

Japan joins PICES Marine Metadata Federation

By Bernard A. Megrey, S. Allen Macklin and Toru Suzuki

Summary

Significant progress has been made over the past year to connect PICES member nations' metadatabase systems into one integrated resource. With this new scientific resource, a user of any one metadata inventory will have the ability to search for data catalogued by any and all other system participants with a single search request. Using modern data management techniques to cross-search separate metadatabases provides the advantages of shared metadata without compromising national ownership, data integrity, or security of national metadata products.

The PICES Marine Metadata Federation, originally formed by the North Pacific Ecosystem Metadatabase (NPEM) and the Korean Oceanographic Data Center (KODC), has been expanded to include registered nodes from Korea's National Fisheries Research and Development Institute (NFRDI), Japan's Marine Information Research Center (MIRC) of the Japan Hydrographic Association (JHA) and Russia's TINRO-Center (**Fig. 1**). The Korean nodes now serve more than 700 Korean metadata records; NPEM serves more than 3000 records; the Russian and Japanese nodes serve a small but growing number of records. All nodes have established English-language XML metadata records in Federal Geographic Data Committee (FGDC) standard format and provide those records using the Z39.50 communications protocol. Access is through a metadata clearinghouse that supplies search and delivery scripts to the user (**Fig. 2**). Presently, the federation uses FGDC's National Spatial Data Infrastructure (NSDI) Clearinghouse (<http://www.fgdc.gov/clearinghouse/clearinghouse.html>). The ultimate goal of this project is to federate the marine metadata holdings of all PICES member countries.

The status of the PICES nodes can be found at <http://registry.fgdc.gov/serverstatus/>, and the nodes can be searched by going to <http://clearinghouse3.fgdc.gov/>.

From then until now – The Federation history

In 2002 and 2003, informal discussions between KODC and NPEM took place with the goal to connect the two metadatabase systems into one integrated resource. Drs. Hee-Dong Jeong (KODC/NFRDI) and Hae-Seok Kang (KORDI) agreed to contribute Korean metadata. At the 2004 PICES Annual Meeting, the Technical Committee on Data Exchange (TCODE) adopted a pilot KODC-NPEM Marine Metadata Federation project as part of its 2005 work plan, and the first PICES node (PICES-NPEM) was registered with the U.S. NSDI clearinghouse. Using partial support from PICES, KODC and NPEM, Drs. Hae-Seok Kang and Kyu-Kui Jung, along with NPEM personnel, developed the first Korean node, with major

progress coming from joint meetings held in August (Seattle, U.S.A.) and October 2005 (Busan, Korea). The KODC node (PICES-KODC) came on-line during the Busan meeting. 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. Recently Korea added a second node for NFRDI called PICES-NFRDI KODC. See *PICES Press* 2006, Vol. 14(1): 8–11 for more details.

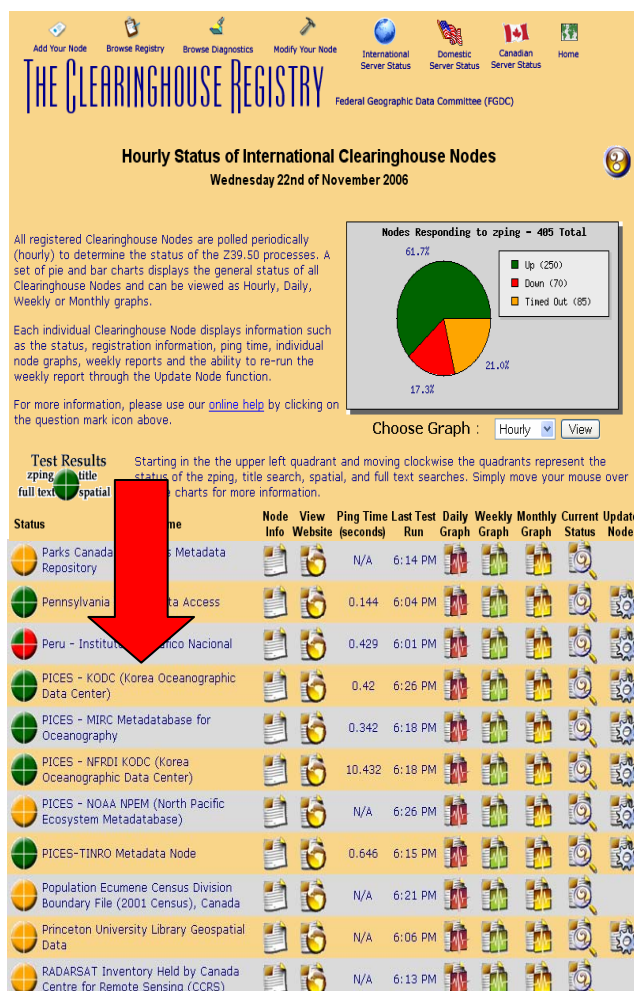


Fig. 1 The arrow points to the current PICES nodes (KODC, MIRC, NFRDI, NPEM and TINRO) of the Clearinghouse Registry. The symbols describe connectivity statistics for all registered sites. At this time, there were 405 sites participating in the clearinghouse (not all shown).

At previous TCODE meetings, Japan expressed an interest to join the Federation. In 2005, in collaboration with Dr. Toru Suzuki (MIRC), TCODE submitted, for partial support by PICES, a Phase II Metadata Federation proposal to expand the Federation to include Japan. Dr. Bernard Megrey and Mr. Allen Macklin also prepared a successful companion proposal ("North Pacific Ecosystem Metadata

Federation: Japan Component”, US\$ 20,000) to the 2006 NSDI Cooperative Agreements Program (CAP). The objective of the NSDI CAP program is to fund innovative projects in the geospatial data community to build the infrastructure necessary to effectively discover, access, share, manage, and use digital geographic data. Monies were used to fund two MIRC/JODC-NPEM workshops that took place in August (Seattle) and October 2006 (Tokyo) and discussed required technical details and technical hurdles, and the means to address and solve problems associated with federating with the Japan Oceanographic Data Center (JODC) and MIRC.

These initiatives take considerable resources. The original KODC-NPEM activity was supported directly at the level of US\$ 4,000 from PICES, US\$ 6,000 from NPEM, and US\$ 6,000 from KODC. Both NPEM and KODC contributed approximately US\$ 35,000 in matching monies. The MIRC-NPEM project was directly funded at the level of US\$ 4,000 from PICES and US\$ 20,000 from NSDI-CAP, with about US\$ 38,700 being contributed in matching monies.

The demonstrated success with these two projects indicates that efforts with other PICES countries should easily scale up with a nominal investment of time and planning.

MIRC-NPEM collaboration

Dr. Suzuki traveled to Seattle in August 2006 for the first MIRC-NPEM planning meeting (Figs. 3–4). 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 2005 PICES Annual Meeting in Vladivostok. An overview of MIRC’s data holding and metadata needs followed. Dr. Suzuki informed participants of the hierarchical structure of MIRC, JODC and JHA. He then reviewed 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 (CSRs) provide information for each observational cruise including date/time, research area, abstract, purpose, and contact information. Therefore, the CSRs contain much of the core metadata elements that will serve as the basic source of the PICES-MIRC metadata base.

The first requirement for the 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).

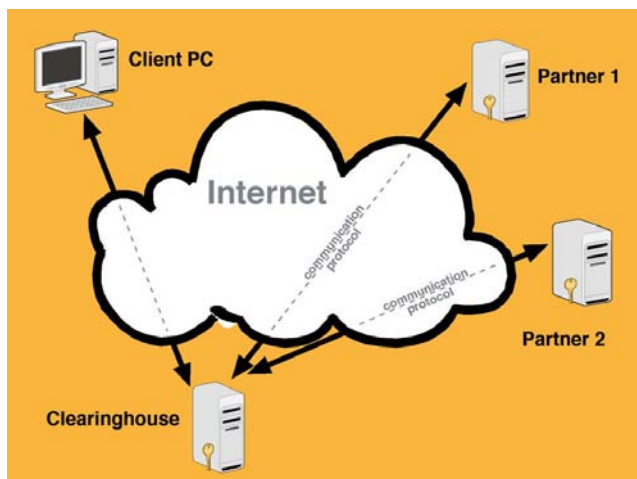


Fig. 2 An Internet metadata federation requires a clearinghouse and partners sharing a communication protocol and serving metadata using the same standard.

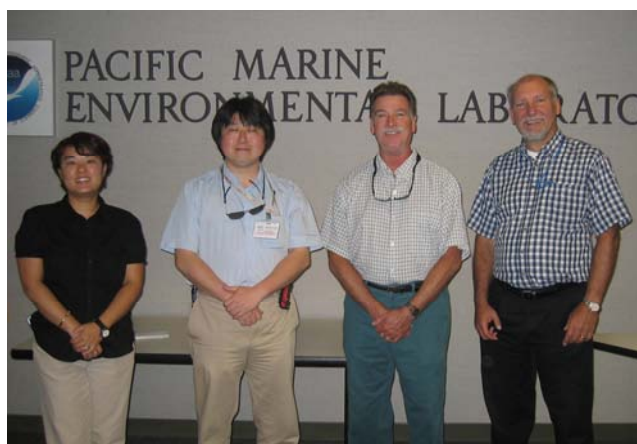


Fig. 3 The MIRC-NPEM Federation Team inside the Pacific Marine Environmental Laboratory, (August 2006, Seattle); left to right: Kimberly Bahl, Toru Suzuki, Allen Macklin and Bernard Megrey.



Fig. 4 MIRC and NPEM representatives working at the Alaska Fisheries Science Center (August 2006, Seattle); left to right: Bernard Megrey, Kimberly Bahl, Toru Suzuki and Allen Macklin.

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 validate an XML metadata record from a JODC CSR.

The second requirement for the Federation is to supply a common communication protocol: Z39.50. Ms. Bahl provided specific instructions on 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 on-going 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. This year, MIRC will submit proposals for three-year projects to the Nippon Foundation to begin in April 2007. Participants of this meeting will work with Dr. Suzuki to develop a MIRC proposal. The proposal will provide support for on-going MIRC participation in the PICES Federation, primarily through development of a MIRC metadatabase. The meeting ended with presentation of MIRC plans to develop a demonstration site using Isite and the XML record created at this meeting, and to register the node at the clearinghouse.

Japanese node comes on-line

The second planning meeting was held in Japan in October 2006, in conjunction with the PICES Fifteenth Annual Meeting. It started at the Red Brick Warehouse in Yokohama and continued at the MIRC office in Tokyo (Fig. 5). Mr. Norio Baba of NOWPAP also joined the discussions.

Participants reviewed the advantages of promoting the metadatabase in NOWPAP DINRAC (Data and Information Network Activity Center), discussed the relationship between TCODE and DINRAC activities and new opportunities for capacity building, and investigated the utility of an Asian-side metadatabase mirror server. Mr. 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 Korea and Japan may also assist the DINRAC activity.

Dr. Suzuki reported that Isite had been installed on MIRC's server, and the site was registered with the NSDI Clearinghouse as "PICES-MIRC metadatabase" on October 18. He reported that some small problems were encountered during the installation and configuration of the site. 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 submitted the aforementioned proposal to the Nippon Foundation for metadata translation on October 23, 2006.



Fig. 5 Top: Working session at MIRC headquarters (October 2006, Tokyo); left to right: Allen Macklin, Toru Suzuki, Norio Baba and Bernard Megrey. Bottom: Group photo; left to right: Toru Suzuki, Norio Baba, Kimberly Bahl, Allen Macklin and Bernard Megrey.

Emerging standards

The standards upon which the clearinghouse functionality relies are in a state of change. The present NSDI Clearinghouse legacy search gateway will soon be replaced. This legacy gateway is built with proprietary software that is no longer maintained and supported by the vendor, Blue Angel Software. The legacy interface eventually will be replaced with GeoNetwork, a user-maintained, open-source solution with similar and enhanced capabilities compared to the legacy interface. The GeoNetwork gateway (a beta version of the new search gateway is in development) will be implemented in six months to a year. Both the legacy interface and the new GeoNetwork interface rely on the proven Z39.50 communication protocol.

The final change concerns the way metadata are described. The old method used FGDC standards. While proven and well known, this standard has difficulty with biological data. A new international metadata standard, ISO 19115, is emerging, which was built to be compatible with FGDC and to address the deficiencies in describing biological data. Translators that convert from FGDC to ISO 19115 should be easily available.

These and probably other changes will lead to modifications in PICES Metadata Clearinghouse interface standards and will require changes for existing and future clearinghouse servers.

Future work

A new PICES Technical Report “Federation of PICES member countries metadata” has recently been uploaded to the PICES web site in PDF format. This report provides up-to-date information on the emerging changes expected to take place in the clearinghouse interface, as well as specific technical guidance and instructions for anyone wishing to become a partner.

Future plans include the participation of PICES member countries not already federated (*i.e.*, Canada and China) into a PICES Metadata Federation. Such activity may be supported through future 2007 NSDI-CAP funding possibilities. At the 2006 TCODE meeting, a request for Phase III funding (\$4K) to bring China or Canada into the Federation (pending successful funding application elsewhere) was submitted.

With the move of the U.S. Metadata Clearinghouse interface from proprietary to open-source software, the potential exists for PICES to adopt the open-source standard, federate its metadata internally and sever its direct relationship with the U.S. Clearinghouse nodes. Related costs, consequences, benefits and recommendations were discussed at the recent TCODE meeting.

[*This research article is contribution no. 3032 from the Pacific Marine Environmental Laboratory and contribution EcoFOCI-0625 to NOAA’s Ecosystems & Fisheries-Oceanography Coordinated Investigations.*]



Dr. Bernard Megrey (bern.megrey@noaa.gov) co-directs the North Pacific Ecosystem Metadatabase and co-chairs the Data Management and Communications Committee for the Alaska Ocean Observing System. 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 NOAA’s Fisheries-Oceanography Coordinated Investigations (FOCI), he has nearly 25 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. Bernard is a member of the PICES TCODE Committee and MODEL Task Team.

Mr. Allen Macklin (allen.macklin@noaa.gov) also co-directs the North Pacific Ecosystem Metadatabase and co-chairs the Data Management and Communications Committee for the Alaska Ocean Observing System. Allen is a Meteorologist with the Pacific Marine Environmental Laboratory. He is Coordinator for Ecosystems & Fisheries-Oceanography Coordinated Investigations (EcoFOCI), a NOAA research program to sustain fishery resources in the Gulf of Alaska and Bering Sea while maintaining healthy ecosystems. Allen has 30 years’ experience studying Alaskan marine ecosystems and managing associated scientific information.

Dr. Toru Suzuki (suzuki@mirc.jha.jp) is General Manager of the Research Division at the Marine Information Research Center (MIRC), Japan Hydrographic Association. His scientific background is physical oceanography in coastal regions, and his current work includes integration of oceanographic data and information, and the development of quality-control procedures. Toru is a member of the PICES Section on Carbon and Climate and serves on the IMBER Data Management Committee.