

### 3. BACKGROUND

The foundation of the PICES Metadata Federation effort was the creation and development of the North Pacific Ecosystem Metadatabase (NPEM, originally called the Bering Sea Ecosystem Biophysical Metadatabase). When faced with the decision of how to enlarge the contents of the NPEM, the directors sought a solution that would allow NPEM users access to new metadata without having to host those metadata in the NPEM. Prior interaction with the founders of Alaska's Cook Inlet Information Management and Monitoring System (CIMMS) had demonstrated the value of a distributed system.

#### 3.1. North Pacific Ecosystem Metadatabase

The NPEM (Macklin and Megrey 2004, Fig. 3.1, <http://www.pmel.noaa.gov/np/mdb>) is an Internet utility to aid in the understanding, management,

stewardship and utilization of North Pacific Ocean ecosystems. The utility is a browsable, 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. NPEM's goal is to provide free and open access to information that ordinarily would be unavailable to researchers. This practice enables collaborations between investigators and makes the exchange and use of marine science data more efficient.

NPEM began in 1996 as the Bering Sea Ecosystem Biophysical Metadatabase (Megrey and Macklin, 1998) with 3-year support from the National Oceanic and Atmospheric Administration (NOAA) Environmental Services Data Information Management (ESDIM). The

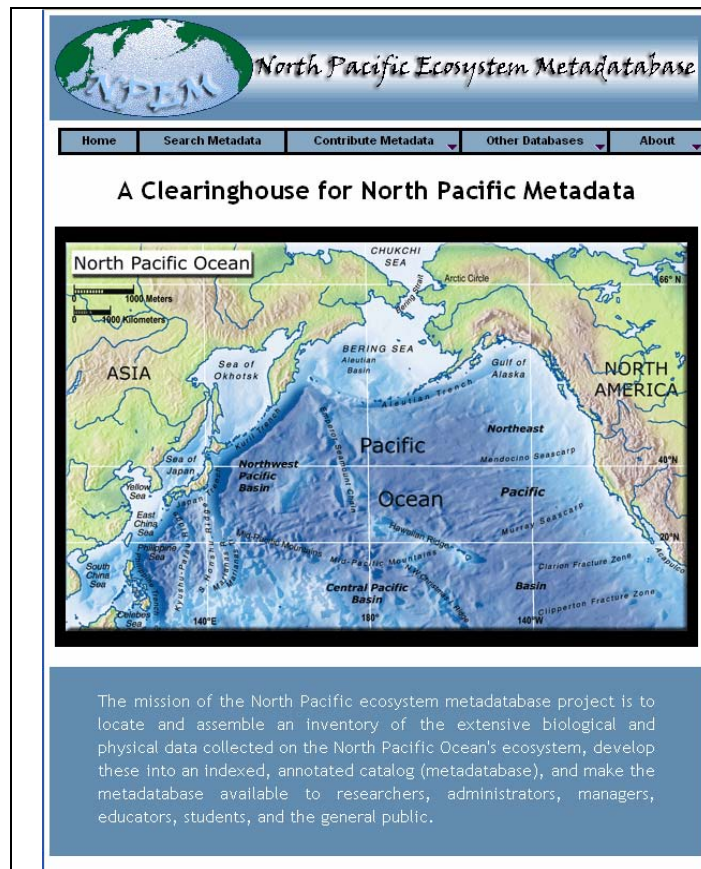


Fig. 3.1 Home page of the North Pacific Ecosystem Metadatabase (<http://www.pmel.noaa.gov/np/mdb>).

metadatabase addressed a deficiency identified by the National Research Council (1996). 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 at that time.

We first developed a schema from the minimum set of Federal Geographic Data Committee (FGDC) descriptive elements, 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, South 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 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, People’s Republic of China, Japan, Republic of Korea, Russia, the United Kingdom, the United States, and 11 international science organizations

nominated existing time series and predictions for determining the 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 the 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.

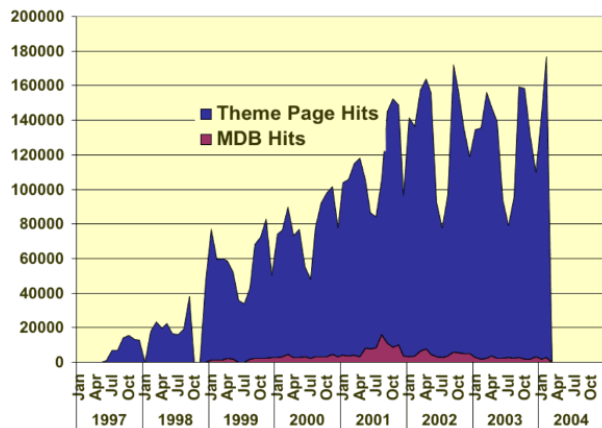
As of December 2006, the NPEM contains 3921 records referencing physical and biological datasets, model output, museum samples, publications, reports, proposals, atlases, and audio and video programs. The regional distribution of these records is shown in Figure 3.2. Most records are from coastal locations. In terms of metadata density, most contributions pertain to the eastern North Pacific Ocean. We suspect that there has been a similar quantity of research performed in the western North Pacific Ocean, 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,



**Fig. 3.2** Regional distribution of metadata records.

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.

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 3.3. 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. 3.3** 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 online, 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 3.2. 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, temporal domain, source, and constraints. There is also a link to the PICES Metadata Federation, and that link is exercised using the principles and instructions contained in this report.

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 Ocean and bordering regions. Toward this end, NPEM established partnership in an existing federation sponsored by FGDC called the National Spatial Data Infrastructure (NSDI) Clearinghouse. The NSDI Clearinghouse requires metadata to be coded using the FGDC standard, and it uses Isite, an instance of the Z39.50 communication protocol, for queries and exchanges. Late in 2003, we began the necessary work, and by September 2005, NPEM became a registered node of the Clearinghouse.

Also late in 2003, we launched plans to implement with other North Pacific marine data centers (e.g., KODC and JODC (Japan Oceanographic Data Center)) "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.

### 3.2. First federation effort: NPEM and the Korea Oceanographic Data Center

Representatives (Fig. 3.4) of the Korea Oceanographic Data Center (KODC) and the NOAA-PICES NPEM exploited a communications technique allowing public

Internet search of their combined metadata collections in a single session. The approach requires that each metadata provider establish English-language XML (Extensible Markup Language) metadata records in the FGDC standard format. The XML records are served using the Z39.50 communications protocol. Access is through a metadata clearinghouse that supplies search and delivery scripts to the user. Presently, the federation uses FGDC's National Spatial Data Infrastructure Clearinghouse (<http://clearinghouse3.fgdc.gov/>), in which KODC and NPEM each have registered nodes.



**Fig. 3.4** The KODC–NPEM Federation Team outside the National Fisheries Research and Development Institute, Busan, Korea; top row, left to right: Bernard Megrey, Allen Macklin, Dan Klawitter, and Hae-Seok Kang; bottom row, left to right: Joon-Yong Yang, Kyu-Kui Jung, Kimberly Bahl, and Hee-Dong Jeong.

With this understanding of requirements to build a federation in hand, NPEM members approached KODC personnel at the Twelfth Annual PICES meeting in Seoul, Republic of Korea, October 2003, with an invitation to cooperate on a joint federation project. KODC expressed interest in federating with NPEM. Informal communications between parties that year culminated in the submission of a proposal from TCODE to the PICES Science Board at the Thirteenth Annual PICES meeting in Honolulu, U.S.A., the following October. PICES agreed to fund, in part, two meetings of KODC and NPEM principals over the coming year to establish the federation, and to promulgate information to other PICES members about joining the federation.

Using partial support from PICES, KODC and NPEM personnel developed the application over a year, with major progress coming from joint meetings held in Seattle, U.S.A., during August 2005 (Fig. 3.5) and Busan, Republic of Korea, in October 2005 (Fig. 3.6). At the latter meeting, KODC joined NPEM as a registered node of the Clearinghouse. To increase its presence in the Clearinghouse, KODC is expanding the information that it serves through prioritized translation of metadata records from Korean to English and their subsequent conversion to the FGDC standard. To facilitate the conversion of DIF metadata records to FGDC, we obtained Excel Visual Basic routines from NASA (National Aeronautics and Space Administration). The routines require modification to work with KODC's modified Directory Interchange Format (DIF) metadata. A listing of the Seattle and Busan meeting minutes and agendas can be found in Appendix 15.1.



**Fig. 3.5** KODC and NPEM representatives working at the Alaska Fisheries Science Center in Seattle in August 2005; left to right: Allen Macklin, Bernard Megrey, Hae-Seok Kang (mostly hidden by laptop), Kyu-Kui Jung, and Kimberly Bahl.



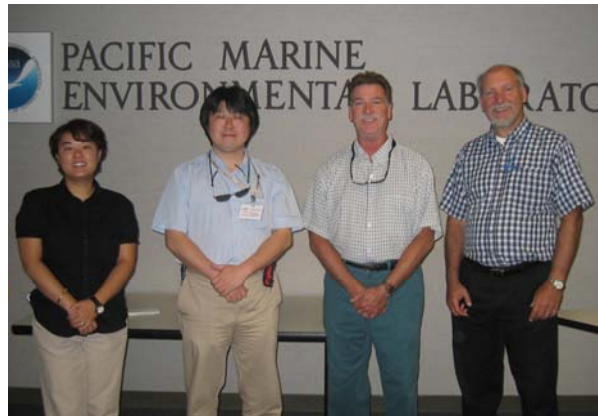
**Fig. 3.6** KODC becomes a registered node of the NSDI Clearinghouse; left to right: Joon-Yong Yang watches on the big screen, Dan Klawitter makes the final keystroke, Hae-Seok Kang talks with the server site in Seoul, and Kyu-Kui Jung looks on.

### 3.3. Continuing effort: Adding Japan's Marine Information Research Center

NPEM personnel recently began working actively with Japan's Marine Information Research Center (MIRC). MIRC provides quality control and value-added product development for the Japan Oceanographic Data Center. The NPEM–MIRC federation is underwritten by FGDC and PICES. A work plan similar to that developed for NPEM–KODC federation is being used.

In August 2006, Dr. Toru Suzuki (MIRC) traveled to Seattle for the first MIRC–NPEM planning meeting (Figs. 3.7 and 3.8). The meeting began with an overview of NPEM, Isite (an application of the Z39.50 protocol) and a history of the NPEM and PICES Federation project. The overview was based on a presentation given at the Fourteenth Annual PICES meeting in Vladivostok, Russia, October 2005. An overview of MIRC's data holding and metadata needs followed. Dr. Suzuki informed participants of the hierarchical structure of MIRC, JODC, and the Japan Hydrographic Association (JHA). He then discussed the varied types of data holdings available through JODC. JODC's data holdings are extremely valuable to scientists working in the North Pacific. They maintain data from several million stations dating back to the early 1800s. JODC Cruise Summary Reports (CSR) provide information for each observational cruise including date/time, research area, abstract, purpose, and contact information. Therefore, the CSR contains much of the core metadata elements that will serve as the basic source of PICES–MIRC metadata base.

The first requirement for federation is to produce FGDC-compliant metadata. Kimberly Bahl, who received training from FGDC last spring, introduced the FGDC metadata content standard and its sections and elements. This gave Dr. Suzuki the rules to write FGDC-compliant metadata records from MIRC information. Ms. Bahl also demonstrated two open-source metadata creation and validation tools, Metavist 2005 and Metadata Parser (MP). These tools allow easy creation of individual metadata records in XML file format (required for any clearinghouse node) and validation that they are FGDC-compliant. Participants used Metavist and MP to create and



**Fig. 3.7** The MIRC–NPEM Federation Team at the Pacific Marine Environmental Laboratory, Seattle, August 2006; left to right: Kimberly Bahl, Toru Suzuki, Allen Macklin, and Bernard Megrey.



**Fig. 3.8** MIRC and NPEM representatives working at the Alaska Fisheries Science Center in Seattle, August 2006; left to right: Bernard Megrey, Kimberly Bahl, Toru Suzuki, and Allen Macklin.

validate an XML metadata record from a JODC CSR.

The second requirement for federation is to supply a common communication protocol: Z39.50. Ms. Bahl provided specific instructions of how to install and configure the Isite application that allows the use of the Z39.50 protocol. The Isite software suite is a free, open-source application available from the FGDC website.

The remainder of the meeting was spent discussing strategies for implementing a Japanese clearinghouse node and dealing with the problems and challenges of locating ongoing funding for the PICES Federation. PICES has been very supportive but has limited resources. At present, funding from within NOAA is unlikely. Despite

numerous efforts, attracting money from international funding organizations has not been successful. There is a possibility that NOWPAP (Northwest Pacific Action Plan) may be able to provide support for a federation. This year, MIRC will request proposals for a three-year project to begin in April 2007. Participants of this meeting will work with Dr Suzuki to develop a MIRC proposal to their funding agency, the Nippon Foundation. The proposal will provide support for ongoing MIRC participation in the PICES federation, primarily through development of a MIRC metadatabase. The meeting ended with a presentation of MIRC plans to build a demonstration site using Isite and the XML record created at the meeting and to register the node at the Clearinghouse.

The second planning meeting was held in Japan in October 2006 in conjunction with the Fifteenth Annual PICES meeting. Meetings were held at the Redbrick Warehouse in Yokohama and continued at the MIRC offices in Tokyo (Figs. 3.9 and 3.10). Mr. Norio Baba of NOWPAP also joined the discussions.

Participants reviewed issues raised at the TCODE meeting which took place a day earlier. These included the advantages of promoting the metadatabase in NOWPAP DINRAC (NOWPAP's Data and Information Network Regional Activity Center). Participants discussed the relationship between PICES TCODE and NOWPAP DINRAC activities and new opportunities for capacity building, and investigating the utility of an Asian-side metadatabase mirror server. Norio Baba said that NOWPAP has worked on metadata capacity building and might be able to invite a specialist from NPEM to collaborate. Dr. Suzuki suggested that representatives from Republic of Korea and Japan may also assist with the DINRAC activity.

Dr. Suzuki introduced the new PICES–MIRC node registered to the NSDI Clearinghouse and reported that Isite had been installed on MIRC's site and registered as 'PICES–MIRC metadatabase' on October 18. He stated that some small problems had been encountered during installation and configuration of the site.



**Fig. 3.9** Group photo at the Marine Information Research Center; left to right: Toru Suzuki, Norio Baba, Kimberly Bahl, Allen Macklin, and Bernard Megrey.



**Fig. 3.10** Working session at MIRC headquarters in Tokyo, Japan, October 2006; left to right: Allen Macklin, Toru Suzuki, Norio Baba, and Bernard Megrey. Kimberly Bahl (photographer, not shown).

Technical issues related to resolving these problems were discussed.

Participants reviewed the progress on the Seattle meeting action plan and amended the plan based on the Japan meeting discussions. Dr. Suzuki planned to submit a proposal by October 23, 2006, to the Nippon Foundation for metadata translation. A listing of the Seattle and Tokyo meeting minutes and agendas can be found in Appendix 15.2.

As the federation grows, the accumulated experience will make it easier for other PICES partners to join. The Pacific Institute of Geography of the Russian Academy of Sciences Far Eastern Branch has already joined the PICES Metadata Federation of its own volition. This report will provide technical guidance for anyone wishing to become a partner.