ICES Oceanography Committee
ICES CM 2004/C:12

Report of the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML (SGXML)

6–7 May 2004
Oostende, Belgium

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

H. C. Andersens Boulevard 44–46 · DK-1553 Copenhagen V · Denmark
Telephone + 45 33 38 67 00 · Telefax +45 33 93 42 15
www.ices.dk · info@ices.dk

Contents

1 Opening of the meeting

The meeting was opened by R. Gelfeld (Co-Chair) and was hosted by the Flanders Marine Institute (VLIZ), Oostende, Belgium. Edward Vanden Berghe welcomed the participants on behalf of VLIZ and also outlined the local arrangements.

Members of the Study Group present were: P. Alenius (Finland), D. Collins (USA), R. Cramer (UK), M. Fichaut (France), R. Gelfeld (USA, Co-Chair), P. Haaring (Netherlands), A. Isenor (Canada, Co-Chair), R. Lowry (UK), N. Mikhailov (Russia, JCOMM ETDMP Chair), G. Reed (Australia), L. Rickards (UK and IODE Chair), H. Sagen (Norway), R. Schwabe (Germany), J. Szaron (Sweden), E. Vanden Berghe (Belgium, GE-BICH Chair) and M. Wichorowski (Poland).

The Intergovernmental Oceanographic Commission (IOC) was represented by P. Pissierssens and V. Vladymyrov. The ICES Secretariat was not represented at this meeting.

Meeting participants not official members of the Group included C. Bamelis (Belgium), S. Belov (Russia), L. Bird (UK), K. Furukama (Japan), F. Hernandez (Belgium), M. Hughes (UK), A. Meerhaeghe (Belgium), G. Moiseenko (Russia), and T. Sakakibara (Japan). Absent members included: T. Carval (France), J. Gagnon (Canada), C. Haenen (Netherlands), F. Nast (Germany), R. Olsonen (Finland), R. Starek (US) and B. Pelchat (Canada).

In total, there were 25 participants representing 14 countries, plus 2 participants from the IOC. 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 and Annex 3) for the Study Group meeting was adopted as a resolution at the 90th Statutory Meeting (C.Res. 2002/2CML). This is the third and final meeting of the Group.

3 Review of 2003 Action Items

A. Isenor (Canada) introduced a discussion focusing on the Action Items identified in the 2003 SGXML Report [1] (for a list of acronyms, see Annex 4; for 2003 Action Items see Annex 5). Action Items 2, 3, 4, 10, 14, 15, and 16 have been completed. These items included construction of a web interface to the BODC parameter dictionary, consolidating the DTD and schema structures for the code XML structure, requesting EU MarineXML review of the Keeley Bricks, evaluation of metadata standards (ISO-19115), and other Group business-related activities.

Action Item 1 (see [1] for complete list of Action Items) involved the cross mapping of 11 parameter dictionaries. This is an enormous activity, involving considerable human resources of specialists in the various dictionaries. Although the 11 mappings are not complete, the UK have spent considerable resources on this effort, and have made excellent progress.

Action Item 5 examined the Geography Markup Language (GML) in comparison to the Keeley Brick structures. This Action Item was only partially completed, as the OGC Web Services (OWS) component of the Action was not completed. Action Item 7 also dealt with the Keeley Brick XML structure and the placement of biological net tow data into this structure. Although not complete, the preliminary placement of biological data into the Brick structure shows potential for the structure to store biological data.

Action Items 8 and 9 dealt with the terminology and a reference model for metadata from the perspective of a virtual oceanographic data centre. Some progress was made on these activities, but this task is also highly complex, requiring much more effort.

Action Item 11 was an examination and mapping of various data structures for cruise level metadata. These structures included CSR, MEDI, EDMED, USNODC DDF and the ISO-19115.

Action Item 12 was reported on during the meeting; no report had been submitted to either of the chairs before the meeting, due to problems in locating the relevant document (ISO-19110).

There was no activity on Action Items 6, 13, and 17.

A. Isenor noted that the Group had completed about 60% of the Action Items as defined in the 2003 meeting report. The completion of those tasks oriented towards the parameter dictionary mappings does mean that considerable progress has been made on this topic. The examination of the structures for metadata was also much needed. Finally, the completion of the Keeley brick application to GML and the testing of the Keeley Brick structure with biological data represent a considerable effort.

4 Presentations and discussions

4.1 SGXML presentation at IODE XVII

R. Gelfeld (United States) opened a discussion on SGXML by first providing the material presented at the IODE XVII in 2003 (NOTE: Links will function properly on the meeting CD version of the final report). An open discussion ensued after the presentation, dealing with some of the accomplishments of the SGXML. It was widely recognized that as a Group, our understanding of XML has increased enormously over the past three years. The restructuring and availability of the BODC parameter dictionary was also noted. This has increased the ocean data community's awareness of the dictionary and has also resulted in data managers consulting the dictionary to obtain and use codes.

The Group discussed the need for tools to support developed XML structures. However, this implies the XML structure is stable. There are also funding issues related to such developments. It was noted that tools may also suggest web services. For example, XML based services for converting Marsden squares to bounding box descriptions. Marsden square descriptions are common in previous metadata descriptions, while bounding box elements are common in more recent metadata descriptions.

4.2 Comparison mapping of metadata standards

M. Fichaut (France) then presented work on the EDMED ISO-19115 XML Mappings. The goal of this effort was to provide metadata information over the web based on both EDMED and CSR descriptions. The technique involved the mapping of both descriptions to an XML ISO-19115 structure, and then creating the web-available information using XSLT.

The EDMED mapping results show most of the fields are found in ISO-19115. A small number of fields are not available in ISO-19115. All mappings are documented in Excel spreadsheet format. The XML schema and an example XML document is also available. At present, about 35 countries use the EDMED structure for metadata. Some are now utilizing this work for their own efforts towards metadata discovery using the web.

The CSR mapping shows about 50% of the CSR fields are not found in ISO-19115. Examples of encountered problems are the Marsden Squares and cruise departure port. Investigations involving Geography Markup Language (GML) are currently underway to evaluate its potential with reference to the CSR.

The Group recognized the need for a central location where the schemas and documentation on such efforts could be made available. The discussion noted the ISO-19135 and 19110 standards. The ISO-19135 is the procedure for creating a registry, while 19110 deals with the content of the registry.

4.3 Distributed data model

N. Mikhailov (Russia) then presented material for a distributed marine data model. The data sources of the model were described as the real data holdings of institutes and labs. The information resource was described as an abstract representation of the data source. These abstractions are the interaction points for the outside contacts requesting data. The metadata were also described as multi-functional, with metadata descriptions for interfacing, for the data, and for other functional aspects. The concept of the distributed marine data model was tested last year on servers in Obninsk, Arkhangelsk, Murmansk, St. Petersburg and Moscow. Results were promising, and the efforts continue at the Russian institutes.

The importance of parameter dictionaries was noted and for metadata attribute descriptions. Metadata attribute systemization was defined in blocks similar to the Bricks. The serviced metadata is described in terms of the relationships between the blocks. The request/response mechanism was described as a request layer and exchange layer. Supporting information was provided in an attached paper.

S. Belov (Russia) then presented information on the construction of the distributed information system. The distributed system was described as an architecture consisting of levels that include the data source, technology and interface levels, and the functional services level. The data source level includes both the local data system and a wrapper between the local system and the distributed system. The technology and interface level includes things like protocols, message exchange and security. This is also the level responsible for the service metadata. The functional services level was noted to have many functions including the coordination of the data coming from different sources, the cooperative use of the data, the request decoding and format generation, end user interfaces and personalization. The present software development for this service uses JBOSS Open source (see http://www.jboss.org/index).

4.4 Report on code mapping – Action Item 1

R. Lowry (United Kingdom) gave a report on the code mapping effort. Action Item 1 from 2003 is now recognized as extremely ambitious. However, considerable progress was made, in large part to funding provided by NERC, under a project named Enabling Parameter Discovery (EnParDis). This effort will fund 1.5 people until October 2004 and another 0.5 from October 2004 to March 2005.

As part of this effort, the BODC dictionary is now available in searchable form on-line. The dictionary may be found at: http://www.bodc.ac.uk:8080/sgxml/ParaSearch.htm.

It is useful to first define exactly what a mapping means. Here, we consider a mapping to be a one-to-one relationship created between codes from two different sources. As an example, one could envisage a spreadsheet of one code contained in a cell being beside the mapped code from the second dictionary.

The project started with the US JGOFS codes. This mapping was completed through a manual mapping. The process showed that manual mappings are laborious and time consuming. It also showed that divorcing units from the codes is a good thing because it drops the number of codes dramatically.

Two IFREMER dictionaries were also mapped. The lessons from this mapping included the realization that codes associated with general method descriptions are easier to map, while codes that are associated with detailed methods cause a problem.

Also noted were the problems caused by codes for sea surface temperature (SST) and sea surface salinity (SSS). This type of code incorporates the z co-ordinate into the parameter description, which is incompatible with the ISO model for geo-referenced data.

In the DONAR (see http://www.donarweb.nl) effort, it must first be realized that DONAR is really a collection of items of information concerning a measurement. It is not a dictionary. Also, the size of DONAR presented a problem, containing 4932 biological parameters, 462 chemical parameters, etc. This was recognized as too large for manual mapping. A semantic model is now being considered for the mapping. The semantic model will deal with atomic items of information that are populated from controlled vocabularies. However, the one-model fits all approach is not practical at the moment. However, a series of semantic models is plausible. This effort is on going.

In a semantic mapping, there is also a role for a dictionary as a registry of valid combinations. Here you are mapping the semantic elements, using the controlled lists. This reduces the workload significantly.

In GCMD, the problem was granularity incompatibility between the BODC dictionary and the GCMD. This effort is on going.

PANGAEA was also a problem. This is because they only have a numeric code and a description. This effort was abandoned.

With the ITIS mapping, it was noted that not all BODC taxon (or taxa) contained in the BODC dictionary were present in ITIS. About 200 additional codes have been sent to ITIS for review and potential incorporation. A browser is currently being developed to exploit the taxonomic grouping of BODC codes. BODC hope to have this available in September 2004.

The ensuing discussion noted that ITIS response to queries and requests indicate a willingness to help, but response is often slow.

SMHI codes have also been mapped to the BODC dictionary. However, some SMHI categories (e.g., ice, humus) are not present in the BODC dictionary.

Canadian efforts mapped parameter codes from three labs: BIO, IOS and MEDS. This mapping extended the exercise to include units and conversions [2]. The Canadian codes being used at the three labs are also available on-line at: http://www.meds-sdmm.dfo-mpo.gc.ca/meds/About_MEDS/standards/.

The issue of unit conversions also has the problem of conversions involving density. Here, it was suggested that the user is important in the conversion exercise, as users tolerate different levels of conversion accuracy. For some users, density assumptions may be used.

It was also noted that a single mapping may not be applicable for all users. For example, some users will want a mapping that actually combines data into categories. For example, a user interested in when blooms occur only requires information on 'chlorophyll' and will therefore want to combine many more specific chlorophyll parameters.

Finally, significant figures will be an issue with conversions. Any conversions should maintain the proper number of significant digits.

4.5 Data Schema Standardization for Coastal Environmental Data Set

K. Furukawa and T. Sakakibara (Japan) presented work on the Data Schema Standardization for Coastal Environment Data Set. This work concentrated on modifying the schema used in the Tokyo Bay Environmental Information Center’s (TBEIC) database system toward the ISO19115 and GML standards. The marine data are currently corrected and distributed by the JODC, while GIS platforms are supported by JGIS. The requirements here are based on Bay wide physical, chemical and biological datasets. The need is to have an integrated database of measurements. The attempt was made to carefully standardize the data based on ISO-19115 and GML.

The developed system consists of a set of spreadsheet forms that are completed by data collectors in the field. Four spreadsheet forms, built from macro-program based templates, are prepared to generate the XML output based on the form content. At present about 10 local government institutions and universities are checking the system. The system and documentation has been provided. The system and documentation will be finalised in July–August 2004.

A committee was established consisting of users, collectors and managers of the data. The funded project is owned by the Ministry of Land, Infrastructure and Transport. The actions of the committee involved defining the data structure (UML), definitions for the data structure, XML schema, sample data, and data exchange tools for XML.

The dictionary part is described using GML tags for things like gml:Point and gml:Name. The observation is then linked to the dictionary via a location name. In the metadata, the user publication information is also stored to reference the use of the dataset.

One key point of the presentation was that SGXML needs to think about the users. The users are looking for tools to support the XML structures and SGXML needs to take this into account.

4.6 Comparing and reconciling the parameter dictionary XML structures

R. Cramer (United Kingdom) then presented the work to consolidate the structures for code mapping. A comparison was conducted of XML schemas between BODC and FIMR. The DTD and XSD structures were also compared. There were differences noted, but primarily in tag naming. The Co-Chairs suggested the editing of the finalised structure be conducted after the meeting, with Group input via email correspondence. The finalised parameter mapping schema is now available.

4.7 Application of XML technology in the NERC DataGrid Project

R. Lowry (United Kingdom) presented an update on the activities of the NERC DataGrid (NDG). This is a UK project to build a grid for data discovery and delivery. The Grid is intended to support functions such as discovery, authentication, authorization, extraction and sub sampling. The NDG is implemented in a fashion similar to the Amazon book search system. The search and discovery is followed by data snippets so the user can decide if the particular result is useful. Results are also presented with similar findings, again modelled after the Amazon system which lists similar books that people selected when selecting the primary search results.

The NDG and the Mikhailov model for a distributed data model have many similar concepts. For example, the NERC wrappers are similar in concept to the Mikhailov integrators.

Also, NDG addresses the multi-definition of metadata by introducing metadata types such as A, B, C, D, etc. Metadata types A and B make the separation between use and discovery. Much of the metadata model can be expressed in 19115, but extensions are needed for the user community. The Mikhailov model describes associated, thematic, service, etc., metadata types.

The NDG data model considers things as variables. Variables may be combined into arrays, and many arrays can fit in a single array. The dataset is the fundamental unit of the data model. The data model incorporates ISO-19103 and 19109. The Grid also uses ISO-19111 and ISO-19108.

Technologies in use include XSLT for preparing metadata records and converting XML structures. This is useful for conversions between metadata types (e.g., to convert metadata B from D). The Open Archive Initiative (OAI) is also being utilized (see http://www.openarchives.org/OAI/openarchivesprotocol.html) for the discovery (type D) metadata harvesting. Use of OAI means that once you have described your XML data structure in a schema, it can be made searchable via OAI. The OAI is similar in concept to the Mikhailov integrator and navigator concepts. Use of OAI will also expose the NGD to other discovery systems.

4.8 Metadata discussion

D. Collins (United States) introduced a discussion on metadata Action Items from 2003. The Group recognizes that metadata occurs in various levels (e.g., the NDG specification of metadata). The terms of the Russia metadata analysis have been defined, and comments are welcomed. For metadata structures, the ISO-19115 must be considered, but we can expect community extensions are required to suit our particular needs. It was suggested that the metadata tag development be a recommendation for future efforts. The ISO-19115 could be the starting point and the required extensions developed from that point. ISO-19110 deals with Methodology for feature cataloguing, and is also of direct relevance for any data management activity where reference is made to geographical features and their attributes.

It was noted that last year’s Action Item dealing with an optimal tag list for metadata is very difficult. One needs to examine all the standards and then merge and come up with the optimal set of tags. This is something for a group, not an individual. This should also be considered for possible recommendation for follow-on activities.

4.9 Application of bricks to current meter and water level data

On behalf of G. Slesser (Scotland), A. Isenor (Canada) presented an XML activity from the ICES Working Group on Marine Data Management (WGMDM). An item on the WGMDM action list for 2003 identified an evaluation of the Keeley Brick 1-d profile structure [3] by placing water level and current meter time-series data into the structure. Fisheries Research Services (FRS) conducted this evaluation.

The evaluation was promising in that the majority of the water level and current meter data was placed in the existing 1-D profile structure. However, some weaknesses and consequently recommendations were made regarding the Bricks and associated documentation. In particular:

- Several inconsistencies were noted in Annex 2 of the documentation [3]. In particular the attributes *institute_code*, *isa* and *order_number* are no longer in the Bricks yet appear in Annex 2. The documentation should be corrected.
- The property attribute in the location_set subelements should have "deployment" and "recovery" in the allowed content. This would address the deployment and recover times of the mooring. Note that recovery and deployment refer to the time in and out of the water. This is different from the start and end time of the mooring data record.
- A mooring detail Brick should be considered. This could be made up of the deployment and recovery times, and information such as instrument depth, sounding, etc.
- The instrument Brick should be capable of storing mooring specific information such as reference number, reference check and magnetic variation. The reference check is a unique identifier used by FRS, Scotland to check the validity of the reference number recorded on magnetic tape by older Aanderaa current meter models. Given the reliability of the Data Storage Units (DSU) now used by Aanderaa current meters the reference check has become redundant.
- It may be important to distinguish the type of mooring. At the moment, the mooring is identified under provenance/platform_type as "mooring". However, there are different types of moorings, such as U-shaped or single-string moorings.

Since this work was also presented to the ICES WGMDM, the MDM report should also be examined for comments regarding the Bricks and XML structure.

4.10 WADI and its XML

P. Haaring (The Netherlands) presented a description of WADI the use of XML in the Water Data Infrastructure (WADI) system. Note that WADI is the successor of DONAR (see http://www.donarweb.nl).

Several governmental institutions are currently using WADI. The WADI approach is to develop one data model to fit all data. The database contains measured and derived data, for salt and freshwater. It is not initially addressing operational data. As well, it does not contain model data. In the WADI system, XML is a core component. WADI XML uses the RDF standard.

4.11 The bricks applied in GML

A. Isenor then presented the results of a Canadian investigation into applying Keeley Bricks to Geography Markup Language (GML). It was noted that the investigation actually considered the 1-D (i.e., data with one independent variable such as z for profile data, or t for a time-series) profile structure resulting from last year's effort. The 1-D structure was represented in GML, utilizing many of the GML objects. In a few cases, non-GML objects were created to address specific functionality.

The results showed several important issues. For example, nomenclature is a problem because the geography community have a specialized perspective on the data, which is different from the perspective of the ocean data community. Another important difference is that GML stores data in comma separated strings, while the initial Brick implementation packaged each data item in an XML element. The lack of GML schema validation for the coverage.xsd schema also reduces ones faith in the language. A report on this work is currently in production.

4.12 IOC XML register

G. Reed (Australia) presented a proposal for the development of an IOC Registry. A registry is the information system on which registers are maintained. A registry may take the physical form of a database. The registry is a process and system for cataloguing other systems that would be recognized in some official capacity. The term "register" refers to a partition of the registry.

The IOC Registry could be set-up in the Oostende IOC Project Office. It is proposed that the registry be ISO-19135 compliant. An example first entry could be the S-57 object catalogue. S-57 is used for encoding electronic navigation charts, and is maintained by International Hydrographic Office (IHO). There are currently two relevant standards for registry's, ISO-19135 and ISO-19126. Other potential loads to the registry could be the BODC parameter dictionary, and the Ocean Biogeographic Information System (OBIS) schema (see http://www.iobis.org/).

OBIS is an example of XML being used in a marine science data management capacity. OBIS uses the DiGIR system (see http://digir.sourceforge.net) to communicate with its distributed data contributors.

The resulting discussion focused on the timetable of establishing such a registry. It was recognized that IOC would need to fund a programmer to set up the interfaces for such a registry. A management body that dealt with submission would also need to be established.

4.13 Exploring biological net tow data in Keeley Brick 1-D profile structure

E. Vanden Berghe (Belgium) and A. Isenor (Canada) presented preliminary results of placing biological net tow data into the previously developed profile structure [3].

A biological database was used to determine the structure inter-relationships of the data. The database described topics such as species, gender, gear, trips, visits, events, and people. One realization was that the in-house database contains considerable information that is not required in an international data exchange structure (e.g., people tasking during the field work).

Two Bricks were described in detail: variable_set and data_point. The variable_set Brick was used to define the various properties of the sample, such as gender, biomass, stage of development, etc. The data_point was then used to identify the property and the value.

Two important Brick suggestions resulted from this work. First, the current state of the Bricks needs to be examined from the standpoint of "gear" rather than the current situation of instruments and sensors. Instruments and sensors appear more applicable to electronic equipment, while many other forms of sampling deal with more mechanical devices, such as nets. Second, the data_set levels resulting from the "level" attribute in the data_set_id element needs to be examined. In biological sampling, a single sub-sample is often used for multiple analyses. This results in the identification of many species, each with identified gender, stage, biomass, etc. In effect, this results in an extra level in the hierarchical structure.

The resulting discussion wondered if the placement of the biological data into the Keeley Brick structure was a use for which the structure wasn't initially intended. It was noted that the Brick XML structure is very generic and therefore allows the placement of the biological data. Some thought the XML structure was in fact too generic, allowing too many possibilities for the placement of the data (Keeley Brick implementation schema is available here). The placement of the data in the XML structure is an optimisation vs. compliance issue (see Section 6.6.2 in reference 3). However, the promotion of duplicated information to higher levels of the structure is a technique for optimization, similar to the normalization process in data modelling.

There was also concern that the biology community will not use XML because it is too complicated. Tools need to be developed to support the use of the XML. The Japanese toolset for the Tokyo Bay Environmental Project is an excellent example. In this case, the detail of the XML is hidden from the user.

The date structure used in the Brick XML implementation was also questioned. The date structure is a subset of the XML Schema Specification for date, and is thus a subset of the ISO standard. The subset was chosen to force user compliance with GMT. The mandatory "Z" in both the date and time elements indicates this.

5 Sub-group activities

In preparation for the final report, the participants were asked to split into three subgroups pertaining to metadata, parameter dictionaries, and data. The three subgroups were asked to review, discuss, and summarize prior accomplishments pertaining to the topic area, and to make recommendations on future activities. 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 (10)
- Parameter Dictionary – R. Lowry (8)
- Data – A. Isenor (7)

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

5.1 Metadata subgroup

The metadata subgroup identified four achievements. It was felt that SGXML has:

- developed a consensus on metadata needs,
- raised awareness of metadata standards, in particular ISO-19115,
- made very good progress on harmonizing individual metadata standards (e.g., EDMED, CSR, NODC DDF, MEDI/DIF) to ISO-19115, and
- made progress on identifying the needs for oceanographic data specific profiles or extensions to ISO-19115.

In terms of possible recommendations to be considered for incorporation into the final report, the metadata subgroup put forth the following ideas:

- The terms being used in the metadata community are beginning to cause confusions. For example, the NDG considers metadata type A, while the Mikhailov model describes system metadata. The terms being used for the various types of metadata need to be compared and consolidated where possible.
- The explicit elements representing the oceanographic extensions to ISO-19115 need to be defined.
- Harvester software needs to be developed to add metadata from distributed repositories to a clearinghouse. With this task, there should also be a comparison of capabilities among the different systems (e.g., OAI, DiGIR). This work should be coordinated with JCOMM/IODE ETDMP and IODE GE-BICH.

5.2 Parameter dictionary subgroup

The parameter dictionary subgroup identified four achievements:

- An XML schema has been developed to map entries from multiple dictionaries to common terms. The schema has been used to support unit inter-comparison as demonstrated in a mapping between BIO, MEDS and IOS Canadian dictionaries.
- SGXML's interest has stimulated the development of the BODC Parameter Dictionary. This is evident by the BODC dictionary population increase from 7982 entries in May 2002 to 14431 entries in May 2004.
- SGXML is responsible for an in depth mapping between BODC and IFREMER dictionaries and BODC and the DONAR/WADI data models.
- The efforts of the SGXML have resulted in significant changes to BODC dictionary structure, including:
    - plain text descriptions being replaced by a semantic model;
    - the complete overhaul of the dictionary classification;
    - improved clarity of descriptions;
    - term definitions incorporated;
    - semantics, including classifications, removed from codes;
    - units are now considered a separate metadata element to parameter description;
    - on-line access to dictionary implemented.

In terms of possible recommendations to be considered for incorporation into the final report, the parameter dictionary subgroup put forth the following ideas:

- That the BODC parameter dictionary be adopted as the standard for marine XML;
- That the BODC dictionary be implemented as a register within the proposed IOC registry;
- That an improved mechanism be established for extension of the dictionary population (e.g., a review college);
- That improved web access be developed for the dictionary, including the development of web services;
- That there be a development of a semi-automated dictionary extension based on a web service;
- That a steering group be created to oversee interoperability standards for marine data. A proposal to establish such a committee should be put to IODE-18.

5.3 Data subgroup

The data subgroup identified two achievements:

- The SGXML has demonstrated that many data types (CTD, XBT, Current meter, Water Level, Underway TS, shipboard ADCP and to some extent biological net tow data) can be stored in XML using a single structure, built from a small set of generic data objects, or Keeley Bricks.
- The SGXML have assisted and influenced the local implementation of the software and schema developments for the Tokyo Bay Environment Information System.

In terms of possible recommendations to be considered for incorporation into the final report, the data subgroup put forth the following ideas:

- Considering biological data, the Ocean Biogeographic Information System (OBIS) should be examined and evaluated for potential use for XML-based data exchange.
- An effort should be made to consolidate GML, the Keeley Bricks, and the Japanese schemas into a single Marine XML, taking into account the mandatory content identified in the ICES WGMDM guidelines. Based on the outcome of the first recommendation, OBIS may also be considered in this consolidation.
- A demonstration project should be initiated to use the single schema developed in the previous recommendation, and demonstrate the Marine XML using a variety of data types and developed tools.

6 Action Items resulting from the meeting

As noted previously, this will be the last meeting of the SGXML. The Action Items resulting from the meeting are listed below. It is intended that these Action Items be completed by 31 December 2004.

Action 1: The SGXML Co-Chairs will formally ask ICES to recognize the Group until April 2005, the time of IODE XVIII.

Action 2: The UK members will continue supporting the SGXML mailing list until April 2005, the time of IODE XVIII.

Action 3: The SGXML Co-Chairs will draft a final report of the Group, to be reviewed and approved by the membership. The final report will summarize the efforts and accomplishments of the Study Group during its 3-year history. The report will also make recommendations for future activities. The report will be provided to ICES and IOC with the recommendation it is made public on respective web sites.

Action 4: In consultation with ICES and IODE, the Co-Chairs will create a proposal that identifies Groups, existing or to be created, as the potential recipients of the recommendations made in the SGXML Final Report.

Action 5: Using the results of Action 4 and 5, the Co-Chairs will prepare a submission for the IODE XVIII meeting in 2005.

Action 6: M. Fichaut will continue the effort to map the CSR to the ISO-19115 standard.

Action 7: D. Collins will examine the effort to harmonize the FGDC standard towards the ISO-19115.

Action 8: Existing crosswalk files, in XML, XLS or manipulations in XSL will be placed on the MarineXML website (see http://marinexml.net./).

7 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 Belgium hosts for extending warm hospitality to the Group while in Oostende. R. Gelfeld then closed the meeting by thanking all of those who had participated.

8 References

1. Report of the ICES-IOC Study on the Development of Marine Data Exchange Systems Using XML, ICES CM 2003/C:12, May 2003.
2. Isenor, Anthony W. 2003. XML Based Manipulation of Codes Exchanged Between Data Systems, DRDC Atlantic TM 2003-132, August 2003.
3. 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, May 2003.

9 Annexes

Annex 1 SGXML – List of participants

Name	Address	Telephone no.	Fax no.	E-mail
Pekka Alenius	Finnish Institute of Marine Research P.O. Box 33, (Lyypekinkuja 3) 00931 Helsinki Finland	+358 9 613 941 +358 9 613 94439 (direct)	+358 0 61394494	pekka.alenius@fimr.fi Web page: http://www.fimr.fi
Christophe Bamelis	Belgian Sea Fisheries Department (CLO-DVZ) Ankerstraat 1 8400 Oostende Belgium			christophe_bamelis@pandora.be
Sergey Belov	Russian National Oceanographic Data Centre 6 Korolev St., Obninsk, Kaluga Region, Russian Federation, 249020			belov@meteo.ru
Laura Bird	British Oceanographic Data Centre Proudman Oceanographic Laboratory 6 Brownlow Street Liverpool L3 5DA United Kingdom			labi@bodc.ac.uk
Donald W. Collins	U.S. National Oceanographic Data Center 1315 East West Highway, 4th Floor, Silver Spring MD, 20910, U.S.A.	+1 301 713 3272 extn 154	+1 301 713 3302	donald.collins@noaa.gov
Ray Cramer	British Oceanographic Data Centre, Proudman Oceanographic Laboratory 6 Brownlow Street Liverpool L3 5DA United Kingdom	+44 151 795 4897		rnc@bodc.ac.uk http://www.bodc.ac.uk
Michele Fichaut	Institut Francais pour le Recherche et l'Exploitation de la Mer (IFREMER), Centre de Brest, TMSI./IDM/SISMER, BP 70, 29280 Plouzane France	+33-2-98-22-4643	+ 33-2-98-22-4644	michele.fichaut@ ifremer.fr

Name	Address	Telephone no.	Fax no.	E-mail
Furukawa, Keita	National Institute for Land and Infrastructure Management Ministry of Land, Infrastructure and Transport 3-1-1, Nagase, Yokosuka 239-0826, Japan	+81 (468) 44 5023	+81 (468) 44 1145	furukawa-k92y2@ysk.nilim.go.jp http://www.nilim.go.jp
Robert D. Gelfeld, (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.	+1 301 713 3295 extension 179	+1 301 713 3303	robert.gelfeld@noaa.gov http://www.nodc.noaa.gov
Pieter Haaring	National Institute for Coastal and Marine Management (RIKZ) P.O. box 20907 2500 EX The Hague The Netherlands	Telephone direct: +31703114570 Telephone general: +31703114311	+31703114321 Fax direct: +31703114500	p.a.haaring@rikz.rws.minvenw.nl http://www.rikz.nl
Hernandez, Francisco	Flanders Marine Data and Information Centre Flanders Marine Institute Vismijn, Pakhuizen 45-52, B-8400 Oostende, Belgium			Francisco.hernandez@vliz.be http://www.vliz.be
Michael Hughes	British Oceanographic Data Centre Proudman Oceanographic Laboratory 6 Brownlow Street Liverpool L3 5DA United Kingdom			mhug@bodc.ac.uk
Anthony W. Isenor (Co-Chair)	Defence R&D Canada - Atlantic PO Box 1012 9 Grove Street Dartmouth, Nova Scotia Canada B2Y 3Z7	(902) 426-3100 ext. 106	(902) 426-9654	anthony.isenor@drdc-rddc.gc.ca http://www.atlantic.drdc-rddc.gc.ca/facilities/facilities_e.shtml
Roy Lowry	British Oceanographic Data Centre Proudman Oceanographic Laboratory 6 Brownlow Street Liverpool L3 5DA United Kingdom	+44 151 795 4895		rkl@bodc.ac.uk http://www.bodc.ac.uk
Angelino Meerhaeghe	Management Unit of the Mathematical Model North Sea (MUMM) Royal Belgian Institute for Natural Sciences Gulledelle 100 B-1200 Brussel Belgium	+32-(0)2-773 21 27	+32-(0)2-770 69 72	Angelino.meerhaeghe@mumm.ac.be
Nicolay Mikhailov	Russian National Oceanographic Data Centre 6 Korolev St., Obninsk, Kaluga Region, Russian Federation, 249020	7-084-397-49-07	7-095-255-22-25	nodc@meteo.ru

Name	Address	Telephone no.	Fax no.	E-mail
Georgiy Moiseenko	Russian Federal Research Institute of Fisheries and Oceanography (VNIRO) 17, V. Krasnoselskaya Moscow, 107140 Russia			georgem@vniroinfo.ru georgem@vniro.ru
Peter Pissierssens	Head, Ocean Services UNESCO-IOC 1, rue Miollis 75732 Paris Cedex 15 France	+33 1 4568 4046	+33 1 4568 5812	p.pissierssens@ unesco.org http://ioc.unesco.org/iode
Greg Reed	RAN METOC Maritime Headquarters Wylde Street Potts Point NSW 2011 Australia			greg@aodc.gov.au
Lesley Rickards	British Oceanographic Data Centre, Proudman Oceanographic Laboratory, 6 Brownlow Street, Liverpool L3 5DA, United Kingdom.	+44 151 795 4897	+44 151 652 3950	ljr@bodc.ac.uk http://www.bodc.ac.uk http://www.oceannet.org
Sagen, Helge	Institute of Marine Research Norwegian Marine Data Centre Nordnesgt 50 5817, Bergen Norway	+47 55 23 8500		helge.sagen@imr.no www.imf.no
Sakakibara, Tsuneki,	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			tsuneki_sakakibara@ kkc.co.jp http://shiba.iis.u-tokyo.ac.jp/sakaki
Reinhard Schwabe	Deutsches Ozeanographisches Datenzentrum (DOD), Bundesamt für Seeschiffahrt und Hydrographie Bernhard-Nocht-Str. 78 D-20359 Hamburg, Germany			reinhard.schwabe@bsh.de http://www.bsh.de
Szaron, Jan	Swedish Meteorological and Hydrological Institute, Oceanographic Unit, Nya Varvet 31, SE - 426 71 Vastra Frolunda, Sweden	+46 (0)31 751 8971	+46 (0)31 751 8980	jan.szaron@smhi.se http://www.smhi.se

Name	Address	Telephone no.	Fax no.	E-mail
Vanden Berghe, Edward, Manager,	Flanders Marine Data and Information Centre Flanders Marine Institute Vismijn, Pakhuizen 45–52, B-8400 Ostend, Belgium	+32 59 342130	+32 59 342131	wardvdb@vliz.be http://www.vliz.be
Vladymyrov, Vladimir Programme Specialist Ocean Services	UNESCO-IOC 1, rue Miollis 75732 Paris Cedex 15 France			v.vladymyrov@ unesco.org http://ioc.unesco.org/iode
Wichorowski, Marcin	Institute of Oceanology PAS, Powstancow Warszawy 55, 81-712 SOPOT, PL	+(48 58) 551 72 81,	+(48 58) 551 21 30	wichor@iopan.gda.pl http://www.iopan.gda.pl

Annex 2 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%

Annex 3 Detailed meeting agenda for SGXML meeting

Thursday 6 May

0900–1000	Opening greetings by Bob Gelfeld Welcome and local arrangements by Edward Vanden Berghe Review meeting schedule and items for discussion by Bob Gelfeld Review Action Items from last years meeting by Anthony Isenor/all participants
1000–1030	Discussion – The SGXML Vision and the Future (Introduced by Bob Gelfeld)
1030–1100	Coffee break
1100–1130	EDMED ISO-19115 XML Mappings (Michele Fichaut)
1130–1230	Reference Common Data Model (Nick Mikhailov)
1230–1345	Lunch
1345–1430	Practical testing of the common data model – distributed information system construction (Sergey Belov)
1430–1500	Report on Code Mapping, Action Item 1 (Roy Lowry)
1500–1530	Coffee break
1530–1615	Data Schema Standardization for Coastal Environment Data Set - Tools and Sample Data (Keita Furukawa and/or Tsuneki Sakakibara)
1615–1630	Comparing and Reconciling the Parameter Dictionary XML Structures (Ray Cramer)
1630–1715	XML and the NERC DataGrid (Roy Lowry)

Friday 7 May

0900–0945	Discussion - Metadata Investigation Action Items (Introduced by Don Collins)
0945–1000	Time-series Data in Brick XML Structure (George Slesser and Anthony Isenor)
1000–1030	Donar XML Update (Pieter Haaring)
1030–1100	Coffee break
1100–1130	GML Structures based on Brick Content (Anthony Isenor)
1130–1200	Proposed Implementation of a Feature Type Catalogue (Greg Reed)
1200–1230	Exploring Biological Net Tow Data in Keeley Brick 1-D Profile Structure (Anthony W. Isenor and Edward Vanden Berghe)
1230–1345	Lunch
1345–1500	Break up into individual Working Groups on:

- Parameter Dictionaries
- Point Data Investigation
- Metadata Investigation

We would like the working groups to:

- review, discuss, and summarize prior accomplishments of the Group with respect to the Working Group topic
- recommend future activities

We hope the output of the Working Groups will contribute directly to the final report of the SGXML.

1500–1530	Coffee break
1530–1630	Working Group presentations (10 minutes each) followed by open discussion
1630–1700	Closing and summation of Meeting

Annex 4 List of acronyms and terms

Acronym or Term	Description
AODC	Australian Oceanographic Data Centre
BIO	Bedford Institute of Oceanography (Canada)
BODC	British Oceanographic Data Centre
CSR	Cruise Summary Report
DDF	Data Documentation Form (USNODC)
DIF	Directory Interchange Format
DiGIR	Distributed Generic Information Retrieval
DONAR	Data Opslag NAtte Rijkswaterstaat Or in English: Data Storage Wet (Water related parts of) Rijkswaterstaat.
DSU	Data Storage Unit
DTD	Document Type Definition
EDMED	European Directory of Marine Environmental Data
EnParDis	Enabling Parameter Discovery
ETDMP	Expert Team on Data Management Practices
EU	European Union
FGDC	Federal Geographic Data Committee (USA)
FIMR	Finnish Institute of Marine Research
FRS	Fisheries Research Services
GCMD	Global Change Master Directory
GE-BICH	Group of Experts on Biological and Chemical Data Management and Exchange Practices (IODE)
GIS	Geographical Information System
GML	Geography Markup Language
GOOS	Global Ocean Observing System
ICES	International Council for the Exploration of the Sea
IFREMER	Institut Francais pour le Recherche et l'Exploitation de la Mer
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange
IR	Information Resource
IOS	Institute of Ocean Sciences (Canada)
ISO	International Organisation for Standardization
ITIS	Integrated Taxonomic Information System
JCOMM	Joint Commission on Oceanography and Marine Meteorology
JGIS	Japan Geographic Survey Institute
JGOFS	Joint Global Ocean Flux Study
JODC	Japan Oceanographic Data Center
MEDI	IOC Marine Environmental Data Information Referral Catalogue system
MEDS	Marine Environmental Data Services - Canada
MML	Marine Markup Language
NDG	NERC DataGrid
NERC	Natural Environment Research Council
NODC	U.S. National Oceanographic Data Centre
OAI	Open Archive Initiative
OBIS	Ocean Biogeographic Information System
OGC	Open GIS Consortium
OWS	OGC Web Services
SGXML	ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML
SMHI	Swedish Meteorological and Hydrological Institute
TOR	Term of Reference
UML	Unified Modelling Language
USNODC	United States National Oceanographic Data Center
VLIZ	Vlaams Instituut voor de Zee (Flanders Marine Institute)
WADI	WAter Data Infrastructure
WDCA	World Data Centre for Oceanography/Silver Spring
WGMDM	Working Group on Marine Data Management
WMO	World Meteorological Organisation

Acronym or Term	Description
XML	Extensible Markup Language
XSD	XML Schema Definition
XSL	Extensible Stylesheet Language
XSLT	Extensible Stylesheet Language Transformation

Annex 5 Action Items from 2003 Meeting

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.

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.

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)

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.