Intergovernmental Oceanographic Commission
Reports of Meetings of Experts and Equivalent Bodies

IODE Steering Group for MEDI

Third Session

Drexel University
Philadelphia, USA
11-13 September 2006

UNESCO 2006

Paris, 25 September 2006
English only

Abstract

The Marine Environmental Data Information Referral Catalogue (MEDI) is a directory system for datasets, data catalogues and data inventories developed by IODE. The IODE Steering Group for MEDI was established to support the MEDI system. During its Third Session the Steering Group discussed the implementation of a marine community profile of the ISO 19115 metadata standard. The marine community profile developed by the Australian Ocean Data Centre Joint Facility will be disseminated for comment through a new IODE metadata discussion list. The Steering Group will review comments received and recommend a marine profile for MEDI. The Steering Group agreed that governance of vocabularies used by MEDI should be responsibility of the MarineXML group.

The Steering Group also discussed cooperation with the JCOMM Water Temperature Metadata (META-T) Pilot Project and agreed that SG-MEDI can assist the META-T Pilot Project by providing discovery metadata tools and format and also with the technical governance of controlled vocabularies.

TABLE OF CONTENTS

Page

1. INTRODUCTION

The Session was opened on Monday, 11 September 2006 at 08:30 at Drexel University, Philadelphia, USA, by Greg Reed, Chairman of the MEDI Steering Group (SG-MEDI). Mr Reed welcomed the participants to the meeting and thanked the organisers of the 7th International Conference on Hydroscience and Engineering (ICHE-2006) for providing the venue for the meeting. The Chair expressed his regret that two members of the Group, Ms Jixiang Chen and Dr Malika Bel Hassen, were unable to attend the meeting. The list of participants is given in Annex II.

The Chair introduced the Provisional Agenda for the meeting and invited comments. The Group adopted the Agenda, as given in Annex I.

2. CHAIRMAN'S REPORT

The Chairman outlined the objectives of the meeting. He recalled that the Marine Environmental Data Information Referral Catalogue (MEDI) is a directory system for datasets and data inventories developed by IODE. MEDI has been operational within IODE for over thirty years. The MEDI authoring tool was introduced in 2001 and has been used by a number of IODE data centres and is used during IODE data management training courses and capacity building activities. However, the MEDI tool is built on older versions of JRE and Tomcat and is no longer supported and this has resulted in problems for new installations. The Eighteenth Session of IOC Committee on International Oceanographic Data and Information Exchange (IODE-XVIII) adopted Recommendation IODE-XVIII.3 (Annex III). The task for the Steering Group is to implement these recommendations. In addition, the Group should work more closely with other activities such as JCOMM and SeaDataNet.

3. REPORTS FROM STEERING GROUP MEMBERS

The Steering group members reported on current metadata developments of relevance to MEDI.

Edward Vanden Berghe described the Access to Biological Collection Data (ABCD), a proposed standard XML schema for distributed data retrieval from collection databases. ABCD is a common data specification for biological collection units, including living and preserved specimens, along with field observations that did not produce voucher specimens. ABCD is intended to support the exchange and integration of detailed primary collection and observation data.

Roy Lowry discussed the NERC Data Grid (NDG) discovery metadata which uses an intermediate metadata schema (MOLES). XQuery is used to generate discovery metadata documents in various formats. Documents are posted on OAI by data hosts and are centrally harvested into an eXist XML database. MOLES is currently based on DIF but there are plans to move to one or more ISO19139 profiles. Dr Lowry also described the SeaDataNet discovery metadata. The SeaDataNet project has inherited three heavily populated legacy discovery metadata resources, namely, EDIOS (observing programme datasets) from EuroGOOS, EDMED (European marine datasets) from SEA-SEARCH, and ROSCOP (Cruise Summary Reports) from SEA-SEARCH and ICES. SeaDataNet will harmonise structures and vocabularies as a migration towards a single intermediate metadata schema and develop marine profile exports for operational (EDIOS), cruise (ROSCOP) and general-purpose (EDMED) datasets, accepting that there will be information loss. Some of the concerns of the SeaDataNet project include (i) the size and complexity of the job that needs to be done, (ii) looming deadlines, and (iii) the maintenance of content integrity.

Don Collins described the NODC archive management system (AMS) which consists of the Accession Tracking Database (ATDB). The NODC metadata repository uses the FGDC standard database that can export metadata to a geospatial one-stop. Metadata can also be exported to GCMD/MEDI and other clearinghouses after review and publication. Other NODC activities include the National Coastal Data Development Center (NCDDC) which provides coastal data resources with a tool to develop, validate, manage and publish metadata records via secure internet access. The Metadata Enterprise Resource Management Aid (MERMAid) allows users and data providers to establish unlimited metadata databases (see www.ncddc.noaa.gov/). NOAA's Coral Reef Information System (CoRIS) provides a single point of access to coral reef information and data products, especially those derived from NOAA's Coral Reef Conservation Program. A metadata portal for the coral data collections and information is managed at NODC for the Coral Reef Conservation Program office. NODC plans for 2007 include the update of metadata for historical data collections, review and revise controlled vocabularies, improve export of descriptive metadata from processing activities.

Greg Reed discussed current metadata developments in Australia. A marine community metadata profile of the ISO 19115 international standard has been developed under the sponsorship of the Australian Ocean Data Centre Joint Facility (AODCJF). The Marine Profile has been developed in accordance with the rules established by the international metadata. The Marine Profile has been developed as a subset of the international standard and includes all ISO 19115 core metadata elements and selected non-core elements. In addition, the Marine Profile has defined supplementary metadata elements, codelists and vocabularies to meet the needs of the marine community to support the documentation and discovery of marine resources. The AODCJF is also developing a metadata entry and search tool as open-source software. This metadata tool will support the Marine Community Profile and will assist with standardised metadata creation and submission for a range of client groups describing marine data and services. It will be deployed across all AODCJF agencies and will also serve as a tool for university partners to use when submitting metadata to the AODCJF distributed network.

Elanor Gowland briefed the Group on the establishment of the JCOMM Pilot Project for the collection of real-time metadata regarding sea surface temperature and water temperature profile data (META-T). The Pilot Project will consider the following metadata: (i) information intrinsic to the platform or its instruments and which are not varying during the platform operational life-time; (ii) type of probes; (iii) sensor accuracy, as obtained from manufacturers or through pre-deployment calibration; (iv) drogue type of Lagrangian drifter; (v) information from Argo floats, which is cycle dependent and is presently appended to float profiles in technical files; (vi) quality information flags associated to observations; and (vii) quality information such as standard deviation or bias estimated after platform deployment. The following categories of metadata will be considered: (i) metadata required for real-time distribution along with the observational data; (ii) metadata required for real-time applications but made available separately from the observations; and (iii) metadata made available in delayed mode. The META-T Pilot Project will facilitate the use of ocean temperature information by increasing the quality and quantity of metadata for real time as well as delayed mode activities and the Pilot Project would like to work with SG-MEDI to develop META-T.

Melanie Meaux discussed some of the common metadata standards and how profiles can impact interoperability. Metadata tools can assist with creation of metadata and assure compliance. The GCMD DocBUILDER is a stand alone metadata authoring tool for Earth science data and services. There are a number of controlled vocabularies for marine data, including the BODC Parameter Discovery and Usage Vocabulary, Global Change Master Directory Keywords, CF Standard Names, AGU Index Terms, IOOS Core Variables, JGOFS

Flux Study Parameters, U.S. GLOBEC Thesaurus, and OBIS Taxonomic Categories. The GCMD currently has twelve sets of controlled vocabularies.

John Graybeal (MBARI) described the Marine Metadata Interoperability Project (MMI), created in 2004 with a mission to create a community of metadata-aware scientists and data managers. The project has established a web site (http://www.marinemetadata.org), hosted a vocabulary mapping workshop, and developed tools for mapping vocabularies (VINE - Vocabulary Integration Environment tool). The MMI Steering Committee has been established and its membership includes international participants. Following the successful Advancing Marine Vocabularies workshop held in 2005, MMI will host a Sensor Metadata Interoperability workshop from 19-20 October 2006, in Portland, Maine, USA. The workshop will focus on the content standards needed for data systems to work with and exchange sensor metadata. Five content standards for sensor metadata in observing systems - FDGC, ISO, DIF, sensorML, and transducerML - will be applied to real-world problems and evaluated.

Rainer Lehfeldt (BAW) described the NOKIS project which is implementing information infrastructures as part of an Integrated Coastal Zone Management in Germany (http://nokis.baw.de/). NOKIS has establish a metadata infrastructure for the German North Sea and Baltic Sea coastal regions, based on ISO 19115. The NOKIS profile has been developed to include all ISO recommended core plus additional elements. A metadata tool has also been developed which provides interoperability across all participating institutions..

4. MARINE METADATA PROFILE

The IODE Committee recommended (IODEXVIII.3) the development of a marine profile of ISO19115. The Australian Marine Community Profile of ISO 19115 was discussed. This profile has been circulated to the NDG community and the SeaDataNet community for comment. The Group agreed that the Marine Community Profile should be circulated to a wider audience for further comment on its suitability for use by the international community. One way for the profile to be circulated to the IODE community would be through a dedicated metadata discussion list. The Chair will ask the IODE Project Office to establish a moderated metadata discussion list to be operational by early October so the Marine Community Profile can then be circulated. Subscribers to the list will be requested to submit comment by mid-December 2006. Subscribers to other relevant discussion lists will be asked to subscribe to the IODE metadata discussion list. Roy Lowry will contact the SeaDataNet and CF lists, John Graybeal to contact the MMI list and post information on the MMI website, Edward Vanden Berghe to contact the GEBICH list, Greg Reed to contact IODE about circulating to the JCOMM/ETDMP. The Steering Group will be responsible for the governance of the email discussion list and will be moderated by a small core of the Group. The Chairman will prepare a one-page document describing the Marine Profile with an example in XML to be circulated with the profile. The Group will review the comments received by email in early 2007 and a recommendation on the adoption of an agreed profile will be presented to IODE XIX in March 2007.

5. VOCABULARIES

Roy Lowry discussed governance of controlled vocabularies. Vocabularies are collections of entries representing, through standardised descriptions, instances of an entity. The ISO 19115 standard uses codelists to constrain the population of metadata elements and vocabularies can be used to constrain the population of keyword lists. Governance of vocabularies has two parts, (i) content governance decides what changes are necessary to the entries of a vocabulary (i.e. additions, modifications to existing entries and, in exceptional circumstances, deletions); and (ii) technical governance which covers controlled vocabulary

storage, version management and serving. Content governance makes the decisions and technical governance implements them.

Content governance is often absent in many vocabularies and changes are often made based on unplanned decisions with definitions being either absent or poor in quality. The inclusion of definitions forces research and the application of quality thought. Some of the issues to be addressed in technical governance of vocabularies include:

- "Broken" systems. Where terms have no keys, the term itself becomes embedded in user schemas, in relational databases and XML documents. Local lists of terms can be used to constrain usage.
- Versioning problems. BODC, MEDS, SEA-SEARCH and others have been serving vocabularies for years without versioning. Ontology maintenance will be unmanageable without versioning and version histories.
- Valid terms not explicitly included in lists. The GCMD Science Keywords are made up of CATEGORY > TOPIC > TERM > VARIABLE. The DIF validator allows for variable to be absent but Science Keyword list does not include these truncated entries so interoperability can be compromised as a result.
- Stale version serving. "Stale" version incidents can be caused by locally cached copies, no regular updates, and no web server synchronisation.

NERC Data Grid has established a vocabulary server to directly access the master database (http://www.bodc.ac.uk/products/web_services/vocab) and a demonstration client that can be used for list downloads (http://vocab.ndg.nerc.ac.uk/client/vocabServer.jsp).

Content governance can be achieved through a moderated e-mail discussion list. The Group agreed that vocabulary governance for MEDI should be the responsibility of MarineXML group. Roy Lowry will send a message to the relevant lists publicising the marine profile and vocabularies and will set up a parallel to the CF email discussion list.

6. MEDI AUTHORING TOOL

The Chair described the current activities in Australia to develop a metadata entry and search tool (MEST). The MEST will be developed based on the GeoNetwork open source software (http://geonetwork-opensource.org/) and will use the Marine Community Profile. The Chair agreed to keep the Group informed on progress of the MEST development and provide details when tool is available for testing. The Group can then evaluate the suitability of the tool for the IODE community.

7. OTHER BUSINESS

The Group discussed cooperation with the JCOMM Water Temperature Metadata (META-T) Pilot Project. The Group noted the difference between MEDI and META-T. Whereas MEDI provides discovery metadata, META-T defines a very broad set of metadata for a very specific application of use metadata. There is no possible generic solution for META-T. The META-T project needs to define the problem to provide a better understanding of the aims of the Pilot Project.

SG-MEDI can assist by providing discovery metadata tools and format and also with the technical governance of controlled vocabularies. John Graybeal noted that the OOSTethys project (http://www.oostethys.org) that is proposing the "system of systems" concept for the ocean observing community could provide a possible solution for META-T.

8. MEDI ACTION PLAN

The Steering Group agreed to the following actions:

i. Request IODE to establish a moderated metadata discussion list (Greg Reed to contact IODE Project Office). The mailing list should be operational by early October.
ii. Circulate the Marine Community Profile through the metadata discussion list by mid-October with a request for comment by mid-December.
iii. Subscribers to other discussion lists will be asked to subscribe to the IODE metadata discussion list. Roy Lowry will contact the SeaDataNet and CF lists, John Graybeal to contact the MMI list and post information on the MMI website, Edward Vanden Berghe to contact the GEBICH list, Greg Reed to contact IODE regarding circulating to JCOMM/ETDMP.
iv. SG-MEDI will be responsible for the governance of the email discussion list and will be moderated by a small core of the Group.
v. Greg Reed to prepare a one-page document describing the Marine Profile with an example in XML (by mid October). This will be distributed to the discussion list.
vi. The Steering Group will review comments received on the Marine Profile (early 2007) and a recommendation on the adoption of an agreed profile to be presented to IODE XIX in April 2007.
vii. Greg Reed will prepare the meeting report and forward to the Chair of the JCOMM Data Management Coordination Group (DMCG) for the next meeting of the DMCG on 10 October 2006.
viii. Vocabulary governance will be the responsibility of MarineXML group. Roy Lowry will send a message to the relevant lists publicising the marine profile and vocabularies. Roy will also set up a parallel to the CF email discussion list.
ix. Development of the Metadata Entry & Search Tool (MEST) will be monitored by the Chair who will inform the Group when a test version is available.

9. CLOSURE

The Chairman thanked all participants for their contribution to the meeting of the MEDI Steering Group. The Third Session of the IODE Steering Group for MEDI was closed on Wednesday 13 September at 13:00.

ANNEX I

AGENDA

1. OPENING OF THE SESSION
2. CHAIRMAN'S REPORT
3. REPORTS FROM STEERING GROUP MEMBERS
4. MARINE METADATA PROFILE
5. VOCABULARIES
6. MEDI AUTHORING TOOL
7. OTHER BUSINESS
8. PREPARATION OF THE MEDI ACTION PLAN
9. CLOSURE OF THE MEETING

ANNEX II

LIST OF PARTICIPANTS

STEERING GROUP MEMBERS

Greg REED (Chairman)
Executive Officer
Australian Ocean Data Centre Joint Facility (AODCJF)
AUSTRALIA
Email: greg@metoc.gov.au
URL: http://www.aodc.org.au

Donald W. COLLINS
US National Oceanographic Data Center
1315 East West Highway
Silver Spring, MD 20910-3282
UNITED STATES
Email: donald.collins@noaa.gov

Roy LOWRY
Technical Director
British Oceanographic Data Centre (BODC)
Joseph Proudman Building
6 Brownlow Street
Liverpool L3 5DA
UNITED KINGDOM
Email: rkl@bodc.ac.uk

Edward VANDEN BERGHE
Manager, Flanders Marine Data and Information Centre
Flanders Marine Institute
Vismijn, Pakhuizen 45-52
B-8400 Ostend
BELGIUM
Email: wardvdb@vliz.be

STEERING GROUP MEMBERS (not attending meeting)

Malika BEL HASSEN
Chercheur, Institut National des Sciences et Technologies de la Mer (INSTM)
28, rue 2 Mars 1934, Carthage
2025 Salammbo
TUNISIA
Email: belhassen.malika@instm.rnrt.tn

Jixiang CHEN
Marine Data Center
National Marine Data and Information Service (NMDIS)
93 Liuwei Road, Hedong District
Tianjin 300171
CHINA
Email: cathy_chen@mail.nmdis.gov.cn

JOINT COMMITTEE ON MARINE METEOROLOGY (JCOMM)

Elanor GOWLAND
Global Marine Data Manager and QEv co-ordinator
Met Office Edinburgh
Saughton House
Broomhouse Drive
Edinburgh EH11 3XQ
UNITED KINGDOM
Email: elanor.gowland@metoffice.gov.uk

OBSERVERS

John GRAYBEAL
Monterey Bay Aquarium Research Institute (MBARI)
7700 Sandholdt Road
Moss Landing, CA 95039-9644
UNITED STATES
Email: graybeal@mbari.org

Rainer LEHFELDT
Managing Director,
German Coastal Engineering Research Council
Federal Waterways Engineering & Research Institute (BAW)
Wedeler Landstrasse 157
22559 Hamburg
GERMANY
Email: rainer@baw.de

Melanie MEAUX
NASA/GCMD
NASA Goddard Space Flight Center
Greenbelt, MD 20771
UNITED STATES
Email: mmeaux@gcmd.gsfc.nasa.gov

ANNEX III

RECOMMENDATION IODE-XVIII.3

MARINE ENVIRONMENTAL DATA INVENTORY (MEDI) PROGRAMME

The IOC Committee on International Oceanographic Data and Information Exchange,

Recognizing the value of a metadata directory system for databases, data catalogues and data inventories to a broad user community, including IOC programmes such as GOOS and related activities within other global and regional programmes,

Recalling Recommendation IODE-XVI.1 that established the MEDI Steering Group,

Further recalling that IODE-XVII recommended that MEDI should be ISO compliant,

Recommends that the MEDI Steering Group collaborate with national and international metadata initiatives to ensure a metadata system, that meets international standards, is available to the IODE member states and that the MEDI Steering Group be reconstituted with the following terms of reference:

(i) identify and document current metadata management best practice,
(ii) in collaboration with national and international initiatives, as well as the IODE Groups of Experts, recommend a marine profile of ISO19115 and develop relevant vocabularies to describe marine datasets,
(iii) make available an ISO19115 compliant metadata entry tool to the IODE community to ensure metadata is uniformly collected.

Recommends that the membership of the Steering Group shall include Australia (AODCJF), Belgium (VLIZ), China (NODC), Tunisia (INSTM), UK (BODC) and USA (NODC) and other experts as appropriate, and shall be coordinated by Mr. G. Reed (Australia);

Further recommends the continued incorporation of MEDI as a training module in all IODE training activities and capacity building products;

Urges Member States to use the MEDI system as appropriate and promote its use to the widest possible audience.