CHAPTER 20

FishBase as Part of an Oceania Biodiversity Information System

Rainer Froese and Maria Lourdes D. Palomares

ABSTRACT

FishBase is an electronic encyclopedia on fish being developed at the International Center for Living Aquatic Resources Management with the cooperation of the Food and Agriculture Organization of the United Nations and funded by the European Commission. To date, FishBase contains key information for about 12,000 fish species globally, including about 2,800 species of the Oceania region. FishBase is available on CD-ROM for personal computers with Windows 3.1 or later. A design for an Oceania Biodiversity Information System is presented, and the integration of FishBase into such a system is discussed.

INTRODUCTION

FishBase, a large biological database on fish, is being developed in a joint project between the International Center for Living Aquatic Resources Management (ICLARM) and the Food and Agriculture Organization of the United Nations (FAO) funded by the European Commission. FishBase contains key information on fish populations such as nomenclature, ecology, population dynamics, aquaculture, genetics, physiology, and occurrence of fishes (Froese 1990, 1993; Palomares et al. 1991; Pauly and Froese 1991a, 1991b; Froese, Palomares, and Pauly 1992; Palomares and Pauly 1992a, 1992b; Pauly, Palomares, and Froese 1993; Froese and Pauly 1994b). It was conceived as a "tool" to help fisheries researchers and managers to better understand and manage their natural resources.

FishBase was designed to answer, for example, the following questions for the fishes of a given country:

- What is the current scientific name and classification?
- What is the international [FAO, American Fisheries Society, FishBase) common name?
- What is the global commercial importance?
- What key information (life history, population dynamics) is available?
- What actually are these key parameters?
- What are the research gaps?
- What species are (or can be) used for aquaculture?

- What species are introduced?
- What species are of interest to sports fishers?
- What species are of interest to the ornamental trade?
- What species are threatened?
- What species are endemic?
- What species are dangerous to humans?

FishBase also offers a structure for national information such as occurrence, importance, use, or restrictions in a given country. Such information can only be obtained from national researchers and managers, and the FishBase Project has started collaborative efforts with Hawaii, Taiwan, and Mexico to identify the most efficient procedure for such data exchange.

MATERIALS AND METHODS

FishBase contains more than 1,000 data fields organized in 50 tables, totaling half a million records. More than 200 procedures access this information and create a variety of outputs. FishBase has been moved recently from a DOS (DataEase) to a Windows (Microsoft Access) database software, an undertaking that was much more difficult than we had anticipated. For a decent performance, FishBase now requires at least a 486 PC with at least 8 megabytes of memory. FishBase can be run from the CD-ROM on which it is distributed; however, it will be slow. There are options to install the data on the hard disk (faster; 120 megabytes required) and access the pictures on the CD-ROM, or to install everything on the hard disk (fastest; 430 megabytes required). FishBase can also be installed on a local area network (LAN) so that several persons can use it at the same time.

FishBase makes use of published literature (e.g., journal articles, technical reports, theses) and recent revisions of fish species or families such as those produced by FAO. Key information on any of the topics mentioned earlier is extracted and entered by a group of research assistants (all with graduate degrees in marine biology) based at the ICLARM headquarters in Manila. Collaborators worldwide participate in data entry through data collection forms. Some collaborators have developed similar electronic databases on specific topics (e.g., Eschmeyer, this volume) and have opted to make their databases available by incorporation and distribution through FishBase. Contributions are given proper citations and acknowledgments in FishBase: every datum is attached to the reference where it stems from, and every record bears a "stamp" identifying who entered, contributed, modified, or validated the information.

The quality of the information in the database is monitored in-house through several loops of verification by the research assistants who check each other's work and also through collaborating scientists who offer to check the information in FishBase for their different fields of expertise.

The Project has purchased a CD-ROM recorder with software (~US$8,000) to produce CD-ROMs in-house, allowing for continuous creation of updates. The production time for one CD-ROM takes about one hour, and a blank CD-ROM costs about US$20.

RESULTS AND DISCUSSION

As of October 1994, the FishBase team has been able to incorporate more than 12,000 species extracted from more than 7,000 references, representing half of the estimated 25,000 species of fish in the world (Figure 20.1). Of these, over 2,800 are from marine waters in Oceania, extracted from more than 2,000 references.

In addition, FishBase has become the repository of several outstanding collections—the first four of which have been compiled in-house:

- the largest collection of population dynamics data (growth parameters, natural mortality, length-weight relationships, maximum ages and sizes) entered in more than 4,000 records covering over 1,000 species;
- the largest collection of ecological data (prey, 4,400; predators, 800; diet composition, 700; and food consumption, 150);
- the largest available collection of electrophoretic data entered in 3,600 records for about 50 populations/strains;
- the largest collection of common names (by country and language) entered in about 50,000 records for more than 8,000 species;
- the genera of recent fishes (from Eschmeyer 1990; see also Eschmeyer, this volume);
- the largest collection of data on fish metabolism (from Thurston and Gehrke 1991) entered in 7,400 records for about 300 species;

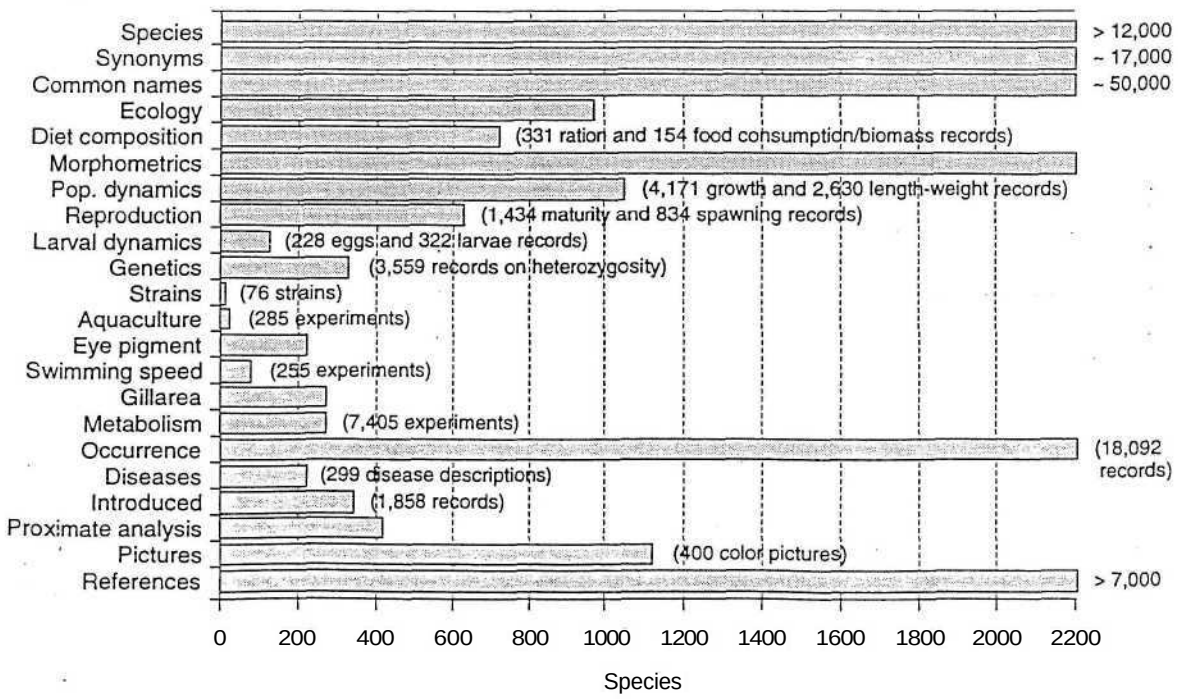


Figure 20.1 Content of FishBase as of October 1994 in number of species covered and/or number of records

- the largest collection of data on larval dynamics (from Houde and Zastrow 1992) covering about 100 species;
- the largest collection of data on introduced fishes (mainly from Welcomme 1988) entered in 1,815 records covering about 360 species; and
- about 1,000 line drawings or color pictures of adults, larvae, eggs, or diseases.

The WINMAP software, a low-level geographic information system that forms part of the FishBase package, can be used to display a variety of maps based on the occurrence and abundance records in the database. This feature is specially useful in fish biodiversity research. Maps showing the global occurrence of members of an order, family, subfamily, genus, or a species can be created.

When representative levels of occurrence and abundance records are reached, it will be possible to use FishBase to determine a theoretical status of threat for certain species or species groups by combining trends in distribution and abundance with a set of biological characters such as environmental tolerance, fecundity, longevity, and age at first maturity.

The CD-ROM version of FishBase is distributed at nominal costs to fisheries institutions worldwide. Special emphasis will be given to developing countries, some of which will be supported in purchasing the necessary hardware and receiving training on how to use FishBase and related analytical tools.

The success of FishBase will depend on its acceptance (= use) by researchers, managers, teachers, and students. It also depends on collaboration of experts in the various fields to ensure quality of the information. Collaborators are entitled to a free copy of FishBase. Others can order the FishBase CD-ROM and manual from ICLARM for US$95 (including air mail).

RECOMMENDATIONS FOR A MARINE/COASTAL DATABASE SYSTEM FOR OCEANIA

Before a database system can be designed, one has to be clear about the purpose it should serve. The objectives mentioned in the background information for the workshop call for a "data management system ... [to assist in a] ... long-range program to conserve biodiversity" and to "summarize information on the systematics and taxonomy of nearshore marine and coastal species in Oceania." We quote the abstract from Froese and Pauly (1994a):

> In order to improve our understanding of aquatic biodiversity, it is suggested to assemble in a single database the huge amount of existing data on the occurrence of aquatic species in space and time. Such data are available in museum collections, research vessel surveys, tagging studies, the scientific literature, and a variety of other sources, often in digitized form. The database would be distributed on CD-ROM with annual upgrades. It would preserve data which might otherwise be lost; it would provide baseline data on biodiversity from historic data sets; in combination with data derived from existing biological, oceanographic, and meteorological databases it would allow for analyses of biodiversity which are currently not possible; and it would guide the ongoing efforts towards collection of data that are most useful for analytical models. We suggest to establish a network of institutions that hold relevant data and are willing to share it.

The database structure suggested by Froese and Pauly (1994a) is presented here again, in more detail and with some modifications. The database for aquatic biodiversity should consist of the following tables (Figure 20.2):

The Specimens table is the center of the database. It contains one record per specimen encountered and is similar to existing collection databases. It contains the following fields: genus, species, taxonomic group, catalog code, record type (museum, survey, literature, tagging), collector, station code, number of specimens, length, weight, sex, live stage, abundance, . . . , remarks, contributor code.

Taxonomic group, genus, and species fields of the Specimens table are linked to taxonomic dictionaries such as Eschmeyer's Genera and forthcoming Species databases for fishes or CoralBase for corals (see Eschmeyer, and Navin and Veron, respectively, this volume). The dictionaries ensure the validity and proper spelling of the scientific name and provide the classification into higher taxa. The same fields will also provide the link to biological databases such as FishBase.

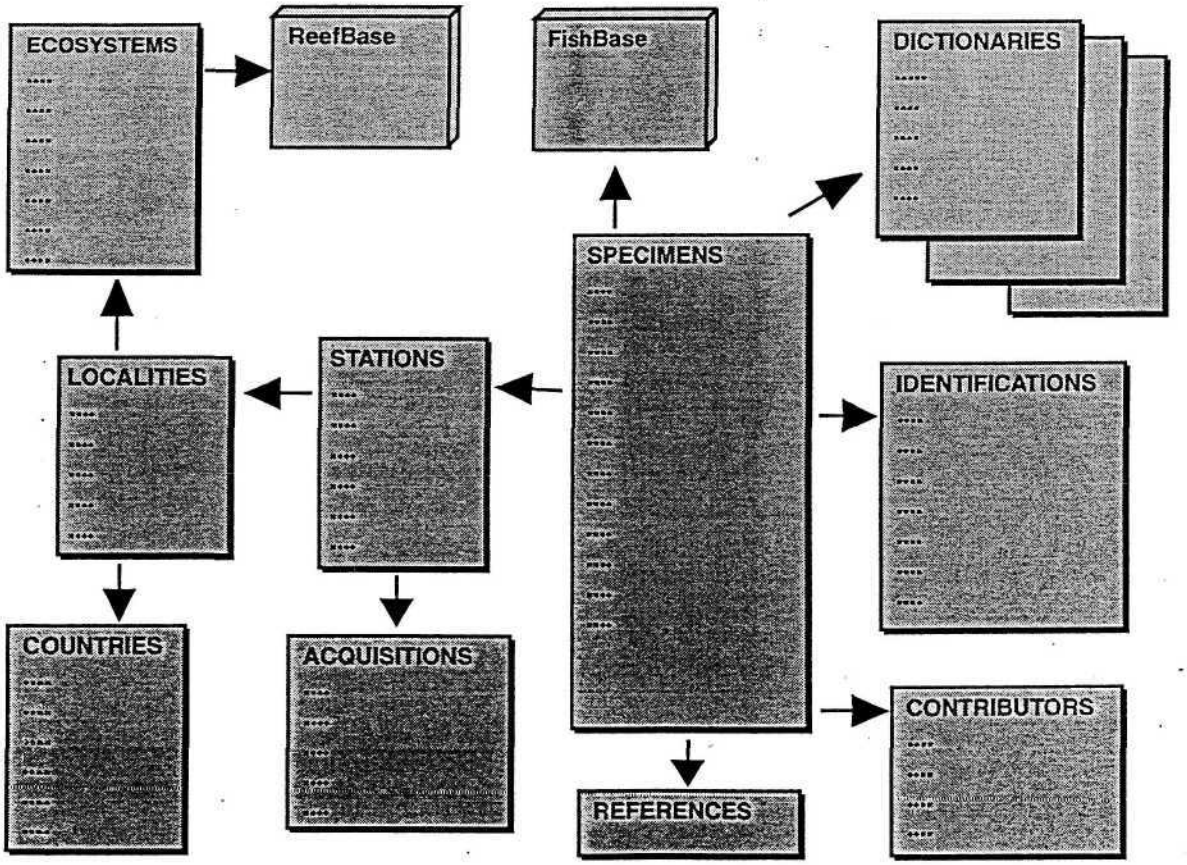


Figure 20.2 Structure suggested for the Oceania Biodiversity Information System

Through the taxonomic group and catalog code, the Specimens table is linked to the Identifications tables that are specific for each taxonomic group and contain the following fields: taxonomic group, catalog code, identifier, date, identified as, length, head length, . . . (other group-specific morphometric counts or measurements), remarks, reference. These tables keep track of the taxonomic work and verify identifications where no specimen is preserved (e.g., during surveys).

The station code links the Specimens table to the Stations table containing the following fields: station code, station name, locality code, acquisition code, date, time, latitude, longitude, depth, altitude, temperature, salinity, gear, remarks, reference. The Stations table, which is similar to those used in trawl surveys, facilitates data entry because station data are entered only once for all specimens encountered.

The locality code links the Stations table to the Localities table, which is a gazetteer of aquatic localities and contains the following fields: locality code, locality name, locality type (river, lake, lagoon, reef, seamount), locality description, ecosystem code, country code, latitudinal range, longitudinal range, other names used, remarks. This table should provide links to ecosystem databases such as ReefBase.

The acquisition code links the Stations table to the Acquisitions table, which contains the following fields: acquisition code, expedition/survey name, date from, date to, collector, donor, reference, remarks. This table is similar to the acquisition file maintained in museums. It facilitates data entry in that this information need not be repeatedly entered for each station of a survey.

The country code links the Localities table to the Countries table, which contains the following fields: country code, country name, area, shelf area, coastline, population, coastal population, languages, . . . , reference. This table standardizes country names and provides additional information that might be useful in comparative studies. This table already exists in ReefBase and FishBase.

The ecosystem code links the Localities table to the Ecosystems table, which contains the following fields: ecosystem code, ecosystem name, ecosystem type, latitudinal range, longitudinal range, surface area, drainage area, depth range, average depth, salinity range, surface temperature range, average surface temperature, . . . , remarks, reference.

The ecosystem classification should closely follow the one described by Holthus and Maragos, this volume. For reefs, this table can draw on information contained in ReefBase (McManus et al., this volume).

The contributor code in the Specimens table establishes the link to the Contributors table, which identifies the collaborating individuals and institutions and contains the following fields: contributor code, name, institution, address, e-mail, remarks.

A References table will be linked to the references mentioned in the other tables and will contain the full citation.

This structure, consisting of more than ten tables with many code fields, will be largely hidden from the user who will just click on buttons and select options from choice lists to access the desired information.

As Froese and Pauly (1994a) have pointed out, "A species name, a date, and a locality do not seem to be much of a base for sophisticated analytical models. However, one has to realize that these three bits of data each represent a vast amount of information: the species name actually provides us with all the biological information on an organism; the locality leads to all the ecological information available for a site; and the date provides us with information about seasonal and historical environmental conditions." As a result of subsequent discussions, the source (i.e., a reference or the name of the provider of the information) is added as the fourth bit of core biodiversity data. The source gives an indication of the reliability of the information, leads to other reports and species identified by the collector, and gives feedback on the experience of the collector.

Thus, the proposed system will be able to answer, for example, the following questions:

- What species occur in a certain country or ecosystem?
- Where and when has a certain sensitive species last been collected?
- What surveys have been made in a certain ecosystem or country?
- What species have been collected during a certain survey?
- What are all the ecosystems in which a certain species occurs?
- What biological characters are shared between species of a certain ecosystem?
- What is the percentage of indicator species in a given area compared to other areas?

A crucial point of the suggested system is the quality of the identifications. It will be necessary to determine expert centers for each species group that will classify the probability of correctness (probably correct, doubtful, probably misidentified) for each new record. There will probably be a need to include experts outside of Oceania.

Froese and Pauly (1994a) suggested, "Start this exercise with a [international] project which, if successful, will be turned into a permanent activity of an appropriate international body, similar to the institutionalized gathering of meteorological, oceanographic, and recently also coral reef data."

Acknowledgments—We thank the members of the FishBase team—Daniel Pauly, Susan Luna, Crispina Binohlan, Armi Torres, Liza Agustin, Maria Teresa Cruz, Pascualita Sa-a, Roberto Cada, Rodolfo Reyes, Jr., Emily Capuli, Rachel Atanacio, and Portia Bonilla—for their dedicated work. We also thank our collaborators throughout the world who believed in our vision.

Note: This is ICLARM Contribution No. 1104.

REFERENCES

Eschmeyer, W. N. 1990. *Catalogue of the genera of recent fishes.* California Academy of Sciences, San Francisco.

Froese, R. 1990. FishBase: An information system to support fisheries and aquaculture research. *Fishbyte* 8(3): 21-24.

———. 1993. Study group report on FishBase. ICES C. M. 1993/L:6.

Froese, R., M. L. D. Palomares, and D. Pauly. 1992. Draft user's manual of FishBase. ICLARM, Manila, Philippines.

Froese, R., and D. Pauly. 1994a. A strategy and a structure for a database on aquatic biodiversity. In *Data sources in Asian-Oceanic countries,* ed. J.-L. Wu, Y. Hu, and E. F. Westrum, Jr., 209-20. Committee on Data for Science and Technology, Ann Arbor, Michigan.

———, eds. 1994b. *FishBase user's manual.* ICLARM Software No. 7, Manila, Philippines.

Houde, E. D., and C. E. Zastrow. 1992. Ecosystem- and taxa-specific dynamic and energetic properties of fish larvae assemblages. *Bull. Mar. Sci.* 53(2): 290-335.

Palomares, M. L. D., J. Moreau, P. Reyes-Marchant, R. Froese, and D. Pauly. 1991. FishBase: Une base de données sur les poissons. *Fishbyte* 9(2):58-61.

Palomares, M. L. D., and D. Pauly. 1992a. FishBase as a worldwide computerized repository of ethnoichthyology or indigenous knowledge on fishes. Paper presented at the International Symposium on Indigenous Knowledge and Sustainable Development, 20–26 September, International Institute for Rural Reconstruction, Silang, Cavite, Philippines. .

———. 1992b. Use of FishBase for assembling information on the fishes of Cambodia, Laos, and Vietnam. Paper presented at the 3d Asian Fisheries Forum, 26–30 October, World Trade Center, Singapore.

Pauly, D., and R. Froese. 1991a. FishBase: Assembling information on fish. *Naga, ICLARM Q.* 14(4): 10–11.

———. 1991b. The FishBase project, or how scattered information on fish can be assembled and made useful for research and development. *EC Fisheries Cooperation Bulletin,* December.

Pauly, D., M. L. D. Palomares, and R. Froese. 1993. Some prose on a database of indigenous knowledge on fish. *Indigenous Knowledge and Development Monitor* 1(1): 26–27.

Thurston, R. V., and P. C. Gehrke. 1991. Respiratory oxygen requirements of fishes: Description of OXYREF, a data file based on test results in the published literature. In *Fish physiology, toxicology, and water quality management. Proceedings of an International Symposium, Sacramento, California, USA, 18-19 September 1990,* ed. R. C. Russo and R. V. . Thurston, 95-108. U.S. Environmental Protection Agency EPA/600/R-93/157.

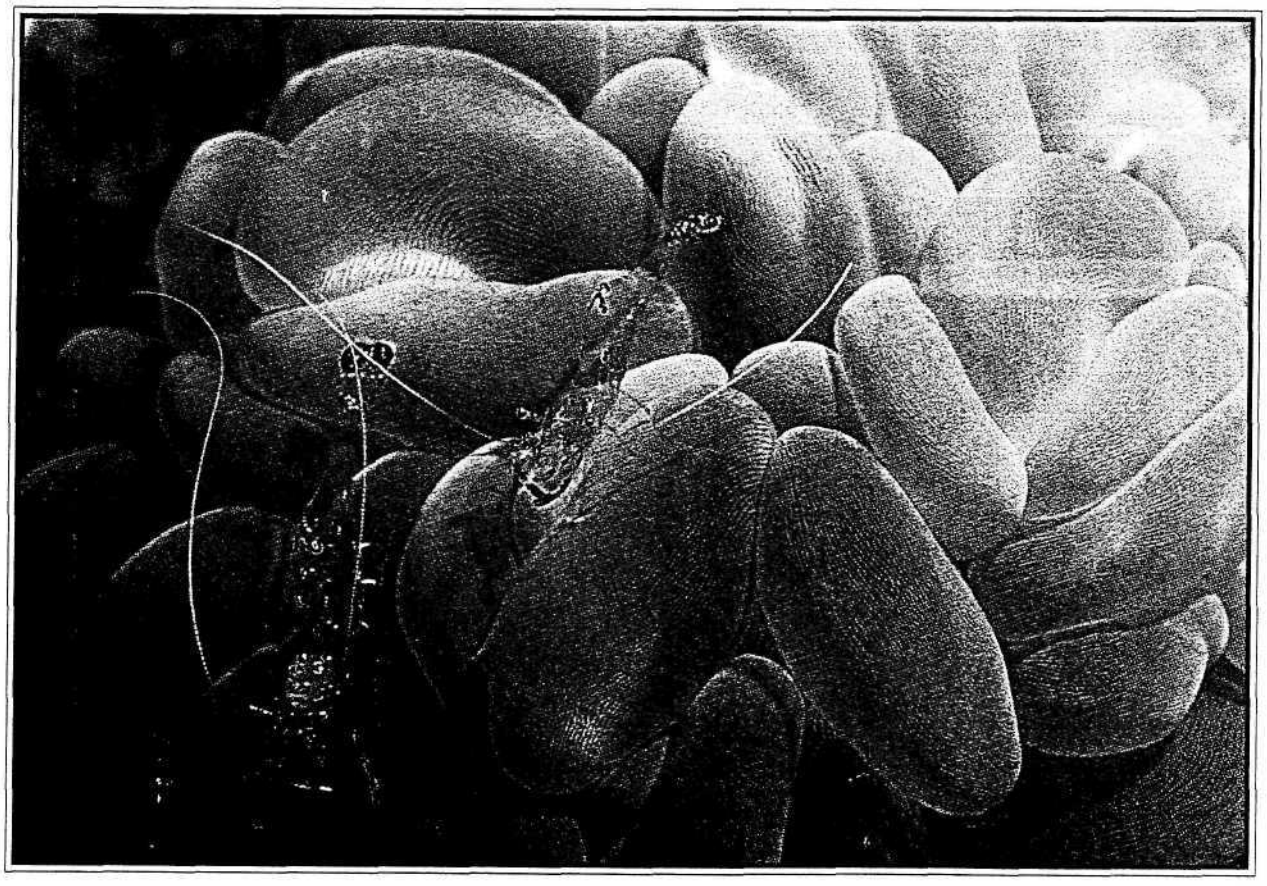
Welcomme, R. L. 1988. International introductions of inland aquatic species. FAO Fish. Tech. Paper 294, Rome.

Marine and Coastal Biodiversity in the Tropical Island Pacific Region

Volume 1
Species Systematics and Information Management Priorities

edited by

James E. Maragos, Melvin N. A. Peterson,
Lucius G. Eldredge, John E. Bardach, and Helen F. Takeuchi



Program on Environment
East-WestCenter
Honolulu, Hawaii

Ocean Policy Institute
Pacific Forum/CSIS
Honolulu, Hawaii

Pacific Science Association
c/o Bishop Museum
Honolulu, Hawaii