tracking of data from collection through to the generation of integrated global and regional datasets. XML can support the metadata describing the data collection, quality control and subsequent processing. The IOC, through its IODE programme, is involved in the development of Marine XML applications. One of the tasks of the IODE Group of Experts on the Technical Aspects of Data Exchange (GE-TADE) is to develop Marine XML as a mechanism to facilitate format and platform independent information, metadata and data exchange. In addition the IOC/IODE metadata system, MEDI, uses XML to validate and transfer metadata. The IOC/IODE also hosts a 'community portal' web site for Marine XML (http://www.marinexml.net) that provides a central location for document distribution and general discussion forum. Cooperation between IOC and ICES on the development of a Marine XML commenced in 2002 with the creation of the ICES-IOC Study Group on the Development of Marine Data Exchange Systems using XML (SGXML). This Group met for the first time in April 2002 and developed a work plan to guide the investigations into how XML technology could be used in an oceanographic context. Another project involving IOC/IODE is the EU Marine XML project, 'MarineXML: a pre-standardization development for marine data interoperability using XML', will demonstrate that XML technology can be used to develop a framework that improves the interoperability of data for the marine community and specifically in support of marine observing systems.The project will develop a prototype of an XML-based Marine Mark-up Language (MML). In another development, the XML 'brick' concept developed by MEDS Canada, employs a small number of generally defined structures that can be assembled in different ways to reflect the structures of a variety of data collected over a broad spectrum of disciplines. This brick structure represents a basic building block for packaging data and metadata.

Keywords: XML; Ocean; Data transfer; Interoperability; IODE.

Introduction

The eXtensible Markup Language (XML) was developed by the World Wide Web Consortium (W3C) to improve data transfer over the Internet. With the emergence of XML as a data transfer protocol, the global oceanographic community now has available a mechanism to support the exchange of marine data. Unlike earlier standardization initiatives, XML can support existing data formats and information systems while providing maximum benefits when included in the development of new systems. Thus existing investments in instruments, computing systems and staff training are protected, and all users are able to benefit substantially from the standardization and interoperability of data that resides in such a framework. XML changes the way data moves across networks, it encapsulates data inside custom tags that carry semantic information about the data. A Marine XML structure will encapsulate marine data and provide an efficient means to store, transfer and display marine data.

There are a number of reasons for using Marine XML:

- Exchange of data. A major strength and source of potential of XML is that it facilitates the exchange of data between different applications and operating systems. One of XML's strongest points is its ability to provide data interchange. Because different organizations (or even different parts of the same organization) rarely standardize on a single set of tools, it takes a significant amount of work for two groups to communicate. XML provides a clean separation of structure and content, allowing XML implementations to easily recreate the intended structures from an exchanged document. XML is potentially the answer for oceanographic data exchange, as long as all sides agree on the markup to use.
- Extensibility. Extensible means that it is not a fixed format like HTML. While HTML tags must follow pre-set standards, new XML tags can be created by anyone at any time. XML will allow groups of people or organizations to create their own customized markup languages for exchanging information in their domain. Examples of existing industry-specific XML include music, chemistry, electronics, linguistics, engineering and mathematics.
- Plain txt. Since XML is not a binary format, files can be created and edited with a standard text editor making it useful for storing small amounts of data. At the other end of the spectrum, an XML front end to a database makes it possible to efficiently store large amounts of XML data. XML provides scalability for anything from small configuration files to an industry-wide data repository.
- Data identification. The XML standard specifies how to identify data, not how to display it. HTML, on the other hand, describes how things should be displayed without identifying the content. Because the different parts of the information have been identified, they can be used in different ways by different applications.
- Stylability. XML allows data content to be transmitted without assuming a particular display, but provides convenient mechanisms to transfer portrayal rules. Since XML is inherently style-free, different style sheets can be used to produce output in postscript, PDF, or any other format.
- Hierarchical. XML documents are hierarchical in structure. Hierarchical document structures are, in general, faster to access because you can drill down to the part you need, like stepping through a table of contents.

At the Sixteenth Session of the IOC Committee on International Oceanographic Data and Information Exchange (IODE) in November 2000 it was proposed to develop an XML interoperability framework based on a ‘Marine Mark-Up Language’ (MML) specification. The development of MML interoperability framework will have considerable benefits to the oceanographic community, supporting the objectives of data interoperability and providing a framework on which:

- data providers can operate within the MML framework using their own (existing) data structures and systems,
- application developers can build common programs that support multiple user formats and data standards, and
- users can access and fuse heterogeneous data sets and access processing tools across the Internet or on local networks.

A number of international initiatives are currently in progress which will contribute to the development of the MML framework. These include:

- the ICES/IOC SG-XML
- XML 'Brick' concept
- the EU Marine XML Project
- the MEDI metadata authoring tool
- GE-TADE activities
- establishment of a MarineXML community portal site

Each of these initiatives will be discussed here.

Study group on the development of marine data exchange systems using XML

The Intergovernmental Oceanographic Commission (IOC) and the International Council for the Exploration of the Sea (ICES) have jointly established the Study Group on the Development of Marine Data Exchange Systems Using XML (SG-XML). The Terms of Reference for the Group are:

a) develop a framework and methodology for the use of XML in marine data exchange in close consultation with IOC and the Marine XML Consortium;
b) develop a workplan that within 4 years will lead to published protocols for XML use in the marine community;
c) explore how to best define XML tags and structures so that many ocean data types can be represented using a common set of tags and structures;
d) test and refine these common tags and structures using designated case studies i.e.:
   ii) Point (physical/chemical) data (profile, underway, water sample);
   iii) Metadata (cruise information, building from the ROSCOP/Cruise Summary Report);
   iiii) Marine biology data (integrated tows (e.g., zooplankton-phytoplankton tows), demonstrate the use of taxonomy).

The first meeting of the Group, in April 2002, resulted in the initial development of a plan to guide an investigation into how XML technology might best be used in an oceanographic context. From an IOC/IODE perspective, the requirement was to design a framework for an XML structure that data centres can use. An action plan was developed for the intersessional period to investigate the following issues:

- Parameter Dictionary. Create, evaluate and discuss intersessional work on SGXML parameter dictionary including the population of the dictionary for distribution via a defined XML structure. Complete the XML web distribution of the parameter dictionaries and determine the applicability of the XML structure for other dictionaries.
- Point Data Investigation. Evaluate the usefulness of the generalized Keeley brick approach with application to various point data types. Evaluate the point data structure from the perspective of the IODE data centres. Provide biological and taxonomic input to the Keeley bricks.
- Metadata Investigation. Evaluate the usefulness of linkages to other metadata standards and on the implications of a generalized metadata model to existing models. Define a general

metadata model that will include the definitions of EDMED, MEDI and CSR.

XML bricks

The XML 'brick' concept, developed by R. Keeley of MEDS, employs a relatively small number of generally defined structures that can be assembled in different ways to reflect the structures of a variety of data collected over a broad spectrum of disciplines. This brick structure represents a basic building block for packaging data and metadata.

Bricks can describe data units, measured (or computed) variables, space and time location of the measurement and data quality flags. In addition, bricks can also describe the supporting information such as sampling techniques, species taxonomy, etc. The current working list of bricks is as follows:

Brick	**Definition**
Analysis_method	Information about the analysis employed
Archive_information	Information attached by the recipient of the data
Availability	A marker to control access to the data
Calibration	Information about instrument calibration
Comment	General comments on the data
Data_dictionary	Used to identify the dictionary in use in the data file
Data_point	Contains a single data value and supporting information
History	Processing history of the data
Instrument	Information about the instrument used to make the measurements.
Location	This used to record the x, y, z, or t values that are fixed for a given set of observations.
Provenance	The originator of the data
Quality	A marker providing the assessment of data quality
Quality_testing	Information about how the data quality assessment was made
Record_id	An identifier used to distinguish data
Sample_property	Describes properties of sampling for contaminants
Sampling	Information about the sampling methods used
Sampling_method	Used to store details of the sampling
Sensor	Identifies sensor specifics
Storage	Storage techniques employed
Taxonomy	The full taxonomy of any life form from which a sample was taken
Units	The units of measurements
Variable	Information about the variables measured

As an example, the Provenance brick holds information about the originator of the data. The Provenance brick, where data is collected from a single source, would take the form:

```
provenance
    ├ originator
    ├ project
    ├ country
    ├ date_received
    ├ platform_name
    ├ agency
    ├ originator_identifier
    ├ data_grouping
    ├ description
    └ platform_type
```

The elements of the Provenance brick are:

Element	**Definition**
Originator	The name of the originator
Project	A general name that the dataset may be associated with. For example, the project name may be related to an international or national program such as WOCE, CLIVAR.
Country	The country of origin for the dataset
Date_received	The date that the record was created
Platform_name	The name of the platform serving for the data collection
Agency	The agency from which the data came. This should be as specific as possible.
Originator_identifier	The identifier used by the originator by which the data are identified
Data_grouping	A way to identify groupings of data. This can be used, for example, to identify stations that belong to sections.
Description	An open description of the dataset
Platform_type	The type of platform (e.g. moored buoy, ship, airplane, etc.)

The following example of the application of bricks to XML is based on a common oceanographic data collection strategy, the cruise. A cruise is made up of a number of stations at which one or more variables are measured over the water column. An XML file assembling these data might look something like this:

```
<collection>
        <comment/>                    - comments relevant to the entire collection
        <provenance/>                 - origins of the collection
        <variable/>                   - variables reported
        <units/>                      - units of measurement
        <instrument/>                 - instruments used
        <quality_testing/>            - description of QC undertaken
        <location/>                   - x,y,t location of the event
        <event>                       - first event
                <comment/>            - comments pertinent to the event
```

```
            <data_point/>               - value of the measured variable
                  <quality/>            - quality of the value
            <data_point/>
                  <quality/>
                  <comment/>            - comment appropriate to the data value
      </event
      <event/>
</collection>
```

This is a simple example where all of the information about variables, units, quality test procedures and so on apply to the entire collection. Using a basic set of bricks, it is possible to build an XML file in many different ways for a single collection of data. A set of guidelines has also been developed for the construction of bricks. Further details are provided on the Marine Environmental Data Service (MEDS) web page at http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/ICES/web%20xml/SSF-xml.htm.

EU marine XML project

The aim of the EU Marine XML project, *MarineXML: a pre-standardization development for marine data interoperability using XML*, is to demonstrate that XML technology can be used to develop a framework that improves the interoperability of data for the marine community and specifically in support of marine observing systems. The project will develop a prototype of an XML-based Marine Mark-Up Language (MML) to show the integration between MML and data supported by other established standards. These include the International Hydrographic Organization (IHO) S-57 standard, the OpenGIS Consortium's (OGC) Geographic Mark-Up Language (GML) standard and proprietary data formats such as those from marine instruments such as ADCP, expendable bathythermographs and Argo floats etc. It will specifically demonstrate how this MML approach supports data interoperability, widens data re-use and improves end-to-end data management in marine observing systems. This MarineXML is to be developed in partnership with international agencies, government departments and organizations responsible for data standards to ensure that the research meets the needs of key stakeholders with interests in global ocean observing systems.

The MarineXML consortium participants project partners are:
- HR Wallingford UK (Coordinator)
- UK Marine Information Council UK
- SevenCs DE
- Nansen Environmental and Remote Sensing Centre NO
- Central Laboratory of the Research Council UK
- Rijkswaterstaat NL
- Flanders Marine Institute BE
- Social Change Online AU
- Swedish Meteorological and Hydrological Institute (EuroGOOS) SE
- International Oceanographic Data & Information Exchange Committee (IODE)

The MarineXML project has identified four key measurable project objectives:

Objective 1. To produce a prototype marine data ontology framework for interoperability

The project outcome will align a conceptual Marine Mark-Up Language (MML) specification with other XML and related standards. This will be achieved through research into a marine data ontological framework. Ontologies (semantic structures) define how entities *exist* and their relationship to other entities. Accordingly this ontological framework will research how different marine data *co-exist* from the perspective of their interoperability. This work, while essential to the success of the project, may be the least visible to the user community. In the way that HTTP provides an underlying Internet Protocol, invisible to all but the most technical users, the ontologies developed in MarineXML will also be largely invisible. However, this framework will be used to provide the initial MML interoperability protocols and will be incorporated into the test bed demonstration systems. Specific functions (applications) will then be built on this underlying protocol to create usable systems and processes in a similar manner to an e-mail package that is built on the HTTP protocol.

Objective 2. To produce working demonstrations of the data interoperability framework

The project will develop a working prototype test bed of the interoperability framework in Objective 1 that test and demonstrate the functionality and interoperability of heterogeneous data sets from disparate agencies. It is intended that a hybrid version of the Electronic Chart Display and Information System (ECDIS) will be the basic data management platform of these demonstration prototypes. An XML version of an ECDIS will be developed that will support the automatic ingestion of data from a number of sources including near to real time measurement. It will provide the platform to integrate this data and prepare it, using XML based technologies such as XSLT (eXtensible Stylesheet Language Transformation), for ingestion by other XML and non-XML compliant applications and models for analysis, display and dissemination.

Objective 3. To develop a prototype MML specification

Using the outputs of the demonstration in Objective 2, the interoperability framework can be further developed as the basis of the MML specification. To be effective, the demonstration systems must be able to create an interoperability environment and this environment will form the embryonic MML specification. While the project will not develop the full MML specification, it will produce and document an initial framework.

Objective 4. To advance the standardization of a Marine Mark-up Language

MarineXML will seek to ensure the standardization of a MML by liaising with the International Marine XML Panel (under the auspices of the IOC) during the project and provide support and advice for the Panel to further the standardization process beyond the end of the project.

Additionally, the dissemination activities centred around the demonstration systems will aim to attract stakeholder organizations to participate in the MML specification development.

The work to be accomplished by the project is broken down into six work-packets:
- **WP1: Co-ordination.** Project co-ordination and management
- **WP2: Exploitation and Dissemination**. The dissemination, exploitation and post-project development of MML
- **WP3: Standards Review**. Analysis of relevant marine data standards and standards outside the marine sphere that have structures capable of supporting marine data with minor modification and adaptation.
- **WP4: Ontologies and Glossaries**. Development of ontological structures, application of glossaries to marine terms and the application of generic ontological model of marine data.
- **WP5: Test Bed Demonstration**. Includes test bed design and data source selection, test bed application development and test bed refinement
- **WP6: MML Specification Development**. Analysis and prototyping of the MML specification requirements and modifications to the MML specification resulting from 'lessons learned' during the test bed development. Final MML Specification Creation

The MarineXML Project will not result in the creation of a full MML specification but will address the underlying framework issues of interoperability between existing and emerging standards and will provide a technical basis for the development of full specification. Work is expected to commence on the project in December 2002.

Other initiatives

Marine Environmental Data Inventory (MEDI)

The IODE programme of the IOC has developed the MEDI metadata authoring tool to encourage data collectors and scientists to produce metadata descriptions for their datasets. The tool is browser-enabled and operates in a client-server configuration. Clients can access MEDI on a local network or over the internet. MEDI uses the Directory Interchange Format (DIF) developed by NASA's Global Change Master Directory (GCMD). A DIF consists of a collection of fields which detail specific information about the data. The DIF contains those fields which are necessary for users to decide whether a particular data set would be useful for their needs. Metadata records are stored as DIF-XML files and XML is used to transfer MEDI-DIF and GCMD-DIF records.

IODE Group of Experts on Technical Aspects of Data Exchange (GE-TADE)

This group has defined the main mission of its work plan as *'The development of an End-to-End Marine Data Management Framework'*. The present situation with a large number of data formats in existence, covering many different data types with many countries and centres having different computer systems results in a high level of duplication of effort. There is a need for a framework to underpin the full cycle of marine data from collection through to product development and dissemination. One of the Group's objectives is the development of marine

XML as a mechanism to facilitate format and platform independent information, metadata and data exchange. A limited pilot project to demonstrate an activity within the E2EDM framework is currently in progress at the Russian NODC.

MarineXML community portal

IOC has registered the marinexml.net domain and established a community portal to provide a discussion forum for MarineXML activities. This is a dynamic web site and registered users are encouraged to submit items for inclusion on this page. The site contains details of current developments detailed in this paper together with reference material.

References

Keeley, R.J. 2001 The XML Bricks Concept - Version 3. Available: http://ioc.unesco.org/marinexml/files.php?action=viewfile&fid=3&fcat_id=22 [14 November 2002]

Marine XML Portal. Available: http://marineXML.net [14 November 2002]

Marine Environmental Data Inventory (MEDI). Available http://ioc.unesco.org/MEDI [14 November 2002]

Oceanographic Bricks Applied in an XML Environment . Available http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/ICES/web%20xml/SSF-xml.htm [14 November 2002]

Report of the ICES-IOC Study Group on the Development of Marine Data Exchange Systems Using XML, Helsinki, Finland. 15–16 April 2002. Available: http://ioc.unesco.org/marinexml/files.php?action=viewfile&fid=4&fcat_id=22 [14 November 2002]

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 187-193
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Use of XML technology in the Baltic Sea fisheries database

Peter Sandbeck, Brian James Cowan and Teunis Jansen

Danish Institute for Fisheries Research
Jægersborgvej 64-66, DK-2800 Lyngby, Denmark
E-mail: pes@dfu.min.dk

Abstract

Aggregated commercial catch sampling data is used to estimate the fish stocks of cod (*Gadus morhua)* in the Baltic Sea. The present system for data validation, uploading and downloading of data via the Internet is presented. XML technology is used to check the files before uploading. The way to configure different kinds of data validation and data cardinality in an XML schema file is discussed. XML schema files can be used to check for cardinality between record types, mandatory data, value ranges, string patterns and enumeration data consistency. The column separated data exchange format used in many present marine data exchange systems i.e. by the International Council for the Exploration of the Sea (ICES) is compared to the XML format and the benefits of using XML technology are discussed. Finally a model for a future system using XML technology and Webservices in a distributed Internet based datawarehouse solution is presented and discussed.

Keywords: XML schema; Baltcom; Validation; Webservice; Datawarehouse.

Introduction

In three projects financed by The European Commission, Directorate-General FISH, fisheries discard data from the fishery in the Baltic Sea in the period 1995-2001 has been sampled (Degel and Jansen, in prep.). All countries around the Baltic Sea participated in the projects. The overall aim of the projects was to improve the quality of the stock assessment of Baltic cod. In the second project a simple database system was developed, and in the third project a web based datawarehouse was developed in order to make the data handling much easier and the data validation more consistent. The datawarehouse named Baltcom was placed at the Danish Institute for Fisheries Research on a Windows 2000 server with Internet Information Server and Microsoft SQL Server software. The web based user interface was developed with .Net technology using Visual Basic and Active Server Pages (ASPX).

The data exchange format used during the projects was a column separated format with nested record types (one to many relations) as used by ICES. The participating countries all have programs, which request data from the national databases and write the data into files with the ICES format. In order to check the data in a consistent way before they are archived in the Baltcom database, it was decided to use XML technology to configure the data validation rules, thus the configuration was done in an XML schema file. XML is an international standard developed by the W3 consortium (W3C). A program component which converts a file in ICES format into XML format was developed. By converting the ICES format into XML format we got the possibility of using a standard XML application programming interface (API), which is part of most modern programming software, to check and upload the data to the Baltcom datawarehouse. The W3C is still developing the XML standards and the XML schema semantics has been improved since the first version of the Baltcom datawarehouse application. The same is true for the standard XML API used in Visual Basic. Therefore a new version of the Baltcom datawarehouse application is under construction using the new version of XML schema semantics. This paper focusses on using the new version of XML schema technology in order to configure and perform various data validation checks in the Baltcom datawarehouse application.

Baltcom data exchange

The Baltcom data upload system is shown in Fig. 1. The data is archived in a common datawarehouse, which is accessed via the Internet. The program which uploads and downloads the data runs in an Internet browser. The national database manager makes a zipped file in the agreed data exchange format, which is an ICES format type. This file is then uploaded via the Internet based interface program. The upload program first converts the column based exchange format into an XML format file according to a predefined XML schema file. The data is then validated using the XML schema validation rules. If the data is erroneous the data is not uploaded, otherwise the data is archived in the Baltcom datawarehouse.

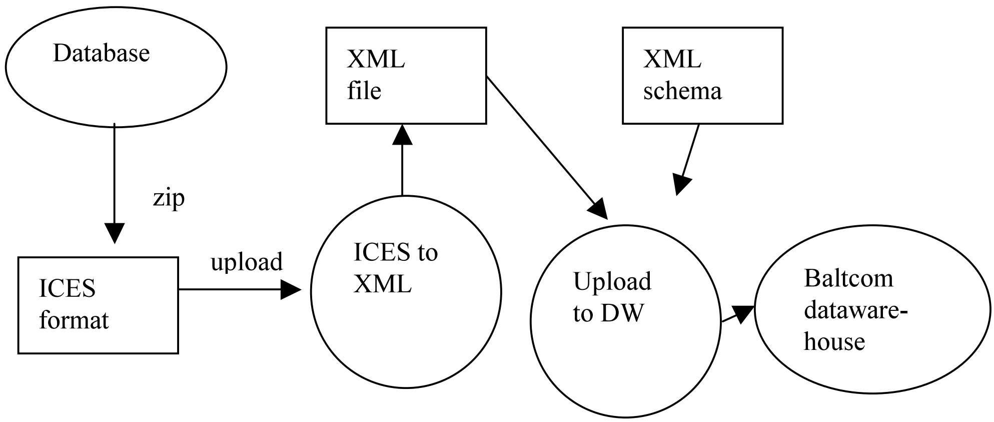


Fig. 1. The Baltcom data upload system.

The URL to the Baltcom system is http://www.Baltcom.org. A password to the system can be requested by mailing the administrator: administrator@baltcom.org.

ICES format to XML format conversion

The commercial catch sampling data used in Baltcom includes 4 different record types (EU-study 98/024). There is a one-to-many relation between record type 1 (haul information) and record type 2 (length frequency data). Record type 3 (sex-maturity-age-length keys) is related to a journey not to a specific station and therefore it is not directly related to record type 1. Record type 4 (extended gear information) is not used in the present Baltcom datawarehouse. Each record type consists of a set of data fields. Each data field has a defined data type, and it might have a range of valid values, a specific char pattern or a set of valid enumerations. A part of the record type 1 data format specification is shown in Table I.

Table I. Part of an exchange format specification (from EU-study 98/024). All elements are mandatory. Type given as length (first digit) and alphanumeric/numeric (sec. digit)

Specifications for record type 1 (Haul information)

Position	name	Type	Range	Comments
1 - 2	Record type	2A		Fixed value HH
3 - 5	Country	3A	See appendix I	ICES alpha code
6 - 7	Year	2N	00 to 99	
8 - 11	Journey	4N	1 to 9999	National coding system
12 - 14	Station No.	3N	1 to 999, 0	Seq. numb. by journey
15 - 16	Month	2N	1 to 12	
17 - 18	Day	2N	1 to 28/29/30/31	
19	Sampling type	1A	H, S	Harbour or sea sampling
20 - 22	Vessel length	3N	1 to 999	Overall length in m

In the XML format the haul information record type (record type 1) has been split into two record types, journey and station. In this way record type 3, the sex-maturity-age-length keys (SMALK) can be related to the journey record. Thus the structure of the data in the XML file is this:

```
<Journey>
        <Station>
                <Length frequency></Length frequency >
        </Station>
        <SMALK></SMALK>
</Journey>
```

There is a one-to-many relation between Journey and Station, between Station and Length Frequency, and between Journey and SMALK. These relationships are reflected in the XML schema file and hence in the XML data files. The ICES format files are converted to this XML format and then validated using the XML schema.

The advantages of using XML format for data exchange instead of column based formats are:

- data in XML format is more readable;
- data validation can be configured in an XML schema file;
- the format can easily be changed;

- uploading, downloading and data validation procedures can be programmed using the built in standard XML API.

XML schema validation

The ICES format specifications have all the information needed to build an XML schema file with data validation configuration. Each record type and each field type is defined as either a complex type or a simple type. The schema is defined in a top down manner starting with the topmost record types in the record tree and ending with the definitions of all the single field types. In XML terminology a schema file is denoted an XML namespace. Namespaces can be global like the main W3C XML namespace or private like the present Baltcom schema. The Baltcom schema could be made global by placing it on a website. The references to the XML namespaces used in an XML file are placed at the top of each XML file.

The following shows some examples of how different kinds of data validation are defined in the Baltcom XML schema file.

Cardinality

The attributes “minOccurs” and “maxOccurs” are used to define the cardinality between records. The default values are 1 for minOccurs and 1 for maxOccurs. In the example below, a journey record may hold from 1 to an infinite number of “Station” records and zero or one “SMALK” record.

```
<xsd:complexType name="JType">
        <xsd:sequence>
                <xsd:element name="DateStart" type="xsd:date"/>
                <xsd:element name="DateEnd" type="xsd:date"/>
<xsd:element name="Station" type="SType" maxOccurs="unbounded"/>
<xsd:element name="SMALKs" type="SMALKsType" minOccurs="0"/>
        </xsd:sequence>
        <xsd:attribute name="Country" type="tCountry" use="required"/>
        <xsd:attribute name="Journey" type="tJourney" use="required"/>
        <xsd:attribute name="Year" type="tYear" use="required"/>
</xsd:complexType>
```

Primary key fields

The fields which make up the primary key for a record type are defined as attributes rather than normal fields (see example above). In this way the primary key data in the XML file will be shown in the header target of each record. The “use=required” attribute makes the data mandatory.

Range check

Range checks are defined with the “minInclusive” and “maxInclusive” attributes.

```
<xsd:simpleType name="tYear">
        <xsd:restriction base="xsd:int">
```

```
            <xsd:minInclusive value="1900"/>
            <xsd:maxInclusive value="3000"/>
      </xsd:restriction>
</xsd:simpleType>
```

String pattern

String patterns can be defined with the “pattern” attribute.

```
<xsd:simpleType name="tRectangle">
      <xsd:restriction base="xsd:string">
            <xsd:pattern value="[0-9]{2}[A-Z]{1}[0-9]{1}"/>
      </xsd:restriction>
</xsd:simpleType>
```

In this example a geographical rectangle is defined as two digits followed by one letter followed by one digit.

Enumeration

Some data may only have a defined set of values. These values are defined as enumerations.

```
<xsd:simpleType name="tSamplingType">
      <xsd:restriction base="xsd:string">
            <xsd:enumeration value="H"/>
            <xsd:enumeration value="S"/>
            <xsd:enumeration value="M"/>
      </xsd:restriction>
</xsd:simpleType>
```

What is missing is the possibility to check for data dependencies between data fields. For example in the journey record, it is not possible to check that the end date is bigger than or the same as the start date. This feature will properly be included in future XML schema semantics. However today this kind of data validation has to be done by the upload program in a traditional way.

Distributed internet based databases

The present Baltcom datawarehouse solution operates with one website and one datawarehouse, which all partners have to use. However new Internet technologies make it possible to build more advanced database solutions. A proposal for a distributed database system is presented in Fig. 3.

Here each institution has a database with a structure, which is different from the other institutions databases, which is usually the case. When a common data exchange format is defined in an XML schema, each institution builds a datawarehouse based on that data structure and loads the data from the database into the datawarehouse. Thus the datawarehouses at all the institutions participating in the project have the same structure, and data can easily be transferred to another datawarehouse in XML files. The request for data is done directly by calling a webservice at the source website, which makes the data request, sends the data back to the calling webservice, which validates and uploads the data in the target datawarehouse. A webservice is a program on a website, which can be called from another program at another website. There are other technologies, which can do the same i.e. Corba and BizTalk technology.

The advantages of this solution are that data can be pulled to the requesting institution at any time. When common data have to be used the data access is fast because data are stored locally. Finally data at each site can automatically be updated when they are updated at the source database. Thus this kind of solution will probably be seen more in future projects.

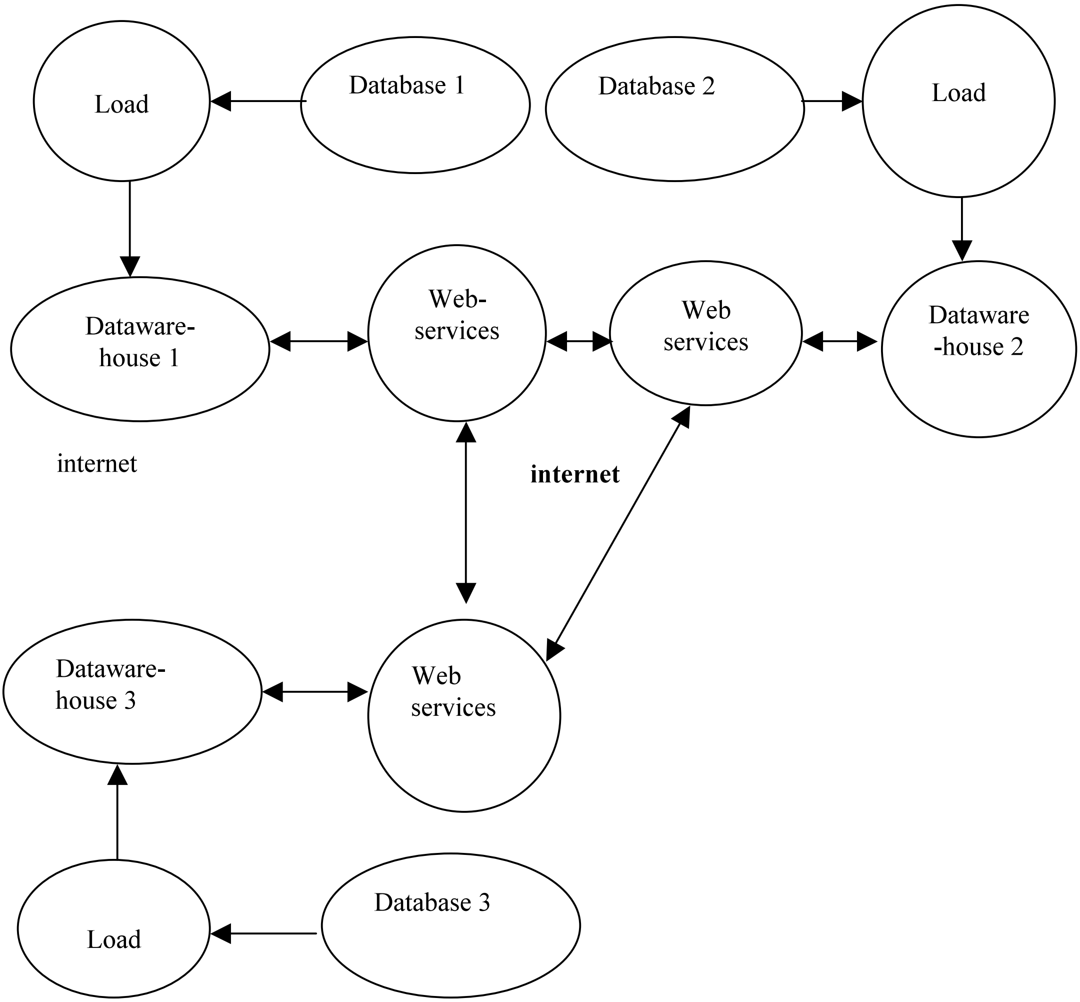


Fig. 2. Distributed internet based database system.

Conclusion

In the Baltic Sea Fisheries database XML technology has been used for data exchange and data validation. The project has revealed some general advantages of using XML for these purposes. Column based data exchange formats like ICES formats can easily be converted into XML formats with nested record types. Data in XML format is more readable and the files can be handled directly by most modern application programming interfaces.

The XML data exchange format for a specific project can be defined in an XML schema, thus ensuring the consistency of the XML exchange format used. The XML schema can also be used to configure the validation of data. Several kinds of checks can be performed like cardinality between record types, mandatory data, value ranges, string patterns and enumerations. However some kinds of data validation like comparisons of data values between fields can still not be performed.

Modern Internet technologies use XML as a standard data interface and configuration format. XML is an integrated part of new technologies like .Net, webservices and others. With these technologies one can build advanced solutions like distributed database systems where data is pulled from a source database on one website to a target database on another website by an application.

References

Degel H. and T. Jansen. (in prep.). BaltCom database. An internet based data warehouse for fisheries data in Baltic fisheries. ICES Working Paper. Baltic Fisheries Assessment Working Group.

EU-study 98/024. EU-study project No. 98/024. Version 7 of 14/6-01. Livingston, David: Essential XML for web professionals, Prentice Hall 2001 - W3C: http://www.w3c.org

SESSION: CASE STUDIES

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 197-205
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

RIVELA: database for the research on Venice and the lagoon

Pierpaolo Campostrini[1], Caterina Dabalà[1], Stefania De Zorzi[1] and Renzo Orsini[2]

[1]CORILA
San Marco 2847, I-30124 Venezia, Italy
E-mail: campostrini@corila.it

[2]Computer Science Department, University of Venice
via Torino 155, I-30172 Venezia, Italy

Abstract

CORILA is running a complex Research Programme on the Venice Lagoon, involving several research institutions of different fields. The integrated and interdisciplinary management of information constitutes a challenging goal of the Programme. RIVELA is a relational database of which the main aim is to store and make available the data and results of the present Research Programme, but it is also open to consider previous and different data sets collected on the Venice Lagoon, even for different purposes. The vision is to achieve, in a near future, a distributed database made of many archives, each independently maintained by different administrations, with a complete data sets visibility and browsing from any position. The data model specifically built for RIVELA required to interview all the data producers, coping with the different ways of storing the relevant information, in order to find a data structure effectively able to handle data of various types and origins, but not too complicated, to persuade to use the system. The interdisciplinary use of data is of paramount importance in the environmental research, and specifically in the Venice case. The scientists' co-operation from the first stage has been considered fundamental to achieve the project goal. The RIVELA relational database structure is made of two parts, static and dynamic. All the interactions with the database occur via the web. The data input phase requires first a CSV file preparation on the user's computer, following easy rules, then to connect to the CORILA web site to send the file. A consistency check is automatically performed before the actual database population. The access to data occurs via the web, through guided interactive searches and GIS queries. The query result is produced in a tabular format, which can be shown either in the web page, or downloaded for further elaborations. The result visualisation on GIS maps will be provided too.

Keywords: Environmental database; GIS; Web interface; Data model; Venice Lagoon.

Introduction: the research on the Venice Lagoon and the many environmental data archives

The Consortium for the Researches on the Venice Lagoon System (CORILA) is an association of the University of Venice, the University of Padua, the Venice University Institute of Architecture (IUAV) and the National Research Council of Italy (CNR). It was established in 1998 to promote and to co-ordinate the scientific research on Venice and its Lagoon. The current Research Programme (2000-2004) has a value of € 10 million and is mainly funded by the Italian Special Law for Venice, via the Ministry for Teaching, University and Research. The programme involves an international network of 70 research institutions and it is organized in 11 research lines of different disciplines, grouped in four thematic areas: Economy, Architecture, Environmental Processes and Data Management.

The Venice Lagoon represents a very complex and delicate environment of unique cultural and natural value. The urban and lagoon environments require continuous interventions, both for maintenance and for remediation of errors made in the past (*e.g.* pollution in the sediments due to past industrialisation) and for coping with global changes problems (mainly sea level rise). Moreover, in order to continue to exist as a living city, Venice should maintain and develop economic activities (*e.g.* port, industries, fishery), while it is subjected to an increasing tourist pressure. Therefore, from the management needs, several specific and general questions rise, how to maintain all these activities and interventions in a 'sustainable framework' for the safeguarding of Venice: some of these questions are challenging the scientific community.

In fact, since many years, a wide research in several branches is occurring on the Venice Lagoon system, producing a huge amount of data, reports, publications. However, in the past, there was not a systematic effort in the scientific research to put together all the pieces of information, making available tools which would permit an easy access to all the data. Nevertheless, very long historical series of data have been recorded, by single private or public institutions and in the last years some of them have been made available, mainly through the efforts spent by a private academic institution, *Istituto Veneto di Scienze, Lettere ed Arti*, based in Venice (*www.istitutoveneto.it*).

Independently from the scientific community, the Ministry of Public Works, who has the competence by law for the Venice Lagoon safeguarding, through its local office *Magistrato alle Acque*, is running large monitoring activities and is maintaining the larger data set: however, presently the access to data is strictly regulated and possible only after a specific authorisation, while an easier access is foreseen in the near future. Other Administrations collect data for specific purposes (*e.g.* air or water quality control, tide level forecast, fishery health checks, etc.), but there is not at present a real structured information exchange, even where this would permit relevant mutual aid for the administrations themselves and would improve the phenomena comprehension by the scientific community.

Generally speaking, the problem of data archiving and sharing is presently felt as one of the most relevant for properly assessing the safeguarding of Venice, not only by scientists, but also by administrators and lay people. In this framework, the CORILA's database RIVELA is intended to be a first step, mainly oriented towards the scientific community, but it is expressly built open to a wider integration with other archives, for realising, in a near future, a really accessible-to-all distributed information system of the Venice Lagoon.

The need of a common data model for interdisciplinary research

RIVELA is a relational and flexible database for storing and managing, in permanent and secure archives, information from the researches on the Venice Lagoon, independently from the actual research executor. RIVELA is designed to facilitate the provision of the research results to the scientific community, the decision makers and the general public.

Presently, it is populated mainly by the data produced by the first CORILA's Research Programme 2000-2004. It is well-known that the main problem in planning, structuring and organizing an environmental database is the management of the mass of data which derive from different disciplines (Michener and Brunt, 2002). In fact, the complexity and the richness of the research programme ask for a carefully designed data management, both for long term archiving of data and for the immediate and interdisciplinary data usage by researchers of many different sciences. Therefore one of the most important objectives of RIVELA is the definition of a common data model able to integrate and manage interdisciplinary data (*e.g.* the interaction between chemical parameters, information about the biota and physical data).

The development of such a database required, in the planning phase, the lengthy task of organizing the existing knowledge, to obtain an accurate description of environmental matrices, sampling localities, parameters to be measured, types of sample, data acquisition methodologies, and a lot of other ancillary information, to be stored in the database in addition to the main data. Thus, the classification of records in RIVELA is the result of a complex interaction with the research groups to collect and define, in a common structure, the different ways to store the relevant information.

The RIVELA database structure

The RIVELA relational database consists of two main components (Fig. 1): a *static* part, relative to the results of auxiliary support data, and a dynamic part, relative to surveys performed in field or laboratory activities. The database engine is Oracle 9i.

The static part contains the following information:

- Research Groups administrative data and list of current activities (Research Projects, Work Packages, Activity);
- Geographical location of data (Zones, Environmental Units, Localities);
- Data types (Matrices, Types of Sample, Parameters);
- Data acquisition methodologies (Method, Apparatus).

The dynamic part contains four interconnected fundamental entities: *Measurements, Samples, Stations* and *Sampling Activities*. A *Measurement* is the value of a parameter deriving from a certain Sample, which has a precise spatial location (*Station*) and a temporal location. The Samples are, in turn, classified according to type (*Sample type*), depending on the environmental matrix.

RIVELA applications

To ensure the maximum flexibility of accessing RIVELA, all the applications have a web interface. This allows the researchers to work remotely both for loading and for querying the data. Moreover, RIVELA is designed to offer both textual and GIS interfaces. A web application allows the visualization of the static part of the database, and the suggestion to the database administrator of modifications or additions to the following lists: Parameters, Environmental Units, Localities, Substratum, Unit of Measurement.
The other two main web applications available are for data input, and for data search and extraction through guided queries.

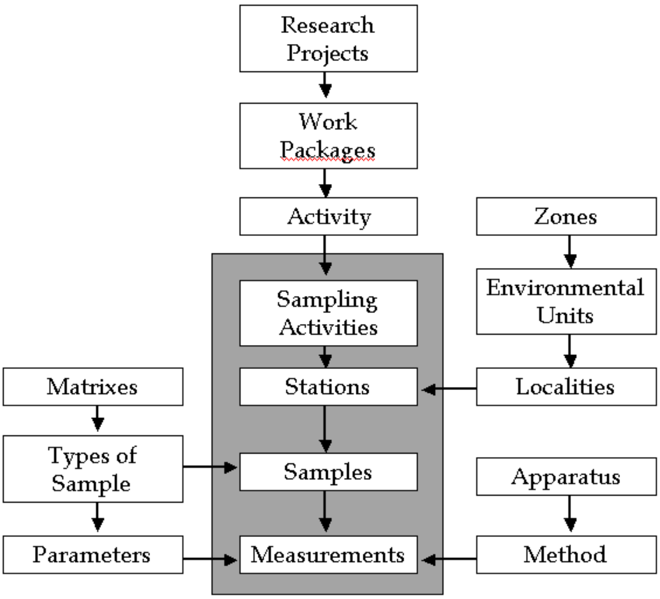


Fig.1. Simplified scheme of the RIVELA structure. The dynamic part is in gray framework.

Data input

The use of web application for data input, even if largely suggested, presently is not so frequently implemented in scientific data bases (Bobak, 2000; Michener and Brunt, 2002). The presence of many users of different institutions represented a strong incentive to use this approach.
In order to feed RIVELA, the scientific coordinator of a research unit, properly identified with username and password, accesses an appropriate web page containing both a series of fields to be filled in manually and buttons to send one or more files in CSV format, which contain the data previously prepared on his own computer. Information on the methodologies used must also be inserted, and auxiliary data like bibliographies, images, etc. can be provided too.

Three formats for data files are foreseen, according to the following types of measurements:

- a set of measures relative to a unique sample of a unique station in a unique sampling activity;
- a set of measures relative to a set of samples and measurements related to a unique station via a unique sampling activity;
- data relative to a unique sampling activity with many stations and many measurements.

In the data files, data are organized in tabular form, where, as a general rule, the rows represent different values of a set of parameters measured in a specific sample and/or at a particular time; the columns represent all the values taken for each parameter during measurements of the samples or over time. So, the number of the columns is variable, as each column contains all the measured values of each single parameter. The file can be prepared with any program managing tables, like MS Excel or MS Access, or any text editor, or can be produced by appropriate software converting data from some measurement device, and saved in the standard CSV format. Each file sent to the database is processed by an automatic loading program, which carries out a set of controls of validity (presence of the coordinates in the correct reference frame, correct name of the parameters etc.) and indicates by e-mail any errors, both to the user and to the administrator. In fact, this control is considered of paramount importance to guarantee the data quality (Brackett, 2000; Michener and Brunt, 2002).

The structure scheme of the data input is represented in Fig. 2. An important feature of the input system is that the CSV file containing the input data is permanently stored in the database too. Moreover, all data in the database maintain an information which can be used to trace back to the original input file. These two facts make possible the following fundamental feature of RIVELA: *it is always possible to correct the database whenever an error is discovered in an input data file*, by 'undoing' the relative insertions, and 'redoing' the new correct ones.

Guided queries

RIVELA is a relational database, and therefore all the usual ways of accessing the data are available to users: for instance, by using directly the SQL query language, or through "individual productivity" tools, such as spreadsheets, like Excel, or databases, like Access or FileMaker, to extract sets of data in tabular format. Moreover, two additional access and extraction ways have been provided: a guided interactive search and a GIS query tool (Fig. 3).

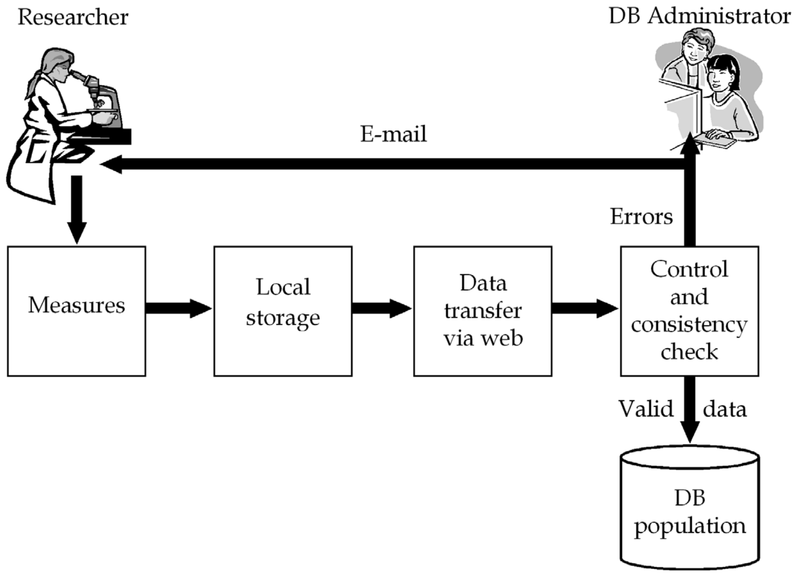


Fig.2. Scheme of RIVELA data input.

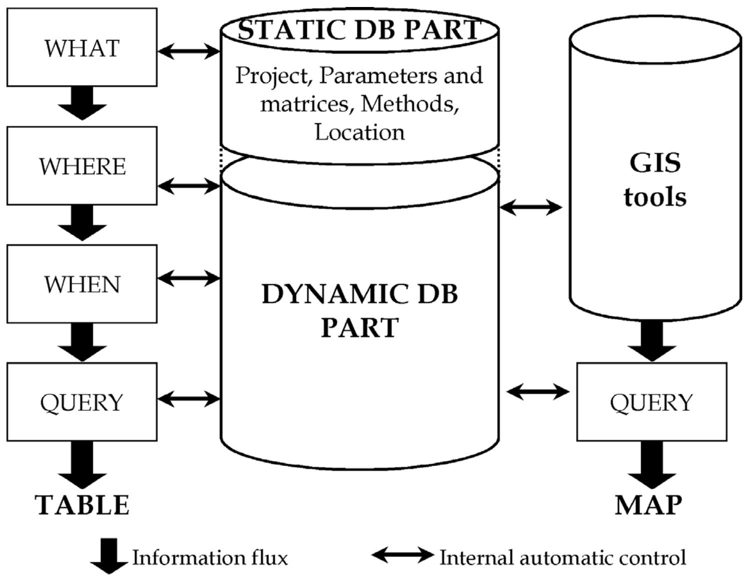


Fig. 3. Scheme of a guided query.

In formulating an interactive interrogation, the user can choose between different query modes:

- by writing a free text to select a parameter;
- by choosing among all the parameter types, listed and divided in different matrices;
- by choosing among all the parameter types, classified by the corresponding type of sample.

Then, the user is guided towards:

- selection of time and location;
- selection of attributes required to be shown on the report table.

The spatial location can be indicated either by selecting a locality, or by specifying directly the geographical coordinates of the area of interest.
The attribute selection page allows, by check-boxes, an easy-to-read representation of the report table in the screen window.

Finally, the query result is produced in a tabular format, which can be shown either in the web page, or downloaded, in CSV format, for further elaborations. On the web page, each record is always linked to its ancillary information, which can be explored by traversing all the relations among the data.

Integration with a Geographical Information System

A proper and effective correlation between the environment data, the landscape and the geographical representation is of emerging importance in all information systems used in environmental applications (Hof and Bevers, 2002) and it has been considered essential in RIVELA, where for each stored piece of information, a specific temporal and geographic location is retained. This allows the 'database engine' to be used by a number of GIS applications.

A set of base maps, stored in the database, is provided as the default layout: these are active maps and allow the user to select any geographical object.

The Venice Lagoon is displayed at different scales, to allow users to find easily the relevant information.

All information stored in the database is in principle displayable overlaid on the Venice Lagoon maps. Presently, RIVELA and the GIS tools are integrated only in the local CORILA network, where the GIS applications run on specific workstations and make use of the data stored in the database server. The GIS applications are based on Intergraph GeoMedia© systems family. The system is organized in a client-server architecture.

Many data classifications and correlations can be visualized (Fig. 4), and queries can be expressed by selecting an area with an input device.
The GIS-web interface, that will allow to display the content of RIVELA on the georeferenced maps, is presently under construction. This interface will provide the user with different types of result visualisations, modelling and statistical analyses.

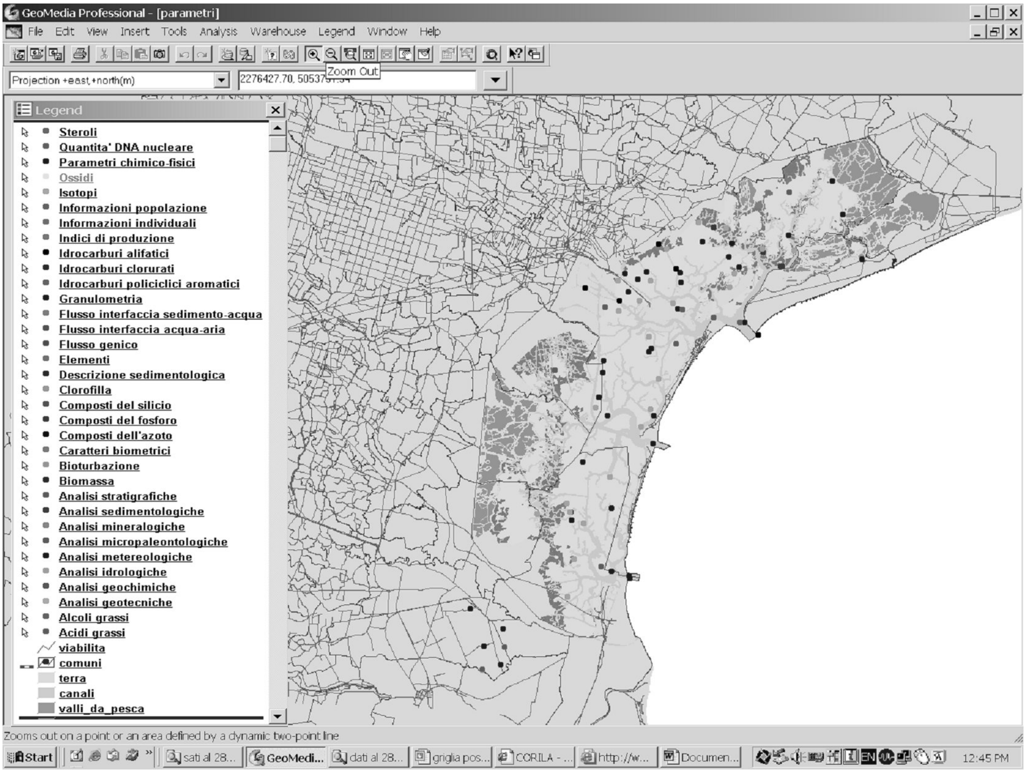


Fig. 4. Example of a GIS query: acquisition sites of the CORILA research program.

Conclusions

RIVELA is a flexible and open archive for scientific researches on the Venice Lagoon, growing with the CORILA Research Program. It is developed in strong connection with the final user, *i.e.* the CORILA researchers and the scientific community. While the use of a relational structure allows the use of standard tools for managing the database, the web interface allows scientists to remotely load and query the data in a very simple manner.

The aim of RIVELA is to become an open archive able to store, in a standard format, the information related to the scientific researches, developed on the Venice Lagoon. Its main goals are:

- to avoid any data loss;
- to permit interdisciplinary data analysis;
- to allow wide dissemination and easy access to data.

The researchers will find in RIVELA a data warehouse, able to give added value to one's own data, accessing larger data pools and using advanced instruments, both for analysis and visualization.

RIVELA is under continuous development: next works will concentrate on the web GIS interface, on a data engine to search information into textual documents, and on an interface for statistical and spatial analyses.

Acknowledgement

The authors wish to thank all the researchers interviewed and the other members of the CORILA staff, who patiently collaborated to the definition of the data model. Dr Roberto Rosselli, Director of the Servizio Informativo of Magistrato alle Acque, for useful discussions and for data and material generously made available, and Mr. Giovanni Gatto for the help given during his thesis.

References

Bobak Angelo R.. 2000. Database distribuiti e multi – database. Jackson Libri

Brackett Michael H. 2000. Data resource quality: turning bad habits into good practices. Addison – Wesley

Hof John and Michael Bevers. 2002. Spatial optimization in ecological applications. Columbia University Press

Michener Willam K. and James W. Brunt. 2002. Ecological data (Design, management and processing) – Blackwell Science Ltd

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 207-215
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Recent advances in oceanographic data management of the Mediterranean and Black Seas: the MEDAR/MEDATLAS 2002 database

C. Maillard[1], E. Balopoulos[2] and the MEDAR Group

[1] IFREMER/SISMER
Brest, France
E-mail: Catherine.Maillard@ifremer.fr

[2] National Centre for Marine Research
Anavyssos, Greece

Abstract

An integrated database of temperature, salinity and biochemical parameters has been developed for the Mediterranean and Black Seas, from data collected for several decades by about 150 laboratories of 33 countries, through a wide co-operation of Mediterranean and Black Sea countries. These data have been reformatted at the common MEDATLAS format, and checked for quality according to a common protocol based on the international IOC, ICES and EC/MAST recommendations, with automatic (objective) and visual (subjective) checks. The resulting database includes 284 371 temperature profiles, and several thousands of profiles of each biochemical parameter (oxygen, nutrients, chlorophyll), whose number decreases from phosphate, (20808 profiles) to total nitrogen (153 profiles). This represents the most comprehensive data set available for the region, even if the spatial coverage of each parameter is not homogeneous, and in areas such as the middle of the deep basins and the Lybian shelf, remains limited.

These data have been compiled in each country by the national oceanographic data centre or the national agency designated for international data exchange to the Intergovernmental Oceanographic Commission (IOC). They have been afterwards reformatted at the common MEDATLAS format, and checked for quality according to a common protocol based on the international IOC, ICES and EC/MAST recommendations, with automatic (objective) and visual (subjective) checks. Data selected with acceptable quality flags have been interpolated at 25 horizontal levels and objectively analysed to produce the gridded climatological fields, vertical sections and horizontal maps, by using a variational model for objective analysis. Finally, all the metadata (cruise inventory), observed data, gridded data, maps, documentation and software are published on a set of four CD-ROMs for easy access to the data.

There are plans to maintain and develop further this data management system, especially to enlarge the number of processed parameters, to improve the quality control procedures of the biochemical parameters taking into account the analysis of recent good quality data sets and to facilitate the on line access to the distributed data. To carry out such data management tasks and insure long term data safeguarding and accessibility, it is important to enhance the infrastructure of professional data centres, and strong links with scientific institutes. This will not only ensure a faster data circulation, but also produce scientifically validated data products, one of the major requirements from scientists and non-scientific users.

Keywords: Mediterranean; Black Sea; Temperature; Nutrients; Database.

Introduction

As the marine biological ecosystem is the most sensitive to any climatic change, the availability of basic oceanographic data like temperature, salinity, oxygen concentration and nutrients are necessary for monitoring the system. For the Mediterranean and Black Seas, temperature, salinity, oxygen, nitrate, nitrite, ammonia, total nitrogen, phosphate, total phosphorus, silicate, H2S, pH, alkalinity, and chlorophyll-a vertical profiles have been collected for several decades by about 150 laboratories of 33 countries, most of them from the bordering countries. However, many of these data remained dispersed in scientific laboratories and national data centres at different formats and various levels of documentation.

To facilitate the access to these dispersed data, an EU concerted action MEDAR/MEDATLAS II (MAS3-CT98-0174 & ERBIC20-CT98-0103) was initiated for developing a joint comprehensive database through a wide co-operation of Mediterranean and Black Sea countries. This action follows two previous pilot projects (MEDATLAS and MODB) where first compilations of temperature and salinity data were made, and represents a regional contribution to the global GODAR project (Global Ocean Data Archaeology and Rescue; Levitus *et al.*, 1998). As the biological data requirements are increasing, it was decided that the update should include the biochemical parameters that were necessary for the ecosystem modelling and for which a sufficient quantity of data was available to allow quality checks to be performed. This compilation has been made in the perspective to get long time series of observations and also climatological statistics, both being necessary for the monitoring and the modelling of the region. This work has been carried out by the MEDAR network, which includes 25 governmental institutes as international organizations from Europe and Mediterranean countries (Fig.1) and the World Data Centre A (USA).

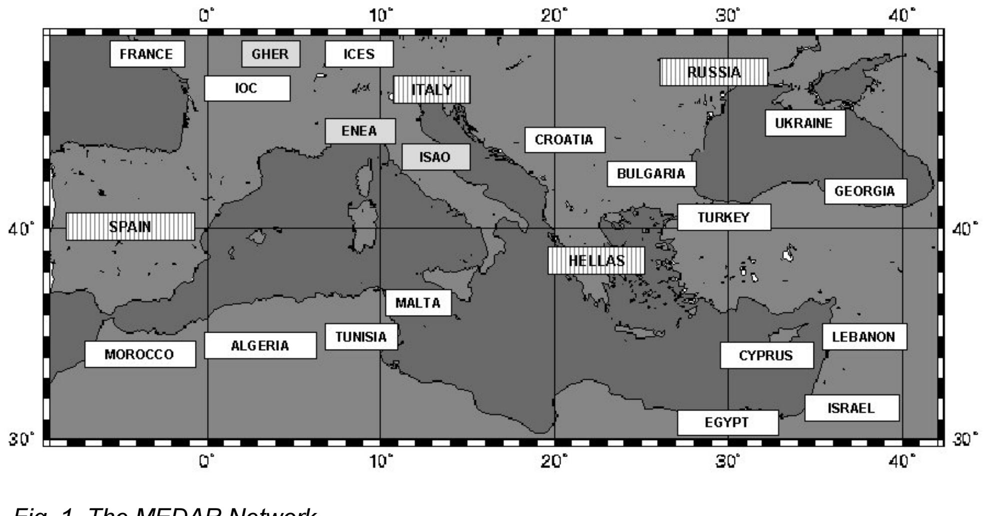


Fig. 1. The MEDAR Network.

This paper is focused on the observed in-situ biochemical data and the final data product on CD-ROM, the climatological analysis being described in Rixen *et al.* (2001 and 2004).

Each NODC is represented by its country name, the regional coordinating centres being hachured. The modelling centres are in grey. ICES and IOC are intergovernmental organisations.

Data processing

The partners were mainly the National Oceanographic Data Centre or Designated National Agencies for International Oceanographic Data and Information Exchange (IODE) of UNESCO Intergovernmental Oceanographic Commission (IOC), and had the duty to compile and safeguard copies of the data sets collected at sea by the scientific laboratories of their country. Information on the experimental conditions, cruises and sources laboratories have been compiled at the same time, but for old historical data, this information was not always available. Nevertheless, the availability of long time series is an important expected result, and even with limited information on the experimental conditions, these data are valuable and have been integrated in the database.
The result of the data rescue, after the elimination of duplicates is given in Table I (a: by type of instrument; and b: by parameter). The biochemical data come from the bottle casts, except for the oxygen concentration data, which are also measured by CTD.

Table I. Content of MEDAR 2002 database

a) by type of instrument

Data type	Nb of profiles
CTD	35 679
Bottle	88 323
MBT and XBT	161 848
Thermistors	29

b) by parameters

Parameter	Nb of Profiles	Parameter	Nb of Profiles
Temperature	284 371	Nitrite	10 508
Salinity	118 009	Ammonium	5 239
Oxygen	44 928	Chlorophyll	4 672
Phosphate	20 761	Alkalinity	2 548
Silicate	15 920	Total Phosphorus	2 381
PH	14 512	H2S (*)	1 843
Nitrate	10 572	Total Nitrogen	153

For the nutrients, and even for phosphate (Fig. 2), which is considered by the biologists as a control parameter of the biota in this region, the data coverage is unfortunately not very homogeneous and the centres of the main basins and the Lybian shelf, have a poor data coverage.

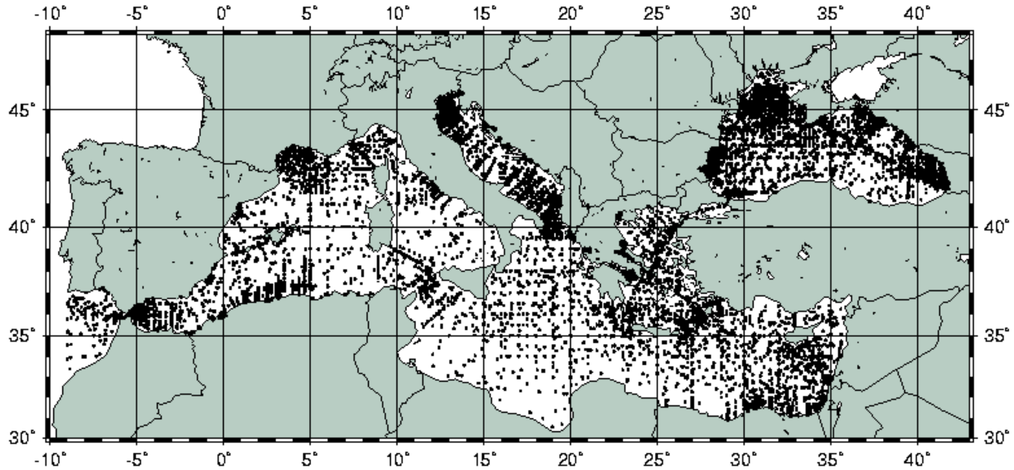


Fig. 2. Location of the 20 761 phosphate profiles of MEDAR 2002 Database.

Quality checks of the observed data

To insure compatibility and coherence of data from so many different sources, it was necessary to define a common protocol for data formatting (including the mandatory metadata) and checking for quality. The existing international standards published by UNESCO (1993) give general guidelines for formatting the data and for implementing quality checks. However they cover essentially the temperature over the global ocean, which needed to be better adjusted for such contrasted regional seas like the Mediterranean and Black Seas. Moreover, very little was available for the quality checks of the biochemicals and further preliminary work has to be done to define a common protocol for data processing (MEDAR Group, 2001).

The common exchange format is the auto-descriptive MEDATLAS format developed for the previous MEDTLAS 1997, MODB et MTPII-MATER projects. The data of the same type (e.g. CTD, bottle casts, or Xbt) collected during one cruise are copied in one file including metadata (cruise and source laboratory information, parameters names, codes and units) in addition to the data points. After the reformatting, the data were submitted to the series of tests defined in the protocol, in three steps:

- QCO: automatic check of the format and completeness of mandatory information;
- QC1: automatic and visual check of the stations headers (coherence of date, time, latitude, longitude, ship velocity);
- QC2: automatic and visual check of the data points (constant profiles detection, broad range checks by comparison with regional limits, narrow range checks by comparison with climatological means, spike detection, vertical stability, Fig. 3).

As a result, a quality flag was added to each numerical value (international GTSPP flag scale). The flag value increases with the severity of the anomaly encountered, and is coloured accordingly on the screen of the operator performing the checks.

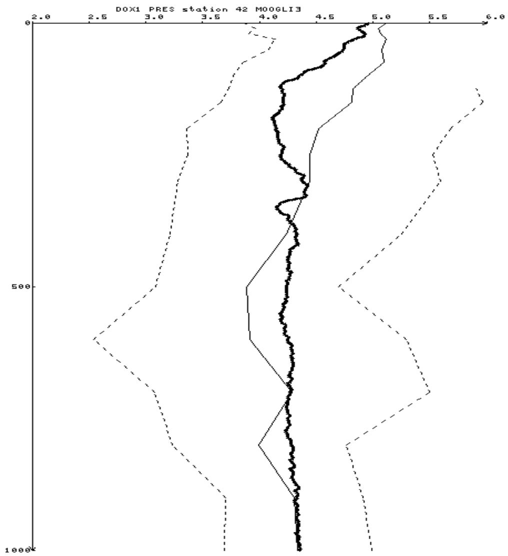


Fig. 3. Check of a set of oxygen profiles of a cruise.

The current profile is plotted in thick line (on the work screen it is coloured according to the quality flags values). The nearest climatological profile is plotted in grey and the expected range of variability (corresponding to 3 climatological standard deviations in deep sea) in dotted lines. All the profiles of the same cruise will be superimposed afterwards to check internal coherence of the data.

Basically, checking the data means to compare them with the existing statistics: minimum and maximum regional values, climatological means and standard deviations (3 to 5 times, depending on the depth of the station). For the temperature and salinity, the previous MEDATLAS (MEDATLAS Group, 1997) climatology was available, and for the biochemicals, the global ocean climatology (Levitus *et al.*, 1998), which may differ from the real local mean, as in Fig. 3. For the biochemicals, it was in fact necessary to better adjust the minimum and maximum acceptable values in subregions and to recheck some values after the first MEDAR climatological computations. One of the main difficulties was that the broad range check was too broad at the first iteration to eliminate erroneous data, including unit errors in historical data. Another difficulty issued from the coastal data that may be good, even outside the limits.

These checks were performed by the national coordinators in four Regional Data Centres (RDC) respectively for the Black Sea, the Mediterranean Eastern Basin, the Central Mediterranean and the Mediterranean Western Basins. After regional quality control of the data, each RDC sent the data sets to the Global Assembling Centre for final checking and merging in the final integrated database.

The quality control of the biochemical data was critical and more difficult to perform than temperature and salinity, because they are based on preliminary knowledge of the statistics, and that these statistics are based on less available information. Some further work has been undertaken, on the methodology, for example to include the check of the Redfield ratio in the check list. But a large effort is still devoted to improve the statistics. New information results from the project climatology itself (Rixen *et al.*, 2004) and on further data analysis made by Manca *et al.* (2003) on the MEDAR data and data from a recent set of good quality data collected during the MTP-II/MATER project in the Mediterranean (Maillard *et al.*, 2002).

Integration of the database and preparation of the data product

Twenty-five vertical levels were selected for computing the climatological computations. These levels correspond to the usual 'standard levels' of the IOC manual, with a few levels added in the more variable upper layer for consistency with the World Ocean Atlas (1998). Only the data points with acceptable quality flags were selected and interpolating at the standard levels, by using the SELMEDAR software (Fig. 4), which is also available for any user on the published database.

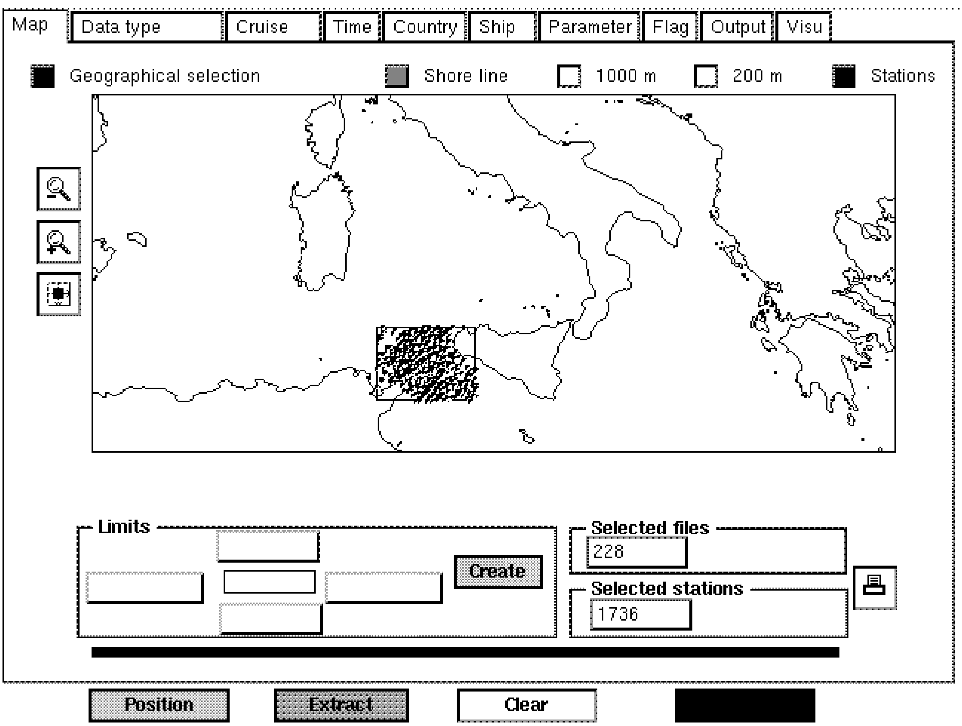


Fig. 4. User interface for data selection, visualisation and interpolation SELMEDAR software.

The climatological computations were made as described in Rixen *et al.* (2001 and 2004) and reviewed each time by regional experts and modellers. These computations have been tested first on preliminary subsets of data not fully qualified. They have been made twice on the final data set, because it appeared some remaining problems in the data and the correlation parameters of the objective analysis model needed better tuning. These iterations from the climatology to the data and vice versa, have been necessary before getting acceptable data and climatological fields, especially for nutrients where the quality control methodology was not fully stabilized during the project.

Among all the annual, seasonal and monthly gridded climatological fields computed and published on the analysis web site (http://modb.oce.ulg.ac.be/Medar/), only the fields given in Table II have been selected for the publication of the database, due to the paucity of data in some areas or period of the year to insure a sufficient level of quality in the results. However, when no other information is available, the non qualified climatological fields can have some utility as first guess of the parameters.

Table II. Content of the qualified MEDAR 2002 climatology

Parameters	Qualified climatology
Temperature, salinity	Annual, seasonal, monthly
Oxygen, silicate, phosphate Hydrogen sulphide (H_2S) in the Black Sea	Annual and seasonal
Nitrate, nitrite, Ph, ammonium, alkalinity, chlorophyll	Annual only
Total phosphorus, total nitrogen	No climatology

All the documentation, qualified gridded data, maps and data requests for observations can be accessed from a distributed WWW site: www.ifremer.fr/medar and are published on a set of four CD-ROMs (MEDAR Group, 2002) described in more detail in Fichaut *et al.* (2002). It includes also a cruise inventory, GIS developed by the RIHMI – World Data Centre, Russian Federation (Vyazilov and Puzova, 2001), QCMEDAR quality control software developed by IEO, Spain (Garcia *et al.*, 2001) and Ocean Data View (ODV), a software to process and plot data developed for the international WOCE experiment at the Bremerhaven University (Schlitzer, 2000). The observed database includes the data files by cruises at MEDATLAS format, and the SELMEDAR software, which allows to extract data at MEDATLAS format and at the ODV import format.

Conclusion

The integrated database of biochemical parameters collected from many laboratories around the Mediterranean and Black Sea since the beginning of the century, represents a unique and comprehensive data product that can have been developed only thanks to a wide international cooperation. MEDAR/MEDATLAS II project had made possible, not only to safeguard and provide an easy access to this important set of data, which in many times, were in danger to be lost, but also to compute improved annual and seasonal gridded climatologogical fields. These are particularly valuable for regional ecosystem models initialisation and following up of the environmental conditions. In addition to the production of the database, sharing expertise and know-how among the partners, has contributed to the improvement and standardization of data management methodologies, and to the overall enhancement of the MEDAR network as distributed data management infrastructure, trained in data qualifying, processing, mapping, archiving and communication. A particularly important point is the development of quality control procedures for biochemical parameters, based both on the general recommendation of the international authorities like ICES, IOC and MAST, and on the statistics made with the archived data. Finally, MEDAR appears to have successfully met all its initial project objectives.

However many other users requirements exist, which are not covered by the present action. First the database needs to be regularly updated by integrating the data (including real time data) recently collected and other important parameters such as CO2 and sea current. Then an internet on line data access to the distributed data should be provided to the users to get the data continuously updated by the data centres of the network. It is also necessary to continue the improvement of the quality controls, with insertion of the new statistics and methods, which were not available during MEDAR. Relationships between nutrients such as the Redfield ratio should be incorporated in routine checks. However the QC of the nutrients in areas with poor data coverage remains difficult and there is still a need for a better data coverage as the quality checks are based on the pre-existing knowledge of the distributions. This is critical in the middle of the deep basins and along the Southern Mediterranean coasts. Compatibility and standardization with data of the same types, collected at the global ocean scale are also required for many research studies.

It is therefore planned to maintain and develop further this data system, through networking and standards development with all the European and Mediterranean Oceanographic Data Centres within and extended ‘Sea DataNet’ oceanographic data management network. The database will be continuously updated by integrating new data produced by recent projects and real time data produced by operational oceanography programmes. Easy on line access to the semi-distributed data sets will be provided by using the developed standards and new information technology tools. As the MEDAR action provides an innovative and leading system for data exchange, it will represent a key module in the broader pan-European oceanographic data management system.

Acknowledgements

The MEDAR Group acknowledges with thanks the contribution of all the laboratories and scientists who have published data in the data centres and the European Union for the financial support to the project (contracts MAS3-CT98-0174 and ERBIC20-CT98-0103).

References

Fichaut M., M-J. Garcia, G. Maudire, M. Rixen, E. Vyazilov and MEDAR Group. 2002. Added values to oceanographic multi-parameters data sets. Example of European concerted action: MEDAR/MEDATLAS 2 Symposium Ensuring Long-Term Preservation and Adding Value to scientific and technical data. Toulouse 5-7 November 2002, France

Garcia M.-J., A. Nikouline, P. Alcantara and J. Molinero. 2001. QCMEDAR, MEDATLAS Format and Quality Control (V1.1). Users guide. Rap. Int. Instituto Espanol de Oceanografia IEODC.

Levitus S., T.P. Boyer, M.E. Conkright, T.O' Brien, J. Antonov, C. Stephens, L. Stathoplos, D. Johnson and R. Gelfeld. 1998: NOAA Atlas NESDIS 18,World Ocean Database 1998, US Gov. Printing Office, Wash., D.C., 346p + set of CD-ROMs.

Maillard C., E. Balopoulos, A. Giorgetti, M. Fichaut, A. Iona, M. Larour, A. Latrouite, B. Manca, G. Maudire, P. Nicolas and J.-A. Sanchez-Cabeza. 2002. An integrated system for managing multidisciplinary oceanographic data collected in the Mediterranean Sea during the basin-scale research project EU/MAST-MATER (1996-2000). Journal of Marine Systems 33-34:523-538.

Manca B., M. Burca, A. Giorgetti and C. Coatanoan, M.-J. Garcia and A. Iona. 2003. Physical and biochemical averaged vertical profiles in the open Mediterranean regions: an important tool to trace water masses climatology and to validate incoming data from operational oceanography. Submitted to J. Mar. Syst.

MEDAR Group. 2001. Medar-Medatlas Protocol (Version 3). Part I. Exchange format and quality checks for observed profiles. Rap. Int. IFREMER/TMSI/IDM/SIS002-006. 50p.

MEDAR Group. 2002. MEDAR/MEDATLAS 2002 Database. Mediterranean and Black Sea cruise Inventory. Observed data, analysed data and climatological atlas (4 CD-ROMs).

MEDATLAS Group. 1997. A Mediterranean hydrographic atlas from a composite quality checked temperature and salinity data (3 CD-ROMs). IFREMER Ed.

Rixen M., J.-M. Beckers, J.-M. Brankart and P. Brasseur. 2001. A numerically efficient data analysis method with error map generation. Ocean Modelling 2(1-2):45-60.

Rixen M., J.-M. Beckers, C. Maillard and the MEDAR Group. 2004. A hydrographic and biochemical climatology of the Mediterranean and Black Seas: some pittfalls. p. 227-243. In: Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002. Vanden Berghe E., M. Brown, M. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). IOC Workshop Reports 188 (UNESCO, Paris). x + 308 pp. – also published as VLIZ Special Publication 16.

Schlitzer R. 2000. Electronic atlas of WOCE hydrographic and tracer data now available. Eos Trans. AGU 81(5):45.

UNESCO. 1993. CEC/DG XII, MAST and IOC/IODE Manual of quality control procedures for validation of oceanographic data. UNESCO Manual and Guides 26. 436p.

Vyazilov E. and N. Puzova. 2001. Oceanographic research vessel cruises for the Mediterranean and Black Sea. RIHMI-WDC Rep.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 217-226
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Application of light attenuation measurement for the determination of vertical plankton distribution in seawater

Atanas V. Palazov

Institute of Oceanology, Bulgarian Academy of Science
PO Box 152, 9000 Varna, Bulgaria
E-mail: palazov@io-bas.bg

Abstract

The investigation of the horizontal and vertical distribution of phyto- and zooplankton communities is extremely important for the assessment of the ecological status of the marine environment, especially in the conditions of increasing anthropogenic influence. There is much evidence that during the last 20 years the Black sea coastal ecosystem is under the severe impact of eutrophication. As a response a dramatic increase of phytoplankton blooms was registered, accompanied by serious changes of its species succession; biological cycles and biomass, with corresponding alterations in zooplankton too. The high time-space variability, the patchiness of its horizontal distribution and vertical aggregation stress the necessity of reliable express methods for monitoring to be developed.
The increased abundance of plankton under the influence of high eutrophication of the Black Sea as a stratified basin, results in a dramatic change of the optical properties of the water masses especially in the coastal regions. Light attenuation in the water could be used for determination of vertical plankton distribution, abundance assessment and community aggregation down the water column.
A specially constructed device based on measuring of an optic system directed light beam attenuation in water has been used. The relationship between light attenuation (extinction) and total plankton abundance and chlorophyll 'a' fluorescence is evaluated in laboratory experiments on sea water samples with modelled plankton biomass. A good relationship between extinction coefficient as a measure of light attenuation and total phyto- and zooplankton biomass as well as chlorophyll 'a' fluorescence has been established. The experimental results were used for calibration of the device for *in-situ* application.
Two series of *in-situ* measurements were accomplished in the region of Varna Bay canal connecting Varna Lake and Varna Bay. During the first one five points were sampled and during the second – three points. At each point light attenuation was measured at depths from 0.5m to 8.5m with a step of 1m and water bottle samples were taken at three depths 0.5, 4.5 and 8.5m. The samples were processed by classical methods. In addition the chlorophyll 'a' (as a measure of total phytoplankton biomass) was analyzed on a Turner Design Fluorometer (model 10-000R).
As a result of the study a well-expressed and steady relationship between vertical distribution of extinction coefficient and chlorophyll 'a' measurement has been established. This gives ground to suggest that light attenuation could be used as an express method for determination of vertical plankton distribution in seawater. Applying 'express' and 'classic' methods of measurement in parallel could obtain more detailed results.

Keywords: Plankton; Chlorophyll 'a'; Optical properties; Light attenuation; Seawater.

Introduction

The investigation of the horizontal and vertical distribution of both phyto- and zooplankton communities is extremely important for the assessment of the ecological status of the marine environment especially in conditions of increasing anthropogenic influence. There is much evidence that during the last 20 years the Black sea coastal ecosystem is under the severe impact of eutrophication. As a response, a dramatic increase of phytoplankton blooms is registered, accompanied by serious changes of its species succession, biological cycles and biomass (Moncheva, 1991), with corresponding alterations in zooplankton too. The high time-space variability, the patchiness of its horizontal distribution and vertical aggregation stress the necessity of reliable express methods for monitoring to be developed (Cachru *et al.*, 1989).

The application of some hydro-optical measurements proves to give satisfactory results Karabashev *et al.* (1986). Among them the most reliable seems the assessment of light attenuation in the water column. Karabashev *et al.* (1986) for example have constructed and experimentally tested an autonomous underwater transparencymeter based on the measurement of collimated light beam attenuation. Li (1980) also focusses on light attenuation in the construction of an underwater transparencymeter. Lieberman *et al.* (1984) has established a good relationship between light transmission and chlorophyll ‘a’ fluorescence in the coastal waters off South California.

This paper presents the results of the application of light attenuation measurement for the determination of vertical plankton distribution in seawater.

Methods and materials

Method of measurement

The method of the plankton biomass determination is based on the light attenuation measurement from a source of a collimated beam and an optical system. At a distance from the light source, called optical base, there is a sensitive element for measuring the illumination. The transmission coefficient K is determined by comparison of the luminous flux passed through the air and sea water:

$$K = F / F_o$$

Where: Fo = the luminous flux through the air and F = the luminous flux through the water. The relationship between the illumination E and the normal incidence luminous flux F on a surface S is:

$$E = F / S$$

As the surface of the light beam in this case is constant then:

$$K = E / E_o$$

Where: Eo = illumination in the air and E = illumination in the water.

The transmission coefficient K is closely related to light absorption and dispersion in seawater and is identified also with the extinction coefficient. At a proper selected source power and optic base the illumination due to the natural sun light in seawater (both direct and dispersed) can be ignored.

Hydro-optical device

The panel diagram of the hydro-optical device used for light attenuation measurement Palazov (2001), is shown on Fig. 1. Power source (1) supplies the collimated light beam source (2), which consists of a lamp, a reflector and an optical system. At a given pathway L (optical base), a dry photocell (sensitive element) is fixed (3), which changes its resistance proportionally to illumination. A measuring device (4) measures dry photocell resistance. The optical base L of the hydro-optical device can be changed to obtain optimal measuring condition.

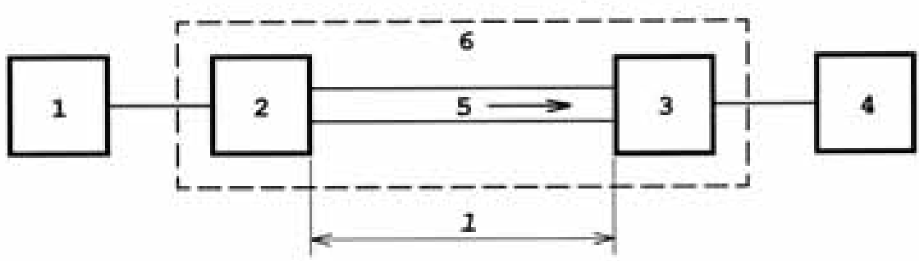


Fig. 1. Panel diagram of the hydro-optical device.

The device was constructed in the Institute of Oceanology, Varna and calibrated in the air by illuminometer type PU 150 through controlling the intensity of light source and measuring the illumination of the cell. The calibration curve and the corresponding equation are plotted on Fig. 2. As expected it is a power function.

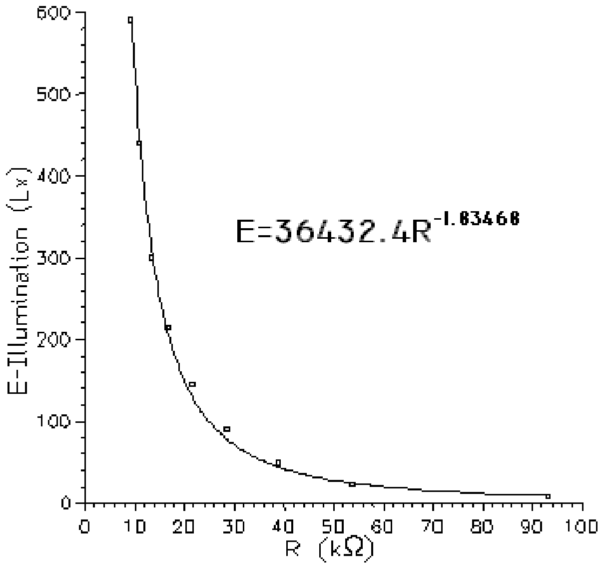


Fig. 2. Calibration curve of the hydro-optical device.

Optical device specifications are:

Range:	from 0 to 3200Lx
Resolution:	0.5Lx
Accuracy:	± 1Lx
Output signal:	max 6V
Resistance of the sensitive cell:	from 1 to 100kΩ

Laboratory experiments

The laboratory experiments with a constructed hydro-optical device were conducted on natural seawater with experimentally controlled phyto- and zooplankton biomass (Slabakov *et al.*, 1996; Palazov, 2001). The plankton was sampled in Varna Bay (by a Jeddy net for zooplankton) on the day of the experiment and diluted with filtered sea water several times, step by step, until the biomass becomes negligible. For each step, light extinction, total biomass and chlorophyll ‘a’ fluorescence were measured and estimated. Extinction coefficient was estimated by means of the constructed hydro-optical device, zooplankton biomass by Dimov's method (1959) and phytoplankton biomass by Soumia's method (1978). The total biomass is the sum of phyto and zooplankton biomass. Chlorophyll ‘a’ fluorescence (as a total phytoplankton biomass parameter) was measured by a Turner Design Fluorometer (model 10-OOOR).

The experimental medium covered the range of the natural plankton concentration. Temperature and salinity in the experimental bath were controlled during each experiment. The influence of temperature and salinity on the light extinction was also experimentally tested and was found to be practically negligible. The salinity range was between 10-35 in experimental series decreasing stepwise by 5 units and the temperature range was between 5-25°C, set by 5°C intervals.

The results of the two experiments are presented in Figs 3 to 6. On Fig. 3 the relationship between extinction coefficient and total plankton biomass is depicted and on Fig. 4 -the relationship between extinction coefficient and chlorophyll 'a' fluorescence for the first experiment.

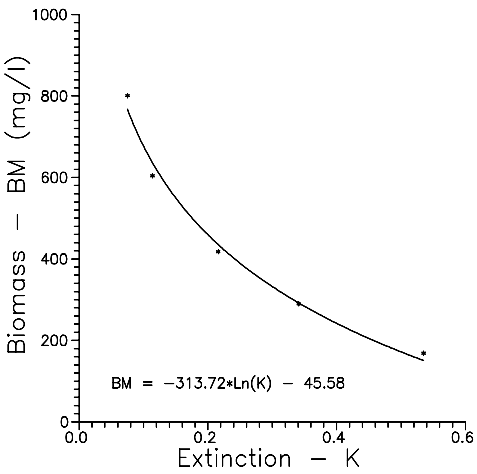


Fig. 3.
Relationship between extinction coefficient and total plankton biomass (first experiment).

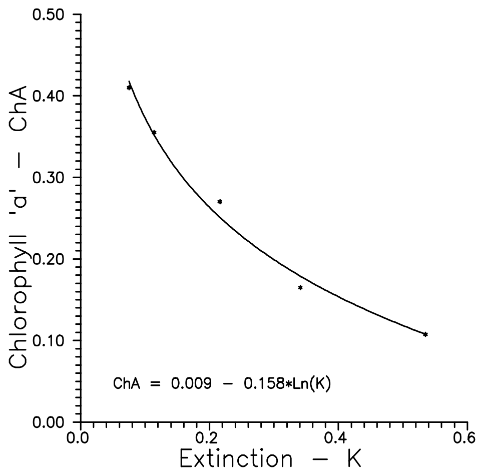


Fig. 4.
Relationship between extinction coefficient and chlorophyll 'a' fluorescence (first experiment).

Of the second experiment the relationship between extinction coefficient and total plankton biomass is presented in Fig. 5 and the relationship between extinction coefficient and chlorophyll 'a' fluorescence – in Fig. 6.

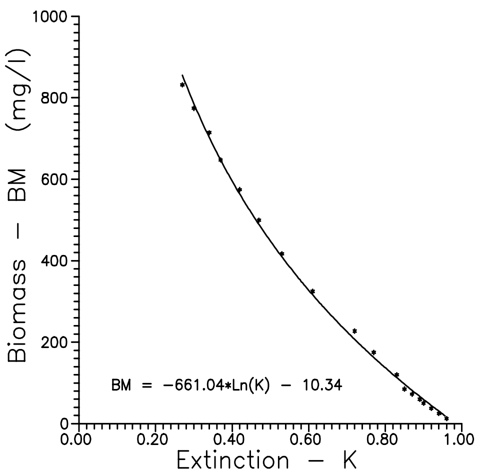


Fig. 5.
Relationship between extinction coefficient and total plankton biomass (second experiment).

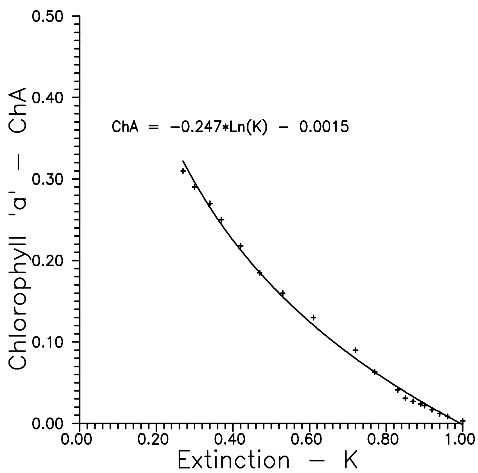


Fig. 6.
Relationship between extinction coefficient and chlorophyll 'a' fluorescence (second experiment).

In-situ *measurements*

Two series of *in-situ* measurements were accomplished in Varna Bay and in the canal connecting Varna Lake and Varna Bay. During the first one five stations were sampled and during the second – three stations, selected to cover a wide range of plankton concentration. The stations from 1 to 5 and from 6 to 8 are located along a line from open sea to Varna Lake. The scheme of sampling stations is shown in Fig. 7.

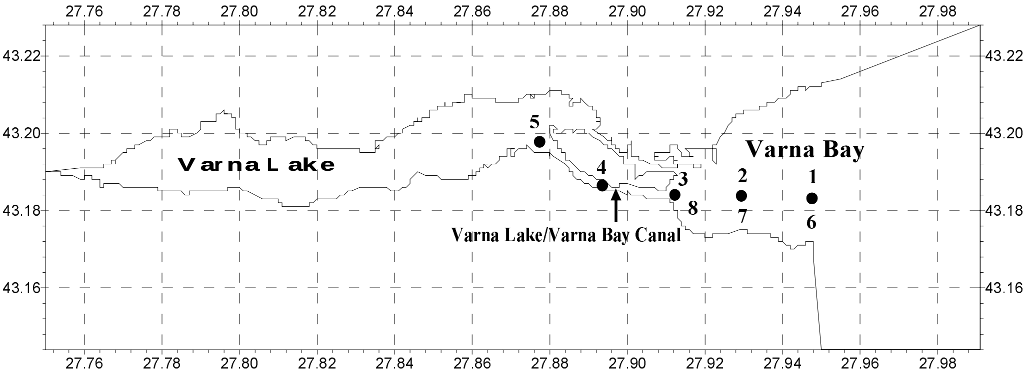


Fig. 7. Scheme of sampling stations.

At each station light attenuation was measured at depths from 0.5 to 8.5m with a step of 1m down the vertical profile. Water samples were taken by water bottles at depths of 0.5, 4.5 and 8.5m in parallel. The samples were processed in the biological laboratory of the Institute of Oceanology, Varna, on the day of sampling. The phytoplankton biomass was estimated by the method of Sournia (1978). Chlorophyll 'a' was estimated by the method of SCOR UNESCO (1966). The chlorophyll 'a' fluorescence (as a total phytoplankton biomass parameter) was measured on a Turner Design Fluorometer (model 10-000R).
The taxonomic composition of phytoplankton in Varna Bay at the time of the first experimental series is presented in Table I.

Table I. Phytoplankton taxonomic composition

Class/Species	Numerical abundance [cells.l^{-1}]	Biomass [ml^{-1}]
Bacillariophyceae		
Ditylum brightwellii	15000	1.530
Skeletonema costatum	320000	0.096
Thalassiosira parva	27000	0.054
Cyclotella caspia	14000	0.008
Thalassionema nitzschioides	8000	0.010
Cerataulina pelagica	7000	0.046
Pseudosolenia calcar-avis	1000	0.330

Class/Species	Numerical abundance	Biomass
Pseudonitzschia delicatissima	2000	0.001
	394000	2.075
Dinophyceae		
Scripssiella trochoidea	2000	0.026
Prorocentrum minimum	166000	0.332
Prorocentrum micans	1000	0.013
Gyrodinium fusiforme	200	0.009
Oxyphysis oxytoxoides.	13000	0.585
	182200	0.965
Cyanophyceae		
Phormidium sp.	5000	0.001
TOTAL:	581200	3.041

The results of the first series of measurements (stations 1 to 5) are presented in Table II. Table III lists the results of the second series of measurements (stations 6 to 8).

Table II. Results of the first series of *in-situ* measurements

Point	1	1	2	2	3	3	4	4	5	5
Depth	1- K	Chl*10	1- K	Chl*10	1- K	Chl*10	1- K	Chl*10	1- K	Chl*10
0.5	0.2087	0.130	0.2819	0.260	0.3262	0.300	0.3195	0.285	0.2742	0.250
1.5	0.2094		0.2812		0.3376		0.2673		0.2533	
2.5	0.2032		0.2483		0.3016		0.2533		0.2232	
3.5	0.1912		0.1944		0.2388		0.2008		0.2164	
4.5	0.1732	0.079	0.1699	0.090	0.2086	0.150	0.1732	0.150	0.211	0.140
5.5	0.1597		0.1415		0.1952		0.1631		0.2032	
6.5	0.1529		0.1500		0.1511		0.1700		0.1900	
7.5	0.1597		0.1326		0.1500		0.1335		0.1600	
8.5	0.1424	0.048	0.1250	0.040	0.1400	0.080	0.1200	0.098	0.1550	0.095

1-K: measured extinction coefficient subtracted from one.
Chl*10: measured chlorophyll ‘a’ fluorescence multiplied by ten.
The transformations were made for better illustration of the results (Figs 8 to 15).

Table III. Results of the second series of *in-situ* measurements

Point	6	6	7	7	8	8
Depth	1- K	Chl*10	1- K	Chl*10	1- K	Chl*10
0.5	0.2117	0.140	0.2954	0.250	0.2352	0.165
1.5	0.2277		0.3070		0.2410	
2.5	0.2086		0.3170		0.2381	
3.5	0.2613		0.3195	0.288	0.2694	0.250
4.5	0.2673	0.160	0.3213		0.3213	
5.5	0.2742		0.3100		0.3006	
6.5	0.2468		0.3006		0.3041	
7.5	0.2388		0.3177		0.3177	
8.5	0.2200	0.120	0.3133	0.280	0.3300	0.300

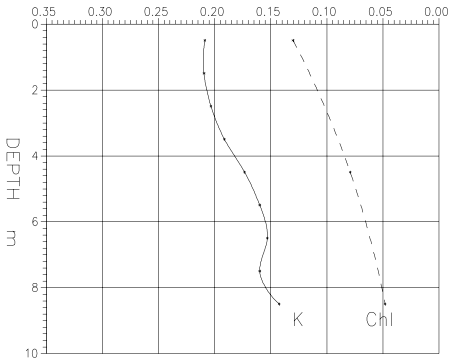


Fig. 8. Station 1.

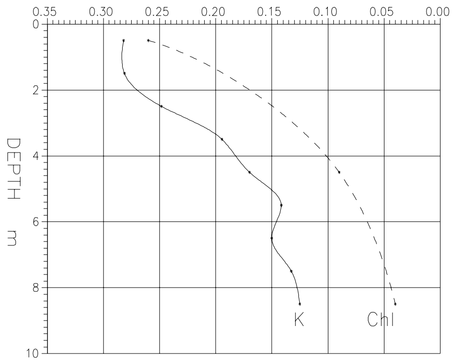


Fig. 9. Station 2.

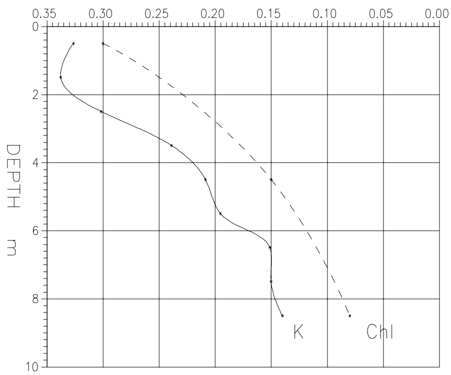


Fig. 10. Station 3.

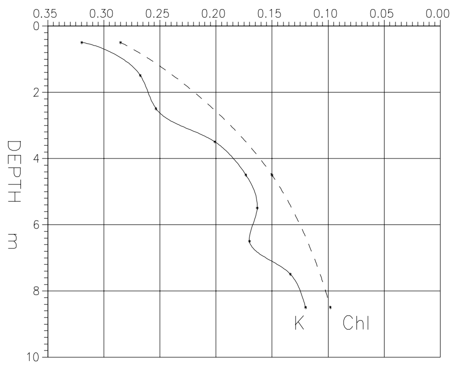


Fig. 11. Station 4.

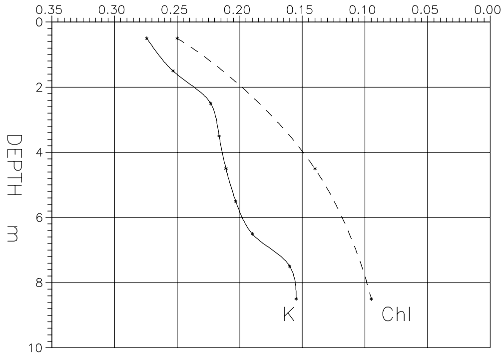


Fig. 12. Station 5.

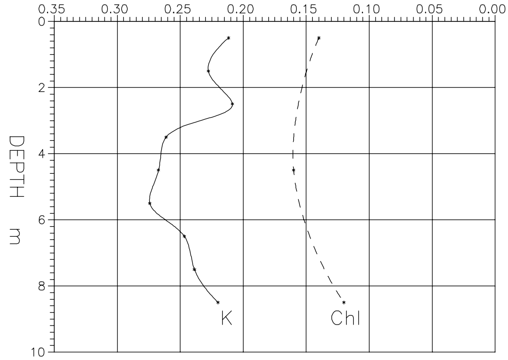


Fig. 13. Station 6.

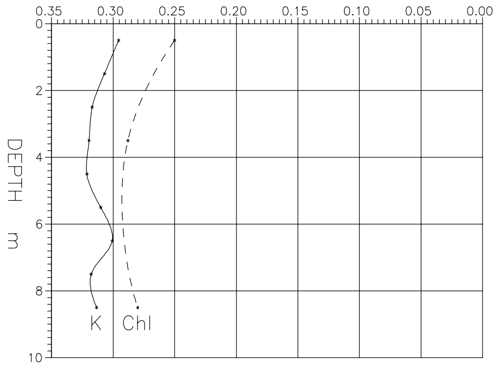


Fig. 14. Station 7.

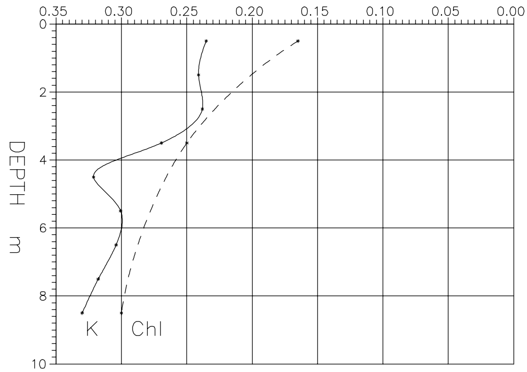


Fig. 15. Station 8.

K (solid line): measured extinction coefficient; Chl (dashed line): measured chlorophyll ‘a’ fluorescence.

Results and conclusions

During the laboratory experiments a well-expressed steady relationship between extinction coefficient and both total plankton biomass and chlorophyll ‘a’ fluorescence was established. The relations are logarithmic, which is in conformity with the results of other investigations (Lieberman *et al.*, 1984). The differences between the two experiments may be related to the differences of the species composition (different size composition) as well as to the differences in the phyto/zooplankton ratio. A probable reason could be the presence of the Ctenophore *Mnemiopsis leidyi*, which has not been estimated.

The results from *in-situ* measurements manifest generally a good agreement between variation of extinction coefficient and chlorophyll 'a' fluorescence. During previous experiments it was demonstrated that the phytoplankton influence on the extinction is much stronger than that of zooplankton (Palazov, 2001).

The established relationships between the total plankton biomass and chlorophyll 'a' fluorescence with extinction coefficient give ground to consider the method applicable for determination of the phyto/zooplankton ratio.

Acknowledgements

The study was supported by the Bulgarian National Scientific Research Fund (Project B-67) and CESUM-BS Project (ICA1 – IT-2000-70038). The author would like to thank Dr. Slabakov, Dr. Konsulov and Dr. Trayanov for the research collaboration and especially Dr. Snejana Moncheva for the valuable help and comments during the preparation of the paper.

References

Cachru M.S., S. Chesmann and E. Allickas. 1989. Complete equipping for investigation of the space distribution of phytoplankton. Moscow. Oceanology XXIX(4):670-674.

Karabashev G.S., A.F. Culeshov and S.A. Shanaaev. 1986. Autonomous transparencymeter. Moscow. Oceanology XXVI(6):1028-1032.

Li M.E.1980. Submersible autocollimated transparencymeter. p.2911-295. In: Optical methods for studying oceans and land basins. Talin Inst. of Thermo and Electrophysics AS USSR.

Lieberman S.H., G.D. Gilbert, P.F. Seligman and A.W.Dibelka. 1984. Relationship between chlorophyll-a fluorescence and underwater light transmission in coastal waters off Southern California. Deep Sea Research 31(2):171-180.

Moncheva S. 1991. Eutrophication/Plankton blooms/Hypoxia. Rep. Int. Workshop on the Black Sea: Focus on the Black Sea Shelf, Varna, 30 Sept.-4 Oct., Coast. Res. Center WHOI

Palazov A.V. 2001. An express method for plankton abundance detection in sea water. Proceedings of the Institute of Oceanology 3:180-184. Varna.

Slabakov Ch., S. Moncheva, A. Palazov, A. Konsulov, T. Trayanov and N. Nickolov. 1996. A possible express method for plankton abundance in sea water. p.253-258. In: Coastal management and habitat conservation. Salman A.H.P.M., M.J. Langeveld and M.Bonazowntas (Eds). EUCC, Leiden, The Netherlands.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 227-243
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

A hydrographic and biochemical climatology of the Mediterranean and the Black Sea: some statistical pitfalls

Michel Rixen[1], Jean-Marie Beckers[2], Catherine Maillard[3] and the MEDAR Group

[1]Southampton Oceanography Center, European Way
SO14 3ZH, Southampton, United Kingdom
E-mail: myr@soc.soton.ac.uk

[2]GHER B5, Institut de Physique, University of Liège
B-4000 Liege, Belgium

[3]SISMER, IFREMER, Centre de Brest
Boîte Postale 70, F-29280 Plouzané, France

Abstract

The aim of the MEDAR/MEDATLAS II project was to archive and rescue multidisciplinary in-situ hydrographic and biochemical data of the Mediterranean and the Black Seas through a wide cooperation of countries and to produce a climatological atlas of 12 core parameters, which include temperature and salinity, dissolved oxygen, hydrogen sulfure, alkalinity, phosphate, ammonium, nitrite, nitrate, silicate, chlorophyll and pH. Gridded fields have been computed using the Variational Inverse Model and calibrated by Generalised Cross Validation, by making usual assumptions on the statistical distribution of data and errors. They have been produced for both entire Mediterranean and Black Seas and several additional sub-basins including the Alboran Sea, the Balearic Sea, the Gulf of Lions and the Ligurian Sea, the Sicily Strait, the Adriatic Sea, the Aegean Sea, the Marmara Sea and the Danube shelf area at climatic, seasonal and monthly scale when relevant. Inter-annual and decadal variability of T/S for both basins has been computed as well. The resulting atlas is made available free of charge at http://modb.oce.ulg.ac.be/Medar and on CD-Rom. We review here the different biases that might occur when one of the statistical hypotheses is not satisfied, and suggest further possible improvements to the analysis method.

Keywords: Objective analysis; Statistical pitfalls; Mediterranean Sea; Black Sea.

Introduction

The Mediterranean and the Black Sea have been studied actively for several decades (e.g. Brankart and Pinardi, 2001; Karafistan *et al.*, 2002), generating huge amounts of data that remain in danger of being lost if not archived. New multivariate sensors, in growing number and spatio-temporal resolution produce crucial information which require careful but efficient integration. This information is becoming an essential part of the numerous toolkits used by oceanographers, engineers, managers, navies and authorities to continuously monitor the ocean state and variability, to infer possible climate changes, etc. (e.g. Beckers *et al.*, 2002).

The aim of the MEDAR/MEDATLAS II project is to archive and rescue multidisciplinary in-situ hydrographic and biochemical data of the Mediterranean and the Black Sea through a wide cooperation of countries. The growing interplay between scientific disciplines requires multidisciplinary integration of this information which includes, among many other parameters, temperature and salinity, dissolved oxygen, hydrogen sulfur, alkalinity, phosphate, ammonium, nitrite, nitrate, silicate, chlorophyll and pH.

The project was divided into several tasks. First, a global inventory was compiled using existing data sets of the core parameters and their cruise reports from the WDC-A, the WDC-B, the ICES and the MEDATLAS I inventories. Duplicates were eliminated by careful cross-checking. Secondly, the data sets were assembled regionally for the Western, Central and Eastern Region and the Black Sea sub-areas, transcoded in the common ASCII human readable MODB/MEDATLAS format and quality checked by regional experts. Finally, a global integration ensured the assembling and consistency of the regional data sets. The vertical profiles were interpolated on 25 standard vertical levels, chosen according to the vertical distribution of the data (Maillard *et al.*, 2004).

Raw in-situ datasets are difficult to interpret and higher level products are required to offer a more complete and synthetic view of the Mediterranean and Black Sea hydrographic and biochemical systems. The computation of climatological or gridded fields is also justified by the need to provide initial or boundary conditions to numerical models (e.g. Beckers *et al.*, 2001) and is subject to the choice of an adequate analysis method.

In the following sections, we give some details on the objective analysis method, we review potential problems that may appear when the statistical hypotheses of the analysis are not fully satisfied and we provide some insight into the results of the new climatology. Finally, some conclusions and perspectives are drawn.

Methodology

Instead of using the classical objective analysis scheme (OA) (e.g. Bretherton *et al.*, 1976), gridded fields have been computed using the Variational Inverse Model (Brasseur, 1991; Brasseur *et al.*, 1996; Brankart and Brasseur, 1998; Brankart and Pinardi, 2001; Rixen *et al.* 2001), shown to be statistically equivalent to OA.

The basic idea of variational analysis is to determine a continuous field approximating the data and exhibiting small spatial variations. In other words, the target of the analysis f is defined as the smoothest field that respects the consistency with the observed values over the domain of interest. It is also referred to as a spline interpolation method. Expressed in mathematical terms, the analysis is obtained as the minimum of a variational principle (in a two-dimensional, horizontal space):

$$
\begin{aligned}
J[\phi] &= \sum_{i=1}^{N_d} \mu_i \left[d_i - \phi(\mathbf{x}_i)\right]^2 + \|\phi - \phi_b\|^2 \\
\|\varphi\|^2 &= \int \left(\nabla\nabla\varphi : \nabla\nabla\varphi + \alpha_1 \nabla\varphi \bullet \nabla\varphi + \alpha_0 \varphi^2\right) d\mathbf{x}
\end{aligned} \qquad (1)
$$

The integral extends over the whole domain. The first contribution in J represents the distance between the data and the target field at the exact position of the observations. The weights mi are determined according to the confidence in the data. In principle, the weights could be adjusted to every observation individually, but in practice it is difficult to decide whether one observation is more reliable than another.

The second contribution in J is a measure of the smoothness of the target field. The coefficients a1 and a2 fix the weights of the lower derivatives in the smoothing operator. In practice, all observations at a given level are selected to perform the minimization in a horizontal plane. The reconstruction of a three-dimensional scalar field is then obtained as a superposition of several analyses at different standard depths. This reduction is made possible because the data profiles describe the vertical structure of the sea reasonably well and do not produce hydrostatic instabilities. For bio-chemical data, the structure of profiles can however be more complex.

The reference field ϕ_b (or background field) has been computed by a semi-normed analysis (leaving out the underived term in J) (Brankart and Brasseur, 1996). It has indeed been shown that in data-void areas this method produces more realistic fields than using a simple constant value or a linear regression of the data. The variational principle is solved using a finite element technique. The main advantage is a numerical cost almost independent of the number of data analyzed (Rixen et al, 2001), as compared to optimal interpolation. The mesh also easily takes into account the complexity of the basin geometry by automatically prohibiting correlations across land barriers (Brankart and Brasseur, 1998).

By converting the dimensional weights of the variational principle into non-dimensional quantities, and imposing that the correlation function be the Modified Bessel function of order 1, the calibration of the free parameters m, a1 and a2 reduces to the choice of a characteristic length scale L and a signal-to-noise ratio (S/N=e2/s2), which represents the ratio of the standard deviation of the signal e to the standard deviation of observational errors s. A Generalized Cross-Validation (GCV) technique has been implemented to compute the statistical parameters L and S/N out of the raw data (Brankart and Brasseur 1996). The method consists in successively eliminating one measurement from the full database and performing analyses with the remaining data. The variance of the misfits between these reconstructed fields and the corresponding eliminated data is then considered as a statistical indicator of the quality of the analysis, with respect to S/N and L. In practice, for climatological analyses, the quality of the analysis is not very sensitive to the correlation length and has thus been fixed a priori, according to a reasonable choice of features to be represented in the domain of interest.

By analogy with OA, the statistical error field is obtained by using the stiffness matrix K of the analysis and two transfer operators T1 and T2 (Rixen *et al.*, 2001)

$$e^2(x) = \varepsilon^2(x) - T_1(x)K^{-1}T_2(x)c(x) \qquad (2)$$

where c is the correlation function, i.e. the Modified Bessel function of order 1.

Statistical pitfalls and related solutions

The variational principle in eq. (2) relies on several statistical assumptions. Among others, it assumes that the frequency distributions of the noise and the signal are gaussian [assumption 1], that the statistics are homogeneous and isotrope [assumption 2] and that the noise is uncorrelated [assumption 3]. Asssumption 1 is generally speaking related to the amount of available information as an observed phenomenon will usually show a gaussian frequency distribution as long as the amount of data is sufficient. Assumption 2 is generally valid except the fact that hydrodynamics usually stretch the features along circulation paths. This problem can be easily circumvented by adding an advection constraint into eq. (2) for example (e.g. Brankart and Brasseur, 1996). This problem occurs frequently in coastal areas and has to be dealt with carefully. Assumption 3 is also generally valid unless they are systematic biases in the measurements that can be more easily detected.

Problems related to assumption 1

In Figs. 1 and 2, we show, respectively the distribution of temperature data in the Alboran basin and in the Ionian Sea at 200m over a 1°x1° square box. Note that even at 200m the data distribution is not strictly gaussian for the Ionian case. This problem may then lead to biases in the background field and hence in the signal itself and subsequently in the solution.

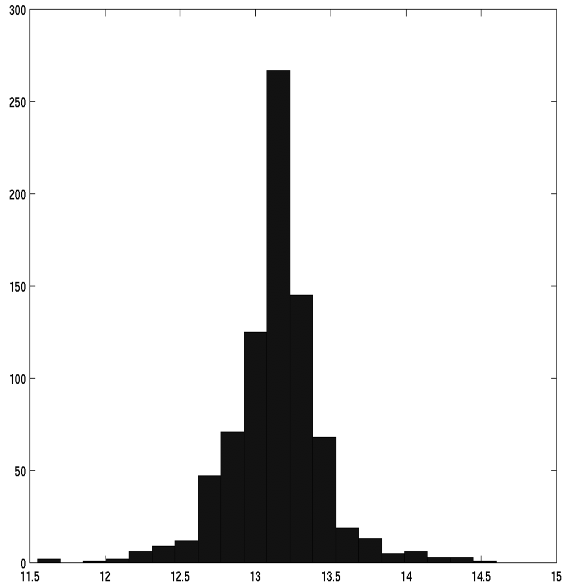


Fig. 1. Distribution of temperature data in the Alboran basin at 200m over a 1°x1° square box. Note the gaussian shape of the frequency distribution.

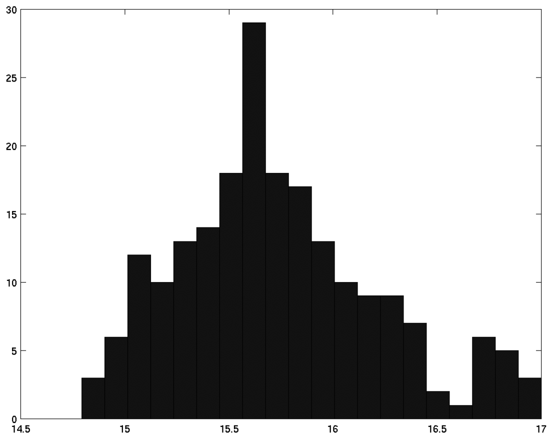


Fig. 2. Distribution of temperature data in the Ionian basin at 200m over a 1°x1° square box. Note the bi-modal shape of the frequency distribution.

A second problem that can appear is the weak spatial coverage in certain areas, where the solution field has to be extrapolated either from adjacent available stations or eventually from upper/lower available values when a 3D interpolation is considered. For the salinity for example, although the coverage including all seasons is very good (see Fig. 3), there might be areas void of data for specific months, even at the surface. In the case of silicate for example (see Fig. 4), the distribution is very patchy and concentrated in areas of interest, like the Gulf of Lions, the Ligurian Sea, the Gibraltar Strait, … but the areas in between have to be extrapolated from these patches.

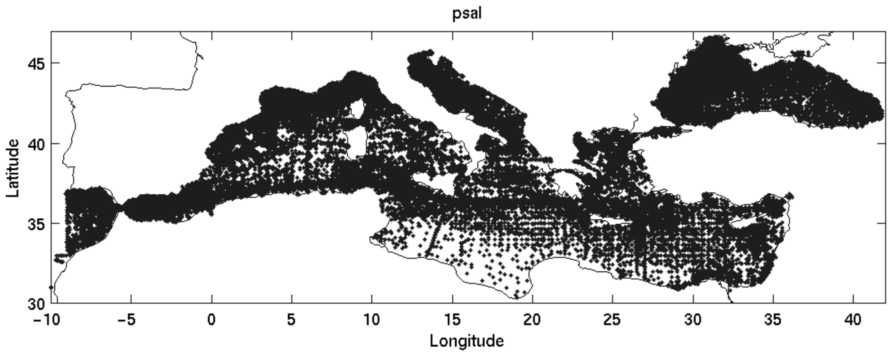


Fig. 3. Station distribution of salinity. Note the nice data coverage except along the Tunisian coast. The monthly distribution (not shown) is however more patchy.

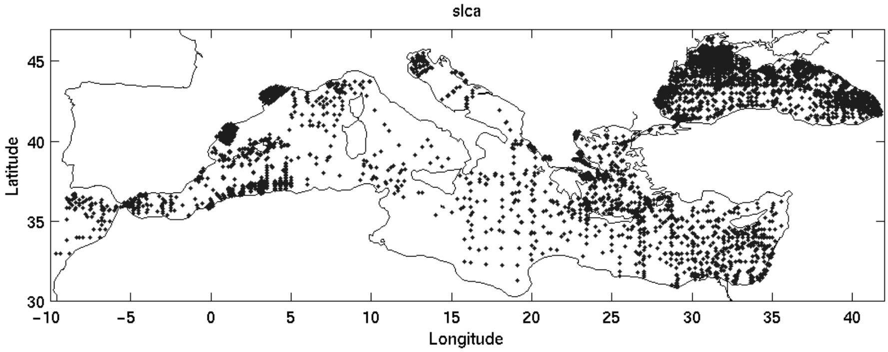


Fig. 4. Station distribution of silicate. Note the patchiness of the distribution and some large areas void of data where extrapolation will be needed in the analysis.

On the vertical, the availability of data rapidly decreases from the surface towards the bottom as seen from Fig. 5 for temperature for example. As the typical correlation length in the Mediterranean varies between 100 and 500km, a homogeneous distribution would require at least ~1200 to 50 data for each horizontal level. As the geometry of basin is very complex, these figures are however surely too optimistic.

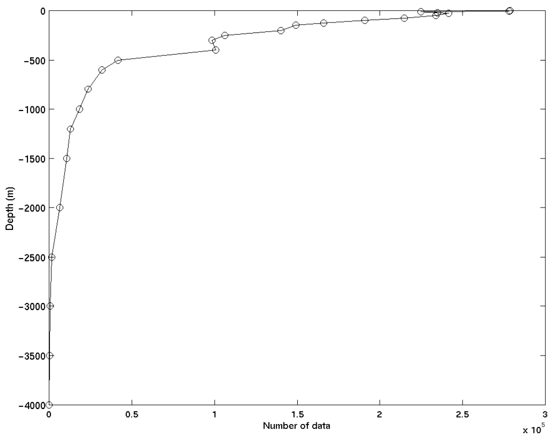


Fig. 5. Number of temperature data as a function of depth. Note that although the number of stations at the surface is impressive, it rapidly decreases towards the bottom.

The MEDAR database includes data over the 1900-2002 period, but most of the data have been collected after 1945 (see Fig. 6). A climatic analysis will then not properly represent the mean state of the century but most likely a state close to the period referring to the second half of the century.

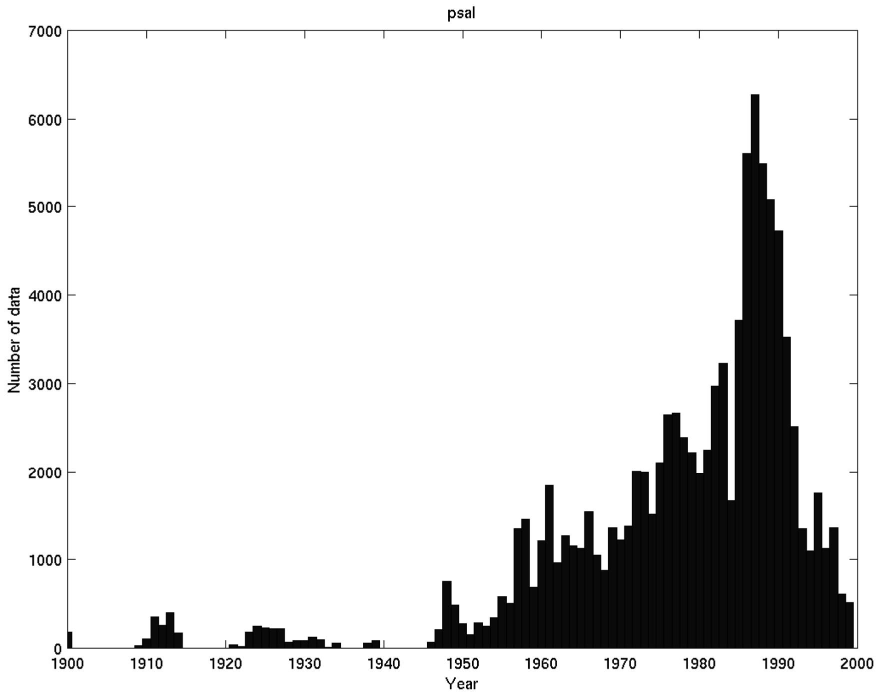


Fig. 6. Yearly distribution of salinity profiles from 1900 to 2000. Note the data-void periods during the two world wars. Note also the constant increase in available data during the second half of the century and the decrease of the availability from 1990 on due to the delay of releasing the information.

For some parameters, the distribution over months is not uniform, and seriously biased towards high summer frequencies and low winter frequencies (see Fig. 7 for example). A climatic temperature analysis for example may then bias the mean towards higher temperature locally.

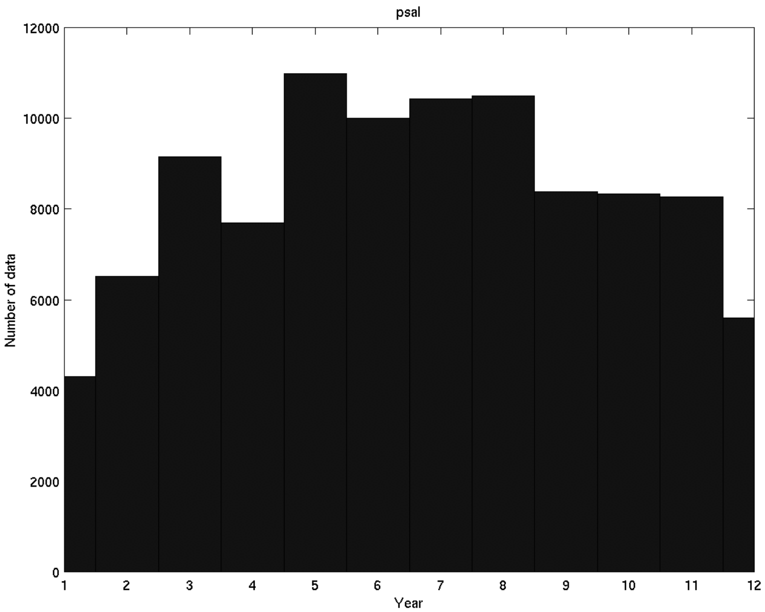


Fig. 7. Monthly distribution of salinity profiles. There are more data available in summer than in winter.

For the Ionian sea, temperature has been sampled during the following years mainly during the following months, where '1' stands for 'January', '2' for February, etc.:

- 1980 Months 2
- 1986 Months 3, 4, 9
- 1988 Months 7
- 1990 Months 10, 11
- 1992 Months 5
- 1994 Months 1

Similarly, for the Levantine basin, the available data are:

- 1984 Months 10
- 1986 Months 8, 9, 10, 11
- 1988 Months 3, 8, 9
- 1990 Months 7, 10, 11
- 1994 Months 1, 2

In both cases, the available data could result in the analysis in an increase of temperature for the period 1980-1988 followed by a cooling period during 1990-1994.

A final possible bias is illustrated in Fig. 8 where we show a vertical section for ammonium, a parameter whose values are at the limit of instrumental errors and for which very few data are available. In this case the 2D horizontal analysis results in a 'layered' solution, although the original profiles are smoother.

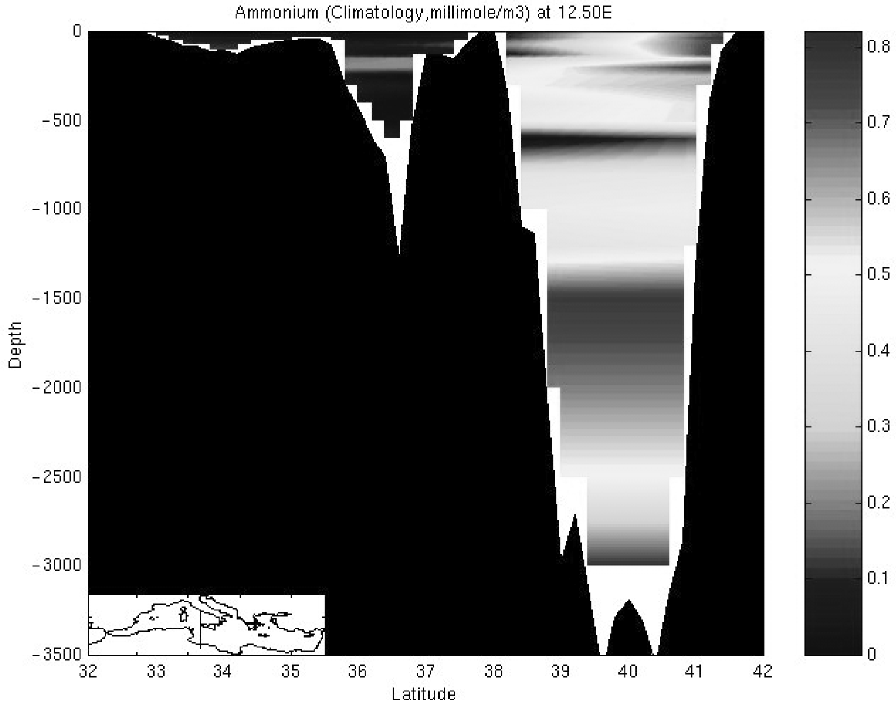


Fig. 8. Vertical section of ammonium in the Tyrrhenian Sea: an illustration of a possible bias due to the horizontal analysis approach. An obvious improvement would be to extend the analysis scheme to a 3D method.

Problems related to assumption 2

For the preliminary version of the climatology, a preprocessing quality check (QC) was applied to the raw data in order to eliminate values out of 3 standard deviations on bins 5-10 times bigger than the analysis grid. This method provided a final QC on the data and ensured that most potential bulls' eyes were removed in the analyses when the statistical distribution of data was typically gaussian. This was surely the case for parameters like T or S at levels close to the surface. However, at deeper levels or for parameters less frequently sampled, this pre-processing was rejecting a significant amount of useful information when the statistical distribution was not strictly gaussian. As the major discrepancies were only found for coastal areas or areas with shallow waters, this preprocessing was not considered anymore.

Coastal data are sometimes the only information available but they might be significantly affected by coastal processes, including resuspension of sediments and anthropogenic activities. In preliminary versions of the climatology, coastal data have been used and have led to misinterpretation of the resulting gridded fields, especially regarding nutrient parameters. There is thus a difficult compromise to be made between the use of all data and non-coastal data only. Coastal data closer than 15 km from the coast or in areas shallower than 50m were rejected. These values have been chosen a priori but correspond roughly to the horizontal and vertical radius of influence of land and bottom respectively, when focusing on basin and regional processes. It was effectively shown that this preprocessing was far more efficient in removing artificial features in the analyses.

In Fig. 9, we show the station distribution for phosphate at 30m depth. The 'big dots' stations (3317) have been rejected by the pre-processing either because they are closer than 15km from the coast, either because the associated sounding depth is shallower than 50m. The 'small dots' stations (12261) are those that are effectively used in the analysis. In this example, rejection occurs mainly along Algeria, Spain, Egypt and Israel, around the Balearic islands, in the Aegean and Alboran Sea and in the Northern Adriatic. The impact of this pre-processing is shown in Fig. 10. On the top figure, the phosphate field using all available data shows clear biases in the analysis with high values extending anomalously far from the coast. Indeed, the interpolation 'spreads' these errors off-shore by the correlation function influence. The analysis of phosphate without the rejected data (middle figure) shows a clear reduction of these biases. The differences between the two fields (bottom figure) are located in areas where the data have been rejected. The corrections are particularly visible in the Almeria-Oran area, in the Adriatic, the Aegean, around the Balearic island, and along Egypt/Israel and Spain. But one notes also significant changes in the open ocean. For temperature or salinity, the artificial influence of rivers or other local processes is also reduced. This pre-processing thus prevents coastal data from creating artificial features far from the coast.

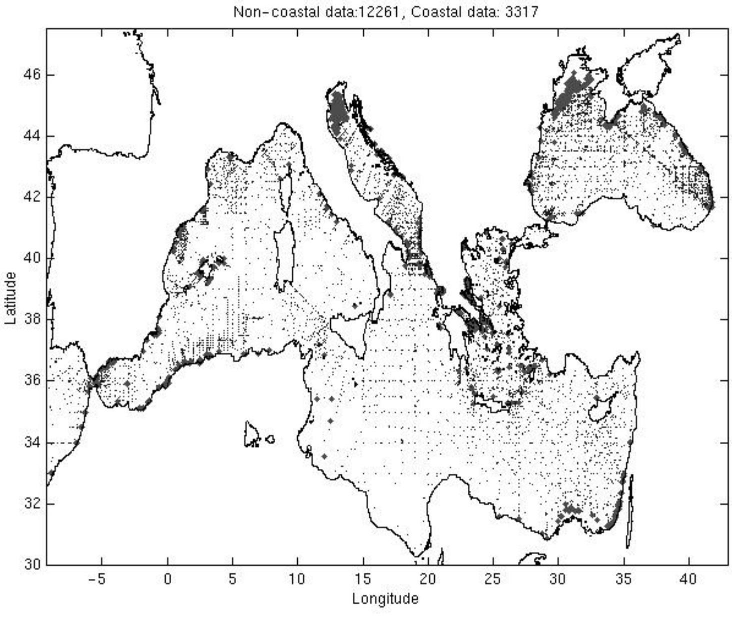


Fig. 9. Station distribution of phosphate at 30m depth used in the analysis (small dots). Coastal data and data with sounding shallower than 50m (big dots) have been rejected by the pre-processing.

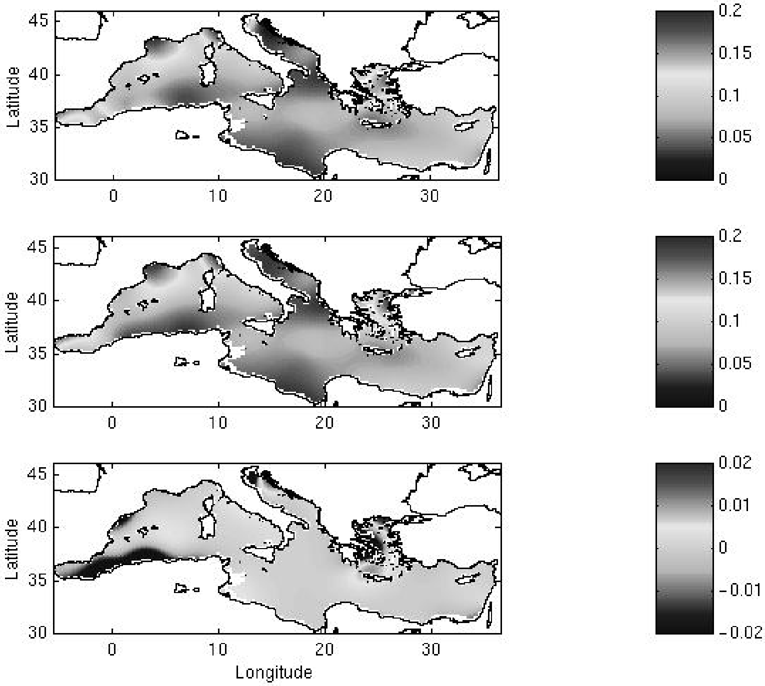


Fig. 10. Phosphate field (millimole/m3) at 30m depth. Top: all available data are used. Middle: analysis without coastal data. Bottom: difference between both analyses.

Problems related to assumption 3

Obvious instrumental error can sometimes be easily detected as seen for example in the T/S diagram (Fig. 11) for the Ionian Sea in the area 36-37ºN, 19-20ºE. Unless these data are flagged properly, that will induce serious biases in the analysis.

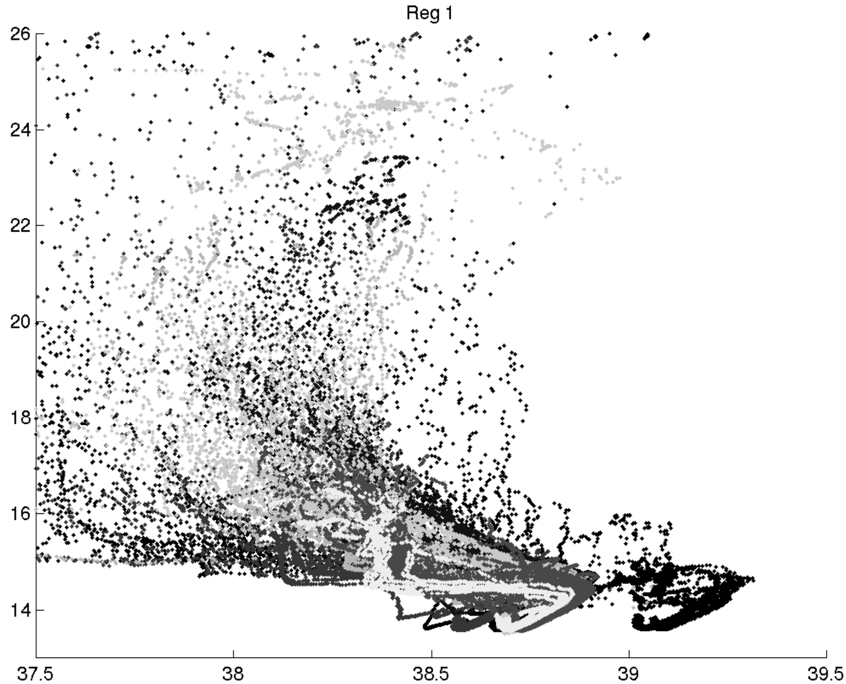


Fig. 11. T/S diagram for the Ionian Sea in the area 36-37 ºN, 19-20ºE. The different gray scales correspond to different field experiments. Note the clear shifts in salinity due to calibration problems. Most of these problems have been identified and corrected for the final MEDAR database.

Results

Gridded fields have been produced for both the Mediterranean and Black Sea basins and several additional subbasins including the Alboran Sea, the Balearic Sea, the Gulf of Lions and the Ligurian Sea, the Sicily Strait, the Adriatic Sea, the Aegean Sea, the Marmara Sea and the Danube shelf area at climatic, seasonal and monthly scale when relevant. Inter-annual and decadal variability of T/S for both basins has been computed as well. It should be noted that for a given temporal scale, all data have been used, but the observational error standard deviation has been modified according to Brankart and Pinardi (2001), by imposing a temporal correlation length: 3 months for a seasonal analysis, 1 month for a monthly analysis, 10 years for a decadal analysis and 1 year for an annual analysis.

In practice, the usual statistical hypotheses of OA are not fully satisfied, either because some (potentially big) areas are void of data, or because the distribution of the signal and/or noise is not gaussian (except for T/S and maybe dissolved oxygen). Therefore, the spatial correlation length has been increased artificially for consistency of the gridded fields at basin and regional scale. As a result, a posteriori, only gridded fields that have been computed by 'robust' data distributions have been selected. They include:

- temperature and salinity at climatic, seasonal and monthly scale
- dissolved oxygen, phosphate and silicate at climatic and seasonal scale
- nitrate, nitrite, alkalinity, pH, ammonium at climatic scale

The resulting atlas is made available free of charge at 'http://modb.oce.ulg.ac.be/Medar' and on CD-Rom, where the fields are found in Netcdf format (compressed with the gzip tool), and figures in Jpeg format. Results are organized in a 'tree' way: first identify the domain, then the parameter, then the relevant period. To illustrate the climatology, some examples of the Mediterranean and the Black Sea are shown below. For the corresponding color figures, the reader is invited to visit the web site.

Figs. 12 and 13 show horizontal sections of climatological temperature and alkalinity at 100m depth in the Mediterranean, with a clear signature of the Mediterranean inflow, the Alboran Western and Eastern gyres, the gradient of temperature from west to east, the Rhodes gyre, the outflow of the Black Sea. The continuous evaporation taking place over the Mediterranean, together with the regional topography and local intermittent formation of deep water and intermediate water are responsible for these features, well reported in the literature.

Fig. 14 and 15 show horizontal sections of climatological salinity and oxygen at 50m depth in the Black Sea, with a clear signature of the typical dipole cyclonic circulation, and the saltier and less oxygenized in the core of these cyclones. This general circulation of the Black Sea and the associated distribution of bio-geochemical parameters are resulting from the atmospheric forcing over the basin.
Provided physical and biochemical processes are sampled adequately, i.e. with enough accuracy in time and space, they will appear in the analyzed fields. The examples shown are of course only a tiny fraction of exploitable results of the climatological atlas that has been produced. The whole climatology currently represents some 2000 fields, 20000 figures and 600 animations.

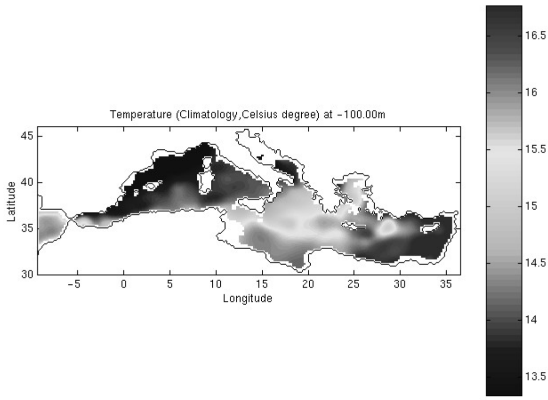


Fig. 12. Climatological temperature in the Mediterranean at 100m depth.

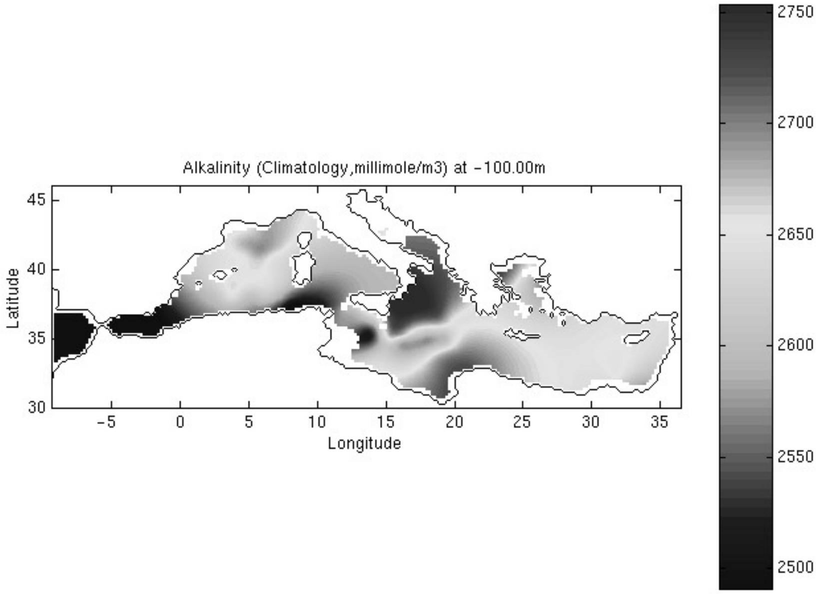


Fig. 13. Climatological alkalinity in the Mediterranean at 100m depth.

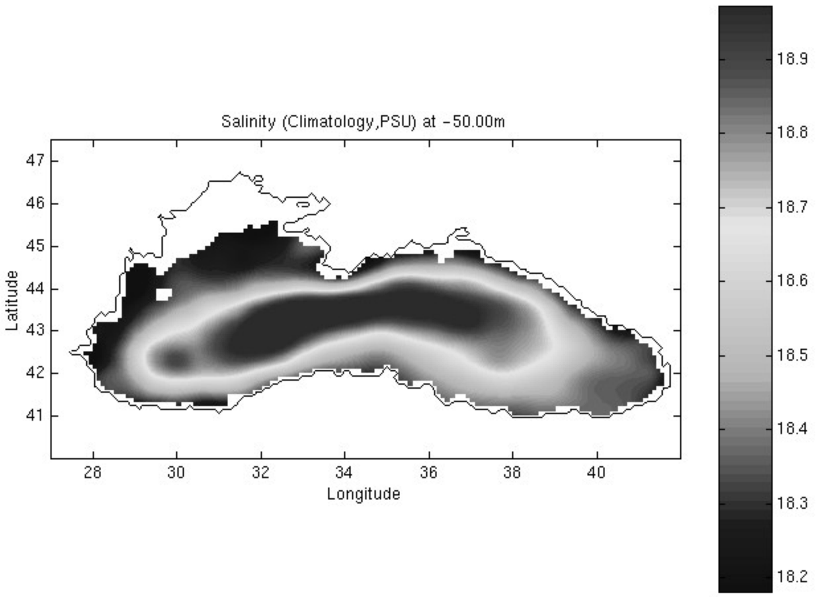


Fig. 14. Climatological salinity in the Black Sea at 50m depth.

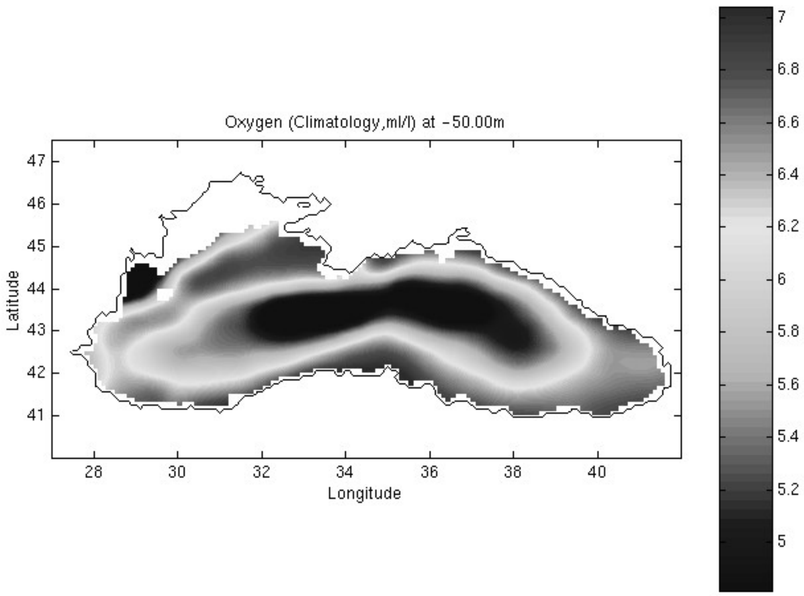


Fig. 15. Climatological oxygen in the Black Sea at 50m depth.

Conclusions

In this paper, we have described the methodology used in the gridding task of the MEDAR/MEDATLAS project. The critical technical choices made during the project have been justified following a constant feedback between data providers, regional centers, the global assembling center and the data analysis center. This feedback loop, actually 3 rounds, has allowed to refine the pre-processing steps and the data analysis technique in order to enhance the quality of the climatology throughout the project.

The quality of the analyses is highly dependent upon the validity of the underlying statistical hypotheses. In the future, more powerful interpolation schemes should investigate the usefulness of multivariate and/or 3D interpolation schemes that effectively take into account the vertical correlation present in the profiles of raw parameters. Systematic decimation of the data has not been possible as this method might result in other biases when the data coverage is very weak.

The issue of coastal data is particularly critical, especially for biochemical parameters, as they reflect anthropogenic activities. The scale of influence of the related processes must thus be correctly mapped. For this purpose, the rejection of coastal data and data with shallow sounding is particularly useful, as shown by the changes reflected in the analyses.

Some results have been presented and we expect that both the datasets and the gridded fields produced by state-of-the-art statistical techniques will generate a new era of scientific investigation of the Mediterranean and the Black Sea, including investigation on the general circulation, on possible climate changes, on anthropogenic impacts on the seas, on regional eco-hydrodynamic processes, ...

Acknowledgements

We gratefully acknowledge the organizers and sponsors of 'The Colour of the Ocean Data' symposium. Many thanks to all data providers over the Mediterranean and the Black Sea. The FNRS provided valuable supercomputing assistance. The first author currently benefits from a EC Marie Curie Fellowship at the Southampton Oceanography Center.

References

Beckers J.-M., M. Rixen, P. Brasseur, J.-M. Brankart, A. Elmoussaoui, M. Crepon, C. Herbaut, F. Martel, F. Van den Berghe, L. Mortier, A. Lascaratos, P. Drakopoulos, P. Korres, N. Pinardi, E. Masetti, S. Castellari, P. Carini, J. Tintore, A. Alvarez, S. Monserrat, D. Parrilla, R. Vautard, and S. Speich. 2002. Model intercomparison in the Mediterranean. The MedMEx simulations of the seasonal cycle. Journal of Marine Systems 33-34:215-251.

Brankart J.-M. and P. Brasseur. 1996. Optimal analysis of in situ data in the Western Mediterranean using statistics and cross-validation. Journal of Atmospheric and Oceanic Technology 13(2):477-491.

Brankart J.-M. and P. Brasseur. 1998. The general circulation in the Mediterranean Sea: a climatological approach. Journal of Marine Systems 18:41-70.

Brankart J.-M. and N. Pinardi. 2001. Abrupt cooling of the Mediterranean levantine intermediate water at the beginning of the 1980s: observational evidence and model simulation. Journal of Physical Oceanography 31:2307-2320.

Brasseur P. 1991. A Variational Inverse Method for the reconstruction of general circulation fields in the Northern Bering Sea. Journal of Geophysical Research 96(C3):4891-4907.

Brasseur P., J.-M. Beckers, J.-M. Brankart and R. Schoenauen. 1996. Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of an historical dataset. Deep-Sea Research 43(2):159-192.

Brasseur P. and J. Haus. 1991. Application of a 3-D variational inverse model to the analysis of ecohydrodynamic data in the Northern Bering and Southern Chuckchi seas. Journal of Marine Systems 1:383-401.

Bretherton F. P., R.E. Davis and C.B. Fandry. 1976). A technique for objective analysis and design of oceanographic experiment applied to MODE-73. Deep-Sea Research 23:559-582.

Karafistan A., J.-M. Martin, M. Rixen and J. Beckers. 2002. Space and time distributions of phosphates in the Mediterranean sea. Deep-Sea Research 49(1):67-82.

Maillard C., E. Balopoulos and The MEDAR Group. 2004. Recent advances in oceanographic data management of the Mediterranean and Black Seas: the MEDAR/MEDATLAS 2002 database. p. xxx-xxx. In: Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus, P. Pissierssens (eds). Proceedings of 'Colour of Ocean Data' Symposium, Brussels, 25–27 November 2002. Paris, UNESCO/IOC, VLIZ, 2004. IOC Workshop Report 188 – VLIZ Special Publication 16.

Maillard C., M. Fichaut, G. Maudire, C.R. Coatanoan, B. Boudjelall, N. Eddalia, J.-M. Beckers, M. Rixen, G. Kortchev, G. Zodiatis, H. Dooley, S. El-Agami, I. Oliounine, E. Balopoulos, A. Iona, A. Lykiardopoulos, P. Karavregekis, I. Gertman, Y. Tsehtik, B. Manca, A. Giorgetti, A. Mosetti, G. Manzella, N. Pinardi, M. Zavatarelli, S. Lakkis, A. Drago, A. Orbi, J. Larissi, S. Zizah, N. Mikhailov, E. Vyazilov, A. Kuznetov, N. Puzova, M.-J. Garcia, M. Ozyalvac, F. Berkay, A. Suvorov, A. Khaliulin, R.

Gelfled, C. Sammari, V. Diaconu, K.Z.S. Bilashvili and D. Vlado. 2002. MEDAR/MEDATLAS 1998-2001: A Mediterranean and Black Sea oceanographic database and network. Bollettino di Geofisica Teorica ed Applicata. (submitted).

Rixen M., J.-M. Beckers, J.-M. Brankart and P. Brasseur. 2001. A numerically efficient data analysis method with error map generation. Ocean Modelling 2(1-2):45-60.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 245-255
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

A regional GIS for benthic diversity and environmental impact studies in the Gulf of Trieste, Italy

Vivianne Solis-Weiss[1], Pietro Rossin[2], Floriana Aleffi[2], Nicola Bettoso[2] and Serena Fonda Umani[3]

[1]Instituto de Ciencias del Mar y Limnologia, Universidad Nacional Autonoma de México, UNAM
Apdo Postal 70-305, Mexico DF 04510, Mexico
Present address (2)
E-mail: solisw@mar.icmyl.unam.mx

[2]Laboratorio di Biologia Marina
Via A. Piccard 54, 34010 Trieste, Italy

[3]Università degli Studi di Trieste
Via Weiss 2, 34100 Trieste, Italy

Abstract

The aim of this study was to build a database and apply GIS techniques with data obtained from 1966 to 2001 in the Gulf of Trieste, in order to determine the diversity, distribution and evolution in time and space of the local macrofauna in this heavily anthropized region for management purposes.

A custom-built relational database was created compiling biological, physico-chemical and sedimentological data from 450 stations, 187 sampling sites and 20 projects. A total of 278,770 organisms from 691 species (Polychaetes: 145,950 individuals, 276 species; Mollusks: 100,432 ind., 198 sp.; Crustaceans: 16,962 ind., 109 sp.; Echinoderms: 10,181 ind., 42 sp.; 'Others': 5,245 ind., 66 sp.) were included. Taxonomy was updated in all cases. Feeding guilds, biocoenotic characterizations and ecological groups were included for all the species where information existed.

All the surface analyses were performed using Surfer 7, Didger 2, Idrisi 32 and Cartalinx 1.2. The vectorial map supplied is in Idrisi vector file. The reference system is UTM 33-N. Distributions of the specific abundance maps were interpolated using the inverse distance to a 2 power gridding method.

The GIS analysis allowed us to individuate the main stress factors affecting the macrozoobenthic communities and determine the 'sensitivity areas' based on intensity and persistence in time of the different disturbances (urban development, industrial and port activities, mariculture, fishing and tourism), as evidenced by their effect on the benthos. The changes in time and space of the benthos composition indicate the evolution of the bottom conditions in the Gulf and helped establish the resilience of the system.

Keywords: GIS; Macrobenthos; Database, Gulf of Trieste.

Introduction

In modern times, databases constitute a basic tool and often the starting point for any ecological analysis and as such are being increasingly used either by government, scientific or private consultant companies for management purposes everywhere. They are built with a collection of as complete as can be obtained sets of data which can be then combined to get integrative interpretations of specific problems. In the case of ecological assessments they can be ideally complemented with Geographical Information Systems (GIS) for easier visualization of the different situations under study. Considering the classical geographical map as the primitive (or first) GIS, today, thanks to specialized software, the GIS represents the newest data processing technology for graphic representation of accurate geographical data. Launched during the early seventies in the USA, the GIS took some time to expand to Europe, but from the 80's, it has been increasingly used in Italy, not only by scientists, but also by public or private entities as well as common citizens (Baherenberg *et al.*, 1984; Nijkamp and Rietveld, 1984; Van Geenhuizen and Nijkamp, 1997).

The proved capacity for synthesis and comprehensive visualization of complex interactions among the different components of an ecosystem, makes the GIS an increasingly essential tool for adequate management of the territory as it enhances the potentiality of analysis of the existing information.

However, the application to marine ecosystems of GIS techniques to evaluate biodiversity or sensitivity areas in environmental impact studies, for protection of coastal zones and in management policymaking in general, is still very limited worldwide. Italy is no exception to this, not only in the number of existing GIS (the first and apparently most complete was produced in the Apulian region (http://siba-gis.unile.it/bis)) but in scope, since most of them were carried out in small localities to solve or interpret local problems.

Many parameters can be used to build a database for particular purposes, and in the case of a GIS, in the land studies, the fundamental units for analysis are constituted by the organisms living on the ground (human population of course and vegetation). For the marine realm, the same types of units, i.e. the organisms living in/on the substrate, are the most useful for environment assessments and management or protection studies. In fact, the benthic communities, composed of the organisms living for most or all of their lives in close contact with the bottom, are commonly used for studying or monitoring the environment and especially the effects of pollution on natural biota: their limited motility, high diversity, differential and fast responses to specific kinds of stress or their interruption, and their relatively long life spans makes them ideal for the historical record of what is affecting the area and how (Pearson and Rosenberg, 1978; Borja *et al.*, 2000). They can thus constitute, by the analysis of their composition, and preferably with their evolution in time (replacements, eliminations, alternate dominances, etc.), good indicators of the recent history and updated status of the habitat under study.

The Gulf of Trieste, located at the extreme northeastern end of the Adriatic Sea (45°40'N 13°30'E), is an ideal site to use these tools for environmental impact assessment: it is a large (about 600 km², 100.5 km of coastline) and shallow area (25 m at its deepest point), easily accessible all year long, heavily anthropized since at least 2000 years and which can boast to have several of the oldest records of marine benthic species (Stossich, 1876). The diversified and actively expanding activities that are carried out today there, potentially conflict with each other, either ecologically, economically or both: 1) fisheries, mariculture and tourism need clean and healthy waters, but cause stress on the environment; 2) industry (siderurgic, chemical, naval shipyards, etc), a large thermoelectric plant, agricultural by-products from rivers run-off, urban development and port activities whose discharges and/or refuses' final destiny is the Gulf of Trieste definitely impact the area; 3) maritime traffic (large commercial ships and oil tankers) and dredging to ensure navigation (in front of the lagoons and in Muggia Bay), constantly remove the bottoms in these shallow waters, a considerable stress for the benthic biota. The only official ecological haven, the small marine reserve of Miramare, needs constant protection (Fig. 1).

The aim of this study was to build a database and apply GIS techniques with data obtained from 1966 to 2001 in the Gulf of Trieste, in order to determine, for management purposes, the diversity, distribution and evolution in time and space of the local macrofauna in this heavily human-impacted region.

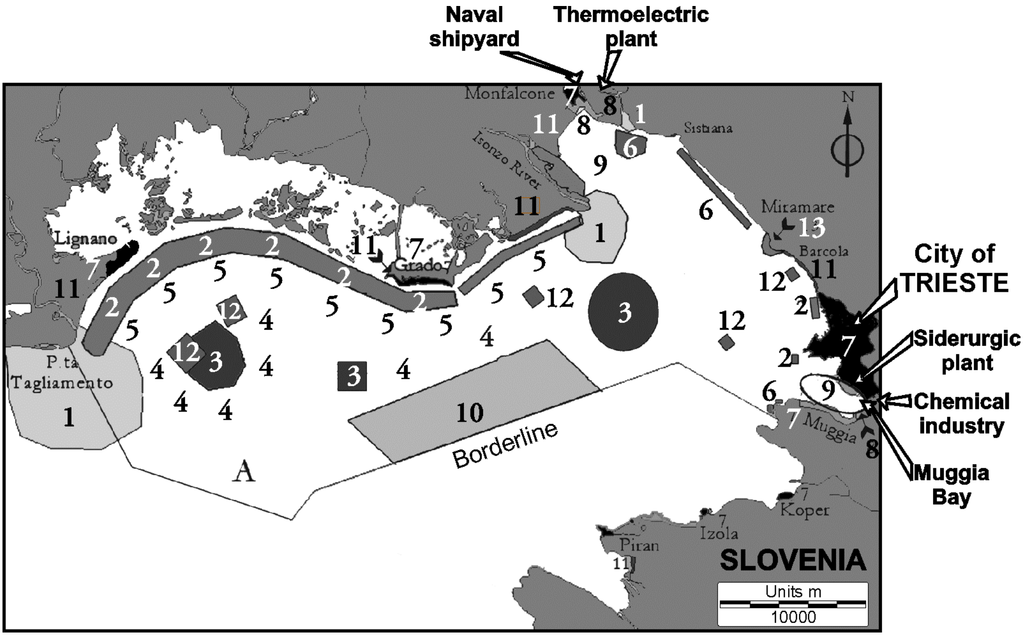


Fig. 1. Study area showing the different human activities and the sensitivity areas. (1) river estuaries, (2) zones of discharge of fishing gear cleaning, (3) discharge of dredged muds, (4) deep fishing area, (5) hydraulic dredges fishing area, (6) mussel farms, (7) urban centers, (8) industrial zones, (9) navigation, canal (Dams' zone at NW end), (10) 'promiscuous fishing rectangle' area, (11) tourist zones, (12) submarine urban sewage ducts, (13) natural marine reserves.

Methodology

A custom-built relational database was created compiling biological, physico-chemical and sedimentological data from 20 projects done between 1966 and 2001.
Taxonomic names were updated from original data in all cases where nomenclature has changed. Biocoenotic characterizations are based in Pérès and Picard (1964), Ecological Groups in Grall and Glémarec (1997) as in Borja *et al.* (2000) and Feeding Guilds in Fauchald and Jumars (1979) and Bachelet (1981).

All sampling sites (stations) were georeferenced (UTM 33N) either originally or processed by us, after careful positioning using original navigational data.

The original vectorial maps supplied were in Arc/Info export or Surfer format, with coordinates both in UTM 32-N and Lat-Long, scale 1/250000. The shoreline and bathymetric data were digitalized from the chart N°39 (Da Punta Tagliamento a Pula – UTM projection) 1:100000 (45°20'N) published by the 'Istituto Idrografico della Marina'. The resulting vectorial maps are in Idrisi vector files with UTM 33-N as the reference system.

The raw data are stored in the database as originally collected but for processing and interpretation purposes, as attempted in this case to exemplify its potential use, comparison between samples done in such a large time interval (35 years), coming from different projects with different objectives was not directly possible, so that we resorted to different techniques to homogenize the data. To solve the problem of different types of sampling gear used (Van Veen grabs and Charcot dredges), and different volumes of sample obtained, all abundances were adjusted to 50 litres. The effect of the different sizes of sieves used to wash and select the macrofauna for the different projects was minimized with the elimination of all organisms thought to pass through a 2mm sieve, the largest used. To evaluate the effect of seasonal variability, hierarchical classification based on Bray-Curtis similarity coefficient was run between samples taken at the same sites but at different dates. The resulting dendrogram showed that the stations are in general more related to the areas of sampling than to the sampling year or month (Solis-Weiss *et al.*, in prep.).

All the surface analyses were performed using Surfer 7, Didger 2, Idrisi 32 and Cartalinx 1.2. To obtain the maps for the distribution of the specific richness, Shannon-Wiener diversity index, Pielou's Evenness index and *Corbula gibba* distribution, an interpolation using the inverse distance to a 2 power gridding method offered by Surfer 7 (Golden Software, Inc.) was carried out and the stations chosen are not repeated in time.

Results and discussion

Database

The basic problems encountered in building the database were related to the high quantity of data that needed to be linked or associated to each other: a total of 450 stations, 187 sampling sites and 20 projects are included. In all, 278,770 organisms from 691 species were incorporated to the database. The Polychaetes constitute the dominant group with 145,950 individuals, in 276 species; Mollusks follow with 100,432 ind. and 198 sp., then Crustaceans with 16,962 ind., and 109 sp., and Echinoderms with 10,181 ind., and 42 sp. All other groups, being significantly less abundant and diverse (5,245 ind., of 66 sp. from Sipunculida, Ascidiacea, Porifera, Cephalochordata and Pisces) are grouped under the 'Others' category.

The Database basic structure, in Microsoft Access 2000, includes seven tables (Abundances, Stations, Projects, Samples, Species and 2 Indexes) connected by joins (Fig. 2) and thus allows swift searches with a minimum of space requirements.

In the 'Stations' Table, all data identifying the sampling stations are compiled: an identifying unique code number, the original name, geographical position and type of sediment. The 'Comments' are related to specific problems found in the data. An inner join connects it to the 'Samples' Table, where all available sampling information is recorded for the 450 stations: date, depth, type of sampler (Van Veen grab, Charcot dredge), mesh diameter used to wash and sort the sample, sampled volume and residual volume. 'Comments' are related to specific problems found in the sampling or in the compiled data. An inner join connects it to the 'Abundance' Table where the number of individuals of every species found in every sample is stored. In the 'Projects' Table, the complete name of every project involved, its corresponding identification code name and the files containing the original data (without any elaboration) are recorded and an inner join connects it to the 'Samples' Table. The 'Species' table contains the taxonomic list of the 691 species included in the Database organized by Phylum, Class, Order, Family and Genus-species, the latter in alphabetical order. 'Comments' are mainly related to synonymies, with date or other useful information. This table is likewise connected to the 'Abundance' Table.
Indexes 1 and 2: these two tables contain the bibliographic references and definitions of the indexes and categories used for the listed species related to each species. The Biocoenotic Characterization, Feeding Guild(s) and Ecological Group for each species where information is available, are compiled there.

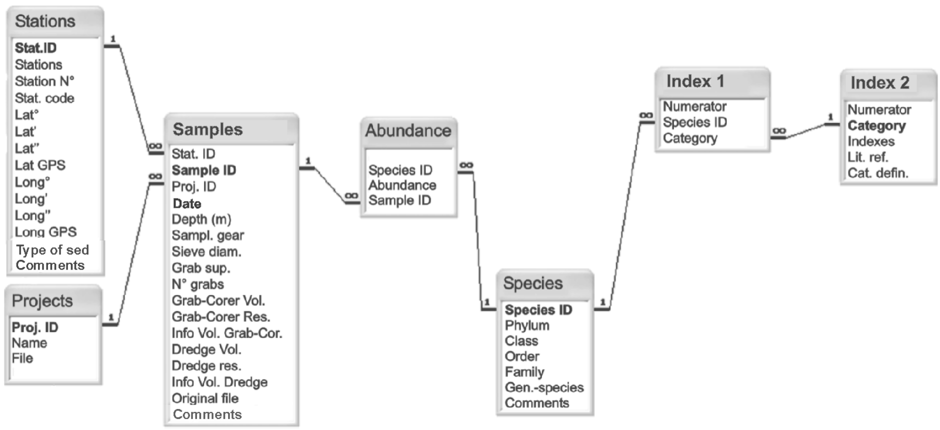


Fig. 2. Diagram of the composition of the Tables of the database.

Biocoenotic characterizations include: DC=detritic coastal bottoms, DE=detritic muddy bottoms, excl=species characteristic exclusive of 'x' biocoenosis, HP= Posidonia meadows, IETP=invertebrates of very polluted water, LEE=euryhaline and eurythermal, lre=wide ecological distribution, MI=unstable bottoms, minut=minute sediments, mixt=mixed sediments, ND=no data found, PI=pollution indicators, pref=species characteristic preferential of 'x' biocoenosis, sab tol=tolerant to sands, SFBC=fine well sorted sands, SFS=surficial fine sands, SGCF=coarse sands under bottom currents, sspr=ecological significance not specified, STP=very polluted sands, SVMC=muddy sands in sheltered areas, vas str=strictly muddy sediments, vas tol=tolerant to muds, VP=deep muds, VTC=terrigenous muds.

Feeding Guilds include: C=carnivores, G=grazers, O=omnivores, SDF=surface deposit feeders, SF=suspension feeders, SSDF=subsurface deposit feeders.
Ecological Groups include: I=species very sensitive to organic enrichment, intolerant to pollution, II=species indifferent to enrichment, III=species tolerant to enrichment, slightly unbalanced environments, IV=second-order opportunistic species, slight to pronounced unbalanced environments, V=first–order opportunistic species, pronounced unbalanced environments.

The database has a user-friendly interface and the user is guided trough the choice of operation to be performed. Complex types of queries are also possible and constant updating is being carried out.

Analysis of the macrobenthic communities

Different parameters characterising the macrobenthos distribution in the Gulf of Trieste were analysed using this database and GIS techniques to try to individuate the stress factors affecting the area.

In Fig. 3, species richness is schematically shown (maximum 108sp., minimum 0sp., average 34sp.): the richest regions were found to be in the centre of the Gulf (87 to 108), followed by the zones of discharge of the submarine urban sewage ducts, the Isonzo estuary and the Barcola and Sistiana areas (around 70). The poorest zones were found in the industrial zone of Muggia Bay and the port of Trieste (0 to 38sp.), as well as in the western zone of the study area (22 to 43sp.).

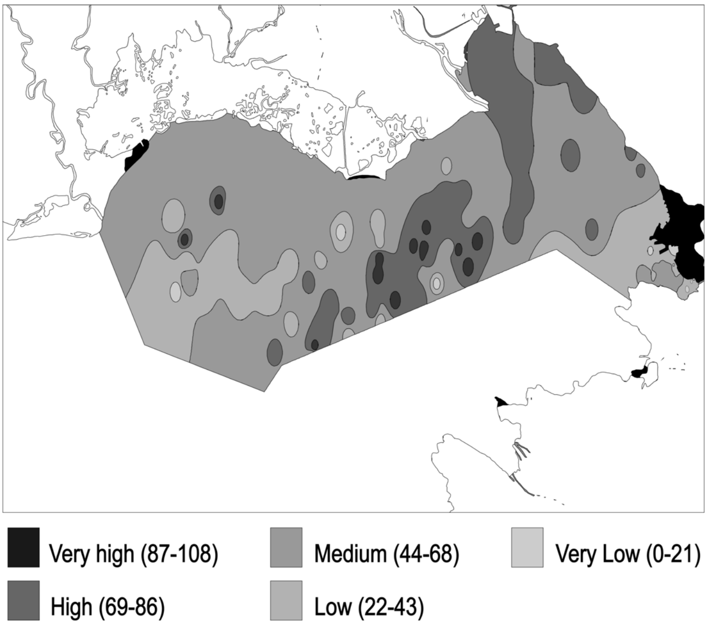


Fig. 3. Image of the species richness in the study area.

Based on the general accepted assumption that ecosystems in equilibrium (or at least healthy) are generally diversified, the poorest zones can be presumed to be impacted, either naturally or anthropically. In this case, it is easy to suspect that the industrial zone and the commercial port are the main source of impact in the area of Muggia Bay: in fact, from heavy metals concentrations in the sediments to water quality measurements (UNITS, 1977; Adami *et al.*, 1997), severe anthropogenic pollution can be clearly established there. In the western zone, however, the impoverished fauna results are most probably caused by the sampling method, since in those hardened bottoms (for sediment distribution in the Gulf of Trieste see Brambati *et al.*, 1983), the performance of the dredges and grabs is limited.

The submarine zones of urban sewage discharges (Fig. 1) at the end of long underwater ducts (from 1 to 7km long), quite surprisingly were found to have a quite rich and varied fauna (from 49 to 83 species and from 934 to 1636 ind./50 l) especially in the western duct zones (Solis-Weiss *et al.*, in prep.), implying that these discharges do not constitute a negative impact on the local benthic communities.

The analysis of the vectorial map representing the Shannon-Wiener diversity index values (Fig. 4), shows that the highest diversities (2.8 to 3.4) were found again in the centre of the Gulf around the deep fishing zone, followed by the submarine zones of urban sewage discharges, in the mussel (*Mytilus galloprovincialis*) farms' areas (north-eastern end) and around the tourist beaches of Sistiana and Barcola (Fig. 1). Lowest values were found in the industrial zone of Muggia Bay, in front of Grado and in some southern zones of the study area.

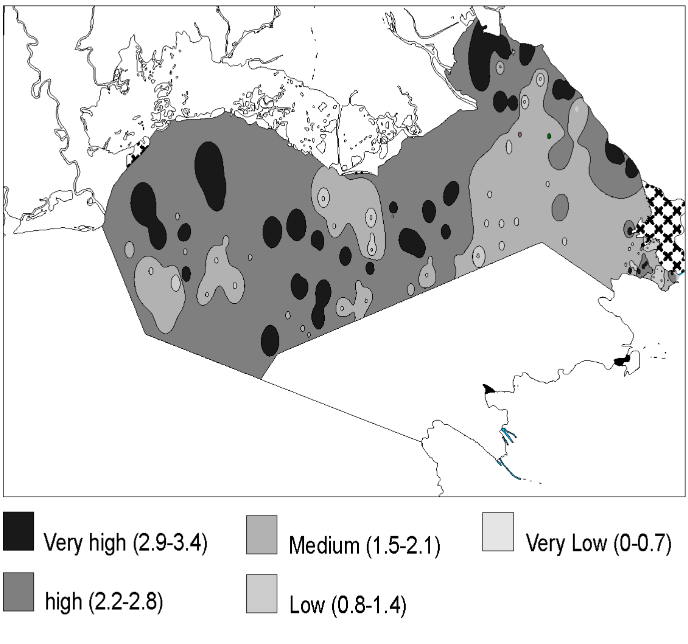


Fig. 4. Image representing the Shannon-Wiener diversity index values.

When overlapping Pielou's Evenness vectorial map (Fig. 5) with the precedent, we could see that in the centre of the Gulf, Evenness values are quite low (0.6 to 0.7) while in the mussel farms' areas, the beaches region, and the ducts zones, Evenness is high (0.87 to 0.94). This means that in the central Gulf, besides high values of diversity there should be strong dominances to explain the low Evenness values and that in the beaches and mussel farms' areas, diversity values are based in an equilibrium between the number of species and abundance which are generally not too high (around 45 species and 500 individuals on average).

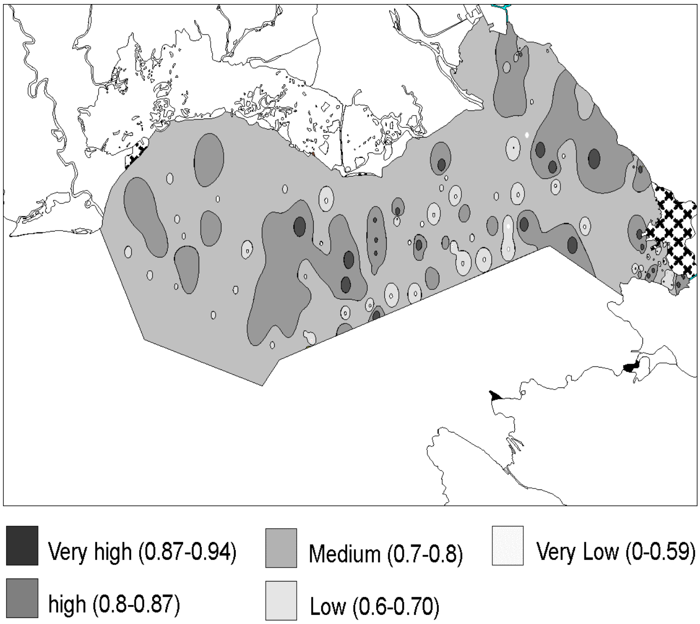


Fig. 5. Image of Pielou' Evenness index values.

To complement these results, we looked at the composition and distribution of the dominant species. The keystone species for the whole area is undoubtedly the mollusk *Corbula gibba*, which by itself represents almost 23% of the total fauna. Indeed in Fig. 6, we can see that *C. gibba* is most abundant in the centre of the Gulf (up to 4000 organisms in a single sample) and in the central areas of Muggia Bay. *C. gibba* is a well-known indicator of sediment instability (Orel *et al.*, 1987; Elias, 1992; Aleffi and Bettoso, 2000). Its ecological group (III), reinforces the perception of habitat unbalance. Other strong dominances are found there for a few species (the polychaetes *Nothria conchylega* 12,765 ind., almost 10% of the total of the fauna, *Eunice vittata* and *Maldane glebifex*).

The central part of the Gulf of Trieste, which also happens to be the deepest (about 25 m), is an area of active fishing, in detritic bottoms, of species like *Callista chione*, *Pecten jacobaeus*, *Solea vulgaris* and *Sepia officinalis*. The powerful hydraulic dredges and trawls used to fish them, have differentially affected the benthic populations (besides the commercial mollusks themselves), by constantly removing the natural bottoms, described there as muddy sands (Brambati *et al.*, 1983). The very high numbers of *C. gibba* clearly indicate the instability of those bottoms, this being reinforced by the local abundance of the polychaete *Owenia fusiformis*, a species reported to thrive in dredged bottoms or sands (Solis-Weiss *et al.*, 2001). It is however most interesting to note the high number of rare species present, as indicated by the high diversity (up to 3.4) and species richness values (87 to 108) of this area. Those constitute a potential pool for restoration of the original communities if the fishing pressure would diminish and thus the bottom instability and sediment texture would return to their original condition.

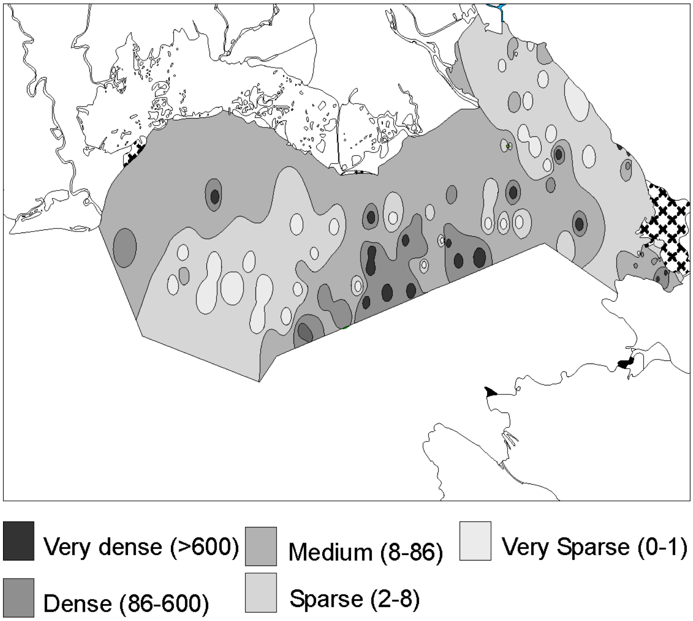


Fig. 6. Image of the keystone dominant species Corbula gibba.

In Muggia Bay, *C. gibba* is also found as dominant, especially in the southern region of the Bay, the species being eliminated from the highly polluted areas of the industrial zone (north and northeast) and found more abundantly in the region where the large ships entering the Bay manoeuvre and thus remove the shallow bottoms, or where dredging is taking place constantly to ensure a suitable depth to the navigation canal (Solis-Weiss *et al.*, submitted). In this last region, we found also large numbers of *O. fusiformis*. The areas closer to the industrial zone are characteristically very poor in abundance, diversity and species richness. In fact the only two stations found azoic in the whole area are located there. A decreasing gradient is evident in species richness when moving from the cleaner waters of the entrance of the Bay to the eastern and northern areas of the industrial zone, following the main current direction. The environmental pressures in this area, besides the very active industrial activity of the last 100 years, included the direct discharges of sewage from the city of Trieste (until the 80's) and water stagnation induced by the artificial closure of the Bay by a three dams' system built in 1909.

A survey of the database, comparing rough data from different years, indicatively could signal that there have been improvements since the ecological laws of 1976 were enforced in the area at the beginning of the 80's. In particular in Muggia, an improvement is clear after the sewage discharges stopped and there are more controls on the industrial discharges: the fauna show some recovery, both in composition and diversity (Solis-Weiss *et al.*, submitted). The ducts' areas are another interesting particularity of the region, since even though they discharge treated sewage, organic matter and other by-products should be present, they do not seem to harm the fauna, quite the opposite, and this is being now investigated with more detail (Solis-Weiss *et al.*, in prep.).

Conclusions

The analysis of the different maps showing the distribution of the ecological parameters measured and the dominant species distribution, allowed us to individuate the areas of more impact, or 'sensitivity areas' in the Gulf of Trieste, each one characterized by different kinds of impacting activities.

The central region of the Gulf is subjected to intensive trawl fisheries; in the Bay of Muggia there is a large industrial and commercial port area, subjected until recently to sewage discharge, with problems of induced stagnation and, in addition, active maritime traffic; the submarine ducts in turn, discharge, away from shore (1 to 7km), sewage waters (treated to a certain extent), so organic enrichment is obvious in the output areas, but since trawls are forbidden there, only more selective and much smaller means of fishing are employed, so that in practice they also constitute zones of protection and concentration for benthic organisms and fishes. The estuary of the Isonzo is a zone of discharge for agricultural by-products and Mercury (Hg) coming from up-stream. The beaches areas are presumably impacted by tourist activities. The mussel cultures of the northern areas are representative of a lucrative commercial activity, now recovering from a crisis period, which needs clean waters, and where controls are constant to guarantee the quality of the product. An additional sensitivity area is where shallow clams' fisheries are performed and the refuse of all kinds of fishing is directly discharged to sea, in front of the Grado and Marano Lagoons. However, this last zone was not evaluated, since sampling data from these shallow waters (up to 5m deep) is not available.

The fact that the fauna shows recovery shortly after strong deleterious impacts are stopped could indicate a high resilience of the system in general. The changes in time and space of the dominant species indicated the evolution of the bottom conditions in the Gulf and emphasized the resilience of the system.

References

Adami G., F. Aleffi, P. Barbieri and A. Favretto. 1997. Bivalves and heavy metals in polluted sediments: a chemometric approach. Water, Air and Soil pollution 99:615-623.

Aleffi F. and N. Bettoso. 2000. Distribution of *Corbula gibba* (Bivalvia, Corbulidae) in the northern Adriatic Sea. Annales 21:173-180.

Bachelet G. 1981. Données préliminaires sur l'organisation trophique d'un peuplement benthique marin. Vie Milieu 31:205-213.

Baherenberg G., M. Fischer and P. Nijkam. 1984. Recent developments in spatial data analysis. Gower, Aldershot.

Borja A., J. Franco and V. Pérez. 2000. A marine biotic index to establish the ecological quality of softbottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40:1100-1114.

Brambati A., M. Ciabatti, G.P. Fanzutti, F. Marabini and R. Marocco. 1983. A new sedimentological textural map of the northern and central Adriatic Sea. Bollettino di Oceanologia Teorica Applicata 4(1):267-271.

Elias R. 1992. Quantitative benthic community structure in Blanca Bay and its relationship with organic enrichment. PSZNI: Marine Ecology 13:189-201.

Fauchald K. and P. Jumars. 1979. The diet of worms: a study of polychaete feeding guilds. Oceanography and Marine Biology an Annual Review 17:193-284.

Grall J. and M. Glémarec. 1997. Using biotic indices to estimate macrobenthic community perturbation in the Bay of Brest. Estuarine and Coastal Shelf Science 44:43-53.

http://siba-gis.unile.it/bis. Visited November 2002.

Nijkamp P. and P. Rietveld. 1984. Information systems for integrated regional planning. North-Holland. Amsterdam.

Orel G., R. Marocco, E. Vio, D. Del Piero and G. Della Seta. 1987. Sedimenti e biocenosi bentoniche tra la foce del Po ed il Golfo di Trieste (Alto Adriatico). Bulletin d'Ecologie 18:229-241.

Pearson T.H. and R. Rosenberg. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanography and Marine Biology: an Annual Review 16:229-311.

Pérès J.M. and J. Picard. 1964. Nouveau manuel de bionomie benthique de la Mer Mediterranée. Recueil des Travaux de la Station Marine d'Endoume 31(47):5-137.

Solis-Weiss V., P. Rossin, F. Aleffi, N. Bettoso, G. Orel and B. Vriser. 2001. Gulf of Trieste: Sensitivity areas using benthos and GIS techniques. p.1567-1578. In: Proc. Fifth International Conference on the Mediterranean Coastal Environment. Vol. 1-3. Ozhan E. (Ed.).Medcoast, Middle East Technical University, Ankara, Turkey.

Solis-Weiss V., F. Aleffi, N. Bettoso P. Rossin, G. Orel, and S. Fonda-Umani. Effects of industrial and urban pollution on the benthic macrofauna in the Bay of Muggia (industrial port of Trieste, Italy). (submitted to Science of Total Environment, August 2002).

Solis-Weiss V., F. Aleffi, P. Rossin and G. Orel. The macrozoobenthic communities of the Gulf of Trieste: a study in time and space. (in prep.).

Stossich A. 1876. Breve sunto sulle produzioni marine del Golfo di Trieste. Bollettino della Società Adriatica di Scienze 2:349-371.

UNITS. 1977. La Baia di Muggia - Stato ed evoluzione dell'ambiente. Università Degli Studi di Trieste.Comune di Muggia. Trieste. 155p.

Van Geenhuizen M. and P. Nijlamp. 1997. L'architettura dei sistemi informative nella gestione di una città complessa. p.1-66. In: La città e le sue scienze. Vol. 4. Le metodologie delle scienze della città. Bertuglia C.S. and F. Vaio (Eds). Angeli, Milano.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 257-270
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Diagnosis of environmental impacts on the Mexican coastal zone with a comprehensive ad-hoc database

Vivianne Solis-Weiss[1] and Alejandro Granados Barba[2]

[1]Instituto de Ciencias del Mar y Limnologia, UNAM
Apdo Postal 70-305. Mexico, DF 04510, Mexico
(present address: Laboratorio di Biologia Marina, Via A. Piccard 54, 34010 Trieste, Italy)
E-mail: solisw@mar.icmyl.unam.mx

[2]Centro de Ecologia y Pesquerias, Universidad Veracruzana
Apdo. Postal. 663 CP 91000, Xalapa, Veracruz, Mexico

Abstract

Mexico, with 11,500 km of littorals and almost 3.106 km² of marine Economic Exclusive Zone is ideally located for a large array of coastal habitats to thrive. Its main sources of revenue: petroleum, tourism and shrimp catch, are all closely related to the coastal zone. The current booming development of this area which boasts now the highest population growth in the country, contrasts with the lack of knowledge, adequate legislation or ecological protection measures.

The Mexican government financed this project in order to have an integrated knowledge of what data already existed to assess water quality in its coasts. A custom-built database would help in defining the necessary policies to ensure its adequate development and recognize what needs to be further done in research.

Integrated in the RAISON software and based on a comprehensive bibliographic search, 126 parameters were chosen as indicators of water quality and divided in: Physico-chemical (32), Persistent Organic Compounds (34), Hydrocarbons (21), Metals (24), Biological (10) and Geological (5). Of those, 119 were recorded from water, 108 from sediments, 71 from organisms, and they can be divided by States (17).

The main activities were divided in: Fishing and Aquaculture, Petroleum extraction, Tourism, Port activities, and Urban and Industrial development. Anthropogenic impact, as contamination by fecal coliforms, was ubiquitous and sometimes found in very high concentrations and/or already incorporated into sediments.

The information was then valued and analyzed to assess environmental impacts. Since norms only exist for water parameters in Mexico, international legislation and literature reports were used for sediments and organisms. A diagnosis of the coastal localities emerged, which combined with our evaluation at the State level, allowed us to make a general diagnosis of the coastal zone in Mexico.

Keywords: Mexican coastal zone; Coastal pollution; Environmental policies; Database.

Introduction

Mexico is a country that ranks 12th in the world for its extensive coastline of about 11,500 km (Merino-Ibarra, 1990). Its littorals, in contact with both the Atlantic (Gulf of Mexico and Caribbean, 32%) and the Pacific (Mexican Pacific and Gulf of California, 68%) (Fig. 1), include tropical as well as subtropical and, to a less extent, temperate zones, with an impressive marine

Exclusive Economic Zone (EEZ) of about 3.106km2. Several biogeographic provinces are found in Mexico: the Californian and Panamian provinces in the western coasts (Brusca, 1980) and the Carolinean and Caribbean to the east (Briggs, 1995). Thus, a large variety of climatic conditions exist which, complemented by a high diversity of geomorphologic features, make for a wide spectrum of marine habitats to thrive, among the most valuable: about 130 coastal lagoons (Lankford, 1977; Contreras Espinosa, 1993) (12,500 km2) and estuaries (16,000 km2), islands (6,606 km²), coral reefs, extensive mangrove areas (more than 600,000 ha (Lot and Novelo, 1990), seagrass beds and beautiful sandy and rocky beaches. In turn, these contribute to a rich variety of fauna and flora (largely unknown), as well as a vast array of renewable and non renewable exploitable resources.

Fig. 1. Study area: the Mexican coastal zone with its main coastal ports and cities, visited sites and boundaries of the coastal states.

In contrast with this extension, variety and richness, the development of the coastal zone in Mexico has been remarkably low, to the point of being qualified about twenty years ago as a country that lived with its back turned to the sea. This situation had historical roots since, from the start, human settlements in the coastal zone were more hazardous than in the high plains due to the proliferation of infectious diseases, typical of the tropical humid areas covered by rich vegetation, pirate attacks, etc. Then, following the Spanish conquest, the economic development of colonial Mexico was directed towards more profitable activities such as mining (it is to this day the first silver producer in the world), cattle breeding and agriculture, all done in high lands away from the sea. To this day, the three most important cities in Mexico (Mexico City, Guadalajara and Monterrey) are all far away from the sea.

However, since the discovery of the rich oilfields in the Gulf of Mexico about 30 years ago, followed by their continuously increasing exploitation, the growth of fisheries and aquaculture and the thriving tourist ‘industry’ directed mostly at its spectacular beaches, the development of the coastal zone has been in constant expansion. Nowadays, these three activities, all closely related to the coastal zone, are the most lucrative in the country and the first sources of foreign currency for Mexico.
However, the high benefits already obtained by these rich resources coupled with the high potential for further development of these and other activities, also imply definite dangers for the environmental health of the country and its sustainable development in the face of the high economic profits at stake and conflicting interests arising from their contemporaneous exploitation.

In fact, this booming growth has proceeded until very recently with very few controls: the lack of knowledge, adequate legislation or guiding plans for the harmonious development of the zone was striking considering not only the resources and potential but also the growing interest at international levels for sustainable development and protection of the environment first addressed with most governments’ commitment at the Rio Earth Summit UNCED Conference (1992). The Mexican government, recognizing these problems, and as signatory of international treaties such as Agenda 21 and the OCDE, considered as a priority in its official Program for the Environment 1995-2000 (Poder Ejecutivo Federal, 1996), to carry out this project named: ‘Environmental diagnosis and development of a database for the Coastal Zone of Mexico’ through the National Institute of Ecology (INE), a branch of its Environment Ministry.
The INE in fact, among its duties, is in charge of formulating, conducting and evaluating the official policies at the national level for the protection of the environment and to ensure preservation and restoration of the ecosystems, their rational and sustainable use and their correct management.
The aim of this project, the first of its kind in Mexico, was thus to have an integrated state of the art knowledge of what had been done in the country to assess water quality of the coastal waters, in a custom-built database, which would help to define the necessary policies to ensure an adequate development of that area and could also be used to recognize what needs to be done in research to complement this effort. The results would be available to all interested parties (scientific, public or private). Water quality was chosen as the indicator, because it can help to assess the local situation as well as that ‘upstream’, in the global sense. We are aware of the implicit reductionist view in this approach, but it was retained necessary at this point since this was the first attempt to a complete overview of the problem at the national level. Besides grouping all the available information on the subject, the study aimed also at an evaluation of the sources and of the data gathered and a diagnosis at the States level which, when grouped, can give a synthetic idea of the state of water quality and research needs of the whole coastal zone as it is known today.

Methodology

A comprehensive bibliographic search was performed in the 69 libraries of research centers, government agencies and public institutions around the country (48 cities and resort centers) where available published data existed on water quality of the coastal zone. The 42 coastal cities and centers of coastal development considered as the most important in the country were visited

(Fig. 1), and the bibliographic data (taken from 1990 to 1997) were complemented with interviews of key officials of public agencies or involved individuals (from businessmen to fishermen); a photographic survey of areas of particular interest to check the real state of the coast completed the overview. Data from before 1990 were not retained necessary to compile because of the intrinsic changing characteristics of the parameters under study, but were taken into account for analyses purposes.

In the absence of a satisfactory and universally accepted definition of 'coastal zone', and especially its boundaries, for the purposes of this project the coastal zone was defined as the interface between marine and terrestrial habitats, with the boundaries set, towards sea, at the end of the continental shelf and towards land up to where marine influence can be assessed (which is not always a clear-cut boundary). 'Water quality' is another hard to define parameter, when seawater is involved, since otherwise it amounts to determine its drinking quality. However, for seawater, very large variations in the concentration of many compounds are normal (as a single example, consider salinity in estuaries). For water quality assessment, in this case we resorted to comparison between our data and published official permissible concentrations of different parameters (from coliforms to phosphates or hydrocarbons), and in accordance with the intended use (recreational, industrial, port activities…) at national or international levels

The RAISON software (Regional Analysis by Intelligent Systems on a Microcomputer) developed by Environment Canada, already used by the ONU for the GEMS studies (Global Environmental Monitoring System) was chosen to manage the information, in order to integrate the database, with an Excel Spread Sheet and a graphic editor to visualize the results in maps and figures.

All data were validated and only incorporated into the database if the following quality criteria were met: published, georeferenced, dated and available for confrontation. The data were organized by parameters divided in: Physico-chemical, Persistent Organic Compounds (POC), Hydrocarbons, Metals, Biological and Geological, and separated by source of the data: water, sediment, and/or organisms: They are also presented by locality and State. The references used were listed also by state, to facilitate consultation for selected areas (see internet site cited below) in addition to a general section.

The main official normative references regarding water quality in Mexico which were used to analyze and evaluate the impacts of the different parameters known to influence water quality in coastal waters were:

1. the NOM-001-ECOL-1996 (Norma Oficial Mexicana (= Official Mexican Norm) (published officially Jan 6, 1997), which is the only one with legal status (Diario oficial de la Federacion, 1996);
2. the 'Reglamento para la Prevencion y Control de la Contaminacion de las Aguas' (= Regulations for prevention and control of water pollution) (published officially March 29, 1973);
3. the 'Criterios Ecologicos de la Calidad del Agua' (= Ecological criteria for water quality) (CE-CCA-001/89, 1989) (officially published Dec. 2, 1989).

At all sites where studies existed, comparison of the values measured and the official permissible concentrations was performed in order to assess the water quality of the particular

Table I. List of parameters used for the assessment of coastal waters quality in the Mexican coastal zone

	PHYSICO-CHEMICAL
1	Ammonium
2	Biochemical Oxygen Demand
3	Chemical Oxygen Demand
4	Colour
5	Dissolved Oxygen
6	Dissolved Solids
7	Fats and Oils
8	Nitrates
9	Nitrites
10	Organic Dissolved Nitrogen
11	Organic Dissolved Phosphorus
12	Organic Matter
13	Organic Particulate Nitrogen
14	Organic Particulate Phosphorus
15	pH
16	Phenols
17	Phosphates
18	Redox Potential
19	Salinity
20	Sampling Depth
21	Substances Active to Methylene Blue
22	Sulphites
23	Sulphures
24	Suspended Solids
25	Total Alcalinity
26	Total Depth
27	Total Nitrogen
28	Total Organic Carbon
29	Total Phosphorus
30	Transparency or Turbidity
31	Water Renewal time
32	Water Temperature
	POC'S
1	2,4,5-T Derivatives
2	2,4-D Derivatives

11	DDE
12	DDT
13	Dieldrin
14	Durban
15	Endosulfan
16	Endosulfan I alfa
17	Endosulfan II beta
18	Endosulfan Sulphate
19	Endrin
20	Endrin aldehyde
21	Heptachlorine
22	Heptachlorine Epoxid
23	Lindane
24	Malathion
25	Metoxychlorine
26	p,p´-DDD
27	p,p´-DDE
28	p,p´-DDT
29	Parathion
30	Pertane
31	Ronnel
32	TDE
33	Toxaphene
34	Triazine Derivatives
	HYDROCARBONS
1	2,6 Dimetylanthracene
2	Acenaphtene
3	Acenaphtylene
4	Anthracene
5	Aromatic Polynuclear Tot Hydrocarbons
6	Benzo(a)anthracene
7	Benzo(a)pyrene
8	Benzo(b)fluoranthene
9	Benzo(ghi)perylene
10	Benzo(k)fluoranthene
11	Crisene

20	Saturated Hydrocarbons
21	Toluene
	METALS
1	Aluminium
2	Antimonium
3	Arsenic
4	Barium
5	Berillium
6	Boron
7	Bromine
8	Cadmium
9	Cyanure
10	Cobalt
11	Copper
12	Chromium
13	Fluorine
14	Iron
15	Lithium
16	Magnesium
17	Manganese
18	Mercury
19	Molybdene
20	Nichel
21	Silver
22	Lead
23	Vanadium
24	Zinc
	BIOLOGICAL
1	Clorophyll A
2	Faecal Coliforms
3	Faecal Streptococci
4	Heterotroph bacteria
5	Hydrocarbonoclastic Bacteria
6	Indicator Species
7	Primary Productivity
8	Respiration

Table I (cont.)

3	Aldrin	12	Flourene	9	Total Coliforms
4	Arsenic	13	Fluoranthene	10	Total Streptococci
5	BHC alfa	14	Indo(1,2 Cd)pyrene	**GEOLOGICAL**	
6	BHC beta	15	Naphtalene	1	% Gravel
7	BHC gamma	16	Petroleum (Aliphatic Hydrocarbons)	2	% Muds
8	Carbamates	17	Petroleum Hydrocarbons	3	% Sand
9	Clordane	18	Phenanthrene	4	Sediment Type
10	Cumafos	19	Pyrene	5	Sedimentation Rate

site. Since these norms only exist for water parameters in Mexico, for values recorded in sediments and organisms we resorted to international legislation and literature reports.

To evaluate the degree of impact of any given parameter, the averages of reported values for each parameter were confronted with the reference value (highest permissible monthly average value) or Norm (defined for each one in the NOM-001-ECOL-1996) to calculate the number of times that these parameters surpassed the Mexican Norm (NOM), according to the intended use of the local water body, *i.e.*: different concentrations of any parameter are considered for bathing and tourist activities, or fishing and aquaculture than for industrial or port areas. The results were then grouped by class (from 0= no impact, average values inferior to the Norm, 1= low impact, average values equal to norm, 2= moderated impact, average values equal to two times the norm, 3= intense impact, average values equal to three times the norm, 4= severe impact, average values equal to four times the norm, 5= extreme impact, average values equal to or larger than 5 times the norm).

With these results, regional maps were constructed which show the impact of the different sets of parameters according to the intended use of the coastal areas evaluated. Since all regions and all uses where data existed are compiled and shown in the figures, in this paper it would be impossible to show them all, so we will only use a few of them when necessary as examples of the results. However, all the tables and figures used for this project, which would also be too many to show here, can be directly consulted in the INE site at http://sepultura.semarnat.gob.mx/dggia/zcoster/index.html,where they have been inserted since 2001. The same stands for the literature used and the research centres visited, for the same reasons of space constraints.

Results

A total of 126 parameters were chosen as the main indicators of water quality. Of these, 32 correspond to Physico-chemical parameters, 34 to POC's, 21 to Hydrocarbons, 24 to Metals, 10 to Biological and 5 to Geological; 119 of them were recorded from water, 108 from sediments and 71 from organisms.

379 publications were found on the subject, from which 186 (49.1%) were formal publications (129 in national journals and 57 in international journals),165 (43.5%) theses (123 B Sc level, 34 MSc and 8 PhD) and 28 (7.4%) Technical Reports. From the total, 221(58.3%) publications

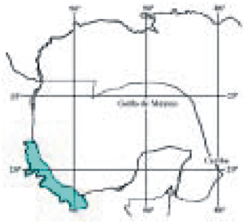
refer to the Pacific Ocean (161 in the Northern Pacific and 60 in the Southern Pacific), 145 (38.3%) to the Gulf of Mexico, and 13 (3.4%) to the Caribbean. Of those, only 159 publications met the quality standards required for incorporation into the Database (92 or 57.9% from the Pacific, 59 (37.1%) for the Gulf of Mexico and 8 (5.0%) for the Caribbean. This means that the quality of the analyses done until then was very often inadequate. Also, quite often, we found out that the information required can only be obtained at the personal level and is never found in publications.

There are 17 States in Mexico (out of 32) with coastal boundaries (Fig. 1). For each State, information of general interest was included and schematically presented such as: surface, population, cities, climate, land use (%), coastal ecosystems present and their surface, number of research centers, natural protected areas, ports, main fishing resources and production (Fig. 2.).

The literature results for that State were then presented, evaluated (number and type of publications) and commented (Fig. 3). An analysis of the main parameters measured in the different important sites of the State is then presented along with maps with all the georeferenced sampling stations considered (Fig. 4). Finally, a diagnosis is made at the State level based on the analysis of the main impacts found to influence its water quality.

From water measurements, the states where more records were found were Oaxaca and Tabasco (40 parameters each: 17 and 18 respectively for physico-chemical data, 4 and 7 for metals, 15 and 13 for hydrocarbons, no data for POC's, and Sinaloa (30 parameters: 20 for physico-chemical data, 5 for POC, 5 for biological, no data for metals or hydrocarbons).
Those with less records were Baja California Norte (17: 10 physico-chemical data, 3 for metals, 4 biological, no data for POC's or hydrocarbons), Yucatan (14: 13 for physico-chemical data , 1 for biological), Baja California Sur (13: 11 physico-chemical, 2 biological, no data for POC, hydrocarbons or metals) and Guerrero (11 parameters, 9 for physico-chemical data and 2 for biological).
From sediments, physico-chemical records are also the most abundant. The States with more data are: Tabasco (26 parameters: 2 physico-chemical, 6 metals, 15 hydrocarbons and 3 geologic), Veracruz, (25 parameters: 1 physico-chemical, 8 metals, 13 hydrocarbons, 1 geologic and 2 biological), Oaxaca (25 parameters: 5 metals, 16 hydrocarbons and 4 biological).
The states with less records were: Michoacan and Guerrero, with no data at all from sediments, and Tamaulipas, (2 parameters: 1 metal and 1 geologic).

Measurements made from organisms were scarce, and all related to metals and hydrocarbons: from these Veracruz with 20 parameters (8 metals and 12 hydrocarbons) and Oaxaca (19 parameters: 4 metals and 15 hydrocarbons) were the ones with more data, while Sonora, Jalisco, Michoacan, Guerrero, Chiapas, Yucatan and Quintana Roo had no data at all.
Physico-chemical parameters measurements were those which were more often found in the Mexican coastal zone since they were recorded from 43 coastal water bodies. Our results showed that 'extreme' impacts existed due to high concentrations of total suspended solids in some water bodies of Sonora, Jalisco, Colima, Campeche and Quintana Roo; the same was true for Phosphates, Nitrites, Nitrates Sulphurs or Phenols, in some areas of Sonora, Jalisco, Guerrero, Oaxaca and Quintana Roo; and with Substances Reactive to Methylene Blue (indicating detergents) in some water bodies of Sonora, Sinaloa, Colima, Guerrero, Oaxaca, and Quintana Roo.



Veracruz

Ecosistemas Principales:

- Laguna Pueblo Viejo
- Laguna de Tamiahua
- Laguna de Tampamachoco
- Laguna Grande
- Laguna Verde
- Laguna Farallón
- Laguna la Mancha
- Laguna Mandinga
- Laguna Camaronera
- Laguna de Alvarado
- Laguna Pajaritos
- Laguna Catemaco
- Laguna Sontecomapan
- Laguna Ostión
- Río Tuxpan
- Río Cazones
- Río Tecolutla
- Río Misantla
- Río Actopan
- Río Jamapa
- Río Papaloapan
- Río Coatzacoalcos
- Río Tonalá
- Río Panuco

Áreas Naturales Protegidas Costeras:

- Reserva Especial de la Biosfera Sierra de Santa Martha"
- Reserva Especial de la Biosfera"Volcán de San Martín"
- Parque Marino Nacional "Sistema Arrecifal Veracruzano"

Población por Municipio Costero (1995)

Veracruz	425,140 hab.
Coatzacoalcos	259,096 hab.
Papantla	171,167 hab.
San Andrés Tuxtla	137,435 hab.
Boca del Río	135,060 hab.
Tuxpan	127,622 hab.
Martínez de la Torre	113,560 hab.
Alvarado	48,490 hab.
Pueblo Viejo	48,054 hab.
Agua Dulce	46,404 hab.
Catemaco	44,321 hab.
Actopan	41,884 hab.
Angel R. Cabada	34,312 hab.
Soteapan	28,888 hab.
Ursulo Galván	28,158 hab.
Tamiahua	27,398 hab.
Alto Lucero de Gutiérrez B.	27,331 hab.
Tecolutla	25,730 hab.
Cazones de Herrera	23,621 hab.
La Antigua	23,529 hab.
Mecayapan	22,764 hab.
Lerdo de Tejada	20,810 hab.
Vega de Alatorre	19,412 hab.
Tampico Alto	13,604 hab.
Pajapan	13,073 hab.
Nautla	9,599 hab.

Coordenadas Extremas:

22°28'-17°09' Lat. N. 93°36'-98°39' Long. O.

Superficie:

68,940.27 km^2, que constituye el 3.7% del total nacional.

Longitud de la Línea de Costa:

745.14 km

Superficie de la Plataforma Continental:

22,935 km^2

Superficie de Lagunas Litorales:

1,166 km^2

Amplitud de la Plataforma Continental:

Máxima: 56 km frente a Las Barrillas.
Mínima: 22 km frente a la barra de la Laguna de Tamiahua.

Tipos de Climas Dominantes:

A(w) cálido subhúmedo–52.30%
Am cálido húmedo–27.76%

Población Total:

6,737,324 habitantes (1995)

Ciudades Costeras Principales:

- Tuxpan
- Veracruz
- Coatzacoalcos
- Alvarado

Uso del Suelo: (Hectáreas (1994))

Agrícola	4,458,249.5
Pecuario	2,119,330.3
Urbano	72,792.1
Industrial	60,677.5
Forestal	292.6

Recursos Pesqueros Principales:

141,486 toneladas en volumen de captura en peso desembarcado a nivel nacional (1996)

Tilapia	16,086
Ostión	17,996
Lebrancha	4,674
Jaiba	4,809

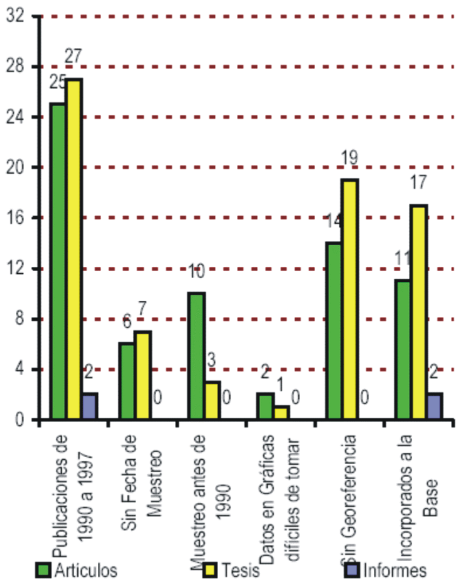
Puertos Principales:

- Veracruz (Industrial, Mercante, Pesquero, Turístico).
- Tuxpan (Industrial, Mercante, Pesquero).
- Coatzacoalcos y Pajaritos (Industriales, Mercantes, Pesqueros).
- Alvarado (Pesquero).

Centros de Investigación Costera:

- Facultad de Biología de la Universidad Veracruzana (UV).
- Instituto de Ecología, A.C.
- Instituto Tecnológico del Mar (ITMAR) Veracruz.
- Instituto de Investigaciones Oceanográficas del Golfo de México y Mar Caribe, SEMAR.
- Centro Regional de Investigación Pesquera (CRIP-Veracruz).

Fig. 2. Example of the presentation of data by State (here Veracruz): General data.



Zona Nerítica			
AGUA	máx.	mín.	prom.
NH4 mg/L	3.383314	0	1.128981
PO4 mg/L	0.370808	0	0.104417
GyA mg/L	15.73	8.27	12
NO3 mg/L	1.496122	0	0.22766
NO2 mg/L	0.10626	0	0.032066
OD mg/L	10.9	2.15	6.287357
pH	8.265	7.074	8.024031
PM m	200	0	26.43913
Cl a mg/L	0.002671	0.000805	0.001625
CF NMP	34100	136	17118
CT NMP	50100	433	25266.5
SEDIMENTOS	máx.	mín.	prom.
Cd µg/g	9.59	8.3	8.885
Cu µg/g	7.47	6.16	6.981667
Cr µg/g	20.97	19.71	20.27667
Fe µg/g	100.57	43.69	69.61667
Mn µg/g	14.081	0.59	12.00333
Ni µg/g	40.46	33.7	37.265
Pb µg/g	105.73	72.8	88.23667
Zn µg/g	14.03	9.75	12.06333

Fig. 3. Example of the presentation of data by State (here Veracruz): Literature evaluation and analysis of impacts.

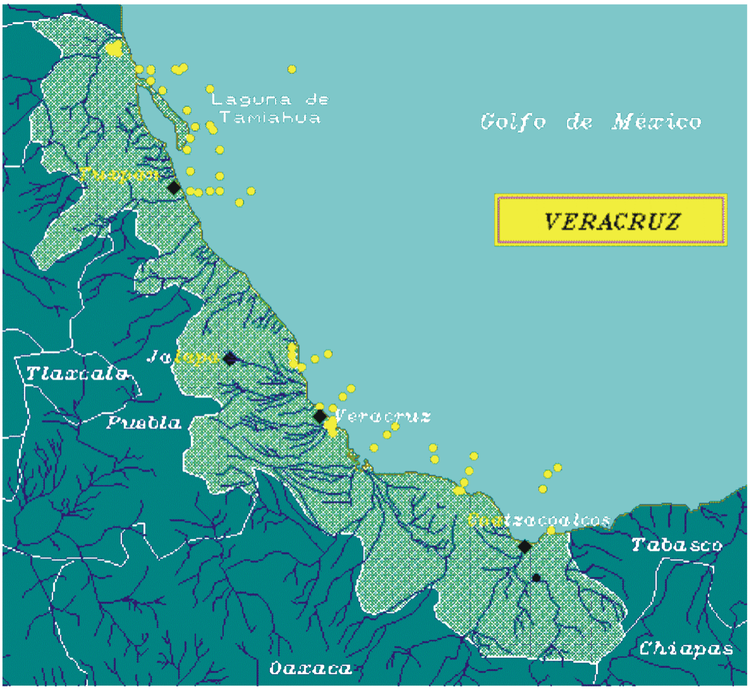


Fig. 4. Example of the presentation of data by State (here Veracruz): sites sampled in the neritic zone, georeferenced stations.

Important concentrations of Fats and Oils were recorded from Bahia de Banderas (Jalisco), Manzanillo (Colima), and Cancun Lagoons (Quintana Roo) where, in addition, Phenols are significantly present. Dissolved solids and low levels of Oxygen were found in front of Chetumal (Quintana Roo). Impacts are considered ‘extreme’ in Acapulco Bay where very low concentrations of dissolved Oxygen were found, in addition to suspended solids in abundance as well as Fats and Oils.

Data about POC’s and Hydrocarbons were definitely scarce and the records came from small sampling areas. This only enabled us to report ‘moderate’ impacts: in the case of POC’s they are associated to intense agricultural activities (like in Sonora). This scarcity of information is troubling since those compounds are widely used in agriculture and their high environmental toxic potential is already well documented. Hydrocarbons are only recorded from areas where petroleum industry or related activities prevail, like Veracruz, Tabasco and Oaxaca: they were already present in water, sediments and organisms in Salina Cruz, Oaxaca, all water bodies of Tabasco and all lagoons and estuaries in southern Veracruz.

Metals constituted another category where data were scarce: only found in 16 water bodies of the following States: Baja California, Sonora, Sinaloa, Nayarit, Colima, Oaxaca, Tamaulipas, Veracruz, Tabasco, Campeche and Yucatan. In most records, local impacts were qualified from ‘intense’ to ‘extreme’, since significant concentrations of Cadmium, Copper, Chrome, Lead and Zinc were reported. In the Veracruz lagoons and in Laguna La Cruz (Sonora), impacts reached ‘severe’ levels whereas in Bahia Magdalena (Baja California), Pabellon-Altata (Sinaloa), Bahia de Manzanillo (Colima) and the neritic zone in Campeche, impacts reached ‘extreme’ levels for considerable concentrations of Cu, Zn, Ni, Cr, Ba and Al.

Biological parameters: records basically regarded coliform concentrations (faecal and total), whose levels were qualified as ‘intense’ to ‘severe’ in most of the 30 water bodies where records exist, especially along the Veracruz and Sinaloa coasts. ‘Extreme’ levels of impact were recorded in Baja California, Sonora, Guerrero, Colima, Oaxaca (in Salina Cruz alarming concentrations were recorded), Tamaulipas, Campeche and Quintana Roo. Other records concerned heterotroph and hydrocarbonoclastic bacteria.

Practically no records are available for studies performed in organisms or their tissues, even for POC’s or hydrocarbons: the few data gathered enabled us to consider the impact as ‘moderate’ for both parameters. Baja California and Terminos Lagoon (Campeche) were the only sites where data existed for POC’s and Veracruz and Oaxaca (Salina Cruz port) were the only ones for hydrocarbons, despite the heavy oil-oriented status of the Tabasco and Campeche coasts which are, at the same time, highly praised for their shrimp, oyster and fish production.

Data on metals recorded from organisms were only found in 15 water bodies, in the coastal regions of Baja California Sur, Sinaloa, Nayarit, Colima, Oaxaca, Tamaulipas, Tabasco and Veracruz. Their impact reached levels from ‘moderate’ to ‘extreme’, the latter in the Veracruz lagoons, while Cd, Cu, Cr, Pb, Zn and Fe were found in high concentrations in selected organisms of Tabasco coastal lagoons (Carmen & Machona).

Discussion

From the compiled results and the visits done to the main coastal cities, resorts and industrial centres of the Mexican coasts, we could divide the main activities done in the coastal zone in: Fishing and Aquaculture, Petroleum extraction, Tourism, Port activities, and Urban and Industrial development.

From these, fishing, aquaculture and tourism require a clean environment to survive and flourish, while creating potential damage to it, especially in the case of aquaculture (the by-products of aquaculture can irreversibly damage entire coastal areas) and secondly tourism. The other activities are always in conflict with a healthy environment and considerable efforts have to be made in order to limit the damages. These conflicts are particularly acute in Mexico, because of the especially high economic benefits derived from the first three activities mentioned: the potential and demonstrated fast profits of the short term have to be constantly weighed with the known dangers and real damage in the long term. The conflicts of interests become even more delicate to manage when two or more activities are carried out contemporaneously; to take only one example, see the offshore oil extraction in the continental shelf of Tabasco and Campeche in the richest zone of shrimp fisheries in the country.
With the results obtained from the measured parameters, regional maps were built which showed, for each coastal city or locality studied, the degree of impact of the different parameters on water quality for the different activities performed in the area, as is exemplified in Fig. 5.

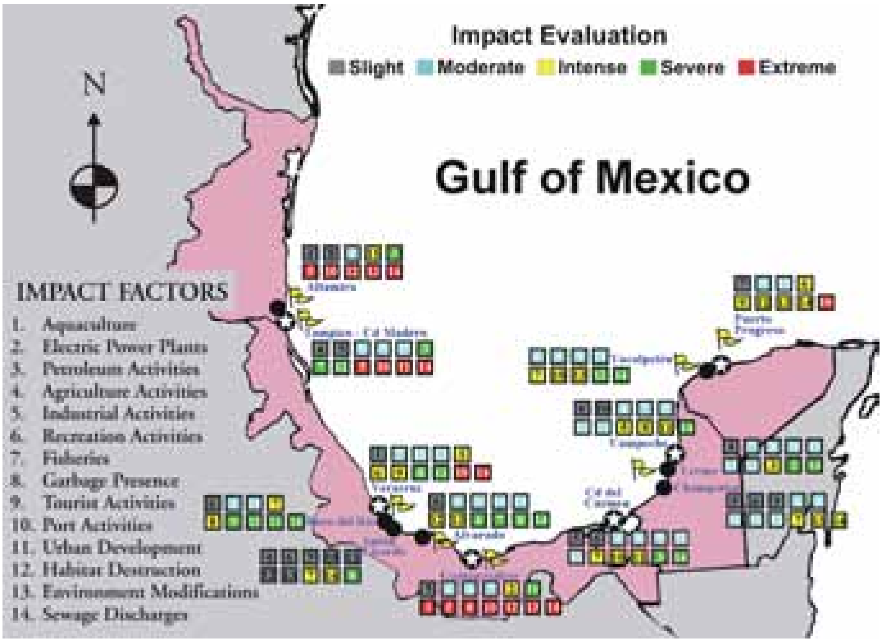


Fig. 5. Example of impacts evaluation in the Gulf of Mexico region (Tamaulipas, Veracruz, Tabasco, Campeche and Yucatan).

The graphic information provided by these regional maps, where impacts are evaluated for each site and each group of parameters, complemented by the analysis and diagnosis of water quality arranged and presented as exemplified in Fig. 3, for each State individually and the information contained in the database complemented by the in situ observations, enables us to present an overview of the situation of the whole Mexican coastal zone as follows:

- Studies about water quality were generally limited in scope and area covered, whatever the parameter or the category under study. More importantly, no time sequences could be observed which means that there is no continuity in the research projects; monitoring is rare, or if carried out, the results are not available to the public.
- The lack of communication between government agencies and public universities resulted in redundant efforts in well known areas (*i.e.* Veracruz) or left uncovered important zones (*i.e.* Yucatan, Chiapas)
- When data are available, pollution by coliforms is always present, be it in coastal zones, rivers, estuaries, bays, lagoons, beaches and sometimes in a large section of the neritic zone. This type of pollution is alarming in the large coastal cities and tourist centres, where in addition, low levels of dissolved oxygen were commonly detected. This problem is certainly due to inefficient or inadequate wastewater treatment.
- Physico-chemical parameters were more often studied than the other categories. Problems were detected close to urban, tourist and industrial centres where high concentrations of Fats and Oils as well as of dissolved and suspended solids and low values of dissolved oxygen were reported. By contrast, nutrients never represented a problem as pollutants, but discontinuity in those studies, poor sampling effort and selection of the sites (by exclusion of some 'hot spots', where information either does not exist or is not available to the public), most probably mean that future (urgently needed) studies will not confirm these results.
- The lack of information on POC's except for selected zones can be considered a serious problem: their presence in sediments and organisms, in the few studies available was alarming, since the toxicity of these commonly used products in agriculture is well known; their effects on the environment, in this case the surrounding coastal zone, should be better documented and monitored.
- In the industrial areas especially where petroleum exploitation or related industries are involved (Veracruz, Tabasco, Campeche, Oaxaca) pollution by compounds derived from those activities was notorious in water, sediments and organisms. However, the situation for the presence of hydrocarbons is similar to that of the POC's: too few records are available. In addition, some industrial areas such as Coatzacoalcos (the largest petrochemical plants in the country) as well as many ports are generally considered out of bounds for the public, so that independent analyses or evaluations can only exceptionally be carried out (this is an international problem).
- Metals are also generally poorly studied, probably due to the high costs involved in the necessary technology used for their determination. However, in selected zones like the shrimp farms areas of the Sonora coasts or the ports and areas used in activities related to the petroleum industry in Tabasco and Campeche, with adequate sampling effort, studies have shown impacts that go from 'severe' to 'extreme', again in water, sediments and organisms.
- Urban, tourist and industrial development has proceeded with no planning or controls to avoid potential damage to the coastal environment. Generally the 'charge capacity' of the aquatic systems is ignored and abuses are frequent. The intrinsic characteristics of each region are not taken into account when thinking about their future development, often

provoking social discomfort and unwanted pressures on the local populations when changes occur (such as new industries, ports or marinas, construction booms for tourism, etc.).

- When measures to protect the environment are considered, mitigation and corrective measures are commonly preferred to prevention and conservation. This was remarked especially regarding the oil industry. In addition, quite often, no measures at all are enforced, or are ineffective for the problem considered. In some cases, we observed for example waste water treatment plants which although installed, had never functioned.
- According to our in situ observations of environmental impacts together with the existing (valued) data, we can say that the areas with greater environmental damage are the big tourist resorts such as Acapulco, Ixtapa-Zihuatanejo and Cancun regarding environment modification, sewage discharges and coliform presence. The industrial ports like Coatzacoalcos, Lazaro Cardenas, Tampico-Madero and Altamira show 'extreme' impacts in port activities, urban development, environment modification and sewage discharges. The less polluted areas were found to be zones like San Felipe, Bahia Kino (Baja California) and in general the coastal zone of Yucatan.

As part of the effort of the government to act in accordance with international trends in environment protection, an increasing number of areas are being declared 'protected areas', with different degrees of protection enforced. They were 26 in the coastal zone when the Database was finished (see internet site and Fig. 3), but since then a few have been added, the latest in the Caribbean sea, protecting the coral reef areas of Cozumel (1999). In 2002, the total protected areas in Mexico, counting coastal and inland were 117 (SEMARNAP/INE, 2000).

Conclusions

The most important contribution of this study was the systematization of the whole set of available information until then dispersed and not contextually evaluated, and make it available to any interested party. Then, with the evaluation and analysis of the data contained in the database, it was possible to make a diagnosis of the water quality at the State level, and have an approximate idea not only of the coastal environment water quality but also to highlight the shortcomings, lacks and/or needs in scientific information, to evidence critical areas that need special attention and to see that corrective measures are now preferred to the more effective protection and prevention. All this will help to organize the needed research and data gathering in a systematic frame which, in addition to scientific results, will help formulate the necessary laws to protect and conserve the natural resources of the coastal zone which are now the richest source of wealth for Mexico, as well as ensure their sustainable exploitation.

The excellent initiative of declaring ever more protected areas should be continued, with all the sectors of society involved in the protection of their environmental patrimony.

Acknowledgements

Thanks are due to the INE (Mexican government) for financing this project, especially Dr A. Fernandez Bremauntz, and B. E. Lezcano Bustamante. We would also like to acknowledge all students of the Laboratorio de Ecologia Costera, ICML-UNAM, Mexico, who directly collaborated in this project, with data mining and analyses, in particular, M.A. Garcia Salgado, L. D. Ortiz Lozano, M.E. Zamudio Resendiz, M. Hermoso Salazar, and A. L. Gutierrez Velazquez. We are grateful to all the Mexican institutions, private or public that collaborated with us in providing the necessary data, and their researchers or members who in any way helped us to complete this study. Dr P. Rossin provided technical assistance and helped with artwork for which he is deeply thanked.

References

Briggs J.C. 1995. Global biogeography. Elsevier Science. Amsterdam. 452p.

Brusca R.C. 1980. Common intertidal invertebrates of the Gulf of California. 2nd Ed. University of Arizona Press. Arizona. 513p.

Contreras Espinosa F. 1993. Ecosistemas costeros mexicanos. CONABIO-Universidad Autonoma Metropolitana Unidad Iztapalapa. Mexico. 415p.

Diario Oficial de la Federacion. 1989. CE-CCA-001/89, 1989. Acuerdo por el que se establecen los criterios ecologicos de calidad del agua del 2 de diciembre de 1989.

Diario Oficial de la Federacion. 1997. Norma Oficial Mexicana NOM-ECOL-001-1996. Diario Oficial de la Federacion 6 de enero de 1997. Mexico.

Lankford R.R.1977. Coastal lagoons of Mexico: their origin and classification. p.182-215. In: Estuarine process, circulation, sediments, and transfer of material in the estuary. Cronin L.E. (Ed.). Academic Press, Inc. New York, 1977.

Lot A. and A. Novelo. 1990. Forested wetlands of Mexico. p.287-298. In: Ecosystems of the world 15, forested wetlands. Lugo A.E. et al. (Ed.), Elsevier, Amsterdam.

Merino-Ibarra M. 1990. El manejo de la zona costera mexicana: una evaluacion preliminar. p.137-155. In: El manejo de ambientes y recursos costeros en America Latina y el Caribe. Departamento de Asuntos Cientificos y Tecnologicos de la OEA y el Honorable Senado de la Nacion Argentina (Ed), Buenos Aires, Argentina.

Poder Ejecutivo Federal. 1996. Programa de medio ambiente 1995-2000. Secretaria de Medio Ambiente, Recursos Naturales y Pesca. Mexico. 171p.

SEMARNAP/INE. 2000. Caracteristicas Generales de las 117 Areas Naturales Protegidas. Direccion Tecnica de Areas Naturales Protegidas/Unidad Coordinadora de Areas Naturales Protegidas/INE-SEMARNAP, 2000. 9p.

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 271-278
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

The present and future of an integrated database on oceanology of the Southern Scientific Research Institute of Marine Fisheries and Oceanography (YugNIRO, Kerch, Crimea, Ukraine)

B.G. Trotsenko, E.V. Romanov and B.P. Panov

YugNIRO, Southern Scientific Research Institute of Marine Fisheries and Oceanography
2, Sverdlov Street, 98300 Kerch, Crimea, Ukraine
E-mail: island@crimea.com

Abstract

A brief description of the Southern Scientific Research Institute of Marine Fisheries and Oceanography (YugNIRO) status and activities is presented. The state of archives dating from the middle of the 20th century is characterized, as well as measures for its safety storage, processing and future use are under consideration. Four principal datasets specified by their origin, means of QC, methods of analysis and data presentation, namely: environmental, ichthyological and hydrobiological, fisheries statistics, and references are described in short. A listing of 15 databases available at present, though accomplished to different levels of extent, is also given.

The value of the databases originates from the method of research: collection of the information on the state of the ecosystem was carried out simultaneously with oceanographic surveys, which were complemented with meteorological observations. This multipurpose nature of primary data collected offers a scope of opportunities for further research of the marine ecosystems. On the one hand, this opens a way to analyzing the situation in the synoptic spatial-temporal scale, i.e. to define the state of fish population and its behaviour in relation to the water structure peculiarities. This type of data obtained may serve as a ground for expanding knowledge of ecosystems' mechanism of functioning (scientific aspects), and for implementing operative regulation of fishery (closed seasons/areas, quota management, etc.) and scientific advice to improve the efficiency of fisheries fleet operations (short-term forecasts), administrative and commercial applications tailored to rational and sustainable use of MLR. Data obtained by standard methods in terms of repeated (through a number of years, in the same geographic regions) integrated surveys allow to access the long-term dynamics of the state of population of certain species against long-term fluctuations of the oceanographic and meteorological modes. Besides the science-related aspects, such data maintain applied information on changes in the overall stocks and estimations of the Total Allowable Catches in the different regions, as well as the development of middle- and long-term forecasts for fisheries about catches of particular commercial species in the future.

Keywords: Databases; Informational support; Marine living resources.

Introduction

YugNIRO, the Southern Scientific Research Institute of Marine Fisheries and Oceanography, formerly (till 1989) the Azov and the Black Seas Scientific Research Institute of Marine Fisheries and Oceanography (AzCherNIRO), is the principal governmental research institute under the State Committee for Fisheries of Ukraine involved in integrated marine fisheries research. Its basic responsibility is to provide information support to fisheries on the current state and dynamics of marine living resources (MLR), marine environment, marine pollution, as well as to perform control over the fishing fleet activity, and to carry out data collection.

Such kinds of information are backgrounded by the studies on marine commercial fish populations in different areas of the World Ocean and of oceanological conditions defining the state of population and behaviour of commercial species, executed by the three departments: the World Ocean Fisheries Resources, the Azov and the Black Seas Marine Living Resources, and the Department of Fisheries Oceanology.

Archives

Since the time of the YugNIRO foundation, 80 years ago, the history of its intensive research of the World Ocean living resources has resulted in a solid database on a variety of subjects. This information defines the current state of commercial species and their natural environment. The database ‘Cruises’ includes records made by over 950 research, scouting, and multipurpose (research and fishing) ocean-going (to high seas and to EEZs of coastal countries) expeditions. It is also being planned to systematize the available material for the Black Sea expeditions.

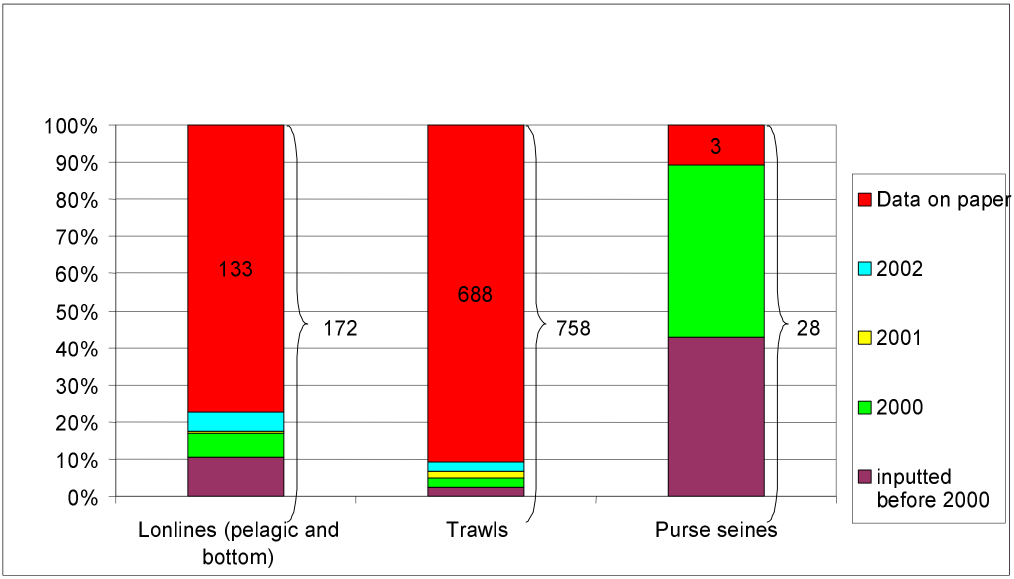
Archives of the institute’s expeditions are available since the late 1950s only, data collected earlier have not been saved. They cover primary research data from the Indian, Southern, Atlantic and Pacific Oceans, the Azov, the Black, the Mediterranean and the South-China Seas. Those are records on tens of thousands of fishing operations (hauls, tows, sets), on hundreds of thousands of biological analysis, and size frequency data of fishes, cephalopods, shellfishes, on over 70 thousand profiles of water temperature obtained by bathythermographs of different types and by about 50 thousand deepwater Nansen’s bottle oceanographic stations, including temperature measurements, salinity data, information on dissolved oxygen and nutrients concentrations accompanied by meteorological data. In recent years the archives have been updated by data collected on commercial species at sampling sites scattered all over the Crimean Peninsula coastline.

A substantial part of the above-mentioned unique materials is being stored in paper archives, that include logbooks of fishing vessels, records of ichthyologic, hydrobiological and oceanographic surveys started yet in the middle of the 20th century, and this by no means may provide the archives’ security and efficient application. At present, the accumulated materials exist mostly in single copies of original cruises logbooks, records of biological analysis, and records of oceanographic and hydrometeorological observations. They are not guaranteed to be safe in case of natural cataclysms and anthropogenic disasters, nor are they protected against natural ageing of paper and other media.

The only feasible possibility for preservation and efficient use of the archive is its transfer into electronic media, creation and maintenance of an integrated database. To follow this objective, YugNIRO is digitizing the archived data and transferring the information into electronic media. This advancement has become partly possible, due to and within the framework of several projects conducted under the auspices of IOC of UNESCO, particularly the GODAR Project, World Data Center for Oceanography-Silver Spring, NMFS. However, unfortunately YugNIRO’s own insufficient resources are inadequate, and thus they do not allow to provide the necessary volume of work in due time; proper requirements for secure data storage can not always be observed; and the software support to data processing is not always updated accordingly.

The biggest advancement in turning archives into digital data has so far been made in the section of oceanographic and meteorological information, for this database has been formalised since the very origin of data collection. The state of digitizing of the fisheries archives is illustrated in Table I. Materials of biological and ichthyological surveys in some cases do not allow any kind of formalization. In particular, formalization of footnotes and extended comments made in cruises logbooks is not possible, while they present valuable information for researchers. Such information may be scanned and stored as digital images in 'registers of field studies', to be included in the final integrated electronic product.

Table I. World Ocean Cruises Archive (number of cruises)



Databases

We distinguish four principal datasets specified by their origin, means of quality control, and methods of analysis and data presentation:

- environmental data (physical and chemical parameters of water body (SD2, MBT, EBT, CTD), meteorological parameters (METEO), and data on marine pollution) obtained in YugNIRO cruises and integrated with WODB 98, WOCE Global Data 98, TOGA, COADS 2001, GOSTAplus);
- ichthyological and hydrobiological data collected by research/scouting cruises (data on haul/set operations, composition of catch by species, biological analyses, size frequencies, ichthyoplankton and zooplankton surveys, etc.);

- fisheries statistics (total catch/effort by time/area strata and by type of vessels/gears; daily radio reports database ('OCEAN'), fishing operation database ('RIF'), and fisheries logbooks database for some areas/fisheries);
- reference datasets (our original metadata on YugNIRO cruises, sampling methods and methodology, fishing vessels of Ukraine, and such well known and widely spread databases as ASFA, GEBCO 97 Digital Atlas, FAO FishStat, FishBase 98, FAO World Fisheries and Aquaculture Atlas, AGRIS).

Available at present, though accomplished to different levels of extent, are the following databases:

1. Data on oceanography and hydrometeorology (including original research records by YugNIRO, and information obtained on exchange basis through international scientific cooperation).
2. Data on research/scouting fishing.
3. Data on commercial purse-seine tuna fisheries collected by observers aboard fishing vessels.
4. Data from tuna purse-seine vessels' logbooks and logbooks records extracts.
5. Database of daily radio reports from the purse seine tuna fishing fleet.
6. Data on research and scouting long-line catches.
7. Data of daily reports by vessels of the several fishing companies from Soviet Corporation 'Yugryba' (a. Fishing operation database 1985-1988 (RIF)). Data on catch and effort (trawl sets, hours of trawling, sets positions) by types of vessels; b. Daily radio reports database 1993-1997 (OKEAN). Data on catch and effort (number of fishing days, noon position of vessels).
8. Zooplankton tows database.
9. Data on commercial invertebrates.
10. Data of statistical reports about catches by the Ukrainian fishing companies of fish and other MLR in the Azov-Black Sea basin.
11. Data of monthly reports on catches at sampling sites and during research expeditions by YugNIRO in the Azov and the Black Seas.
12. Data of daily observations at the sampling site 'Yurkino'.
13. Data on the structure and capacity of the fishing fleet of Ukraine (fishing vessels inventory).
14. Meteorological data of research expeditions by YugNIRO, 'Yugribpoisk' fishing company, and scientific observers from the icthyolological and fishing logbooks.
15. Data on the pollution of the Azov and the Black Seas water body and bottom sediments, and of tissues and organs of commercial species by contamination (heavy metals, components of oil and chlor-organic compounds – pesticides and polychlorinated).

Tables II and III demonstrate structures of 'SD2' Database (Nansen's Bottles Data from YugNIRO original cruises, 1950-2001) and 'OCEAN' Database (Daily radio reports 1993-1997).

Discussion

The above-described databases, in our opinion, present a substantial/essential part of the global knowledge about fishing resources in particular areas of the World Ocean and of the Azov and the Black Sea basin, also being a unique database in Ukraine.

Table II. 'SD2' (Nansen's Bottles Data from YugNIRO original cruises, 1950-2001) and 'OCEAN' Data Base (Daily radio reports 1993-1997)

Record 1

No	Parameter	Position	Length	Description
1.	File id	1	1	
2.	Form number	2-7	6	
3.	Record type	8	1	
4.	Count	9-10	2	number of records type 2 + 1
5.	Ship code	11-14	4	(from SHIPS.txt)
6.	Country code	15-16	2	
7.	Station number	17-21	5	
8.	Lat. (degr., min.)	22-25	4	DDMM
9.	Lat. Hemisphere	26	1	N or S
10.	Long. (degr., min.)	27-31	5	DDDMM
11.	Long. Hemisphere	32	1	E or W
12.	Year	33-34	2	
13.	Month	35-36	2	
14.	Date	37-38	2	
15.	Start time of station	39-42	4	HHMM
16.	End time of station	43-46	4	HHMM
17.	Hour zone	47-48	2	
18.	Maximum depth	49-52	4	whole meters
19.	Cruise number	53-54	2	
20.	Start year of cruise	55-56	2	

Record 2

No	Parameter	Position	Length	Description
1.	File id	1	1	
2.	Form number	2-7	6	
3.	Record type	8	1	
4.	Line number	9-10	2	
5.	Time	11-14	4	HHMM
6.	Depth	15-18	4	whole meters
7.	T°C	19-22	4	degree to hundredths
8.	S‰	23-27	5	less to thousands
9.	Δ	28-31	4	less to hundredths
10.	O2 (mg-at/l)	32-34	3	whole
11.	O2 (%)	35-38	4	less to tenths
12.	PH (B)	39-41	3	less to hundredths
13.	Alk (mg-ekv/l)	42-44	3	less to hundredths
14.	PO4-P (mg-at/l)	45-47	3	less to hundredths
15.	P (total) (mg-at/l)	48-50	3	less to hundredths
16.	SIO3-SI (mg-at/l)	51-53	3	whole
17.	NO2-N (mg-at/l)	54-56	3	less to hundredths
18.	NO3-N (mg-at/l)	57-60	4	less to hundredths
19.	NH4-N (mg-at/l)	61-63	3	less to tenths

Table III. Structure of "OCEAN" Data Base (Daily radio reports 1993-1997)

No	Field Name	Type	Width	Dec	Remark
1.	DATA	Date	8		Date
2.	SUDNO	Character	4		Radio call
3.	RAJON	Character	4		Fishing area
4.	SHIROTA	Numeric	6		Latitude
5.	DOLGOTA	Numeric	7		Longitude
6.	ORUDIE	Character	4		Fishing gear
7.	K_TRAL	Numeric	2		Number of sets
8.	VR_TRAL	Numeric	3	1	Duration of set
9.	GLUBINA	Numeric	4		Depth (m)
10.	VID1	Character	3		Species code
11.	VILOV1	Numeric	4	1	Catch (kg)
12.	VID2	Character	3		Species code
13.	VILOV2	Numeric	4	1	Catch (kg)
14.	VID3	Character	3		Species code
15.	VILOV3	Numeric	4	1	Catch (kg)
16.	VID4	Character	3		Species code
17.	VILOV4	Numeric	4	1	Catch (kg)
18.	VID5	Character	3		Species code
19.	VILOV5	Numeric	4	1	Catch (kg)

The value of this database originated from the method of research: collection of the information on the state of the ecosystem (estimations of the stock abundance, food, localization/distribution of aggregations, and corresponding biological analysis) were carried out simultaneously with oceanographic surveys, which were complimented with meteorological observations. This multipurpose nature of primary data collected offers a scope of opportunities for further research of the marine ecosystems. On the one hand, this opens a way to analyzing the situation in the synoptic spatial-temporal scale, i.e. to define the state of fish population and its behaviour in relation to the water structure peculiarities. This type of data obtained may serve as a ground for expanding the knowledge on ecosystem functioning mechanism (scientific aspects), and for implementing operative regulation of fishery (closed seasons/areas, quota management, etc.) and scientific advise to improve efficiency of fisheries fleet operations (short-term forecasts), administrative and commercial applications tailored to rational and sustainable use of MLR. Data obtained by standard methods in terms of repeated (through a number of years, in the same geographic regions) integrated surveys allow to access the long-term dynamics of the state of population of certain species against long-term fluctuations of the oceanographic and meteorological modes. Besides the science-related aspects, such data maintain applied information on changes in the overall stocks and estimations of the Total Allowable Catches in the different regions, as well as the development of middle- and long-term forecasts for fisheries about catches of particular commercial species in the future.

The available databases and current YugNIRO monitoring activities are the principal basis for the fulfilments of Ukrainian obligations in the framework of international conventions, agreements and membership in the international organizations involved in fisheries regulations CCAMLR, NAFO, CITES, which means monitoring of the state of fisheries ecosystems in the high seas areas of active Ukrainian fisheries, monitoring of the fishing fleet activity, and development of sound management advice on the use of fisheries resources, forecasts of resource state, as well as estimation of the Black Sea and the Sea of Azov species stocks (European sprat, European anchovy, grey mullets, sturgeons, flatfishes, etc.) for fisheries regulation and setting fishing quota by the State Committee of Fisheries of Ukraine.

Thus, when developing the database, we faced a rather serious problem of arranging different types of data, also with consideration of the data interrelationships. Volumes of information that differs by contents and by forms, suggest a very specific role of experts in charge of the final integrated product development, as well as a requirement to ensure the prospective users' access to different levels of adequate information presented in the most explicit way on the specific matters of the user's enquiry.

A possible solution to the problems stated above is the proposed design of the GIS-based integrated database. This is supposed to be a comprehensive Fisheries Science/Knowledge Database, which will contain both primary data, obtained from surveys, and analytical ones and descriptive information, as a result of investigations then compiled into aggregated descriptions of certain species, their biological specifications and features of behaviour determined by the water structure of the region. The system should harmonize all scientific and fisheries data that presently exist and will be assimilated in the future.

The main objectives of the current stage of the database development are:

- to provide adequate hardware to store, process and analyze scientific and fisheries data available;
- to develop and maintain an intranet which would have access to other communication networks;
- to develop and arrange a databank with a compulsory access to the global network;
- to provide data search and data transfers of archives by YugNIRO into digital formats (primarily of the information in danger of extinction because of its paper aging under effect of time);
- to integrate into the database those materials that have been collected in expeditions by other research institutions and other databases maintained as a result of a number of projects performed;
- to pursue quality control and information correction of data entered into the database;
- to maintain adequate software facilities for archiving, processing of integrated oceanological, biological and fisheries data, and distribution of final data and information;
- training of personnel.

The next step, in our opinion, should be done to provide links to other administration's fisheries management information systems such as vessel registrations details, licensing of fishing vessels, vessels surveillance, quota management, market prices, etc. Such vessels surveillance system, based on VMS as a part of a general system, is presently being created at the division of the State Committee of Fisheries.

Conclusion

On the basis of the above-said information it will be possible to develop guiding decisions regarding priorities and strategies of sustainable use and development of MLR, on ecological monitoring of certain regions of the World Ocean, coastal zone management and sustainable development improving health and productivity of aquaculture, effective MLR management through the provision of better information. The above-mentioned specific datasets and substantial experience of personnel would make a good background for provision of MLR-risk management solutions even in cases of uncertainty.

Acknowledgements

- COD Organizing Committee
- Dr Edward Vanden Berghe – Chair of Organizing Committee
- Dr Sydney Levitus – the GODAR Project Leader
- Galya Korshunova – FORTRAN Programmer
- Natalya Romanova – FoxPro DB Programmer
- Svetlana Pronenko – FoxPro DB Programmer
- Dr Inna Petroune – English proofing

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 279-289
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Using the integrated information technology based on GIS for marine environmental data management and creation of reference books of the hydrometeorological conditions

Alexander Vorontsov[1], Nickolay Mikhailov[1], Yury Nalbandov[2] and Valentin Tuzhilkin[3]

[1] RIHMI-WDC, NODC of Russia
6 Korolyov St., Obninsk Kaluga Reg., 249035 Russia
E-mail: vorv@meteo.ru

[2] P.P. Shirshov Institute of Oceanology of RAS
23 Krasikova, Moscow, 117851 Russia

[3] Moscow State University, Faculty of Geography
Vorobjovy Gory, Moscow, 119899 Russia

Abstract

The efficiency and validity of the world ocean investigations and decisions made in the course of exploitation of marine resources depend considerably on the level of information support for this activity. This makes it necessary to treat a full technological cycle of data management, from acquisition of observational data to provision of an end-user with complex information on all environmental aspects. Now the Russian NODC is developing such technology based on modern geographical information system (GIS) to successfully support the marine environment data management and climatic research. The paper analyzes the concept, architecture and development state of information technology. The paper examines the practical samples of the specialized integrated information systems. The technology of making environmental reference books based on GIS-technology is considered in the paper. The paper discusses the results of analysis of modern electronic reference books on the Black Sea hydrology. Shown are the fragments of a new electronic guide to the Black Sea hydrology, created on the basis of GIS-technology.

Keywords: Information system; Technology; GIS; Hydrometeorology.

Introduction

Due to the requirements imposed on the information support of the world ocean investigation and exploration, it is necessary to consider a complete technological data management cycle, *i.e.* from collection of observation results to providing an end user with complex and validated information required for proper understanding of natural processes and phenomena and making correct decisions. To describe the *'end to end'* scheme of data management, the term *'integrated information technology'* (hereafter referred to as IIT) is often used (Pospelov, 1983; Fedra, 1997).

The application domain (AD) of IIT is defined by the requirements imposed on the information support for marine environment investigation and exploitation. Generally these requirements may be divided into several classes of tasks. To perform these tasks, IIT should handle data and information on a broad spectrum of various AD objects. It generates a need to create fairly complicated sets of tools to realize IIT. These tools may be divided into components and functional subsystems. The components (methods, general design decisions, data description languages, codes and codifies, instrumental software systems, etc.) form the basic level of IIT tools and comprise the environment for the development of technology functional elements. The functional subsystems are created from the components and intended to perform appropriate information activities.

The first (input) block of the technology, the archived data bank subsystem (ADB), accumulates data on marine environment, systematizes and converts them to internal information standards. Datasets formed in ADB arrive at the next block - integrated data bank (IDB). The major function of IDB is to provide integration on the basis of a more complicated and unified data model which considers both AD and functional requirements. The outcome of the IDB work consists in the complex data base (results of observations and calculations, textual data, topographical and thematic maps and others), maintained in actual conditions to be able to *'feed'* the next block – a subsystem of problem-oriented applications (POA).

In a broad sense, POA may be represented as a set of specially selected (to solve a specific task) subject – oriented data, knowledge obtained earlier and applied programmes realising methods and models of calculation of environmental characteristics. POA is oriented to obtain new information useful for selection of reasonable and economically profitable design decisions related to marine environment exploration.

The aforementioned subsystems are interconnected being developed on the basis of unified components. And whereas the ADB subsystem interacts with other subsystems on the basis of information standards (formats, metadata structure and others) only, IDB and POA subsystems have a higher level of interconnection on the basis of the client-server architecture using GIS, DBMS and WWW technologies.

In both cases the optimum organization of the data management represents the greatest complexity. In this connection when designing IIT, the aspects of representation and connections of various data types were considered, including:

- metadata - «data about data»;
- factographical data - observed, derived, calculated and modeled data;
- spatial data - electronic marine navigation and thematic maps;
- image data (graphics, figures, photo);
- textual data (documents, description, papers);
- software modules realizing particular algorithms of data processing or modeling to obtain information production.

Data management

The data management in IDB and POA subsystems is based on exchange procedures of elementary data units in client-server architecture. Thus the POA subsystem including GIS applications and certain part of the data is on the client side, and the IDB subsystem supporting the basic database is placed on the server side. The data depending on their type are placed both on the client (POA) side and on the server (IDB) side. The server side supports factographical, image and textual data as tables of DBMS or through the references to data files. The spatial data-electronic marine maps, including semantic layers of thematic maps and certain part of specialized textual and image data, are placed on the client side as an internal database of GIS applications.

If necessary GIS application has an opportunity to contact to the basic database, which, in turn, is under management of a IDB subsystem. For this purpose, GIS application addresses to the server side of technology, where the basic data base is placed under DBMS management. To use data from the basic database, the GIS-application provides ODBC links (path, names of the users, passwords, etc.) with DBMS, which are preserved as long as a session of a GIS application is running.

The GIS application data management is carried out using a special link table. This table represents a nucleus of the standard module, which is named as Navigator and which is used in nearly all GIS-applications. The Navigator module and link table were constructed on the unified classification of marine objects and are applied in all IIT subsystems.

In this case spatial data management is provided by internal GIS application tools using attributive tables. As a result, the data present in a different form (factographical, textual, graphic, spatial) and placed in the external and internal (relative to GIS) data base are jointly used by applying a Navigator link table and an internal attributive table of GIS application. The example of the interaction of the tables mentioned is shown below:

The data management in IDB and POA subsystems is based on exchange procedures of elementary data units in client-server architecture. Thus the POA subsystem including GIS applications and certain part of the data is on the client side, and the IDB subsystem supporting the basic database is placed on the server side. The data depending on their type are placed both on the client (POA) side and on the server (IDB) side. The server side supports factographical, image and textual data as tables of DBMS or through the references to data files. The spatial data - electronic marine maps, including semantic layers of thematic maps and certain part of specialized textual and image data, are placed on the client side as an internal database of GIS applications. For example, the internal attributive GIS-table, each string which corresponds to the information about one coastal station is shown in Fig. 1.

If necessary GIS application has an opportunity to contact to the basic database, which, in turn, is under management of an IDB subsystem. For this purpose, GIS application addresses to the server side of technology, where the basic data base is placed under DBMS management. To use data from the basic database, the GIS-application provides ODBC links (path, names of the users, passwords, etc.) with DBMS, which are served as long as a session of a GIS application is running.

The GIS application data management is carried out using a special link table. This table represents a nucleus of the standard module, which is named as Navigator and which is used in nearly all GIS-applications. The Navigator module and link table were constructed on the unified classification of marine objects and are applied in all IIT subsystems.

In this case spatial data management is provided by internal GIS application tools using attributive tables. As a result, the data present in a different form (factographical, textual, graphic, spatial) and placed in the external and internal (relative to GIS) data base are jointly used by applying a Navigator link table and an internal attributive table of GIS application.

A key feature of the POA is attaching of user programs, which carry out calculations of various characteristics and modeling of natural marine processes. A special input and output parameter definition language was developed for the interaction between the POA shell and the user software. The language includes a number of constructs allowing parameters specification required for the user software to be set up and run. At the same time the language allows one to describe the user interface of an application (buttons, menu bar, sliders, etc.).

The practical realisation of IIT

In the last few years the considered approach to the construction of the integrated information technology has been used to develop several specialized information systems intended to support petroleum exploration and gas deposits in the Arctic seas (Odisharia *et al.*, 1997; Odisharia *et al.*, 1998). In 1999 the work on the IIT development acquired a large-scale character owing to the development of a unified system of information about the World Ocean within the framework. The essential development of IIT components, created on the basis of Web-technology was noted.

IIT methodology is also tested in the Black Sea region through the construction of IAS *'Blue Stream'*. The model of the POA subsystem of IAS related to access and visualisation of marine environmental data in the region under study has been created in GIS media using high level algorithmic languages. It includes: data base, special software environment developed for the model and standard software GIS.

The test data base consists of spatial and factographical datasets. Space-oriented data are represented by topographical and bathymetry maps (hydrometeorology, geocreology end geomorphology, biota and others) of the Black Sea and adjacent land areas. Factographical data include temperature and salinity climate sets, as well as three components of current vectors for each month of a year.

To provide a quick access to data, data visualisation, processing and analysis as well as calculation and visualisation of additional environmental characteristics, several GIS applications were realised within the IAS model. These applications contain two groups of modules: modules extending standard possibilities GIS as far as map visualisation and attributive information describing maps is concerned; and modules ensuring acquisition of new information in GIS media with the help of factographical data processing in a mode set up by a user and subsequent visualisation of results.

The following modules are included into the first group:
- module of quick positioning of a map;
- module of construction of explanatory legend for a map or a diagram;
- module of construction of a scale rule in a map window;
- module of construction of a grid for selected area of a map;
- module of search and representation of factographical data.

The second group includes the following modules:
- module of calculation and visualisation of climate horizontal water temperature and salinity fields;
- module of calculation and visualisation of vertical profiles and time series of water temperature and salinity;
- module of calculation and visualisation of climate horizontal current vectors;
- module of calculation and visualisation of trajectories of tracers released from pressing grid points at standard horizons;
- module of calculation of spatial and time statistical characteristics by factographical datasets.

The creation of reference books of the hydrometeorological conditions

One of the major aspects of a problem of researches of seas is accumulation of the information on the sea environment for the long-term period and reception of the generalized data on a mode of the sea as the various reference books distinguished by the big variety both on a set of parameters, and on scales of processing of the information.

Requirements of today have changed our representation about reference books, - convenience of use, efficiency of their preparation and a possibility of new information technologies allow to speak in the majority the about electronic reference books as all variety of handbooks can be united and unified.

The modern computer technology allows to unite at a functional level of a database, methods and models of calculations, standards and manuals, system and application programs as the integrated information media based on commercial GIS and DBMS for reception of the complex information on a condition of the natural environment. All this allows creating the reference book more mobile by way of inputting the new data and reception of the updated calculating and modeling characteristics.

Thus, it is possible to define, that in modern understanding the reference book in electronic kind (ERB) can and should include not only information base (the observing, processing, modeling and help data), but also a processing and modeling set of programs of reception of regime characteristics and results hydrodynamic modeling.

Basis ERB is technological media storage, management, processing and formation of output production as cartographical, text, tabulated and graphic materials. (GIS, DBMS, Internet/intranet, etc.) problems of creation ERB various ways allow to solve development of new information technologies. The new generation electronic handbook should work both in network variant, and in the variant which has been written down on a compact disc.

ERB today is the integrated information technology of reception of a full set of characteristics of the marine environment in client-server variant for which the powerful program complex of preparation of output production on the central server is developed, it is formed full DB (a set of observational and regime materials as the tabulated, graphic, text data; metadata and the thematic maps for each type of data).

It is obvious, that development of new information technologies problems of creation ERB, as allow to solve a minimum, two various ways. The first can be named off-line or static (electronic analogue before published regime and climatic reference books), the second - on-line or dynamic (realization 'in alive' calculations and reception of statistical and modeling characteristics on the basis of an existing database).

Now two variants of basis ERB are realized: as DBMS- and GIS-applications.
Further we shall consider practical realization IIT as the information system functioning in GIS media. The system is developed as an open software product allowing for modernization and enhancement of functionality. Developments involve the modern information technologies (problem-oriented programming, DBMS and GIS) and are conducted in the distributed environment of the client–server architecture.

The basic idea of the integrated marine information system (IMIS) is to develop 'transparent' interfaces between a variety of marine environmental information resources coming from different sources such as observational networks, historical databases, analytical synthesizing, mathematical methods and models for providing access to available data and new information with a view of improving our knowledge of marine phenomena and facilitating to efficient and effective marine-related planning and decision making.

The IMIS (Batalkina *et al.*, 1999) development is based on the integration principle (sometimes it is called 'end to end' technology).

Therefore in a general form IMIS is the integrated set of subsystems designed to provide user services. Each subsystem can include several invariant functional blocks. The client-server architecture has been chosen as the basis for IMIS implementation: the server is responsible for data accumulation and integration in the distributed database and the client executes intranet applications designed to implement applied tasks (within the scope of the predetermined subject). The client is also responsible for access to data and information products under appropriate information protection conditions.

It is noteworthy, that along with specific features typical for concrete environment-related systems there are standard models typical for all types of systems. The main thing is that the system should be built on the multi-tier principle. Individual subsystems such as the integrated database subsystem can be considered as individual layers. Smaller layers such as the specialized software, the data flow management system, the ancillary software needed to ensure communications and to provide access to the functional part of the system can be further considered within individual subsystems.

The advantage of the multi-tier principle is that it allows for modifications of individual components in one layer leaving other layers unchanged and ensures formal specification of interfaces between the layers to enable independent development of information technologies and related software. In this case open standards will permit smooth replacement of software modules (*i.e.* replacement of a server or a DBMS). In addition the multi-tier approach will enhance reliability and stability of the system.

Generally when information systems are developed a number of interrelated structural elements are identified such as components, functional subsystems and workstations. In this case the base level of the system is represented by the components constituting methodical, organizational and technological environment for development of the system, and the functional level is represented by thematic subsystems responsible for specific information activities. The subsystems in their turn can be divided into functional blocks-elements implementing specific tasks. The subsystems and individual blocks are created through development of the appropriate applications on the basis of the system components ensuring the unification of the system implementation tools.

The basis of IMIS is formed by the set of components (scientific-methodical, information-linguistic, program-technological, etc.) creating environment for interaction between technological blocks and information flows. The scientific-methodological component is responsible for thematic orientation of the system determining specific features of its operation. The information component includes the information fund and data unification and encoding tools used to represent data in the databases of the system. The program-technological component ensures IMIS operation on the basis of the DBMS and GIS applications. The system components are combined in the form of several subsystems.

The first (input) block of IMIS - the integrated database (IDB) subsystem - accumulates the necessary data on objects of the domain, systematizes and converts them into internal standards for representation of declarative (factographic, textual and spatial) and procedural data (models) on marine environment. The IDB core is formed by the complex database (DB) updated on a regular basis to *'feed'* the next block – the subsystems of problem-oriented applications (POA).

In a broad sense POA can be conceived as a set of specially selected (to solve a specific problem) thematic data, knowledge obtained previously and applied programs implementing methods and models of calculating marine environment characteristics. The POA information-technological complex is aimed to process, analyse and interpret data, and to represent information products. The separate IDB/POA block is responsible for user access to data and information as well as dissemination of information products. In terms of functionality the calculation/model block should be noted. It includes calculation procedures integrated into IDB and a number of analytical modules from POA.

Noteworthy also is the section of topographical bases, which includes marine maps of a various scale. The topographical bases are applied in thematic tasks as a basis for representation of metadata (reference information) at the maps of the areas selected.

In fact IMIS is the combination of various IDB/POA elements interconnected by information-technological interfaces when the output data flow of one source (technology) serves as a *'transparent'* input data flow for the other.

Structurally the test version of IMIS includes the specialized DB, operating in the DBMS environment; the DBMS-application block performing a number of functions in the DBMS environment; the basic GIS-application, within which the user interface and basic data visualization and field construction functions are implemented.

The general framework of IMIS is shown on Fig. 1.
Data management in the system is based on the database design and on object (or logic item) management. In the system by the object of management the elementary domain component resulting from structuring is meant. It is presented in the form of an individual item, which can originally be described by:

- data kind (pollution, hydrology, etc.), groups of parameters and individual parameters within the data kind;
- data type (metadata, data on space-time distribution, outcome of observation, generalizations etc.);
- data presentation type (map-based layer, covering, graphics, textual ASCII or DOC-description, dbf-file).

Object management has three steps:

- interaction between communication tables and files-catalogs by the use of an identifier;
- file-catalog transfer by the use of an identifier;
- interaction between reference tables and catalogs.

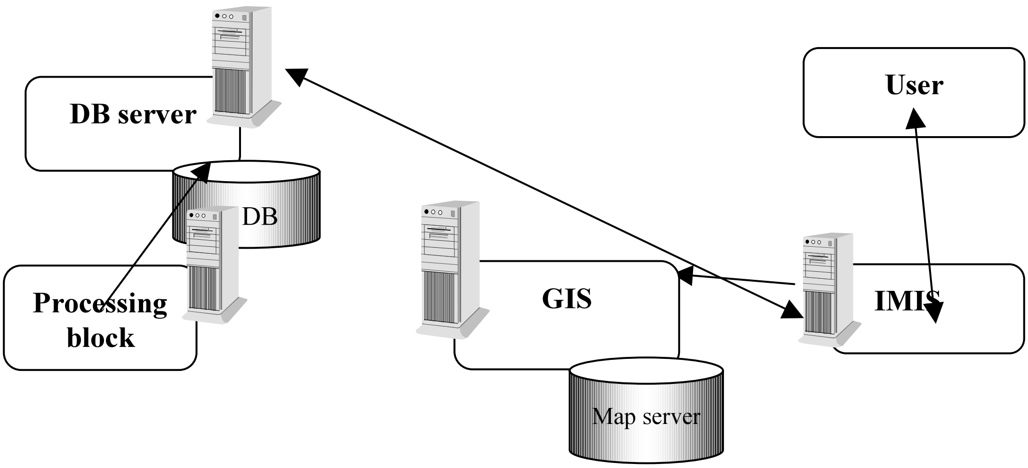


Fig. 1. The framework of IMIS.

Structurally the IDB subsystem includes a database and a set of technical and software-technological data management and processing tools. The following types of input objects should be included into IDB:

- characteristics of environmental conditions;
- data collection, accumulation and processing technologies, methods and models including calculations and modeling of environmental characteristics.

IMIS includes a number of tools and modules:

- the system shell in the GIS ArcView 3.2a environment responsible for IMIS operation control and maintenance of intersystem links;
- the DBMS-application for the XOZ graphs which allows the XOZ graphs to be plotted by a random sample;
- the DBMS-application for XOt graphs which allows a parameter time variation to be plotted by the data being selected;
- the XOY isoline tool in the form of the GIS-application based on Spatial Analyst 2.0, which allows fields to be constructed in a specific standard (default functions) and by the choice of a user;
- the table data visualization module – a tool to compose and display the tables of a specified type;
- the textual data visualization module – a tool to compose and display the textual data in the *.doc and *.txt format;
- the graphics visualization module – a tool to compose and display graphics of specified formats;
- the file Navigator - a tool to manage files of specified extensions (selection, viewing, processing etc.);
- the module to save data in a specified format;
- the standard GIS-function-based module to generate reports.

On Fig. 2 working panels of IMIS (managing panel, display in separate windows metadata and a thematic map) are shown.

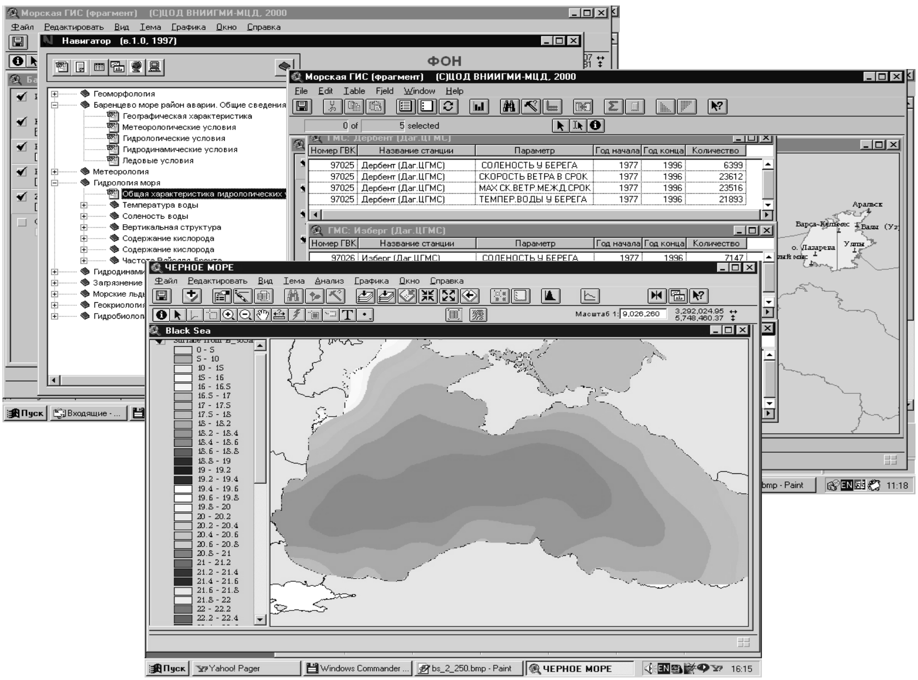


Fig. 2. Working panels of IMIS.

Conclusion

Today IMIS is in the starting model stage. The basic functional task is to visualize the initial and processing data generated in the DBMS environment and to put into operation some problem-oriented applications in the GIS environment.

In the future the full-function information system should be developed, including its ability to operate in the distributed mode. And ERB on the marine environment are realized as a complex set of the problem-oriented appendices which work should be carried out on the server of IMIS, and the user should receive only results.

One of variants of such creation of appendices is based on application of technology Taxxi - the most advanced embodiment of the concept of the objective surface offering uniform decision of problems of distributed systems connected to development. This technology provides the low traffic, minimizing volume of the sent data that comfortably allows to work with services even at access on lines with low throughput.

Convenience of application of the given technology is caused also by that the user computer is protected from the non-authorized access, as from network Taxxi Communicator (client part) receives only passive XML-descriptions, instead of active Java-applets. It means, that there is no basic opportunity of automatic installation of another’s executed code.

The complex of appendices for preparation of output production as ERB on marine environmental now is created as the separate project.

The basic components are:
- the control service of access;
- the managing center;
- the set of application;
- the local database;
- service of auto updating of the data.

For an example on Fig. 3 panels of the managing centre and the application for sample of the data are shown.

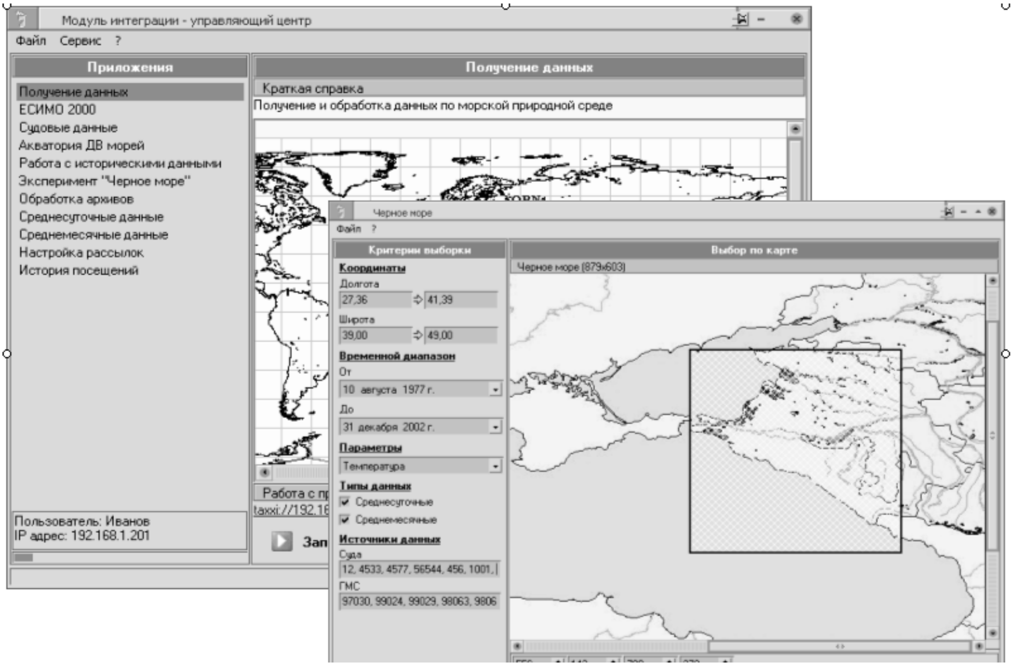


Fig. 3. Managing centre and the appendix for sample of the data.

References

Batalkina S., N. Mikhailov, V. Tuzhilkin and A. Vorontsov. 1999. Marine Information System in Oceanography and Meteorology of the Black Sea. p.234-236. In: Proc. International Conf. Oceanography of the Eastern and Black Sea, 22-26 February, 1999, Athens, Greece.

Fedra K. 1997. Integrated environmental information systems: from data to information, p.367-378. In: Integrated approach to environmental data management systems. Kluwer Academic Publisher, Dordrecht.

Odisharia G.E., L.V. Shershneva, A.S. Tsvetsinsky, N.N. Mikhailov, B.V. Arkhipov, A.A. Vorontsov, S.A. Batalkina and V.S. Tuzhilkin. 1998. On application of GIS technologies for design of Jamal gas production and transportation sites, ArcReview, 1998, 4(7):12-13.

Odisharia G.E, A.S. Tsvetsinsky, N.N. Mikhailov and G.I. Dubikov. 1997. Specialized information system on environment of Yamal Peninsula and Baydaratskaya Bay. Proc. Int. Offshore and Polar Eng. Conf. (ISOPE-97), Honolulu, USA, 1:574-581.

Pospelov G.S. 1983. Artificial intellect: modern information environment - Vestnik AN USSR 6:31-42.

PANEL DISCUSSION

Vanden Berghe E., M. Brown, M.J. Costello, C. Heip, S. Levitus and P. Pissierssens (Eds). 2004. p. 293-307
Proceedings of 'The Colour of Ocean Data' Symposium, Brussels, 25-27 November 2002
IOC Workshop Report 188 (UNESCO, Paris). x + 308 pp.
– also published as VLIZ Special Publication 16

Summary report of the panel discussion

In the context of ocean data management, scientists, data managers and decision-makers are all very much dependent on each other. Decision-makers will stimulate research topics with policy priority and hence guide researchers. Scientists need to provide data managers with reliable and first quality controlled data in such a way that the latter can translate and make them available for the decision-makers. But do they speak the same 'language'? Are they happy with the access they have to the data? And if not, can they learn from each other's expectations and experience?

The last time slot of the symposium was used for a panel discussion. The main objective of the panel was to bring about a discussion between data centre managers on the one hand, and users of the data, in the first place the scientists, on the other.

There were two panel members from each of the data management and the scientific communities, and from international organisations. The panel discussion was divided into two parts; the first part consisted of short opening statements by the panel members, based on the opening questions listed below; each panel member covered 6 questions: three general, three specific for the community (s)he represented. The second part was dedicated to open debate.

Panel members

- Chair: Savi Narayanan, MEDS, Canada
- Representatives from data centres:
    - Lesley Rickards, BODC, UK
    - Catherine Maillard, IFREMER/SISMER, France
- Representatives from the science community:
    - Peter Herman, NIOO/CEME, The Netherlands
    - Neville Smith, Bureau of Meteorology, Australia
- Representatives from international organisations:
    - Alan Edwards, EU
    - Peter Pissierssens, IOC/IODE

Synthesis

We have tried to summarise and synthesise the main points that were discussed during the conversations in this section. It should be noted that the debate was a very lively one, and much of the input that is now part of the synthesis below came actually from the members of the public.

Changing role of data centre

Changes in technology have been leading to changes in the role of data centres. There is a trend to move away from the traditional data centre, with its main task of archiving datasets, to become more service-oriented.

Data centres can look towards libraries for inspiration to redefine their role; libraries provide expertise and guidance in cataloguing. Archives are grey and dusty, libraries are active and open; data centres should strive to resemble the latter rather than the former. Data management needs an equivalent to the 'Web of Science': a mechanism to bring up a list of relevant, available, quality controlled, peer reviewed datasets.

There is a need to create data and information products; not only towards other data managers and scientists, but also towards the policy makers and the society at large. These products will assist in increasing the visibility of the data centres, thus assisting in attracting both funding for further activities, and data submissions from scientists.

Some traditional roles of data centres remain important: long-term stewardship of data, integrating datasets, documenting and redistributing datasets, development of standards and standard operational procedures…

Bridging the gap between scientists and data managers

Both data centres, and data and information management procedures are very poorly known by marine scientists. In most university programmes, there is no training on data management, no information on data centres, data management procedures… Data management is perceived too much as an IT topic. There is a need to investigate how to put data and information management on the curriculum of academic institutions. This would result in a better knowledge of the data centres, and an increased quantity and quality of data submissions.

Data managers should actively seek collaboration with scientists. If data managers have a background in science, it is possible to establish a relationship of trust with the scientists, a smoother collaboration, and a greater input of the data managers in the development of data collection. The involvement of the data managers in the planning of projects from a very early stage makes 'End to end data management' a reality.

EU has the mandate and the funds to support and improve training for scientists in data management, and could be playing a role in this.

Creating incentives for scientists to submit data to data centres

To a large extent, data centres are dependent on scientists to submit data. Especially in view of the extent to which scientists are not aware of the role or even the existence of data centres, this is a potential problem. Several actions can be taken in this respect.

- Creating awareness about importance of data management, by *e.g.* including data and information management in the curriculum of universities;
- Requirement for data management written into project condition for funding – is already the case for EU proposals, and happens for short-term data management;

- Developing peer review and quality control procedures, to assess usefulness of a dataset, and making a dataset citeable, so that a scientist's contribution of data to a data centre can be measured, and taken into account for career advancement.

Need for long-term activities

Datasets often result from projects, which usually have a limited time-span. Data management on short term, within the time span of the project, is usually no problem: scientists do need data management to produce the deliverables to the project; moreover, making provisions for data management is a prerequisite to have a proposal accepted in the first place. There is an obvious need for activities beyond the duration of the data-generating project, to assure continued availability of the data. This always has been, and probably should remain, one of the tasks of data centres.

Funds for long-term data management should not come from research budgets, but rather from operational networks or other mechanisms. Several initiatives of the EU are relevant in this respect. Within Framework 6, there is a possibility to fund the operations of large 'Networks of Excellence' that will operate on time spans much longer than a typical project. The Global Monitoring for Environment and Security (GMES) initiative is another potential mechanism.

Duplication of efforts

A certain degree of duplication is unavoidable, and is a fundamental aspect of the scientific process. There has to be room for experimentation, different attempts at solving the same problem. After some time, however, experimenting should stop and be replaced with one or a couple of strategies.

Undesirable duplication can partly be stopped during the project proposal review process. One of the objectives of the Networks of Excellence, as proposed by the EU, is to increase communication between partners of the network, raising awareness of each other's activities, and hence decrease the probability of duplication.

Need for peer review of datasets, and for standard practices

There has to be peer-review, as a way to measure and recognise progress, to recognise value and expertise, and as a foundation for standards and accepted procedures. Standards and audit procedures are needed to allow objective peer review. Developing these standards is a task for the data centres.

Peer review is a way to increase the compliance with standards. Countries, or even institutions or scientists, could be tempted to work along principles that were developed locally; obviously, these will fit local needs, and are usually much faster to develop. Doing so, however, can lead to fragmentation, and hamper data exchange.

Difference between biological and physical data management

The problems of biological and physical data management are different: physics datasets are often high volume and low complexity; biology datasets are low volume but high complexity. Taxonomy brings a 5^{th} dimension to ocean data management.

The lower level of standardisation in biology makes importance of proper documentation with the datasets even greater.

Commonalities are more important than differences: both biological and physical data management need for long-term activities; quality control and peer review; creation of data products

Involving the developing countries

Participation of developing countries in global programmes is the best way to transfer expertise. Global programmes can operate at several levels, so that they can serve both global and local needs.

Internet access is a problem in many third-world countries, and assisting with connectivity and basic telecommunications should be made a priority in any capacity-building programme. Where internet is available, the bandwidth is often very limited, making it virtually impossible to download large volumes of data. As long as this problem remains, data should also be distributed on alternative carriers such as CD ROM or DVD. Data warehousing and brokering can assist in locating and selecting relevant datasets, thus limiting the volumes of data to be downloaded.

Also funds to purchase hard- and software, and expertise to maintain the systems, are a factor that is more limiting in developing countries. The data management community should provide platform-independent software that is open source and runs on hardware that is compatible with technological expertise available. Reliable and stable standards should ensure that data are available in a form that can be handled by these tools. Capacity building programmes should be organised making use of these tools and standards.

Suggestions for actions

Investigate how data and information management can be made part of the curriculum of marine sciences in academic institutions;

Develop standard operational procedures and a peer review process to allow an objective assessment the quality of datasets;

Guide the user community directly to relevant, and quality controlled datasets, by setting up portal sites;

Create integrated data products, to increase the visibility of the data centres;

Distribute data not only through the internet, but also on CD or DVD;

Assist third-world countries with basic telecommunications, Internet access, and data warehousing;

Create a collection of open source, platform-independent software for the benefit of third-world countries, and organise capacity building programmes around these.

Panel starting questions

Data centre representatives:

1. What do you see as the role for data centres in managing data from the global science programs?
2. What are the challenges you see that data centres need to face individually and collectively?
3. What added value comes from managing data in data centres, rather than in the originating institutions? What should a data centre have on offer to be more than just a convenient data archive?

Scientists:

1. What are your expectations from the global network of data management systems?
2. Can the global network meet your expectations now, with some changes or with radical changes?
3. What governance structure would ensure effective and efficient management of global data, assuring and documenting data quality, securing data for future generations, and providing easy access to integrated multi-disciplinary data?

International organizations:

1. What is the role of international organizations to address the data management requirements?
2. What do you think are the major challenges that the international organizations face in global data management?
3. What changes should be implemented at the international level to better deliver the global data management mandate?

All:

1. What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?
2. What do you see as the main differences in data management practices between biological and physical oceanographers? What can be done to bridge these differences?
3. If you had three wishes to improve global data management, what would they be?

Lesley Rickards

What do you see as the role for data centres in global science programmes?

- Partnership with scientists to create integrated project datasets
- (Scientific) quality control and documentation of data is crucial
- Long term stewardship and archival of datasets: archival; scientists move on to the next piece of work
- Infrastructure and IT expertise
- Development of standards

- Access to other datasets

What are the challenges you see that data centres face individually and collectively?

- Adequate funding: data management is left to last, when money has run out
- Working collaboratively whilst responding to national remit
- Increasing data diversity (many different parameters being measured)
- Working to common standards
- Adapting to new technology (distributed systems, XML, etc.)
- Developing systems to deliver (near) real-time data
- Effective data dissemination mechanisms

What added value comes from managing data in data centres? What should a data centre have to offer to be more than just a convenient archive?

- Integrated datasets and value-added products
- Long term stewardship of data
  - End-to-end data management
  - Pre-cruise to final data product
- Real-time through to delayed-mode
- Effective dissemination mechanisms
  - *e.g.* CD-ROMs, DVDs, internet/web-based
- Advice on quality control & standards
- Use of new technology
  - *e.g.* distributed systems, XML, etc.

What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?

- Ensure data and information are available in the most appropriate form for both – remembering that the same solution is not always appropriate (*e.g.* web vs CD-ROM)
- Sharing of expertise and working together
- Ensuring high-quality capacity building programmes
- Encouraging the participation of developing countries in global programmes (with local/regional applications) *E.g.* GLOSS: serves both local and global needs, at different levels
- Platform independent software

What do you see as the main differences in data management practices between biological and physical oceanographers? What can be done to bridge these differences?

- Physical oceanographers often more computer literate (*e.g.* develop their own software for processing, quality control and analysis)
- Some measurement techniques more standardised
- Biological datasets may be more complex and require much more supporting documentation to describe collection methods, etc.
- Development of flexible data storage mechanisms to accommodate both biological and physical data
- Provision of (appropriate) software tools
- Provision of integrated datasets

If you had three wishes to improve global data management, what would they be?

- Proper collaborative efforts, building on existing standards and practices – not reinventing the wheel each time a new project comes along – *e.g.* WOCE (continued into CLIVAR?) Danger of loosing expertise in the intervening period.
- Recognition of the value of data management and stewardship by the scientific (and operational?) community. Proper recognition of the value of data management and data stewardship long-term archival from the scientific world and the operational communities, so that it doesn't fall off the bottom of the list each time
- Adequate funding also after termination of short-term programmes

Catherine Maillard

Role of the data Centres in managing data from the global science programs

- To offer a perennial (long term stewardship of the data) and high technology level infrastructure and professional staff (to make sure the hard- and software available is matched to the latest developments of technology) to contribute to the project data management structure
- To disseminate the existing international standards and expertises as a minimum basic requirement for data management, especially for the meta-data and the quality control assessments – avoid re-invention of standards
- To provide standardized complementary datasets to the projects – *e.g.* data produced by other projects, historical data

Challenges that the data centres need to face individually and collectively

- To be professional (first priority is data dissemination/redistribution; first priority of scientists is to do science, not sharing data) and operational to insure that all the data are
    - o safely archived
    - o checked for quality
    - o timely processed and delivered - disseminated by internet and other adapted communication systems
- To return products to the data providers and the other users
    - o integrated observations datasets from various sources
    - o value added synthetic products
    - o contributions to scientific papers (like happens in *e.g.* OCL) – data managers should not compete with scientists! Data managers have to contribute and be visible, but in a very specific way.
- To convince the data collectors and the supporting agencies that it is valuable to spend a low percentage of the overall cost of the data collection for the archiving. Comparison of cost of ship time with money for data management

What added value comes from managing data in data centres?

- Providing a perennial archiving system able to cope with the changes in the media technology
- Insuring that the data documentation is in accordance to the international standards and the data organization not 'person dependent' = main difference between data manager and scientist
- Avoiding the dispersion of the data in hundreds of laboratories or scientific teams with different data management methodologies

- Assembling long time series of data of the same type
- Providing timely access to data through dedicated data servicing teams
  - providing data in real time
- Optimising the costs and manpower

What to do to reduce the impacts of technological differences between developing and developed countries?
- To insure one WWW link to all: that is possible, at least with limited speed
  - also distribute data on alternative carriers
- To insure the participation of both developing/developed countries in the development of common databases
- The data policy should insure a wide data circulation
- The developing countries should also contribute to the international effort
- The job training on practical issues should enhance the mutual sharing of expertise

Differences in data management practices between biological and physical oceanographers?
- More complicated datasets in biology:
  - a 5th (or even 6th) dimension with the taxonomy, genetics…
- Necessity of linking environmental parameters, bio-chemistry and biological observations in different compartments water column, SPM, sediment, biota ..
- Less practises of data exchange and computer technology among the biologists, including the documentation of the coordinates
- The bio observations are frequently too dependent on the measurement technology to be inter-comparable

Three wishes
- A common work on the standards for various parameters, not only for temperature and salinity
- A better acknowledgement of the data submission from the experimentators: the submission of a data set that pass the QC at a data centre should be recognized as a 'rank A' publication
- A better public awareness of the data legacy and the 'data librarian duty': how to follow the climate change without long time series of good data?

Peter Herman

What are your expectations from the global network of data management systems?
- Easily accessible, highly resolved ecological data
  - Geo-referenced
  - Consistent taxonomy
  - Auto-ecological information (as in FishBase, AlgaeBase…)
  - Well-documented methods
  - Physical and chemical data (depth, light, chlorophyll, nutrients, sediment composition, physical stress…) linked
- Spatio-temporal variation represented

Can the global network meet your expectations now, with some changes or with radical changes?

- More, especially coastal, datasets would need to be made available
- Emphasis should not only be on species occurrences, but also on density, biomass, functional types,...
- Digested physical, chemical and biogeochemical information would be very worthwhile (+ access to primary data)

What governance structure would ensure effective and efficient management of global data, assuring and documenting data quality, securing data for future generations, and providing easy access to integrated multi-disciplinary data?

- Distributed databases to avoid problems of (incomplete, faulty) duplication
  - o It is hard to transfer biological data from one database to another
- Citable databases? Or: storage of databases in conjunction with publications? In any case: an incentive for the researcher is needed
- Peer review of database quality
  - o Even more important for biological data than for physical data
- Reliable exchange protocols, including taxonomic standards

What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?

- Reliable, stable standards. Data managers are trying too hard to keep up with recent technological developments, at the expense of stability, and the opportunity to bring developing countries up to the level needed to contribute to and to take advantage of global data management
- A comprehensive, consistent set of open-source software according to these standards
- Time to develop systems and build capacity before everything changes again

What do you see as the main differences in data management practices between biological and physical oceanographers? What can be done to bridge these differences?

- The huge variety of variables (every species being one, plus all the measurements on these species) in biological studies
- What can be done: continue to decrease biodiversity?

If you had three wishes to improve global data management, what would they be?

- More involvement of scientists as data suppliers/users, through appropriate incentives
- Technology that promotes active participation of non-wizards, *e.g.* in distributed databases, instead of xml-ing them out
- Comprehensive portal sites that do not lead to metadata-basis nonsense, but directly to data and information

Neville Smith

What are your expectations from the global network of data management systems?

- Efficient and effective communications
- Standard for marine metadata
- Leadership in quality control, quality assurance and data set integrity: assembly
- Data communications, transport

- Responsibility for archives
- Partnerships with science - data and information services

Can the global network meet your expectations now, with some changes or with radical changes?
- There are community attitude changes that are required
- Moving away from ocean data archiving, to ocean data services
- It requires collaboration and cooperation across a broad community
- Questions of strategy and tactics
- Do we trust in natural evolution? [No.]
- Will the sum meet expectations? [I don't know]
- Is it possible to productively harness initiatives

Governance?
- A disciplined approach
- Responsibility for infrastructure
- It seems a single structure can stretch across all of the ocean/marine disciplines.
- It must enjoy the same status and recognition as a science program and work seamlessly between science and operations/applications.

Developing and developed countries?
- Using practices and tools that are commonly available
- Data warehousing and brokering
  - Production of data products by data centres; interpretation of data, filtering of data
- Assisting in basic telecommunications

Differences between biological and physical data management
- Physical data has simple characteristics c.f. biological
  - Biology: complex structure but low volume
  - Physics: simpler structure, but high volume
- Physical community should regard biological community as an important client
- Marine XML a common tool
- Assembly standards: assurance
- Innovative transport and access methods
- Commonalities are more important than differences
  - Need for peer review system

Three wishes
- Community attitude: make scientists and data managers work as one
- Dynamicism and integration; moving towards a more service-based system
- Frontier projects – do things in a different way, not necessarily with new technology.

Alan Edwards

What is the role of international organisations in addressing data management requirements?
- To provide a suitable framework for effective data management (framework is more than just policies, it also has to include at the very least funding)
  - This should acknowledge and protect the rights of ownership

  - o Take account of commercial considerations (not necessarily at the point of delivery, but somewhere in the loop)

What are the major challenges that international organisations face in global data management?

- Compliance
  - o It is one thing to have a ‘policy’; it is another to have everyone respect this.
  - o The carrot and the stick: funding. But there should be more than this; it should be a partnership
- Rapid technological change
  - o This can lead to policies/legislation being redundant before they are adopted.

What changes should be implemented at the international level to deliver better global data management?

- Proper investment in data management.
- Recognition that good data management serves multiple purposes, especially the provision of information products. Products should be seen to contribute to societal needs by politicians
- Actual mechanisms to allow users to properly define what information products they need and in what formatHow can international organisations help to advance Data Management?
- By establishing a proper legislative and / or operational framework to support data management. [*e.g.*, within the 6^{th} Research Framework Programme and the GMES Initiative (Global Monitoring for Environment & Security)]
- By ensuring adequate resources for data management initiatives are available.

Peter Pissierssens

What is the role of international organizations to address the data management requirements?

- To facilitate and promote the management and sharing of data at the global level
- To develop and promote standards
- To assist countries in acquiring the necessary capacity to manage their national data and to participate at equal level at the international level
- To promote cross-sectoriality

What do you think are the major challenges that the international organizations face in global data management?

- Countries preferring to use their own standards rather than adopting those internationally agreed upon
  - o Problem of inertia: it takes too long to agree upon a standard; by the time the standard was agreed upon, it might be obsolete
- Countries placing national short-term gains over long-term global benefits
- Countries unable or unwilling to contribute resources for international initiatives

What changes should be implemented at the international level to better deliver the global data management mandate?

- Countries to actually implement internationally agreed-upon standards
- Countries to actively contribute to international programmes

What data management practices can be employed to reduce the impacts of technological differences between developing and developed countries?

- Wrong question: we need to close the divide between developed and developing in terms of technology availability. We cannot stop technological development so we need to ensure global access to the technology (problem of stability of the technology!!)
- Possibly concentrate on low-cost, platform-independent and open source technology that is therefore more easily accessible (see examples from COD)
- No science without library, no science without communication; spend money on internet access! This was one of the priorities of the Odinafrica project

What do you see as the main differences in data management between biological and physical oceanographers? What can be done to bridge the differences?

- The main difference is the time it takes to process the samples; moving from delayed to real-time is difficult
- To some extent technology could assist (image processing & recognition…)

If you had three wishes to improve global data management, what will they be?

- Real implementation of agreements and standards
- Real commitments and giving more attention to international over national interests
- Coordination and cooperation between all stakeholders and data centres
  - o Do not give up on international initiatives; it is possibly easier to develop isolated solutions, but this would lead to a multitude of systems that would not communicate

General discussion

Don Robertson: Data management left to last minute/dollar; remedy: price data management into contract; last money is only paid on submission of data; this is a simple contractual agreement

Lesley Rickards: legislation would be more of a problem in government departments; in the BODC, it is mainly a matter of intellectual property rights, which are acknowledged in the data centre's standard procedures (the originating scientist's name gets attached to the data set).

Catherine Maillard: Data from other country's EEZ does not get distributed; other types of data with restrictions are eg data relevant to national security

Franciscus Colijn: Too much reinvention of wheels; there should be mechanisms during project proposal evaluation to avoid duplication

Savi Narayanan: Are there plans to solve problems of duplication and of life cycle planning with the funding organisations?

Alan Edwards: Part of the FP 6 mechanisms is created to alleviate the problems of fragmentation and the need for long-term activities. Large Networks of excellence created to avoid duplication, increase communication, and increase integration

Peter Pissierssens: Problem that projects are limited in time, have to produce deliverables after 3-4 years; projects are seen as discrete events, with a beginning and an end. What to do with needs for long-term activities?

Alan Edwards: Research budgets are not appropriate mechanism for very long term activities. There is a tendency to increase project duration, but even this will not cover adequately the needs of long-term activities. This has to be done by funds for *e.g.* operational networks, or through activities like GMES.

Jerry Miller: Most innovative work has been done by a combination of a small amount of own data with large amounts of other people's data. No programmes are funded unless full sharing of data. Sharing implies data management, and this obviously requires funding. Proposals hardly ever include Principle Investigator for Data management.

Peter Herman: EU proposals have to include funds for data management, and during the review process this requirement is checked. Often during the negotiation phase the amount of resources allocated to data management is increased.

Peter Pissierssens: In how far is DM part of academic curriculum, of the scientific culture? Do scientists find their way to data centres? Do data managers visit the scientific institutions?

Savi Narayanan: Scientists do not know about data centres.
Murray Brown: 10 year ago, people did refuse to share data. Since then things have changed. There are now contractual obligations to submit data, and this involves data sharing and management. In this process, scientists have learned to work together with data managers. Duplication of efforts is not necessarily bad in its own, and unavoidable as part of the scientific process. It is not possible to make a big master plan to stamp out all duplication.

Neville Smith: We should not strive at a one-size-fits-all; 'experimenting' and duplication is indeed part of the scientific process. But there is a time that experimentation should stop, and replaced with one or a couple of strategies. Data sharing is a means to an end, not a goal in itself. Data sharing has to bring added value. Standards' sharing is on another level, and useful in its own right.

Savi Narayanan: Brings up Peter's point again.

Karin Stocks: It is necessary to train people in data management; but it is also necessary to have professional recognition for people spending time on data management task, *e.g.* reviewing quality of datasets and web sites, so that data managers can get professional recognition, can put data management activities on their CV.

Daniel Davis: There is not enough incentive for scientists to work together with technologists; ocean science is lagging behind in technology.

Catherine Maillard: Peer refereeing of datasets has drawback: it weighs on the time span on which the data can be delivered.

Edward Vanden Berghe: Standard operational procedures by which dataset is generated can be quality-controlled, rather than data set itself, thus avoiding delays in data delivery.

Bob Branton: Standard approaches are difficult; it is often easier/one is tempted to devise one's own solution.

Tim Deprez: No training on data management in the academic curriculum, no information given on data centres, data management procedures (cataloguing, archiving…). Scientists are computer literate, and working with ease with Excel… but not with RDBMSs like Access. Also, scientists don't know about 'big' data centres, but know about little ones.
Bernard Avril: Mathematics, statistics… are part of the academic curriculum, but not data management; it is perceived too much as an IT topic. Awareness of data centres: through libraries? Librarian is helpful as a paradigm to data manager.

Peter Pissierssens: Necessity to put data and information management on curriculum of academic institutions. By having students trained in Data and Information Management, they will realise the importance of data management. It will no longer take a stick to have them submit their data to data centres, but their awareness of the importance of data management, and the role data centres can play in this, will serve as a carrot.

Closing remarks

Savi Narayanan: Data management is more than an IT problem, there are concepts involved; there is a need to inform scientists about these concepts, and create awareness of their importance. Scientists who attended the panel discussion should take these ideas to their home institutions, and discuss possible ways to bring data and information management of the curriculum of their institution.

Lesley Rickards: Scientists should make sure to organise their own data; this would make the work of data centres much lighter, and avoid the danger of data loss. Also: most senior BODC staff have a background in science, and hold PhDs; this allows to establish a relation of trust with the scientists, a smoother collaboration, and a more creative input of the data managers in the development of data collection.

Peter Herman: Data management needs an equivalent to the web of Science: a mechanism to bring up a list of relevant, available, quality controlled data. A portal site has to provide access to data, preferably using a common interface.

Neville Smith: Data centres have to move away from their traditional role of data archiving centre, towards a data service centre. They can take an example from libraries, which provide expertise and guidance in cataloguing. There has to be peer-review and citation for datasets, as a way to measure and recognise progress, to recognise value and expertise, and as a foundation for standards and accepted procedures. Peer-review and an audit system will allow enforcing standards and orderly progress in what we do.

Catherine Maillard: A new type of data centre, more service-orientated, is emerging.There should be more efforts from the side of the data centres to ensure that data centres have enough visibility. It is difficult to avoid overlapping tasks.

Peter Pissierssens: Data managers have to find new ways of communicating with scientists; this symposium has been a very good opportunity to start this dialog. Both scientists and data managers are making data and information products; a close collaboration between the two groups will produce synergies, allowing the complementarity between the two skill sets to come into play.

Alan Edwards: Summarises the main points brought up during the discussion panel. Training: EU has money and the mandate to support ad improve training for scientists in data management. Library view of data centre: Archives are grey and dusty; libraries are active and open. Requirement for data management written into project condition for funding – is already the case for EU proposals, and happens for short-term data management. For long-term data management, often no funds are available anymore.

Savi Narayanan: Long-term data management is a challenge.

The colour of the ocean data is changing. There is a long tradition of standardized data management and exchange in the 'blue', physical sciences oceanography. These standards are only now emerging for geological, chemical and biological data. Moreover the shallow sea areas get more and more attention as highly productive biological areas that need to be seen in close association with the deep seas. How to fill in the gap between the deep 'blue' and the shallow 'green' biological data management was a major focus of the international symposium 'The Colour of Ocean Data'. This event took place at the Palais des Congrès (Brussels, Belgium) on November 25-27, 2002 and was jointly organized by the Flanders Marine Institute (VLIZ), the Belgian Science Policy, IOC/IODE and the Ocean Biogeographic Information System (OBIS/CoML).

ISSN 1377-0950

THE COLOUR OF OCEAN DATA
PROCEEDINGS 2004

IOC Workshop Report No. 188
VLIZ Special Publication No. 16