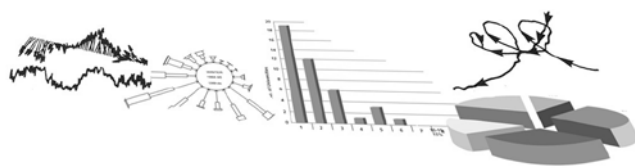


Marine Data Management



*We can do more,
but can we do better?*

Jan Seys
Peter Pissierssens
Ward Vanden Berghe
Jan Mees

The planning and implementation of research, and the efficient management of the resulting data often appear to be two widely separated worlds. Data managers consider the careful collection, management and dissemination of research data as essential for the effective use of research funds. Many researchers, on the other hand, consider data management as technical, boring and an (un)necessary evil; so data management is often insufficiently planned, or not planned for at all, and is assigned a low priority. This is unfortunate, as there is much of social relevance and applicability in the colourful world of oceanographic data management. Our objective is to guide you through some of the many initiatives related to marine data management and to present the main players. We focus mostly on physical and biological oceanographic data (Boxes 2 and 3 overleaf), less on hydrographic, chemical and geological data. We also discuss the new trends and developments that will determine the future of this field.

Tackling a growing problem

The social relevance of measurement and sampling at sea, and the need to disseminate the results as widely and in as user-friendly a manner as possible, cannot be overestimated. More services and products useful to industry, the general public and policy makers, could, and should, be extracted from databases. The oceans cover two-thirds of the Earth, and about half the world population live in coastal areas, so monitoring

the health, resources and 'tantrums' of the global ocean is no luxury. There are many applications of data-management that relate to climate and weather, safety at sea and along the coast, fisheries, offshore activities, management of the seas, etc. Let us focus on a few examples.

Meteorology and coastal defence

The weather has a tremendous impact on our lives. To a large extent, weather is 'produced' at sea, and the distribution of heat stored in the upper layers of the ocean is of great importance for both long-term and daily weather patterns. A good knowledge of meteorological conditions, and of how they are developing above the ocean, therefore makes a substantial contribution to timely prediction of storms and other unfavourable weather. Nowadays, through good measuring networks, and systems for making data available swiftly (in real time or near real time), it is possible to avoid a great deal of human suffering.

However, on a long-term basis it is also important to monitor sea-level changes: it is expected that by 2100 sea-level will have risen about 38–55 cm as a result of the greenhouse effect and the predicted rise of 1.5–6 °C in the Earth's temperature. To monitor this trend effectively, and to protect coastlines, we need more than a global network of sea-level stations. It is just as important to estimate the change in sea-level that will occur as a result of wind, atmospheric pressure patterns, rise and fall of land masses, and changes in ocean current patterns; and for these measure-

Box 1: Marine data management: a working definition

First, we need to distinguish 'data' from 'information'. 'Data' are observable, raw 'values' that result from research or monitoring activities; these values can be numerical (as in temperature or salinity measurements) or nominal (as in species lists for a particular region). The term 'information' is commonly used to mean data that have already been processed and/or interpreted results. In that sense, so-called 'metadata', i.e. data about data (e.g. by whom, at what time, where and how the results were collected) can be considered a special kind of 'information'.

Marine or oceanographic data management is the process of entry, quality control, archival and dissemination of research or monitoring data collected in the world's seas and oceans. We usually make a distinction between operational data (i.e. data continuously collected by networks of buoys and measuring platforms, transmitted by satellite or radio) and archival/service data (i.e. data collected by research teams and research projects, including experimental, field and modelling data). Data collected automatically on board research vessels are considered to be 'semi-operational data'.

ments to be of use, good data management, quality control and fast data availability are essential. In addition, experts expect that rapid warming of the climate will lead to shifts in – and increased intensity of – heat waves, droughts, floods, storms and other severe weather phenomena. Global warming also affects natural climate variability on time-scales of days to decades, by influencing atmospheric and ocean circulation (see below).

Predicting El Niño

El Niño is a large-scale periodic climatic anomaly, typified by a temporary warming of the surface waters of the eastern Pacific Ocean. Because the phenomenon was discovered by fishermen along the west coast of South America and nearly always started during the Christmas period, it was known as El Niño (the Boy Child). Important El Niño events occurred in 1982–83, 1986–87, 1991–92 and especially 1997–98. The name La Niña (the Girl Child) has been given to the cold phase that follows some El Niños (e.g. 1988–89).

It was initially thought that the effects of El Niño were limited to South American coastlines, where they dealt heavy blows to fisheries, but it was soon realized that their impact went much further afield; a strong El Niño is accompanied by heavy rainfall over the centre of the Pacific Ocean, the western part of South America and the southern part of the United States. Droughts then occur in Indonesia, Australia, southern Africa and north-eastern Brazil. All this is caused by fluctuations in the pressure difference between the Indonesian Low and the South Pacific High, known as the Southern Oscillation (SO). A marked decrease in the pressure difference causes the usually strong easterly winds over the tropical Pacific Ocean to weaken, leading to suppression of upwelling of cold, nutrient-rich water along the coast of South America.

El Niño events affect fisheries, agriculture, ecosystems and weather patterns (and thus human health and safety) far beyond the tropical Pacific. It is estimated that approximately 125 million people were affected by the 1997–98 El Niño event (the worst recent El Niño) and the material damage amounted to approximately US \$30 billion. Particularly destructive were the forest fires in Indonesia, the powerful cyclones that struck the west coast of Mexico, and the floods that destroyed harvests in East Africa.

The damage could have been much worse, had the El Niño event not been predicted six months in advance, thanks to the TOGA/TAO network of 72 measuring buoys that became operational in the tropical Pacific Ocean in the early 1990s. These buoys register meteorological and oceanographic data at the surface, and water temperature to a depth of 500m. Not surprisingly, the timely prediction of such a dramatic phenomenon created considerable 'El Niño hype': many extreme weather events were (wrongly) attributed to El Niño. The fact that prediction of El Niño was made possible by temperature measurements at depth (beyond the reach of satellite-borne sensors) and that the phenomenon was studied on a global level, led to the development and implementation of new and global measuring networks at

sea. Between 1990 and 2002, the World Climate Research Programme (WCRP) established the World Ocean Circulation Experiment (WOCE). This mega-project, with its 300 floating buoys and numerous basin-wide hydrographic sections, collected more temperature and salinity measurements over a period of eight years than had been collected during the previous 100 years.

The newest and most ambitious undertaking so far is Argo. This is a global network of (eventually) 3000 autonomous and freely floating profilers that should be operational by 2006. The project has been promoted by the Global Ocean Data Assimilation Experiment (GODAE), the Climate Variability and Predictability Project (CLIVAR) and the global observation systems GCOS (Global Climate Observing System) and GOOS (Global Ocean Observing System). Since the year 2000, Argo floats have been deployed in all the oceans at intervals of ~300 km. A float is submerged to a depth of about 2000 m, is transported by slow, deep currents for about 9 days, then slowly ascends to the surface, all the while measuring temperature and salinity. Having surfaced, the float transmits the collected data to a satellite, and another descent–ascent cycle begins. The average life-span of a float is estimated at four to five years.

Argo data are already available and, after a rigorous quality control phase of five months, will be freely accessible to all; see <http://argo.jcommops.org/>. This measuring network will not only provide more insight into the ENSO (El Niño–Southern Oscillation) system but will also greatly improve our knowledge of other climatic anomalies (such as those affecting the Arctic and the Antarctic, the Pacific Decadal Oscillation and the North Atlantic Oscillation).

Predictions needed for the safety of shipping

Tides, storms and currents are among the factors that determine the safety of shipping and other activities at sea. Predictions of these, by means of calculations with mathematical models and measurements made from satellites, buoys and other measuring platforms, have become commonplace in the context of shipping. But less ordinary events can also be explained through oceanographic databases. A good example is the occurrence of huge waves which appear from nowhere – accounts of which were often dismissed as 'sailors' tales'. However, it is becoming increasingly accepted that these 'rogue', 'freak' or 'extreme' waves are not only real but are the cause of many unexplained shipping disasters. The project MaxWave (within the EU Framework 5) brought together 11 European research teams to seek a scientific explanation of this phenomenon; see <http://w3g.gkss.de/projects/maxwave/>. The team also sought to identify the probability of such a 'wall of water' and to investigate how new ships can acquire better protection against such events.

Management of living and non-living resources
Management of living as well as non-living resources requires good knowledge and professional data management. Since the UN Conference on Environment and Development, held in

Box 2: The main players in the management of oceanographic data

- The Intergovernmental Oceanographic Commission (IOC) of UNESCO was founded in 1960 to promote oceanographic research and contribute towards the establishment of systematic ocean observation platforms (ships and, later, satellites), the necessary technological development and the transfer of knowledge (<http://ioc.unesco.org>). The Secretariat, as the executing organ of the IOC, is based in Paris and has been headed since 1998 by Dr Patricio Bernal, a Chilean oceanographer. The IOC has 129 Member States, which make up the IOC Assembly. Of these, 36 constitute the Executive Council. As far as data management is concerned, the IOC aims to ensure that oceanographic data and information – collected through research, monitoring and observation – are used efficiently and are distributed to the widest possible audience.
- This objective has been expressed through establishment of the International Oceanographic Data and Information Exchange (IODE) network that comprises more than 60 oceanographic data centres in as many countries. Most of these are National Oceanographic Data Centres (NODCs) – such as the Flanders Marine Institute, VLIZ – or Designated National Agencies (DNAs). Some of these were given special responsibility for specific regions or data types and are called Responsible National Oceanographic Data Centres (RNODCs). Together, the NODCs, DNAs and RNODCs supply their data to those World Data Centres (WDCs) dedicated to oceanography, which are based at Silver Spring (USA), Obninsk (Russian Federation) and Tianjin (China). The IODE network has also established a number of groups of experts and steering teams to provide advice to the IODE Committee or assist with the implementation of projects. They include the Group of Experts on Technical Aspects of Data Exchange (GETADE; now merged into the ETDMP of JCOMM – see below), the Group of Experts on Marine Information Management (GEMIM) and the Group of Experts on Biological and Chemical Data Management and Exchange Practices (GEBICH).
- The International Council for Science, originally the International Council for Scientific Unions (ICSU), was established in 1931 as a worldwide umbrella of scientific councils, academies and societies/institutions (<http://www.icsu.org>). ICSU maintains a network of 73 national and 27 international member groups and involves itself with everything related to science and society. It mobilizes funds and knowledge, publicizes research through meetings and publications, promotes constructive debates and participation of as many scientists as possible around the world, and facilitates interaction between disciplines and researchers of developed and developing countries. To this effect, ICSU coordinates and initiates important international and interdisciplinary programmes in close cooperation with other international organizations. Examples related to oceanography include cooperation with WMO and IOC in the World Climate Research Programme (WCRP) and with WMO, IOC and UNEP in GOOS and GCOS. ICSU was also instrumental in developing the system of 40 World Data Centres, established during the International Geophysical Year 1957–58 (two of which deal with oceanography; see under IOC above).
- The World Meteorological Organization (WMO) (<http://www.wmo.ch>) is a specialized UN agency, established in 1951. Its headquarters are in Geneva (Switzerland), and it was the successor to the International Meteorological Organization (IMO), whose origins date back to 1853. The WMO has 185 Member States, and is responsible for global cooperation in meteorological and hydrological observations and services (including systems for rapid data exchange, standardized observations and uniform publication of observations and statistics). The backbone of WMO is the WWW or 'World Weather Watch', a global data and information network of measuring stations, managed by Member States and using nine satellites, plus approximately 10 000 land-based, 7000 ship-based and 300 fixed and floating measuring buoys with automatic weather stations. The WMO plays a leading role in a number of international programmes and cooperation agreements related to climate change (such as the World Climate Programme which supports GCOS and the Intergovernmental Panel on Climate Change (IPCC); see below).
- Of the three operational observation systems, GOOS and GCOS, mentioned above, are important for operational oceanography. In addition, the climate module of GOOS is identical to the ocean component of GCOS, so the two systems can be considered Siamese twins. GOOS was established in the early 1990s under the co-sponsorship of the IOC, WMO, UNEP (UN Environment Programme) and ICSU. It was established in response to a clear need for a global measuring system, amplified by the call of the World Climate Conference, the IPCC in 1990, and the UNCED Conference in Rio in 1992. Since then a substantial number of regional sub-programmes of GOOS have been created (EuroGOOS, MedGOOS, Black Sea GOOS, GOOS Africa, etc.), and a number of existing initiatives have been absorbed by GOOS. GOOS includes two main groups of operations: (1) measuring systems in the open ocean, specifically to support services at sea, weather prediction, and monitoring of climate change; and (2) measuring systems in coastal areas, aimed at the study of the health and sustainable development of these areas. GOOS was initially established on the basis of existing observing systems, but it also developed its own pilot projects such as GODAE (which includes Argo – see p.21).
- The WMO/IOC Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) is a relatively new (1999) intergovernmental body of experts which provides the international coordination, regulation and management mechanism for an operational oceanographic and marine meteorological observation, data management and services system. To a large extent, it unites the common activities of the IOC and WMO and attempts to improve integration of expertise in oceanography and meteorology. Technical aspects of data management are discussed by the JCOMM/IODE Expert Team on Data Management Practices (ETDMP), taking over this role from the IODE GETADE (see above), but expanded to cater also for the needs of the WMO. This initiative also has close ties with the previously mentioned organizations.
- In the north-east Atlantic region, ICES (<http://www.ices.dk>) has been active since 1902 in the coordination and promotion of marine research, particularly in research related to living resources. Its special Working Group on Marine Data Management (WG-MDM) aims to optimize the data flow between individual research groups, and to develop a data management system that delivers products useful for fisheries policy advice and the management of living resources in the North Atlantic. In this regard, the ICES Oceanographic Data Bank acts as the data centre for OSPAR. The OSPAR Convention is an international agreement concerned with monitoring the environmental quality of all marine waters in the north-east Atlantic region (<http://www.ospar.org>).

Rio de Janeiro in 1992, monitoring of biodiversity has been considered necessary for assessing the health of ecosystems. Many new initiatives have been taken, especially in oceans and seas, to fill knowledge gaps with regard to living organisms.

Management of fishing resources should be based upon the data available. For the north-east Atlantic, ICES (see Box 2) plays a crucial role. Based upon the catch data collected by its member states, ICES annually advises the EU Commissioner for Fisheries about how much of each species can be caught in the coming year. Political negotiations within the European Fisheries Board lead to the so-called TACs (Total Allowable Catches) and the resulting quotas per species and per country.

Exploitation of non-living resources, such as sand, gravel, oil, gas and manganese nodules, is also well documented in databases, which in turn are of great value for managing future use of these resources.

Who is at the helm?

Without attempting an exhaustive review, in Boxes 2 and 3 we highlight a few of the main players in the domain of oceanographic data management, concentrating mainly on physical and biological aspects. A number of projects, organizations and cooperating teams focus on the collection and management of a wide range of operational data and/or other data streams; others specialize in particular areas or activities.

Data centres in evolution

Changes in technology and changes in society are both forcing data centres to rethink their role and modus operandi. Another trend is the increased interest in biodiversity and the need to set up management and monitoring programmes to study marine (and other) biodiversity. Human-induced world-wide changes, such as global warming, will no doubt affect our living resources; one of the challenges of the new data centres is to integrate biological and physicochemical data and make both data types available for combined analysis. These and other developments were discussed at the 'Colour of Ocean Data' Symposium, held in Brussels in 2002. The last part of the symposium was dedicated to a panel discussion, in which the changing role of data centres was discussed. What follows is a brief overview of the most important trends and issues that were identified.

There is a trend away from the traditional data centre, with its main task of archiving datasets, towards becoming more service-orientated. 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 and peer-reviewed datasets.

Any mechanism for finding data – i.e. 'data discovery' – is meaningless (and very frustrating!) if it is not linked to a system for data distribution, through which the scientist or interested lay-

Box 3: Biological marine data

Biological marine data management covers a special group of initiatives and cooperation agreements relating to biological-taxonomic databases. The increased attention given to biodiversity has necessitated the realization of easily accessible and complete species databases, because the concept of 'number of species' is the most practical and widely used measure of biodiversity. Some of the larger and better-known initiatives are OBIS (Ocean Biogeographic Information System: <http://www.iobis.org>), ITIS (Integrated Taxonomic Information System: <http://www.itis.usda.gov>) and Species-2000 (<http://www.sp2000.org>). In addition to catalogues of species, these databases may also contain synonyms, distribution data, and information about ecology, vulnerability, economic use, etc.

OBIS, together with the initiatives 'History of Marine Animal Populations' (HMAP) and the 'Future of Marine Animal Populations' (FMAP) and others, forms the backbone of the 'Census of Marine Life' (CoML) programme. ITIS focusses on the biota of North America and now contains ~ 320 000 species names, of which 186 000 are unique. Species-2000 is a species list with global scope that now has about 860 000 entries, of which ~ 300 000 are unique species. The European Register of Marine Species (ERMS) (<http://www.vliz.be/vmdcdata/erms>) began as a European MAST project and produced the first extensive list of marine species (~ 30 000) for Europe. In addition, it contains a database of 800 taxonomists in 37 countries, a bibliography of 600 identification guides, and an overview of the collections of marine species available in European museums and institutions. Also noteworthy are the many initiatives of the ETI (the Expert Centre for Taxonomic Identification, based in Amsterdam) which has earned respect through its management of biodiversity information (<http://www.eti.uva.nl>). Every year, ETI produces about ten CD-ROMs with highly quality-controlled taxonomic information, and for this task it can count on the help of no less than 1500 taxonomic experts worldwide. Since 1991, ETI has produced about 90 CD-ROMs of which about 25 are related to marine/coastal organisms.

person can access actual data. Setting up both data discovery and data distribution mechanisms is made possible by recent developments in internet and database technology.

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, etc. Datasets often result from particular projects, which usually have a limited time span. Short-term data management, within the time span of the project, is usually not a problem: scientists need data management to produce useful results; moreover, making provisions for data management is often a prerequisite for getting a proposal accepted in the first place. After the project ends, however, there is a danger that detailed knowledge about the collected data can disappear, together with project staff. It is the mandate of data centres to work together with project staff, to ensure that data are properly documented, and that the integrity of the data themselves are safeguarded after completion of the project.

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.

There is a need to create data and information products, not only for other data managers and scientists, but also for policy makers and society at large. These products will assist in increasing the visibility of data centres and demonstrate the usefulness of data management to a larger audience. As such, it may assist in attracting funds for further activities as well as data submissions from scientists.

Unfortunately, marine scientists are generally very poorly informed about data centres and about data and information management procedures. There is a need to investigate how to put data and information management on the curriculum of academic institutions. This would result in a greater awareness of data centres, and an increased quantity and quality of data submissions. Data managers should actively seek collaboration with scientists. Involvement of data managers in the planning of projects at a very early stage, and more input in the development of data collection makes 'Beginning-to-end data management' a reality.

There has to be peer review, as a way of measuring and recognizing progress, for recognizing value and expertise, and as a foundation for standards and accepted procedures. Standards and audit procedures are needed to allow objective peer review. Peer review, in turn, is a way of improving compliance with standards. Countries, or even institutions or scientists, could be tempted to work along principles that were developed locally; obviously, these fit local needs, and are usually much faster to develop, but having a variety of practices can lead to fragmentation and hamper data exchange. Developing standards is a task for the data centres.

Management of physicochemical and biological data has to be better coordinated. 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. The lower level of standardization in biology makes the importance of proper documentation of the datasets even greater. However, commonalities are more important than differences: both biological and physical data management need long-term archives; both need quality control and peer review; managers of both have a responsibility to create data and information products.

Marine data management is – or should be – essentially a global business. Participation of developing countries is essential if we want to build a complete picture of what is happening with our oceans and climate. Several problems hamper this participation. Internet access is difficult in many third-world countries, and where internet access is available, the bandwidth is often very limited, making it virtually impossible to download or upload large volumes of data. Also funds to purchase hard- and software, and expertise to maintain the systems, are factors that are 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 the 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 organized, making use of these tools and standards.

The future

Managers of marine data are facing major challenges. First, there is the incredible increase in the volume of data, especially in the area of remote sensing. Second, there is the great diversity in the types of data that have to be handled: physicochemical, geological, meteorological and biological data, all have to be integrated, and analyses and information products have to draw on all of them. Last but not least, there is a major discrepancy between the scale at which data are typically gathered, and the scale at which the data and information are needed. With very few exceptions, projects collect data and information on local scales, and over short time-spans. Humanity has brought on itself problems such as global warming and consequent sea-level rise, depletion of fish stocks, and pollution, which have generated a need for data and information on a global scale; integration of all available local datasets is the only way to create a data- and information base to support global decision-making.

Modern data management is inseparable from information technology. Recent developments in technology assist in coping with both the diversity and volume of data flows. The internet provides means to exchange data at no – or very low – cost. Electronic publishing is more and more the method of choice for communicating research results and other information. Database systems are becoming more sophisticated, allowing scientists and data managers to concentrate on subject matter rather than technical nitty-gritty. Computer systems are becoming faster, hard disk and other storage space is becoming cheaper, and information technology is making it possible to conduct data management, and devise information products, that could only be dreamed of just a couple of years ago.

The main challenge for data managers is now to remain in control of developments, and not to let marine data management become technology-driven. Obviously, recent technical developments should be monitored, and put to good use whenever and wherever relevant. But it is more important to continuously re-evaluate what the role of data centres should be, rather than how objectives are being realized. The real issues for data management are standardization, collaboration and enabling knowledge-based decision-making. Obviously, we can do 'more'. But can we also do 'better'?

Jan Seys is Information Officer at the Flanders Marine Institute (VLIZ); Jan Mees and Ward Vanden Berghe are respectively Director and Manager of the Flanders Marine Data and Information Centre (hosted by VLIZ).* Peter Pissierssens is Head of the Ocean Services Section at the Intergovernmental Oceanographic Commission of UNESCO.†

*Flanders Marine Institute (VLIZ), Vismijn Pakhuizen 45–52, B-8400 Oostende, Belgium. Email: info@vliz.be

† IOC/UNESCO, 1, rue Miollis, 75732 Paris Cedex 15, France, Email: p.pissierssens@unesco.org