

Ocean Biodiversity Informatics – an emerging field of science

Edward Vanden Berghe¹ and Mark J. Costello²

¹Flanders Marine Institute (VLIZ), Wandelaarkaai 7, B-8400 Oostende (Belgium)
wardvdb@vliz.be

²University of Auckland; Leigh Marine Laboratory, Box 349, Warkworth (New Zealand)

Abstract

‘Ocean Biodiversity Informatics’ (OBI) heralds a new era in biological research and management that is revolutionising the way we approach marine biodiversity research. OBI uses computer technology to manage marine biodiversity information (capturing, storing, searching for, retrieving, visualising, mapping, modelling, analysing and publishing data). This allows more users better and faster access to biodiversity information than ever before. The global nature of phenomena such as climate change, over-fishing, and other changes in ecosystems, would not have been recognised had it not been for informatics-aided analyses.

The prospect of data mining and exploration on a global scale is enough to gladden the hearts of marine scientists across the world, as marine biology embraces the computer age. Access to global data through OBI will allow for worldwide gap analysis resulting in new perspectives on current research, the promotion of collaborations between research groups and real data sets for teaching purposes, to mention just a few of the potential benefits. OBI is an initiative of the 21st century and will make conventional marine biodiversity research more dynamic and comprehensive, with a range of constantly evolving online tools.

Background

We know very little about the biodiversity in the world’s oceans. But one thing is sure: the diversity of the type of data and information that is stored in data systems around the world is increasing dramatically. While well-managed databases with global coverage used to be restricted to geophysical sciences, this is no longer true. In two important meetings, the first in Hamburg in 1996, the second in Brussels in 2002, biologists have discussed how to take an example from the physical oceanographers, and to formulate plans on how to work together to integrate individual databases. In the workshop held in Hamburg in 1996, discussions were held on how to improve the quantity and quality of chemical and biological data available to the scientific community. The specific purpose of the workshop was to provide recommendations to guide management of chemical and biological oceanographic data by the Programme on International Oceanographic Data

and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission (IOC).

Biological data managers met with their physical oceanography colleagues during the 'Colour of Ocean Data' symposium held in Brussels, 25-27 November 2002. While it was realised there that the needs of biological data managers were different from those of physical oceanographers, it was stressed that commonalities are more important than differences. Some applications were presented that demonstrated the power of collaboration across disciplines.

Data on marine species and their environment (be it regional or on a global scale) is the 'fuel' on which OBI operates and therefore complements the traditional disciplines of taxonomy, ecology and biogeography. However, many scientists are ignoring pleas from international scientific organisations, including the International Council for Science (ICSU), to make their data public. The irony is that most data collections are paid for, directly or indirectly, by public funds. Taxonomists have led the way with regard to public accessibility of data, as type specimens are lodged in museums for the common good. Molecular sequence data is to be submitted to international repositories, where the data is publicly available, before research on these sequences is accepted for publication in scientific journals. It is suggested that there should be a protocol for scientists where ecological data would be made available in a similar way. OBI will make more data available to more people more quickly than ever before, including the repatriation of data and information collected in developing countries.

A change in biological science culture to one of open access to primary data is essential for accountability of research. A greater recognition of the value of biodiversity data accessibility by the scientific community, publishers, funding agencies and employers is vital. This change in culture is currently underway. Many marine biologists are actively gathering knowledge, as they have been doing for a long time. What is new is that many of these scientists are willing to share their knowledge, including basic data, with others over the Internet. The World Wide Web makes it easy and cheap to share data in a global community. It is compared to the invention of printing centuries before, and like it, will bring more information and opportunities for discovery to people all over the world. Our challenge now is to try and manage this trend, avoid confusing users with a multitude of contradicting sources of information, and make sure different data systems can be and are effectively integrated.

Meta-databases and other methods of data discovery will certainly gain more and more importance as the number of studies, and the number of scientists conducting these studies, increases. Such methods of data discovery, and better communication between and among scientists and data managers, are essential for avoiding unnecessary duplication of effort. Comprehensive inventories of collected data should facilitate projects of data archaeology and rescue, and make sure knowledge about the existence of data does not disappear.

One of the most challenging aspects of any informatics system is quality assurance. This is particularly important when you consider all the possible uses to which data can be put, and the number of steps from the original point of data collection to the end use.

However a data or information system is designed, its continuity and development depend on support from the scientific community. This community includes contributors, evaluators of funding applications, users and science policy makers. While not long ago it was difficult to find any information at all, now the Internet user is confronted with a large number of possible sources, but without an indication of the quality of the proffered information. Again, metadata can play a role in this; it should contain information on who collected the data, and for which purpose. The first can be an indication of quality, or lack thereof. The latter can be an indication of 'fitness for use', *i.e.* an evaluation of whether the data would be suitable for a particular analysis.

Integrating data from different sources brings its own set of problems. One very important aspect is that one has to make sure that the same terms are used in the different component datasets – for taxonomy, geography and measured parameters. Several initiatives exist to compile lists of taxonomic names – some for specific taxa, or for a restricted geographical area, some global in geographic and taxonomic scope. In order to facilitate data integration, we need a world register of marine organisms, containing all taxonomic names, including recent synonyms and often-used misspellings. Such a register, of which the content should be managed by the taxonomic community, would allow 'translating' between different sets of names now in use for different data sets.

Developments in technology have made possible new approaches to data sharing and dissemination. Distributed databases are becoming a reality, and the advantages of a distributed system now far outweigh the extra cost of technical complexities to create them. The Census of Marine Life (CoML), and its data management programme Ocean Biogeographic Information System (OBIS), is making this happen for marine biogeography. The Ocean Biogeographic Information System (www.iobis.org) is already publishing almost 10 million location records for 61,000 marine species from a global network of over 100 databases. OBIS is the information component of the Census of Marine Life, the largest ever global marine biology discovery programme (www.coml.org).

The main challenge is now to remain in control of developments, and not to let OBI become technology-driven. Obviously, recent technical developments should be monitored, and implemented when they bring a significant improvement, either in efficiency or in functionality. But what we do is much more important than how we do it. The real issues for data management are standardisation, collaboration and enabling knowledge-based decision-making.

OBI, the conference

The conference topics were restricted to marine biological data management – taxon-based, biogeography but also environmental, non-taxon based data management. Specific objectives were to

- Learn how and why researchers have used large-scale marine biodiversity databases to make major discoveries about the functioning and state of ocean ecosystems.

- Bring together biological data managers to discuss the present state, and progress, in this field since the meetings in Hamburg (1996) and Brussels (2002).
- Discuss standards and protocols for data exchange. Take note of new developments such as Distributed Generic Information Retrieval (DiGIR) and OBIS, and discuss how this will influence biological data management in general.
- Provide an opportunity for biological data managers to find out what is happening at IODE National Oceanographic Data Centres and marine research agencies from around the world.
- Discuss potential gaps and overlaps in the taxonomic and geographic scope of existing data systems. How can we, as a community, ensure that we are covering the whole field, and that no taxonomic groups are left behind? How can we make maximal use of resources, and avoid overlaps?
- How do we integrate data from separate databases into large datasets that will enable us to provide answers on the global cover and long time scales that we need?

The conference was organised as a series of five consecutive thematic sessions, with a final session dedicated to a panel discussion. Each of these thematic sessions had a corresponding poster session. In an opening session, several acknowledged experts were asked to review the field. Themes for the sessions were

- Opening session
- Information system development
- Taxon-based systems:
- Geography-based systems
- Analysis
- Panel discussion:
 - What is our target audience, and how effectively are we reaching it?
 - How do we integrate individual databases into datasets that allow large-scale, long-term analyses? What is the role of international organisations such as the International Council for the Exploration of the Sea (ICES), the IOC and the Food and Agriculture Organization of the United Nations (FAO) in this? What is the role of CoML and OBIS, and of GBIF? Which others have a role to play?
 - How do we avoid overlap and duplication of effort? How do we avoid gaps in taxonomic/geographical coverage?
 - What mechanisms are we using now, or do we plan to use in the future, to disseminate data? What should we do to persuade data providers to make their data available?

All sessions had ten oral presentations, except for the opening session with three scientific presentations, and the session on Analysis with 11 speakers – a total of 44 oral scientific presentations. In parallel, there was a permanent poster sessions, with a total of 56 poster presentations. A full list of titles of all presentations and posters is given on page 13.

During the opening session, participants were welcomed by Prof. Dr Peter Ehlers, President of the Federal Maritime and Hydrographic Agency (BSH), Hamburg/ Rostock. In his address, Dr Ehlers stressed the importance of integration of information over different disciplines; we have to optimise marine data management, not only biological data, but through a multi-disciplinary approach. Mrs H. Imhoff welcomed the participants on behalf of the German Federal Ministry of the Environment; it was thanks to the financial support of her Ministry, and her own personal interest in this matter, that we were able to use the nice conference facilities. Mrs Julie Gillin, Data Manager of ICES, and Mrs Lesley Rickards, Chair of the International Oceanographic Data and Information Exchange programme of the Intergovernmental Oceanographic Commission (IOC/IODE) addressed the meeting and stressed the importance of this conference for their respective organisations. Dr Mark J. Costello and Dr Edward Vanden Berghe reviewed the objectives of the meeting, and discussed expected outcomes.

The second part of the opening session consisted of three scientific talks, giving a broad overview of different aspects of OBI. Arthur Chapman gave a review of issues of quality control and the principles of data quality; he discussed the concept of 'Fitness for Use', meaning that data perfectly acceptable for one purpose might be useless for another. Donald Hobern of the GBIF Secretariat talked about new and existing technologies for data integration using distributed systems – the basis of emerging global systems for biodiversity information. The third review was from Dr Daniel Pauly, on fisheries impact on global marine biodiversity and ecosystems. In his talk, Dr Pauly illustrated how large, integrated datasets can bring insights that can't be learned from individual datasets.

The main point of discussion during the Panel Discussion was about the advantages of data sharing, and mechanisms to stimulate data custodians to make their data openly available. A draft text had been circulated from the beginning of the conference, and was edited during this final panel session. The resulting text was approved by all participants of the meeting, and is reproduced in this volume.

There were 168 participants in the conference, coming from 39 different countries; the complete list can be found on page 8. Not unexpectedly, best represented was Germany with 38 participants. This, together with the generous support from the German Government and the host institution, clearly demonstrate that OBI is taken serious in Germany. The US and UK were runner-up with 19 and 18 participants, respectively. Financial support from the EU made it possible to invite people from developing countries, and brought us participants from India, Mauritius, Seychelles, Tunisia and the Philippines. Extra support from the World Data Centre for Oceanography, Silver Spring, allowed us to support participants from South America.

Several papers, based on presentations or posters of the conference, were published as a Theme Section of the Marine Ecology Progress Series (Vol. 316, 2006). The table of contents of the Theme Section is given on page 18. Thanks to financial support from MarBEF, all papers in this Theme section are Open Access, and can be freely downloaded from the internet, either from the site of MEPS (<http://www.int-res.com/abstracts/meps/v316/>), or from the MarBEF web site (<http://www.marbef.org/modules.php?name=moa&lvl=Ref§ion=1&refid=100285>).

All conference presentations, publications and other documents remain available on the site of the Flanders Marine Institute (VLIZ)(<http://www.vliz.be/obi>).

Organisers

Local host: Bundesamt für Seeschifffahrt und Hydrographie (BSH, Germany)

Organising committee: Maren Fabricius (BSH), Edward Vanden Berghe (IOC/IODE and VLIZ), Peter Pissierssens (IOC/IODE), Mark J. Costello (OBIS), Friedrich Nast (BSH)

Scientific Committee: Mark J. Costello (OBIS), Fred Grassle (CoML), Syd Levitus (World Data Centre for Oceanography, Silver Springs), Peter Pissierssens (IOC/IODE), Tony Rees (OBIS), Lesley Rickards (British Oceanographic Data Centre, Chair IODE), Edward Vanden Berghe (VLIZ, Chair IODE GE-BICH), Sunhild Wilhelms.