ICES CM 2008/R:06

Pacific-wide marine metadata discovery, management and delivery: The PICES Metadata Federation

Bernard A. Megrey and S. Allen Macklin

The member countries of the North Pacific Marine Science Organization (PICES) separately maintain vast quantities of marine ecosystem data. To support detection and prediction of ecosystem change in the North Pacific Ocean, it is beneficial to discover data holdings with a single search, rather than having to access each country's records, perhaps stored in different languages and formats. We report on the creation of a PICES "metadata federation" of member countries (Canada, People's Republic of China, Japan, Republic of Korea, Russian Federation, and the United States of America). Through (1) English-language coding of metadata using the Federal Geographic Data Committee standard; (2) acquisition, installation and configuration of ANSI Z39.50-1995 (ISO 10163-1995) open-source communications software on a public-access server; and (3) registration with a clearinghouse, it is possible for any metadata-serving agency to become part of the PICES Metadata Federation. The federation enables an Internet user to search the collected metadata holdings of any or all members, thus providing access to information across national holdings in a single search. To date, metadata collections from Japan, the Russian Federation, Republic of Korea, USA and China are federated. This activity supports PICES' goals to promote and co-ordinate marine scientific research in the northern North Pacific and adjacent marginal seas; to advance scientific knowledge about the ocean environment, global weather and climate change, living resources and their ecosystems, and the impact of human activities on them; and to promote the collection and rapid exchange of scientific information on these issues.

Keywords: metadata, data sharing, ecosystems, informatics

Contact author Bernard A. Megrey: National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA [tel: +1 206 526 4147, fax: +1 206 526 6723, e-mail: bern.megrey@noaa.gov].

1. Introduction

1.1. Data sharing, international boundaries and large marine ecosystems

Marine science and marine management organizations have a need to know, first-hand and quickly, pertinent details of marine ecosystems to facilitate planning and management. The North Pacific Marine Science Organization (PICES), for example, produces North Pacific Ecosystem Status Report (PICES, 2004) review and summarize the status and trends of the marine ecosystems in the North Pacific, and consider the factors that are causing, or are expected to cause, change in the near future. The first report served as a pilot project. This report was based largely on geographic locations and subjects for which time series data or information were readily available. The report also identified locations and subjects where data were collected but were not yet available.

For the most part, marine ecosystem data are contained in archives of various national ocean data centers, academic institutions and other agencies. Some of these bodies serve some or all of their

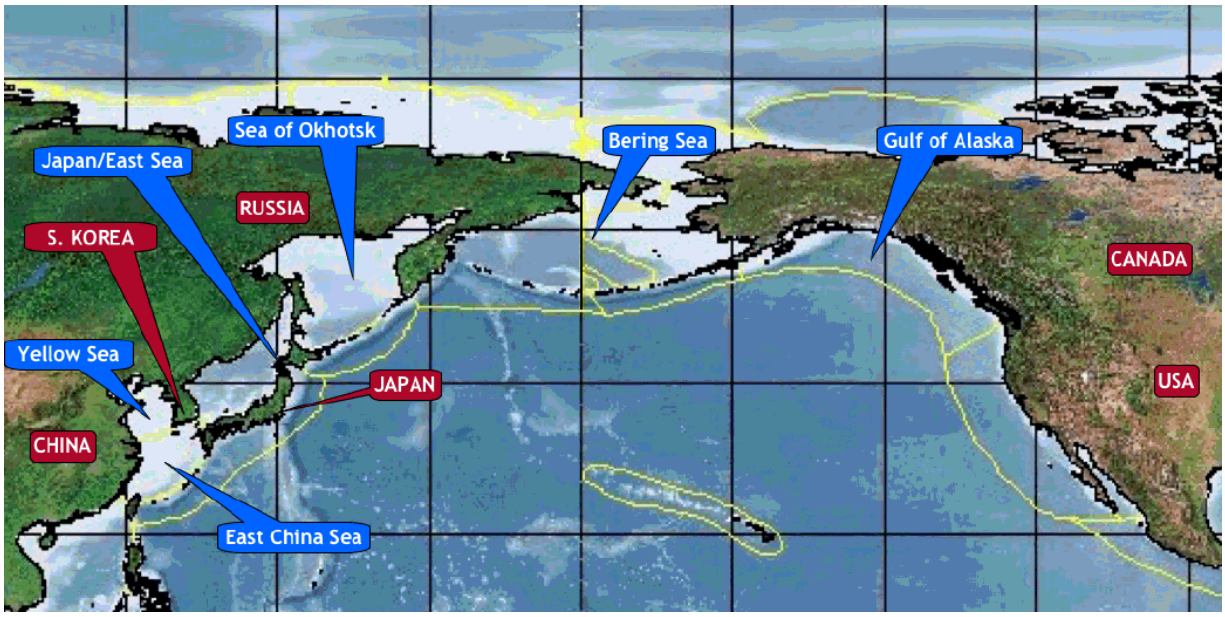

Fig. 1.1 The North Pacific Ocean and its marginal seas (blue labels), PICES member nations (red labels), and Large Marine Ecosystems (yellow boundaries).

archived information through the World-Wide Web. Many data, however, remain hidden from public search and use, especially when countries are not obligated by charter to share their data archives. Even discovering the data that are available on line can be an onerous task. For example, obtaining information on the East China Sea requires access, at least, to marine data of the bordering nations: People's Republic of China, Republic of Korea and Japan. The situation is repeated for many other marginal seas of the North Pacific Ocean.

More and more, ecosystem management focuses on Large Marine Ecosystems (LMEs, Fig. 1.1). LMEs are regions of the ocean encompassing coastal areas from river basins and estuaries to the seaward boundaries of continental shelves and the outer margins of the major current systems. They are relatively large regions on the order of 200,000 km^2 or greater, characterized by distinct bathymetry, hydrography, productivity, and trophically dependent marine populations. On a global scale the 64 LMEs produce 95% of the world's annual marine fishery biomass. Within LME waters, most of the global ocean pollution, overexploitation, and coastal habitat alteration occur. Yet, vital information on these processes in not centrally available, nor even easily available, for public scrutiny, let alone international management efforts. Often, shared ecosystems span national boundaries, have overlapping Exclusive Economic Zones (EEZ) and can produce conflicts due to national interests and resource management disagreements. For example, the Japan/East Sea (Fig. 1.1) ecosystem includes the EEZ's Japan, China, Russia and Korea. A method is needed for sharing national information about marine ecosystems, independent of political boundaries.

The PICES Technical Committee on Data Exchange (TCODE) is working to address this problem. TCODE has among its terms of reference: (1) *to identify the data management requirements of PICES* and (2) *to develop strategic plans to meet these requirements*. Part of TCODE's working plan for the last several years has been the PICES Metadata Federation project. This project attempts to alleviate the shortcomings of international ecosystem data sharing by creating a "federation" of marine ecosystem data servers using standard data cataloging and communication protocols. This technology is not new, but its application is overdue in the field of marine ecosystem management. Similar technology to collect disparate sources of information have been in existence for many years, used mainly by library and museum systems. Several examples of scientific clearinghouses exist and the method described in this paper makes use of such a clearinghouse.

1.2. Data sharing, international boundaries and large marine ecosystems

The objectives of the PICES Metadata Federation project are to create standardized metadata descriptions of national, institutional and agency databases and to serve those descriptions in a World-Wide Web-based, one-stop environment with search and delivery capability. Such a federated or distributed system has many benefits.

Federation is a process of joining for mutual benefit. For example, suppose Provider 1 produces bread, and Provider 2 produces butter. In a non-federated system, a consumer wanting bread would have to get it from Provider 1, and butter would only be available from Provider 2. However, if Providers 1 and 2 are willing and able to cooperate, each provider can maximize the distribution of its own product by also offering it through the other provider. This is a federation.

The federation promotes efficiency for the provider and the consumer. Each provider effectively boosts its product line by having available more products without actually having to produce them. The consumer benefits by being able to locate more products without having to know more providers.

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. That solution was to federate with the providers of the new metadata.

2. Implementation

2.1. Federation model

A recommended model for federation was that used by the National Geospatial Data Clearinghouse (http://clearing house3.fgdc.gov/). The approach requires that each metadata provider establish English-language XML (Extensible Markup Language) metadata records in the Federal Geographic Data Committee's (FGDC) standard format (Federal Geographic Data Committee, 1995). The XML records are served using the Z39.50 communications protocol (National Information Standards Organization, 2002) which is a client-server protocol for searching and retrieving information from remote computer databases. Z39.50 specifies

procedures and formats for a client to search a database provided by a server, retrieve database records, and perform related information retrieval functions. The protocol addresses communication between information retrieval applications at the client and server; it does not address interaction between the client and the end-user. Access is through a metadata clearinghouse that supplies search and delivery scripts to the user (Fig. 2.1). All software is open-source and freely available.

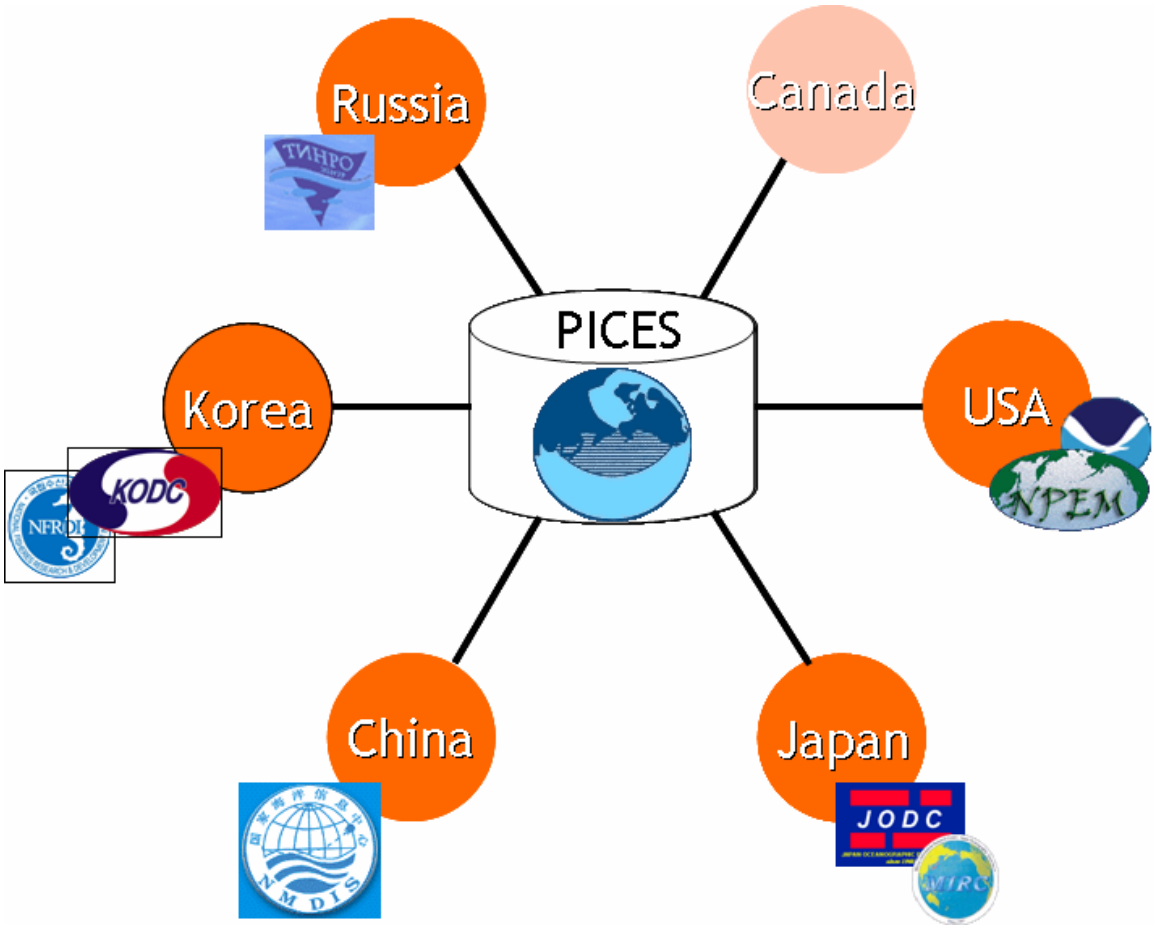

Fig. 2.1 The North Pacific Ocean and its marginal seas (blue labels), PICES member nations (red labels), and Large Marine Ecosystems (yellow boundaries).

2.2. Building the federation

With this understanding of requirements to build a federation in hand, NPEM members approached Korea Oceanographic Data Center (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.

During summer and fall 2005, representatives of KODC and the NOAA–PICES NPEM exploited this communications technique allowing public Internet search of their combined metadata collections in a single session, with the result that KODC and NPEM each registered nodes.

NPEM personnel then began working with Japan's Marine Information Research Center (MIRC) in 2006. MIRC provides quality control and value-added product development for the Japan Oceanographic Data Center (JODC). The NPEM–MIRC federation was underwritten by National Spatial Data Infrastructure-Cooperative Agreements Program (NSDI-CAP) FGDC and PICES. A work plan similar to that developed for NPEM–KODC federation was used, and Japan joined the federation in October 2006.

Meanwhile, scientists at Russia's Pacific Research Institute of Fisheries and Oceanography (TINRO) followed the lead and established a node in the federation of their own volition. Another proposal was written to the NSDI-CAP program to federate the National Marine Data and Information Service (NMDIS) of the State Oceanic Administration of China. NMDIS maintains and develops the national marine database resulting from China's marine observations establishments. After two workshops NMDIS joined the Federation in 2008.

Workshops were held with each partner where metadata training took place as well as technical instructions on how to install and register a clearinghouse node. As part of the project, a PICES Technical Report (Megrey et al, 2007), downloadable from the PICES web site in PDF format (http://www.pices.int/publications/tech_reports/tech_rep_1/Federation_Report_13April07_Final.pdf), was prepared to help assist future partners. 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.

3. Emerging Standards

The standards upon which the clearinghouse functionality relies are in a state of change, with new standards emerging. The present NSDI Clearinghouse legacy search gateway will soon be replaced. The legacy search gateway is built with propriety software that will no longer be maintained and supported by the vendor, Blue Angel Software. The legacy interface will be replaced with GeoNetwork (http://geonetwork-opensource.org/), a user-maintained and open-source solution with similar and enhanced capabilities compared to the legacy interface. In

combination with the opensource GeoServer (http://geoserver.org), GeoNetwork provides portal services that actually implement a metadata, data, information and services clearinghouse, data services components to access spatial content and allow data processing, and portal service to process geospatial information and prepare it for presentation to the user. The GeoNetwork gateway will soon be implemented. Both the legacy interface and the new GeoNetwork interface rely on the proven Z39.50 communication protocol.

Recently, PICES rented a server to consolidate all metadata clearinghouse nodes onto one machine (Fig. 3.1). Some national laboratories were not able to overcome the security issues involved with opening a server port directly to the internet without the benefit of firewall protection. Some nations have a node within their labs and a mirror node registered on the rented server (i.e. TINRO and NPEM). Future plans include the participation of PICES member countries not already federated (Canada) into a PICES Metadata Federation. With the move of the US metadata clearinghouse interface from proprietary to open-source software, PICES is planning to adopt the open-source standard, federate their metadata internally and sever its direct relationship with the USA Clearinghouse nodes.

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 is emerging, ISO 19115, 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 will 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.

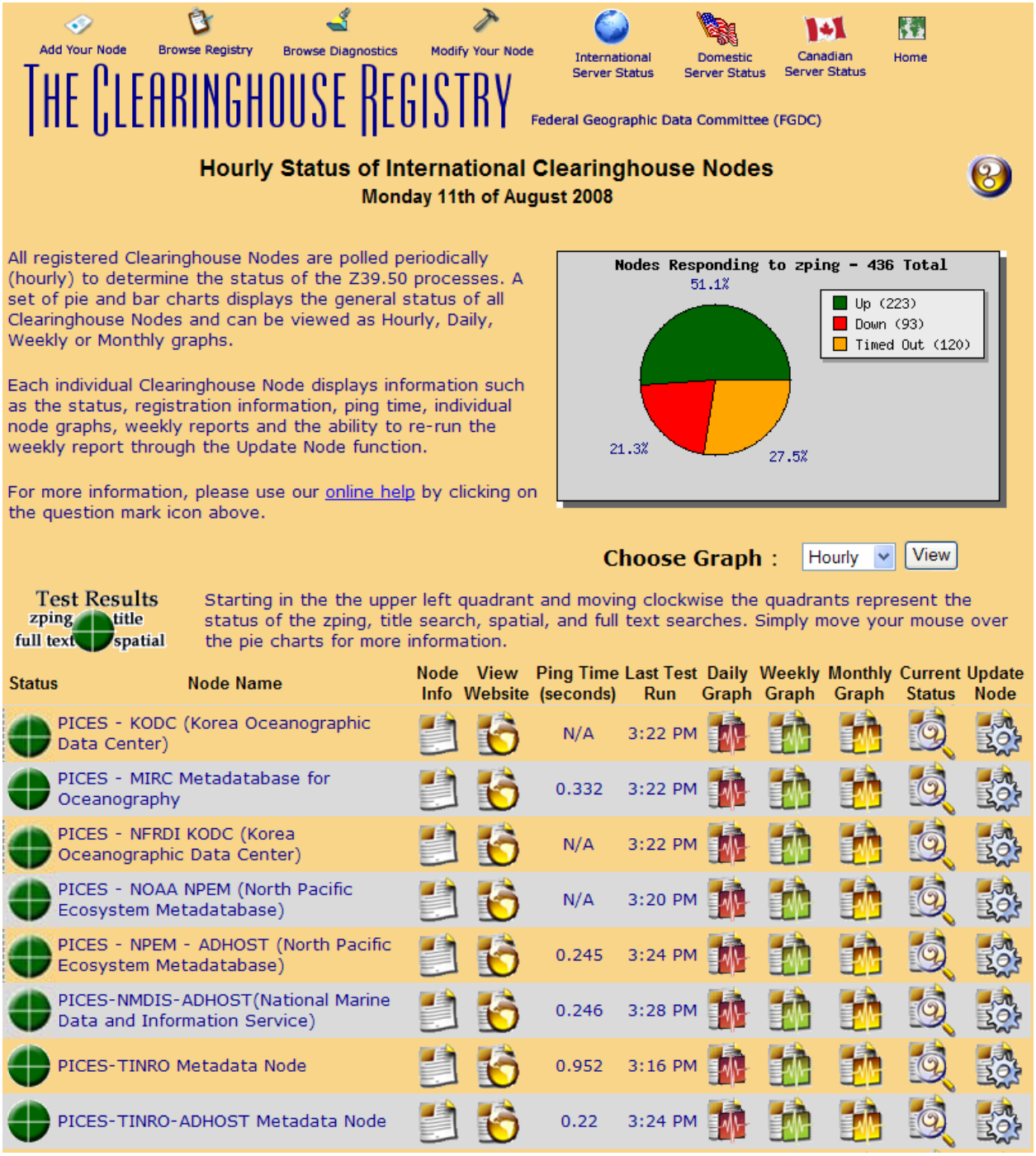

Fig. 3.1 A listing of PICES Metadata federation nodes registered at the NSDI Clearinghouse.

4. Summary

Significant progress has been made 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 Metadata Federation has been expanded to include registered nodes from the Korea's Korean Oceaographic Data Center (KODC) and National Fisheries Research and Development Institute (NFRDI), Japan's Marine Information Research Center (MIRC), Russia's TINRO laboratory, and China's National Marine Data and Information Service (NMDIS). (Fig 2.1). The Korean nodes now serve more than 700 Korean metadata records; NPEM serves more than 3000; the Russian, Japanese and Chinese 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 serve those records 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://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/.

References

Federal Geographic Data Committee. 1995. Content standards for digital geospatial metadata workbook (March 24). Federal Geographic Data Committee, Washington, D.C.

Megrey, B.A., Macklin, S.A., Bahl, K., and Klawitter, D. (eds.). 2007. Metadata federation of PICES Member Countries. PICES Technical Report No. 1.

National Information Standards Organization. 2002. Information retrieval (Z39.50): Application Service Definition and Protocol Specification : National Information Standards Organization. ISSN 1041-5653. http://www.loc.gov/z3950/agency/Z39-50-2003.pdf (viewed 13 August 2008).

PICES. 2004. Marine Ecosystems of the North Pacific. PICES Special Publication Number 1. 278 pp.