Oceanography Committee **ICES CM 2003/C:12**

Report of the

ICES-IOC Study Group on the Development of Marine Data Exchange Systems using XML

Gothenburg, Sweden
26–27 May 2003

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

Palægade 2–4 DK–1261 Copenhagen K Denmark

TABLE OF CONTENTS

Section **Page**

1 OPENING OF THE MEETING

The meeting was opened by R. Gelfeld (Co-Chair) and was hosted by the Swedish Meteorological and Hydrological Institute (SMHI), Gothenburg, Sweden. Lars Andersson (Head of Data and Reporting) welcomed the participants on behalf on SMHI. L. Fyrberg outlined the local arrangements.

Members of the Study Group present were: P. Alenius (Finland), D. Collins (USA), R. Cramer (UK), J. Gagnon (Canada), R. Gelfeld (USA, Co-Chair), P. Haaring (Netherlands), A. Isenor (Canada, Co-Chair), R. Lowry (UK), N. Mikhailov (Russia), R. Olsonen (Finland), G. Reed (IOC Consultant), L. Rickards (UK), H. Sagen (Norway), R. Starek (US), J. Szaron (Sweden), and E. Vanden Berghe (Belgium). The ICES Secretariat was not represented at this meeting. Absent members included: C. Haenen (Netherlands), T. O'Brien (US), and B. Pelchat (Canada). Meeting participants not official members of the Group included M. Fichaut (France), K. Furukama (Japan), S. Sagan (Poland), T. Sakakibara (Japan), R. Schwabe (Germany), S. Belov (Russia), and M. Wichorowski (Poland). A complete list of names, addresses and contact points of participants can be found in Annex 1.

2 ADOPTION OF THE AGENDA

The agenda (Annex 2) for the Study Group meeting was adopted as a resolution at the 90th Statutory Meeting (C.Res. 2002/2CML). This is the second meeting of the Group.

3 REVIEW OF 2002 ACTION ITEMS

A. Isenor introduced a discussion focusing on the Action Items identified in the 2002 SGXML Report (for a list of acronyms, see Annex 3). Action items 1, 3, 4, 5 and 6 have been completed. These items included defining the elements and attributes of the parameter dictionary structure in written form (action 1), a DTD (action 3) and a schema (action 4). As well, the investigation involving the application of the Keeley Bricks to point data (action 5) and the initial development of a taxonomic brick (action 6) have been completed.

Action item 2 involved the mapping of parameter dictionaries to the SGXML defined structure. A total of 5 groups completed this mapping exercise, with each group introducing small variations on the defined structure to meet particular needs. It was noted that the Russian mapping was very different from the SGXML defined structure.

Action item 7 dealt with reviews of the Keeley bricks. No reviews were received. The development of a generalized ocean data model (action 8) including definitions for EDMED, MEDI and CSRs was not completed and thus the follow on reviews (action 9) were not completed. The mapping between EDMED, MEDI and CSR (action 10) and the follow-on review of the mapping (action 11) were also not completed. The development of an IOC Marine XML portal (action 12) was completed.

A. Isenor noted that the Group had completed about 50% of the action items as defined in the 2002 meeting report. Although not terribly impressive, the completion of those tasks oriented towards the parameter dictionary representations in XML does mean that considerable progress has been made on this topic. As well, the completion of the Keeley brick application in an XML environment for point data represents a considerable effort.

4 REVIEW OF IODE XVII

G. Reed presented **a review of SGXML discussions at IODE-XVII** (see Annex 4 for slide presentations). The IOC Committee on International Oceanographic Data and Information Exchange held its Seventeenth Session (IODE-XVII) at UNESCO Headquarters between 3 and 7 March 2003.

R. Gelfeld presented a report on First Session of SGXML at IODE-XVII. He outlined the terms of reference for SGXML and summarised the major lines of action and the proposed work plan for the Group.

The IODE Committee established a sessional working group tasked to:

(i) establish an informal group of interested data centres to evaluate the usefulness of the XML brick structure;

(ii) further develop the XML brick concept and extend the range of data structures to include, for example, time series datasets; and

(iii) explore how the brick structure maps to an object oriented database view of ocean data.

The IODE Committee recommended that an informal group of interested data centres be invited to work closely with the Canadian XML team on the further development of the XML brick concept and invited interested member states to evaluate the usefulness of the *XML Brick* structure and map their existing point data to the XML brick structure and report to the SGXML. The Committee also tasked the SGXML to consider at their next meeting to describe their '*vision*' of XML and to report back to IODE-XVIII.

The representative of FAO informed the Session about the FIGIS (Fisheries Global Information System) project that had been involved in the development of an XML-based mark-up language related to fisheries. The committee proposed IOC and FAO collaborate in the field of XML and invited FAO to participate in the SGXML. The representative of WMO and the JCOMM co-president expressed strong interest in cooperation with the SGXML as this issue is most relevant to some of their activities.

G. Reed then described the MarineXML community portal site that has been established by IODE to provide a discussion forum for Marine XML activities. This is a dynamic web site allows members to submit items for inclusion on the site. The MarineXML site is at http://marinexml.net

5 EU PROJECT REVIEW AND AODC ACTIVITIES

G. Reed also provided a review of the **EU Marine XML project**. This project, which commenced in February 2003, aims to demonstrate that XML technology can be used to improve data interoperability for the marine community. MarineXML will realise a prototype Marine Markup Language (MML). The project has identified four key project objectives:

- Objective 1. To produce a prototype marine data ontology framework for interoperability that will align a conceptual Marine Mark-Up Language (MML) specification with other XML and related standards. This will be achieved through research into a marine data ontological framework.
- Objective 2. To produce working demonstrations of the data interoperability framework by developing a working prototype test bed of the interoperability framework that will test and demonstrate the functionality and interoperability of heterogeneous datasets from disparate agencies. ECDIS will be the basic data management platform of these demonstration prototypes.
- Objective 3. To develop a prototype MML specification by using the outputs of the demonstration in Objective 2, the interoperability framework will be further developed as the basis of the MML specification. While the project will not develop the full MML specification, it will produce and document an initial framework.
- Objective 4. To advance the standardisation of a Marine Mark-up Language to ensure the standardisation of a MML by liaising with the SGXML during the project and provide support and advice to further the standardisation process beyond the end of the project.

The work plan for the project is divided into six work packages (WP):

- WP 1 - Coordination. Objectives are to ensure the overall progress of the project according to the contract signed with the European Commission and to maintain a continuous and effective interface with the European Commission.
- WP 2 - Exploitation and Dissemination. Objectives are to disseminate the developments and findings of MarineXML to interested stakeholders and organizations, to develop an exploitation plan for identified exploitable project deliverables, and to ensure post-project development and standardisation of a MML.
- WP 3 - Standardisation Review. Objectives are to create alignment between the work of MarineXML and existing and developing standards, reducing the risk and effort of developing MarineXML by adopting appropriate prior and ongoing work. This will be achieved by reviewing existing approaches to XML-based interoperability to identify appropriate basic frameworks and examining methods to ensure standardisation of a MarineXML design process.
- WP 4 - Framework Development. Objectives are to determine how ontologies can be used to support interoperability between XML structures and other standards by developing a generic model for an ontologies based interoperability framework and providing the theoretical framework for MML.
- WP 5 - MML Test-bed demonstration. Objectives are to develop a test bed application based on a GOOS relevant data flow, to demonstrate improvements in the ability to access data across disparate 'formats' and extract common elements (fields) using the initial MML specification, and to demonstrate practicality and efficiencies of MML structures through the automation of data access, data flow and data processing activities.
- WP 6 - MML Specification Development. Objectives are to utilise the ontological structures developed in WP 5 and create a specific marine version as the basis of an embryonic MML specification in XML, to build the necessary tools and capabilities to implement and use MML, specifically for the test bed applications.

The deliverables of the EU Marine XML Project are a working test bed, an outline MML specification and a route-map for development post project.

G. Reed invited any interested organisations to join the project as affiliate members. More information about the project can be found at http://marinexml.net.

G. Reed, representing the **Australian Oceanographic Data Centre (AODC)**, presented an evaluation of the *XML Brick* structure. AODC believes that XML should be self-describing and as such the XML file should be self-explanatory for a common user. Any proposed XML file should avoid, as much as possible, codes, abbreviations, look up tables etc. It was noted that the use of NetCDF format is increasing in the marine community. Within Australia, NetCDF is the most common format in BoM and CSIRO, and global programs such as WOCE and Argo have adopted this format. An XML encoding has been developed as a vehicle for data in NetCDF format. NcML is a schema for an XML metadata file that describes the NetCDF file contents (e.g., dimensions, variables, etc.). AODC expressed a need for a Common Data Dictionary and it was anticipated that SGXML would develop the common dictionary structure for all marine data and define the mandatory bricks for all oceanographic data. ISO standards should be adopted for all data in XML format and it was noted that the proposed Keeley bricks for date-time and position do not accord with ISO standards 8601 and 6709 respectively.

The Keeley XML file size is a concern and should be considered as an important issue along with database storage, data retrieval, searching, querying and sub-setting. It was suggested that one option could be to have the metadata in XML file structure packed in a top level, separate from the data. As an example, AODC has created an XML file for XBT profile data. A cruise comprising six full profiles has been generated in AODC internal XML format and five data points from the first 3 profiles of the same cruise generated in the proposed "Keeley Bricks" XML format. A general cruise/station metadata consisting of 65 lines (1 depth-pressure (DEPTH) consists of 17 lines, 1 data point (TEMP) consists of 17 lines) would result in file sizes of approximately 3,500 lines for a shallow (100 m) XBT depth-temperature profile and approximately 15,700 lines for a typical (460 m) XBT depth-temperature profile.

The SGXML expressed their appreciation for the evaluation of the XML brick structure. This feedback will be valuable for any future development of the Brick concept.

6 A VISION FOR SGXML

R. Gelfeld introduced a discussion on a vision for this Study Group. A. Isenor and R. Gelfeld had, several weeks ago, distributed possible options for the Group to consider.

The discussion that followed identified several key points for the Group to consider over the next two days. It was noted that at ICES Study Group has a limited duration, and that under the present situation, this Group should be considering a two year lifespan. It was also noted that others had criticized the Group as being excessively technical. Many in the ocean data community do not understand XML and it appears that we have not adequately explained the benefits of XML.

It was noted that many of the technical details of XML should be hidden from the users. The users should see the benefit of XML in the tools and applications, not in the details of the XML structure.

The Group also recognised the need for standards, but more importantly, the adoption and use of the standard by the participants. Other efforts (e.g., the NERC DataGrid) are looking at standards, and it was recognised that a standard in the topic area of ocean metadata does not exist.

Ultimately, the Group is looking to XML as a means of enhancing interoperability. Some thought interoperability could be simply summarized as conformance to standards.

7 OTHER XML PRESENTATIONS

A series of short presentations were given to update participants on various XML developments in other organisations. These presentations, which predominantly served to initiate a series of discussions, are summarised below (summaries are delimited by the bold text). Slide presentations are provided in Annex 4.

K. Furukawa and T. Sakaibara jointly presented XML activities of the **Japanese National Institute for Land and Infrastructure Management and University of Tokyo Center for Spatial Information Service**. This work is related to water quality impacts on biological activity in Tokyo Bay. Here the problem was a multitude of data collectors all providing water quality sampling data in various spreadsheet formats. A spreadsheet based tool was developed that allowed the collectors to digitize their water quality sampling data in the spreadsheet, and by clicking a button, provided an output XML document in a standard structure. The XML document is then passed to the National Institute. In this way, the Institute now deals with a single XML structure rather than a multitude of spreadsheet structures.

R. Lowry introduced a discussion on **parameter dictionaries**. Lowry believes the community should consider a dictionary as dynamic, not static. He noted that many organisations within SGXML had generated the XML that contained the single organizations dictionary, but merging between XML instances is a problem. Many IT personnel consider this merging as a knowledge management problem, which means ontology. Although the term ontology is not fully understood within SGXML, it is used in this context to mean the knowledge required for a mapping of terms between dictionaries.

For such a mapping, the problems include the syntax. Aggregation software needs to be developed. Then the problem becomes the maintenance of the ontology. Finally, it was noted that others have adopted the BODC dictionary. However, this also introduces problems such as the procedures related to correcting errors in the dictionary and the propagation of those corrections to the user community.

R. Starek presented activities of the **US Navy**. It was noted that the Meteorological and Oceanographic (METOC) section exchanges data twice a year with US NODC. The XML related project now nearing completion involves modelling, remote sensing, data collection and the fusion of these data types into products. The XML structures are online at the Dept. of Defence Repository. A central entry site can be found at http://pao.cnmoc.navy.mil.

The data modelling for this project was conducted in ErWin. The created entity-relations model can be used to automatically create the XML structures using XMLAuthority software. The data requests to the system may be single requests, or subscription based and is implemented in a web services paradigm using SOAP and WSDL. Standard metadata may be outputted as a FGDC, or ISO 19115, etc. depending on user request. The retrieval methods are either synchronous or asynchronous and some data are returned as XML using the Joint METOC Brokerage Language Response. In this system, information discovery is not a problem that this system needs to solve as the users are not browsing for data, but rather have specific requests.

N. Mikhailov and S. Belov presented Russian NODC XML efforts related to the **distributed data centre concept**. For this work, the need for a unified parameter dictionary (UPD) was stressed as well as the need for a common model of ocean data. In the Russian model of a distributed ocean data centre, a user defines a request for data. The request is sent to navigator software, which has a catalogue available of the distributed data centre holdings. This catalogue, or information resource (IR), is used to identify those data centres that could supply data for the request, and sends individual requests to interrogator software at those centres. The interrogator software receives the request and queries local databases. The return set from the local databases is packaged in the interrogator, and then returned to the navigator. In this model, the development of a common exchange format is one problem.

Some of the resources for the described model are now available over the Internet. S. Belov demonstrated the available components. A dataset was selected from a web page that represented the IR. At the moment, the IR is created manually, when in the final product it will be generated by the interrogators. The IR is an XML catalogue.

In the demonstration, the navigator was running on the local host machine. The data selection was sent over the Internet to the interrogator on a computer in Obninsk. The request is based on the IR, which is an XML catalogue. Ultimately, the request can be based on x,y,z,t specification rather than a data set level request, which was demonstrated here. The interrogator processed the request on the computer in Obninsk, and returned the dataset to the local host computer.

A. Isenor then presented the results of the **Canadian application of the Keeley Bricks** in an XML environment (a brief report is given in Annex 5). The Canadian efforts concentrated on defining those bricks necessary for building the profile data structure. This structure, which is more correctly for 1-diminsional data, (i.e., data with one independent variable such a z for profile data) was used to contain CTD, XBT and float data. Full definitions of the bricks and components were developed. The rules defined during the development were presented and some interesting details of the implementation where given. All report, schema, example profile XML documents and brick definitions are available on the MEDS web site at:
http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/ICES/web%20xml/SSF-xml.htm

Some members wondered if the EU Maine XML Project should consider the bricks as part of the standards review, recognizing that the bricks are not really a standard. Most members thought other standards should be used or incorporated into the bricks. An example could be GML. However, the detail in GML was recognized as a potential problem (e.g., the GML schema is about 700 pages in length). It was also commented that one value to the brick effort is the incorporation of an ocean data perspective into a structured environment. It was noted that the 1-dimensional case presented here would include water level data. In this case, datums would need to be contained in the structure. It was thought that the datum information could be contained without altering the XML structure.

D. Collins presented the Group with information on **metadata standards**. It was noted that the ISO 19115 standard was approved in March 2003. There are many available standards, and mappings do exist between the standards. However, some were sceptical whether or not the mappings actually work properly.

T. Sakakibara informed the Group that he was a member of the TC 211 committee that produced the ISO 19115 standard. R. Starek noted that when considering the interchange of metadata, the Group should be aware of SEDRIS, a modelling and simulation synthetic environment interchange model. SEDRIS™ technologies provide the means to represent environmental data (terrain, ocean, air and space), and promote the unambiguous, loss-less and non-proprietary interchange of environmental data. http://www.sedris.org/. As well, UDDI web service may be useful to consider. The Universal Description, Discovery and Integration (UDDI) protocol is one of the major building blocks required for successful Web services. UDDI is the language in which Web services may be described. The UDDI descriptions then represent a data source for web searching of available services. UDDI enables companies and applications to dynamically find and use Web services over the Internet (http://www.uddi.org/).

8 SUB-GROUP ACTIVITIES

It was evident that when considering XML in an ocean data context, this Study Group has been focusing on three areas of interest: parameter dictionaries, metadata and data. To expedite and provide focus for these activities, it was decided to split the participants into three smaller subgroups and explore these topics in detail. Participants were allowed to join the sub-group of their choice.

A facilitator was appointed for each sub-group. The facilitators were:

Metadata – D. Collins (15)
Parameter Dictionary – R. Lowry (4)
Data – A. Isenor (4)

The number in brackets following the facilitators name indicates the number of participants in that sub-group.

Participants were asked to think about the future direction of SGXML, and in particular the goals and objectives of SGXML for the particular topic area of the sub-group. Then, identify action items appropriate for the goals and objectives.

8.1 SGXML Vision Statement

The sub-groups identified goals and objectives to varying degrees. However, after reviewing the provided information for the sub-groups, a general vision starts to emerge.

The SGXML Vision may be stated as follows:

The ICES/IOC SGXML will utilize or establish international standards to promote the seamless exchange of data from distributed data sources, by using a single parameter dictionary, well-defined and explicitly tagged metadata, and a common XML data structure, packaging all content and providing to the client datasets and software tools that are platform independent or web enabled.

It was thought that the exchange of data using XML would be limited by file size implications. Although compression can reduce the file sizes during the transfer, the processing requires uncompressed files. It was noted by the data group that this might be a problem for exchanges between large data centres. However, it was pointed out that this is the old paradigm of transfer, and that the frequency of transfer also impacts the size of the transferred files. A more frequent transfer would reduce the file sizes.

9 ACTION ITEMS RESULTING FROM THE MEETING

The meeting structure was predominantly a series of discussions. Considering this, it was considered better to summarise the Action Items in a single section as during the discussion the items were only loosely defined. Here, the Action Items are presented under four topic headings.

9.1 Parameter Dictionary

Action 1: *R. Lowry will post a version of the BODC dictionary on the marinexml.net site. This will be used to establish mappings from BODC dictionary to the following dictionaries:*

RNODC dictionary – N. Mikhailov
USNODC dictionary – R. Gelfeld
PANGAEA dictionary – R. Lowry
Canadian MEDS dictionary – J. Gagnon
USJGOFS dictionary – R. Lowry
DOD (German) dictionary – R. Schwabe
SMHI dictionary – J. Szaron
Netherlands dictionary – P. Haaring
GCMD dictionary – E. Vanden Berghe
IFREMER dictionary – M. Fichaut

Action 2: *R. Cramer will construct a web interface for accessing the BODC dictionary.*

Action 3: *R. Cramer will compare and reconcile the parameter dictionary XML structures as defined by the DTD and schema.*

9.2 Point Data Investigation

Action 4: *G. Reed will formally request to the EU Marine XML project that the scope and content of the bricks (not the XML syntax) be reviewed in the standards review process.*

Action 5: *A. Isenor will determine which parts of the bricks can be substituted with components from OWS, GML and other accepted international standards.*

Action 6: *J. Gagnon will identify and construct the ocean cruise oriented bricks.*

Action 7: *A. Isenor and E. Vanden Berghe will attempt application of the brick / XML structure to 3-d data (e.g., net tow) and identify lacking bricks.*

9.3 Metadata Investigation

Action 8: *Define common terminology for metadata. (R. Starek and N. Mikhailov)*

Action 9: *Create a reference model for the abstraction of metadata (R. Starek and N. Mikhailov)*

Action 10: *Evaluate existing metadata standards by examining ISO19115 to identify elements specific to ocean community needs (R. Starek and N. Mikhailov)*

Action 11: *Complete a comparison mapping of CSR (M. Fichaut), MEDI (G. Reed), EDMED (L. Rickards), USNODC DDF (D. Collins) to the ISO 19115. This is an update of the T. Sakakibara spreadsheet that listed ISO 19115 elements.*

Action 12: *Evaluate the catalogue standard ISO 19110 for application to ocean datasets (E. Vanden Berghe and D. Collins)*

Action 13: *Initiate development of an optimal metadata tag list. (G. Reed)*

9.4 Other Items

Action 14: *G. Reed will create categories on the marinexml.net site for the Working subgroups.*

Action 15: *A. Isenor will check SGXML Yahoo site for content that should be moved to the marinexml.net site. The marinexml.net site will now be used as the forum for communication.*

Action 16: *G. Reed will identify proper procedures for adding new members to the SGXML within the IOC community of countries.*

Action 17: T. Sakakibara will provide G. Reed with the spreadsheet software for in-the-field water sample collection and data reporting in XML.

10 ANY OTHER BUSINESS

G. Reed noted that IOC would be very interested in obtaining the software demonstrated by the K. Furukawa and T. Sakakibara. This software has potential application in developing countries. T. Sakakibara agreed to provide this software.

It was suggested that the next meeting be held on 6–7 May 2004 at Oostende, Belgium. R. Gelfeld and A. Isenor thanked E. Vanden Berghe for volunteering to host the next meeting.

The proposed 2002/2003 Terms of Reference for the Study Group were briefly discussed and are presented here in Annex 6.

11 MEETING CLOSURE

A special thank you was extended to the new participants that considered the activity of the Study Group sufficiently important to attend. R. Gelfeld and A. Isenor thanked the Swedish hosts for extending warm hospitality to the Group while in Gothenburg. R. Gelfeld then closed the meeting by thanking all of those who had participated.

ANNEX 1: NAMES AND ADDRESSES

Names, addresses and contact points for participants.

Alenius, Pekka
Finnish Institute of Marine Research,
P.O. Box 33, (Lyypekinkuja 3),
00931 Helsinki,
Finland
Tel (operator): +358 9 613 941
Tel (direct): +358 9 613 94439
Fax: +358 0 61394494
e-mail: pekka.alenius@fimr.fi
Web page: http://www2.fimr.fi/ or www.fimr.fi

Belov, Sergey – *IOC*
Russian National Oceanographic Data Centre
6 Korolev St., Obninsk, Kaluga Region,
Russian Federation, 249020
e-mail: belov_sergey@meteo.ru

Collins, Donald W. – *IOC*
U.S. National Oceanographic Data Center
1315 East West Highway, 4th Floor
Silver Spring MD, 20910, U.S.A.
Tel: +1 301 713 3275 extn 154
Fax: +1 301 713 3302
e-mail: donald.collins@noaa.gov

Cramer, Ray (see note *)
British Oceanographic Data Centre,
Proudman Oceanographic Laboratory,
Bidston Observatory, Prenton,
Merseyside, CH43 7RA,
United Kingdom
e-mail: rnc@bodc.ac.uk
Web page: http://www.bodc.ac.uk

Fichaut, Michele – *MDM*
Institut Francais pour le Recherche et
l'Exploitation de la Mer (IFREMER),
Center de Brest,
Departement IDM,
BP 70,
29280 Plouzane
France
Tel: 33-2-98-22-6663
e-mail: michele.fichaut@ifremer.fr

Furukawa, Keita – *IOC*
National Institute for Land and
Information Management
Ministry of Land, Infrastructure and Transport
3-1-1, Nagase,
Yokosuka 239-0826, Japan
Tel: +81 (468) 44 5023
Fax: +81 (468) 44 1145
e-mail: furukawa-k92y2@ysk.nilim.go.jp
Web page: http://www.nilim.go.jp

Gagnon, Jean – *MDM*
Marine Environmental Data Service (MEDS),
Department of Fisheries and Oceans,
200 Kent Street, 12th Floor,
Ottawa, Ontario K1A OE6,
Canada
Tel: +1 613 990-0260
Fax: +1 613 993-4658
e-mail: gagnonj@dfo-mpo.gc.ca
Web page: http://www.meds-sdmm.dfo-mpo.gc.ca/

Gelfeld, Robert D. (**Co-Chair**)
U.S. National Oceanographic Data Center/
World Data Center - A Oceanography,
1315 East West Highway, 4th Floor,
Silver Spring MD, 20910-3282,
U.S.A.
Tel: +1 301 713 3295 extn 179
Fax: +1 301 713 3303
e-mail: robert.gelfeld@noaa.gov
Web page: http://www.nodc.noaa.gov

Haaring, Pieter - *Observer*
National Institute for Coastal and Marine Management (RIKZ)
P.O. box 20907
2500 EX The Hague
The Netherlands
Telephone direct: +31703114570
Telephone general: +31703114311
Fax direct: +31703114500
Fax general: +31703114321
e-mail: p.a.haaring@rikz.rws.minvenw.nl
Web Page: http://www.rikz.nl

Isenor, Anthony W. (**Co-Chair**)
Defence R&D Canada - Atlantic
PO Box 1012, 9 Grove Street
Dartmouth, Nova Scotia
Canada B2Y 3Z7
Ph: (902) 426-3100 ext. 106
Fax: (902) 426-9654
e-mail: anthony.isenor@drdc-rddc.gc.ca

Lowry, Roy (see note *)
British Oceanographic Data Centre,
Proudman Oceanographic Laboratory,
Bidston Observatory, Prenton,
Merseyside, CH43 7RA,
United Kingdom
Tel: +44 151 653 1519
e-mail: rkl@bodc.ac.uk
Web page: http://www.bodc.ac.uk

Mikhailov, Nicolay – *IOC*
Russian National Oceanographic Data Centre
6 Korolev St., Obninsk, Kaluga Region,
Russian Federation, 249020
Tel: 7-084-397-49-07
Fax: 7-095-255-22-25
e-mail: nodc@meteo.ru

Olsonen, Riitta – *MDM*
Finnish Institute of Marine Research,
P.O. Box 33, (Lyypekinkuja 3),
00931 Helsinki,
Finland
e-mail: riitta@fimr.fi
Web page: http://www2.fimr.fi/ or www.fimr.fi

Reed, Greg – *IOC*
Intergovernmental Oceanographic Commission (IOC),
1 Rue Miollis
75732 Paris Cedex 15
France
Tel: 01 45 68 3960
e-mail: g.reed@unesco.org

Rickards, Lesley (see note *)
British Oceanographic Data Centre,
Proudman Oceanographic Laboratory,
Bidston Observatory, Prenton,
Merseyside, CH43 7RA,
United Kingdom
Tel: +44 151 653 1514
Fax: +44 151 652 3950
e-mail: ljr@bodc.ac.uk
Web page: http://www.bodc.ac.uk;
http://www.oceannet.org

Sagen, Helge
Institute of Marine Research
Norwegian Marine Data Centre
Nordnesgt 50, 5817, Bergen
Norway
Tel: 47 55 23 8500
e-mail: helge.sagen@imr.no

Sagan, Slawomir (WG) – *MDM*
Institute of Oceanology PAS,
Powstancow Warszawy 55,
81-712 Sopot,
Poland
Tel: +(48 58) 5517 283 x211,
Fax: +(48 58) 5512 130
e-mail: sagan@iopan.gda.pl
Web page: http://www.iopan.gda.pl

Sakakibara, Tsuneki - *Observer*
University of Tokyo
Center for Spatial Information Science
Cw-502, 4-6-1, Komaba, Meguro-ku
Tokyo 153-8505,
Japan
Tel and Fax: +81-3-5452-6417
e-mail: sakaki@skl.iis.u-tokyo.ac.jp
Web page: http://shiba.iis.u-tokyo.ac.jp/sakaki

Schwabe, Reinhard – *MDM*
Deutsches Ozeanographisches Datenzentrum (DOD),
Bundesamt für Seeschiffahrt und Hydrographie
Bernhard-Nocht-Str. 78
D-20359 Hamburg, Germany
e-mail: reinhard.schwabe@bsh.de
Web page:
http://http://www.bsh.de/Oceanography/DOD/htm

Starek, Robert
Commander, Naval Meteorology and Oceanography Command
1100 Balch Blvd., Code N522
Stennis Space Center, MS 39529-5005
USA
DSN Phone: 828-5993
Commercial Phone: 228-688-5993
e-mail: starekr@cnmoc.navy.mil

(*) After September 15, 2003 the contacts will be:

British Oceanographic Data Centre,
Proudman Oceanographic Laboratory,
6 Brownlow Street,
Liverpool L3 5DA,
United Kingdom.

Phone Numbers:

Cramer: +44 151 795 487
Lowry: +44 151 795 4895
Rickards: +44 151 795 4897

Szaron, Jan
Swedish Meteorological and Hydrological Institute,
Oceanographic Services,
Nya Varvet 31,
SE - 426 71 Vastra Frolunda,
Sweden
Tel: +46 (0)31 751 8971
Fax: +46 (0)31 751 8980
e-mail: jan.szaron@smhi.se
Web page: http://www.smhi.se

Vanden Berghe, Edward – *IOC*
Manager, Flanders Marine Data and Information Centre
Flanders Marine Institute
Vismijn, Pakhuizen 45-52, B-8400 Ostend, Belgium
Tel: +32 59 342130
Fax: +32 59 342131
E-Mail wardvdb@vliz.be
Web Page: http://www.vliz.be

Wichorowski, Marcin
Institute of Oceanology PAS, Powstancow Warszawy 55,
81-712 Sopot,
Poland
Tel: +(48 58) 55 72 81,
Fax: + (48 58) 51 21 30
e-mail: wichor@iopan.gda.pl
Web page: http://www.iopan.gda.pl

ANNEX 2: 2002/2003 TERMS OF REFERENCE

2CML An **ICES-IOC Study Group on the Development of Marine Data Exchange Systems using XML [SGXML]** (Co-Chairs: R. Gelfeld, U.S.A. and A. Isenor, Canada) will meet in Gothenburg, Sweden from 26-27 May 2003 to:

a) create, evaluate and discuss inter-sessional work on parameter dictionaries including the population of the dictionary for distribution via a defined XML structure;

b) evaluate inter-sessional work on point data structure;

c) evaluate the usefulness of the generalised Keeley brick approach with application to various point data types;

d) report on the investigation into other available existing standards (e.g. geographers through the Open GIS consortium, taxonomy, ISO standards, metadata standards (MEDI, GFDC, EDMED, etc), utilising what has already been built;

e) evaluate inter-sessional work on metadata;

f) evaluate the usefulness of linkages to other metadata standards and on the implications of a generalised metadata model to existing models;

SGXML will report by 30 April 2003 for the attention of the Oceanography Committee.

Supporting Information

Priority:	The future of marine data management, processing and exchange in an interoperable environment lies in the use of virtual data systems and exploiting web technologies. If ICES does not participate in these developments its ability to receive, process and disseminate data in the form required by the user community will be negatively impacted.
Scientific Justification:	a) The XML web distribution of the parameter dictionaries should be completed and the usefulness of the exercise for cross mapping of parameter dictionaries needs to be assessed. The applicability of the XML structure for other dictionaries should also be determined b,c) The generalised point data structure needs to be critically evaluated from the perspective of the international data centres. The applicability of the abstract Keeley bricks needs to be evaluated d) The metadata problem is common to many organisations and considerable effort has been made by these other organisations. The usefulness of these efforts needs to be evaluated within the context of ocean data transfer e,f) Progress on the generalisation of the metadata model needs to be evaluated. The generalised model needs to be considered within the context of existing models
Relation to Strategic Plan:	The Group is set up to provide members of the ICES scientific community, efficiently and effectively, with the support they need to meet the scientific goals. The ICES Vision goes far beyond the capacities of any single organization. Networking in partnership with Member Countries, other IGOs, and scientific NGOs will enable ICES to foster valuable cooperation, coordination and collaboration. By so doing, ICES will not duplicate activities already carried out by others. Rather, this will provide a new, value-added dimension. Enhanced interdisciplinary knowledge and networking will benefit the entire science community.
Resource Requirements:	No specific resource requirements beyond the need for members to prepare for and participate in the meeting.
Participants:	Participation in the XML Study Group is open to any individual or group, internal or external to ICES.
Secretariat Facilities:	None.
Financial:	None specific.
Linkages To	This is important to work on data integration which is of direct interest to ACE

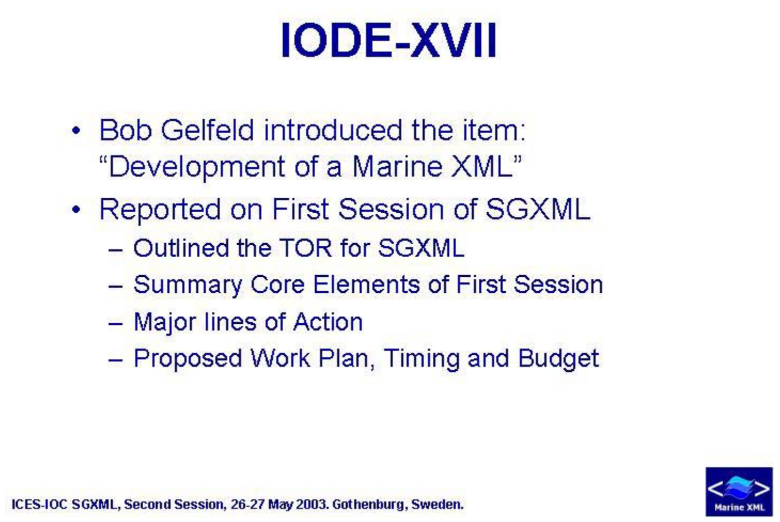
Advisory Committees:	
Linkages To other Committees or Groups:	The ICES working group on marine data management (WGMDM).
Linkages to other Organisations	The WMO/IOC Joint Commission on Oceanography and Marine Meteorology (JCOMM) has also expressed an interest in XML. A Marine Consortium is currently being formed to address XML internationally (IODE, 2000). The Consortium's goal is to develop a free and open specification for a Marine XML that will be used in all exchanges of ocean data. ICES has been asked to become a Consortium member. Several ICES countries have joined or are about to join the Consortium (Belgium (Flanders), Netherlands, UK). In addition IOC/IODE is about to join and EuroGOOS is considering membership.
Cost Share	ICES 100%

ANNEX 3: LIST OF ACRONYMS AND TERMS

Acronym or Term	Description
AODC	Australian Oceanographic Data Centre
Argo	The array for real-time geostrophic oceanography
BODC	British Oceanographic Data Centre
BoM	Commonwealth Bureau of Meteorology
CSR	Cruise Summary Report
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DDF	Data Documentation Form (NODC)
DIF	Directory Interchange Format
DTD	Document Type Definition
ECDIS	Electronic Chart Display and Information Systems
EDMED	European Directory of Marine Environmental Data
FGDC	Federal Geographic Data Committee (USA)
FIGIS	Fisheries Global Information System
GCMD	Global Change Master Directory
GIS	Geographical Information System
GML	Geography Markup Language
GOOS	Global Ocean Observing System
ICES	International Council for the Exploration of the Sea
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange
IR	Information Resource
ISO	International Organisation for Standardization
JCOMM	Joint Commission on Oceanography and Marine Meteorology
MEDI	IOC Marine Environmental Data Information Referral Catalogue system
MARC	MAchine Readable Cataloguing record
MEDS	Marine Environmental Data Services - Canada
MML	Marine Markup Language
NERC	Natural Environment Research Council
NODC	U.S. National Oceanographic Data Centre
OGC	Open GIS Consortium
OWS	OGC Web Services
RNODC	Russian National Oceanographic Data Centre
SEDRIS	Source for Environmental Representation and Interchange
SGXML	ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML
SMHI	Swedish Meteorological and Hydrological Institute
SOAP	Simple Object Access Protocol
TOR	Term of Reference
UDDI	Universal Description, Discovery and Integration
WDCA	World Data Centre for Oceanography/Silver Spring
WGMDM	Working Group on Marine Data Management
WMO	World Meteorological Organisation
WOCE	World Ocean Circulation Experiment
WOD	World Ocean Database
WSDL	Web Services Definition Language
XML	Extensible Markup Language
XSL	Extensible Stylesheet Language

ANNEX 4: PRESENTATION SLIDE SUMMARIES

SGXML discussions at IODE-XVII: G. Reed



IODE-XVII

- The committee established a sessional working group tasked:
  i. to establish an informal group of interested data centres to evaluate the usefulness of the XML brick structure;
  ii. to further develop the XML brick concept and extend the range of data structures to include, for example, time series datasets; and
  iii. to explore how the brick structure maps to an object oriented database view of ocean data.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

IODE-XVII

- The Committee recommended that an informal group of interested data centres be invited to work closely with the Canadian XML team on the further development of the XML brick concept.
- The Committee invited interested member states to evaluate the usefulness of the *XML Brick* structure and map their existing point data to the XML brick structure and report to the SGXML.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

IODE-XVII

- The Committee tasked the SGXML to consider at their next meeting to describe their '*vision*' of XML and to report back to IODE-XVIII

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

IODE-XVII

- The representative of FAO informed the Session about the FIGIS (Fisheries Global Information System) project that had been involved in the development of an XML-based mark-up language related to fisheries.
- Proposed IOC and FAO collaborate in the field of XML. The Committee invited FAO to participate in the SGXML

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

IODE-XVII

- Both the representative of WMO and the JCOMM co-president expressed strong interest in cooperation with the SGXML as this issue is most relevant to some of their activities

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

IODE-XVII

- IODE Programme and Budget for 2003/2005 includes $5,000 each year from RP funds to support SGXML

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

marineXML.net

- IODE has established a community portal site to provide a discussion forum for Marine XML activities
- This is a dynamic web site which means that members can submit items for inclusion on the site
- Features of the Marine XML site are:

marineXML.net

Uses "knowledge object" concept

- 5 different types of knowledge that can be used:
  - i. html pages with embedded imagery and hyperlinks;
  - ii. documents that can be uploaded to the server by the author and downloaded by users. This will constitute a searchable e-library;
  - iii. links to external web sites or web pages;
  - iv. events: a list of upcoming events displayed on the homepage with links to internal or external web sites or web pages;
  - v. forum: web-based discussion threads on the relevant subject category.

marineXML.net

Anyone can register to become a member of the site

- Simple registration process enables a user to save his or her details and become a member of the site
- Encourages community involvement and participation

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

marineXML.net

All members of the community can submit content

- Once a member has registered, he or she is able to submit content to the site.
- The site editor receives notification and can approve the content

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

marineXML.net

Members are informed of additions to the site

- When new content is added to the site, members can choose to receive an email that will notify the title of the item and a link to the page.
- You do not have to keep visiting the site to see if anything new has been added

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

EU Marine XML Project: G. Reed

EU Marine XML Project

Using XML technology for marine data interoperability

Greg Reed, IOC

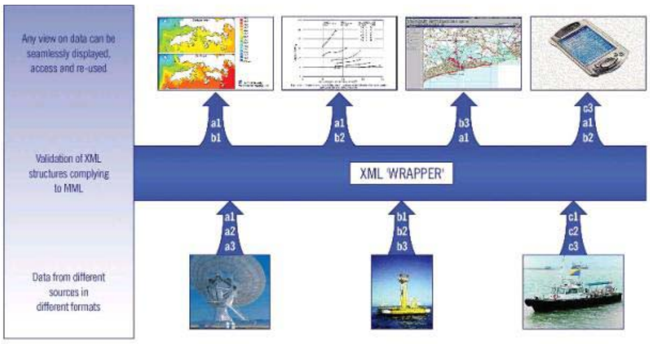
ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Overview

- Aim of MarineXML is to demonstrate that XML technology can be used to improve data interoperability for the marine community
- MarineXML will realise a prototype Marine Markup Language

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

MML Concept



Objectives of the Project

- MarineXML has identified four key project objectives:
  1. To produce a prototype marine data ontology framework for interoperability
  2. To produce working demonstrations of the data interoperability framework
  3. To develop a prototype MML specification
  4. To advance the standardisation of a Marine Mark-up Language

Objectives of the Project

- Objective 1. To produce a prototype marine data ontology framework for interoperability
  - This task will align a conceptual Marine Mark-Up Language (MML) specification with other XML and related standards. This will be achieved through research into a marine data ontological framework.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Objectives of the Project

- Objective 2. To produce working demonstrations of the data interoperability framework
  - Develop a working prototype test bed of the interoperability framework that will test and demonstrate the functionality and interoperability of heterogeneous data sets from disparate agencies. ECDIS will be the basic data management platform of these demonstration prototypes.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Objectives of the Project

- Objective 3. To develop a prototype MML specification
  - Using the outputs of the demonstration in Objective 2, the interoperability framework will be further developed as the basis of the MML specification. While the project will not develop the full MML specification, it will produce and document an initial framework.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Objectives of the Project

- Objective 4. To advance the standardisation of a Marine Mark-up Language
  - MarineXML will seek to ensure the standardisation of a MML by liaising with the SGXML during the project and provide support and advice to further the standardisation process beyond the end of the project.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Project Partners

1. HR Wallingford (UK)
2. Marine Information Council (UK)
3. SevenCs (Germany)
4. Nansen Environmental and Remote Sensing Centre (Norway)
5. Central Laboratory of the Research Council (UK)
6. Rijkswaterstaat (Netherlands)
7. Flemish Marine Institute (Belgium)
8. Social-change On-line (Australia)
9. Swedish Metrological and Hydrological Institute (EuroGOOS)
10. IODE (Intergovernmental)

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Project Workplan

- WP 1 Co-ordination (HRW)

  Objectives

  - Ensure the overall progress of the project according to the contract signed with the European Commission.
  - Maintain a continuous and effective interface with the European Commission.
  - Establish a pre-project consortium agreement and post-project exploitation agreement.
  - Steer the project within allotted time and budget to ensure completion of the project deliverables.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Project Workplan

- WP 2 Exploitation and Dissemination (IODE)

  Objectives

  – Disseminate the developments and findings of MarineXML to interested stakeholders and organisations.

  – Develop an Exploitation Plan for identified exploitable project deliverables.

  – Ensure post-project development and standardisation of a MML

Project Workplan

- WP 3 Standardisation Review (NERSC)

  Objectives

  – To create alignment between the work of MarineXML and existing and developing standards, reducing the risk and effort of developing MarineXML by adopting appropriate prior and ongoing work.

  – To review existing approaches to XML-based interoperability to identify appropriate basic frameworks.

  – To examine methods to ensure standardisation of a MarineXML design process.

Project Workplan

- WP 4 Framework Development (VLIZ)

  Objectives
  - To determine how ontologies can be used to support interoperability between XML structures and other standards
  - Develop a generic model for an ontologies based interoperability framework
  - Provide the theoretical framework for MML

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Project Workplan

- WP 5 MML Test-bed demonstration (7CS)

  Objectives
  - Develop a test bed application based on a GOOS-relevant data flow.
  - Demonstrate improvements in the ability to access data across disparate 'formats' and extract common elements (fields) using the initial MML specification.
  - Demonstrate practicality and efficiencies of MML structures through the automation of data access, data flow and data processing activities.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Project Workplan

- WP 6 MML Spec. development (CCLRC)

  Objectives
  - To utilise the ontological structures developed in WP 5 and create a specific marine version as the basis of an embryonic MML specification in XML
  - Build the necessary tools and capabilities to implement and use MML, specifically for the test bed applications

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Project Deliverables

- Delivery
  - working test bed, outline MML specification and a route-map for development post-project.
- Principles in realising deliverables
  - Work with on-going initiatives
  - Work with standards bodies
  - Ensure post-project sustainability

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

Progress

- Kick-off Meeting held 16-19 February 2003
- Technical Meeting to discuss WP3/4 on 20th March 2003 at VLIZ, Oostende
- Project Meeting planned for 19-20 June 2003 at 7Cs Hamburg

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

AODC Comment on Keeley Bricks: Greg Reed

AODC Comments to the XML Bricks document by Robert Keeley, Anthony Isenor, Joe Linguanti, dated 20 January 2003

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

1. XML should be self-describing and as such the XML file should be self-explanatory for a common user. The proposed XML file should avoid, as much as possible, codes, abbreviations, look up tables etc.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

2. Use of netCDF format is increasing in the marine community.
   - Within Australia netCDF is the most common format in BoM and CSIRO.
   - Global programs such as WOCE and Argo have adopted it.
   - An XML encoding (NcML) has been developed as a vehicle for data in netCDF format.

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

3. Need for a Common Data Dictionary.
   - It is anticipated that SGXML will develop the common dictionary structure for all marine data and define the mandatory bricks for all oceanographic data

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

4. ISO standards should be adopted for all data in XML format.
   – The proposed Keeley bricks for date-time and position do not accord with ISO standards 8601 and 6709 respectively

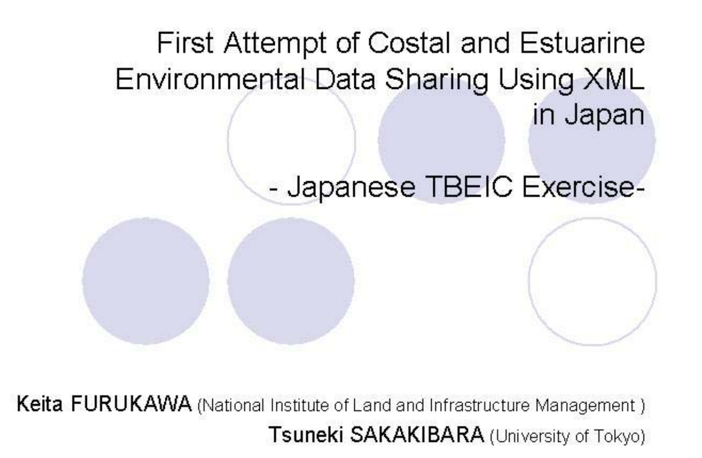
ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

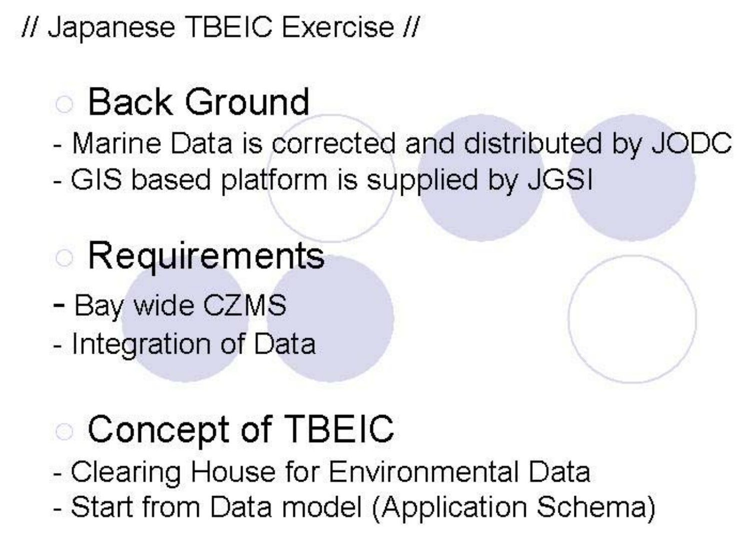
5. The Keeley XML file size is a concern and needs to be considered as an important issue along with database storage, data retrieval, searching, querying and subsetting.
   – Perhaps having metadata in XML file structure packed in a top level, separate to the data itself is a better option

ICES-IOC SGXML, Second Session, 26-27 May 2003. Gothenburg, Sweden.

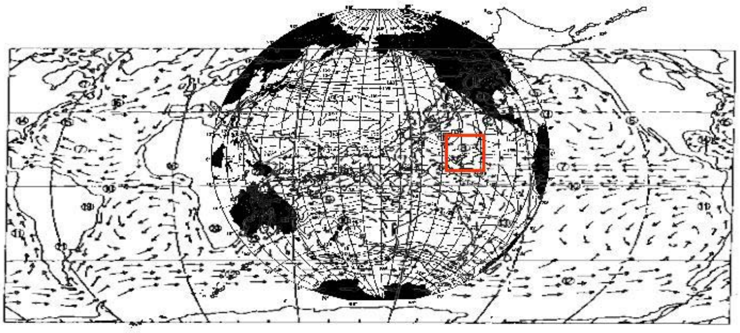
6. AODC has attempted to adopt a "Keeley Bricks" idea of block structure to create XML file for XBT profile data.
   - A cruise comprising **six full profiles** is provided in AODC internal XML format and **five data points** from the first 3 profiles of the same cruise are provided in the proposed "Keeley Bricks" XML format. It still needs more work to be done, but it gives the structure to work with. It shows the basic outline of how we can put our data into the proposed brick structure.
   - aodc_xbt.xml
   - bricks_xbt.xml

6. Some rough calculations on the size of XML files based on the data point by data point Keeley Brick structure where:
   - General Cruise/station Metadata consists ± 65 lines
     - 1 depth_pressure (DEPTH) consists of 17 lines
     - 1 data_point (TEMP) consists of 17 lines
   - gives file sizes of:
     - ± 3,500 lines for a shallow (100 m) XBT depth-temperature profile
     - ±15,700 lines for a typical (460 m) XBT depth-temperature profile

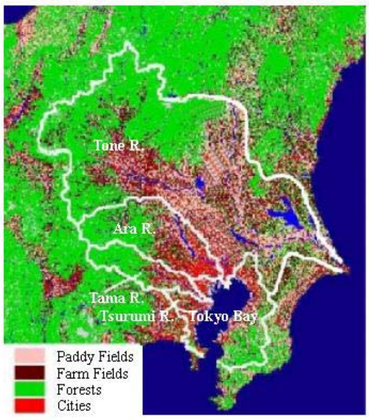

First Attempt of Costal and Estuarine
Environmental Data Sharing Using XML
in Japan
- Japanese TBEIC Exercise-
Keita FURUKAWA (National Institute of Land and Infrastructure Management)
Tsuneki SAKAKIBARA (University of Tokyo)


// Japanese TBEIC Exercise //
Back Ground
- Marine Data is corrected and distributed by JODC
- GIS based platform is supplied by JGSI
Requirements
- Bay wide CZMS
- Integration of Data
Concept of TBEIC
- Clearing House for Environmental Data
- Start from Data model (Application Schema)

Introduction -Where is Japan?



Tokyo Bay Watershed



- Watershed district

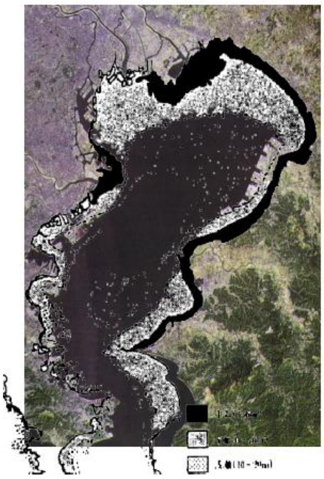
..pop/.. million

..area/......

..industries/20-30%

.....

Tokyo Bay



Area : 1000 km²

Mean Depth □15 □

River Discharge

: 8-12 x 10^9 m^3/y

Residence Time: 1.5 month

DIN load :1.1 x 10^5 t/y

Sewage :60 %

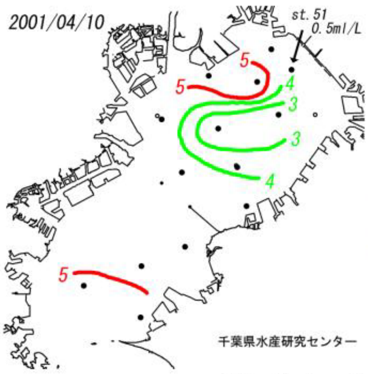
Tidal Flat:16 km^2 / 150km^2

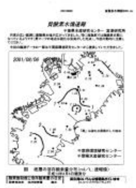
Natural Coastal Line

(1940s / 1980s)□40km / 800km

Data Sharing

Hypoxic water / Blue tide

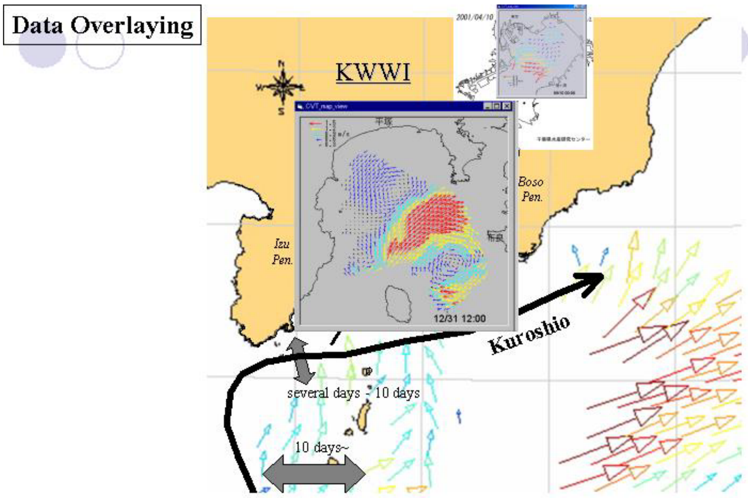




- Env. Res. Center, Chiba Pref.
- Env. Agency, Tokyo
- Fishery Association for Trawl net

DO distribution (mg/L)

Monitoring for Temporal and Local Processes
e.g. Wind driven Temporal Vertical Circulation



Monitoring for Process Interaction
between Oceanic Current and Internal Circulation in the Bay

Needs of TBEIC

- Developing the Platform for Sharing Environmental Information Concerned with Tokyo Bay
  - Standardization of Environmental Data
    - "Water Quality", "Bottom Materials", "Subpopulation", "Weather", "Marine Phenomenon"
  - Developing the Place for Data Sharing
    - Clearinghouse
- Tokyo Bay Environment Information Center (TBEIC)
  - Data Sharing with XML
  - Based on ISO/TC211 and GPL3.0

Steps of developing TBEIC data standard

- 1. Making an AS of TBEIC
  - Making an AS based on ISO/TC211
- 2. Encoding from the AS to a XML-Schema
  - Making XML-Schema with GML 3.0
- 3. Developing XML data based on the XML-schema
  - Conversion to XML by Excel macro-program
  - We have established a Study Group (SG) for this discussion
    - Members : researchers and private companies

.Last year, we dealt with only "Water Quality" data

Making the AS for TBEIC

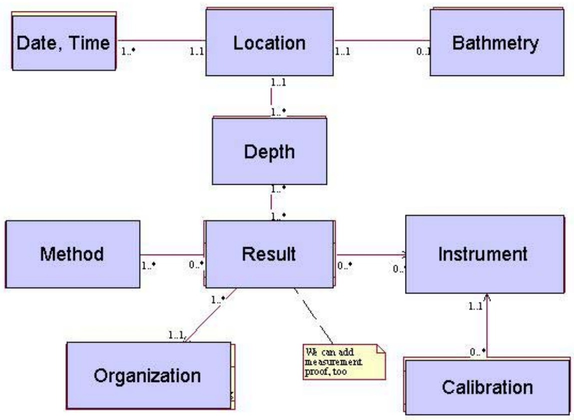
- Making data definition (AS) by UML class diagram
- Way of discussion
  - finally…
  - Describing the TBEIC AS by UML
  - however…
  - UML requires some skill to understand the diagrams
  - so…
  - We made "Definition Table (and Model diagram)"
    - Clarify important elements and their relationships

Definition Table

- Location: Monitoring Station
  - Spatial Location
  - Temporal Location
  - Total Depth
  - Depth
- Result: Monitoring Item
  - Item
  - Observation Method
  - Instrument
  - Observation Organization
  - Calibration Resurt

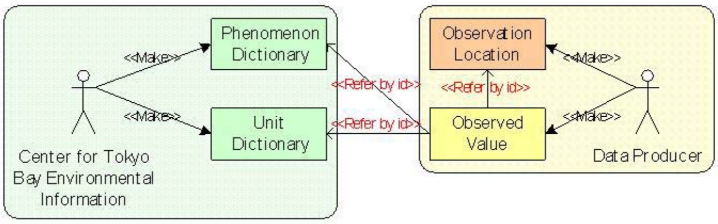
Class Diagram

- The class diagram made from the Definition Table agreed by SG

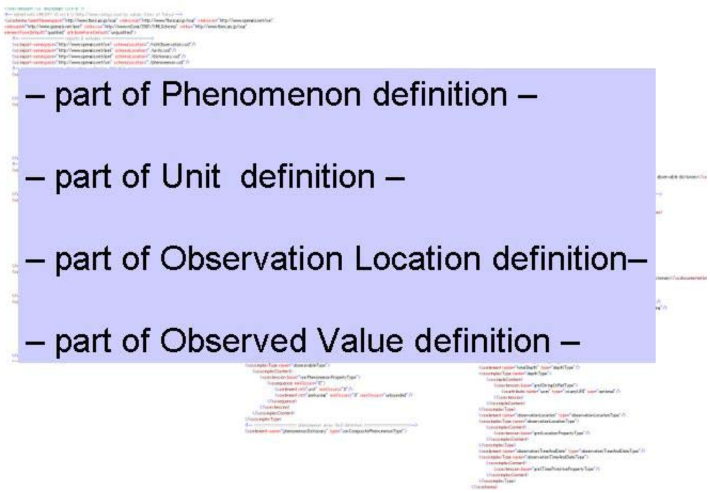


Development a XML-Schema prototype

- Prototyping a XML-Schema based on the AS of TBEIC
  - Using GML3.0
  - Refer "Observations and Measurements" (additional document of GML3.0)
- We divided into four XML files to describe a set of environmental data



XML-Schema



Development of XML files

- We developed sample XML files from the XML-Schema made as a prototype.
- Developed files
  - The file provided by TBEIC
    - A dictionary of observation items : tbeicPhenomenonDictionary.xml
    - A dictionary of Unit : tbeicUnitDictionary.xml
  - The file made by data developers
    - Observation point : observationLocation.xml
    - Observed value : observedValue.xml

Coming Next…

- Accomplishing the XML-Schema of “Water Quality”
  - With verification using the XML files
- Expand to the other categories
  - bottom materials, and others
- Development of a program to generate XML files
  - Now developing a macro-program on Excel
    - Already developed version : without parsing
    - Next version : with parsing and verification
- Implementing quality evaluation and the description based on ISO/TC211

SGXML Parameter Dictionaries

Roy Lowry & Ray Cramer
BODC

SGXML 2002

- The Plan
  - Schema/DTD for XML merged dictionary
    - Root based on the real-world entity, not the parameter code
    - Extended by Anthony to implement inter-dictionary transforms
  - Map existing dictionaries into this schema

SGXML 2002

- Progress
  - Snapshot instances of BODC, Sismer, BIO, IOS and MEDS dictionaries produced as XML documents conforming to this schema
  - Russian dictionary produced in a totally different XML structure
  - XSL produced to facilitate browser display of schema-conformant documents
  - Tomcat-based dynamic XML document generator under development
  - BIO/IOS/MEDS dictionary snapshots merged into a single document with transformations

Dictionary Interoperability

- Problems
  - Local parameter dictionaries cover the current data holdings of the maintaining organisation
  - Therefore, they are dynamic
  - Dynamic distribution of an organisation's dictionary as an XML document is relatively straightforward
  - Dynamic distribution of a document containing merged dictionaries with transforms is far from straightforward

Dictionary Interoperability

- Possible Solution
  - Dynamic XML instances of local dictionaries conforming to a common agreed structure to be strongly encouraged and a catalogue maintained
  - These could be regularly harvested and centrally integrated into the extended agreed merged dictionary format
  - Result posted as an XML document. Not fully dynamic, but updated regularly
  - How???

Dictionary Interoperability

- Possible Solution
  - Automated dictionary integration has to be based on an ontology that:
    - Specifies what entities may be merged 'asis'
    - Specifies what entities may be merged following a transform
    - Specifies the transforms in a manner that may be dynamically implemented

Dictionary Interoperability

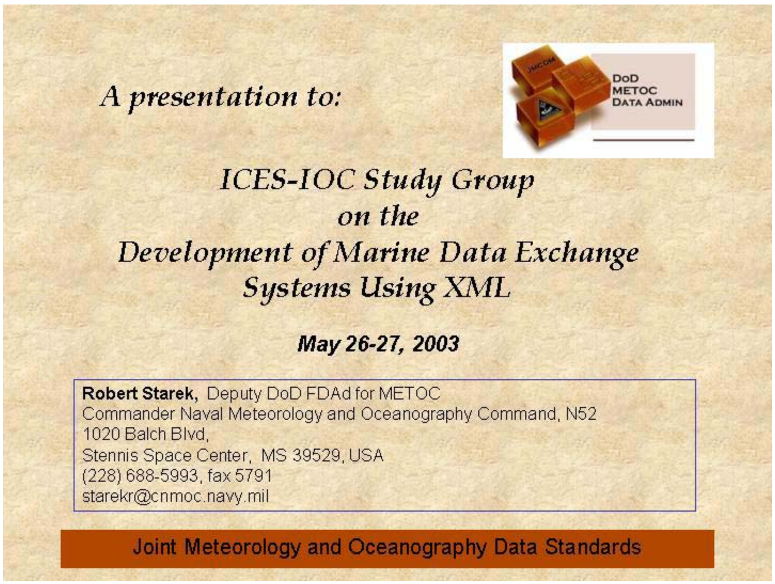
- Implementation Mechanism
  - Dictionary harvesters monitor component dictionaries and inform the ontology keeper of changes
  - Necessary extensions added to ontology
  - Merged dictionary generated and posted

Dictionary Interoperability

- Problems
  - Agreed syntax for expressing transforms needs to be developed
  - Data aggregation software needs to be developed
  - Who maintains the ontology?

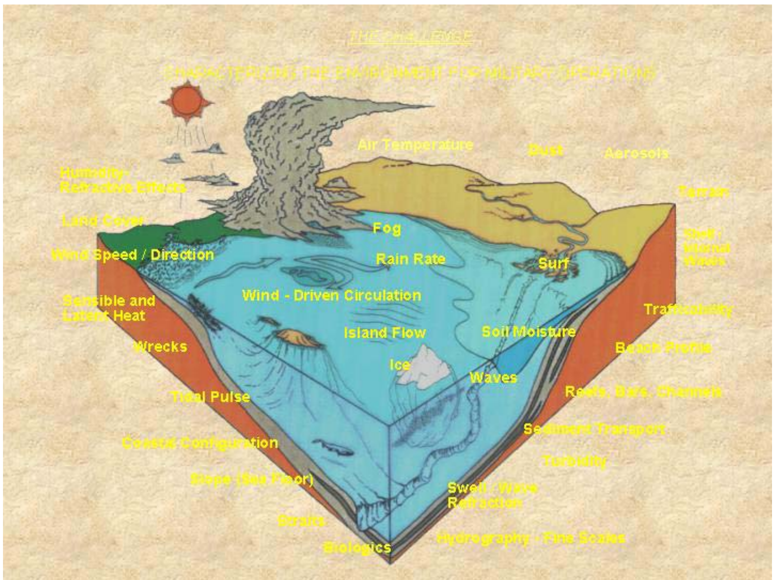
Parameter Discovery

- Currently an area of interest in BODC
- Approach of 'Agreed Parameter Groupings' (APG) taken for EU SeaSearch Common Data Index
- Hierarchical discovery mechanism
  - Discipline
  - APG
  - Parameter group
  - Parameter code
- Top 2 layers could sit above merged dictionary entities

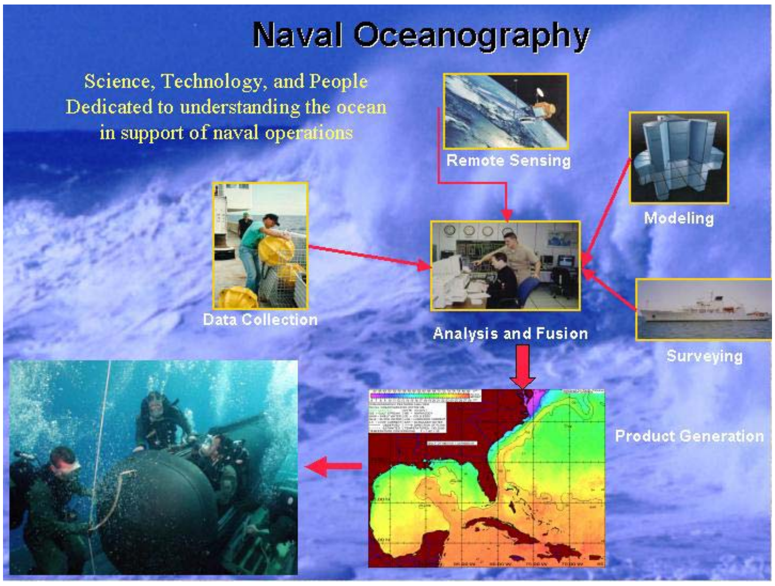

A presentation to:
DoD
METOC
DATA ADMIN
ICES-IOC Study Group
on the
Development of Marine Data Exchange
Systems Using XML
May 26-27, 2003
Robert Starek, Deputy DoD FDAd for METOC
Commander Naval Meteorology and Oceanography Command, N52
1020 Balch Blvd,
Stennis Space Center, MS 39529, USA
(228) 688-5993, fax 5791
starekr@cnmoc.navy.mil
Joint Meteorology and Oceanography Data Standards


Topics
DoD
METOC
DATA ADMIN
• Who we are
• Background
• Approach
• Accomplishments
• XML
• Data Request Language
Joint Meteorology and Oceanography Data Standards

COMNAVMETOCCOM
"Has the global responsibility of characterizing the "Environmental Battlespace" for the operating forces of the Navy and the Department of Defense"
DoD
METOC
DATA ADMIN
Three Disciplines of Naval METOC
Meteorology
Oceanography
Global Geospatial Information and Services
"Any time. . .
Anywhere"
Joint Meteorology and Oceanography Data Standards



THE CHALLENGE
CHARACTERIZING THE ENVIRONMENT FOR MILITARY OPERATIONS
Air Temperature
Dust
Aerosols
Humidity-
Refractive Effects
Terrain
Land Cover
Fog
Shelf / Internal Waves
Wind Speed / Direction
Rain Rate
Surf
Sensible and Latent Heat
Wind - Driven Circulation
Trafficability
Island Flow
Soil Moisture
Wrecks
Beach Profile
Ice
Waves
Tidal Pulse
Reefs, Bars, Channels
Sediment Transport
Coastal Configuration
Turbidity
Slope (Sea Floor)
Swell / Wave Refraction
Straits
Biologics
Hydrography - Fine Scales



Introduction

DoD METOC DATA ADMIN

- Department of Defense Directive 8320.1
- Office of Secretary of Defense (OSD) Principal Staff Assistant (PSA) for Science and Technology (S&T) sponsorship
- METOC Community of Interest (COI)

Joint Meteorology and Oceanography Data Standards

Background

DoD METOC DATA ADMIN

DoD Data Administration objectives:

- Aggressively implement Data Administration
- Standardize data elements to enhance interoperability
- Adopt existing standards before creating new ones
- Promote data standardization in the DoD in a consistent manner in collaboration with DoD, other Federal Agencies, and under the terms of international treaties and agreements
- Coordinate standards for information, information processing, and telecommunications

Joint Meteorology and Oceanography Data Standards

Background

DoD METOC DATA ADMIN

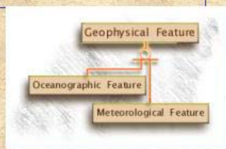
Rapid Data Standardization (RDS):

- Suggested a highly interactive collaborative process to achieve a critical mass
- Recommended forums of collaborative decision makers called "Stakeholders"
- Promoted model based collaboration as a way to achieve standardization goals more rapidly
- Promised expedited DoD formal approval where RDS was successfully applied

Joint Meteorology and Oceanography Data Standards

Getting Started

- Recognized that getting the right Subject Matter Experts (SMEs) involved was the key to success
- Adopted a tailored version of collaborative RDS that maximized the efficiency of SME collaborative time
- Developed a prioritized long-term schedule of collaborative sessions (one full week about every two months)
- Obtained ASD(DDR&E) and DISA approval of long-term approach
- Initiated collaborative sessions in the Spring of 1995



Joint Meteorology and Oceanography Data Standards

METOC DA Standardization Process

- Process is documented and model-based (assists in keeping participants on the same page)
- Enhances SME productivity and data standards quality
- Places emphasis on planning and facilitation of sessions, keeps the scope of sessions manageable
- Assures that SMEs have adequate read-ahead materials and enough notice so they can plan their participation
- Provides travel funding for SMEs
- Develops Strawman models based on reverse engineered METOC systems
- Builds the Joint METOC Conceptual Data Model by integrating session models (bottom up and top down combination)
- Utilizes contractor support to assist in the tedious DISA submissions further insulating the SME from the DoD approval process

Joint Meteorology and Oceanography Data Standards

JMCDM Configuration Management Process

DoD METOC DATA ADMIN

- JMCDM growth drove the Configuration Management (CM) requirement
- CM assures orderly growth sanctioned by the METOC data community
- Hold periodic review sessions to deliberate change proposals
- Developed a web application to facilitate CM

Joint Meteorology and Oceanography Data Standards

Software Support Tools

DoD METOC DATA ADMIN

- Made every effort to utilize available COTS and GOTS software
- The key COTS tool for IDEF1X modeling was ERwin
- Utilized the DISA Groupware for Windows tool to support collaboration of metadata
- Developed custom software to fill the gaps between COTS and GOTS where an adequate return on investment could be assured

cuda

Joint Meteorology and Oceanography Data Standards

Comprehensive Utilities for Data Administrators (CUDA)

DoD METOC DATA ADMIN

- Streamlines submissions to the Defense Data Model (DDM)
- Provides web publishing of JMCDM and Strawmodels
- Provides Groupware integration files
- Provides interactive review of standards and maintenance of change requests
- Provides web-based CM
- Provides mapping/matching support
- Negates the need for SMEs to own a copy of ERwin

Joint Meteorology and Oceanography Data Standards

Accomplishments

DoD METOC DATA ADMIN

- Completed 30+ collaborative sessions since 1995
- JMCDM has grown to 670 approved entities and 2,950 approved attributes
- JMCDM is base-lined with only a few additions yet to be incorporated
- Provides over 110 functional views in the disciplines of Meteorology, Oceanography, Astrometry, Space Weather, Geodesy, and others
- JMCDM has been used by many organizations involved in database development or modernization efforts
- JMCDM providing logical basis for developing truly joint Defense Information Infrastructure (DII) Shared Data Environment (SHADE) physical database segments
- Regular JMCDM web site users expanded to 150+

Joint Meteorology and Oceanography Data Standards

Lessons Learned

DoD METOC DATA ADMIN

- A structured process for performing data modeling is absolutely necessary
- Strong SME involvement is mandatory
- Good planning, advance preparation, and skilled facilitation are key to successful collaborative sessions
- A good technical support staff and software support tools are very important to maintaining high productivity in standardization work
- Continuous education for stakeholders and nurturing of cooperative attitudes is important
- Maintaining top management support is a continuous challenge
- Benefits are obvious but difficult to quantitatively measure
- Bridging the gap between logical and physical data standards is a new challenge

Joint Meteorology and Oceanography Data Standards

METOC DA Future

DoD METOC DATA ADMIN

- Maintain and improve the JMCDM
- Expand use of JMCDM and CUDA
- Develop joint database segments with a high potential for reuse
- Develop a joint standards-based data exchange infrastructure
- Establish and manage a standards-based METOC XML name-space for the DoD XML Registry
- Maintain DISA and PSA support for long-term joint DA work
- Support DISA initiatives to improve processes and procedures that enhance interoperability and data reuse

Joint Meteorology and Oceanography Data Standards

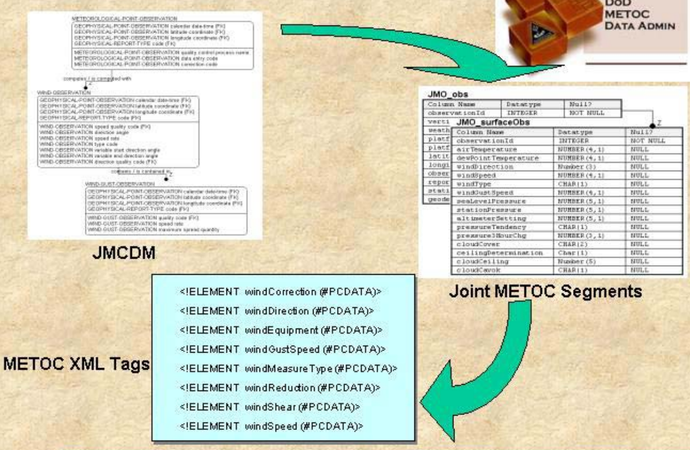
DoD METOC DATA ADMIN

METOC XML

Joint Meteorology and Oceanography Data Standards

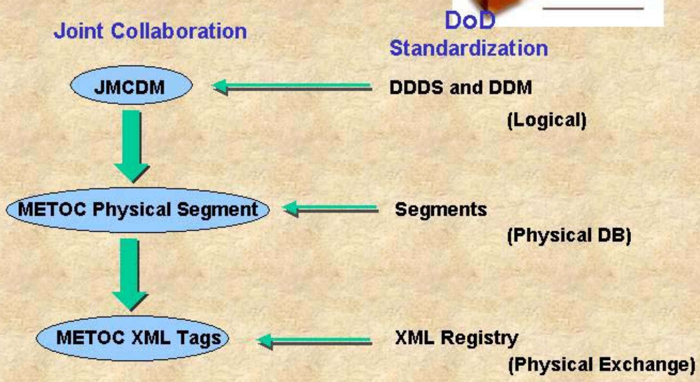
METOC XML Development Flow

DoD METOC DATA ADMIN



Joint Meteorology and Oceanography Data Standards

Joint Standards Lineage

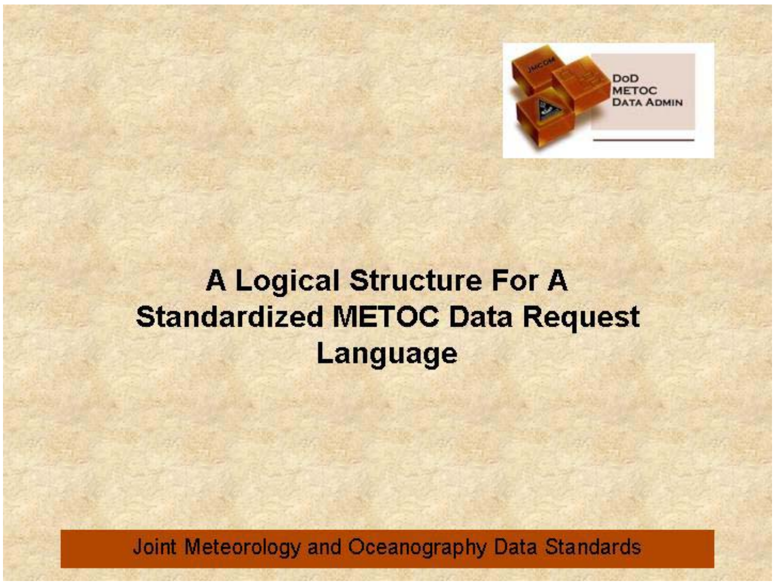


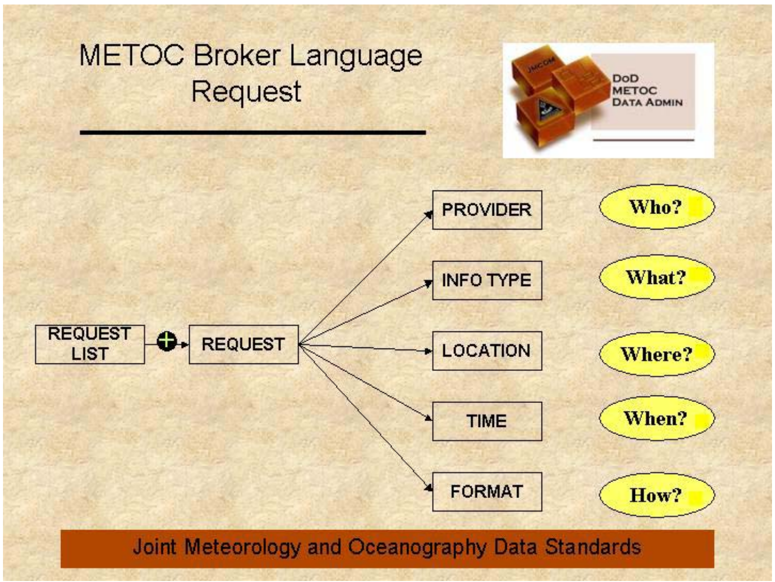
Joint Meteorology and Oceanography Data Standards

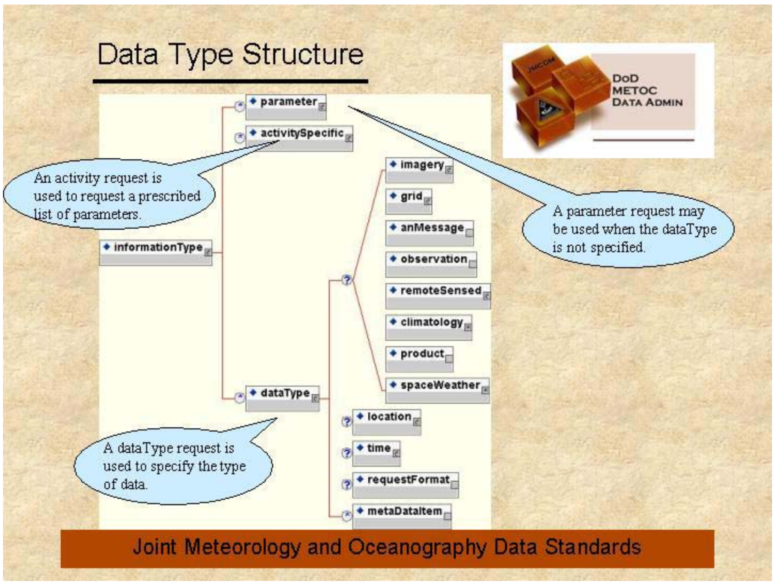
Where Are We Now?

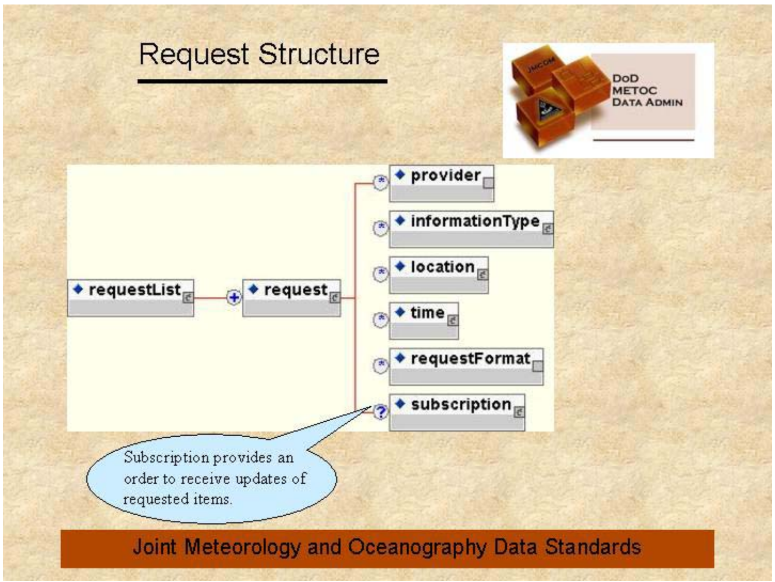
- METRICS
  - 92 Element Tags estimated, 133 currently
  - 556 Attribute Tags, 664 currently
  - 10 Schemas/DTDs, 15 currently
- Still have approximately 5 segments ahead.

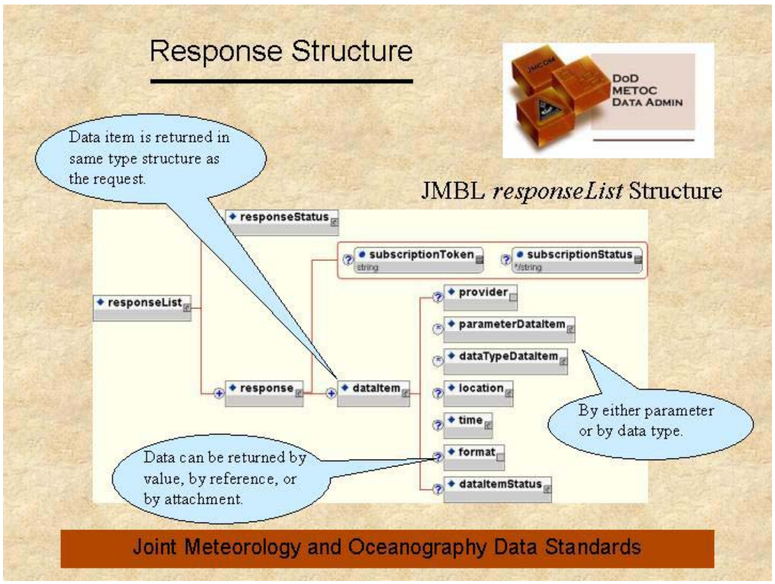
Joint Meteorology and Oceanography Data Standards


DoD
METOC
DATA ADMIN
A Logical Structure For A Standardized METOC Data Request Language
Joint Meteorology and Oceanography Data Standards


METOC Broker Language Request
DoD
METOC
DATA ADMIN
REQUEST LIST
REQUEST
PROVIDER
INFO TYPE
LOCATION
TIME
FORMAT
Who?
What?
Where?
When?
How?
Joint Meteorology and Oceanography Data Standards


Data Type Structure
DoD
METOC
DATA ADMIN
parameter
activitySpecific
An activity request is used to request a prescribed list of parameters.
informationType
imagery
grid
anMessage
observation
remoteSensed
climatology
product
spaceWeather
A parameter request may be used when the dataType is not specified.
dataType
location
time
requestFormat
metaDataItem
A dataType request is used to specify the type of data.
Joint Meteorology and Oceanography Data Standards

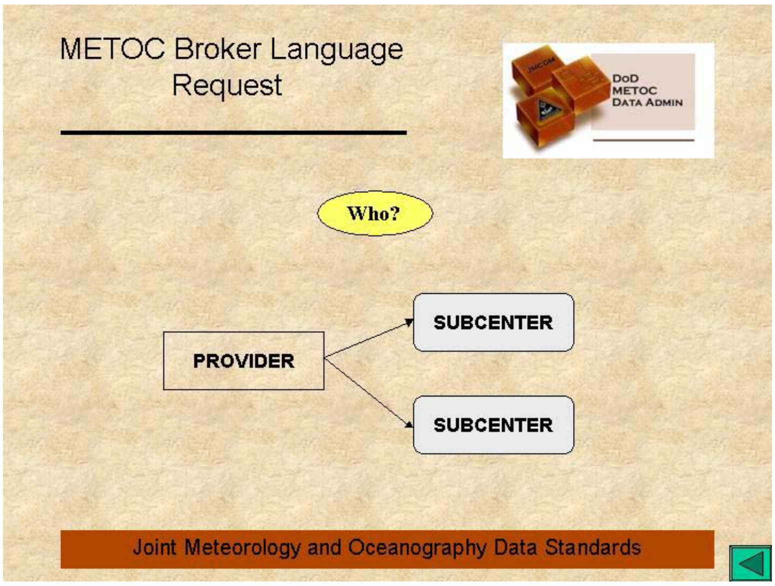

Request Structure
DoD
METOC
DATA ADMIN
requestList
request
provider
informationType
location
time
requestFormat
subscription
Subscription provides an order to receive updates of requested items.
Joint Meteorology and Oceanography Data Standards

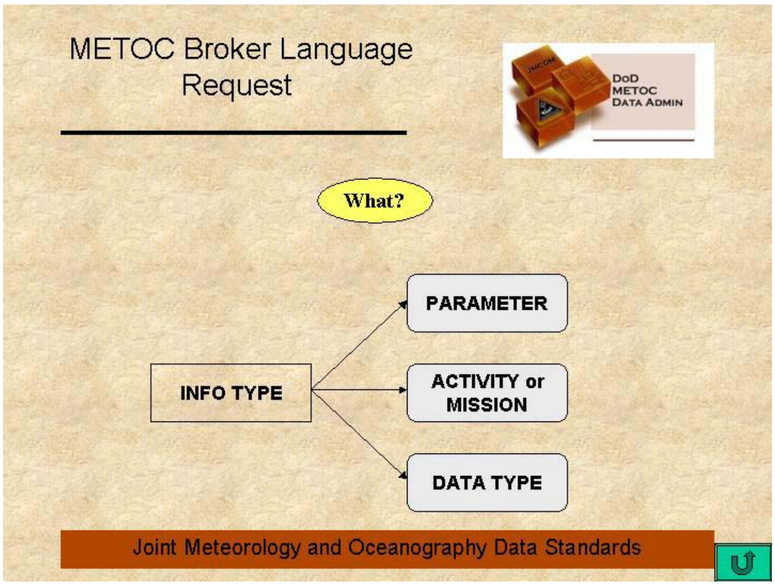


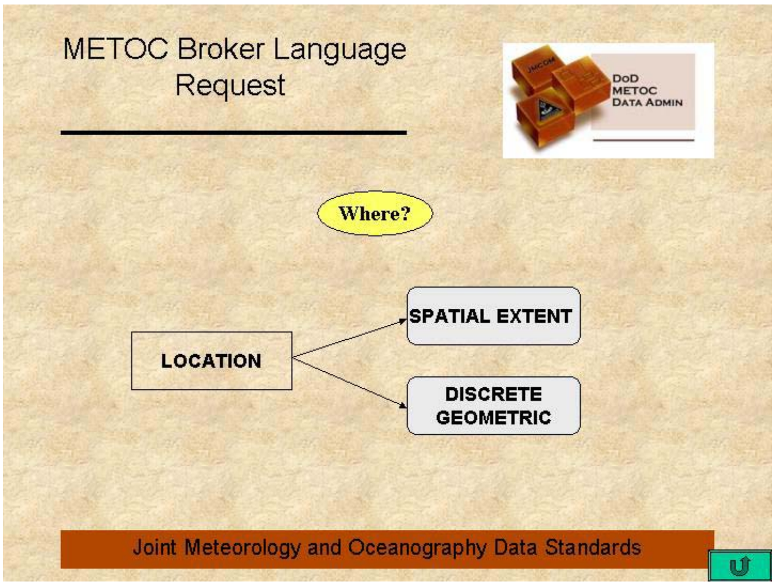
Proposed Technical Standards

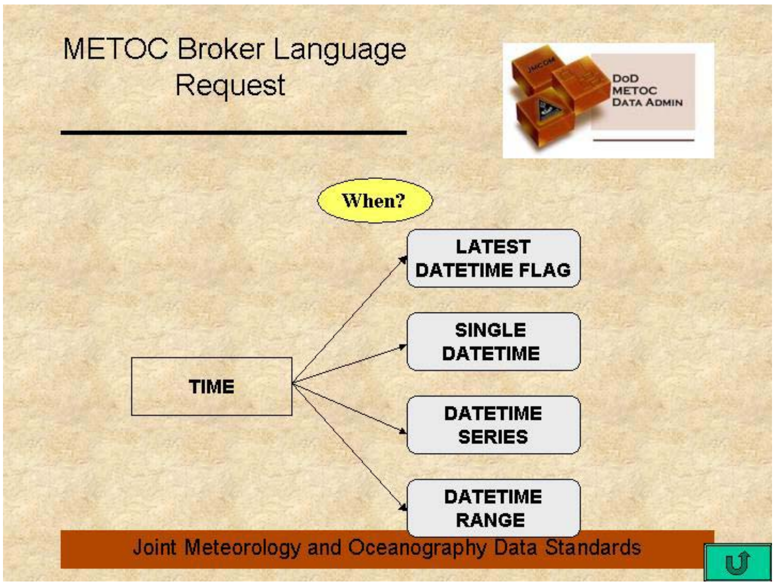
- Client METOC request will be in the form of an XML document
- METOC request parameters must be JMCDM approved tags
- Lower level protocols will be TCP/IP and HTTP
- Messaging and service description will be SOAP and WSDL

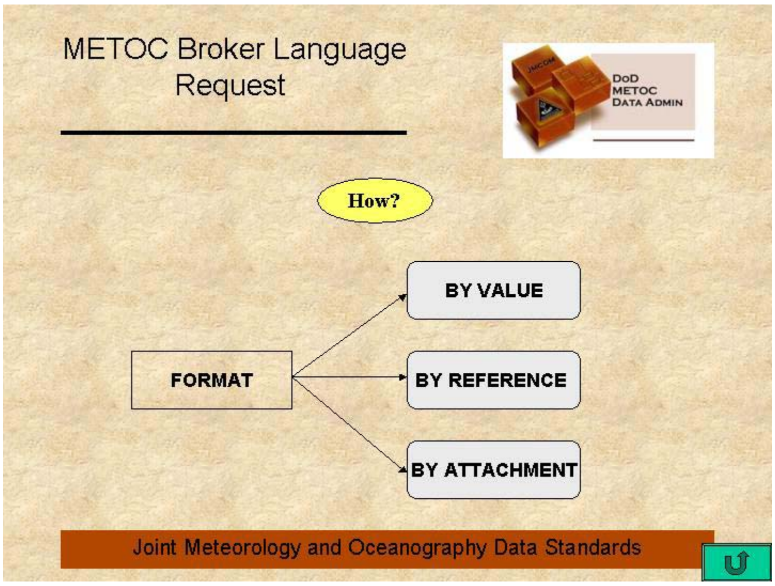
Joint Meteorology and Oceanography Data Standards

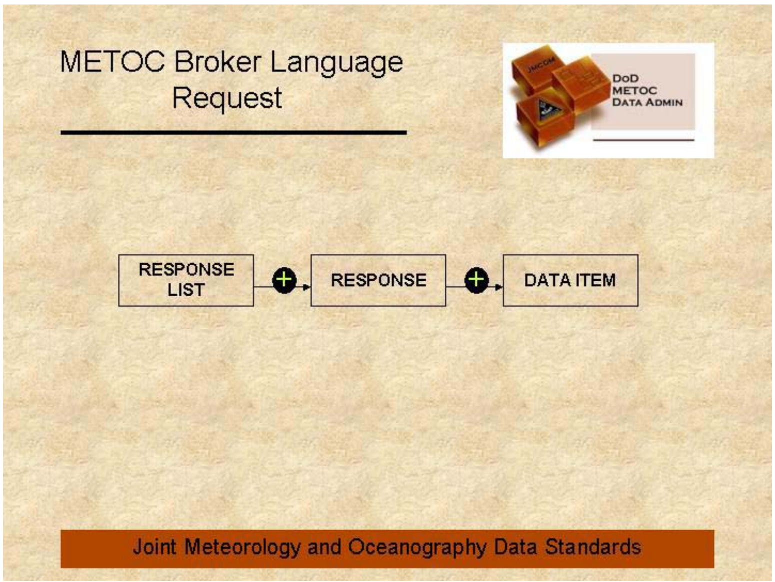

METOC Broker Language Request
DoD METOC DATA ADMIN
Who?
PROVIDER
SUBCENTER
SUBCENTER
Joint Meteorology and Oceanography Data Standards

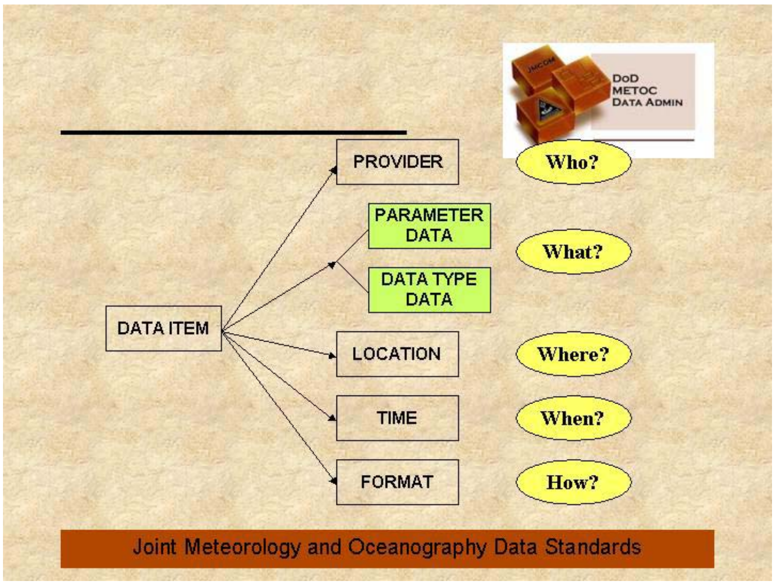

METOC Broker Language Request
DoD METOC DATA ADMIN
What?
INFO TYPE
PARAMETER
ACTIVITY or MISSION
DATA TYPE
Joint Meteorology and Oceanography Data Standards


METOC Broker Language Request
DoD
METOC
DATA ADMIN
Where?
LOCATION
SPATIAL EXTENT
DISCRETE GEOMETRIC
Joint Meteorology and Oceanography Data Standards


METOC Broker Language Request
DoD
METOC
DATA ADMIN
When?
TIME
LATEST DATETIME FLAG
SINGLE DATETIME
DATETIME SERIES
DATETIME RANGE
Joint Meteorology and Oceanography Data Standards

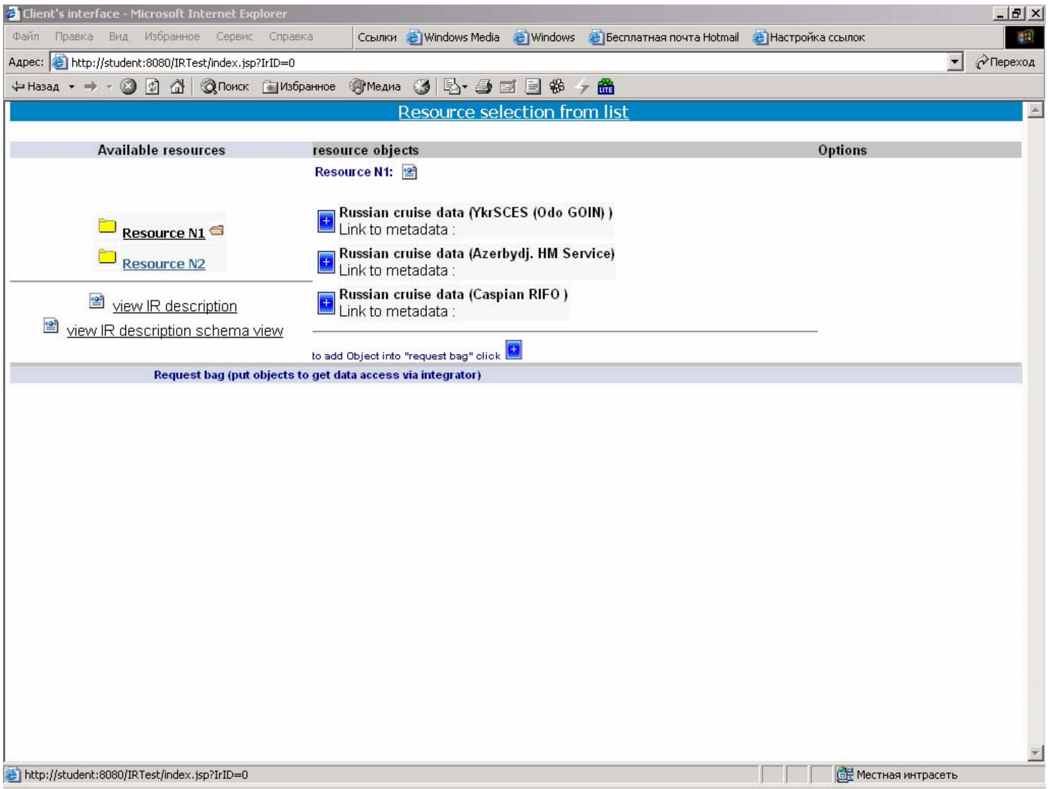

METOC Broker Language Request
DoD METOC DATA ADMIN
How?
FORMAT
BY VALUE
BY REFERENCE
BY ATTACHMENT
Joint Meteorology and Oceanography Data Standards

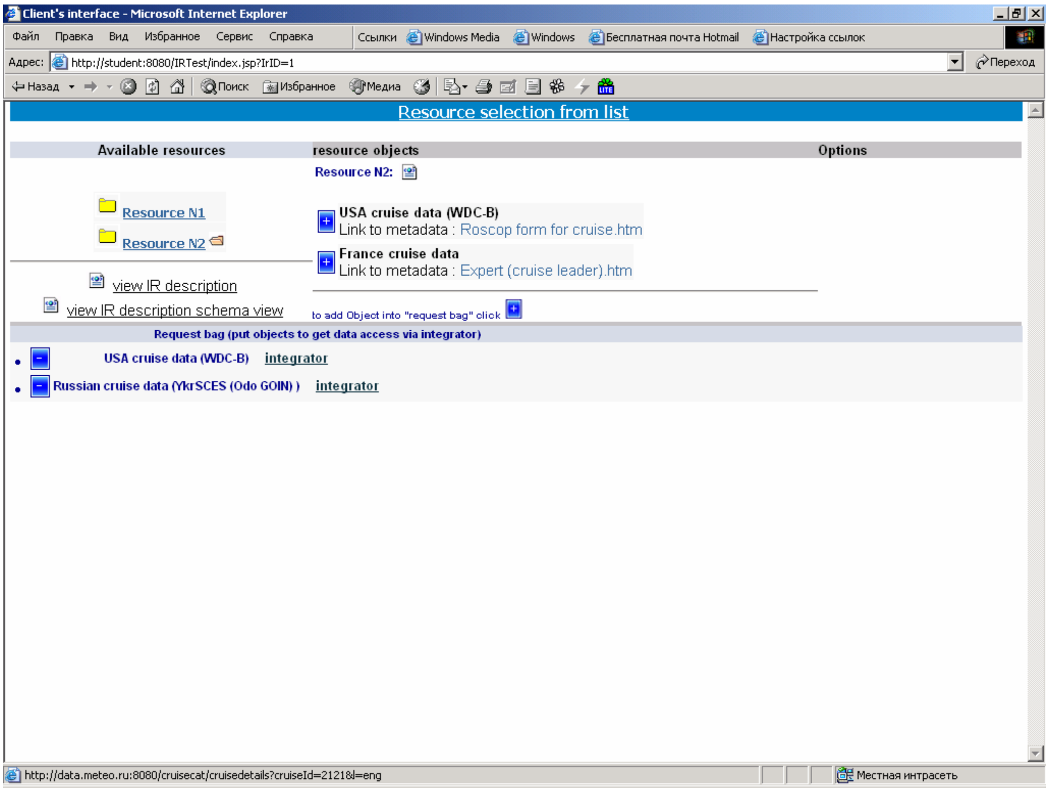

METOC Broker Language Request
DoD METOC DATA ADMIN
RESPONSE LIST
RESPONSE
DATA ITEM
Joint Meteorology and Oceanography Data Standards

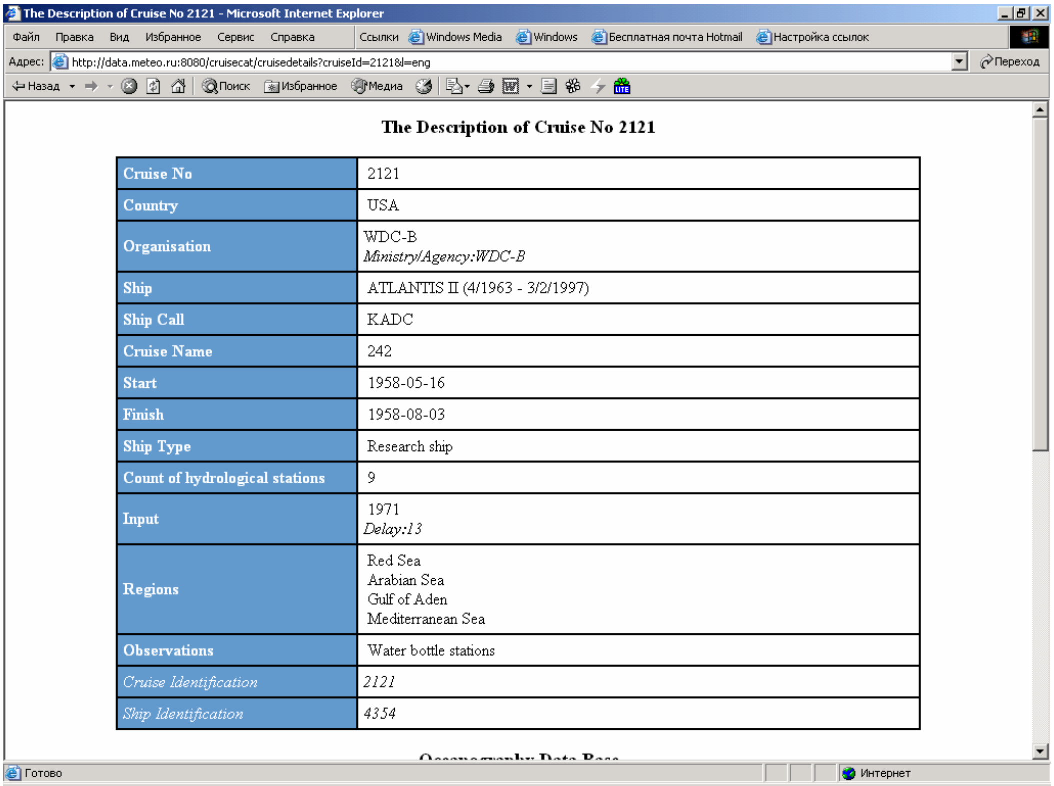

DoD
METOC
DATA ADMIN
DATA ITEM
PROVIDER
PARAMETER DATA
DATA TYPE DATA
LOCATION
TIME
FORMAT
Who?
What?
Where?
When?
How?
Joint Meteorology and Oceanography Data Standards

The first level selection interface is based on the IR object index (first screen snapshot). Here, the user selects an IR unit and receives a list of IR objects (second screen snapshot). The second level selection interface displays the IR object title and relevant metadata. The third level is a communication layer between the IR object and the integrator. The Integrator provides the link to the data, represented as file or as working web-system (e.g., web-site, web-portal) application. The Integrator shall describe the users request and the transport-level software shall form a system request, addressed into the distributed system node. The final stage is to obtain the answer and present it to the end-users.

In present version, software provides functions similar to a shopping cart. The user may collect the resources they need and place those resources in a “cart”. At present, the interaction between the local software and the Integrator is simulated, because of its complex semantic and technical nature. In the final version, software will hide all interaction processes between the selection interfaces, integrators and transport procedures. Screen snapshots 3 to 5 show the IR testing software, currently working at RNODC web-server (see http://data.meteo.ru:8080/IRTest/index.jsp).

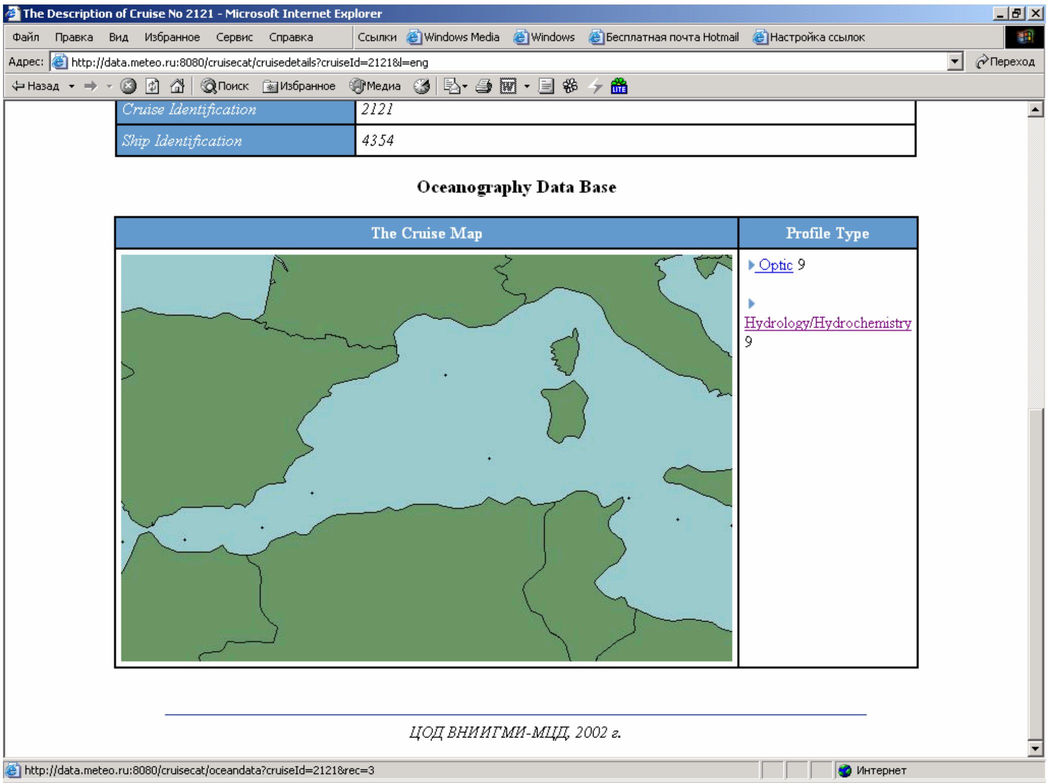






The Description of Cruise No 2121

Cruise No	2121
Country	USA
Organisation	WDC-B *Ministry/Agency:WDC-B*
Ship	ATLANTIS II (4/1963 - 3/2/1997)
Ship Call	KADC
Cruise Name	242
Start	1958-05-16
Finish	1958-08-03
Ship Type	Research ship
Count of hydrological stations	9
Input	1971 *Delay:13*
Regions	Red Sea Arabian Sea Gulf of Aden Mediterranean Sea
Observations	Water bottle stations
Cruise Identification	*2121*
Ship Identification	*4354*



The Description of Cruise No 2121 - Microsoft Internet Explorer

Адрес: http://data.meteo.ru:8080/cruisecat/cruisedetails?cruiseId=2121&l=eng

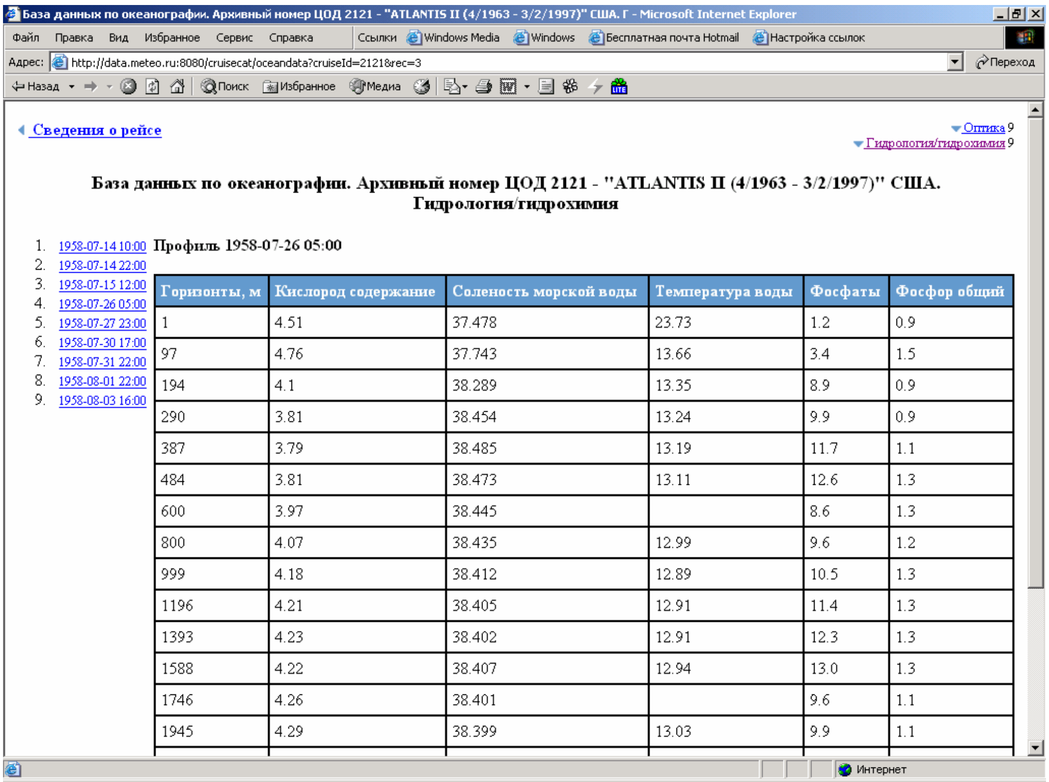
Cruise Identification	*2121*
Ship Identification	*4354*

Oceanography Data Base

The Cruise Map	Profile Type
	Optic 9 Hydrology/Hydrochemistry 9

ЦОД ВНИИГМИ-МЦД, 2002 г.

http://data.meteo.ru:8080/cruisecat/oceandata?cruiseId=2121&rec=3 Интернет



База данных по океанографии. Архивный номер ЦОД 2121 - "ATLANTIS II (4/1963 - 3/2/1997)" США. Г - Microsoft Internet Explorer

Адрес: http://data.meteo.ru:8080/cruisecat/oceandata?cruiseId=2121&rec=3

Сведения о рейсе

Оптика 9
Гидрология/гидрохимия 9

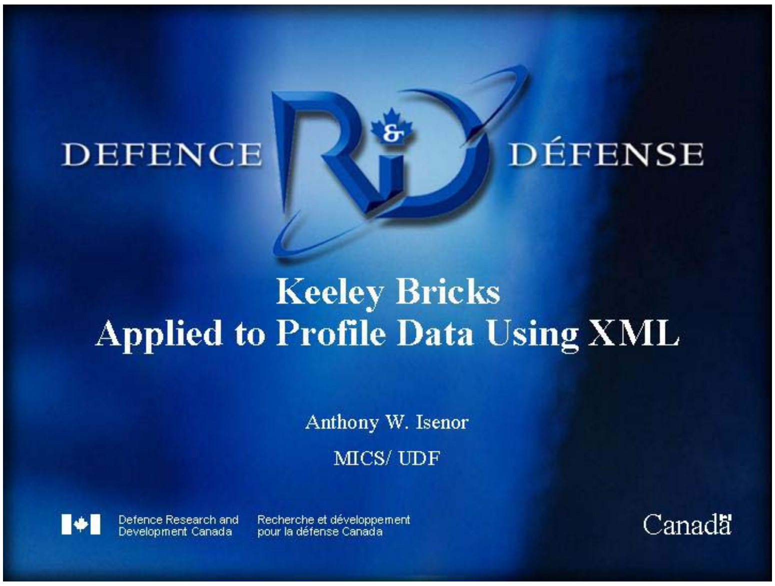
База данных по океанографии. Архивный номер ЦОД 2121 - "ATLANTIS II (4/1963 - 3/2/1997)" США. Гидрология/гидрохимия

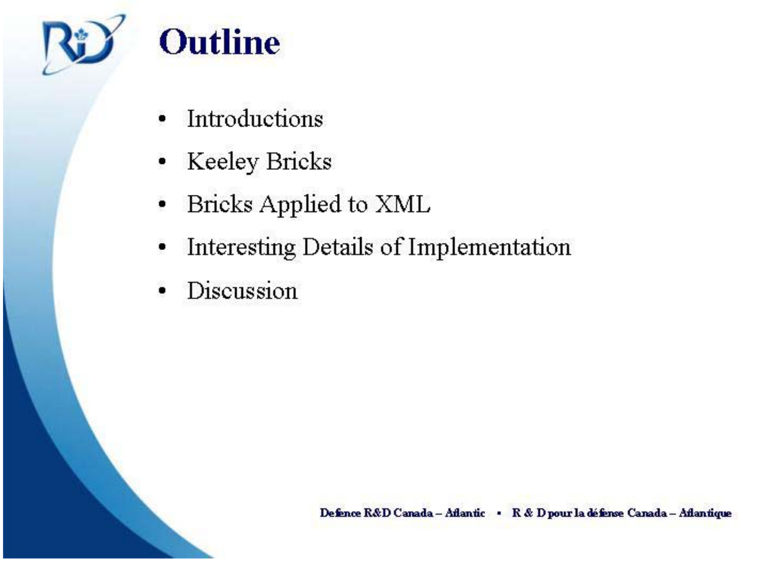
1. 1958-07-14 10:00
2. 1958-07-14 22:00
3. 1958-07-15 12:00
4. 1958-07-26 05:00
5. 1958-07-27 23:00
6. 1958-07-30 17:00
7. 1958-07-31 22:00
8. 1958-08-01 22:00
9. 1958-08-03 16:00

Профиль 1958-07-26 05:00

Горизонты, м	Кислород содержание	Соленость морской воды	Температура воды	Фосфаты	Фосфор общий
1	4.51	37.478	23.73	1.2	0.9
97	4.76	37.743	13.66	3.4	1.5
194	4.1	38.289	13.35	8.9	0.9
290	3.81	38.454	13.24	9.9	0.9
387	3.79	38.485	13.19	11.7	1.1
484	3.81	38.473	13.11	12.6	1.3
600	3.97	38.445		8.6	1.3
800	4.07	38.435	12.99	9.6	1.2
999	4.18	38.412	12.89	10.5	1.3
1196	4.21	38.405	12.91	11.4	1.3
1393	4.23	38.402	12.91	12.3	1.3
1588	4.22	38.407	12.94	13.0	1.3
1746	4.26	38.401		9.6	1.1
1945	4.29	38.399	13.03	9.9	1.1

Интернет


DEFENCE
DÉFENSE
Keeley Bricks
Applied to Profile Data Using XML
Anthony W. Isenor
MICS/ UDF
Defence Research and Development Canada
Recherche et développement pour la défense Canada
Canada


Outline
• Introductions
• Keeley Bricks
• Bricks Applied to XML
• Interesting Details of Implementation
• Discussion
Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Introductions

- J. Robert Keeley
  MEDS, Ottawa
  DFO

- Joe Linguanti
  IOS, Sidney
  DFO

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Keeley Bricks

- Two principles
  - Exploit the natural hierarchy found in ocean data types
  - Construct different structures from these data objects
- Profile Data Study
  - 20 bricks

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Keeley Bricks

- analysis_method
- availability
- calibration
- comment
- data_dictionary
- data_point
- data_set_id
- depth_pressure
- history
- instrument
- latitude
- ldate
- longitude
- provenance
- quality
- quality_testing
- sampling
- sensor
- units
- variable

Keeley Bricks

- data_set
  - availability
  - comment
  - data_point
  - data_set_id
  - provenance
  - quality
  - quality_testing
  - variable_set ⇐
  - location_set
  - history_set
  - data_set

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Keeley Bricks

- variable_set
  - analysis_method
  - calibration
  - comment
  - instrument
  - sampling
  - sensor
  - units
  - variable

Bricks Applied to XML

- Pure bricks – have content
- Compound bricks – wrappers

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- data_set
  - availability
  - comment
  - data_point
  - data_set_id
  - provenance
  - quality
  - quality_testing
  - variable_set
  - location_set
  - history_set
  - data_set

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- Pure bricks – have content
- Compound bricks – wrappers
- Needed Rules for Construction

"It is not our abilities that show what we truly are… it is our choices"

… Albus Dumbledore

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- Why do we need rules?

```
<ldate property="start">2003-03-28Z</ldate>

<ldate>
  <property>start</property>
  <value>2003-03-28Z</value>
<ldate>

<ldate>
  <property>start</property>
  <year>2003</year>
  <month>3</month>…
```

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- Some “Rules”
  - Treat attributes as categorical variables (lists)
  - Compound bricks will have no defined attributes/elements
  - Define various formats (time, date, latitude, …)

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- variable_set
  - analysis_method
  - calibration
  - comment
  - instrument
  - sampling
  - sensor
  - units
  - variable

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- units brick

```
units {pt_code, pt_link, received_units, stored_units}
 ├ conversion
 ├ reference
 └ variable_name
```

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- units – stores information about the units used to record a measurement

Attribute	*Definition*	*Occur*
pt_code	A unique code for the variable.	1
pt_link	Used to allow links to be made between 2 different measurements made from the same instrument, or two of the same measurements of variables	0,1
received_units	The units in which the data were originally received.	0,1
stored_units	The units in which the data have been archived. Normally this would be in SI.	1

Bricks Applied to XML
pt_code

- the code as tagname → difficult to define schema due to occurrences
- the code as tag content → separates code from value
- the code as attribute content → Yes

Bricks Applied to XML

- A data value.

data_point {**pt_code**, pt_link, statistic, typing}

<data_point pt_code="TEMP">1.25</data_point>

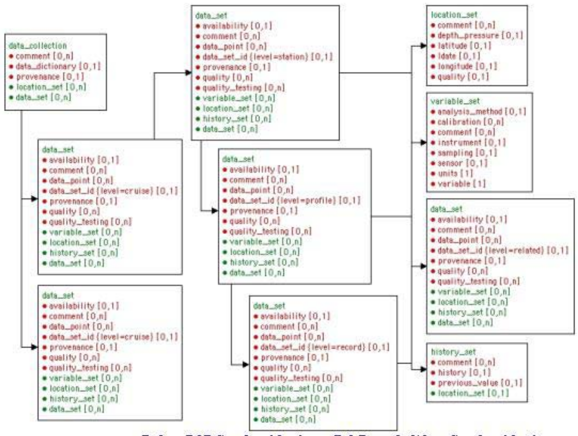
Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML

- location_set
  - comment
  - depth_pressure
  - latitude
  - ldate
  - longitude
  - quality

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Bricks Applied to XML



Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Implementation

- MEDS / IOS
  - Fortran
- BIO
  - Matlab

- MEDS / IOS
  - Fortran using msxml.dll
- BIO
  - XSLT
  - Java using data binding

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Interesting Details of Implementation

- Specific tag names are a complication
- In-house formats are lacking
- Parameter Code Mapping
- Null Values are not XML friendly

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Discussion

- Comments?
- Questions?
- Discussion?

Defence R&D Canada – Atlantic • R & D pour la défense Canada – Atlantique

Metadata Standards Review and Progress

Donald W. Collins
US NODC

Current Metadata Standards

- ISO (Approved by ISO in March 2003)
- MEDI/NASA Data Interchange Format (DIF)
- Cruise Summary Report
- US NODC Data Documentation Form
- US Federal Geographic Data Committee
  - Content Standard for Digital Geospatial Metadata
- EDMED
- ODAS (Ocean Data Acquisition System)
- MARC21 (Machine Readable Cataloging – Libraries)

Current Status of Mapping

- Spreadsheet with mapping between FGDC, CSR, MEDI/DIF, DDF
- EDMED not yet included
- ODAS – CSR in formative stages
- Other efforts?

Issues

- Granularity: what are we trying to describe?
  - ISO, FGDC, CSR, DDF, MEDI/DIF
  - ODAS
  - EDMED
- ?

ANNEX 5: A MODEL OF THE DISTRIBUTED MARINE INFORMATION RESOURCES - APPROACHES AND DECISIONS

Second meeting of the IOC/ICES SGXML, 26-27 May, 2003

Nickolay Mikhalov, Evgeny Vyazilov, and Sergey Belov

Russian Research Institute of Hydrometeorological Information - World Data Center (RIHMI-WDC), Obninsk, Russia

Background

The conceptual vision of an integrated data management system at a strategic level is given in several GOOS documents (GOOS DM -2001 [1], GCOS/GOOS/GTOS – 2000 [2], GCOS DM-2000 [3], COOP DM-2002 [4], EuroGOOS DM-2000 [5], WMO Future System [6], etc.). The preliminary technical, conceptual approach and decisions on the distributed information system including the "end to end" data management aspects were considered in GETADE/ETDMP documents [7]. This document also covered the entire data management cycle, from the initial collection of marine observations to the potential development of value-added information products as required by a wide range of end users.

In addition, there are now a number of developments dealing with the integration of various blocks of data management practices. This includes the co-ordinated and joint processing of multidisciplinary data from various data sources (VODS-1999; CoastBase-2001; OBIS-2002; IWICOS-2002, ESIMO-2002 (Russian Unified System on World Ocean Conditions[8]), SearSearch-2003 and others). Such developments have been mainly carried out independently, and realized with different visions and understanding on the methods and tools used for the integration of data, information, different titles and abbreviations.

Despite obvious confusion with the names and definitions, the mentioned and other examples have a general purpose - to integrate (in a broad sense) inhomogeneous and geographically distributed data and products on the basis of the World Wide Web. This purpose requires the fulfilment of two main tasks:

1) developing the integration technology in a form of a technological "umbrella" over existing data sources. This will allow users access to numerous data flows/sets/bases in a unified information space, with "single stop shopping" [4];
2) mapping this technology on the data centre network, when each data centre will support certain data and products. This provides the interface to local data and products according to the standards of the integration technology.

Considering this development strategy, it is possible to make a basic conclusion that today we should concentrate our attention on the technological component of the future unified marine system. This will form part of the integration technology for non-uniform and geographically distributed data sources, commonly referred to as "integrated system", or "end to end data management", or "virtual ocean data centre", etc.

For further considerations, it is necessary to precisely determine what is an integration technology. First of all, an integration technology is a Web-based technology. Therefore, development should be based on the solutions and standards of the WWW. The WWW is a network hypermedia system and the key architectural element of Web-technology is a resource or service [9-11]. By taking into account the above-mentioned WWW issues, an integration technology is **a set of rules, standards and tools, to support a Web-based, distributed, marine information resource system.** The key object of an integration technology is the "**information resource"** which has two connected views:

(i) low technological view – the software and transport/protocol are developed in the context of web-technologies (e.g., WSDL, SOAP, XML-RPC, JMS, application and original servers, etc.), which are developed by a number of international and corporate working groups (e.g., W3C, OASIS and other) [12,13]. Alongside the W3C solutions for the web components, there are a number of solutions developed in various applied spheres of research. These technologies are designated by the term "Grid" and are often associated with virtual organizations [14]. The GRID protocols and standards are integrated in the GlobusToolkit. The GlobusToolkit is used in frames of large-scale projects for various subject domains such as: NERC Grid Physics Network project; European Data Grid; Particle Physics Data Grid; Network for Earthquake Engineering Simulation Grid. The GlobusToolkit, in many respects, are similar Web technologies. There are certain distinctions between GRID and Web technologies, which, may be overcome at the Internet 3 level.

(ii) high problem-oriented semantic view - the structures and formats, links, software and other tools which reflect the specific marine information resources and are developed for the required functionality.

In this document, the approaches and design decisions for the **distributed marine information resource model (IR model)** are considered. The IR model is the basis for the integration technology and here we consider the IR model as an abstraction of the architectural software/information elements required to establish a web-based distributed marine information resource system. The IR model contains definitions of elements and their roles, basic agreements on interconnections between elements, features of their realization, etc. Parts of the IR model elements (especially regarding the metadata), which are developed for ESIMO construction [14], are now considered in more detail. The descriptions of their structure and contents and also the examples of their practical testing with XML use are given.

1. The general definitions and agreements

1.1. The information resources

The definitions and specifications of the IR model elements should take into account the requirements in the context of a web-technology. This includes the requirements determined by the data nature (information resources) and the functional requirements.

1.1.1. The resources, services and client-servers

The information resource in a web-based distributed system (W3C definition [15]) represents a "service", ensuring that the IR representations are in the form of a:

- Web-site - a set of the HTML, GIF, Jpeg, etc. with a host page or with direct reference to separate components of a set;
- ftp server for transfer of data files;
- files in local file system;
- software applications: 1) independent application providing the complete access to the information resource; 2) system applications (considered the integration technology) - realizing various services and developed on the basis of the system (integration technology) agreements and means;
- telephone and other (according to the specifications W3C[10]).

Any resource (service) is unequivocally identified by a WWW physical identifier on a URI scheme [16] and the semantic (logic) identifier in the distributed marine information resource system (under the IR model). The resources can be static or dynamic (updated in time).

Considering information resources as a "service" and using W3C specifications, it is possible to reduce the information interaction in the distributed information resource system to interaction between clients and servers [9,11], where:

server (or service provider) - represents a set of resources (services) and also is the representation of the distributed marine information resource system node (in a broad sense, the system node is software, telecommunications, organizational structure. This includes data centre specialists and local dataware, databases/datasets, web-pages, web-applications, agreed rules to support defined resources, etc., that provide generation of system services);

client (service requester) - the directing of requests to server(s) through the appropriate system navigation mechanism and tools, receiving a result from the servers (directly or through the aggregation of separate results from several servers).

A server can operate as the client. Use of the client - server approach as the basis of the integration technology means separate consideration of the information interface aspects between the clients and servers (resource sets) from aspects of data and information management and storage in system nodes on local data source level (database, system files, etc.).

1.1.2. The semantic view on the information resource

The information resources (semantic vision) are any operational and delayed-mode information on the marine environment (metadata, data and production) provided to the end-users. The marine information resources have a hierarchical organization and multiple nature of semantic representation (Fig.1).

2. The architectural components of the integration technology

The basis of the architectural design decision is a universal interface approach. All kinds and forms of interaction between system nodes (servers) and all clients are realized on the basis of the interaction interface on the technological and semantic levels. The technological universality of the interaction interfaces is achieved by using common standards (XML, WSDL and other W3C standards). The semantic universality of interfaces means use of unified techniques of information resource exchange, unified dictionary of parameters, code tables, formats, structures, etc. The most effective way to achieve universality at all is to use technological standards for the development of semantic means of interface unification (marine XML as the example).

As a whole, the integration technology has a complex architecture. This architecture consists of various components which provide all kinds and forms of interaction between the system nodes (servers) and clients using the software, protocols and other tools of web-technological and semantic levels. However, here we will consider the simplified vision of the integration technology architecture, where details of web-system components will be taken into account only when it is required for specifications of the structures and interaction rules of the IR model.

Thus, from this point of view, the integration technology (IT) architecture consists of software and information components (Fig.2.).

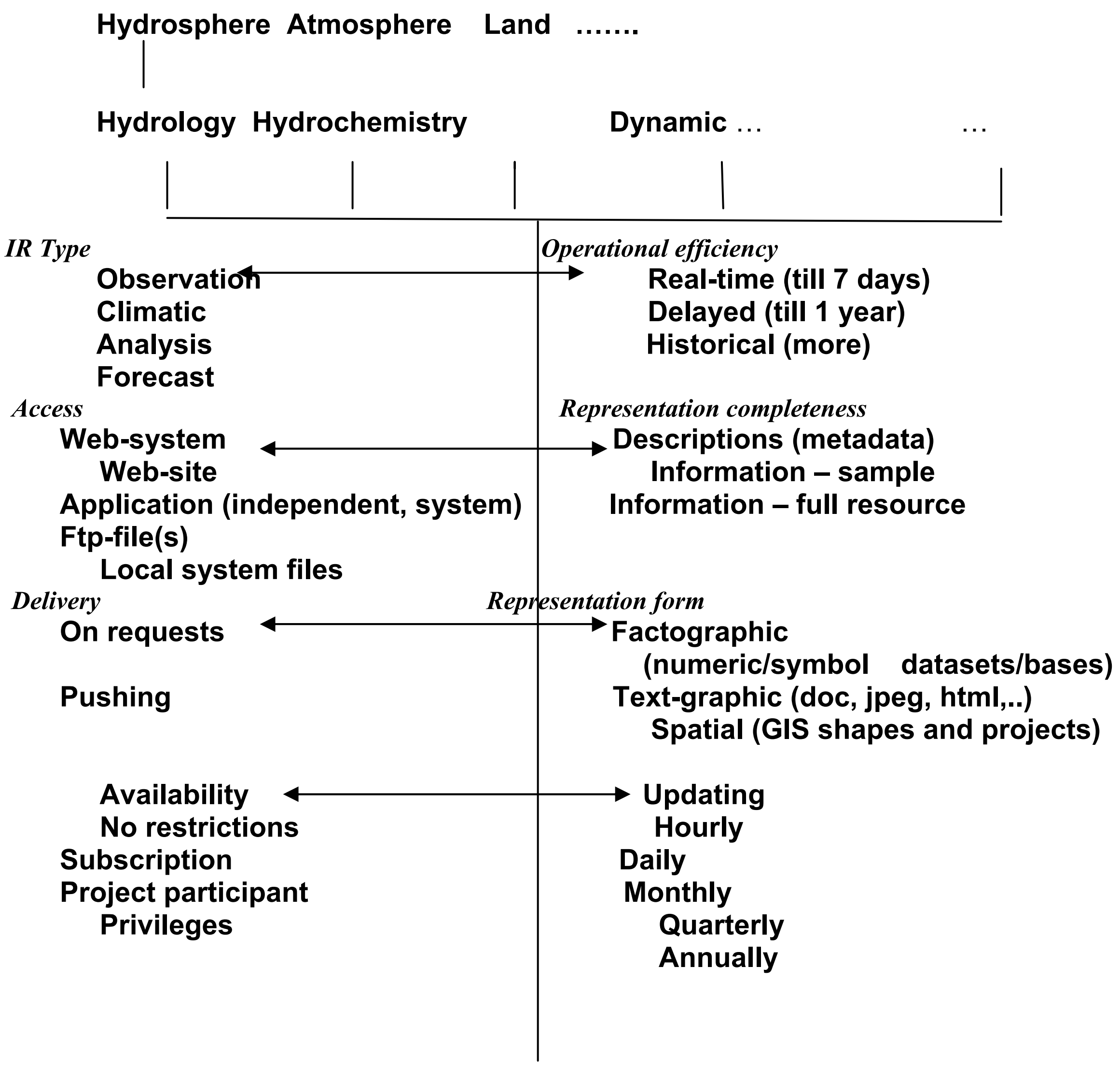


Fig. 1. The information resource representations in distributed information systems (semantic view).

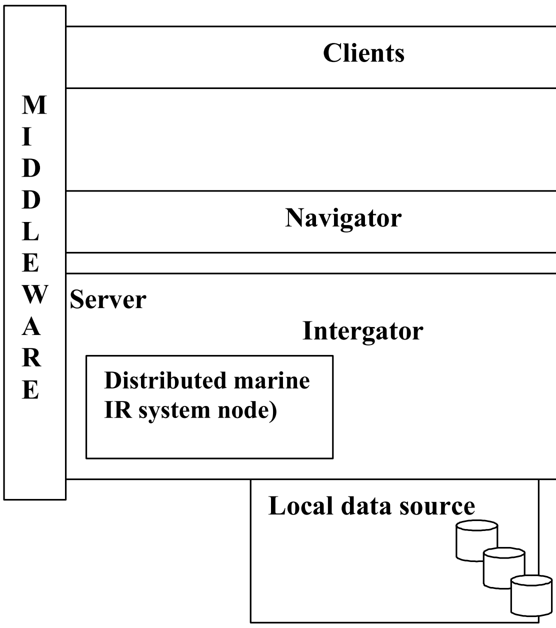


Fig.2. The architectural components of the integration technology.

2.1. Software components

Software components include:

Middleware is the lowest level of the IT architecture (connectors - W3C definition and other client-server software), ensuring operation and interactions between clients and servers, including addressing the services, packing and messaging, communications and interaction. Connectors can be realized in the various interaction standards (e.g., CORBA, COM, etc.).

Navigator (Discovery Agency on W3C definition [11]) provides a number of coordinated system data management functions such as:

- decoding of user requests in common request format (request resolver);
- finding, binding (using middleware elements) with appropriate servers and transmission of formalized requests to server services using the integrators and aggregation of their implementation results (requires provider);
- providing the processing, viewing, saving and other user-oriented functions (request publisher).

Integrator. The integrators ensure connection between the local data sources (a resource set) and the system UIS. Conceptually, there are two types of integrators:

- the integrators to bind the existing services (e.g., when a system node consists of Web-pages, Web-based applications, etc.) on the local level.
- the integrators providing the connections with the local data sources (e.g., DBMS, data file system, etc.). This also deals with the conversion of data coming from a data source to the common exchange data format and forming the representation of local dataware as a system service (resource).

Client applications – a browser and (or) problem-oriented application which is used to access the Navigator. This application deals with full system services to find, process, view and represent the information resources to the end-users.

2.2. Information components

The Information components (actually the IR model) provides information links and interactions between clients and servers, and between various servers in a complex request case on base of principal approach. Any information resources which are generated by the system nodes (servers) interact through the system Information components.

The design decisions on the IR model are based on the following agreements:

1) the IR model should reflect the semantic nature and the web-technological view on the integration technology using structural and functional information resource elements;
2) the IR model should consider the information interfaces between the clients and servers separately from aspects of interfaces on the local data source level. This means, that we do not oblige ourselves to integrate the existing local data sources with numerous formats and structures by means of a physical data reformatting to a unified information resource format. Rather, the integration technology concept is based on common information resource specifications which create a virtual environment to transfer and exchange information resources by system services;
3) we will reduce the possible amount of structures defining the typical forms and structures of information resource representations.

2.2.1. Information resource elements

The structural information resource elements (Fig.3.) include;

1) **the information resource (IR unit)** - any titled collection which has a unique WWW URI identifier and semantic system identifiers. The information resource is a representation of the single "service" (according to URI) and is considered a unit of uniform information space of the system, that is inherited in all other decisions, rules and standards of the IR model. The information resource examples include:

   - the Russian NODC cruise ocean data (for area, R/Vs ant etc.) on ftp;
   - the Russian NODC time series ocean data on OWC (s) from a database;
   - textual and graphic information on hydrochemical climate of a given region on the Web-site with host page;
   - the Russian NODC digital prognostic hydrometeorological information (GRID/GRIB) in a data file.

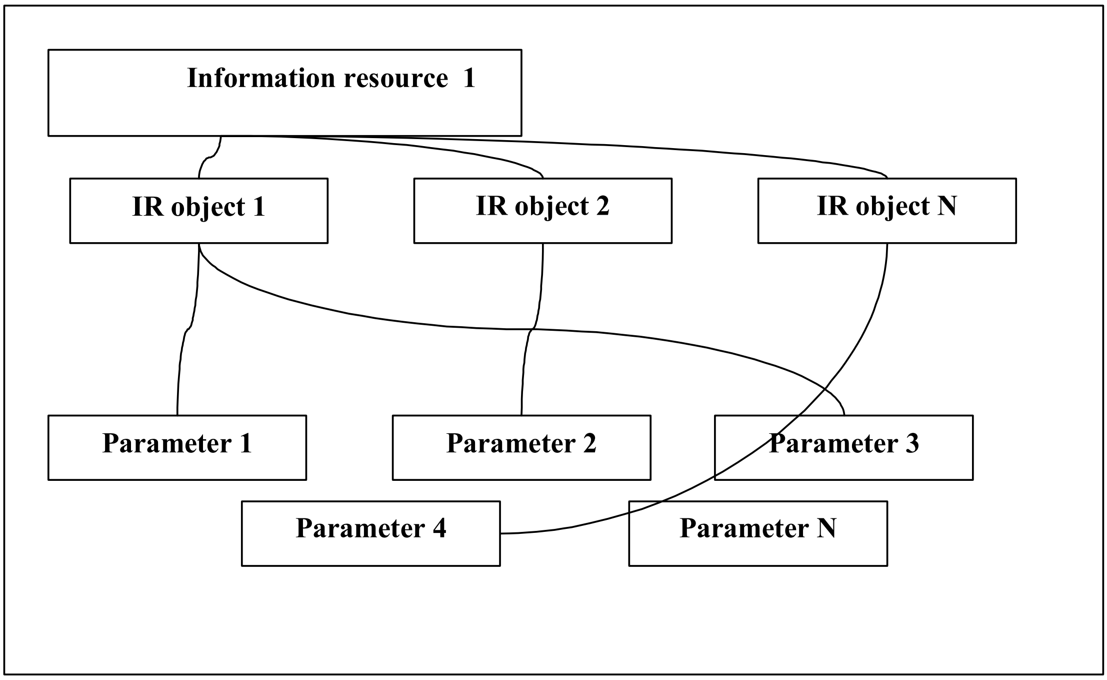


Fig.3. The structural information resource elements

***2*) information resource object (IR object)** – set of data (information) inside the single IR, defined on the semantic attributes and using the following agreements:

- the IR object represents a certain self-valuable and self-sufficient logical IR element. In most cases, the access, manipulating, processing and viewing by the client applications take place in relation to the object.
- the IR object inherits the unique identifying (as WWW service and semantic unit of information space) properties of a single IR it belongs to;
- the IR object has its own unique semantic identifier, which provides interaction inside the IR unit during the process of data management and processing on the applied level.

Examples of the IR object include:

- the single cruise ocean data in the cruise ocean data on ftp;
- the time series data on single OWC in the OWS ocean data;
- fragments (textual and graphic separately) of information on hydrochemical climate of given region on a Web-site with host page,
- GRID/GRIB field (for one deep level or one date/time) of a data file with digital prognostic hydrometeorological information.

1) **IR parameter –** the IR elementary component is a marine environmental parameter (more precisely, parameter values), which:
   1. has a place in system marine parameter classification spheres (hydrosphere, sea surface, atmosphere, land, cryosphere, etc.), or processes/phenomena (e.g., hydrosphere resources including currents, waves, etc.);
   2. has various contents: observation data; climatic, analysis and forecast data/information products;
   3. can be identified by parameter code and has a link to appropriate metadata.

Examples of the IR parameter include:

- value parameters placed in above mentioned IR objects.

2.2.2. The functional information resource elements are designed for managing of above-mentioned structural IR elements and include:

1) Information resource **inventory;**
2) information resource **request;**
3) common data exchange **format.**

The above-mentioned structural information resource elements can be represented as **metadata** and data. The IR metadata includes **system and content metadata**.

The **system metadata** are a key part (skeleton) of the uniform information space, providing the interface between clients and servers for the binding, accessing and manipulation of the resource representations. The system metadata includes metadata categories, nominated on a functional basis:

- system metadata provided semantic integrity of the IR - the unified parameter dictionary and common (as to UIS) code tables;
- system metadata provides technological integrity and consists of the IR descriptions on the above-mentioned IR element level:
  1) the IR unit metadata – IR unit descriptions ***(IR Inventory*)** which is managed under UIS,
  2) the IR object metadata – IR object descriptions (***IR object Inventory***) which are inside the IR unit;
  3) the IR element metadata – IR element description (***IR parameter Inventory***) which are inside the appropriate IR object.

The system metadata are intended to:

- provide links between thematic metadata and actual data;
- support the UIS integrity including the monitoring of UIS information contents and states;
- present a realization of the end-user interface representing the coordinated information (on IR unit, IR object and IR element levels; where, when, what are in system UIS).

The **content metadata** represents the traditional data about data, which includes the information for understanding and interpreting the data. This metadata category is managed by an independent metadata system: for example, the data sets/databases - MEDI/GCMD/EDMED, the projects - EDMERP, the experts – OceanExpert, etc. Along with the independent metadata system, the content metadata (all or part) are stored together with the actual data and in most cases the content metadata are managed by the independent systems. In this connection, we note the other major function of system metadata - to supply the integrity for the content metadata. To realize this concept the IR element inventories have attributes to support the needed links (see section 3) and it can be explained using the example of the IR unit – "the Russian NODC cruise ocean data":

- this IR unit inventory will include a link on data set description in Russian metadata format or MEDI format;
- the IR object inventory titled "Russian NODC cruise ocean data" will include a set of cruise ocean data descriptions (cruise by cruise) including links on organization, expert, project, ship and other thematic metadata to provide the end-user needed information about Russian ocean cruises under this IR unit;
- the IR parameter inventory will have links on the service (information or software application) which can give additional information for time-space distribution, method, instrument, etc., of separate parameters (temperature, salinity, oxygen, etc.) in each ocean R/V cruise (the IR object) which was included in the IR unit.
-

Common data exchange format. The most complex task of the IR model construction is the modelling of a data exchange format. First, we will be limited to consideration of the following three types of data representation:

- factographical - structured numeric/symbol values which are placed in a DBMS, data file system, etc.;
- textgraphical - textual documents, pictures, plans, photos, etc.;
- spatial - electronic vector maps, including the thematic layers (geological, hydrometeorological, ice, etc.).

For factographical representations we will use only three typical structures:

- the data with variable geocoordinates (mobile observing/measuring platforms, type "data points");
- the data with geographical coordinates fixed in space (e.g., coastal stations, fixed buoys, etc.) and by means of artificial fixing of coordinates in time (e.g., ocean data on OWS or any fixed point of sea, type "data time series");
- the data distributed in space on a certain geometrical method (modelling/calculating results and other grid data. These data have a type "data grid").

For representation of the spatial data there are two structures:

- the vector data (as complete GIS project or separate GIS layer in terms ESRI shapes);
- the graphic images of vector data.
-

For representation of the text-graphic information the two structures used are:

- The textual messages;
- The electronic documents with text and graphics.

Each of the listed structures includes two forms of representation:

1) Identification (with the account IR unit, the IR object, IR parameter):
2) values (numerical and symbolical data or references to files – textgraphic, vector, HTML).

The information resource request. To have interaction on needed information resources between the agents of the system (clients, services), we should define standards for the request exchange format which will include structured time, space and other criteria on information resource selection from the UIS. The IR requires a standard, that should be inside the above-mentioned agreements. This standard should include rules on the IR indexes and common exchange format.

The specifications of defined structural elements in the IR model are considered in section 4. These specifications, in combination with the UDP, make a linguistic basis for the uniform information space, and integrating various and multiple structural and geographical representations of the information resource.

3. The unified parameter dictionary and common code table

Unified Dictionary of Parameters (UDP) is designed for unified and standardized data and information representation which are placed in various existing system nodes. As an example, there is a Russian UDP which is developed under ESIMO[8]. The Russian UDP is based on the information resources and classification scheme listed below:

- Sphere (cryosphere, hydrosphere, land surface, lithosphere, low atmosphere, surface of sea);
- Process (phenomenon);
- Observations/measurements parameters;
- Statistic (derived variable).

The UDP, with the mentioned basic attributes for parameters, provides a number of additional attributes. The essentials for the standardization of data access and exchange on the basis of various sources of marine information includes: units and observation methods, time and space scale, derived variables and other. A list of dictionary attributes and a brief description is given below:

Attribute name	***Description***
ID PARAMETER	Unique identifier
PARAMETER TYPE	The UDP provides description as single parameters as groups of parameters, bounded with obtaining conditions – methods, equipments, etc.
GF3_CODE	Parameter code according to BODC/GF3 code
SPHERE	Sphere code (3 digits number from the code table CODETYPE)
SPHERE_NAME_RUS	Sphere name (Russian)
SPHERE_NAME_ENG	Sphere name (English)
PROCESS	Process code (3 digits number from the code table CODETYPE)
PROCESS_NAME_RUS	Process name (Russian)
PROCESS_NAME_ENG	Process name (English)
STATISTIC	Statistic code – derived variable, obtained from parameter value, used for information production description (3 digits number from the code table CODETYPE)
STATISTIC_NAME_RUS	Statistic name (Russian)
STATISTIC_NAME_ENG	Statistic name (English)
NAME_RUS	Full parameter name (Russian)
NAME_ENG	Full parameter name (English)
NAME_OUT_RUS	Abbreviation for review (Russian)
NAME_OUT_ENG	Abbreviation for review (English)
ACCURACY	Parameter accuracy, format - 9999.9999
UNIT	Parameter unit code from code table UNITS_M, represented in string format
UNIT_NAME_RUS	Parameter unit name brief (Russian)
UNIT_NAME_ENG	Parameter unit name brief (English)
UNIT_SHORT_NAME_RUS	Unit name full (Russian)
UNIT_SHORT_NAME_ENG	Unit name full (English)
TIME_SCALE	Time scale code from code table CODETYPE
TIME_SCALE_NAME_RUS	Time scale name (Russian)
TIME_SCALE_NAME_ENG	Time scale name (English)
SPACE_SCALE	Space scale code from code table CODETYPE
SPACE_SCALE_NAME_RUS	Space scale name (Russian)
SPACE_SCALE_NAME_ENG	Space scale name (English)
MODULE	Equation for computed (measured) value
VERT_DIMENS_MIN	Minimum value (range)
VERT_DIMENS_MAX	Maximum value

Attribute name	*Description*
CODE_NAME	Code from international code system
METHOD	Parameter method (observation/measurement, calculation)
METHOD_NAME_RUS	Method name (Russian)
METHOD_NAME_ENG	Method name (English)
DESCRIPTION	Plain text with parameter description and link with additional information
CREATED	YYYY-MM-DD When parameter was created
MODIFIED	YYYY-MM-DD When parameter was modified

The UDP and code tables are updated via a DBMS and there is developed software for transferring data from the database into an XML-file, as well as access to the UDP via a browser (this can serve as an example of a system service). The Russian UDP on-line software has the following features:

- programming language - Java (Servlets + JSP + EJB technologies).
- middleware – application server Jboss - Tomcat 3.0.4 and DBMS Oracle 9i;
- software access - http://data.meteo.ru:8080/udopweb/eng/index_en.jsp.

The software realizes two main functions: search and viewing of the parameter descriptions, and saving the selected parameter descriptions in a user "bag" in XML and ASCII-files.

4. The structure of functional information resource elements

As we defined previously, the IR model includes the following functional elements:

1) information resource inventories;
2) information resource request;
3) common data exchange format.

These elements of the IR model consist of specification sections and blocks (by analogy with the Keeley bricks [see 17]):

The section is a component of the IR element specification including specification blocks. The section:

- can be used for various structural elements of the IR model;
- can be missing in the IR element description [M (0)], or can have several description samples within one IR element description [M (1, n)].

The **block** is an elementary part of the IR element specifications. The block:
- can be used for various structural elements of the IR model;
- can be missing in the IR description [M (0)], or can have several description samples within one IR description [M (1, n)].
- can be a key element (KEY) or a characteristic (CHA);

The IR element specifications includes 19 blocks inside 6 sections (Annex 1). The 6 sections include:

- IR identification (URI and system identifiers, other);
- IR representation properties (technical);
- IR (semantic) contents (what, where, when);
- IR interconnection with semantic metadata;
- IR exchange data (the IR fragment) specification
- IR life cycle.

4.1. The IR unit inventory structure

The IR unit inventory includes the following sections:

- IR identification;
- IR access specifications;
- IR (semantic) contents (what, where, when);
- IR interconnection with semantic metadata;
- IR life cycle.

Structure:

Section *Gen_Res* : IR identification (M[1])
Block Id_Res – IR identification

Section *Tspec_Res* : IR representation properties
Block ID_Inv – Integrator link specification (M[1])
Block Spec_Res - Technical characteristics of IR representation (M[1])

Section Cont_Res : IR content

Block ID_PTerm according to UDP (Unified dictionary of parameters) (M[1:n])
Block Loc_Res - Geographical region (M[1])
Block Loc_Resadd - additional description for geographical region (M[1:n])
Block Time_Res - time characteristics for IR (M[1])
Block HeightDeep_Res - IR characteristics for height above surface and sea (ocean) depth

Section *Connect_Res* : IR links specification
Block Metadata_Dec [M (0:n). – IR links description with metadata objects (project, equipment, methods; set/flow data); ROSCOP/ODAS and other

Section Life_cycle : IR life cycle
Block Status_Res – IR status [M (0:1)].
Block Provider_Res –IR provider information [M (1)]
Block Creator_Res – Information about IR creator (can be different from provider) [M (0:n)]
Block Life_Res – IR life cycle dates

Example:


```
<?xml version=“1.0”?>
<T>
 <ID_Res>
  <Id_R>90_RIHMI_IR1</Id_R>
  <Id_L>00001</Id_L>
  <Title_R>Resource N1</Title_R>
 </ID_Res>
 <ID_Res>
  <Id_Service>rus_IBD_cruise_data_01</Id_Service>
  <Id_URI>http://data.meteo.ru:8080/IRTest/</Id_URI>
 </ID_Res>
 <Spec_Res>
  <Frequency>on event</Frequency>
  <Forma>table</Forma>
  <Method>application</Method>
  <Constraint>all_user</Constraint>
 </Spec_Res>
 <ID_Inv>
  <Form>XML</Form>
  <URI>http://data.meteo.ru:8080/IRTest/catalog/IR_o1.xml</URI>
```

```
 </ID_Inv>
 <ID_PTerm>
  <Keyword>russian, cruise,platform,profile,hydrology/hydrochemistry</Keyword>
  <Category>Earth science</Category>
  <Topic>Hydrosphere</Topic>
  <Term>Hydrochemistry</Term>
  <Detailed_Term/>
  <N_Variable/>
  <C_Variable/>
 </ID_PTerm>
 <Loc_Res>
  <Latitude_Max>+90</Latitude_Max>
  <Longitude_Max>+180</Longitude_Max>
  <Latitude_Min>-90</Latitude_Min>
  <Longitude_Min>-180</Longitude_Min>
 </Loc_Res>
 <Loc_Resadd>
  <Location_Name>North Atlantic Ocean,Northeastern Atlantic Ocean,Mediterranean Sea,Black Sea</Location_Name>
  <Lgeo>http://data.meteo.ru:8080/shapes/region.shp</Lgeo>
 </Loc_Resadd>
 <Time_res>
  <Start_Date>1930</Start_Date>
  <Stop_Date>2002</Stop_Date>
 </Time_res>
 <Metadata_Dec>
  <Identifier_M/>
  <Form>XML</Form>
  <M_URI>http://data.meteo.ru:8080/resource/objects.db</M_URI>
 </Metadata_Dec>
 <Provider_Res>
  <Identifier_Org>26</Identifier_Org>
  <Title_Org>RIHMIWDC</Title_Org>
  <Creator>Belov S.V</Creator>
  <Telephone>0843974953</Telephone>
  <Email>belov_sergey@meteo.ru</Email>
 </Provider_Res>
</T>
```


4.2 The IR object inventory structure

The IR object inventory includes following sections

- IR object identification;
- IR object content;
- IR object specification for links with metadata and other IR (derived IR);
- IR life cycle

Structure:

Section *Gen_Res* : IR identification (M[1])
    Block Id_Res – IR identification

Section Cont_Res: IR content
    Block ID_PTerm according to UDP (Unified dictionary of parameters) (M[1:n])
    Block Loc_Res - Geographical region (M[1])
    Block Loc_Resadd - additional description for geographical region (M[1:n])

Block Time_Res - time characteristics for IR (M[1])
Block HightDeep_Res - IR characteristics for height above surface and sea (ocean) depth
Block Type_Res - IR characteristics for requested IR content (forecast, climate,..), time and spatial scale

Section *Connect_Res*: IR links specification
Block Metadata_Dec *[M (0:n)*. – IR links description with metadata objects (project, equipment, methods; set/flow data); ROSCOP/ODAS and other

Section *Life_cycle*: IR life cycle
Block Status_Res – IR status [M (0:1)].
Block Provider_Res –IR provider information [M (1)]
Block Creator_Res – Information about IR creator (can be different from provider) [M (0:n)]
Block Life_Res – IR life cycle dates

Example:


```
<Object_Description>
- <ID_Res>
<Id_R>cruise14536</Id_R>
<Id_L>90_RIHMI_obj1</Id_L>
<Title_R>Russian cruise data (YkrSCES (Odo GOIN) )</Title_R>
  </ID_Res>
- <ID_Res>
<Id_Service>rus_IBD_cruise_data_01</Id_Service>
<Id_URI>http://data.meteo.ru:8080/cruisecat/cruisedetails?cruiseId=14400&l=eng</Id_URI>
  </ID_Res>
- <ID_PCode>
<N_Variable>pH</N_Variable>
<C_Variable>41</C_Variable>
  </ID_PCode>
- <ID_PCode>
<N_Variable>oxygen</N_Variable>
<C_Variable>47</C_Variable>
  </ID_PCode>
- <ID_PCode>
<N_Variable>temperature</N_Variable>
<C_Variable>229</C_Variable>
  </ID_PCode>
- <ID_PCode>
<N_Variable>salinity</N_Variable>
<C_Variable>227</C_Variable>
  </ID_PCode>
- <Loc_Res>
<Latitude_Max />
<Longitude_Max />
<Latitude_Min />
<Longitude_Min />
  </Loc_Res>
- <Loc_Resadd>
<Location_Name />
<L-geo />
  </Loc_Resadd>
- <Time_Res>
<Start_Date>1983-02-16</Start_Date>
<Stop_Date>1983-05-23</Stop_Date>
  </Time_Res>
- <HeightDeep_Res>
<Height_Max>0</Height_Max>
<Deep_Max>500</Deep_Max>
  </HeightDeep_Res>
- <Metadata_Dec>
<Identifier_M />
<Form />
```

```
<URI />
  </Metadata_Dec>
  </Object_Description>
- <Object_Description>
- <ID_Res>
<Id_R>90_RIHMI_obj2</Id_R>
<Id_L>cruise_4631580</Id_L>
<Title_R>Russian cruise data (Azerbydj. HM Service)</Title_R>
  </ID_Res>
- <ID_Res>
<Id_Service>rus_IBD_cruise_data_02</Id_Service>
<Id_URI>http://data.meteo.ru:8080/cruisecat/cruisedetails?cruiseId=32747&l=eng</Id_URI>
  </ID_Res2>
- <ID_PCode>
<N_Variable>Nitrite</N_Variable>
<C_Variable>38</C_Variable>
  </ID_PCode>
- <ID_PCode>
<N_Variable>pH</N_Variable>
<C_Variable>41</C_Variable>
  </ID_PCode>
- <ID_PCode>
<N_Variable>Oxygen</N_Variable>
<C_Variable>47</C_Variable>
  </ID_PCode>
- <ID_PCode>
<N_Variable>Chlorine</N_Variable>
<C_Variable>54</C_Variable>
  </ID_PCode>
- <Loc_Res>
<Latitude_Max />
<Longitude_Max />
<Latitude_Min />
<Longitude_Min />
  </Loc_Res>
- <Time_Res>
<Start_Date>1985-08-20</Start_Date>
<Stop_Date>1985-08-21</Stop_Date>
  </Time_Res>
- <HeightDeep_Res>
<Height_Max />
<Deep_Max />
  </HeightDeep_Res>
- <Metadata_Dec>
<Identifier_M />
<Form />
<URI />
  </Metadata_Dec>
  </Object_Description>
```


4.3 The structure of IR parameter inventory

The structure is under consideration.

4.4. The information resource request specifications

The request structure defines a request for one information resource. The request structure consists of 3 logical segments:

- IR identification in request;
- Request content;

- Request address.

The IR request is based on the model blocks. The structure is under consideration.

4.5. The common data exchange specifications

The common data exchange structure defines a request for the IR fragment as a response of a single information resource. This common data exchange structure consists of 4 logical segments:

- IR identification in response;
- IR content in response;
- IR structure in response;
- Responce data.

The common data exchange structure is based on the model blocks and is now under consideration.

5. The functioning scheme of the system as a virtual ocean data centre

Practically, the distributed marine information resource system must function as a ***Virtual Oceanic Data Center (VODC) providing forming, support and use of*** the World ocean unified information space (fig.4.). Functioning of a VODC is considered below.

The data centers (IODE and other which were involved in the distributed marine information resource system) provides functioning of VODC nodes, including support of:

- Software components (middleware environment, integrators and etc.);
- information components (system and content metadata, information resource).

The Client (end-user or any other system node) sends the request to the Navigator, which in turn transforms the client request to the elementary request appropriate for the system nodes whose information resources satisfy the client request. The local software (integrators) execute requests and transmit the results to the Navigator. The navigator compiles the complete response to the request and transmits the data to the user.

The basic function of the Navigator is to analyze a user request in the context of the VODC IR content, to define the list of system nodes able to meet the request and to ensure interaction with these system nodes.

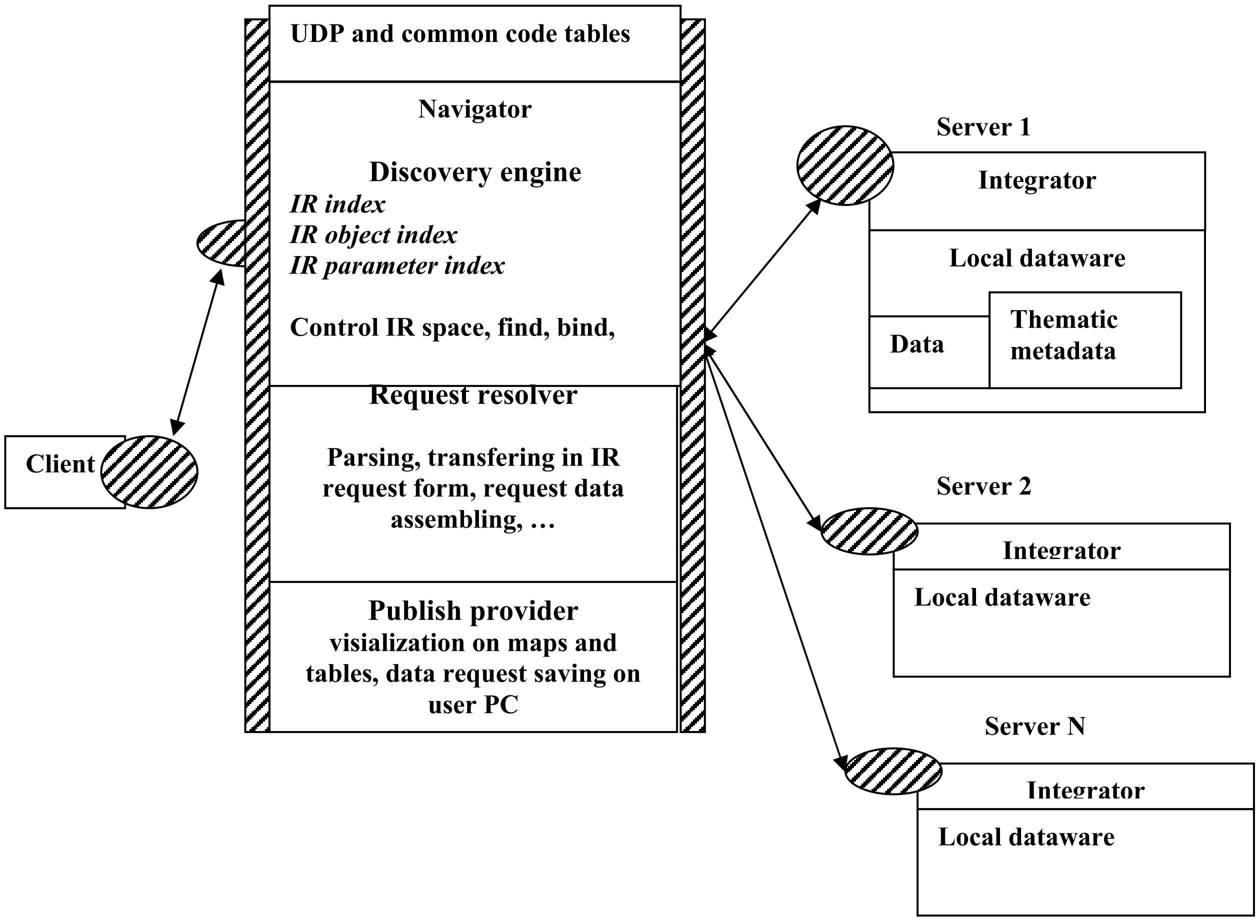


Fig. 4. The data management scenario of the VODC functioning.

The function of the Integrators is to execute specific elementary requests for data. These requests arrive at the system node from the Navigator as a result of a user request. The processing and decomposition steps are:

(i) an elementary request comes to the system node Integrator and is transmitted to the Request Decoding Service to be converted to appropriate semantics for the local dataware;
(ii) the selected data are converted from the local data structure (format) to the system common exchange format (e.g., XML, ASCII);
(iii) the selected data are sent to the Navigator to be incorporated into the integrated IR fragment meeting the user request specifications.

6. IR model testing and the role of XML

At the first stage, the practical realization of the IR model assumes the interconnection between the model elements and creation of a test software version. The test version of the IR model includes:

1) the IR unit index, created for two single IRs - Russian ocean cruise and Ocean cruises of other countries;
2) the IR objects index, created for a number of cruises for each IR unit.
3) these IR elements and IR request format, described based on XML.
4) the applied software (Java, JSP and HTML). This realizes the functions of the Navigator (to find the IR units and the IR objects on request) and Integrators (to provide to IR) are located on a demonstration PC imitating the single entry into the system;
5) actual ocean data of defined IRs are on web-server of RIHMI-WDC (powered by JBoss).

Annex 1 - The Specifications of the Information Resource Description

Section 1: Id_Res - the IR identification.

Block: Id_Res – IR identification (semantic) in the given distributed information resource system (M[1])

Id_R -unique system semantic identifier (country code, brief name of organization, a IR number in system node, e.g. 90_RIHMIWDC_1)

Id_L - Unique local identifier for interconnection with local dataware in the data center

Title_R – IR brief title

Block: Id_Res – IR identification in Web-space (M[1])

ID_Service – Web-service type (Web-site, applications, local system, …) (table 8.)
Id_URI - IR identifier using URI scheme с учетом типа Web-service

Section 2 : Tspec_Res - the IR access specification

Block Spec_Res - the characteristics of the IR representation (M[1])

Frequency - Frequency of update (daily, monthly, yearly, others- table 13)
Forma – form of IR representation (factographic, text-graphic, space)
Method – access method to IR (Web-resource, Web-site, application, local system, …)
Constraint - user restrictions (free access, authorization and other, table 10)

Block ID_Inv – Integrator link to the IR specification (M[1])

Forms – format link description (“Table” – ASCII file with delimiters “;”, “XML” – XML-file, etc)
URI – full URI –address of the IR description depending on form

Section 3: Cont_Re – the IR content.

Block ID_PTerm according to UDP (Unified dictionary of parameters – tables 1-2) (M[1:n])

Category – Earth Science and etc.
Topic – Ocean-, atmo-, lito-, techno-, sociosphere and etc.
Term – Processes (wave, current, level, …), or type of information (scientific, legal, methodical)
Detailed_Term – detailed information (e.g. wave – wind and swell)
N_Variable – Parameter name (from UDP)
C_Variable – Parameter code (fro UDP)

Block Loc_Res - *Geographical location of the IR* (M[1])

Latitude_Max – Latitude max
Latitude_Min – Latitude min
Longitude_Max – Longitude max
Longitude_Min - Longitude min

Block Loc_Resadd - additional description for geographical region (M[1:n])

Location_Name – Geographical region of IR (IHB)
L-geo – Links on ship-file (contour of region)

Block Time_Res - time characteristics for the IR (M[1])

Start_Date – Date of beginning of observations and data summarizing and processing; time of beginning of forecast operation
Stop_Date – Date of finishing of observations and data summarizing and processing, time of finishing of forecast operation

Block HightDeep_Res - IR characteristics for height and depth

Height_Max – max height (meters), surface =0
Deep_Max – max depth (meters),ocean bottom – “bottom”

Block Type_Res - IR characteristic by content

Id_Type – **observation data, climate, analysis, forecast and other – table 9**
Id_Stream –characteristics of the time period from observation to presentation-table 12
Id_Form – Completeness of IR presentation (description only, description and sample – table 11
Spatial_Coverage – Spatial dimension of IR (fixed point, 1°,2°, …, table 4)
Temporal_Resolution – Temporal dimension of data – table 5

Section 4: Connect_Res – the IR link specification

Block Metadata_Dec– IR links on the metadata objects (project, equipment, methods) *[M (0:n).*

Identifier_M – unique identifier metadata object and (or) other IR
Form – format representation (“Table” – ASCII file with delimiters “;”, “XML” – XML-file, etc)
M_URI – full URI –address on metadata object

Section 5: Dec_Res –

Block Cont_Res– general description of exchanged IR fragment [M (1)].

C_error - return code (error table)
N_par – number of strings in exchanged IR fragment

Block Str_Res – exchanged IR fragment structure description *[M (0:n)].*

C_Variable – parameter code from UDP
T_ Variable – parameter type - "KEY(I)" или "KEY(U)" или "MIT" или "CHA"

Section 6 Life_cycle: IR lifecycle

Block Status_Res – IR status *[M (0:1)].*

Version – IR version
Status – IR status

Block Provider_Res –IR provider information *[M (1)]*

Identifier_Org – abbreviation of the IR organization -provider (15 chars)
Title_Org – org –provider full name
Creator – author name/surname
Telephone- phone
E-mal – e-mail

Block Creator_Res – Information about IR creator (can be different from provider) *[M (0:n)]*

Identifier_Org – abbreviation of IR organization -creator (15 chars)
Title_Org – org –creator full name
Telephone- phone
E-mal – e-mail

Block Life_Res – IR Lifecycle dates

DIF_Creation_Date – Date of IR description entry
Last_DIF_Revision_Date - Date of IR description last update

System Code Tables for Information Resource Description

1. Sphere

Code	Sphere
100	**Lithosphere**
111	Land surface
121	Under land surface soil layer
200	**Hydrosphere**
211	Sea surface
222	Under sea surface layer
223	Bottom layer
300	**Cosmo sphere**
310	Solar Physics
320	Ionosphere
400	**Atmosphere**
410	Low atmosphere
420	High atmosphere
430	Stratosphere
440	Stratopause
450	Troposphere
460	Tropopause
470	Ionosphere
500	**Atmosphere – Hydrosphere**
600	**Cryosphere**
700	**Biosphere**
800	**Sociosphere**
900	**Technosphere**

2. Processes and phenomena

Code	Science, process, phenomenon
100	**Lithosphere**
110	Marine Geophysics
111	Marine Sediments
200	**Hydrosphere**
210	Sea Surface Height
220	Ocean Waves
224	Ocean Acoustics
230	Ocean Circulation
242	Ocean Chemistry
243	Salinity/density
244	Ocean Pressure
245	Ocean Temperature
246	Ocean Optics
270	Coastal Processes
300	Atmosphere
310	Solar Physics
414	Convection
430	Paleoclimate
440	Atmospheric Phenomena
550	Ocean Heat Budget
600	Cryosphere
610	Snow
620	Fresh ice
650	Sea Ice
700	**Biosphere**
701	Aquatic Habitat
702	Ecological Dynamics

703	Fungi
704	Microbiota
705	Terrestrial Habitat
706	Vegetation
707	Wetlands
720	Human Dimensions
730	Agriculture
740	Zoology
800	**Sociosphere**
900	**Technosphere**

3. Temporal dimension

Code	Description
00	Initial data
01	Daily data for a specific month and specific year
02	Monthly data for a specific year
03	Seasonal data for a specific year
04	Annual data
05	Climatic long-term monthly data
06	Climatic long-term seasonal data
07	Climatic long-term daily data
08	Climatic long-term annual data
09	Decade data for a specific month and a specific year
10	Decade data for a specific year
11	Climatic long-term decade data
21	Forecasting for a day
22	Forecasting for a week
23	Forecasting for a month
24	Forecasting for a season
25	Forecasting for a year
99	Unknown

4. Spatial dimension

Code	Description
00	Non regular
Geographical region	
01	10 degree latitude zone
02	10 degree longitude zone
Regular trapezium	
10	10 degree zone
11	5 degree zone
15	2.5 degree zone
16	1 degree zone
17	30 minute zone
18	15 minute zone
Irregular trapezium	
30	2° latitude, 3° longitude zone
31	2.5° latitude, 1.5° longitude zone
40	15’latitude , 30’longitude zone
41	20’latitude , 30’longitude zone
42	30’latitude , 30’longitude zone
Fixed point	
50	*Fixed point on the land ,regular time series*
51	*Fixed point in the estuary, regular time series*
52	*Fixed point on the coast, regular time series*
53	*Fixed point in the open sea, non-regular time series*
Grid	

Code	Description
60	20’*30’*grid*
61	30’*30’ *grid*
62	5^0*5^0 *grid*
63	10°*10° *grid*
64	1°*1° *grid*
65	5°*10° *grid*
67	5’*5’ *grid*

5. Regularity of observations

Code	Description
00	Unknown
10	< 1 hour
11	30 “
12	1 ‘
13	2 ‘
14	5 ‘
15	10 ‘
16	15 ‘
17	30 ‘
20	1 hour
30	< 1 day (time)
40	Daily observations (1 time/day)
50	Weekly observations (1 time/5-7 days)
60	Ten-day period observations (1 time/10 days)
70	Monthly observations
80	Seasonal observations (2-4 time/year)
99	Non regular observations

6. Logical units of data flow

Code	Description
00	Not known
10	*Initial data*
11	Data from a cruise
12	Month data from a coastal station
13	Data from satellite flight
14	Hourly observation from a buoy
15	Time of observation for GTS
16	Time of observation for voluntary ships
20	*Data transformation*
21	Square
22	Time series
23	Sections
30	*Data interpolation*
31	Vertical *interpolation* on standard horizons
32	Vertical *interpolation* on isopycnics
33	Space *interpolation*
34	Time *interpolation*
40	*Data calculation*
41	*Calculation of* parameters from an oceanographic station
42	Statistical characteristics of parameters
43	Statistical characteristics of c*alculated* parameters
50	*Metadata*
51	Information on data sets
52	Information on observing platforms
53	Cruises descriptions
54	Space-time coverage data

70	*Forecasts*
71	Short-range
72	Long-range
73	Very-long-range

7. Form of IR presentation

Code	Description
Fact	Digital data
Text	Textual
Space	Spatial
Graphic	Graphical
Sound	Sound
Mixed	Mixed

8. Web-service type

Code	Description
http	Web-site (World Wide Web)
ftp	FTP server
File	Files in a local computer
rlogin	Application with a data base
Clogin	Application – integrator (converter)

9. Resource type

Code	Description
Observ	Initial data
Sum	Statistical and climatic data
Analysis	Diagnostic and analytical information
Forecast	Forecasting information

10. IR Availability

Code	Description
all_user	Free
middle_user	Access through registration or/and payment
high_user	Access for VIP
esimo_user	Access for project partners

11. IR completeness

Code	Description
Exz	Samples
Dec	Description
f_inf	Full information, regular update

12. Operation of data

Code	Description
Rt	Real time data (delay up to 7 days)
Dt	Delay up to 1 year
Ht	Historical data (delay for more than 1 year)

13. The frequency of updating

Code	Description
Hour	Hourly
Day	Daily
Month	Monthly
Season	Seasonal
Year	Annual

14. Presentation form

Code	Description
Map	Map
Graphic	Graphic
Images	Images
Tables	Tables
Db	Data base
Txt	Text
Mix	Mixed forms

References

1) GOOS, Data and Information Management Strategy and Plan, June 2001.
2) GCOS/GOOS/GTOS Joint Data and Information Management Plan, Version 1.0, May 2000, GCOS – 60/GOOS – 70/WMO/TD No. 1004.
3) GCOS, Data and Information Management Plan, version 1.0, 1997.
4) The Integrated, Strategic Design Plan for the Coastal Ocean Observations Module of the Global Ocean Observing System, 2002.
5) EuroGOOS data management pan, 2000.
6) Second meeting of the Inter-programme Task Team on future WMO Information Systems, Final report, 2000
7) The Basic Elements of the End to End Data Management Strategy. Informal meeting of JCOMM Expert Team on Data Management Practices (EDTMP) (Brussels, 28-29 November 2002).
8) Unified System of Information on the World Ocean Conditions (ESIMO, Russian Federation). Web-portal - http://data.meteo.ru:8080/resource/index_en.jsp. 2003.
9) Representation state transfer (REST). Doctoral thesis of R. T. Fielding., 2000.
10) Architectural principles of World Wide Web(draft). W3C Technical Architecture Group, 2002.
11) Principled design of the modern Web architecture. R. T. Fielding and R.N. Taylor .
12) Web characterization terminology and definition sheet. W3C working draft, 1999.
13) Glossary for the OASIS Web-service Interactive Applications (WSIA/WSRP), 2002.
14) Grid Physics Network project. http://eu-datagrid.web.cern.ch/eu-datagrid/
15) Web Services Glossary. W3C working draft, 2002.
16) Uniform resource identifiers (URI). Generic syntax. W3C RFC 2396, 1998.
17) Isenor, Anthony W., J. Robert Keeley and Joe Linguanti. 2003. Developing an eXtensible Markup Language (XML) Application for DFO Marine Data Exchange via the Web, DRDC Atlantic ECR 2003-025, Defence R&D Canada – Atlantic.

ANNEX 6: BRIEF REPORT ON BRICK APPLICATION TO XML

Canadian Investigations of the Keeley Bricks With Application to Profile Data Transfer Using XML

Anthony W. Isenor
Defence R&D Canada – Atlantic
anthony.isenor@drdc-rddc.gc.ca
DRDC Atlantic SL 2003-029

Introduction

The 2002 SGXML meeting in Helsinki, Finland, outlined various action items for the Group. In particular, Action Items 5 and 7 under subsection 6.2 [1] identified a specific point data investigation. This report briefly describes the development resulting from this investigation.

The development involved three Canadian data centres: the Institute of Ocean Sciences (Sidney, B.C.), the Marine Environmental Data Service (Ottawa, Ontario) and the Bedford Institute of Oceanography (Halifax, N.S.). A detailed report of the development is available [2, 3]. Here, a summary is presented and details of the data structure are described.

Background

The 2002 Helsinki meeting introduced the SGXML to the concept of generic data objects for packaging ocean data. These objects were called Keeley Bricks. The original concepts behind the bricks were based on the work of Bob Keeley, Marine Environmental Data Service, Canada.

The Keeley Bricks are essentially data objects that group associated data and information. The term "bricks" is used because these objects resemble construction bricks in that they can be assembled in different ways to produce different structures. This represents the two fundamental principles behind the Keeley Brick concept: 1) exploit the natural grouping of data and information into well-defined objects, and 2) arrangement of the bricks in various ways to imitate the natural structures found in a variety of ocean data types.

A Canadian investigation has constructed the initial set of bricks required to define a structure appropriate for ocean profile data. This work included full brick definitions. The work also extended into an application of the bricks in an XML environment. This resulted in brick definitions that focused on the associated XML syntax. Thus, the brick definitions have moved beyond the abstract, to specific implementation details related to elements and attributes in an XML environment.

The bricks resulting from the Canadian development are summarised in Table 1. Both pure and compound bricks are identified. Compound bricks are specific to the XML application. Compound bricks are wrappers around other compound or pure bricks. Pure bricks contain data and information content. In the XML application, this content is stored in the elements and attributes.

The profile data structure resulting from the development is shown in Figure 1. Here, compound bricks are identified in green (also identified by the trailing string "_set") while pure bricks are red. The figure also shows the XML occurrence of the brick. The square brackets enclose the minimum and maximum occurrence of the brick using the notation [i, j] with i representing the minimum and j the maximum.

An example of encapsulating water sample data in the XML structure is provided in Annex 1. Only one record of data is shown. The data record contains five data values.

Detailed information on the bricks and the profile data structure is available [3]. In this brief report, some of the specific implementation decisions are described. The reader is assumed to have an intermediate knowledge of XML [4, 5, 6].

Parameter Codes

Parameter codes represent the coding of information related to the variables in a dataset. Codes are often stored in systems called parameter dictionaries. Within the XML profile data structure, the dictionary can be described by name using the data_dictionary brick. This brick is contained in the data_collection compound brick at the highest level of the profile structure (see Figure 1).

At the individual XML datum level, the specification of a parameter code may occur in one of three forms:

1) the code as a tag name,
2) the code as tag content, and
3) the code as attribute content.

The parameter code existing as a tag name was not considered viable. In essence, this option would allow for many thousands of tags, as there are many thousands of oceanographic parameters. This option would also eliminate the usefulness of the schema validation process. This is because one would not be able to create a schema that identified all possible combinations of parameters.

Considering option 2), the parameter code could easily exist as content for the tag. However, if the data value is also stored in tag content, then the direct connection (or

Table 1. The list of bricks and definitions. Pure and Compound bricks are described in the text.

BRICKS	Brick type	DEFINITIONS
analysis_method	Pure	Stored information about any physical, chemical or biological analyses carried out on the data
availability	Pure	Stores information about the possible release of the data to the public
calibration	Pure	Stores calibration information on the instrument, sensor or variable
comment	Pure	Stores general textual information not intended to be used in data retrievals
data_collection	Compound	Used to encapsulate the entire XML file
data_dictionary	Pure	Indicates the specifics of the data dictionary being used within the collection
data_point	Pure	Used to store any type of data or metadata value
data_set	Compound	Used to encapsulate a dataset at a defined level of granularity
data_set_id	Pure	A numeric or text identifier for a particular data set
depth_pressure	Pure	Store the z coordinate of the data
history	Pure	Processing history of the data
history_set	Compound	Used to encapsulate history information
instrument	Pure	Information about the instrument used to make the measurements.
latitude	Pure	The y coordinate of the data
ldate	Pure	The time coordinate of the data
location_set	Compound	Used to encapsulate the x, y, z, t values.
longitude	Pure	The x coordinate of the data
previous_value	Pure	Information about the value before it is changed
provenance	Pure	The originator of the data
quality	Pure	A marker providing an assessment of data quality
quality_testing	Pure	Information about how the data quality assessment was made
sampling	Pure	Information about the sampling methods used
sensor	Pure	Identifies sensor specifics
units	Pure	The units of measurement
variable	Pure	Information about the variables measured.
variable_set	Compound	Used to encapsulate all the information required to declare a variable

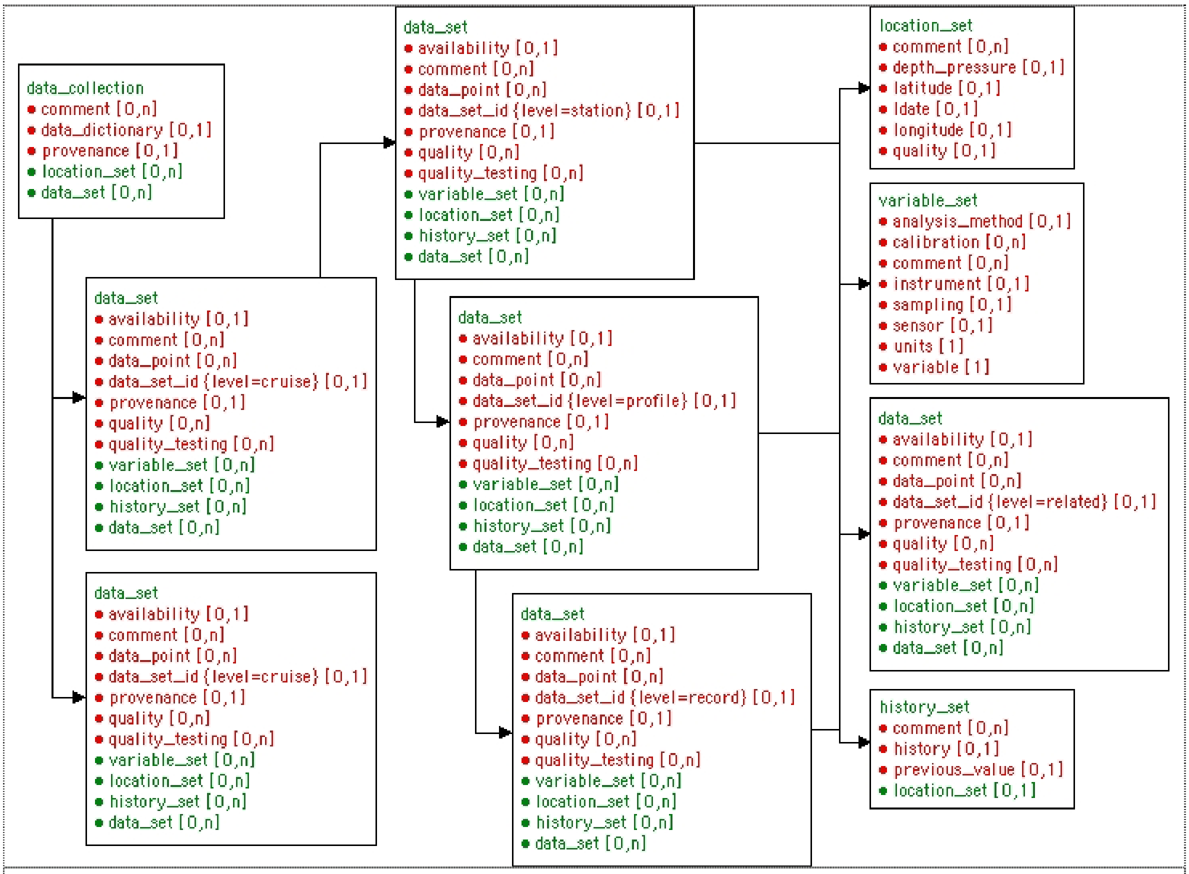


Figure 1. The profile data structure using Keeley Bricks. The red text indicates pure bricks, while the green text indicates compound bricks. Compound bricks may also be identified by the name, which typically ends in the string "_set" (the exception being data_collection, which is also a compound brick). The [i, j] notation indicates the [minimum, maximum] occurrence of the brick. The arrows indicate an expansion of one element into sub-elements in the XML application. Components of the structure are described in the text. The details and definitions of the bricks may be found at [3].

encapsulation) between the code and the data value is lost. This connection could exist if the attribute of the tag contained the data value.

Option 3) was selected for this project. This option packages the code with the data value in a single XML element. As well, the option leaves open the possibility of using a list of allowable parameters for the attribute content.

The pt_code attribute (see Annex 1, data_point tag) is used to store parameter code information in various bricks. As well, a pt_link attribute may be used to provide a counter on the pt_code. The pt_link attribute allows for the same parameter occurring more than once in a dataset, as might occur if oxygen values are determined using two different methods.

The use of the schema in document validation is an important functionality within XML. The schema defines allowable structures for the XML documents. The XML environment provides tools that compare and report on the compliance of a particular XML document as compared to the schema. Content rules may be built into the schema, thereby reducing the requirement on developed software. An example content rule may be a range for latitude between –90 and +90 degrees.

data_set Compound Brick

The data_set compound brick (see Figure 1) is an important construct within the profile structure. It was recognized that the term "dataset" means a variety of things to different people. With this in mind, the data_set compound brick contains an identifier brick, data_set_id, which identifies the particular level of granularity of the data_set.

The data_set compound brick contains the necessary compound and pure bricks to describe:

- the availability of the dataset for distribution,
- free-format comments,
- data points within the dataset,
- a dataset identifier,
- the owner of the dataset,
- the quality tests applied and the results of those tests,
- the variables contained within the dataset,
- spatial-temporal location information for the dataset,
- the processing history of the dataset, and
- a dataset at a finer level of granularity.

A dataset within a dataset exploits the natural hierarchy familiar to oceanographic data collections (e.g., many profiles at a station, and many stations within a cruise). The last bullet in the above list indicates this concept.

There are four bricks of particular importance within the data_set compound brick: data_point, data_set_id, variable_set and location_set. These will now be described in more detail.

The data_point brick is used to encapsulate data or metadata relevant to the particular data_set level. The content of data_point is the actual data or metadata. data_point has four attributes (see Figure 2) which contain the parameter code, a point link, statistical properties and the type of content. Type can be numeric, character, date, time or date/time. The statistic attribute identifies if the data_point is a statistical measurement such as a mean, standard deviation, etc.

data_point {**pt_code**, pt_link, statistic, typing}
Figure 2. The data_point brick. The content for data_point is any data or metadata value for the dataset. Attributes of the brick are shown inside {}. The bold text indicates mandatory content.

The data_set_id brick (Figure 3) allows for identification of the particular dataset. The content of this brick is a unique dataset identifier. The level attribute for the brick identifies the granularity of the dataset. For example, allowable content in the level attribute includes cruise, station, profile, and record. This content was developed for the profile structure. Other content may be required for other data types.

data_set_id {**level**}
Figure 3. The data_set_id brick. The content of data_set_id is a unique identifier for the dataset. Attributes of the brick are shown inside {}. The bold text indicates mandatory content.

The variable_set compound brick contains all the information to describe a variable or parameter used within the XML document. This compound brick (Figure 4) is comprised of bricks that allow for descriptions of:

- analysis methods used on particular variable,
- calibration information for the variable,
- free-format comments,
- information on the instrument used to collect the variable,
- information on the sampling used to collect the variable,
- information on sensors used to collect the variable,
- units of the variable, and
- variable specific information such as name, limits, accuracy and precision.

variable_set
- analysis_method
- calibration
- comment
- instrument
- sampling
- sensor
- **units**
- **variable**

Figure 4. The variable_set compound brick. The vertical, branching line, illustrates the brick structure encapsulated within variable_set. This compound brick contains eight pure bricks. The content of the eight bricks is not shown. The bold text indicates mandatory bricks.

The units brick is of particular importance (Figure 5) within the variable_set compound brick. The units brick contains all the information related to the units of the variable. The pt_code and pt_link attributes are contained within the units brick to allow usage outside the variable_set compound brick. In the profile implementation, the units and variable bricks are both contained within variable_set. This arrangement does not require pt_code and pt_link in the units brick, as the encapsulation within variable_set implies a relationship to the variable brick. However, in anticipation of the units brick being used outside the variable_set compound brick, we include the pt_code and pt_link attributes.

The units brick also has attributes received_units and stored_units. received _units is used to contain the units of the variable as originally received by the agency constructing the XML document. The stored_units attribute contains those units used within the particular XML document.

The units brick also contains sub-elements conversion, reference and variable_name. The variable_name element is similar to pt_code and pt_link in that it is not necessary within the profile implementation. Again, it is included for anticipated use of the units brick outside the variable_set compound brick. The conversion element allows a description of the conversion from the received to the stored units. Finally, the reference element allows for a publication reference for those detailed conversions.

units {**pt_code**, pt_link, received_units, **stored_units**}
- conversion
- reference
- variable_name

Figure 5. The units brick. This brick contains 4 attributes and three elements. The bold text indicates mandatory content.

The location_set compound brick (see Figure 6) contains the position and time specification for the dataset and quality information for the spatial-temporal point. In particular, the main bricks within the location_set compound brick are latitude, longitude, depth_pressure and ldate (a location date).

The content bricks within location_set represent a slight departure from the general philosophy of the Keeley Bricks. The content bricks have very specific tag names. This specific naming was intended to highlight the importance of these parameters.

```
location_set
  ├ comment
  ├ depth_pressure
  ├ latitude
  ├ ldate
  ├ longitude
  └ quality
```

Figure 6. The location_set compound brick. The lack of bold text indicates there is no mandatory content.

Unfortunately, the specific naming of these bricks caused numerous problems and represents one of those areas where improvement could be made to the profile structure and quite possibly the bricks in general. The problems resulted from the added complexity at the record level, when these specifically named parameters are data (see section on Interesting Results).

Conversion Methods

Various XML documents were produced using the developed XML profile structure. All three participating labs provided an assortment of profiling float, XBT, water sample, and CTD data in the XML structure. This required the development of conversion software at two levels: the conversion from in-house to XML structure and the conversion from XML structure to in-house.

The method used for the first conversion is specific to the in-house development environment, and therefore of limited use to the international community. The second conversion method, from XML to in-house format, is more interesting for the community. This conversion was conducted in Fortran using msxml (IOS and MEDS), and in Java using data binding techniques (BIO). As well, eXtensible StyleSheet Language Transformations (XSLT) were used as a preliminary investigative tool with limited success (BIO).

At BIO, the transformation of ODF profile data to XML was conducted using the Ocean Sciences Division (OSD) Matlab-based set of tools called the Oceans Data System (ODS) Toolbox. This suite of tools has been developed by OSD over the past decade and constitutes the primary analysis tool within the Division. The Toolbox is capable of reading ODF files and creating a memory resident ODF cell structure within the Matlab environment.

The BIO transformation from XML to ODF was conducted using two different methods. The first method used XSLT, an XML-based language used to manipulate the structure of an XML document. This method was developed as an exercise in understanding the usefulness and versatility of XSLT and was not considered the most practical method of dealing with the brick objects.

The second BIO method for transforming the XML to ODF utilized Java technology – in particular, data binding using Sun's Java Architecture for XML Binding (JAXB). Data Binding is a technique for linking and automatically creating Java classes based on a constraint file, such as a schema file. The result of the binding is the removal of complicated data access methods via calls to nodes in the XML structure. Instead, the binding creates more descriptive classes based on the brick names. So, classes such as DataCollection were created. As well, binding results in method calls such as object.getProvenance().getDescription() (where object might be DataCollection) to obtain the description data within the provenance brick.

The JAXB-developed classes are really object representations of the bricks. A similar construction of objects may use an object-oriented database. In such a database, the structure, including encapsulation and inheritance in the XML document, can be utilized within the database.

After the classes and methods were created in the binding process, a wrapper program, called ProcessDataCollection, was created. ProcessDataCollection controlled the transfer (unmarshalling) of the XML document into Java objects. ProcessDataCollection also controlled the access of the Java objects to create a new file in the in-house ODF format.

Some Interesting Results

We conclude with several interesting results discovered during the development process.

- During the development, some tags for data or metadata were given specific names (e.g., latitude, longitude, depth_pressure, ldate). Such specific naming is not necessary within the XML environment. These data could be included using the data_point brick and the pt_code attribute.

  This specific naming complicated the BIO XSLT investigation. The latitude, longitude and ldate bricks were sufficiently specific to remove the requirement for the pt_code attribute within these bricks. Although seemingly innocuous, the removal of the pt_code attribute caused problems with the XSLT code for transforming the XML to ODF. Specifically, the pt_code (and pt_link) attribute permitted the sorting of the elements. This was very important for the output ODF structure, as ODF requires the header variable definitions in the ODF file to be in the same order as the variable columns. Using a sort on pt_code for both variable and data_point information resulted in a consistent ordering of the variable information in the ODF file. The removal of the pt_code attribute in the latitude, longitude and depth_pressure bricks meant special consideration was required for these data. This introduced complications in the XSLT code.
- Another realization during the process dealt with the numerous inadequacies with the in-house formats. These formats were developed to maintain local processing and archival systems. The data model on which many were developed was not oriented towards ocean data in a general sense, but more toward specific ocean data types. For example, the ODF format does not have the ability to store information on more than one instrument. This implies that the data in a single ODF file must originate from a single instrument (not always the case). The XML structure, allows unique data types to identify unique instruments. This more generalized structure does not map well to the ODF format.
- The development presented here only provides a data structure for encapsulating ocean data. When exchanging data within a common structure, the problems associated with parameter codes are still present. For example, one institute may refer to particular organic carbon as CPX1 in milligrams/litre while another institute refers to the same data as Carbon:Particulate:Organic in micrograms/litre. Although the identification of the parameter dictionary used within the XML document may be included using the data_dictionary brick, the conversion from one dictionary to another remains a requirement. Developments are ongoing to convert code sets from one dictionary to another using XML [7].
- Null values are not XML friendly. In the old paradigm of data formats, null values were defined to fill the space of the "missing" data value. These null values were typically outside the space of realizable data values. For example, latitude stored in degrees might have an associated null value of –99. The –99 is not in the realizable range of latitude values expressed in degrees. In XML, one may define restrictions to set the allowable limits of data values. If such a schema restriction were placed on latitude for the range –90 to 90, the null value of –99 would not pass validation.

  Also, implementing this restriction would not allow an empty tag to be present in the XML document. In XML, all tag content is used in validation, including empty content.

  This functionality has consequences for the mandatory set of tags defined for the bricks. If, in the schema, a tag is declared as **mandatory** with **restrictions**, then the tag **must** be present with valid content in the XML document. The input data streams therefore must have that content available.
- The XML documents generated within this project provided an opportunity to investigate file size issues. The ODF files, which are ASCII, were compared to the XML documents generated from the ODF content. When comparing six ODF files (bottle, CTD, XBT, float, moored current meter, and underway TS data) to the XML equivalents, the XML representation occupied 600% of the disk space as compared to the ODF files. However, when compressed using common compression software, the XML representation was 40% larger as compared to the compressed ODF files. This indicates that compression of XML files may alleviate XML file size issues.

The Future

This project has shown that XML is a viable exchange mechanism for ocean profile data. The project has also shown that development based on the Keeley Bricks shows promise as an object-oriented approach for developing XML structures for ocean data types. Building on the work of SGXML, the development of code mapping capabilities [7] has also shown that seamless exchange of ocean data can be a reality.

The future of international ocean data exchange will ultimately rest with institute personnel responsible for developing the exchange systems. At this time, the technical problems of exchange are minimal. Potentially the largest problem is developing the interest and initiative to see a vision to completion. The vision may start small, with the linking of a few centres all providing data in a common structure and parameter dictionary understood by the client. It could easily grow to a single portal interface that provides the client with access to data from the international data system. The portal could interrogate local databases and provide datasets to clients in a common structure and dictionary. This has been described [8] as the Mikhailov Model of a distributed ocean data system. It can be a reality.

References

Report of the ICES-IOC Study Group on the Development of Marine Data Exchange Systems Using XML, Helsinki, Finland, 15–16 April 2002, ICES CM 2002/C:12.

Isenor, Anthony W., J. Robert Keeley and Joe Linguanti. Developing an eXtensible Marup Language (XML) Application for DFO Marine Data Exchange via the Web, Defence R&D Canada – Atlantic ECR 2003-025, March 2003.

Canadian XML Brick Website, http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/ICES/ICES_e.htm

For general knowledge see http://www.oasis-open.org/cover/xml.html#applications

For Specifications see http://www.w3.org/

Means, W. Scott and Elliotte Rusty Harold. XML in a Nutshell, January 2001, 0-596-00058-8.

Isenor, Anthony W. 2003. XML Based Manipulation of Codes Exchanged Between Data Systems, Technical Memorandum in preparation, Defence R&D Canada – Atlantic.

Report of the Working Group on Marine Data Management, Helsinki, Finland, 17–19 April 2002, ICES CM 2002/C:11.

Example XML Document


```
<?xml version=“1.0” encoding=“ISO-8859-1”?>
<data_collection xmlns:xsi=“http://www.w3.org/2001/XMLSchema-instance”
xsi:noNamespaceSchemaLocation=“c:\Anthony\Projects\xml\odf_conversion\bricks_v2.xsd”>
  <data_dictionary>
    <dictionary_name>GF3+</dictionary_name>
  </data_dictionary>
  <data_set>
    <data_set_id level=“cruise”>73025</data_set_id>
    <provenance>
      <agency>Bedford Institute</agency>
      <date_created>2003-02-18Z</date_created>
      <institute_code>1810</institute_code>
      <originator_identifier>73025</originator_identifier>
    </provenance>
    <data_set>
      <data_set_id level=“station”>118</data_set_id>
      <data_set>
        <comment>17-JAN-1984 00:00 BOTTLE DATA 73027</comment>
        <comment>17-JAN-1984 00:00 </comment>
        <data_set_id level=“profile”>Q1</data_set_id>
        <provenance>
          <agency>Bedford Institute</agency>
          <data_grouping>BOTL</data_grouping>
          <date_created>2003-02-18Z</date_created>
          <description>Q2</description>
          <originator_identifier>Q1</originator_identifier>
        </provenance>
        <variable_set>
          <instrument type=“bottle”/>
          <sampling pt_code=“PRES”>
            <interval>   0.00000</interval>
          </sampling>
          <units pt_code=“PRES” stored_units=“decibars”/>
          <variable kind=“I” pt_code=“PRES” typing=“R”>
            <decimal_places>4.000000</decimal_places>
            <maximum_value>411</maximum_value>
            <minimum_value>33</minimum_value>
            <null_value>-.99000000D+02</null_value>
            <variable_name>PRES_1</variable_name>
          </variable>
        </variable_set>
        <variable_set>
          <instrument type=“bottle”/>
          <units pt_code=“PSAL” stored_units=""/>
          <variable kind="D" pt_code="PSAL" typing="R">
            <decimal_places>4.000000</decimal_places>
            <maximum_value>34.944</maximum_value>
            <minimum_value>32.725</minimum_value>
            <null_value>-.99000000D+02</null_value>
            <variable_name>PSAL_1</variable_name>
          </variable>
        </variable_set>
        <variable_set>
          <instrument type="bottle"/>
          <units pt_code="DOXY" stored_units="ml/l"/>
          <variable kind="D" pt_code="DOXY" typing="R">
            <decimal_places>4.000000</decimal_places>
            <maximum_value>8.53</maximum_value>
            <minimum_value>7.36</minimum_value>
            <null_value>-.99000000D+02</null_value>
            <variable_name>DOXY_1</variable_name>
```

```
    </variable>
  </variable_set>
  <variable_set>
    <instrument type="bottle"/>
    <units pt_code="SLCA" stored_units="mmol/m**3"/>
    <variable kind="D" pt_code="SLCA" typing="R">
      <decimal_places>4.000000</decimal_places>
      <maximum_value>0.8</maximum_value>
      <minimum_value>0.6</minimum_value>
      <null_value>-.99000000D+02</null_value>
      <variable_name>SLCA_1</variable_name>
    </variable>
  </variable_set>
  <variable_set>
    <instrument type="bottle"/>
    <units pt_code="PHOS" stored_units="mmol/m**3"/>
    <variable kind="D" pt_code="PHOS" typing="R">
      <decimal_places>4.000000</decimal_places>
      <maximum_value>7</maximum_value>
      <minimum_value>1</minimum_value>
      <null_value>-.99000000D+02</null_value>
      <variable_name>PHOS_1</variable_name>
    </variable>
  </variable_set>
  <variable_set>
    <instrument type="bottle"/>
    <units pt_code="TEMP" stored_units="degrees C"/>
    <variable kind="D" pt_code="TEMP" typing="R">
      <decimal_places>4.000000</decimal_places>
      <maximum_value>0.9</maximum_value>
      <minimum_value>-1.75</minimum_value>
      <null_value>-.99000000D+02</null_value>
      <variable_name>TEMP_1</variable_name>
    </variable>
  </variable_set>
  <location_set>
    <ldate property="creation">
      <pdate>2003-02-18Z</pdate>
      <ptime>19:30:53.99Z</ptime>
    </ldate>
  </location_set>
  <location_set>
    <ldate property="original">
      <pdate>1990-11-20Z</pdate>
      <ptime>00:00:00Z</ptime>
    </ldate>
  </location_set>
  <location_set>
    <latitude property="start">  67.00700</latitude>
    <ldate property="start">
      <pdate>1973-09-04Z</pdate>
      <ptime>00:00:00Z</ptime>
    </ldate>
    <longitude property="start"> -26.83800</longitude>
  </location_set>
  <location_set>
    <latitude property="end">  67.00700</latitude>
    <longitude property="end"> -26.83800</longitude>
  </location_set>
  <history_set>
    <comment> </comment>
    <history>
      <application_date>1992-11-05Z</application_date>
```

```
        </history>
      </history_set>
      <history_set>
        <comment>GF3 Name Checking and Code Formatting</comment>
        <comment>Name check performed</comment>
        <history>
          <application_date>2003-02-18Z</application_date>
        </history>
      </history_set>
      <data_set>
        <data_point pt_code="min_depth">  32.66939</data_point>
        <data_point pt_code="max_depth"> 406.51097</data_point>
        <data_point pt_code="sounding"> 424.00000</data_point>
        <data_point pt_code="depth_off_bottom"> 0.00000</data_point>
        <data_set_id level="related"/>
        <variable_set>
          <instrument type="sounder"/>
          <units pt_code="sounding" stored_units="metres"/>
          <variable kind="I" pt_code="sounding" typing="R">
            <decimal_places>5.000000</decimal_places>
            <variable_name>sounding</variable_name>
          </variable>
        </variable_set>
        <variable_set>
          <instrument type="bottle"/>
          <units pt_code="max_depth" stored_units="metres"/>
          <variable kind="I" pt_code="max_depth" typing="R">
            <decimal_places>5.000000</decimal_places>
            <variable_name>maximum depth</variable_name>
          </variable>
        </variable_set>
        <variable_set>
          <instrument type="bottle"/>
          <units pt_code="min_depth" stored_units="metres"/>
          <variable kind="I" pt_code="min_depth" typing="R">
            <decimal_places>5.000000</decimal_places>
            <variable_name>minimum depth</variable_name>
          </variable>
        </variable_set>
        <variable_set>
          <instrument type="sounder"/>
          <units pt_code="depth_off_bottom" stored_units="metres"/>
          <variable kind="I" pt_code="depth_off_bottom" typing="R">
            <decimal_places>5.000000</decimal_places>
            <variable_name>depth off bottom</variable_name>
          </variable>
        </variable_set>
      </data_set>
      <data_set>
        <data_point pt_code="PSAL">32.7250</data_point>
        <data_point pt_code="DOXY">8.5300</data_point>
        <data_point pt_code="SLCA">0.6000</data_point>
        <data_point pt_code="PHOS">5.0000</data_point>
        <data_point pt_code="TEMP">-0.5000</data_point>
        <data_set_id level="record"/>
        <location_set>
          <depth_pressure pt_code="PRES">33.0000</depth_pressure>
        </location_set>
      </data_set>
    </data_set>
   </data_set>
  </data_set>
</data_collection>
```

ANNEX 7: 2003/2004 TERMS OF REFERENCE FOR SGXML

An **ICES-IOC Study Group on the Development of Marine Data Exchange Systems using XML [SGXML]** (Co-Chairs: R. Gelfeld, U.S.A. and A. Isenor, Canada) will meet in Oostende, Belgium from 6-7 May 2004 to:

a) create, evaluate and discuss inter-sessional work on parameter dictionaries including the dictionary mapping analysis, and the reconciliation of the XML structure for dictionary exchange;
b) evaluate inter-sessional work on the point data structure including the investigation into accepted standards for incorporation in the Keeley Bricks and the efforts to apply the Keeley Bricks to 3 dimensional biological data;
c) evaluate inter-sessional work on metadata including reporting on the comparison of metadata standards (ISO, MEDI, EDMED, etc) and the initial development of an optimal metadata tag list;

SGXML will report by 30 May 2004 for the attention of the Oceanography Committee.

Supporting Information

Priority:	The future of marine data management, processing and exchange in an interoperable environment lies in the use of virtual data systems and exploiting web technologies. If ICES does not participate in these developments its ability to receive, process and disseminate data in the form required by the user community will be negatively impacted.
Scientific Justification:	a) The XML web distribution of the parameter dictionaries should be completed and the usefulness of the exercise for cross mapping of parameter dictionaries needs to be assessed. The applicability of the XML structure for other dictionaries should also be determined b) International standards need to be incorporated into the generalized point data structure and evaluated from the perspective of the international data centres. The applicability of the abstract Keeley bricks to other data types needs to be evaluated. c) The metadata problem is common to many organisations and considerable effort has been made by these other organisations. The usefulness of these efforts needs to be evaluated within the context of ocean data transfer. As well, the generalization of the metadata model needs to be evaluated. The generalised model needs to be considered within the context of existing models
Relation to Strategic Plan:	The Group is set up to provide members of the ICES scientific community, efficiently and effectively, with the support they need to meet the scientific goals. The ICES Vision goes far beyond the capacities of any single organization. Networking in partnership with Member Countries, other IGOs, and scientific NGOs will enable ICES to foster valuable cooperation, coordination and collaboration. By so doing, ICES will not duplicate activities already carried out by others. Rather, this will provide a new, value-added dimension. Enhanced interdisciplinary knowledge and networking will benefit the entire science community.
Resource Requirements:	No specific resource requirements beyond the need for members to prepare for and participate in the meeting.
Participants:	Participation in the XML Study Group is open to any individual or group, internal or external to ICES.
Secretariat Facilities:	None.
Financial:	None specific.
Linkages To Advisory Committees:	This is important to work on data integration, which is of direct interest to ACE.
Linkages To other Committees or Groups:	The ICES working group on marine data management (WGMDM).
Linkages to other Organisations	The WMO/IOC Joint Commission on Oceanography and Marine Meteorology (JCOMM) has also expressed an interest in XML. A Marine Consortium has been formed to address XML internationally (IODE, 2000). The Consortium's goal is to develop a free and open specification for a Marine XML that will be used in all exchanges of ocean data. ICES has been asked to become a Consortium member. Several ICES countries have joined or are about to join the Consortium (Belgium (Flanders), Netherlands, UK, Sweden). In addition IOC/IODE and

	EuroGOOS have joined.
Cost Share	ICES 100%