

Planning for OBIS: examining relationships with existing national and international biodiversity information systems

Mark Fornwall

USGS Center for Biological Informatics • Denver, Colorado USA

Introduction

The Ocean Biogeographic Information System (OBIS) grew from a series of workshops funded by the Sloan Foundation to explore the feasibility and goals of a Census for Marine Life. During the workshop that focused on the benthic environment, an online digital atlas of species distributions and associated marine habitats was envisioned (Ausubel, 1999). This atlas could be used to "guide sampling designs for a 'Census of Marine Life' and generate hypotheses concerning the origin and maintenance of diversity of life in the oceans" (Grassle and Stocks, 1999). An information system to provide centralized access to existing datasets and to incorporate new data created by the Census of Marine Life was also proposed. This Ocean Biogeographic Information System would provide Web access to global data and information on marine species, such as occurrence, abundance, distribution, and related physical and chemical information.

Because of the scope and magnitude of this project, identification of key organizations in the international biodiversity information community is an important first step, both to understand pertinent interrelationships among those organizations and to engage them in planning and implementation of the new system. This paper discusses some of the current international biodiversity information initiatives and suggests an approach for implementing OBIS within this community.

Because of the large number of organizations and their acronyms included in this discussion, a glossary containing a brief description of each has been provided.

The International Biodiversity Information Community

An ever-increasing interest in biodiversity over the

past decade is evident from the variety of governmental and intergovernmental initiatives instituted during that period, including the United Nations Convention on Biodiversity (1992), the Summit for the Americas (1996), the Organization for Economic Cooperation and Development recommendation to develop a Global Biodiversity Information Facility (1999), and numerous

national initiatives. Establishment of wide-ranging scientific studies that involve the public in biodiversity issues, such as the All Taxa Biodiversity Inventory of the Great Smoky Mountains (Discover Life, 1998) and the International Biodiversity Observation Year (Mooney et al., 2000), demonstrate the

increasing prominence of biodiversity issues in the scientific and public arenas.

Mirroring the complexity and diversity of the species they seek to understand, the biodiversity information community is a sociologically diverse mix of government agencies and individuals with divergent interests and different backgrounds (academic, industry, government) who have varied views or requirements (PCAST, 1998). A general understanding of this community and its inherent complexities may be gained by examining how biodiversity information is being managed in the United States and looking at the emerging relationships of U.S. information initiatives with information initiatives around the globe. The perspectives gained can be used as a starting point for planning and implementing OBIS.

Global Coordination of U.S. Biodiversity Data Collection and Dissemination

Biodiversity data are collected by a variety of U.S. organizations for two general purposes, scientific under-

*Mirroring the complexity
and diversity of the species
they seek to understand,
the biodiversity information
community is a
sociologically diverse mix.*

standing and natural resource management. Scientific inquiry is undertaken through studies by educational institutions, museums, zoos, and other groups. Funding is acquired through private organizations or the government. Resultant information is made available through the peer review process and constitutes much of our foundational understanding of biodiversity.

Personal preference, organizational policies, availability of funds, and commercialization potential help determine which data are made available. How long such information is accessible is the result of a similar set of criteria that also includes longevity of the storage medium and the technology used to store datasets. Data may or may not "disappear" as funding or individual researcher interest wanes; the format of the information may become obsolete as technologies change.

Biodiversity data and information collected under the auspices of various governmental bodies are generally targeted at local issues such as those associated with land-use planning or invasive and endangered species regulations. The scope and scale of projects vary, and involved researchers frequently represent both Federal and non-Federal sectors. Although data collection activities at regional and national levels may have little or no national coordination or integration (PCAST, 1998), mutual benefits would be gained by leveraging investments to produce national datasets that could be used by scientists, land managers, and the public.

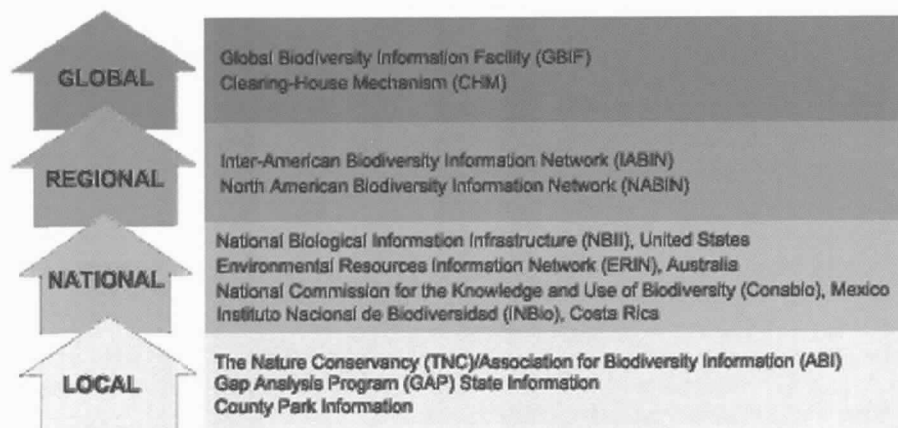
In the United States, the National Biological Information Infrastructure (NBII) strives to do just that. Evolving from a National Research Council (1993) recommendation, the NBII is a partnership of interested organizations that collectively seek to facilitate swift user access to biological databases, information products, directories, and guides maintained by Federal, State, and local government agencies, non-government institutions, and private sector organizations, both in the United States and around the world (Cotter et al., 1996). To accomplish this goal the NBII has assumed a traditional governmental role in fostering the development of an infrastructure upon which the biodiversity community can build. This infrastructure includes an organizational framework for biodiversity information, adoption of information and technology standards, and national and global partnerships to address common issues and to share access to information and data sources. The NBII is also developing a research agenda, in collaboration with a variety of information partners. Figure 1 depicts the emerging schema for coordination of global biodiversity information, with NBII identified as the primary U.S. participant.

On a regional level, NBII partnerships have been formed among neighboring nations who share common species or biodiversity-related concerns, such as the Inter-American Biodiversity Information Network. In December 1996, leaders of the governments of the Americas, meeting at the Santa Cruz (Bolivia) Summit on Sustainable Development, agreed to "seek to establish an Inter-American Biodiversity Information Network, primarily through the Internet, that will promote compatible means of collection, communication, and exchange of information relevant to decision-making and education on biodiversity conservation as appropriate...that builds upon such initiatives such as the Clearing-House Mechanism provided for in the Convention on Biological Diversity, and the Man and the Biosphere Network and the Biodiversity Conservation Information System, an initiative of nine IUCN [World Conservation Union] programs and partner organizations" (Summit of the Americas, 1996). Initial actions have focused on governance issues, coordination with existing data stewards, and adoption of standards for biodiversity data and information.

At the global level are two complementary activities the Clearing House Mechanism (CHM) and the Global Biodiversity Information Facility (GBIF). Established under the Convention on Biodiversity, the purpose of the CHM is to promote scientific and technical cooperation related to biodiversity conservation and sustainable use. The likely result of this initiative will be facilitation of regional biodiversity activities and sharing of meta information about biodiversity information and sites.

The CHM's work will be complemented by the Organization for Economic Cooperation and Development's (OECD) Global Biodiversity Information Facility. The GBIF is envisioned as a distributed facility that is global in scale, implemented regionally and nationally, and open to all countries. It will provide a link to global biodiversity information and will include tools to locate, evaluate, analyze and

Figure 1



Biodiversity Information Infrastructure: Geopolitical Levels

combine information. (OECD, 1999).

Another global biodiversity information initiative that bears mentioning is DIVERSITAS. Created in 1991 by six international scientific organizations—the International Union of Biological Sciences; the Scientific Committee on Problems of the Environment; the United Nations Educational, Scientific and Cultural Organization; the International Union of Microbiological Societies; the International Council for Science and the International Geosphere-Biosphere program—DIVERSITAS is an international program for the study of biodiversity. The basic concept of DIVERSITAS is to promote and integrate biodiversity research to ensure that all areas are receiving attention and to reduce duplication of effort. The program provides syntheses and discussion opportunities for the scientific community as well as for managers and policymakers as they move toward meeting the requirements of the Convention on Biological Diversity (Mooney et al., 1999).

Within this emerging global biodiversity information infrastructure, then, is a model wherein each nation works within its borders to facilitate cooperation between scientific and governmental biodiversity enterprises to promote the establishment of an information infrastructure (in the broadest sense of the word, hardware, people, and policies) and to coordinate these efforts at the regional and global scales. Concurrent with these governmental initiatives are initiatives originating from the scientific sector; these initiatives may subsequently intersect and combine. The question now becomes, “How does OBIS work within this global organization to achieve the best results for its stated goals?”

The OBIS Role in the International Arena

As discussed above, a considerable number of biodiversity information sources and initiatives are currently underway or are proposed, and a great volume of biodiversity data sources (including physical and chemical data) exists. Connections—both literal and organizational—between each relevant biodiversity information system or information source and OBIS should be pursued.

The OBIS role within the global biodiversity information community may include four areas: facilitating global partnerships, providing an organizational framework for ocean biodiversity information, fostering adoption of information and technology standards, and providing access to data and information sources and tools.

Facilitating Global Partnerships

Forging partnerships with other interested individuals or organizations would promote synergy, distribute the overall workload, and result in a more comprehensive service to customers. OBIS could serve as a critical thematic area within the international biodiversity information infrastructure. The sources, range (physical, chemical, and other parameters), and types (numeric,

textual, sound, video, and others) of data to be included are voluminous. Issues regarding precision, accuracy, and other considerations can be addressed only by members of the relevant scientific communities (academia, museums, government, industry, and others). OBIS could be the venue to bring these constituencies together, help develop consensus, and advocate for the needs of the marine science community.

Marine scientists would implement a two-pronged approach, one through their peers (global science community) and a second with funding agencies (geopolitical community). Members of OBIS could then advocate for the needs of marine science within their home countries and through scientific societies. In endorsing those needs, the scientific organizations could help effect change in the global community, influencing the evolution of the global biodiversity information infrastructure to include the needs of the marine biodiversity community.

Using the NBII as a starting point, for example, OBIS could provide access to a critical area of biology (both information and tools) that is currently not well represented within the NBII system. And the NBII could provide OBIS with immediate access to data that it has or will develop for the U.S. biodiversity information infrastructure, as well as access to tools for creating, organizing, manipulating, analyzing, and delivering marine data and information.

Providing an Organizational Framework for Ocean Biodiversity Information

In the same way that a library must consider how best to organize its information collection, the international biodiversity information system must address the questions of how best to organize the vast virtual collection of biodiversity information. An integral element in this process is early identification of just what is needed and where information gaps exist.

Current OBIS design criteria describe a system developed around two thematic areas—species and geographic area—that are efficient and useful approaches to organize, display, and deliver data. But OBIS must also assume a role in identifying critical data needs (such as abundance and appropriate spatial and temporal resolution), as well as in identifying protocols needed to facilitate the collection and categorization of datasets for aggregation. OBIS could also help address vital questions such as: “What steps can the marine science community can take to encourage posting individual data sets at accepted archives, as the gene community is doing?”; “Are there incentives to encourage earlier sharing of datasets via the Web?”; “What about orphan data?”

Fostering the Adoption of Information and Technology Standards

Early establishment and adoption of standards is critical for effective operation of systems that provide

access to distributed datasets. The standards currently being used by OBIS, including Web-based standards for information technology and the Federal Geographic Data Committee standards for spatial data, provide an excellent starting point. But consideration should also be given to including additional standards that address data and information content, particularly in regard to metadata. For example, OBIS could collaborate with Integrated Taxonomic Information System (ITIS) for North American waters and Species 2000 for data from other locations around the globe to ensure completeness and accuracy of those taxonomic authority databases.

Not only is it critical to define organisms in a standardized manner, it is also critical to describe other dataset concepts. This process requires development of a controlled vocabulary that can be used in describing information and improve the performance of search engines. Such a controlled vocabulary for biological data is now under development through a partnership with the California Environmental Resources Evaluation System (CERES) and the NBII. In a preliminary study for Internet search results using metadata produced with this controlled vocabulary and a commercial search engine, a search for "common loon" (*Gavia immer*) was reduced from tens of millions of hits to just a few very relevant datasets. OBIS could participate in this and other innovative information access approaches that could significantly enhance the usability of the system.

Two other current activities that are dealing with biological thesauri may also warrant additional investigation: the International Council of Scientific and Technical Information's Life Science Explorer, and a United Nations Environment Program (UNEP) involving environmental treaties.

Providing Access to Information, Data Sources and Tools

The biodiversity system envisioned would be both distributed and global, and would allow for interaction among dispersed datasets. This approach allows scientists to "control" their data, provides for easy updating by data owners, distributes costs across many system administrators, and generally is well received by dataset originators. Furthermore, the approach chosen for OBIS—relying upon specimen-based data only from recognized expert systematists—will help ensure data quality, a critical consideration for any data set.

As the base of data grows, it will be necessary to develop the ability to search the various datasets and identify only those data that are relevant to the user. Metadata will have to be created for each dataset so that the search engine can effectively identify and deliver only

the desired datasets, as described above. Tools to facilitate the creation of metadata will be needed to encourage cooperation by data holders. Other capabilities could also be offered, such as tools that allow for data analysis or alternate ways to display data.

OBIS should plan for the ability to incorporate species distribution models (as it has GARP [Genetic Algorithm for Rule-set Production]) and similar tools that may emerge by using appropriate standards, as noted above, and maintaining awareness of other efforts. With a number of initiatives currently under way that focus on integrating scientific disciplines and addressing biocomplexity, many more such tools or applications will likely be developed. The challenge that the new information environment creates revolves around not just serving catalogs of data but also must include allowing the user to interact with the data-combining, modeling, comparing, and analyzing-to address specific information needs.

Opportunities for partnering will occur in many sectors. A partnership with the NBII could provide assistance in the development of metadata-creation tools to allow researchers to more easily categorize their datasets, help in the implementation of software tools such as BioBot (a biological search engine that facilitates retrieval and profiling of information on the Internet), and allow participation in NBII Regional and Thematic Nodes (PCAST, 1998). From an organizational perspective, collaboration and coordination with the Global Biodiversity Information Facility will likely prove fruitful as well. By taking these preliminary steps, OBIS will be poised to expand as relevant applications and datasets emerge.

Conclusion

An information system focused on marine biodiversity information is sorely needed. Such a system will play a large role in the development of an overall global biodiversity information infrastructure, both by encouraging better access to relevant data, and through organizational integration with the global system. Early in its formation, OBIS should seek input from the scientific community and from national and international biodiversity information system developers. By collaborating with other biodiversity projects, OBIS will not only gain access to a functioning world-wide information infrastructure and its inherent tools and applications, but OBIS developers will also have the opportunity to influence emerging information standards and tools to include the needs of ocean science.

The challenge that the new information environment creates revolves around not just serving catalogs of data but also must include allowing the user to interact with the data-combining, modeling, comparing, and analyzing-to address specific information needs.

Acknowledgements

My thanks to Fred Grassle of Rutgers University, one of the founders of the OBIS concept, for the opportunity to participate in this important project. I would also like to thank NBII Director Gladys Cotter for her valuable insights and for providing the figure. NBII editor Terry D'Erchia conducted a careful and thoughtful review of the manuscript, and her contributions are appreciated.

REFERENCES

- Ausubel, J.H., 1999: Toward a Census of Marine Life. *Oceanography*, 12(3), 4-5.
- Cotter, G., F.D'Erchia, H. Fleet, A. Frondorf, M. Nyquist, G. Waggoner and J. Waldon, 1996: Developing a Biological Information Infrastructure. In: *Proceedings of the Conference on Scientific and Technical Data Exchange and Integration*. December 15-17, 1997, Bethesda, Maryland.
- Discover Life, 1998: *All Taxa Biodiversity Inventory of Great Smoky Mountains National Park*. World Wide Web: <http://www.discoverlife.org/>

- Grassle, J.F. and K.I. Stocks, 1999: A Global Ocean Biogeographic Information System (OBIS) for the Census of Marine Life. *Oceanography*, 12(3), 12-14.
- Mooney, H.A., C.S. Adam, A. Larigauderie and J. Sarukán, 2000: *International Biodiversity Observation Year*. World Wide Web: http://www.nrel.colostate.edu/-IBOY/mooney_essay.html
- National Research Council, 1993: *A Biological Survey for the Nation*. National Academy Press, Washington D.C., 205 pp.
- Organization for Economic Cooperation and Development (OECD), 1999: *Final Report of the OECD Megascience Forum Working Group on Biological Informatics*.
- President's Committee of Advisors on Science and Technology (PCAST), 1998: *Teaming with Life: Investing in Science to Understand and Use America's Living Capital*. Washington, DC, 86 pp.
- Summit of the Americas, 1996: *Plan of Action for the Sustainable Development of the Americas*, Article 31.
- Summit on Sustainable Development, Santa Cruz de la Sierra, Bolivia. 

GLOSSARY

ABI	Association for Biodiversity	The membership organization for the Information International Network of Natural Heritage Programs and Conservation DataCenters (the Heritage Network). ABI and its member programs, in collaboration with The Nature Conservancy, are working to create comprehensive regional, national, and international data products and services related to biodiversity for use by government, industry, scientists, educators, and the interested public.
BCIS	Biodiversity Conservation	A global information initiative by 12 international conservation organizations and programs.
BioBot		Online search tool designed to help users in a number of biological data discovery areas, including profiling or defining the information of interest.
CERES	California Environmental Resources	An information system developed by the California Resources Agency to facilitate access to a variety of electronic data describing California's rich and diverse environments.
CHM	Clearing-House Mechanism of the Convention on Biological Diversity	A network of 176 countries working together to facilitate access to and exchange of biodiversity information around the world.

Conabio	National Commission for the Knowledge and Use Biodiversity	Promotes and develops scientifically based activities whose aim is to explore, study, protect, or find a sustainable use for biological resources in Mexico.
DIVERSITAS		A partnership of scientists from all parts of the world whose goal is to provide accurate scientific information and predictive models of the status of biodiversity and sustainability of the use of Earth's biotic resources, and to build a world-wide capacity for the science of biodiversity.
ERIN	Environmental Resources Information Network	A national facility coordinated by the Australian Department of the Environment. Provides access to key information on the Australian environment. Integral to supporting requirements for ecologically sustainable development.
GAP	Gap Analysis Program	U.S. Geological Survey program that provides regional assessments of the conservation status of native vertebrate species and natural land cover types, and assists in the application of this information to land management activities.
GBIF	Global Biodiversity Information Facility	Main goal is to produce home pages for all species of organisms, dynamically derived from online databases. GBIF should lead to an accelerated rate of describing new species and to new information markets on biodiversity, and complement the Clearing-House Mechanism of the Convention on Biological Diversity.
IABIN	Inter-American Biodiversity Information Network	An initiative of the Summit of the Americas to provide greater and more useful access to biodiversity information throughout the hemisphere.
ICSU	International Council of Scientific Unions	Comprised of 98 multi-disciplinary National Scientific Members (scientific research councils or science academies) and 26 international, single-discipline Scientific Unions to provide a wide spectrum of scientific expertise, enabling members to address major international, inter-disciplinary issues that none could handle alone.
IGBP	International Geosphere-Biosphere Program	Focused on acquiring basic scientific knowledge about the interactive processes of biology and chemistry as they relate to global change.
INBio	Instituto Nacional de Biodiversidad	Generates knowledge about biodiversity in Costa Rica, communicating and promoting this information in many formats to a broad spectrum of national and international users.

ITIS	Integrated Taxonomic Information System	An authoritative source for biological names for the biota of North America.
IUCN	International Union for Conservation of Nature and Natural Resources	The world's largest conservation organization, bringing together 76 states, 111 government agencies, 732 non-government organizations, 36 affiliates, and more than 10,000 scientists and experts from 181 countries in a unique worldwide partnership. Its mission is to influence, encourage, and assist throughout the world to conserve the integrity and diversity of nature and to ensure that any use of natural resources is equitable and ecologically sustainable.
IUBS	International Union of Biological Sciences	A non-governmental, non-profit organization established in 1919 to promote the study of biological sciences; to initiate, facilitate and coordinate research and other scientific activities that require international cooperation; to ensure the discussion and dissemination of the results of cooperative research and international conferences, and to assist in the publication of their reports.
IUMS	International Union of Microbiological Societies	One of the 20 scientific unions of the ICSU, consisting of three divisions: Bacteriology and Applied Microbiology, Mycology, and Virology.
MABNET Americas	Man and the Biosphere Network	An interdisciplinary program of research and training intended to develop the basis, within the natural and the social sciences, for the rational use and conservation of the resources of the biosphere, and for the improvement of the global relationship between people and the environment.
NABIN	North American Biodiversity Information Network	The NBII program participates on the Steering Committee for the development of NABIN, which seeks to increase access to and integration of biodiversity data from Canada, Mexico, and the United States.
NBII	National Biological Information Infrastructure	A broad, collaborative program to provide increased access to data and information on the nation's biological resources.
OBIS	Ocean Biogeographic Information System	A proposed information system to provide centralized access to existing ocean biodiversity datasets and to incorporate new data created by the Census of Marine Life.
OECD	Organization for Economic Cooperation and Development	An organization of 29 member countries that provide governments with a setting in which to discuss, develop, and perfect economic and social policy.

PCAST	President's Committee of Advisors on Science and Technology	Committee members, appointed by the President of the United States, are drawn from industry, education, and research institutions, and other non-governmental organizations. They provide feedback about Federal programs and actively advise the National Science and Technology Council about science and technology issues of national importance.
SCOPE	Scientific Committee on Problems of the Environment	A recognized authority at the interface between the science and decision-making spheres, providing advisors, policy-planners and decision-makers with analytical tools to promote sound management and policy practices. Encompasses a world-wide network of 40 national science academies and research councils, and 22 international scientific unions, committees, and societies.
Species 2000		Objective of enumerating all known species of plants, animals, fungi, and microbes on Earth as the baseline dataset for studies of global biodiversity.
TNC	The Nature Conservancy	Private international conservation group whose mission is to preserve plants, animals, and natural communities that represent the diversity of life on Earth by protecting the lands and waters they need to survive.
UNEP	United National Environment Program	Established to protect and promote the enhancement of the human environment.
UNESCO	United Nations Educational, Scientific and Cultural Organization	Contributes to peace and security in the world by promoting collaboration among nations through education, science, culture and communication. United Nations charter 1945.

