

**REPORT OF TECHNICAL
COMMITTEE ON DATA EXCHANGE**
(final draft - Jan 11, 2008)

The meeting of the Technical Committee on Data Exchange was held from 16:00-19:15 hours on October 31, 2007. The Chairman, Dr. Igor Shevchenko, called the meeting to order and welcomed the participants. The meeting was attended by 10 TCODE members and 4 observers representing PICES member countries and international organizations (*TCODE Endnote 1*). John Holmes served as 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 Workplan (TCODE
Agenda Item 3)**

- Continue Federated Metadata searching project

All PICES member countries, except Canada, have Clearinghouse nodes and are contributing metadata. The latest country brought online was China. A node was not established at the National Marine Data and Information Service laboratory during the Tianjin meeting in August 2007 due to firewall configuration problems. NMDIS is looking into a solution. A training course on FGDC metadata preparation was completed successfully. A suggestion was put forward and supported that an objective of the Federated Metadata project for the next year should be to report on how the remote server service performs. A Federated Metadata progress report for 2007 will be provided by B. Megrey to the Secretariat for posting on the PICES website. Canadian participation in Federated Metadata 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. J. Holmes (Canada) 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 server service for 1 year to consolidate all PICES Metadata federation on US clearinghouse service into one service was discussed.

TINRO-Center has used Geonetwork and Geoserver software and a short presentation on this topic was provided by I. Shevchenko. The software is written in Java and so can be used in Windows, Linux, Unix environments. A suggestion was to install Geonetwork and Geoserver on consolidation server and test them in this virtual server environment.

B. Megrey (USA) was asked to refocus the proposals to reflect discussions at TCODE and consider making them a joint proposal with MONITOR (*TCODE Endnote 3*).

I. Shevchenko (Russia) reports that in Russia they cannot connect to servers of the several metadata clearinghouses. An experiment discussed in which country representatives on TCODE test to see if this is true for their countries.

- Update TCODE Action Plan

Several minor changes to the plan were discussed and approved by TCODE. The plan was updated and posted on the TCODE website.

**Annual reports (TCODE 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 modeling, publications on marine DM issues, education materials, etc. All reports will be posted at TCODE website.

Canada - One addition to the last year 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 this committee is to guide data management strategies and priorities. This activity includes items of direct interest to TCODE, including metadata and data 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”, with real-time data on-line at: http://www.venus.uvic.ca/data/data_plots.html. Venus is now live at Pat Bay node. Another node is being placed off Roberts Bank in the Strait of Georgia now and another ship is laying cable out to

Endeavour Ridge hotvents.

Japan – The Fishery Agency DM activities are focused on the Sea of Japan. One goal is to put together a modern database and make it searchable. The database implementation is behind schedule at present due to the need to bring in some older data. It is expected to be finished in the next year. Three monitoring lines, off Hokaido, near Tokyo, and in East China Sea, are used to monitor the lower trophic levels every season. These observations will continue to be made to 2010. The data collected are going into a database. There is no a generally accepted approach to metadata bases design. Many old metadata records are in Japanese only. They should be translated into English and this is time-consuming activity.

China – All activities to join PICES metadata federation were completed. The software tools to support the metadata production are tested. Translating metadata from Chinese to English is difficult to arrange.

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. KODC 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 Metadata Federation project was advertised by I. Shevchenko at a workshop held in TINRO-Center. He asked colleagues to use it, especially open-source software. In Russia, the technical problems seem to be solvable easily, the challenge is to get scientists to prepare and use metadata.

United States of America – 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 cross-walk from FGDC to ISO. URLs are available for these reports. Still it is 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 the FUTURE (PICES Future Integrative Scientific Program) (TCODE Agenda Item 5)

Role of TCODE at this stage (development of scientific plan) is not active. TCODE will be actively involved at the implementation planning stage. The current statement in planning concerning data management is not clear on what data management TCODE can contribute. Data management in FISP may include building inventories and improving access to key retrospective datasets, providing online access to the PICES publications, annual meeting abstracts, etc. All these items may be listed in the plan under Data and Information Management Services.

Topic session proposals for PICES XVII (TCODE 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**’ (see *TCODE Endnote 4*). One of the goals 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 metadata is and what can be done with it, how to get people involved in creating and using it. It was suggested to have presentations of data centres on what they

have, can do and how to work better together. Representatives of national data centres and ICES WG on Data and information management should be invited.

Relations with other international programs/organizations (TCODE Agenda Item 7)

Standing list of meetings up to 2010 was discussed (used for PICES travel). The goal was to consider where TCODE could sponsor sessions or events. It was suggested that Pacific Arctic Group (PAG, <http://www.arctic.noaa.gov/aro/pag/>) should be added to this list. The Pacific Arctic Group is a loose confederacy of institutes and individuals having a Pacific perspective on Arctic Science. PAG science themes are climate, contaminants, human dimensions and structure and function of Arctic ecosystems.

Chairmanship of TCODE (TCODE Agenda Item 8)

Bernard Megrey (USA) and Kyu-Kui Jung (Korea) were nominated (I. Shevchenko, seconded by R. Brown) as Chair and Vice-Chair, respectively. Their nominations were not opposed so they were endorsed by TCODE as the incoming Chair and Vice-Chair.

PICES Ocean Monitoring Service Award (TCODE Agenda Item 9)

PICES Ocean Monitoring Service Award was established last year to reward individuals or groups that have made great contributions to PICES activities and the progress of marine science in the North Pacific through their long-termed monitoring operations and/or data management for various ocean conditions and marine bio-resources. I. Shevchenko (Russia) moved forward a proposal to nominate for the next year award the team headed by Bernard Megrey and Allen Macklin for advancing the PICES federated metadata project. The proposal was seconded by J. Holmes (Canada).

Summary of items with financial implications (TCODE Agenda Item 11)

MONITOR – T. Royer

TCODE requests support for the following activities:

- Travel support for an invited speaker to attend the scientific session at PICES XVII (see *TCODE Endnote 4*);
- \$USD 2,490 for outsourcing a remote server for one year (see *TCODE Endnote 3*).

Discussion and adoption of the TCODE Workplan for 2007/2008 (TCODE Agenda Item 13)

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

1. Continue to support HAB – R. Brown (Canada)
2. Organize Scientific and E-poster Session at PICES XVII - H. Garcia, A. Macklin
3. Cooperate with ICES WG on Data and Information Management - B. Megrey, G. Moiseenko, I. Shevchenko
4. PICES Metadata Federation Project

- Remote server proposal – B. Megrey to organize with national representatives;

- Run the PICES clearinghouse nodes performance and experiment on accessing clearinghouses from different locations – B. Megrey to organize with national representatives;

- Report on metadata recording in Canada - J. Holmes

- Update TCODE Action Plan – B. Megrey, I. Shevchenko
- Coordination of Activities with

TCODE Endnote 1

Participation List

Members in attendance:

Canada
Robin Brown
John Holmes

China
Ruguang Yin

Japan
Tomowo Watanabe
Toru Suzuki

Korea
Kyu Kui Jung

Russia
Georgy Moiseenko
Igor Shevchenko

U.S.A.
Hernan Garcia
Bernard Megrey
Tomas Royer

Observers:

Canada
Emmy Wong, DFO/IOS

China
Jixiang Chen, NMDIS

U.S.A.
Kimberly Bahl, University of Washington
Janet Webster, Oregon State University/
IAMSUC

**TCODE Meeting Agenda
Wed. Oct. 31, 2007 16:00-19:30**

Chairman: Igor Shevchenko (Russia)

1. Welcome and introduction of members
2. Adoption of agenda (opportunity to add agenda items under "New business")
3. Review progress on the items from the 2006/2007 Workplan
 - a. Continue to support HAB-S work with HAE-DAT database and required metadata
Responsibility - Robin Brown
 - b. Organize the Scientific Session/E-poster session at PICES XVI
Responsibility - Allen Macklin, Kyu-Kui Jung
 - c. Develop collaboration with ICES Working Group on Data and Information Management
Responsibility - Georgy Moiseenko, Bernard Megrey, Igor Shevchenko
 - d. Continue Federated Metadata searching project
 - Complete Phase II report and promote use of metadata
 - Carry out Phase III including capacity building
 - Investigate utility of Asian-side metadata serverResponsibility - Allen Macklin, Bernard Megrey, Igor Shevchenko, Norio Baba
 - e. Update TCODE Action Plan
Responsibility - Robin Brown & Igor Shevchenko
 - f. Coordinate activities with Monitor Technical Committee
Responsibility - Tom Royer
4. Annual reports - TCODE members for Canada, Japan, People's Republic of China, Republic of Korea, the Russian Federation, and the United States of America
5. Discussion of the FISP (PICES Future Integrative Scientific Program)
6. Topic session proposals for PICES XVII
7. Relations with other international programs/organizations
8. Election of Chairman and Vice-Chairman
9. PICES Ocean Monitoring Service Award
10. Summary of items with financial implications:
 - Proposed inter-sessional meetings for 2007 and beyond
 - Proposed publications for 2007 and beyond
 - Travel support requests
 - Other items
11. New business (additional items added at the meeting)
12. Discussion and adoption of the TCODE Workplan for 2007/2008

A Proposal to Establish a PICES Remote Server

Co-sponsored by TCODE and MONITOR

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.

The Proposal: We propose that PICES outsource (rent) a remote server for one year. The server would serve in the capacity of a 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 address (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 the PICES requirements.
- 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 any PICES 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 x 40GB RAID 1 hard Disk storage
- 100/1000Mbps CAT6 on 100 Mbps Port with 75GB of Tier 1 internet Bandwidth

PICES Uses:

- *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, mapserver data presentation functions and a data distribution interface. We propose one of the virtual servers be allocated as a development server to test the functionality of GeoNetwork and its scientific potential to PICES. The TINRO laboratory uses GeoNetwork and has experience with it.
- *Metadata Clearinghouse*
The PICES metadata federation process has been successful at every incremental step,

individual country laboratories have been enthusiastic about participating and we believe the clearinghouse has proven to be a valuable resource to PICES scientists. The time is right for PICES to consolidate all PICES metadata nodes. We propose that PICES assume the responsibility to provide the federation clearinghouse function, gather valuable Pacific-rim metadata resources under one PICES umbrella of scientific activity, and give it the unique and prominent PICES name it deserves – a recognizable scientific trademark that communicates scientific excellence. We propose that one of the virtual servers be allocated to the production serving of the PICES metadata through the already established NSDI 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 PICES use this server to act in that capacity.
- *Distribution point for CPR value added data products*
MONITOR is proposing to make some CPR value added products available to the PICES community. We propose that we satisfy this need with this server and use part of the CPR library of information in the GeoNetwork test mentioned above.
- *Support for the North Pacific Ecosystem Status Report(ESR)*
Many people contribute to the ESR. We propose to use this server to manage the many aspects of organizing this complex document and, when possible, serve chapters and time series as living documents linked to actual data.

Cost Considerations: There will be no software costs. 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. We have solicited two bids to approximate the cost - one from RackSpace and one from Adhost. These vendors were found on the world wide webs. The bid from Adhost is the lower of the two and it amounts to \$USD195/month plus a \$USD150 one-time setup fee. Using these figures, the annual estimated cost for a 12 month contract totals \$USD 2,490 (12*195+150).

We feel this is a relatively small cost for the services provided given that PICES has supported the Metadata Federation for the past three years at \$USD 4K/yr.

The 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. Sever farms located in the USA 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 the USA 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 metadatabase function, all PICES metadata would be consolidated into one PICES resource. PICES branding of individual country contributions will consolidate the commitment to the international effort. The GeoNetwork could prove to be a valuable web tool, facilitating support PICES scientific activities. Faster PDF downloads, serving valuable CPR data and ESR living documents provide a valuable service to the PICES community.

Benefits to participating Metadata Centers: *Security* - Eliminate security risk of opening a port through laboratory firewalls to the internet to permit connection of the node server by the clearinghouse server. *Redundancy* – 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.

Data Requirements and Data Sources to Support FUTURE

Co- Convenors: Jeffrey Napp USA and Young-Jae Ro (KOREA) – from MONITOR
Hernan Garcia, Allen Macklin (USA) and a representative of China – from TCODE

Numerical models are becoming increasingly complex, attempting to vertically and horizontally integrate ecosystem forcing, processes, and predictions across multiple trophic levels from bacteria to human populations. Data requirements for daily, in season, and annual or decadal predictions differ according to whether we are interested in single species, species assemblages, or multi-trophic level predictions. To add to our challenge, the types of sensors and frequency of measurement varies greatly across ecosystem components, particularly the biology. This session encourages papers and presentations including e-posters that:

- Define and specify the types, frequency and spatial resolution of observational data required for our current numerical models

- Review existing and emerging data sources and technologies capable of supplying biomass; species and functional group information

- Review existing and emerging data sources and technologies capable of integrating these data with physical and chemical information

- Showcase novel data assimilation techniques, formal organization of data or database that facilitate the operational use of observational data to predict the effects of anthropogenic and climate forcing on the major ecosystems of the North Pacific.