Korea and U.S. federate marine metadata collections

By S. Allen Macklin, Bernard A. Megrey, Kyu-Kui Jung and Hae-Seok Kang

Summary

Representatives (Fig. 1) of the Korea Oceanographic Data Center (KODC) and the NOAA-PICES North Pacific Ecosystem Metadatabase (NPEM) have exploited a communications technique that allows 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 (Federal Geographic Data Committee) 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://www.fgdc.gov/ clearinghouse/clearinghouse.html), in which KODC and NPEM each have registered nodes.

Fig. 1 The KODC-NPEM Federation Team outside the National Fisheries Research and Development Institute, Busan, Korea; Back row, left to right: Bernard Megrey, Allen Macklin, Dan Klawitter and Hae-Seok Kang; front row, left to right: Joon-Yong Yang, Kyu-Kui Jung, Kimberly Bahl and Hee-Dong Jeong.

Using partial support from PICES, KODC and NPEM personnel developed the application over the past year, with major progress coming from joint meetings held in Seattle in August 2005, and Busan in October 2005. 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. The greater intention of this project is to federate the marine metadata holdings of all PICES member countries. Japan and Russia have expressed interest in joining the PICES federation.

What is federation?

Federation is a process of joining for mutual benefit. For example, suppose Provider 1 produces Product A, and Provider 2 produces Product B. In a non-federated system, a consumer wanting A would have to get it from 1, and B would only be available from 2. However, if 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.

This 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.

Metadata

In this case, the product is metadata. Metadata, or data about data, describe the content, quality, condition, and other characteristics of data. When seeking a certain kind of data, a scientist will examine a broad body of metadata, eliminating those records that do not satisfy the search criteria. For example, when searching for vertical profiles of ocean properties obtained from hydrographic casts, a scientist might specify the locations and times of the casts, the inclusive depths and the variables measured. The successful search will reiterate these parameters and tell the scientist the location of the data, and how those data can be obtained. 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 principal investigator, species association and study hypothesis.
- Semantic metadata describe 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 are described in a common set of terminology, definitions and format that are the metadata standard. The metadata standard determines what thematic, semantic and syntactic descriptors are catalogued and how they are presented. There are a number of metadata standards, *e.g.*, Dublin Core, Ecosystem Metadata Language, Federal Geographic Data Committee (FGDC), Directory Interchange Format (DIF). KODC uses the DIF standard to catalog data, while NPEM uses FGDC.

Method

To enable an Internet browsing client to search and discover information through a federated metadatabase, four elements must be in place. First, the client must be able to communicate with the clearinghouse using normal web communications. Second, the clearinghouse must offer an Internet communication protocol and associated utilities that permit search and discovery of federated metadata records. Third, each federated partner must maintain metadata records using a standard that is supported by the communication protocol. Fourth, each federated partner must serve the Internet through a server that is conversant in the communication protocol and that hosts metadata records in the proper standard. Figure 2 shows these elements.

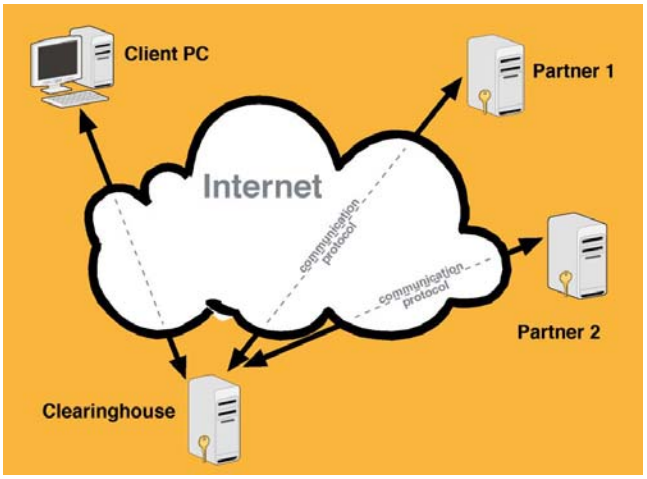


Fig. 2 An Internet metadata federation requires a clearinghouse and partners sharing a communication protocol and serving metadata using the same standard.

Approach

With this understanding of requirements to build a federation in hand, NPEM members approached KODC personnel at the PICES Twelfth Annual Meeting in Seoul, in 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 PICES Thirteenth Annual Meeting in Honolulu 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 member countries about joining the federation.

Also during this exploratory period, NPEM was working to establish 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. Although NPEM was already using the FGDC standard and the installation of ISite was relatively straightforward, security issues slowed progress so much that NPEM was not able to demonstrate this federation feature at the first joint KODC-NPEM federation meeting in Seattle, in August 2005.

At that meeting (Fig. 3), principals presented details of NPEM operations and a tutorial on ISite and its installation. During the discussions, it was revealed that principals do not know the metadata standards used by other PICES-member oceanographic data centers. Perhaps the FGDC/ISite model is not the proper one for the greater PICES federation. Are there web-resident clearinghouses that use different approaches than NSDI? Are there morphing routines available that translate from one metadata standard to another? These were some of the issues that needed clarification before the next meeting in Busan, which was scheduled to occur following the PICES Fourteenth Annual Meeting in Vladivostok. In addition, KODC would translate some key metadata records to English and format them using the FGDC standard. The goal was to demonstrate an NPEM-KODC federation through the NSDI Clearinghouse by the end of the year.

Fig. 3 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.

Following the August meeting, NPEM was able to solve its server security problems and became a registered node of the NSDI Clearinghouse. This utility was demonstrated at

the PICES Fourteenth Annual Meeting in Vladivostok. At this meeting, too, TCODE representatives were instructed to ascertain the metadata standard(s) used by their respective ocean data centers and to query their centers about interest in joining the PICES federation. Several PICES representatives stated that their countries would be interested in joining the federation.

At the October 2005 KODC-NPEM meeting in Busan, the team first reviewed the function and holdings of KODC. NPEM group reported that they had been unable to locate any cost-free, public clearinghouses other than those that use FGDC and Z39.50. Morphing applications from DIF to FGDC are available through NASA. KODC had translated some metadata records to English and was ready to serve them as XML files in FGDC format. With a bit of tweaking of the ISite software and some calls to the server site in Seoul, KODC became a registered node of the NSDI Clearinghouse (http://www.fgdc.gov/clearinghouse/, Figs. 4 and 5).

Fig. 4 KODC becomes a registered node of the NSDI Clearinghouse; left to right: Joon-Yong Yang watches on the big screen, Dan Klawitter just made the final keystroke, Hae-Seok Kang talks with the server site in Seoul, and Kyu-Kui Jung looks on.

Future work

With the initial work completed to become registered nodes of the NSDI Clearinghouse, the focus for the KODC-NPEM federation team is now on increasing the number of metadata records offered by KODC in the English language and adhering to the FGDC format. Because these exercises will require funds beyond the budgets of the two sponsoring groups, the principals will develop proposals to be submitted to agencies and institutions that look favorably on these kinds of projects. To facilitate the conversion of DIF metadata records to FGDC, we obtained Excel VisualBasic routines from NASA. The routines are being examined to determine what modification they might need for the KODC-NPEM work. At some point, it may become desirable for PICES to host its own clearinghouse separate from FGDC NSDI.

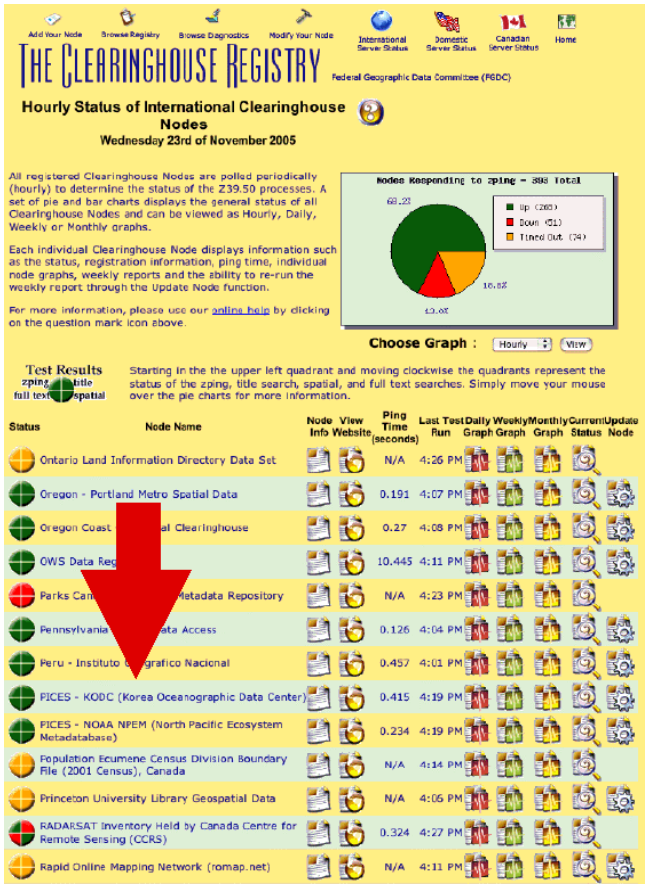
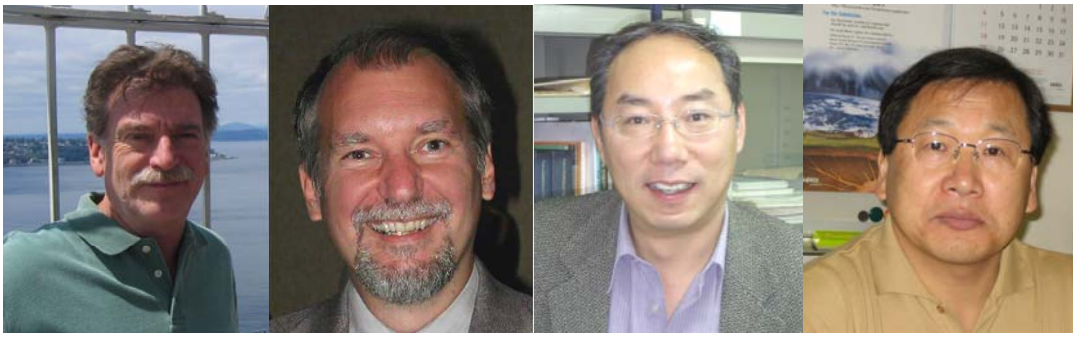


Fig. 5 The arrow points to the current PICES nodes (KODC and NPEM) of the Clearinghouse registry. The symbols describe connectivity statistics for all registered sites. At this time, there were 393 sites participating in the clearinghouse (not all shown).

As the federation grows, the accumulated experience will make it easier for other PICES partners to join. Personnel from the Japan Oceanographic Data Center, the Japan Marine Information Research Center and the Pacific Institute of Geography of the Far Eastern Branch of the Russian Academy of Sciences soon will join the federation process. To aid the ongoing effort, proposals were submitted to the Sasakawa Peace Foundation and to PICES. We encourage interested parties to step forward with further ideas for funding and federation. We anticipate a full PICES federation by the end of the decade.

[*This article is contribution 2882 to NOAA's Pacific Marine Environmental Laboratory and contribution FOCI-0580 to NOAA's Fisheries-Oceanography Coordinated Investigations.*]



Mr. Allen Macklin (allen.macklin@noaa.gov) 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 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 30 years' experience studying Alaskan marine ecosystems and managing information.

Dr. Bernard Megrey (bern.megrey@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. Bernard is a research fisheries biologist with NOAA's Alaska Fisheries Scence Center where he has worked since 1982. As lead investigator for recruitment modeling studies of FOCI, he has nearly 25 years' experience studying the dynamics of exploited North Pacific fish populations, relationships of environment and recruitment variability, and application of computer technology to fisheries research and natural resource management. Bernard is a member of the PICES Technical Committee on Data Exchange (TCODE) and the PICES MODEL Task Team.

Dr. Kyu-Kui Jung (kkjung@nfrdi.re.kr) co-directs the Korea Oceanographic Data and Information Service (KODIS), a metadata system, and manages the Korean Delayed Mode Data Base of the North-East Asian Regional GOOS. Kyu-Kui is a research paleontologist and marine geologist with the National Fisheries Research and Development Institute (NFRDI). As lead investigator for the Korea Oceanographic Data Center (KODC) operated by NFRDI, he has wide experience studying the long-term evolution of marine ecosystems and managing oceanographic data and metadata. Kyu-Kui is a member of TCODE.

Mr. Hae-Seok Kang (hskang@kordi.re.kr) has co-directed the development of the Korea Oceanographic Data and Information Service System for seven years since 1999. Hae-Seok is a manager of the Ocean Data and Information Division of the Korea Ocean Research and Development Institute (KORDI). He has more than 20 years' experience designing and implementing oceanographic data and information management systems. Hae-Seok also serves as a member of TCODE.

PICES Interns

PICES offers sincere thanks to **Mr. Jin-Yong Lee** (Korean Ocean Research and Development Institute), the 2005 PICES intern, who completed his term at the Secretariat at the end of June, and has returned to Korea. We are very grateful for his dedicated work during this past year.

We are pleased to announce that **Mr. Pavel Vorobyov** from the Pacific Scientific Research Fisheries Center (TINRO-Center), Vladivostok, Russia) joined the Secretariat in early January as the Seventh PICES Intern. Pavel helped us out at PICES XIV in Vladivostok last year, and you will have an opportunity to meet him this year at the Symposium on "*Climate variability and ecosystem impacts on the North Pacific: A basin-scale synthesis*" in Honolulu or at the PICES Secretariat office.