I. Interoperable Data Discovery, Access, and Archive

Data Management and Communications Plan for Research and Operational Integrated Ocean Observing Systems

Part III. Appendices

Appendix 7. Biological Data Considerations

Contributed by Dr. Fred Grassle

March 2005

Ocean.US

The National Office for Integrated and Sustained Ocean Observations

Ocean.US Publication No. 6

Contents

Introduction

The Data Management and Communications (DMAC) Subsystem for ocean data from the Integrated and Sustained Ocean Observing System (IOOS) and NSF's Ocean Observatory Initiative (OOI) will accommodate marine biological data from a variety of sources and integrate these databases into a distributed system. Few studies of the marine environment provide information on accurately identified species placed in a context that is spatially and temporally resolved. The general availability of such data, along with modern navigational capabilities and real-time information on ocean processes at grid scales less than one kilometer, would bring about a revolution in management of marine resources by allowing the marine habitats on which species depend for survival to be physically and chemically defined. Present efforts to sample living resources on continental shelves or in deep-sea areas do not adequately consider habitat characteristics and related features of the underwater environment at the time of sampling, and as a consequence managers are unable to relate the abundance of natural resources to changes in the marine environment that alter the amount and quality of available habitats. Such changes in habitat are of serious concern to resource managers and the public because they can profoundly influence the availability of natural resources. Use of real-time data from IOOS will improve the information content and statistical reliability of biological sampling and reduce uncertainty associated with management decisions.

Data that can alert investigators to changes in the physical characteristics of the environment will allow targeted, adaptive sampling to measure specific physical-biological interactions. Understanding interactions among naturally occurring physical and biological factors is key to understanding how human actions may impact the animals and habitats of the oceans. Tools, such as acoustic and optical swath-mapping of the ocean bottom, provide essential information reflecting the distribution of marine organisms in relation to the physical and chemical factors that define their habitats. Each species inhabits a characteristic habitat that is defined by the environment, so that marine populations often respond sharply to ephemeral events such as climate oscillations, the passage of fronts, conditions favoring localized plankton blooms, and myriad other oceanographic phenomena at a variety of scales. Synoptic data on ocean conditions and associated plant and animal species is thus of vital importance to the managers, members of industry and the public whose livelihoods and well being depend on marine resources.

Ocean Biogeographic Information System

The Ocean Biogeographic Information System (OBIS—see http://iobis.org) is being developed to meet observing system needs. It provides international standards and protocols for accessing marine biological data, and can provide a biological subsystem for DMAC. OBIS offers a distributed Web presence for marine biological data in an up-to-date biological context. The many categories of marine biological data are typically maintained separately by appropriate authorities; many will be integrated into a distributed system accessible through a common portal to serve educators, scientists, and the general public. The OBIS portal is already providing a growing arsenal of tools for data analysis, synthesis, and visualization, as well as various access functions. Information from the heterogeneous assemblage of sources listed in the following section will be part of a distributed system accessed and integrated through functions of OBIS.

The OBIS system will allow anyone to click on a point or area of the ocean and obtain information on what lives there. Where does the blue-ringed octopus live? Where and how far does the Atlantic bluefin tuna travel in a year? What habitats have the most coral species, where are these places, and how are they changing? How do deep-sea animals find and distribute themselves with respect to the hydrothermal vents along the mid-ocean ridge? How do we compare the abundance of life from one place to another, and how do we map the myriad patterns of individual movement and behavior that enable each species to survive despite predators and competitors for resources? Answers to these questions are of vital importance to natural resource managers, and to many others who use marine information for work and recreation.

The basic information required is a species name (bioreference) and the latitudes, longitudes, and depth where it is found (georeference); time is also an essential dimension to understand variability within years or changes between years. If the information is from an ecological survey, information on abundance and sampling are necessary. The existing catalog of accessible, bioreferenced, and georeferenced information on life in the oceans is, at present, surprisingly small.

Marine Biological Data

SOURCES

- Government archives—In the United States, NODC and NASA maintain archives for ocean biological data. In Australia this activity is centralized in the newly formed Australian National Oceans Office, and in New Zealand the parallel organization is NIWA.
- Fisheries databases—Commercial fish species data are held for fisheries management areas within each country. International treaty organizations, such as the North Pacific Anadromous Fish Commission, the International Pacific Halibut Commission, and the Inter-American Tropical Tuna Commission have long time series of observations on catches of certain fish species shared among nations. The FAO has an international database that includes part of this information.
- Environmental protection agencies—For the United States, bottom assemblage data and habitat information are maintained by EPA and NOAA.
- Conservation organizations—Species information is used to define hot spots, endangered species, and harmful algal blooms.
- Museums—The major museums of the world have specimen-based databases on the species of the world. The value of these specimens is increasing as new methodologies for morphological and genetic analysis develop.
- Marine Laboratories – A number of marine laboratories, such as the Sir Alistair Hardy Foundation for Oceanographic Studies and the Scripps Institute of Oceanography, have geospatially referenced collections of plant and animal specimens, and related environmental data that span decades.
- Individual scientists—Individual taxonomic specialists and marine ecologists have extensive databases that have not been archived.
- Major oceanographic research programs—Programs such as JGOFS, GLOBEC, and RIDGE have important data sets that need to be accessible through an IOOS portal. OBIS is the data component of the Census of Marine Life (CoML) (http://coml.org) and its field programs including the Oceanic Pacific Pelagic (TOPP) program to track movements of large migratory species (e.g., tuna, swordfish, sharks, marine mammals, turtles, birds, etc.) using individual acoustic tags (http://www.toppcensus.org/). These data are being integrated with open-ocean IOOS data. Coastal listening arrays used to measure movements of individual salmon and other coastal migratory species will become an integral part of coastal observing systems (http://www.vanaqua.org/POST/).

Efforts are in progress to develop a standard classification of coastal habitats and this is one of the goals of the Ocean Biogeographic Information System.

SPECIES

The basic units for biological data are species. The names of these species and descriptions of them are the products of individual scientists whose careers have been devoted to describing and understanding evolutionary relationships among species. Increasingly, such individuals take advantage of DNA or RNA gene sequence data to differentiate among species and to trace their phylogeny. Each species is the unique product of its evolutionary history. Species are classified according to their evolutionary relationships using a well-established, internationally accepted hierarchical system of nomenclature. New species are continually being described (Figure 1 in Grassle, 2000); the hierarchical tree of evolutionary relationships among species and the associated nomenclature must continually be revised to incorporate new information. For this reason, biological data systems require more attention to metadata than do physical data systems.

The names of each species of plant or animal are the key words for information about organisms. Although ideally each species is known by a single, unique name, in practice, a species may be named more than once (creating synonyms) and the same name may be applied to more than one species (creating homonyms). Therefore, biological data systems require name translators that provide accurate scientific names from synonymous and homonymous names. In addition, translators are needed to relate common names to their scientific counterparts. With oversight from the Global Biodiversity Information Facility (GBIF), Catalogue of Life, and organizations such as the Integrated Taxonomic Information System (ITIS), Species 2000, and OBIS, the taxonomic authority for each major group of organisms maintains the accepted list of species.

Biological specimens are stored in museums and/or are maintained in culture collections, depending on the type of organism, to provide reference material for identification. Expert systems for identifications are being developed; identifications, be they based on morphology, on DNA or RNA sequence data, or on some other sort of distinctive feature, use accepted species names. As a minimum quality control and quality assurance measure, the taxonomic authority (the name of the person who originally described the species) and the name of the person identifying the specimen is typically included with each specimen record; such a practice is advisable for each data set based on such a specimen.

However, the units commonly used in biological oceanographic research are frequently not based on species, but instead are based on habitat, taxa above the species level (genus, family, order, etc.), size, chlorophyll biomass, optical or acoustic signatures, and trophic position in food chains or food webs. Such units are often used to quantify ecosystem function, and the precise identity of the component species is not considered germane to the question being studied. However, the relevance of species-level information to the inference of ecosystem function of species assemblages is an important area of research (Kinzig et al., 2001). The ambiguity or outright lack of information on the number and characteristics of species represented may be the product of a particular method of sampling (net tows, optical plankton recorder, flow cytometer, satellite data), which is incapable of making taxonomic determinations. There is also a shortage of expertise for resolving identifications to the species level.

The most accurate species-level information on organisms in the marine environment may, in general, reside in notebooks or computers of individual scientists, or in publications based on these sources. For terrestrial data, agencies and organizations responsible for environmental protection, land use policy, natural resource management, and protected areas, establish basic requirements for archiving the data and provide access to them. By contrast, for the most part, research activities in the marine environment do not have a system of governance to assign responsibilities for accurate environmental information, so marine biological data are insufficiently accessible and are accompanied by too little metadata to meet most user needs.

RELATING DATA ON PRESENCE/ABSENCE, ABUNDANCE, AND BIOMASS TO ENVIRONMENTAL PARAMETERS

Museum accessions usually record the existence of a particular species at a particular location (ideally, but not always, latitude, longitude, and depth). Data on age and/or size of individuals, and life-history stage are often included. Museum records are not designed to provide information about the overall species composition of a locality, since only the species found on a particular field trip or in a particular survey are recorded. However, if the data are arguably the result of an exhaustive comprehensive survey using a consistent sampling design, the possibility exists of inferring the probable presence of common species that may have been absent in a given collection. Thus the interpretation of counts of number of individuals, amounts of biomass, and presence/absence data require description of geospatial sampling patterns, effort expended and type of sampling gear as part of the metadata.

References

Grassle, J.F. 2000. The Ocean Biogeographic Information System (OBIS): an on-line, worldwide atlas for accessing, modeling and mapping marine biological data in a multidimensional context. Oceanography, 13(3), 5–7.

Kinzig, A. P., S.W. Pacala, and D. Tilman (eds.). 2001. The Functional Consequences of Biodiversity, Empirical Progress and Theoretical Extensions. Monographs in Population Biology, 33, Princeton University Press, Princeton, New Jersey. 365 pp.