

The North Pacific Ecosystem Metadatabase

S. Allen Macklin
Pacific Marine Environmental Laboratory, NOAA
7600 Sand Point Way NE,
Seattle, Washington 98115, U.S.A.
E-mail: Allen.Macklin@noaa.gov



Bernard A. Megrey
Alaska Fisheries Science Center, NMFS
7600 Sand Point Way NE,
Seattle, Washington 98115, U.S.A.
E-mail: Bern.Megrey@noaa.gov



Mr. Allen Macklin and Dr. Bernard Megrey co-direct the North Pacific Ecosystem Metadatabase and co-chair Data Management and Communications for the Alaska Ocean Observing System. Allen is a meteorologist with the Pacific Marine Environmental Laboratory. He is the coordinator for Fisheries-Oceanography Coordinated Investigations (FOCI), a NOAA research program to sustain fishery resources in the Gulf of Alaska and Bering Sea while maintaining healthy ecosystems. Allen has 25 years' experience studying Alaskan marine ecosystems and managing information. Bernard is a research fisheries biologist with NOAA's Alaska Fisheries Science Center where he has worked since 1982. As the lead investigator for recruitment modeling studies for FOCI, he has over 20 years' experience studying dynamics of exploited North Pacific fish populations, relationships of environment to recruitment variability, and application of computer technology to fisheries research and natural resource management. [This article is PMEL contribution 2703 and FOCI contribution 500.]

What is the North Pacific Ecosystem Metadatabase?

The North Pacific Ecosystem Metadatabase (NPEM, Fig. 1) is an Internet utility to aid the understanding, management, stewardship and utilization of North Pacific Ocean ecosystems. The utility is a browsable, and searchable, on-line inventory of data and other information. NPEM is dynamic, *i.e.*, it undergoes continuous development to keep its contents up to date, so that users can access current information from which to make decisions. Our goal is to provide free and open access to information that ordinarily would be unavailable to researchers. In this way, we hope to build collaborations between investigators, and to make the exchange and use of marine science data more efficient.

Metadata

Metadata, or data about data, describe the content, quality, condition, and other characteristics of data. For example, metadata for vertical profiles of ocean properties obtained from hydrographic casts might illustrate, minimally, the locations and times of the

casts, the inclusive depths, the variables measured, the location of the data, and the name of the contact person to request access to the data. In general, metadata include thematic, semantic and syntactic descriptors of the data they reference.

- Thematic metadata describe the context of the study that produced the data. Such descriptors can include, *e.g.*, principal investigator, species association and study hypothesis.
- Semantic metadata provide contextual information about the data. Candidate descriptors are measurement type, measurement device, units of measurement, calibration information, etc.
- Syntactic descriptors define the way the data are packaged, *e.g.*, file size, file format, storage mechanism and location.

Metadata in NPEM are described in a common set of terminology and definitions using the Federal Geographic Data Committee (FGDC) metadata standard. The data themselves are not part of the metadatabase and continue to reside with their owner. Each metadata record provides a dynamic link to the data or to the contributor.

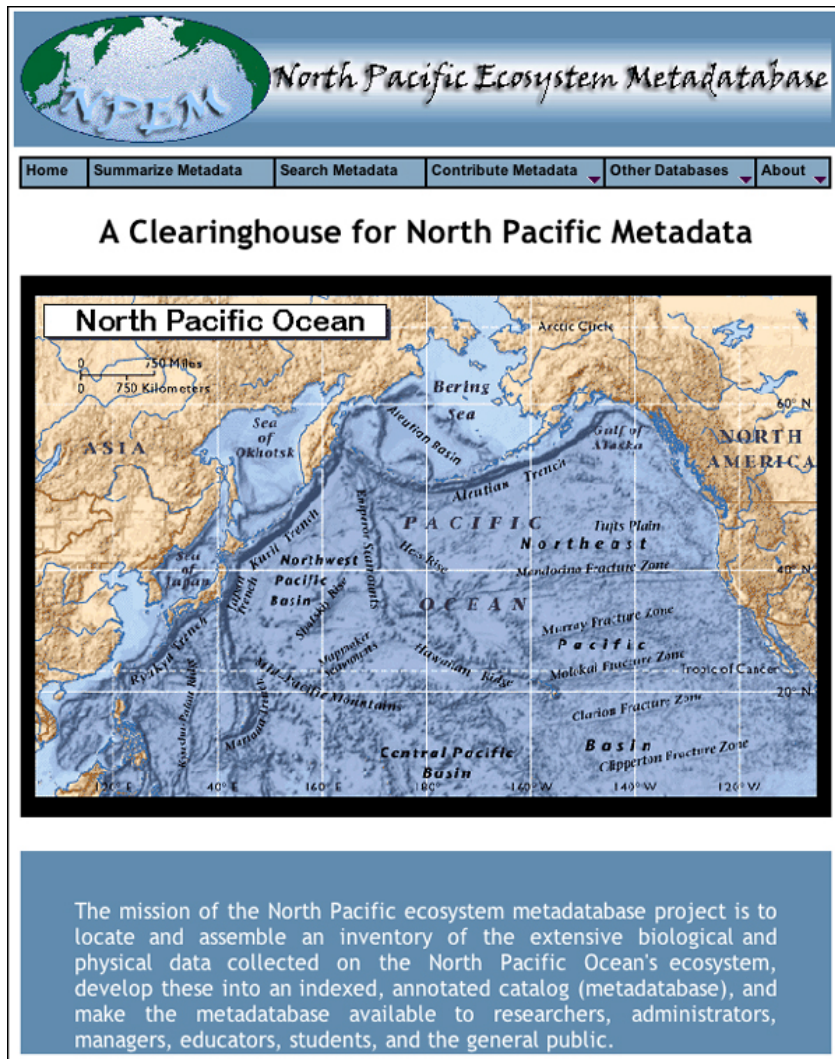


Fig. 1 Home page of the North Pacific Ecosystem Metadatabase.

History

NPEM began in 1996 as the Bering Sea Ecosystem Biophysical Metadatabase with a 3-year support from the National Oceanic and Atmospheric Administration (NOAA) Environmental Services Data Information Management (ESDIM). This earlier metadatabase is described in *PICES Press*, Vol. 6, No. 1 (January 1998). We established the metadatabase to address a deficiency identified in 1996 by the U.S. National Research Council. In its report on the Bering Sea ecosystem, the Council concluded that a directory of data and information sources relevant to the Bering Sea, cataloged in one place, was a critical need. Furthermore, the Council cited the lack of such a database as the one major impediment to studying the Bering Sea. It was clear that scientists had little appreciation of metadata or their importance.

We first developed a schema from the minimum set of FGDC descriptors, designed the database in Microsoft Access, and procured a Windows server as a public interface using Active Server Pages (ASP) scripts. In mid-1997, we published the first call for metadata. We solicited information from scientists, advertised in science newsletters, made national and international presentations, and, through PICES, developed contacts with Canadian, Chinese, Japanese, Korean and Russian marine science institutes. We educated the scientific community on the importance of metadata and indicated the benefits that would accrue to scientists and science as a result of proper metadata specification. We mailed thousands of metadata entry forms to scientists, requesting their metadata. From these efforts, the metadatabase grew to 70 records within a few months. By the end of the third year of funding, there were more than 1000 records populating the

metadatabase. Also in those first years, the metadatabase earned support and endorsement from Fisheries-Oceanography Coordinated Investigations (FOCI), the North Pacific Marine Science Organization (PICES), the *Exxon Valdez* Oil Spill Trustee Council (EVOS) and the North Pacific Marine Research Program, for which the metadatabase was granted funds to be the official program metadata repository.

In March 2001, the metadatabase directors attended a PICES-sponsored, international workshop on “Impact of climate variability on observation and prediction of ecosystem and biodiversity changes in the North Pacific”. Workshop participants from Canada, China, Japan, Korea, Russia, the United Kingdom, the United States, and 11 international science organizations nominated existing time series and predictions for determining status of North Pacific ecosystems. Attendees were amazed at the diversity and quantity of the many data series that were brought forward. Data from western Pacific nations has been particularly difficult to identify and obtain, as much of it is known only locally. The nominated time series from all around the North Pacific rim, basin and marginal seas, have sufficient historical length, accuracy, and likelihood of continuance to be important indicators of climate and climate response. Participants of the workshop recommended that the time series information and scientific contacts identified be recorded and updated in the North Pacific (*i.e.*, Bering Sea) Ecosystem Metadatabase. With this impetus, we again applied to ESDIM, successfully, to expand the Bering Sea Ecosystem Biophysical Metadatabase to NPEM, and that work began in late 2002. Search and display capabilities have been upgraded with this

funding, and the metadatabase is now housed in MySQL and served from a Linux platform.

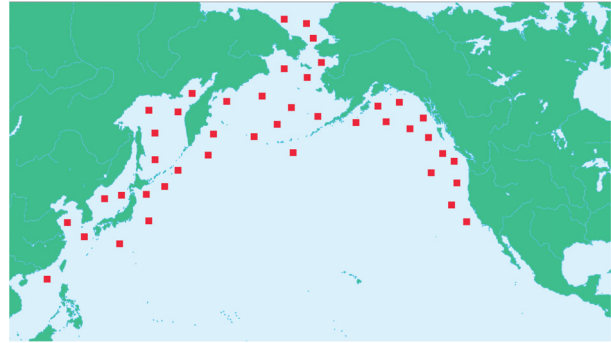


Fig. 2 Regional distribution of metadata records.

As of April 2004, NPEM contains 2746 records referencing physical and biological datasets, model outputs, museum samples, publications, reports, proposals, atlases, and audio and video programs. The regional distribution of these records is shown in Figure 2. Most records are from coastal areas. In terms of metadata density, most contributions pertain to the eastern North Pacific. We suspect that there has been a similar quantity of research performed in the western North Pacific, however results from this research are not as readily available to us. For example, although more than ten Asian institutes have contributed to the metadatabase, these records make up less than 11% of the holdings. Holdings span all biological and physical scientific disciplines, including historical and present information, ranging from atmosphere to open ocean to inter tidal areas. Figure 3 shows the distribution of metadata records by source country and keyword.

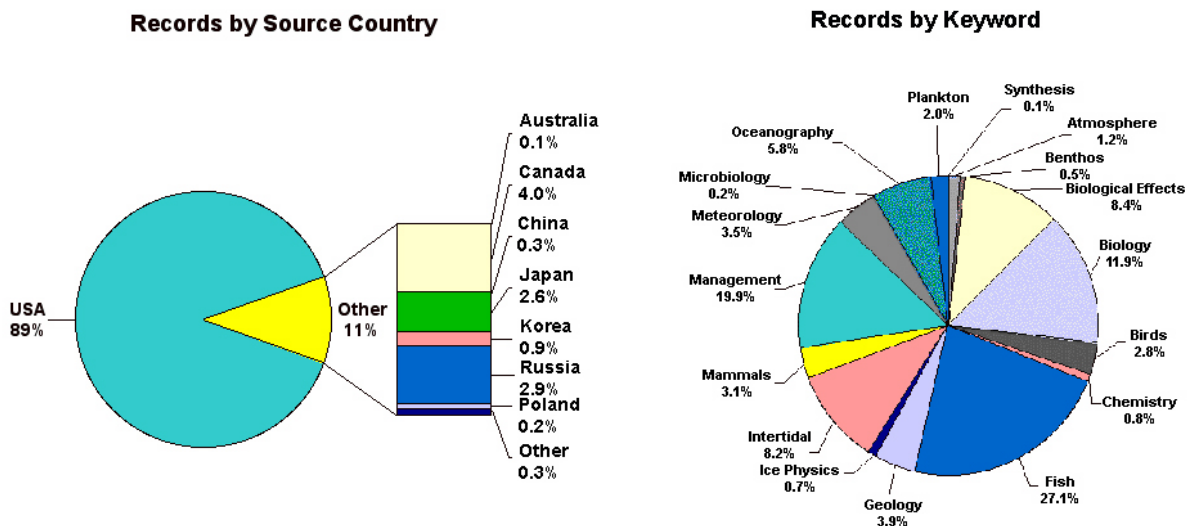


Fig. 3 Distribution of metadata records by source country and keyword.

Access

The North Pacific Ocean Theme Page (<http://www.pmel.noaa.gov/np/>) is the Internet gateway to the metadatabase. The Theme Page and the metadatabase offer a rich suite of environmental information to scientists, students, teachers, managers, and casual users. Since their inception, both the Theme Page and the metadatabase have increased in popularity as shown in Figure 4. Peaks in user activity correspond to important announcements of availability of research funds or other resources. Note the drop in summertime Theme Page usage when U.S. public school is not in session. On average, the metadatabase is exercised about 3500 times a month. This represents about 3% of all Theme Page use. A growth trend from 1997 to about 2002 seems to be leveling.

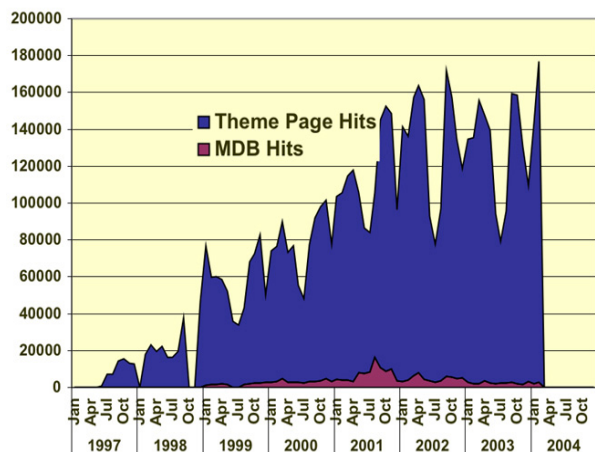


Fig. 4 Time history of Theme Page and metadatabase user activity.

The metadatabase is found through the Theme Page's DATA link or can be accessed directly at <http://www.pmel.noaa.gov/np/mdb/>. Once on-line, a user can learn about the metadatabase, contribute metadata, or search for metadata by time, location, keyword, country of origin, etc. Spatial searches are accomplished through an interactive map display or by direct specification of latitude and longitude. A user is able to build compound searches using any two or more search techniques.

Search results are returned according to user specification. Presently, the options are to return information as a list of metadata record titles or as

dynamically linked icons on a regional map, such as in Figure 2. Clicking on a metadata title or clicking on an icon will display the complete metadata record selected.

Complete metadata records display to the user all FGDC descriptive elements captured within NPEM. These are contributor, citation, description, status, ecosystem components, keywords, spatial domain, time domain, source, and constraints.

Future directions

We continue to archive all metadata associated with the North Pacific Ocean. In particular, we want to increase holdings of Asian metadata to enrich our references to the western North Pacific and bordering regions. To gain recognition in NPEM, a contributor can complete an on-line form through the website. This information is e-mailed to the NPEM coordinator who enters it into the metadatabase.

Late in 2003, we launched plans to implement with other North Pacific marine data centers (*e.g.*, KODC, JODC) "federated searches" or queries that search all metadata sets in separate data locations in a manner that is completely transparent to the user. Using this technique, a user of any of the aforementioned data facilities or of NPEM will be able to search the collection of all subscribing data facilities in a single session. In conjunction with this effort, NPEM will become a clearinghouse node of the National Spatial Data Infrastructure.

During the coming year, NPEM will implement a drill-down keyword thesaurus that is a derivative of the Global Change Master Directory. Metadata contributors will be able to specify keywords for their metadata by scientific discipline using a pull-down menu. Keywords are specified by category, topic, term, variable (optional), and detailed variable (optional and user-provided), *e.g.*, EARTH SCIENCE > OCEANS > OCEAN CIRCULATION > EDDIES > CYCLONIC. Metadatabase users will be able to search for records using the same pull-down menu.

Finally, NPEM will cooperate with various efforts to establish North Pacific Ocean observing systems. NPEM will make metadata available, enter into discussions about metadata standards for description of various sorts of ocean information, and entertain proposals to federate.