

REPORT OF TECHNICAL COMMITTEE ON DATA EXCHANGE



The meeting of the Technical Committee on Data Exchange (hereafter TCODE) was held from 16:00–19:15 hours on October 31, 2007. The Chairman, Dr. Igor I. Shevchenko, called the meeting to order and welcomed the participants. The meeting was attended by 11 TCODE members and 4 observers representing PICES member countries and international organizations (*TCODE Endnote 1*). Dr. John Holmes served as the rapporteur. The Committee reviewed the provisional agenda and adopted it without changes and additions (*TCODE Endnote 2*).

Review progress on items in the 2006/2007 work plan (Agenda Item 3)

PICES Federated Metadata Searching Project

Progress continued to be made with the PICES Federated Metadata Searching Project. All PICES member countries, except Canada, have clearinghouse nodes and are contributing metadata. The latest country brought on-line was China. In August 2007, a node was established at the National Marine Data and Information Service of the State Oceanographic Administration in Tianjin. A training course on FGDC (Federal Geographic Data Committee) metadata preparation was completed successfully. A suggestion was put forward and supported that an objective of the Federated Metadata Searching Project for the next year should be to report on how the remote server service performs. A progress report for 2007 will be provided by Dr. Bernard A. Megrey to the Secretariat for posting on the PICES website.

Canadian participation in the project is probably about a year away, and contributions will be through Canadian efforts to catalogue internal data holdings in a searchable database. The delay is related to the extensive IT reorganization that is currently occurring and the need for time to make necessary software improvements to the existing internal catalogue. Dr. Holmes will report back

to TCODE on progress in Canada with respect to metadata records and how to identify parent records, *i.e.*, prevent duplication in multiple nodes or catalogues.

A proposal to purchase a remote server service for one year to consolidate all PICES metadata nodes was discussed. TINRO-Center has used GeoNetwork and GeoServer software, and a short presentation on this topic was given by Dr. Shevchenko. The software is written in Java and so can be used in Windows, Linux, and Unix environments. A suggestion was to install GeoNetwork and GeoServer on a consolidation server and test them in this virtual server environment. Dr. Megrey was asked to refocus the proposals to reflect discussions at TCODE and consider making them a joint proposal with MONITOR (*TCODE Endnote 3*).

Dr. Shevchenko also reported that in Russia they cannot connect to servers of the several metadata clearinghouses. An experiment was discussed on accessing clearinghouses from different locations to see if this is true for other countries.

Update TCODE Action Plan

Several minor changes to the TCODE Action Plan were discussed and approved. The Plan was updated and posted on the TCODE website (<http://tcode.tinro.ru/pices16.html>).

National annual reports (Agenda Item 4)

TCODE members presented national annual reports. These reports include lists of institutes and agencies, key persons contacts, links to data and metadata sets, ocean observing systems, data and metadata formats and standards, information technologies for collecting, measuring and enumerating marine organisms, marine data management programs that underpin marine science programs, data policies; software applicable in marine ecosystems studies and

TCODE-2007

modeling, publications on marine data management issues, education materials, *etc.* All reports will be posted at the TCODE website.

Canada:

One addition to the last year's report is the website for the National Science Data Management Committee (http://intradev.ncr.dfo-mpo.ca/science/nsdmc/index_e.htm). The goal of the Committee is to guide data management strategies and priorities. This activity includes items of direct interest to TCODE, such as data and metadata inventories, data and metadata standards, data access, and data archaeology.

Canada has established new funding for Cable Underwater Observatories through the academic funding system (not through government departments). There is both a coastal project (VENUS) and an offshore observatory (NEPTUNE). These activities will require a high performance Data Management and Archival System (DMAS) to handle the large amounts of data that may be produced by these systems. The NEPTUNE project has recruited an Assistant Director (Information Systems) to head up this activity. The candidate comes with a very strong background in astronomical data management, so there may be some interesting opportunities for "cross-fertilization" of ideas, approaches and techniques. VENUS is now "live" at the Patricia Bay node, with real-time data on-line at www.venus.uvic.ca/data/data_plots.html. Another node is being placed off Roberts Bank in the Strait of Georgia now and another ship is laying cable out to Endeavour Ridge hot vents.

Japan

The Fishery Agency data management activities are focused on the Japan Sea. One goal is to put together a modern database and make it searchable. The database implementation is behind schedule due to the need to bring in some older data. It is expected to be finished in the next year.

Three monitoring lines (off Hokkaido, near Tokyo, and in the East China Sea) are used to

monitor the lower trophic levels every season, and the collected data are going into a database. These observations will continue until 2010.

There is no a generally accepted approach to the metadata bases design. Many old metadata records are in Japanese only. They should be translated into English, and this is very time-consuming.

China

All activities to join PICES Metadata Federation were completed. The software tools to support the metadata production are being tested. The main difficulty is to arrange translating metadata from Chinese to English.

Korea

Several agencies operate ocean databases and the Argo site and are involved in distribution of the data. A real-time coastal information system has been developed for monitoring aquaculture environments. The system is deployed on fish farms. Measured data (temperature, salinity and dissolved oxygen) play a crucial role in nowcasting/forecasting coastal ocean conditions and in reducing the risk of mass mortality caused by an abnormal change of the water quality. The Korean Ocean Data Center is developing a regional standard for QA/QC of the real-time data acquired from the seas around Korea.

Russia

Federal Ocean State System is now in use but is only available in Russian. The PICES Federated Metadata Searching Project was advertised by Dr. Shevchenko at a workshop held at TINRO-Center. He asked colleagues to use it, especially open-source software. In Russia, the technical problems seem to be easily solvable, the challenge is to get scientists to prepare and use metadata.

U.S.A.

During this year there was a lot of activity in terms of building a national profile for conversion to the new ISO 19115 standard. Several conversion programs were constructed

to do a cross-walk from FGDC to ISO. URLs are available for these reports. It is still not clear if discipline-specific profiles for ISO have been developed.

The WOCE Atlas has been updated in 2007 with physical and oceanographic parameters. This new version is a big improvement on the old one since it is friendlier and searchable.

Discussion of FUTURE (Agenda Item 5)

TCODE has not been actively involved in the development of a Science Plan for a new PICES scientific program, FUTURE (Forecasting and Understanding Trends, Uncertainty and Responses of North Pacific Ecosystems), but this will change at the implementation planning stage. The current statement in planning documents concerning data management is not clear on what data management TCODE can contribute. This activity in FUTURE may include building inventories and improving access to key retrospective datasets, providing on-line access to the PICES publications, Annual Meeting abstracts, *etc.* All these items may be listed in the Implementation Plan under Data and Information Management Services.

Planning for PICES XVII (Agenda Item 6)

TCODE proposed a joint scientific/e-poster session with MONITOR for PICES XVII on “*Data requirements and data sources to support FUTURE*”. One of the goals of this session is to bring to the meeting national data repository managers who manage and provide the data. It is planned to discuss such questions as “what is metadata and what can be done using this information” and “how to get people involved in creating and using metadata”. It was suggested to have presentations of data centers on what they have, can do and how to work better together. Representatives of national data centers and the ICES Working Group on *Data and Information Management* should be invited. After receiving a request from MONITOR to co-sponsor a Topic Session on “*Linking biology, chemistry, and physics in our observational systems – present status and FUTURE needs*” (MONITOR Endnote 6), it was agreed to

incorporate these ideas into the joint session. TCODE nominated Hernan Garcia and Allen Macklin (U.S.A.), Toru Suzuki (Japan) and a representative of China (TBD) to serve as co-convenors.

Relations with other international programs and organizations (Agenda Item 7)

The PICES Standing List of International and Regional Organizations and Programs was discussed. This list is used in part to assist Science Board and the Executive Secretary in decisions regarding sending a representative to meetings of other international organizations and programs. It was suggested that the Pacific Arctic Group (<http://www.arctic.noaa.gov/aro/pag/>) be added to the list. The Pacific Arctic Group (PAG) is a loose confederacy of institutes and individuals having a Pacific perspective on Arctic Science. PAG themes include climate, contaminants, human dimensions and structure and function of Arctic ecosystems.

Elections of TCODE Chairman and Vice-Chairman (Agenda Item 8)

Drs. Megrey (U.S.A.) and Kyu-Kui Jung (Korea) were nominated by Dr. Shevchenko (seconded by Mr. Robin Brown) and unanimously elected for a 3-year term as the new Chairman and Vice-Chairman of TCODE, respectively. Their terms will begin immediately after the closing of PICES XVI. The Committee expressed its gratitude to Dr. Shevchenko for his leadership and valuable contribution to TCODE activities over the last six years.

PICES Ocean Monitoring Service Award (Agenda Item 9)

The PICES Ocean Monitoring Service Award (POMA) was established last year to recognize organizations, groups and outstanding individuals that have contributed significantly to the progress of marine science in the North Pacific through long-term ocean monitoring and management of data associated with ocean conditions and marine bio-resources in the region. Dr. Shevchenko recommended to nominate for the 2008 award the team led by Dr. Megrey and Mr. Allen Macklin

TCODE-2007

for advancing the PICES Federated Metadata Searching Project. The proposal was seconded by Dr. Holmes and supported by the Committee.

Items with financial implications (Agenda Item 10)

TCODE requests:

- Travel support for 1 invited speaker to attend the joint MONITOR/TCODE/BIO scientific/e-poster session at PICES XVII (*MONITOR Endnote 6*);
- US \$2,490 for outsourcing a remote server for one year (*TCODE Endnote 3*).

TCODE work plan for 2006/2007 (Agenda Item 12)

Based on the discussion of all agenda items, the Committee adopted the following work plan:

- Continue to support HAB-S work with HAE-DAT database and required metadata (Responsibility – R. Brown);

- Co-organize scientific/e-poster Topic Session on “*Linking biology, chemistry, and physics in our observational systems – present status and FUTURE needs*” at PICES XVII (H. Garcia, A. Macklin, T. Suzuki);
- Continue collaboration with ICES Working Group on *Data and Information Management* (B. Megrey, G. Moiseenko, I. Shevchenko);
- Continue Federated Metadata Searching Project:
 - Implement remote server proposal (B. Megrey with national representatives);
 - Run the PICES clearinghouse nodes performance and experiment on accessing clearinghouses from different locations (B. Megrey with national representatives);
 - Report on metadata recording in Canada (J. Holmes);
- Update TCODE Action Plan (B. Megrey, I. Shevchenko);
- Coordinate activities with MONITOR (T. Royer).

TCODE Endnote 1

Participation list

Members

Robin Brown (Canada)
Hernan Garcia (U.S.A.)
John Holmes (Canada)
Kyu Kui Jung (Korea)
Bernard A. Megrey (U.S.A.)
Georgiy Moiseenko (Russia)
Thomas C. Royer (U.S.A.)
Igor Shevchenko (Russia, Chairman)

Toru Suzuki (Japan)
Tomowo Watanabe (Japan)
Ruguang Yin (China)

Observers

Kimberly Bahl (U.S.A.)
Jixiang Chen (China)
Janet Webster (U.S.A.)
Emmy Wong (Canada)

TCODE Endnote 2

TCODE meeting agenda

1. Welcome and introduction of members
2. Adoption of agenda
3. Review progress on items in the 2006/2007 workplan
 - a. Continue to support HAB-S work with HAE-DAT database and required metadata (Robin Brown)
 - b. Organize a scientific session/e-poster session at PICES XVI (Allen Macklin, Kyu-Kui Jung)
 - c. Develop collaboration with ICES Working Group on *Data and Information Management* (Georgiy Moiseenko, Bernard Megrey, Igor Shevchenko)

- d. Continue Federated Metadata Searching Project (Allen Macklin, Bernard Megrey, Igor Shevchenko, Norio Baba)
 - Complete Phase II report and promote use of metadata
 - Carry out Phase III including capacity building
 - Investigate utility of Asian-side metadata server
- e. Update TCODE Action Plan. (Robin Brown, Igor Shevchenko)
- f. Coordinate activities with MONITOR (Thomas Royer)
- 4. National annual reports
- 5. Discussion of FUTURE
- 6. Planning for PICES XVII
- 7. Relations with other international programs and organizations
- 8. Election of Chairman and Vice-Chairman
- 9. PICES Ocean Monitoring Service Award
- 10. Items with financial implications
- 11. New business
- 12. TCODE Workplan for 2007/2008

TCODE Endnote 3

A TCODE/MONITOR proposal to establish a PICES remote server

Background

The need for scientific information within the PICES community is increasing and is expected to increase further when FUTURE becomes active. This information can consist of data, metadata, collaboration portals, PICES publications, direct and supporting material for the North Pacific Ecosystem Status report, and a web interface to quickly discover, access, and preliminarily evaluate this information. Even though some of these resources are available today, there is no one-stop PICES web interface to accomplish the tasks described above.

Proposal

We propose that PICES outsource (rent) a remote server for one year to test PICES web resource, and PICES should consider the cost as part of the scientific infrastructure of a maturing and quickly expanding PICES scientific program. For minimal annual costs PICES can rent a capable remote server with the following characteristics:

- Assignment of 16 dedicated IP addresses (needed to register country nodes to the metadatabase clearinghouse); this means one piece of hardware (the rented remote server) could act as 16 different virtual servers which would be more than enough to meet the needs of PICES;
- Multiple login accounts to give PICES users and developers required access;

- Root access for each login (or IP address) to allow installation of custom software and to perform account management;
- FTP access to allow transfer of information, data and software to the PICES remote server;
- Free system repair;
- OS updates and patches;
- 24/7/365 technical support (by phone and email); this means that any PICES member country, regardless of the time zone on which they work, will have access to technical support.

A capable remote server hardware configuration is minimally defined as:

- Windows Server 2003 operating system;
- Server hardware configuration;
- Pentium 2.5-2.8Ghz CPU;
- 1 GB RAM;
- 2 × 40GB RAID 1 hard disk storage;
- 100/1000Mbps CAT6 on 100 Mbps port with 75GB of Tier 1 internet bandwidth.

PICES services

Development test-bed for a PICES GeoNetwork web interface:

GeoNetwork is a free, open-source software system that offers many useful scientific services. It can offer metadata clearinghouse functions, document library management, map server data presentation functions and a data distribution interface. We propose that one

virtual server be allocated as a development server to test the functionality of GeoNetwork and its scientific potential to PICES. The TINRO-Center laboratory has experience in working with GeoNetwork.

Metadata clearinghouse:

The PICES Metadata Federation process has been successful at every incremental step. Participating laboratories have been enthusiastic about their involvement in the project, and we believe the clearinghouse has proven to be a valuable resource to PICES scientists. The time is right for the Organization to consolidate all PICES metadata nodes. We recommend that PICES assume the responsibility of providing the federation clearinghouse function, gather valuable Pacific-Rim metadata resources under one umbrella, and give it the unique and prominent PICES name it deserves – a recognizable trademark that communicates scientific excellence. We propose that one virtual server be allocated for serving the PICES metadata through the already established NSDI (National Spatial Data Infrastructure) clearinghouse, and that one virtual server be assigned to test GeoNetwork as a metadata clearinghouse.

PDF files:

PICES wants to place their library of large PDF files on a server connected to a high bandwidth data server so as to reduce download times by PICES users. We propose that PICES use one virtual server for this task.

Distribution point for North Pacific CPR value-added data products:

MONITOR is proposing to make some value-added products from the North Pacific CPR (Continuous Plankton Recorder) project available to the PICES community. We propose to satisfy this need with a remote server and use the CPR data library in the GeoNetwork test mentioned above.

Support for the North Pacific Ecosystem Status Report (NPESR):

Many people contribute to the NPESR. We propose to use a remote server to manage the

many aspects of this complex project and, when possible, serve chapters and time series as living documents linked to actual data.

Cost considerations

There will be no software costs, as the communication software (Isite) and the clearinghouse and web interface software (GeoNetwork) needed for the metadatabase function are free open-source software.

There will be a direct cost to rent the remote server hardware. To approximate the cost, we have solicited bids from two vendors found on the World Wide Web, RackSpace and Adhost. The bid from Adhost is the lower of the two and is US \$195/month plus a US \$150 one-time setup fee. Using these figures, the annual estimated cost for a 12 month contract totals US \$2,490. This is a relatively small cost for the services provided given that PICES has supported the Federated Metadata Searching Project for the past 3 years at US \$4,000/yr.

Vendors

Many other options are available, but in our evaluation, the two bids from RackSpace and Adhost represent typical costs and the state-of-the-art in server farm service offerings. Server farms located in U.S.A. are preferred because they adhere to the high Tier 1 standards in terms of reliability, security and internet speed and because the internet data infrastructure in U.S.A. meets the highest worldwide standards for reliability. For example, the Adhost Data Center in Seattle offers the following:

- Data Center-physical characteristics,
- Dual access redundant power feeds,
- Power conditioning through redundant UPS (Uninterruptible Power Supply) and power backup units,
- Redundant cable routing system,
- Backup network equipment,
- Zoned dry-pipe pre-action fire suppression system,
- Power backup provided by multi-mW generators fed from diesel reservoirs,

- Two emergency wells to supply water to cooling towers if city water becomes unavailable,
- Data Center- network,
- Redundant Gig-E dedicated leased connections to multiple Tier 1 providers on diverse multiple paths,
- Redundant Cisco routers connected to redundant Cisco switches,
- 100% switch-based 100/1000 Mbps ports connected to Cisco switches.

Impacts on the PICES Secretariat

Since the clearinghouse node will be a rented remote server, there should be no impact on scarce Secretariat resources for the metadata functions. Representatives from each member country will be responsible for maintaining and updating their own individual virtual nodes. Other use functions are already assumed within the Secretariat.

Benefits to PICES

For the metadata function, renting a remote server will consolidate all PICES metadata into one resource. PICES branding of individual country contributions will unite the commitment to the international effort. The GeoNetwork could prove to be a useful web tool, facilitating support of PICES scientific activities. Faster PDF downloads, serving valuable CPR data and NPESR living documents would provide a valuable service to the PICES community.

Benefits to participating Metadata Centers

Renting a remote server will eliminate security risk of opening a port through laboratory firewalls to the internet to permit connection of the node server by the clearinghouse server. Metadata will reside with the federation partner and will also be duplicated on the remote server. This additional backup of the metadata adds an additional layer of security.

