Organism Occurrences in an Ocean Observing System

Daphne G. Fautin
Department of Ecology and Evolutionary Biology, and Natural History Museum and Biodiversity Research Center, University of Kansas, 1200 Sunnyside Drive, Lawrence, KS 66045

Peter Fippinger
Consortium for Oceanographic Research and Education,
1201 New York Avenue, NW, Suite 420, Washington, DC 20005

Abstract - **Integrating data about marine organisms and associated observational information into ocean observing systems is an immense challenge. Identification of few kinds of organisms is possible with techniques used to obtain most physical data. Acoustic sensors and methods have been developed for some species that generate sound, and optic sensing techniques may be possible for others, but even those methods typically cannot distinguish among most species. Genetic sequencing shows promise for species-level identification in the future, but significant hurdles remain identifying appropriate markers, building a library of known sequences, and creating widespread capacity. For now and into the foreseeable future, the occurrence of most types of marine organisms will have to be documented from in-situ and non-real-time sources. One source of such data that is poised to interact with other data types from ocean observing systems is the Ocean Biogeographic Information System (OBIS). OBIS is a distributed, web-based, provider of geo-referenced information on marine organisms world-wide that serves information from museum catalogs, fisheries surveys, the published record, and the like. These data can also provide a historical dimension, thereby contributing to a baseline extending back a century and more; this is highly desirable when environmental change is occurring on a scale of decades. Particularly in an historical context, even point records can be informative. OBIS, and similar systems, must overcome challenges to provide useful products for and from observing systems. Chief among these is interoperability of data management systems and data types.**

I. INTRODUCTION

Providing biological data and products based on biological data is prominent among the rationales for a Global Ocean Observing System (GOOS): the goals of GOOS, according to its website [1], are to "provide information about the present and future states of seas and oceans and their living resources, and on the role of the oceans in climate change." However, most tools and technologies currently used for collection and delivery of real-time data to observing systems are designed to measure physical parameters of surface and near-surface oceanic waters. The GOOS website acknowledges "At this time, apart from the GCRMN [Global Coral Reef Monitoring Network], these measuring systems are concerned primarily with physical observations. However, consideration is now being given to what chemical and biological information is required and how to integrate it with physical data. Living marine resources exist mostly in the coastal zone, but the monitoring requirements for living resources and coastal seas remain under development. The challenge is to develop a high quality, integrated approach to coastal monitoring and forecasting, taking into consideration the needs of resource managers."

We present an overview of some of the challenges in dealing with data about marine organisms and associated observational information, some efforts being made to gather and make available such data, some reasons that integrating such data into an ocean observing system is vital, and some challenges in this enterprise. We focus on approaches being taken by the Ocean Biogeographic Information System (OBIS) [2, 3], one component of the Census of Marine Life (CoML) [4]. CoML aspires "to assess and explain the diversity, distribution, and abundance of marine life in the oceans -- past, present, and future" [4]. Scientists from more than 70 countries are currently engaged in 20 projects that: 1) build historical time series datasets through data mining, 2) undertake field research on biodiversity of marine organisms and explore novel methods of collecting data, and 3) establish predictive methods. Data from all these efforts are to become part of OBIS, which is to be the legacy of CoML, providing a long-term, scalable tool for assessing changes in abundance, distribution, and diversity of marine life through time.

II. CHALLENGES

Among the reasons that organism occurrences do not lend themselves easily to remote and in-situ detection methods in real time that are commonly used for physical parameters are the following.

1) Size: Organisms are tiny relative to features currently detected by remote and in-situ methods. For example, such biologically important organisms as pelagic crustaceans are commonly on the order of millimeters in body length.

2) Heterogeneity: Tens or hundreds of species may occur in a relatively small area of the marine environment, some even living inside others (e.g. intracellular dinoflagellates in reef-forming corals: [5]). This attribute is unlike those such as sea surface temperature (SST), which are relatively uniform over considerable distances in the open ocean. Further, for the very few organisms that can currently be detected by remote means, different (but often related) species may create signals that are indistinguishable. This means that ground-truthing cannot operate as it does for physical parameters such as temperature. The lack of global generalizability for ground-truthing is made clear on a website concerned with remote sensing of coccolithophores [6]: captions of SeaWIFS satellite images from the North

Atlantic are identified as showing blooms of *Emiliania huxleyi*, but those from the South Atlantic could be identified only as "a probable (no ships have ever taken water samples to confirm them there) coccolithophore bloom."

3) Habitat: Animal life extends to the greatest depths of the sea [e.g. 7] but most sensors now in use are restricted to near-surface waters. Still, organismal diversity is greatest in relatively shallow waters (as is density of most species), and most organisms of direct human importance live there. This presents a challenge of another sort – most shallow waters are near shore, where sensors are least effective due to steep gradients and small-scale variability in features such as salinity, water clarity, and depth, and because of political and logistical concerns. These waters are also nearest human populations, and are where most commercial activities occur, and so present special challenges to the GOOS [1].

4) Identification: Taxonomic identifications, commonly at the species level, are needed for many scientific purposes (e.g. understanding ecological interactions and migration routes) and policy matters (e.g. setting fishing limits, identifying invasive species). Many, many species of marine organisms have not yet been described (a process that consists of assigning a scientific name to a unique combination of attributes, conventionally morphological ones but increasingly others such as behavior and genetics). Even for those known to science, remote identification is impossible for all but a very few species that are atypical in many ways, and, for them, precision is not great. The coccolithophore website [6] acknowledges this issue for organisms that photosynthesize: "Satellites can also detect the presence of chlorophyll in the water, but as a rule cannot distinguish which species are responsible…" Guidebooks, which allow identification from images or include keys, are available mainly for species of commercial importance and some of esthetic value, or for particular habitats or small regions of the world (such as Indo-Pacific coral reefs [8, 9, 10] or the coast of central California [11]). Such resources that are increasingly available on the Internet operate in a similar manner. Identification of most marine organisms to the species level (as well as properly describing those new to science) still require the expertise of a taxonomist or parataxonomist [12], and takes an increasingly long time, as taxonomic training declines [e.g. 13].

III. SOME SOLUTIONS

Among technological innovations being developed to resolve organism identifications remotely (without a specimen in hand, and certainly without having to send it to a taxonomist) are acoustic techniques. Some vertebrates, such as whales and fish, make sounds that are thought to be species-specific and that can automatically be matched to an index standard [12]. Sounds of most of the other marine organisms that create auditory signals – for example, the vibrations made when tiny brine shrimp swim are involved in the discharge of the stinging capsules of sea anemones that prey on them [14] -- are of such short range that they would not be useful in observing systems such as GOOS, but large aggregations of pelagic organisms may be detected [12]. However, most marine organisms do not create sound, and echoing technology may not be useful for benthic organisms.

By contrast, all organisms possess nucleic acids. Because the attributes of a species are, by definition, unique in some way, this genetic material, which encodes organism attributes, differs among species. Thus, techniques that can detect these differences should be able to provide species identifications [e.g. 15]. Schander and Willassen [13] review some of the conceptual and operational hurdles to making this potential a reality, but development of a standard library of sequences has begun [15]. Currently, actual samples of tissue from the organism to be identified are required to obtain a sequence, which is then matched to an index standard in the library. Given that many organisms may occupy small areas and volumes (above), and living animals do not emit these compounds into the environment, it is unlikely this technique will have widespread application remotely.

Two field projects of CoML use organisms to carry transmitters that allow their movements to be monitored and, in some cases, also gather data about the animal's immediate habitat and physiology [12]. "The POST [Pacific Ocean Shelf Tracking] project uses newly developed acoustic technology to track the movement of individual animals. Tags implanted in the animals' abdomens send out unique signals, which are picked up by receivers on the ocean floor. The migration path of tagged animals can be reconstructed from the data gathered. Most species and many life history stages can be studied using POST" [16]. The project "TOPP [Tagging of Pacific Pelagics] tests the effectiveness of using animals to gather biological and environmental data, tracking the individuals' movements while recording oceanographic data from their immediate surroundings. According to the TOPP website [17], "TOPP tags have returned massive quantities of data, helping scientists build a rich picture of key travel corridors and "ocean hot spots," or gathering zones where animals feed and breed" [17]. In both cases, the technology is not for making identifications; rather, individuals are tagged because they are known to belong to a particular species. As data accumulate on animal movements, we may learn that (tagged) individuals of certain species or ages tend to go to places where (untagged) individuals of other species go to feed, and so we will be able to infer the presence of other species, or identify areas of high biomass or biodiversity, for example. This approach will provide novel types of data into observing systems and perhaps have unique applications, but only for a tiny minority of marine species (albeit some of economic importance and public concern). The approach does not, however, directly allow monitoring of a particular location to assess which species are present, when, and at what density (unless all individuals were to be tagged).

A recent article concerning OBIS [3] was entitled "Where is what, and what is where?" because the Ocean Biogeographic Information System is designed for just such inventories. OBIS is a distributed system of taxonomically and geospatially resolved data for marine organisms. Collectively, the more than 6 million records from 50 datasets contributed by more than 20 institutions or individuals constitute an Internet-accessible atlas that allows a user to find all occurrence records of a particular species, or to find all species that occur in a particular place. In addition to data from CoML field projects, OBIS serves data drawn from natural history museums, publications (on ecology, biogeography, and taxonomy, among other subjects), fisheries surveys and landings, and the like.

IV. ABOUT OBIS

The OBIS portal assembles these data, but they remain in the custody and under the control of the compiler or owner. Currently the data are communicated directly from the provider to the OBIS portal (at Rutgers University, USA), but a system of 10 Regional OBIS Nodes is being implemented, each responsible for assembling data from a geographical region and passing them on to the portal; data will be served from the nodes and the portal. As described on the OBIS portal homepage [2], software known as DiGIR [18] is the communications medium that "allows the portal to send that query to the data contributors, for the data contributors to translate that query into a search on their local database, and to send the data back to OBIS." DiGIR was developed initially to communicate specimen data from natural history museums to the Global Biodiversity Information Facility (GBIF) [19]. Providers of data to OBIS (and to GBIF) can store their data in any application and label fields as appropriate for their purposes; DiGIR transports data that have been mapped to a common schema. The OBIS Schema is an extension of the Darwin Core version 2, which is the GBIF standard. As stated on the OBIS portal [2], "When the OBIS portal sends queries out to its distributed data contributors, the portal will request data using these fields and needs to have data returned using these fields. The DiGIR software provides the programming to turn an OBIS query into a search on your particular database, but in order to install DiGIR you need to "map" the OBIS schema fields to the fields in your database."

Using these same conventions, data served by OBIS are communicated to GBIF; as its main source of biodiversity data for the marine environment, OBIS is the second largest provider of data to GBIF. Through all of these passages, however, the ultimate provider of the data is credited as being the source of the data. OBIS and GBIF serve largely to allow users to find data and to view data from multiple sources simultaneously.

In addition to "one-stop shopping" for information on distribution of a wide variety of marine taxa, OBIS extends into the fourth dimension, time. An historical perspective is essential to achieve the objectives of GOOS. Knowing what has gone before is vital to its objective forecasting, and given that change has already been under way for some time and change is occurring so rapidly, data assembled since the advent of electronic and remote sensing will provide a very incomplete and truncated baseline from which to make decisions. As a recent report from the US National Academy of Sciences [20, page 73] phrased it: "Geohistorical data are essential for answering many kinds of questions, especially when the aims are either to discriminate between anthropogenic and non-anthropogenic effects or to understand phenomena that cycle or emerge over periods greater than a few years. Such discrimination is extremely difficult without recourse to historic records in the broadest sense … Therefore, a rigorous strategy for evaluating ecological dynamics using geohistorical records – and for integrating geological and biological methods and insights – is essential." In identifying several sources of physical data from the past that should be incorporated into the ocean observing system, the GOOS data management report section 3.2.11 on data archeology [21] recognizes that "Much needs to be done to locate and rescue the balance of the physical data and the biological and chemical data." OBIS is a source of one category of such biological data, those on organism occurrences, which are largely the published record and museum specimens. Sparse, commonly anecdotal, and of undocumented quality, these data lack some features now considered essential. However, they are what exist, and should be used, judiciously, to extend knowledge into the past. Even single occurrence records can be valuable to science as well as in policy decisions by documenting, for example, when conditions favorable to survival of a particular species existed at a certain place, or that a species is not a recent invader of a particular place.

The overwhelming majority of marine species are not of direct economic importance, and so data on them have not been assembled systematically like those related to commercial fisheries [e.g. 22, 23]. Many of the databases served through OBIS deal with such organisms, having been assembled as part of scholarly projects. With increased interest by resource managers in ecosystem-based management, assembling and managing data on the components of ecosystems – primarily a diversity of non-commercial species - will be essential to understanding function and monitoring change. Chapter 19 of the Ocean Commission Report strongly advocates this approach, which, it states [24, page 295] "will provide direct benefits to the ecosystem and create a better mechanism for addressing apparent conflicts between socioeconomic and biological goals."

V. INTEGRATING AND MAINTAINING DISPARATE TYPES OF DATA

Interoperability among data management systems and across data types will be essential to producing useful products and tools. This is recognized among those developing GOOS: with regard to the coastal environment, the GOOS website [1] states "The challenge is to develop a high quality, integrated approach…" It is currently impossible to collect most organism occurrence data in real time; human intervention will be essential into the foreseeable future even for the most visionary technologies being developed. Therefore, a significant challenge remains to integrate the various types of data being gathered in real time by the ocean observing system with what have been and are being assembled by projects such as those of CoML. A simple approach that could suffice for many applications is overlain images. For example, many OBIS data are currently or could be temporally resolved so that maps could be made showing which species occur in various places at various times during a year, or at various stage of the El Niño Southern Oscillation (ENSO) cycle. A more sophisticated approach would be integrating real-time or near-real-time data on such ocean parameters as SST, using filters and thresholds like those that have been designed to assess the bleaching threat on coral reefs [25], coupled with site-specific surveys that could assess if bleaching-prone species occur where threats exist.

The more biological data accessible through OBIS or similar data management systems, the greater the possibility for detecting emergent patterns, the greater the functional utility of observing systems like GOOS, and the clearer the need for a system such as OBIS. The technical issues are immense. Such an ambitious system requires not only in equipment and instrumentation to gather and process these unique data, but ways to find other sorts of relevant data and data management systems to integrate the two. Archiving the large quantity of disparate types of data is essential so patterns can be detected through time, and across space and disciplines. To provide human benefit over the long term also requires investment in building sustainable long-term management to support the maintenance of these unique data – human, logistical, and financial.

Acknowledgments

We are grateful for support from the Sloan Foundation. DGF acknowledges NSF grants DEB 99-78106 and OCE 00-03970.

REFERENCES

[1] What is GOOS?: http://ioc.unesco.org/goos/docs/whatis01.htm -- accessed 26 June 2005.

[2] OBIS: Ocean Biogeographic Information System. http://www.iobis.org – accessed 26 June 2005.

[3] M.J. Costello, J.F. Grassle, Y. Zhang, K. Stocks, and E. Vanden Berghe, “Where is what, and what is where?,” *MarBEF Newsletter*, vol. spring, pp. 20-22, 2005 [at http://www.marbef.org/documents/newsletter/NwsNo2_Feb05.pdf]

[4] CoML: Census of Marine Life. http://www.coml.org – accessed 26 June 2005.

[5] R.W. Buddemeier and D.G. Fautin, “Coral bleaching as an adaptive mechanism: a testable hypothesis,” *BioScience* vol. 43, pp. 320-326, 1993.

[6] Images of *Emiliania huxleyi* blooms. http://www.soes.soton.ac.uk/staff/tt/eh/satbloompics.html – accessed 26 June 2005.

[7] A.F. Bruun, Sv. Greve, H. Mielche, and R. Spärck, *The Galathea Deep Sea Expedition 1950-1952*. London: George Allen and Unwin Ltd, 1956, 296 pp.

[8] G.R. Allen and R. Steene, *Indo-Pacific Coral Reef Field Guide*. Singapore: Tropical Reef Research, 1994, 378 pp.

[9] P.L. Colin and C. Arneson, *Tropical Pacific Invertebrates*. Beverly Hills: Coral Reef Press, 1995, 296 pp.

[10] T.M. Gosliner, D.W. Behrens, and G.C. Williams, *Coral Reef Animals of the Indo-Pacific: Animal Life from Africa to Hawai'i Exclusive of the Vertebrates*. Monterey: Sea Challengers, 1996, 314 pp.

[11] R.I. Smith and J.T. Carlton, eds., *Light's Manual: Intertidal Invertebrates of the Central California Coast*. Berkeley and other cities: University of California Press, 1975, 716 pp.

[12] K.L. Daly, RH. Byrne, A.G. Dickson, S.M. Gallager, M.J. Perry, and M.K. Tivey, “Chemical and biological sensors for time-series research: current status and new directions,” *Marine Technology Society Journal*, vol. summer, pp. 121-143, 2004.

[13] C. Schander and E. Willassen, “What can biological barcoding do for marine biology?,” *Marine Biology Research*, volume 1, pp. 79-83, 2005.

[14] G.L. Watson and D.A. Hessinger, “Cnidocyte mechanoreceptors are tuned to the movments of swimming prey by chemorecrptors,” *Science*, vol. 243, pp. 1589-1591, 24 March 1989.

[15] Barcode of Life: http://www.barcodinglife.org – accessed 26 June 2005.

[16] The Pacific Ocean Shelf Tagging Project: http://www.postcoml.org – accessed 26 June 2005.

[17] Tagging of Pacific Pelagics: http://www.toppcensus.org/web/Background/Overview.aspx -- accessed 26 June 2005.

[18] Distributed Generic Information Retrieval: http://digir.sourceforge.net/-- accessed 26 June 2005.

[19] Global Biodiversity Information Facility: www.gbif.org/ -- accessed 26 June 2005.

[20] Committee on the Geological Record of Biosphere Dynamics, *The Geological Records of Ecological Dynamics: Understanding the Biotic Effects of Future Environmental Change*. Washington: National Academies Press, 2005, 200 pp.

[21] The Global Ocean Observing System Data and Information Management Strategy and Plan: http://ioc.unesco.org/goos/GOOSdm_final.pdf -- accessed 26 June 2005.

[22] California Cooperative Ocean Fisheries Investigations: http://www.calcofi.org/

[23] Plankton subset of World Ocean Database 2001: http://www.nodc.noaa.gov/OC5/WOA01/plankton_intro.html

[24] U.S. Commission on Ocean Policy, *An Ocean Blueprint for the 21st Century*. Washington: Government Printing Office, 2004, 522 pp. [at http://www.oceancommission.gov/documents/full_color_rpt]

[25] Tropical ocean coral bleaching indices: http://www.osdpd.noaa.gov/PSB/EPS/CB_indices/coral_bleaching_indices.html -- accessed 26 June 2005.