Copy No.________

Defence Research and Development Canada | Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

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

Anthony W. Isenor
DRDC Atlantic

Roy K. Lowry
British Oceanographic Data Centre

Prepared for:

The International Council for the Exploration of the Sea
Attention: Einar Svendsen
Chair ICES Oceanography Committee
Oceanography and Climate
Institute of Marine Research
PO Box 1870 Nordnes
N-5817 Bergen, Norway

Intergovernmental Oceanographic Commission (UNESCO)
Attention: Peter Pissierssens
Head, Ocean Services
UNESCO-IOC
1, rue Miollis, 75732 Paris Cedex 15
France

Defence R&D Canada – Atlantic

External Client Report
DRDC Atlantic ECR 2005-005
March 2005

Canada

This page intentionally left blank.

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

Anthony W. Isenor
DRDC Atlantic

Roy K. Lowry
British Oceanographic Data Centre

Prepared For:

The International Council for the Exploration of the Sea
Attention: Einar Svendsen
Chair ICES Oceanography Committee
Oceanography and Climate
Institute of Marine Research
P.O.Box 1870 Nordnes
N-5817 Bergen, Norway

Intergovernmental Oceanographic Commission (UNESCO)
Attention: Peter Pissierssens
Head, Ocean Services
UNESCO-IOC
1, rue Miollis, 75732 Paris Cedex 15
France

Defence R&D Canada – Atlantic

External Client Report
DRDC Atlantic ECR 2005-005
March 2005

Authors

Anthony W. Isenor
Defence Scientist
Co-Chair of the ICES/IOC SGXML

Approved by

D.G. Hazen
Head, Technology Demonstration Section

Approved for release by

Kirk Foster
DRP Chair

List of Study Group Members

The Study Group membership consisted of representation from 14 countries and two organizations. The table below identifies the active membership while Annex 1 provides a complete list of meeting participants. The chairs of other international groups are also identified.

Country or International Organization	Participants
Australia	Greg Reed
Belgium	Francisco Hernandez Edward Vanden Berghe (Chair GEBICH)
Canada	Jean Gagnon Anthony W. Isenor (Cochair SGXML)
Finland	Kati Manni Pekka Alenius
France	Thierry Carval Michele Fichaut (Cochair WGMDM)
Germany	Reinhard Schwabe
ICES	Harry Dooley
Japan	Keita Furukawa Tsuneki Sakakibara
Norway	Helge Sagen (Cochair WGMDM)
Russia	Sergey Belov Nicolay Mikhailov (Chair ETDMP)
Sweden	Jan Szaron
The Netherlands	Pieter Haaring ...continued

United Kingdon	Laura Bird Ray Cramer Michael Hughes Roy Lowry Lesley Rickards (Chair IODE)
United States	Donald W. Collins Robert D. Gelfeld (Cochair SGXML) Robert Starek

Abstract

In 2001 the Intergovernmental Oceanographic Commission (IOC) and the International Council for the Exploration of the Sea (ICES) cooperatively formed a Study Group to examine the application of the eXtensible Markup Language (XML) to marine data exchange systems. The Study Group first met in April 2002 to address issues around the transfer of oceanographic data. The Group has met three times, with the final meeting in May 2004. This document represents the final report of the Group.

The Study Group concentrated its efforts on metadata standards, parameter dictionaries and generic data structures for use in an XML-based language. The Group evaluated several international metadata structures and produced mappings between some structures. In terms of the parameter dictionaries, the Group conducted mappings between several international parameter dictionaries, made structural advances to some dictionaries and attempted to account for dictionary issues imposed by units. The generic data structure development produced about 20 data objects that were then used to create an XML data structure for the transport of ocean environmental data. The structure was applied to one and three-dimensional data sets. The Group has also made numerous recommendations to continue the development of international data exchange systems.

Résumé

En 2001, la Commission océanographique intergouvernementale (COI) et le Conseil international pour l'exploration de la mer (CIEM) formaient en coopération un groupe d'étude pour examiner les applications du langage XML (eXtensible Markup Language) aux systèmes d'échange de données marines. Ce groupe d'étude s'est réuni pour la première fois en avril 2002, afin d'étudier les problèmes liés au transfert des données océanographiques. Le groupe s'est réuni trois fois, la dernière réunion ayant été tenue en mai 2004. Ce document représente le rapport final du groupe.

Le groupe d'étude a concentré son travail sur les normes de métadonnées, les dictionnaires de paramètre et les structures de données génériques destinées à être utilisés dans un langage basé sur XML. Le groupe a évalué plusieurs structures de métadonnées internationales et a établi une correspondance entre certaines structures. En ce qui concerne les dictionnaires de paramètres, le groupe a établi des correspondances entre divers dictionnaires de paramètres internationaux, a effectué des mises à niveau structurelles dans certains dictionnaires et a tenté de prendre en compte les problèmes de dictionnaires imposés par les unités. L'élaboration d'une structure de données générique a produit environ 20 objets de données, qui ont été par la suite utilisés pour créer une structure de données XML pour le transport des données environnementales sur l'océan. Cette structure a été appliquée à des ensembles de données à une et trois dimensions. Le groupe a également présenté plusieurs recommandations en vue de poursuivre le développement des systèmes internationaux d'échange de données.

This page intentionally left blank.

Executive summary

Introduction

In 2001 the Intergovernmental Oceanographic Commission (IOC) and the International Council for the Exploration of the Sea (ICES) cooperatively formed a group to examine the eXtensible Markup Language (XML) with application to the transfer of oceanographic data. The IOC is a commission operating under the United Nations, specifically under the United Nations Educational, Scientific and Cultural Organization (UNESCO). ICES is an independent organization consisting of 19 member states that coordinates and promotes marine research in the North Atlantic Ocean.

The ICES/IOC Study Group on the Development of Marine Data Exchange Systems Using XML (SGXML) first met in April 2002 to address issues around the transfer of oceanographic data. The Group was tasked to develop generic data structures to be used in data transfer and apply these structures to physical, chemical and biological data sets. The Group was also tasked to investigate critical issues related to data exchange, including parameter codes and metadata. The Group concluded its effort in 2004.

Principal Results

The Study Group concentrated its efforts on metadata standards, parameter dictionaries and generic data structures that support data exchange. Metadata standards were evaluated and mappings created between key standards. An XML structure for a general parameter dictionary was created. This structure will assist in the mappings between parameter dictionaries. Finally, a generic set of data structures were developed and tested for use with ocean environmental data sets.

Significance of Results

In a networked environment, the data required by a particular client may reside in many different systems. Ideally, any client request for data should discover, collect and consolidate the available data sets throughout the networked system.

The Group's effort in metadata standards and related mappings provides direct support to the data discovery process. Metadata descriptions of data sets will be critical to the process, as these descriptions will be used during the initial discovery phase. As well, the mappings between metadata standards will permit the use of multiple standards within any single discovery system.

In a networked exchange system, the collection process follows the discovery of data. To aid in the understanding of this process, the Group developed a conceptual design for a software infrastructure that supports the collection and integration of data from the many data nodes in the network. The Groups contribution to this aspect is in the design of a distributed marine resource system.

Finally, the Group considered the consolidation of data from the various data nodes. In the consolidation process, commonality between parameter dictionaries is essential for providing the client with a single set of codes for any consolidated data set. The Group investigated common code systems and mappings between such code sets. As well, a central data structure based on generic data units was developed and tested on numerous ocean data sets. A set of about 20 generic data units, or Keeley Bricks, was used to construct the data structure. The structure was successfully applied to one and three-dimensional oceanographic data.

Future Plans

The Study Group has developed a list of recommendations to be considered by the IOC and ICES parent bodies and made available to the public and other marine organizations via the Marine XML website. These recommendations identify the need for consolidation of metadata terminology, explicit oceanographic extensions to existing standards, and the ability to combine metadata holdings from distributed sources. In terms of parameter dictionaries, the Group recommends the adoption of the British dictionary as the marine community standard and the creation of an international structure and procedures to manage the dictionary. Regarding the case studies, the Group recommends further examination of XML-based biological systems, and the merger of the Canadian and Japanese marine XML structures with application in a demonstration project.

Isenor, Anthony W. and Roy K. Lowry. 2004. Final Report of the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML, DRDC Atlantic ECR 2005-005, Defence R&D Canada – Atlantic.

Sommaire

Introduction

En 2001, la Commission océanographique intergouvernementale (COI) et le Conseil international pour l'exploration de la mer (CIEM) formaient en coopération un groupe afin d'examiner le langage XML (eXtensible Markup Language) en vue de l'appliquer au transfert des données océanographiques. La COI est une commission qui fonctionne sous l'égide de l'Organisation des Nations Unies, et plus particulièrement de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO). Le CIEM est un organisme indépendant, constitué de 19 États membres, qui coordonne et favorise la recherche marine dans l'Atlantique Nord.

Le groupe d'étude conjoint CIEM/COI sur le développement des systèmes d'échange de données marines au moyen de XML (SGXML) s'est réuni pour la première fois en avril 2002, afin d'examiner les problèmes liés au transfert des données océanographiques. Le groupe a reçu pour mandat de développer des structures de données génériques pouvant être utilisées pour le transfert des données, et d'appliquer ces structures à des ensembles de données physiques, chimiques et biologiques. Le groupe a également reçu pour tâche d'examiner les problèmes critiques reliés à l'échange des données, notamment les codes de paramètres et les métadonnées. Le groupe a achevé son travail en 2004.

Principaux résultats

Le groupe d'étude a concentré son travail sur les normes de métadonnées, les dictionnaires de paramètres et les structures de données génériques qui supportent l'échange des données. Les normes de métadonnées ont été évaluées, et des correspondances ont été établies entre les principales normes. Une structure XML pour un dictionnaire de paramètres général a été créée. Cette structure facilitera l'établissement de correspondances entre les dictionnaires de paramètres. Enfin, un ensemble générique de structures de données ont été développées et testées, pour pouvoir être utilisées avec des ensembles de données environnementales sur l'océan.

Importance des résultats

Dans un environnement en réseau, les données requises par un client particulier sont souvent réparties dans de nombreux systèmes différents. Idéalement, une requête de données présentée par un client doit permettre de découvrir, de recueillir et de regrouper les ensembles de données disponibles dans le système en réseau.

Le travail effectué par le groupe sur les normes de métadonnées et les correspondances connexes appuie directement le processus de découverte des données. Les métadonnées

décrivant les ensembles de données seront essentielles dans ce processus, puisqu'elles seront utilisées lors de la phase de découverte initiale. En outre, les correspondances établies entre les normes de métadonnées permettront d'utiliser plusieurs normes au sein d'un système de découverte unique.

Dans un système d'échange en réseau, le processus de collecte suit la découverte des données. Afin de faciliter la compréhension de ce processus, le groupe a élaboré un modèle conceptuel d'une infrastructure logicielle supportant la collecte et l'intégration des données issues de nombreux nœuds de données dans le réseau. À cet égard, la contribution du groupe réside dans la conception d'un système distribué de ressources marines.

Enfin, le groupe a étudié la consolidation des données issues des divers nœuds de données. Dans ce processus de consolidation ou de regroupement, la communité entre les dictionnaires de paramètres est essentielle pour fournir aux clients un ensemble unifié de codes, applicables à n'importe quel ensemble de données consolidé. Le groupe a étudié des systèmes de codes courants, et les correspondances entre ces ensembles de codes. En outre, une structure de données centrale, basée sur des unités de données génériques, a été développée et testée sur divers ensembles de données océanographiques. Un ensemble d'environ 20 unités de données génériques, ou briques de Keeley, a été utilisé pour construire cette structure de données. Elle a été appliquée avec succès à des ensembles de données océanographiques à une et trois dimensions.

Plans pour l'avenir

Le groupe d'étude a rédigé une liste de recommandations qui devront être examinées par les organismes parents, la COI et le CIEM, et qui seront rendues accessibles au public et à d'autres organisations marines via le site Web Marine XML. Ces recommandations font ressortir la nécessité d'une uniformisation de la terminologie des métadonnées, d'extensions océanographiques explicites des normes existantes, et de la capacité de combiner les dépôts de métadonnées issues de sources distribuées. En ce qui concerne les dictionnaires de paramètres, le groupe recommande l'adoption du dictionnaire britannique comme norme de la communauté marine ainsi que la création d'une structure et de procédures internationales pour gérer ce dictionnaire. En ce qui concerne les études de cas, le groupe recommande d'examiner plus en profondeur les systèmes biologiques basés sur XML, ainsi que la fusion des structures marines XML canadiennes et japonaises, avec une application dans un projet de démonstration.

Isenor, Anthony W. and Roy K. Lowry. 2004. Final Report of the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML, DRDC Atlantic ECR 2005-005, Defence R&D Canada – Atlantic.

Table of contents

List of figures

List of tables

Acknowledgements

The investigation carried out under the profile data case study utilized numerous individuals who were not official members of the SGXML. In particular, we thank J. Robert Keeley (Marine Environmental Data Service, Canada) and Joe Linguanti (Institute of Ocean Sciences, Canada) for their contributions to this effort.

The SGXML would also like to acknowledge the support provided by the membership home organisations. Attending and working on international projects requires resources and the support provided by the home organisations is gratefully acknowledged.

Finally, the International Council for the Exploration of the Sea and the Intergovernmental Oceanographic Commission provided the forum and in some cases resources, for the effort. We thank them for their support.

This page intentionally left blank.

1. Introduction

The flow of data is often a critical component of a research activity. Data often flows between collaborating researchers, institutes or data centres in an effort to obtain the most complete data set for addressing a particular research question. When research questions are on a global environmental scale, researchers need to consolidate data from many collaborating partners. This is because any single partner does not have the resources required to make direct environmental measurements on a global scale.

For oceanographic global scale questions, the spatial-temporal data requirements can be very demanding. Global climate questions typically require data sets spanning many decades and one or more of the world's oceans. Fortunately, the assembling of such data sets is made possible by the long-term efforts of the global ocean data management community. This community has been diligently consolidating, quality checking and archiving ocean data for over 100 years.

Of course the global ocean data community is in reality a collection of individual nations, all supporting ocean data management. Individual nations typically identify data centres to manage the ocean data on behalf of the national collectors. Individual collectors provide data to the national data centre, thereby ensuring the quality and long-term storage of the data.

From the national perspective, the national data centres provide the safe keeping for the country's data. Often, these data are collected with public funds and therefore represent a public asset. The data centre provides the infrastructure for managing this public asset.

However, the national centres may also be part of a larger international collection of data centres. In this case, the national centre contributes data to the larger international system. These data transfers support international collaborative efforts that often involve large scale programs or research questions.

The responsibility for the management of the national ocean data asset ultimately lies with the national data centres. However, the challenges and issues being addressed at one data centre are often common across many data centres. The international collection of data centres also provides a forum for discussion and collaboration on such common issues.

Such international collaboration is recognized as important at all levels. Data centres share technical knowledge with colleagues to assist many developments. However, such collaborations are also recognized at the highest political levels. For example, the 2003 Group of Eight (G8) Action Plan [1] reaffirms the political commitment for international cooperation related to global observations. The plan identifies the production of quality ocean data products, the importance of global data reporting, data archiving, data sharing and the filling of existing data gaps. Collectively, many of these data aspects are addressed by the internal data centres and may be termed "data management".

One ongoing data management issue faced by all data centres is related to advances in technology. Data centres are continually examining how technology can assist the centre in their mandate. However, the cooperative nature of the international data systems means that there is considerable sharing of technological information among centres. Typically this information sharing is conducted under the auspices of an international organization. With regards to this report, two international organizations form central roles: the International Council for the Exploration of the Sea (ICES), and the Intergovernmental Oceanographic Commission (IOC).

ICES is an organization consisting of 19 member countries involved in scientific studies in the North Atlantic. ICES coordinate and promote marine research among member states and is the oldest intergovernmental marine–related organization in the world, having been formed in 1902 [2]. In support of marine research, ICES maintains fisheries, environmental and oceanographic databases. The fisheries database maintains bottom trawl survey reports. The environmental database deals with chemical contaminants including trace metals and organics dating back to 1978. The oceanographic database includes cruise reporting information and records of ocean temperature, salinity and an assortment of other parameters, some dating back to 1892 [3].

The second organization noted above was the Intergovernmental Oceanographic Commission (IOC) [4]. The IOC was founded in 1960 under the United Nations Educational, Scientific and Cultural Organization (UNESCO) [5]. The IOC provides member states of the United Nations (UN) with a mechanism for cooperation in ocean related issues. The IOC represents 129 member states.

The IOC coordinates ocean related activities under three main themes: ocean sciences, operational oceanography and ocean services. The ocean science theme deals with broad program areas such as world climate and ecosystem research. Operational oceanography supports individual global scale programs such as the Global Ocean Observing System (GOOS). The ocean services theme deals with programs that support specific outcomes. One program under the ocean services theme is the International Oceanographic Data and Information Exchange system (IODE).

IODE is important because it represents the umbrella organization for international data centres to collaborate on data and information exchange, and product generation. IODE has 65 recognized National Oceanographic Data Centres (NODC) or Designated National Agencies (DNA). These data centres are national centres that are recognized in the IODE system. A small number of these NODCs are also recognized as Responsible National Oceanographic Data Centres (RNODC). The "Responsible" designation indicates that the centre has accepted additional responsibilities related to particular data types or specific geographic regions of the world ocean [6].

As an example, the Marine Environmental Data Service (MEDS) is the Canadian national oceanographic data centre (or Canadian NODC) [7]. As such, MEDS receives ocean data collected by Canadian researchers, institutes and private companies. MEDS quality controls the incoming data, provides long-term storage of the data and builds products from the data. MEDS is also the IODE RNODC for drifting buoy data. This means that MEDS provides the above functions for drifting buoy data collected globally.

Although ICES concentrates its research interests on the North Atlantic, ICES is also the IODE RNODC for data formats [8]. As such, ICES has a recognized interest and expertise in ocean data formats.

1.1 Outline

The two organizations described above have an obvious interest in the collection and distribution of ocean data. As such, they represent a natural pairing with regard to an investigation of structured data transfer. This report describes an IOC and ICES cooperative effort under the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML (or Study Group on XML, SGXML. Note, XML refers to the eXtensible Markup Language).

This report is a consolidation of information and results from the SGXML. The SGXML, which existed for about three years, consisted of 39 participants representing 14 nations and two organizations (Annex 1). The SGXML met three times, producing a report for each meeting. However, more important was the intersessional activities that resulted in considerable advances in the topic areas considered by the Group. This final report attempts to consolidate all previous work of the Study Group, as well as to place into perspective the Group activities relative to other international groups and organizations.

The report first provides a historical outline. This outline reviews the state of activities in the years preceding the SGXML and sets the context for the SGXML Terms of Reference (TOR). The TOR are then introduced and reviewed to provide a starting point for the remaining sections. For those unfamiliar with the XML environment, the very basics of XML are reviewed. The basics of XML will be used throughout the report to describe the results of the various SGXML activities. This is followed by sections that detail the activities within each TOR. Finally, recommendations are made for follow-on activities as well as suggestions for new or existing groups that could address these recommendations.

2. The Trail that Lead to SGXML

It is beneficial to review the history of the SGXML by considering the groups and events that lead to the Study Group formation. This is also useful because it provides an understanding of the SGXML Terms of Reference (TOR) and builds the framework of knowledge and activities that existed at the time of formation.

2.1 A Brief History

The SGXML evolved as a result of activities that took place along several different paths. One such activity path is related to initial XML investigations conducted by the ICES Working Group on Marine Data Management (WGMDM).

In retrospect, the WGMDM investigations were primarily educational in scope. The WGMDM members were unfamiliar with XML but were interested in the potential for XML to simplify data transfers between data centres. XML was first discussed in the WGMDM at the Ottawa 1999 meeting [9]. The Group was generally unfamiliar with XML, and as a topic, XML was only mentioned with reference to its capabilities.

At the WGMDM 2000 meeting in Hamburg [10], XML was again discussed. However, these discussions took on a more technical aspect. Individuals noted that their institutes were interested and actively investigating XML, in particular The Netherlands Institute for Sea Research and The Finnish Institute of Marine Research (FIMR). A sample XML document produced by G. Reed at the Australian Oceanographic Data Centre (AODC) was also reviewed. As well, it was also noted that IOC/IODE Group of Experts on Technical Aspects of Data Exchange (GETADE) was investigating XML. WGMDM decided not to initiate an XML investigation, but rather to wait until the GETADE investigation concluded. However, XML was placed on the TOR for the 2001 WGMDM meeting.

The second activity path is related to the GETADE. The IOC/IODE GETADE was established in 1979 under the IODE committee and was mandated to "*identify technical solutions for the management, exchange and integration of oceanographic data*" [11]. XML fit naturally within the GETADE mandate. In the March 2000 GETADE VIII meeting [12] the Chairman, N. Mikhailov from the Russian NODC, introduced a discussion on data formats and data unification. Included in this discussion was a presentation by G. Reed on background XML information. Reed noted the need to address three XML issues, namely [12]:

(i) which tags will be allowed;

(ii) how tagged elements may nest within one another; and

(iii) how the elements should be processed.

Reed also noted that AODC would be developing a Java based quality control software for use with XML. A reference document [13] provided an overview of XML and listed the numerous advantages to using XML.

It is important to note that GETADE recognized and commented on the need for a standardized data dictionary for use with XML. The data dictionary was considered necessary for the development of the standardized tags for use in the XML-based language. The Geographic Markup Language (GML) was also thought to be potentially useful for the geo-spatial metadata component of ocean data sets.

The ensuing two years saw considerable changes in the international oceanographic community. The GETADE IX meeting [14], held in April 2002, would be the last meeting of the Group. At this time, G. Reed had assumed the responsibilities of GETADE Chairman. Reed had also moved to become a consultant for IOC.

The two years between GETADE VIII and IX saw the formation of the Joint World Meteorological Organisation (WMO)/IOC Technical Commission for Oceanography and Marine Meteorology (JCOMM). JCOMM was established to deal with the integration of marine observing systems, the management of the data from these systems, and the services to support the systems and data management. In essence, JCOMM would focus on operational oceanography and the systems to support it. The first JCOMM session was held in June 2001 and resulted in the formation of the Expert Team on Ocean Data Management, later to be renamed the Expert Team on Data Management Practices (ETDMP). The ETDMP would by chaired by N. Mikhailov from the Russian NODC.

The second session of the JCOMM management committee was in February 2003. At this session, the JCOMM Committee requested that the IODE Committee consider the merger of GETADE with the JCOMM ETDMP. The IODE Committee, which met in March 2003, subsequently requested the review of GETADE and ETDMP work plans, followed by the recommendation and merger of the Groups.

The first session of ETDMP was held in September 2003 [15] (an informal session was held earlier, in November 2002 [16]). The first session noted the merger with GETADE, as well as the efforts of the SGXML and Marine XML. However, ETDMP did not identify specific XML initiatives, but rather wanted to cooperate with existing groups examining XML. This was a very reasonable approach, as three of the 10 ETDMP members were also SGXML members (N. Mikhailov, D. Collins and E. Vanden Berghe).

As noted previously, the April 2000 WGMDM meeting had placed XML on the TOR for the next meeting. The 2001 meeting [17] devoted considerable time to the XML topic. The WGMDM reviewed the very low-level concepts behind XML and also reviewed the AODC Java application for the quality control of XBT profile data in XML. H. Dooley (ICES) outlined the action options for the WGMDM, with the Group eventually agreeing on a proposal to the ICES Council for the creation of an ICES/IOC Study Group on XML. The ICES Council adopted the resolution at the September 2001 Statutory meeting [18].

The Study Group met a total of three times [19, 20, 21] producing meeting reports for 2002, 2003 and 2004. Activities of the Group have been publicized through both ICES and IOC, and interest in Group activities has been wide spread. Within the ICES community, the interest may be gauged by the recent summaries of ICES Group participation. The ICES 2003 Annual Report [22] lists all Groups (Steering, Planning, Working and Study Groups) with country and membership participation. Of the 90 Groups meeting in 2003, only five Groups exceeded the national representation of the SGXML.

There has also been considerable mention of the SGXML efforts within many of the Groups described above. However, other less obvious groups have also been interested in the efforts of SGXML. Other groups following the SGXML activities include the Chilean NODC [23], the North Pacific Marine Science Organization (PICES) Technical Committee for Data Exchange (TCODE) [24], the United Kingdom Environmental Data Network [25], and the joint Steering Committee of the Global Ocean Observing System [26].

2.2 SGXML Initial Terms of Reference

The Terms of Reference (TOR) for the SGXML evolved over the three-year period of the Groups existence. However, it is instructive to examine the initial TOR, as these will form an outline for the document that follows. The initial TOR were approved by the ICES Council in 2001 [18] and were as follows:

2C11 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 be established and will meet in Helsinki, Finland on 15–16 April 2002 to:

a) develop a framework and methodology for the use of XML in marine data exchange in close consultation with IOC and the Marine XML Consortium;

b) develop a workplan that within 4 years will lead to published protocols for XML use in the marine community;

c) explore how to best define XML tags and structures so that many ocean data types can be represented using a common set of tags and structures;

d) test and refine these common tags and structures using designated case studies, i.e;
Point (physical/chemical) data (profile, underway, water sample),
Metadata (cruise information, building from the ROSCOP/Cruise Summary Report),
Marine Biology data (integrated tows, e.g., zooplankton-phytoplankton tows, demonstrate the use of taxonomy).

SGXML will report by 1 June 2002 for the attention of the Oceanography Committee and ACE.

There are several important points made within the TOR. First, the Group recognized the need to first develop a plan (e.g., framework, methodology, workplan) for the Group. At the time, participants recognized that no clear direction currently existed. The Group would first need to define the areas where participants thought XML could be useful.

The IOC and the Marine XML Consortium were also given prominence. The reference to the IOC provided emphasis on the requirement to broaden the multi-national membership. The IOC also had, under the IODE, formed the Marine XML Project. The Project was funded by the European Union (EU) under the 5^{th} Framework Programme. The Project outline [27] includes the production of prototype marine data ontology, working demonstrations of the data interoperability, a developed prototype Marine Markup Language specification and the advancement of the standardisation of a Marine Markup Language. Although no direct working relationship was established between SGXML and the Marine XML Project, five SGXML participants are also active in the Marine XML Project.

The initial TOR also emphasised the quest for common language (e.g., common tags and structures) across many data types. The ocean data community deals with a multitude of data types, but many aspects of the metadata and data are common across these types. By investigating the commonality across data types, the Group hoped to implement the common structures in XML.

The test cases for the common structures were also defined in the initial TOR. The physical/chemical aspect of oceanography would be represented in the point data case study. Metadata was also recognised as an important component and thus given emphasis. Finally, biological data was recognised as a case study. Biological data contains a considerable number of interrelationships and thus provides unique challenges for any developed data structure.

The TOR evolved slightly over the course of the three years; however, the general topics remained the same with one exception. The coding of parameters became a dominant component of the SGXML effort. The parameter code problem was quickly recognised as critical to any data exchange and as such became a central theme of the SGXML effort.

3. XML Basics

The eXentsible Markup Language (XML) was developed by the World Wide Web Consortium (W3C) with the release of the XML specification in 1998. XML is actually a meta-language, a language used to develop other languages. The developed language is thus based on XML, but is not properly named XML.

In the simplest of terms, XML may be used to construct a language using any known computer based character set. XML provides various structures that may be used to capture the data. The simplest XML structures are elements and attributes.

An XML element is similar to a data object. It may contain other elements or attributes. XML syntax used to identify an element is the angle bracket, < and >. For example, the element named cruise would be written as

```
<cruise>
```

The actual text and angle brackets represent the tag. To close an XML element, the / is included in the trailing tag. For example:

```
<cruise>
</cruise>
```

Alternately, an empty tag may be shortened to be:

```
<cruise/>
```

To encapsulate another element inside the cruise element, the syntax would be:

```
<cruise>
  <station>5</station>
</cruise>
```

Here, the leading spaces on the station element are included for clarity. The numeric 5 is the content of the <station> element. By enclosing elements within other elements, one creates a hierarchical data structure.

An attribute for an element maybe included within the starting tag. For example, if the <station> element had an attribute "name" and the name of the station was "Bravo", then the syntax would be

```
<cruise>
  <station name="Bravo">5</station>
</cruise>
```

A namespace may also apply to the developed language. Although namespaces will not be dealt with in detail, they do appear in the schema elements within this report. A namespace may be considered a specifically named topic area for the developed language. For example, if the developed language for this project were “Ocean Data XML” and the namespace “odax” was declared to represent this language, then the namespace addition would be

```
<odax:cruise>
  <odax:station name="Bravo">5</odax:station>
</odax:cruise>
```

Finally, in XML terminology, there are well-formed documents and valid documents. Well-formed means the start and end tags occur in the proper sequence, similar to a stack first-in-last-out feature. Valid means the document agrees with a structure defined in a schema.

There are many more syntactic rules for constructing XML based languages. These rules will not be reviewed here. Those interested are referred to the many on-line resources or published books on the subject (e.g., see [28], [29]).

3.1 Common Tags, Structures and Codes

The SGXML TOR noted above, made reference to a common set of tags and structures that could be used in an ocean data XML. The ability to establish a set of common tags and structures is a strength of the XML environment. This combined with the fact that the common tags do not necessarily need to be directly related to a particular database structure, but rather are transformable into many structures.

The initial SGXML investigations were to examine common tags and structures, and metadata. However, during these investigations the importance of parameter codes quickly became evident. Both the metadata descriptions of data sets and the data content within an XML document depend critically on a description of the data content, which is typically based on a parameter code.

The importance of the parameter code was also recognized in the initial case study involving the point data. The application of a common structure to a data set is only useful if the parameters are described in the data set. If the data exchange is to be successful, the parameters must be known by the receiving party.

Thus, a Group plan began to form around three themes – metadata, codes and common structures.

4. Metadata

Metadata was recognized within the SGXML as critical to the data distribution in a fully networked environment. This is because metadata will be used in any data discovery exercise within a network. For example, consider search software that seeks certain data (e.g., temperature data) within a specific region in x,y,z,t space. If a resource on the network properly describes the available data source, services would be able to discover the data using the metadata descriptions.

This example of a discovery may be made more specific, if the metadata descriptions are available. For example, the discovery could involve particular instruments, sensors or analysis methods. However, the request can only become more specific if the metadata exists to answer the query and if the search techniques can interupt the metadata description.

The metadata investigation examined various metadata structures in use throughout the oceanographic community. For the initial investigations, metadata was considered those data that in some way support the specific ocean data sets. For example, metadata includes information that describes the sensors used to collect data, the data types collected and sampling locations. However, during the SGXML effort, the metadata definition was expanded to include other information important in a networked environment.

It is unlikely that one metadata structure will exist over the network for all data descriptions. Thus, one of the keys to discovery is the ability to search multiple metadata structures, linking these structures to specific search requirements. This means the search requirement can be expressed in terms of the various metadata structures, accounting for the fact that different organizations may make their metadata available in these different structures. Thus, one of the first SGXML tasks was to educate the members on the many different structures for metadata storage. This education would also serve to raise the membership awareness of the importance of metadata.

Also, the SGXML recognized the need to understand past efforts directed toward a common data structure including the metadata. Most notably, was the General Format 3 effort (GF3). The past effort of GF3, together with the metadata descriptions currently available, would form the starting point for the metadata investigation. A brief synopsis of these topics follows in the next sections.

4.1 Previous Work - General Format 3

The SGXML membership recalled past efforts to develop a common format for the transfer of oceanographic data. In particular, General Format 3 (GF3) was an international effort that attempted to construct a general-purpose format for the transfer of ocean data. Developed in the early 1980s, GF3 was a character encoded ASCII or Extended Binary Coded Decimal Interchange Code (EBCDIC) format for the transfer of oceanographic data on magnetic tape [30].

GF3 consists of headers and subsections, similar in structure to the element/subelement nature of XML. In fact, the commonality between the GF3 development and XML was noted in initial XML discussions that lead to the creation of SGXML[1].

The SGXML thought a GF3 review would be useful in the formulation of generic data objects. As such, SGXML reviewed the headers and sections of the GF3 structure [19, see 2002 meeting report]. This investigation noted that the elements resulting from the GF3 "series header" should be considered useful. However, the elements from the GF3 "data cycles" section were weak. Overall, GF3 may be used as a reference for checking any developed structure, but is not useful for defining generic data objects on its own.

The SGXML also reviewed the historical impact of GF3 on the data community. The use of GF3 for the transfer of oceanographic data was not as successful as the GF3 developers originally hoped. However, one aspect of GF3 that obtained considerable use was the code tables. The code tables were one of the first attempts at a common set of international codes. The table was initially conceived as supporting the format, but obtained considerable use in other data structures. At present, there are many extensive code tables, including those commonly called parameter dictionaries, available to the oceanographic community. Many of these are variations of the original GF3 code structure. The parameter codes will be extensively dealt with in Section 5.

4.2 Metadata Structures

As noted previously, metadata descriptions will be critical for the search and discovery aspect of a networked data system. As such, the SGXML membership needed to familiarize themselves with the various metadata structures that may be used within the ocean data community.

Obviously, the ideal situation would be a single metadata structure used by all parties contributing to a data system. However, a single structure is unlikely. The autonomy of the data centres and the need to address different data and political requirements will result in multiple structures. A perfect example of this is the United States (US) Executive Order [31] requiring the use of the Federal Geographic Data Committee (FGDC) standard for all US geospatial data. While the FDGC is mandated in the US, many European centres appear to be utilizing the European Directory of Marine Environmental Data (EDMED). Although different standards are in use, the consistency of use in the different regions should be seen as a positive, as it coordinates the metadata to a single standard within the particular region.

---

[1] Presentation by Anthony W. Isenor at WGMDM Meeting, Birkenhead, UK, April 2001.

Many metadata standards exist for oceanographic data. The following is a list of standards reviewed by the SGXML [19, see 2002 meeting report]. More recently, a more complete Marine Metadata Interoperability Project [32] (MMI) has begun to examine the interoperability of metadata standards. This is an important iniative that was not underway at the time the SGXML was active.

4.2.1 ISO 19115

The International Organisation for Standardization (ISO) Technical Committee (TC) 211 has developed an international metadata structure called ISO 19115 [33]. Approved in July 2003, the standard provides detailed descriptions of the entities and attributes (which comprise over 300 elements) covering the following topics:

- data set access constraints,
- data set maintenance frequency,
- raster, vector spatial representations,
- spatial-temporal reference system,
- distribution details (fees, availability, media, ...),
- spatial extent of the data set, and
- citation, contact and responsible party information.

The ISO 19115 standard defines core metadata components, recommended components and allows community based profiles to be described as extensions to the standard. The ISO 19115 standard is widely seen as the international standard for metadata descriptions.

ISO 19115 is a georeference metadata standard. As such, ISO 19115 does not contain all the necessary fields to adequately describe ocean data sets. These fields will need to be constructed by the ocean data community and made compliant with the ISO 19115 via the user extension capability of the standard.

4.2.2 FGDC

The Federal Geographic Data Committee (FGDC) developed the US Content Standard for Digital Geospatial Metadata (CSDGM). The CSDGM standard is commonly termed the FGDC standard [34]. The FGDC standard has no controlled vocabulary but does allow selection of thesauri to control the vocabulary. This system is highly granular, and in the US it has a growing population of users. In part, this is due to the US mandating its use for geospatial data [31].

4.2.3 EDMED

The British Oceanographic Data Centre (BODC) developed the European Directory of Marine Environmental Data (EDMED) [35] in the early 1990s as part of the European Commission (EC) Marine Science and Technology (MAST) project. The filling of the directory started in 1991 as part of the MAST project. The directory now contains about 2300 data set listings and 500 data centre listings.

EDMED may be considered a high-level inventory. In terms of data sets, EDMED allows descriptions of geographical area, observations, descriptions and parameters associated with a data set. For the data centres, EDMED provides descriptions of the centre, address, country and website address. Representation from 25 European countries is listed in the data centre inventory. No data are accessible through the EDMED system but the EDMED itself is searchable online [36].

4.2.4 DIF

The Directory Interchange Format (DIF) is a metadata structure that was developed in the late 1980s, thus predating both FGDC and ISO 19115. The DIF was originally conceived at a 1987 workshop on Earth Science and Applications Data Systems (ESADS). This was followed (also in 1987) by formal definitions of the format by NASA and other US agencies [37].

The structure has six mandatory fields and 30 optional [38]. Many of the fields have lists of valid content, termed ‘valids’. These ‘valids’ apply to fields such as geographic location, platform types and sensors. In DIF version 7, many of the fields were highly grouped (e.g., address). In version 9, the terms have been split to better align with ISO 19115. For example, the address field has been split to include specific fields for city, province (or state), postal code and country.

The DIF is used in two systems well known to the oceanographic community. The Marine Environmental Data Information Referral Catalogue (MEDI) is an IOC directory system for data sets, catalogues and inventories that utilizes the DIF structure. The system consists of a PC based tool for the recording of DIF information via a graphical user interface [39].

The DIF structure is also used in the Global Change Master Directory (GCMD) [40] developed by NASA. GCMD is a system that uses DIF as its input/output structure.

4.2.5 MARC 21

The US Library of Congress developed the MAchine Readable Cataloguing (MARC) record [41] in the 1960s. MARC is a format for the exchange and use of bibliographic information. As a bibliographic-related format, its primary goal is the identification and description of written material. However, it has been extended to support other material such as maps and music scores.

The MARC data record consists of a designation code and content. In the simplest of records, this means a numeric code identifier followed by the content of the record [42]. However, some record designators contain subfields, which further specify the content. This field/subfield type of structure easily leads to an XML structure.

MARC was not considered a viable option for the metadata information related to oceanographic data. The structure is not widely used in the oceanographic data community and in many ways does not pertain to ocean-related data collections.

4.2.6 Dublin Core

Dublin Core is a standard developed by librarians [43] who were seeking interoperability between metadata collections. The Core represents a consensus of metadata elements. Extended to a set of attributes, it has become an ISO standard [44]. The Core set is now 15 elements. The elements are multi-lingual and could be useful for library-style cataloguing of cruise reports. The Core also allows for extension into a particular field of study.

In terms of oceanographic data discovery, the Dublin Core is not suited because of the lack of geospatial characteristics. Although the geospatial components could be added, the Core was developed for library-type operations and does not easily support oceanographic data. However, the format may have a management role in initiatives to develop oceanographic data sets into citable entities on a par with published papers.

4.2.7 Assessment of Standards

The SGXML reviewed and evaluated the numerous metadata standards. One study [45] considered the geospatial characteristic of the metadata standards. The study examined many of the standards noted above including Dublin Core, ISO 19115, DIF and FGDC. It also examined the Global Information Locator Service (GILS). All the standards were considered based on their minimum sets (or mandatory elements) and the support for geospatial data.

The study indicated that in terms of oceanographic data discovery, the Dublin Core is not suited because of the lack of geospatial characteristics. Although the geospatial components could be added, the Core was developed for library-type operations and does not easily support oceanographic data. It is, however, suited to cataloguing functions associated with the management of oceanographic data sets.

The study also showed that the FGDC and ISO standards were the most relevant to geospatial data sets. It was noted that the FGDC documentation was much easier to understand and is more compact. The ISO standard was difficult to follow in part because of the numerous references to other ISO standards.

Although FGDC meets many of the needs of the geospatial community, the ISO 19115 has been noted to be more complete. Teng [46] indicates the ISO was more complete in the area of maintenance information, data constraint information, catalogue rules information, and the application schema. See [45] for an overview of the Teng comparison and [46] for the details.

4.3 Mapping of Metadata Standards

4.3.1 Existing Mappings

Mappings, or crosswalks as they are sometimes called, are an important method of consolidating information sources in a networked environment. Such mappings provide the automatic systems with the ability to examine metadata sources in a variety of structures. The system queries the metadata source and applies the mapping to relate the source structure to the structure utilized within the system. This allows the system to understand and process the metadata source.

User communities and development teams have provided numerous mappings online. Some of these mappings are noted here:

- FGDC to DIF [47, 48]
- MARC 21 to Dublin Core [49]
- DIF to Dublin Core [50]
- DIF to ISO 19115 [51]

There is also work underway to harmonize the two main geospatial metadata standards, FGDC and ISO 19115 [52]. This work is ongoing by the FGDC and ISO TC 211.

4.3.2 SGXML Mapping Contribution

The metadata standards reviewed in Section 4.2 are in wide use throughout the ocean data community. However, in the immediate future we cannot expect to have all parties agree on one standard. Thus, the mappings between the standards are particularly important.

In a distributed system utilizing multiple metadata standards, any central system needs to be able to utilize the metadata in whatever form provided. As such, there should be mappings between the standards in use at particular labs, to a common standard. The ISO 19115 standard was recognized by the SGXML membership as the choice for the common standard.

The SGXML contributed two mappings of particular importance. The first was the MEDI to ISO 19115 mapping. This mapping shows that nine MEDI elements have no direct mapping to ISO 19115 elements. The unmapped MEDI elements include data set citation reference, attributes for source name, sensor name, project, depth resolution, time resolution, and the metadata information covering last revision date, future review date and revision history. In the ISO 19115, these elements may form part of the community extension requirements.

A second mapping was also conducted from the European EDMED to ISO 19115. This mapping provides all EDMED tags and the mapped ISO 19115 tags. For the ISO standard, both the tag name and the tag number are provided to eliminate any confusion as to the tag being used. The mapping deals with all EDMED tags.

Such mappings are useful for integrating systems over a distributed environment, because the mappings provide systems with the ability to interpret metadata content from sources using different metadata structures. However, mappings are typically incomplete in terms of the semantics. This means the entire set of information contained in one structure is not typically transferable to a second structure.

4.4 The Distributed Data Model

In a broader context, metadata is used to describe some aspect of a data set. The description typically involves the data itself. However, when considering a networked data source, one quickly realises the need for metadata on a variety of topics related to the connection, the data source, etc.

The SGXML metadata investigation realised the importance of the distributed data resource. Fortunately, one SGXML member (N. Mikhailov, Russia) was also involved in the JCOMM ETDMP, and as such, could draw on resources related to ETDMP. In particular, EDTMP was interested in the creation of a system for joining distributed oceanographic data sources.

Work related to the joining of distributed data sources was also part of the SGXML effort. The Russian members [53, 54] provided documentation, demonstrations and example structures for a model of a distributed data system. The documentation also outlines the terminology that could be used for discussion of such a system.

Many aspects of this development are important, but two aspects are particularly relevant to the work of the SGXML. The first aspect is related to metadata and the descriptions of the various types of metadata. The work [53] described four types of metadata:

- Unification metadata – This level of metadata specifies information that is relevant to the integration of the data set into the larger distributed system. This metadata includes dictionaries of metadata attributes, which describe the available metadata. Also included are any parameter dictionaries listing the codes used in the data set. Finally, tools that may be used in the integration of the data set with the larger distributed data set would also be included with the Unification metadata.

- Service metadata - This metadata is related to the distributed system. This type of metadata describes the data content, the location of the data and the access method (including such things as user privileges). Service metadata is used for the navigation, searching and integration of data sources.

- Thematic metadata – This is the type of metadata we are familiar with and typically describe as “data about data”. Thematic metadata describes the features of the observation such as methods of data collection, data processing, data accuracy and data quality.

- Associated metadata – This metadata may be considered an extension of the Thematic metadata. The Associated metadata includes information on the observation platform, measurement techniques and processing techniques. This information would contribute to a fuller understanding of the data, but would not be considered critical to the data set (unlike the thematic metadata).

These metadata descriptions are similar to the descriptions developed under the Natural Environment Research Council (NERC) DataGrid Project (NDG). In this project, six types of metadata were defined [55]:

- Archival (A) – This is defined as the metadata required to support data browse and data delivery services. It includes information such as spatial coverage, access privileges parameter lists, physical data location and storage format.

- Browse, (B) – This is defined as the metadata required to support discovery metadata generation and metadata browse services (the process of locating further data sets of interest by navigation of a structured metadata repository). B metadata includes information such as parameter and spatial coverage and linkage metadata such as data collection activities (cruises, projects, etc.) and data collection instruments. Browse services enhance data discovery, providing more structured navigation through the metadata that overcomes problems such as discovery record format limitations

- Summary (S) – This is the overlap between metadata types “A” and “B”, such as spatial coverage and parameter lists. There is a difference in the level of detail between representations in “A” and “B”, with “A” being more detailed.

- Discovery, (D) – This is defined as the metadata that populates the discovery portals. D metadata is designed to be searched by users or software agents looking for data sets. Discovery metadata comprises totally public domain information encoded in records conforming to established standards such as Dublin Core, DIF or ISO19115. It is a subset of Browse metadata.

- Collection (C) – This ancillary metadata allows the association of things such as publications and annotations with data sets. To date, little consideration has been given to C Metadata within the NERC DataGrid project. However, each NDG metadata entity has been given a unique, persistent identifier that provides a ‘hook’ through which external collections may link to NDG objects and thus data sets.

- Extra (E) – This is defined as metadata provided by systems over and above that held within the core NDG schemas. In practice, “E” metadata will result either because the schemas cannot accommodate the information or because ingestion from existing repositories into the schemas is infeasible (e.g., resource limitations). For example, a PDF cruise report may easily be linked through the “B” metadata to a data set as a URL . Extracting all the relevant information and encoding it into the “B” schema requires significantly more effort. However, the strategic objective is to transform “E” metadata into “B” through schema evolution and ingestion efforts wherever possible.

Many of the fields in the Archive and Browse metadata, such as parameter descriptions and metadata mappings, are populated from controlled vocabularies and reference lists that, wherever possible, conform to accepted standards. These do not form part of the NDG metadata taxonomy as they are external to the project and are referred to as ‘Reference Standards’.

Efforts are underway to relate and compare the Russian metadata types and the NDG types. Initial investigations suggest the following are similar:

- Unification and Reference Standards

- Service and the superset of Archive and Browse. Service also covers the functionality of Discovery.

- Thematic and Browse with an element of Extra

- Associative and Extra with an element of Browse

However, a more detailed examination of the metadata content needs to be conducted to properly relate these types.

The second aspect of the Russian work that is relevant to the SGXML effort was the development of a XML data structure to support the transfer of ocean data in the distributed system. This development [54] recognized the need for multiple XML structures to support various requirements of a distributed system. For example, there will be a need to describe data resources in the distributed system. The development [54] proposes a structure for resource description.

4.5 Metadata Achievement Summary

As a summary, the SGXML identified four group achievements related to metadata. The SGXML has accomplished the following.

- The SGXML developed a consensus on metadata needs. The SGXML membership recognizes the critical role metadata will play in the next generation of ocean data management. In an integrated system, where many centres are connected via a network, the discovery mechanism will be critically linked to the systems ability to search metadata listings that summarize data holdings.

- The SGXML raised the awareness of metadata standards, in particular ISO 19115. Many standards exist for describing metadata. Most of these standards have been developed out of communities of interest to meet the particular community needs. However, the ISO 19115 standard addresses international issues while providing methods for community based extensions to the standard.

- The SGXML made very good progress on harmonizing individual metadata standards (e.g., EDMED, CSR, NODC DDF, MEDI/DIF) to ISO 19115. The SGXML membership recognized the inability of any one standard to obtain total community support. Although the standard may support a wide audience, individual organizations or groups maintain the ability to choose whichever standard is applicable. Thus, to support discovery, the searching software needs to be able to describe the search requirements in various metadata standards. These requirements lead the SGXML to develop mappings between the various metadata standards. The common standard, to which all others were mapped, was the ISO 19115 metadata standard.

- The SGXML made progress on identifying the needs for oceanographic data specific profiles or extensions to ISO 19115. During the process of mapping one metadata standard to another, the SGXML has begun the process of identifying the important components of a metadata standard applicable to the oceanographic data community. This will be an important contribution to defining a standard for use in the ocean data community.

5. Parameter Dictionaries

In most science fields, values of interest may be considered observations, measurements, or model output. These values are typically represented as a value and a descriptor. The value may be numeric or character, with the descriptor typically being a description of the observed, measured or modelled quantity. Often, the description is not a full text description, but rather a short code that is linked to the detailed definition.

A code may be considered some abbreviated form of information that describes the value. A code may be a standard word in a particular language. An example here would be 'temperature'. In more abstract cases, a code may be a combination of letters or numbers that are constructed as a concatenated string that represents measurements, instruments, or procedures. An example here would be TPXB representing temperature from an expendable bathythermograph (XBT). This form of a code, where semantics are included in codes of restricted length, is a non-scalable procedure that will inevitably lead to problems.

In the most general case, a code may be considered as simply a set of hieroglyphics. In all cases, the code has a strict definition associated with it. The definition may be minimal, representing only the name of the parameter. Alternately, the definition may be detailed, including information on collection technique, units, instrumentation, etc. A group of such codes and their definitions represents a parameter dictionary. A more thorough description of codes may be found in [56].

The use of codes provides various advantages in terms of understanding and software development [56]. In a database they provide compaction of information, clarity of definition and may be used as keys within a relational schema. In the multi-lingual environment, the codes also provide single source software with an easy mechanism to present the user with information specific to their language requirements.

5.1 BODC Parameter Dictionary

A set of codes may be managed in many different ways. In the simplest terms, a set of codes could be in a text file, a spreadsheet or a database. Regardless of the management system being used, the set of codes represent a description of the parameters being considered by the local organization. In this report, the managed set of codes is termed a parameter dictionary.

The BODC oceanographic parameter dictionary is by far the most extensive parameter dictionary available to the oceanographic community, containing over 16,500 codes (December 2004). Early in the investigations involving the XML exchange of oceanographic data [57], the importance of parameter codes was recognized. This investigation, which is described in section 6, made clear the importance of using the same or at the least, mappable codes, in any exchange environment. However, the use of a single parameter dictionary is by far the best solution to the exchange of data that identify values using codes.

The SGXML was given a preliminary presentation of the BODC parameter dictionary in 2002. At this time, BODC had made the dictionary available over the web. It was thought that a move to an XML form would serve to enhance and publicize and accessibility of the dictionary. Initiatives are now underway to make the BODC dictionary dynamically available on-line through web services and an RDF-based thesaurus server [58] to supersede the static on-line system developed for SGXML. The BODC dictionary is currently available for download as comma separated fields or as a Microsoft Access® database [59].

The SGXML also considered the mapping exercise to be a knowledge management problem that could involve ontologies. However, initial mappings conducted by SGXML required more direct human resources. In particular, people to compare existing codes from two dictionaries. Considerable progress was made on these mappings in large part due to funding provided by NERC under a project named Enabling Parameter Discovery (EnParDis).

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 cell, which contains a code, being beside the mapped code from the second dictionary. Table 1 shows an example.

Table 1. An example of the Canadian mapping that relates codes from three different parameter dictionaries. A blank cell indicates no code is defined. The Institute of Ocean Sciences (IOS), the Bedford Institute of Oceanography (BIO) and the Marine Environmental Data Service (MEDS) codes are shown.

IOS CODE	MEDS CODE	BIO CODE	FUNCTION
Alkalinity:Total	ALKY	ALKY	Total Alkalinity
Carbon:Particulate:Organic	CPX1		Particulate C = CORG PX
Nitrate	NTRA	NTRA	Nitrate (NO3-N) CONTENT

However, a dictionary is a dynamic entity, changing with the new additions of parameters. It will therefore evolve through time. Mapping large, information rich parameter dictionaries is a labour intensive process that can take months to complete. Map maintenance, tracking changes in each featured dictionary and ascertaining their impact on the map, can be even more demanding on resources. Consequently, mappings between dictionaries often fail to keep current.

It is inevitable that whilst there is considerable overlap between parameter dictionaries, there are some parameters that are unique to each dictionary. For example, in Table 1 there is no BIO code for particulate organic carbon. Consequently, additional entries need to be added to one of the dictionaries if a complete map for the other is to be produced. The approach taken in EnParDis was to populate the BODC Parameter Dictionary to cover all entries in the dictionaries mapped to it.

The project started by manually mapping the BODC Parameter Dictionary to the US Joint Global Ocean Flux Study (JGOFS) codes. This showed that manual mappings are extremely laborious and time consuming and the exercise was never fully taken to completion. It also showed that the more information a dictionary entry contained for a given code, the more likely the requirement for extra entries in the master dictionary. In particular, if the parameter code follows the GF3 practice of carrying information on units then massive dictionary expansion is required to support mapping of biological parameters. Consequently, divorcing units from the information carried by parameter codes is a positive step. Small dictionaries are much easier to both map and maintain. Supporting this requires data formats and metadata schema to include separate fields for parameter and unit codes.

EnParDis next considered mapping two Institut Francais pour le Recherche et l'Exploitation de la Mer (IFREMER) dictionaries to the BODC dictionary. The lessons from this mapping included the realization that codes associated with general method descriptions are easier to map than codes that are associated with detailed information. The exercise also exposed the problems that result if the code descriptions are unclear or ambiguous.

A particular problem was noted with the common practice in oceanographic and climate data management to have specific codes for sea surface temperature (SST) and sea surface salinity (SSS). Besides the obvious problem of accurately defining what is meant by 'sea surface', this type of code incorporates the z co-ordinate into the parameter description. This is incompatible with the ISO model for geo-referenced data and such codes should be avoided.

The largest task undertaken by EnParDis was the mapping of the BODC dictionary to the parameter information held by the Rijkswaterstaat databases conforming to the DONAR [60] data model. 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. As such, three semantic models covering the BODC Parameter Dictionary have been developed and successfully used to complete the mapping for all chemical and biological parameters. Semantic models break the description of each parameter code into atomic items of information that are populated from controlled vocabularies. The dictionary then becomes a registry of valid combinations of these semantic elements. Mapping becomes a two-stage process. The first stage is a mapping between the semantic elements of the two models. The second stage is a mapping of the vocabularies. This normalizes the mapping process, cutting the number of comparisons required by orders of magnitude. Furthermore, the process may be successfully automated. After population issues had been addressed, 90% of the DONAR descriptions were mapped by a simple SQL macro.

Ideally, a single model should describe the dictionary. However, the one-model fits all approach requires such a large semantic element superset that it is too cumbersome for large-scale development work against an existing dictionary. Consequently, a series of smaller semantic models are being used. This effort is on going with the ultimate objective of creating a single model.

The GCMD mapping provided a new set of problems related to granularity. The NASA GCMD DIF format is supported by a parameter vocabulary known as the 'parameter valids'. These are significantly different from the basic codes contained in the BODC Parameter Dictionary. The BODC codes are designed for data mark up. The GCMD parameter valids are designed for data discovery. The important difference between these two types of code is that the latter may describe a group of measurements (e.g., ocean currents), whereas the former may only describe a single measurement (e.g., horizontal current speed). However, the BODC dictionary also includes a code classification (BODC Parameter Groups) that may be considered equivalent to the GCMD parameter valids and a mapping was attempted between them. This revealed a serious problem of granularity incompatibility. As a consequence, the BODC Parameter Groups were totally redefined reducing them from over 2000 to fewer than 300. The mapping of the parameter valids to the revised groups remains to be done.

Mapping to the PANGAEA dictionary [61] was considered, but rejected as too difficult. This is because it contains over 25000 entries with only a numeric code and a plain language description, with no semantic atomization. The manual mapping required would have taken years and consequently this effort was abandoned.

EnParDis also undertook to standardise the taxonomic entities in the BODC dictionary to the Integrated Taxonomic Information System (ITIS) [62] by incorporating ITIS codes into the parameter descriptions. The main problem encountered with the ITIS mapping was that not all 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 has been developed to exploit the ITIS taxonomy to provide a taxonomic grouping of BODC codes. This will be made available through the BODC web site once extensive intranet testing has been completed.

Swedish Meteorological and Hydrological Institute (SMHI) [63] codes have also been mapped to the BODC dictionary. However, some SMHI categories (e.g., ice, humus) are not present in the BODC dictionary. This will be addressed as part of the ongoing BODC dictionary development work.

Canadian efforts mapped parameter codes from three labs: BIO, IOS and MEDS. This mapping extended the exercise to include units and conversions [56]. The Canadian codes being used at the three labs are also available on-line [64]. Efforts related to other projects are currently underway to incorporate or link other dictionaries. This will provide users with the ability to search and identify existing codes, rather than creating new codes.

The issue of unit conversions is also a recognized problem. The Canadian effort offered one conversion method by embedding the conversion in the XML document containing the codes. In the more general case, such conversion factors may be better represented in a separate XML file. However, in oceanographic data some conversions are complicated by the use of water density. Backward conversions are only possible if the density is part of the data stream. However, it was suggested that different users would tolerate different levels of conversion accuracy. For some users, density assumptions may be used for backward conversions. It was noted that care is required when constructing unit conversion systems to ensure that the proper number of significant digits is maintained.

The issue of parameter mappings by systems retrieving data from federated databases was considered. The problem is that the term used to describe a parameter by the portal must be mapped to terms used by databases in the federation. The solution proposed by the Russian system [53, 54] is to underpin the portal by a Universal Parameter Dictionary (UPD) containing the parameter terms available through the user interface. Each term in the UPD is mapped to one or more local database terms through a mapping maintained by the local database management.

Whilst this approach may work for some users for some of the time, it will eventually run into problems, particularly with non-physical parameters. It is probable that different users will require different mappings between UPD terms and local terms. Consider the term 'chlorophyll'. Some users may not be concerned whether this includes chlorophyll-a and chlorophyll-b, but others will be. Ideally, the mapping between UPD and local terms should be under user control.

5.2 A Code Mapping Schema

The importance of parameter dictionaries provided the incentive for the SGXML to explore the representation of these dictionaries in an XML structure. In the first SGXML meeting in 2002 [19], the Group attempted to define an XML structure that was capable of representing groups of codes from multiple dictionaries as categories. Essentially, the Group was modelling an XML structure after a common language dictionary. In this sense, a single term (or category) can have many definitions (local codes).

A schema for the code mapping was drafted at the initial SGXML meeting in 2002. The schema has been revised over the life of SGXML to conclude with the structure shown in Figure 1 (schema provided in Annex 2). The schema evolution was conducted at several labs. In particular, efforts at BODC, FIMR and DRDC / MEDS extended the schema in different ways to accommodate different needs. The reconciliation of the schema evolution has been documented.

- dictionary
  - dictionary_owner [1]
  - dictionary_citation [1]
  - dictionary_description [1]
  - date_structure [0,1]
  - dictionary_entry [1,n]
    - dictionary_term [1]
    - role [1]
    - definition [1,n]
      - instance [1]
      - definition_owner [1]
      - short_name [1]
      - creation_date [0,1]
      - change_date [0,1]
      - methelodgy [0,1]
      - unit_of_measure [0,1]
      - min_value [0,1]
      - max_value [0,1]
      - null_representation [0,1]
      - accuracy [1]
      - authority_citation [1]
      - codeset [0,n]
        - codeset_name [1]
        - code [1]
        - codeset_owner [1]
        - multiplier {pt_link} [0,n]
    - synonym [0,n]
      - synonym_instance [1]
      - synonym_term [1]
      - synonym_owner [1]

Figure 1. The code mapping structure. The heirarchy is shown as a series of indented columns. For example, the <dictionary> element contains five subelements that specify the dictionary_owner, dictionary_citation, etc. Revisions have been made since the initial definition.

The Canadian revision was to address unit manipulations. The schema was extended to include a <multiplier> element and the attribute pt_link [56]. The <multiplier> extension provided the ability to include a multiplication factor to convert one unit to another. A similar element could be added for an offset. The pt_link attribute in <mulitplier> allows the linkage between the <multiplier> element and another <codeset_owner>. In this way, the units may be converted during the code conversion. This was also demonstrated using the profile case study in Section 6 and eXtensible Stylesheet Language Transformations (XSLT) [56].

The BODC and FIMR revisions to the schema structure addressed slight corrections to the initial schema. These corrections included element reordering to address particular language requirements. For example, the <synonym> element represents a language synonym for a particular definition. Thus, the <synonym> element was moved to be within the <definition> element in the structure. Other revisions included a more detailed accounting of dictionary owner information through the use of elements such as <dictionary_owner> and <dictionary_citation>.

However, there are many issues surrounding the implementation of an XML-based process that maps codes. First, there is the issue of maintenance. After the XML document is loaded with a few dictionaries, the maintenance of the document becomes an issue. The document is not intended to be the parameter dictionary system but rather a representation of the dictionaries. Thus, dictionary maintenance would occur in a different system, which would then be responsible for the generation of all or part of the XML mapping document.

A second issue is related to the mapping itself. The initial idea of the schema was that each definition entry element would constitute a universal term (e.g., chlorophyll). The definition would then contain the many local dictionary codes that correspond to that term. In effect, this is using the definition element as a category for the specific codes.

Such categorizations may be useful to users wishing to identify data based on the large grouping of the category. However, the SGXML categorization within the code schema was not necessarily under user control. The ability to also represent user-defined categorizations would be beneficial because it provides a mechanism for the user to form groups natural to their particular investigations.

Also important for the schema is the code set content. A particular problem in this regard is statistical parameters. A typical parameter dictionary for an oceanographic lab contains information on individual measurement values (e.g., temperature, salinity, current speed). However, calculations can represent these values as means, standard deviations, or other such manipulated forms. There is a requirement to represent these statistical values in any developed structure.

It is uncommon for statistical parameters to be included as discrete entities in a parameter dictionary. However, this omission may be easily addressed in dictionaries covered by semantic models by simply adding a ‘statistic’ semantic element to the model. An agreement to do this for the BODC Parameter Dictionary semantic model to cover the requirements of the Russian NODC resulted from a bilateral BODC/RNODC SGXML follow-up meeting in Obninsk in September 2004. Alternatively, the statistical descriptor may be an additional attribute of the value. This is also addressed in Section 6.1.1.

5.3 Parameter Dictionary Achievement Summary

The parameter dictionary subgroup identified four major achievements of the SGXML in relation to parameter dictionaries:

- An XML schema has been developed to map entries from multiple dictionaries to common terms. The schema has been used to support unit inter-conversion 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 and the reduction in the parameter groupings from over 2000 to fewer than 300.
- 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, and
    - on-line access to dictionary instigated.

6. Case Studies

The case studies initially described in the TOR, identified three topics: point data, metadata and biological data. The point data study was to address the physical/chemical oceanographic data types. Metadata was considered a separate topic for study because of its importance in cataloguing available data sets and also because of its anticipated importance in the discovery of data sets. Finally, the biological component was included in an attempt to get the membership thinking about the unique challenges associated with biological data.

The physical/chemical and biological data will be addressed in this section. The metadata issue was not addressed in terms of placing metadata into the generic XML structure as this has been adequately addressed through initiatives developed elsewhere.

6.1 Profile Data – A Structure Based on Keeley Bricks

The data investigation component of the SGXML concentrated efforts on developing generic structures [57] for use in a variety of ocean data types. The initial concept for the generic structures was based on the work of J. Robert Keeley (MEDS) in the 1980s. The initial idea recognized that many data types being delivered to the data centre contained information parts that were consistent across the data types. It was thought that these consistent parts could be formalized into structures, or Bricks. The formal Bricks could then be arranged in multiple ways to address the many structures present in the various ocean data types.

The SGXML wanted to exploit the Keeley concept by the further development of the Bricks. Fortunately for the SGXML, a Canadian-lead interdepartmental investigation was funded by the Canadian Department of Fisheries and Oceans (DFO) under the Science Strategic Funds (SSF) program to develop and apply the Bricks to oceanographic profile data [57].

The Canadian effort fully developed all the Keeley Bricks to address typical oceanographic profile data. Here, profile data is considered one-dimensional data, where one of the four coordinates (e.g., x,y,z,t) may be considered an independent variable. For example, when z is the one independent variable, you have the common depth profile (e.g., XBT profile, CTD profile). When t is the one independent variable, you have a time series (e.g., current meter time series, wind speed time series, etc.).

The Bricks and sub components were developed with full definitions. It is important to note that the Brick concept is independent of the implementation environment. The Bricks, once defined, were then applied to the XML environment. The Bricks and sub components are very well suited to the XML environment, resulting in a smooth application to XML.

The Brick development is fully documented in [57], including many of the critical decisions made during the development process. Only small parts of the document will be highlighted in the following subsections. Attempts are also being made to distribute the Brick and XML implementation to a wider audience [65].

6.1.1 The Defined Structure

The structure developed from the Bricks [57] is shown in Figure 2. The structure utilizes the concept of a repeating and hierarchical <data_set> element. The oceanographic community is familiar with describing groups of data as a data set; however, the exact definition is often ill-defined. In this development, it was thought that the data set definition needed to be formalized to then become part of the data structure.

The data set definition was built around the premise that a data set had one very important feature: it could contain other data sets. As well, the data set must have an identifier and contain data and supporting information. A formal definition evolved that identified a data set as containing the following information [57]:

- a unique identifier either by name or number,
- a history of processing, including processing related to quality testing and results of this testing,
- a definition of the level of availability for the data set,
- parameters or variables,
- data points pertaining to these parameters or variables,
- identification of the data set owner, and
- other data sets.

The resulting structure (Figure 2) shows the importance of the <data_set> element by its repeated use within the structure. The level of the <data_set> element within the structure is described by the <data_set_id> element. The full schema for the structure is given in Annex 3.

Figure 2 shows a collection of boxes that represent the expansion of particular elements within the structure. The left top-most box shows the <data_collection> element, which in turn contains five subelements as indicated by the bulleted items. The occurance of each subelement is indicated by the minimum and maximum values contained in the square brackets. Note that one of the subelements is <data_set>.

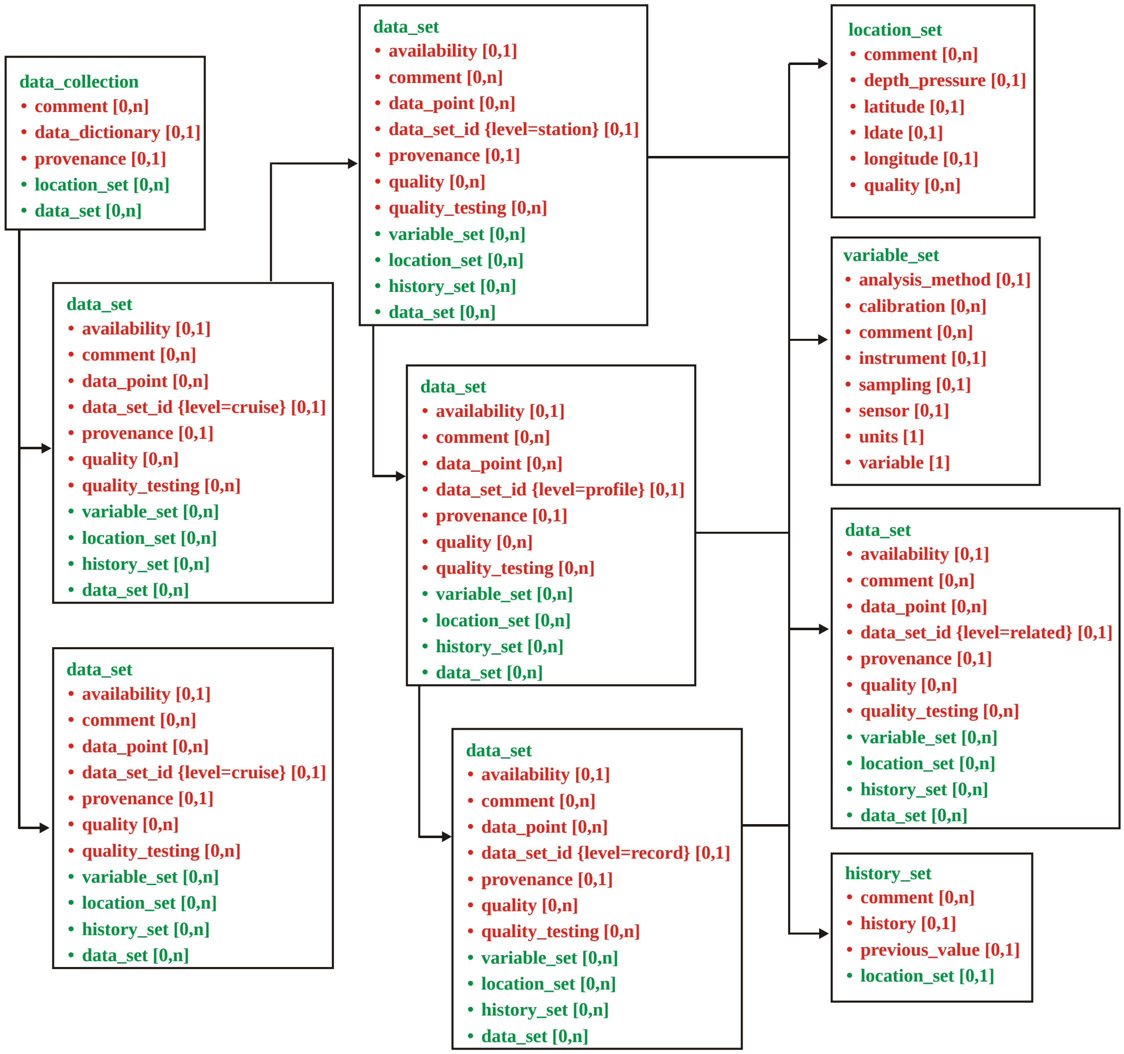

Figure 2. The structure defined for 1-dimensional profile data based on Keeley Bricks. The red text indicates "pure" bricks while the green text indicates "compound" bricks. Definitions of pure and compound bricks are provided in [57].

The <data_set> element is then expanded in the two lower boxes. These two expansions of <data_set> indicate that multiple cruises can be described within any one <data_collection>. The cruise level for a data set is indicated by the {level="cruise"} attribute in the <data_set_id> element.

Each <data_set> element can contain other <data_set> elements defined as {level="station"}, {level="profile"} and {level="record"}. These subsequent levels are described by the three boxes that make up the central column in Figure 2. The <variable_set>, <location_set> and <history_set> are described by the boxes in the right most column of the figure. The {level=”related”} contains ancillary data that may be collected.

As mentioned in section 5.2, statistical values may be identified using attributes of the value. In this profile data development, each <data_point> element contains an attribute “statistic” that indicates the type of statistical value contained in the <data_point> element.

6.1.2 Codes and Links

The issue of oceanographic parameter codes within any XML structure will generate much debate. There are essentially three ways to deal with the codes in an XML environment. The code could be:

- content for an XML element,
- content for an XML attribute, or
- the actual XML element tag name.

The application of the Keeley Bricks to an XML environment considered all three potential placements for the parameter codes. The result was to position the code as attribute content in the <data_point> element. For example,

<data_point pt_code=”TEMP”>3.1</data_point>

The <data_point> element would contain the value of the parameter that was indicated by the code in the attribute content. In this way, the code and value are contained in a single XML element, while still providing the generic capability of the Keeley Bricks.

The linking aspect of the Keeley Bricks was based on the requirement to have links established in the XML document for variables and instruments. The linking uses an XML attribute to provide a link between codes and variable definitions and also between codes and instruments. This provides the XML document creator with the ability to use duplicate codes in a single XML document and also to specify the instrument source for the measurement.

6.1.3 The Hierarchy - Variables and Instruments

When developing a hierarchy, there will often be cases when the structure can be developed in two different ways. The common example is a group of people working on multiple projects. A hierarchy with people at the top, groups the many projects that an individual works on. Here, the project information is repeated across many people. The other option is to group projects at the top. In this case, people are repeated if they work on multiple projects. In either case, information is repeated.

There are indexing efforts underway that attempt to address this scenario [66]. However, the indexing can be rather complicated, it may overly complicate the reasons for the investigation.

In an oceanographic context, the hierarchy problem exists with variables and instruments. A single instrument may sample many variables. As well, many instruments may sample one particular variable type. When represented in a hierarchical form, either case will result in repetition of information. In the Keeley Brick application, the decision was made to place importance on the variable, and to therefore place the variable higher in the hierarchy. The instrument information was thus constructed within or under the variable information.

6.1.4 Application Testing

The Keeley Brick structure, as applied in the XML environment, was then tested using vertical profile data from three Canadian labs: IOS, MEDS and BIO. Each lab developed software to create XML documents that complied with the developed profile schema. As well, each lab developed the software to construct in-house formats from the XML documents.

The software development took place in an assortment of development languages including Fortran, Matlab and Java. The development included an extensive mapping exercise where in-house structures were mapped into and out of the XML structure.

The results indicated that the software development exercise is not difficult nor is it expensive. The difficult part is the intellectual requirements for the mappings of structures and parameter codes, which is required for complete data sharing.

6.1.5 Other Applications

It is worth noting that other case studies have also examined the application of the Keeley Brick structure to oceanographic data sets. The ICES WGMDM have investigated its application to current meter and water level data. The results [67] suggested additional Bricks be created or modified to meet the data stored in the Scottish Executive Environment and Rural Affairs Department (SEERAD) Fisheries Research Services (FRS) data system. However, it is important to note that the XML implementation of the Keeley Bricks is intended to be an exchange structure rather than a storage structure. Thus, only those details important for the exchange or intrepretation of the data set is important to the Brick structure. The in-house storage formats may require more detailed metadata.

As well, investigations extended the initial point data development to include underway temperature-salinity data, water sample data, profiling float data, and acoustic doppler current profiling data (both moored and shipboard).

6.1.6 Bricks and GML

As noted previously, the Keeley Bricks represent data structures that in the above Case Study have been applied in an XML environment. However, the Bricks may also be applied in other environments. For example, the Open Geographic Information System (GIS) Consortium (OGC) Geography Markup Language (GML) may be used as the basis for the implementation of Keeley Bricks.

The SGXML also investigated the application of the Bricks in GML [68]. The study revealed that it is difficult to place the Bricks into the GML structure. However, a more natural application may be to use GML for those parts of the Brick implementation that specifically deal with position information.

6.2 Biological Data – Net Tow

The biological investigation into application of the Keeley Brick structure was given special mention in the initial TOR because of the unique challenges provided by biological data sets. This case study concentrated on biological net tow data, because of its 3-dimensional characteristic.

A data set was identified as the case study for this investigation. The data set originated from the Flanders Marine Institute (or Vlaams Instituut voor de Zee, VLIZ) and was supplied by E. Vanden Berghe (Belgium). The data existed in a Microsoft Access database. Collectively, this database will be referred to as the tow database.

The tow database contains data familiar to many oceanographic data collection experiments. The database contains data on ships, trips made by these ships, visits to particular locations, gear used and samples collected. As an example, Figure 3 shows a single record from the ship table. The record notes the ship name (i.e., Belgica) and an identifier (i.e., 2) assigned to that name.

ID	shipName	description	note
2	Belgica	BMM	

Figure 3. An example record from “ships” table in the tow database.

The complexity associated with biological data starts to manifest itself at the sample level. In the case of net tows, the actual samples are obtained from the material collected in the net. These samples may then be subsampled into smaller units. As well, a single subsample may be analysed by many individuals, yielding different data values from the single sub sample. As well, the analysts may be examining the subsample for the same species or genus. These species may be counted or identified in some way (e.g., by growth stage). Different analysts may also be referring to different reference materiel to perform the identification. All of this information needs to be tracked within the XML structure.

The biological investigation placed the net tow data contained within the database, in the generic XML structure developed previously [57] and shown in Figure 2. The XML document that illustrates the data placement is provided in Annex 4.

The investigation showed that the complicated relationships in the biological data could be addressed in the generic XML structure. However, the multitude of relationships within the biological data means that there are many possible ways to generate the XML document. As well, the flexibility of the XML-based Keeley Brick structure allows for the many different structure possibilities.

These multiple structures are similar to the project-people example provided in Section 6.1.3. For this case study, we provide a specific example of this in the following figures.

Figure 4 shows the data records from two specific tables in the tow database. In the upper table structure, taxonID 6817 indicates genus Crangon (as shown in the records of the second table). It is noted that 15 post larva stage Crangon were found in sample 29, as were 46 Zoea stage Crangon.

In the XML Brick structure, this data can be represented in many different ways. As an example, consider a snippet of an XML document shown in Figure 5. Here we show some of the data described in Figure 4, specifically the Crangon post larva 15 counts and Crangon Zoea 46 counts. These values are indicated by the <data_point> element that contains the “pt_code=number” attribute. The XML also shows pt_links of “1” and “2” indicating two analysts as defined in another section of the XML document (not shown). Also, the record value of “29” indicates the sampleID number while the depth is given as 14.141 (from tables not shown in this document).

This information is acceptable XML structure meeting the validation requirements of the point data investigation (assuming a complete document is included in the validation). However, the record number and depth values are repeated for each <data_set> element.

ID	sampleID	taxonID	StadiumCode	GenderCode	Number
640	29	14666	Medusa, hydro-	no information	4
641	29	1302	unknown	no information	2
642	29	6818	Adult	no information	1
643	29	6817	Postlarva	no information	15
644	29	6817	Zoea	no information	46
645	29	5419	Postlarva	no information	1
646	29	1162	Postlarva	no information	2
647	29	2614	Juvenile	no information	1
648	29	2661	Adult	no information	1

aphiaID	taxonName	note
6817	Crangon	
6818	Crangon crangon	

Figure 4. Example records from tables “records” and “taxa” (bottom) in the tow database.

An alternate representation of the data is also possible in the Keeley Brick XML as shown in Figure 6. This XML snippet shows the promotion of the record number and depth elements to a higher level in the XML hierarchy. This XML document (shown in complete form in Annex 4) also validates against the Keeley Brick schema. Obviously, this structure is more compact than the structure shown in Figure 5.

This issue was previously described (see [57] Section 6.6.2) as the difference between optimization and compliance. Both XML snippets are in compliance with the validation of the XML document against the schema (as shown in Annex 2). However, the XML shown in Figure 6 is optimized. It was stated in [57] that “*every effort should be made to optimize the content within the structure*” although no specific procedure was given to meet this requirement.

The issue of compliance and optimization is similar to the normalization process in database design. During normalization, particular attribute placment within entities seeks to minimize data redundancies. Within the XML structure, this is similar to the movement of an element to a location higher in the XML heirarchy. By moving the element higher, we reduce redundant data within the XML document.

```
<data_set>
  <data_set>
    <data_point pt_code="Genus" pt_link="1">Crangon</data_point>
    <data_point pt_code="Genus" pt_link="2">Crangon</data_point>
    <data_point pt_code="Stage">Postlarva</data_point>
    <data_point pt_code="number">15</data_point>
    <data_point pt_code="biomass">99</data_point>
    <data_set_id level="record">29</data_set_id>
    <location_set>
      <depth_pressure pt_code="DEPT">14.141</depth_pressure>
    </location_set>
  </data_set>
  <data_set>
    <data_point pt_code="Genus" pt_link="1">Crangon</data_point>
    <data_point pt_code="Stage">Zoea</data_point>
    <data_point pt_code="number">46</data_point>
    <data_point pt_code="biomass">99</data_point>
    <data_set_id level="record">29</data_set_id>
    <location_set>
      <depth_pressure pt_code="DEPT">14.141</depth_pressure>
    </location_set>
  </data_set>
</data_set>
```

Figure 5. An example XML snippet containing biological data from the net tow database. The data shows genus Crangon identified by two analysts, as indicated by two pt_link attributes. The growth stage of the genus, the number of identified organisms and the sample record number are also indicated.

The flexibility of the XML structure allows one to describe the biological relationships that are present within the tow database. During early development, the XML structure was considered applicable to any 1-dimensional data [57]. Subsequent applications have shown that the structure is applicable to 3-dimensional data. In all likelyhood, the structure could successfully be applied to four and 2-dimensional data as well.

This flexibility means the structure could be applied in ways that meet the detailed requireemnts of the data set. The alternative would be a more rigid structure, possibly resulting in a single structure for each of the dimensional data categories. These restrictions are certainy possible with the current Keeley Brick approach. Applying such restrictions may result in each dimension having an individual schema. However, this may also clarify the positioning of the elements within the document.

```
<data_set>
  <data_set_id level="record">29</data_set_id>
  <location_set>
    <depth_pressure pt_code="DEPT">14.141</depth_pressure>
  </location_set>
  <data_set>
    <data_point pt_code="Genus" pt_link="1">Crangon</data_point>
    <data_point pt_code="Genus" pt_link="2">Crangon</data_point>
    <data_point pt_code="Stage">Postlarva</data_point>
    <data_point pt_code="number">15</data_point>
    <data_point pt_code="biomass">99</data_point>
    <data_point pt_code="NPOS" pt_link="3">top</data_point>
  </data_set>
  <data_set>
    <data_point pt_code="Genus" pt_link="1">Crangon</data_point>
    <data_point pt_code="Stage">Zoea</data_point>
    <data_point pt_code="number">46</data_point>
    <data_point pt_code="biomass">99</data_point>
    <data_point pt_code="NPOS" pt_link="4">bottom</data_point>
  </data_set>
</data_set>
```

Figure 6. An example XML snippet that is optimized as compared to Figure 5. Here, the record and location information has been moved higher in the heirarchy. The complete document is provided in Annex 4.

6.3 Tokyo Bay Environmental Project

The SGXML also assisted projects conducted by members. During the period of SGXML work, the Japanese SGXML participants utilized the ideas and methods discussed and investigated during the SGXML meetings, applying these ideas to the Tokyo Bay Environmental Information Center Project (TBEIC Project) [69, 70].

The TBEIC Project was initiated to maximize usage of environmental data collected in support of monitoring Tokyo Bay. Tokyo Bay and the surrounding watershed support a human population of approximately 12 million. Such a population base places enormous strain on the Bay, resulting in bio-chemical issues (e.g., eutrophication, red tide occurrence, hypoxic conditions, etc.) and physical issues (e.g., change of flow pattern, floating garbage, etc.).

The Japanese Oceanographic Data Center (JODC) corrects and distributes marine data in support of Japanese oceanographic data collection. However, the integration of numerous data sets was a requirement towards realizing a more cohesive coastal zone monitoring approach to Tokyo Bay. The Tokyo Bay Environmental Information Center was created to support this project.

The TBEIC was created to act as a clearinghouse for Tokyo Bay environmental data. Efforts would be focused on data sharing, data standardization and the construction of the clearinghouse, which itself would support the searching of the data holdings. The data structure needed to support data for water quality, bottom sediment types and bottom quality, biological measurements, meteorological, and oceanographic, as well as any supporting metadata.

The data standardization for the project concentrated on both metadata and data. The Unified Modelling Language (UML) was utilized to construct structures that support the metadata and data. The ISO 19115 and GML structures were also heavily utilized to maintain consistency with ongoing work at the Japan Geographical Survey Institute (JGSI).

The resulting metadata structure detailed much information about the collected data set. The more typical information that details the cruise that collected the data was included in the metadata, but also included were references to papers that utilize the data set. The subject of the data collection exercise and points of contact were also included. Geospatial extents, transfer information and distribution formats were also described.

The TBEIC developed data structure also utilized the GML effort. The developed structure resulted in XML groups that describe information about the observed data (e.g., organization, dictionary, time and location, data values) and explanation information about the data (e.g., units, methods, instruments, calibrations).

The resulting XML grouping of these data were similar in structure to the Keeley Bricks. As examples, one may consider the TBEIC <value> element (see Figure 7) as compared to the Keeley Brick <data_point>. Similarly, one may compare the <instrument> elements from the two efforts.

However, an important difference exists between the Keeley Brick approach and the TBEIC approach. The TBEIC effort made use of considerably more linking within the XML document. For example, Figure 7 shows water temperature described as “item001” with a unitId of “degC”. The degC unit is then described by the <gml:unitDefinition> element. The data value is contained in the element <value>, with the attribute itemId being “item001” indicating that it is a water temperature.

Instruments and methods are described in a similar manner, but are not shown here. However, the full XML document is provided. In this structure the attributes provide the linking mechanism from the value back to the described variable, instrument, method and unit. This method of linking removes much of the hierarchical structure that is present in the Brick structure. The Brick structure utilized the XML hierarchy to capture the intent of the linking.

```
<dictionary>
  <locationList>
    <gml:Point gml:id="loc0001">
      <gml:name>St.1</gml:name>
      <gml:pos>139.9194 35.6361</gml:pos>
    </gml:Point>
  </locationList>
  <itemList>
    <item itemId="item001" unitId="degC" instrumentId="ins0001">
      <name>water temperature</name>
    </item>
  </itemList>
  <unitList>
    <gml:UnitDefinition gml:id="degC">
      <gml:quantityType>Celsius temperature</gml:quantityType>
    </gml:UnitDefinition>
  </unitList>
</dictionary>
<observationLocation locationId="loc0001">
  <time>
    <gml:TimePeriod>
      <gml:begin>
        <gml:TimeInstant>
          <gml:timePosition>2002-08-21T11:28</gml:timePosition>
        </gml:TimeInstant>
      </gml:begin>
    </gml:TimePeriod>
    <valueSet observationId="waterQuality">
      <depthInstant>
        <depthPosition>0.5</depthPosition>
      </depthInstant>
      <totalDepth>
        <depthPosition>6.5</depthPosition>
      </totalDepth>
      <value itemId="item001">20.7</value>
    </valueSet>
  </time>
</observationLocation>
```

Figure 7. An example XML snippet from the TBEIC. The snippet illustrates the linking between <itemlist>, <unitlist> and data as contained in <valueSet>. This is a partial document for illustrative purposes and does not validate. The coloured text is for illustrative purposes and merely indicates the internal relationships between itemId and unitId.

Tool development is an important concept for TBEIC Project. TBEIC has developed tools for the conversion of data contained in a spreadsheet to the developed XML structure. These tools create the necessary internal links, with minimal user awareness of the structure details. Furthermore, such automation ensures the strict adherence to the defined XML structure. Since the initial data types are somewhat constrained by the

Project, the developed structure does not require the flexibility of the XML Brick implementation.

6.4 Case Studies – Major Achievements

The point data subgroup identified three major achievements of the SGXML in relation to data case studies:

- 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 particular software developement associated with the investigation indicated that the software development exercise is not difficult nor is it expensive. The major difficulty was recognized as the mappings of structures and parameter codes, which is required for complete data sharing.

- The SGXML also investigated the application of the Bricks in GML. The study revealed that it is difficult to place the Bricks into the GML structure. However, a more natural application may be to use GML for parts of the Brick implementation that specifically deal with location and position information.

- The SGXML have assisted and influenced the local implementation of the software and schema developments for the Tokyo Bay Environmental Information Center Project. As well, the TBEIC Project provided SGXML with alternate implementation ideas and an actual application example.

7. Recommendations

The previous sections have highlighted the activities and major accomplishments of the SGXML. However, there remains considerable work to fully take advantage of the SGXML efforts. As such, the SGXML wish to make the following recommendations. We hope the various groups identified within each recommendation and in particular IODE18 will consider these recommendations. The recommendations are grouped according to the main topics discussed previously.

7.1 Metadata Recommendations

On the topic of ocean metadata, the SGXML recommend:

1. That a mapping and whenever possible a consolidation of metadata terminology takes place between the Russian metadata model and the NDG model.

   Direct To: ETDMP, NDG, BODC, Russian NODC and MMI

   Justification: The terminology being used in the metadata community is beginning to cause confusion. In a networked environment, there is a requirement for metadata types to support the entire system. Detailed definitions, clear mapping and standardised terminology for common elements of metadata types must be rationalized to avoid potential branching of terms.

2. That definitions be created for the explicit elements representing the oceanographic extensions to ISO 19115.

   Direct To: ICES MDM

   Justification: The ISO 19115 metadata standard holds considerable promise for meeting the needs of the international ocean data community. However, community based extensions are required to address the unique aspects of the community. These extensions should be developed and made available to other ocean programmes.

3. That harvester infrastructure be designed and created, for combining metadata from distributed repositories into an ocean metadata clearinghouse.

   Direct To: Coordinated with ETDMP and IODE Group of Experts on Biological and Chemical Data Management and Exchange Practices (GEBICH).

   Justification: Ocean data centres and labs will be responsible for placing their data assets on the web. However, creating a single coherent catalogue of the total holdings should be the function of an international body. The creation of harvester infrastructure to combine and create the catalogue will move the ocean data community toward an integrated system, where all assets remain managed by the data centres, but are accessible from central locations. This task may also involve a comparison of capabilities among the different systems (e.g., OAI, DiGIR).

7.2 Parameter Dictionary Recommendations

On the topic of parameter dictionaries, the SGXML recommend:

4. That the BODC parameter dictionary be adopted as the marine ocean community standard, including the use of the BODC dictionary in any developed marine XML.

   Direct To: JCOMM and IODE

   Justification: In the process of creating and testing a single XML document structure for marine data, the SGXML recognised the importance of a consistent parameter dictionary between the data provider and receiver. The SGXML made progress toward developing an XSLT structure that allowed the mapping between parameter dictionaries and this mapping was applied to several case studies. However, there remains the need for a central, or common, international dictionary. The most extensive ocean parameter dictionary in existence has been developed by BODC. The BODC dictionary should be adopted as the defacto international standard. As well, the BODC dictionary should be promoted and supported by international organizations and programs.

5. That the BODC dictionary be implemented as a register within the proposed IOC registry.

   Direct To: IOC

   Justification: The IOC has been discussing the establishment of an IOC registry, for the formal approval and registration of standards and specifications that address the needs of the international ocean community. The SGXML support this concept, and if created, would recommend that the BODC parameter dictionary be considered for addition to the registry.

6. That an improved mechanism be established to control the evolution of the dictionary (e.g., a review college), including extension of the dictionary population.

   Direct To: Coordinated between BODC and IOC

   Justification: At present, individuals control the evolution, maintenance and revisions to the BODC parameter dictionary. If the dictionary is to become the de facto international standard, then a management group needs to be established to provide a formal governance framework for the evolution of the dictionary. The established Group would need to meet regularly, establish a mechanism to deal with requests for changes, address user questions or concerns in a timely manner, and actively encourage and promote the use of the dictionary.

7. That improved web access be developed for the BODC dictionary.

   Direct To: BODC

   Justification: User support services for the BODC dictionary need to be improved, to allow users more efficient access to up-to-date dictionary entries. The implemented services and exact methodology needs to be defined and constructed. Coordination with users would help ensure a full range of services is developed.

8. That a semi-automated mechanism for dictionary extension be developed.

   Direct To: BODC

   Justification: The evolution of any dictionary is a critical aspect of the continued use of the dictionary. Part of this evolution is the continual addition of entries as new parameters are measured. Similar dictionaries in other domains, such as the Climate and Forecast (CF) Metadata Convention for NetCDF [71] standard name list, have been unable to respond to requests for extensions within an acceptable timescale. Introducing automation to the process is the only realistic way to overcome the problem. Placing automated tools on users' desks would also allow them to become part of the development thereby gaining a vested interest in its continuation.

9. That a steering group be created to oversee interoperability standards for marine data.

   Directed To: IODE

   Justification: A formal governance framework needs to be established to oversee interoperability standards for marine data. Recommendation six deals with the governace of the dictionary domain, while this recommendation extends beyond the dictionary to include all aspects of interoperability standards development and deployment across marine sciences. Examples of projects that contribute to components of this are presently underway. For example, the MBARI Marine Metadata Interoperability (MMI) project is attempting to deal with metadata standards.

7.3 Case Study or Data Recommendations

On the topic of structures that support data transfer, the SGXML recommend:

10. That the Ocean Biogeographic Information System (OBIS) be examined and evaluated for potential use for XML-based data exchange of biological data.

    Directed To: GEBICH

    Justification: The tow data examined as part of this investigation provided an example of the numerous relationships present in biological data. Although the SGXML did not have sufficient time to evaluate the OBIS, this system may be useful for the distribution of such data. OBIS management may also be interested in such an investigation, especially if conduced by former SGXML members. Such an investigation may provide useful insights if OBIS were to consider moving toward an XML based data exchange structure.

11. That an effort 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.

    Directed To: Canadian and Japanese Development Teams

    Justification: The Canadian and Japanese developments made in conjunction with the SGXML, have many XML structures in common. The Japanese development also utilizes GML structures. The Canadians have investigated porting the Keeley Bricks to GML, but did not investigate the partial use of GML. The Japanese also have developed field tools that support the structure. These tools are particularly useful for private industry collecting marine data. There needs to be an effort to consolidate the Canadian and Japanese structures, to create a unified near-shore and ocean XML structure.

12. That a demonstration project be initiated to use the single schema developed in recommendation 11, to demonstrate the XML structure using a variety of data types and developed tools.

    Direct To: ICES MDM

    Justification: Any developed structure needs to be tested in case studies involving data familiar to the ocean data community. The Japanese development is operational in the Tokyo Bay Project and therefore meets the needs of that particular project. However, the needs of the international data centres must be considered and any single Marine XML structure should also accommodate the data centre requirements. The ICES MDM are appropriately linked to marine data centres to provide a valuable input on the use of the structure in case studies.

8. References

1. G8 - Sommet Evian Summit 2003. Science and Technology for Sustainable Development - A G8 Action Plan. See http://www.g8.fr/evian/english/ (February 2005).

2. The International Council for the Exploration of the Sea. 2005. See http://www.ices.dk/ (February 2005).

3. UNESCO. 2001. IODE Steering Group for Underway Sea Surface Salinity Data Pilot Project. First Session Brest, France, 15-16 November 2001.

4. Intergovernmental Oceanographic Commission. 2005. IOC. See http://ioc.unesco.org/iocweb/index.php (February 2005).

5. United Nations Educational, Scientific and Cultural Organization (UNESCO). 2005. UNESCO.ORG. See http://www.unesco.org/ (February 2005).

6. IODE Responsible National Oceanographic Data Centres. 2005. RNODCs. See http://ioc.unesco.org/oceanteacher/resourcekit/Module1/DataCentre/rnodcs.htm (February 2005).

7. Marine Environmental Data Service. 2005. Marine Environmental Data Service (MEDS) - Le Service des données sur le milieu marin (SDMM). See http://www.meds-sdmm.dfo-mpo.gc.ca (February 2005).

8. The International Council for the Exploration of the Sea. 2005. RNODC (Formats) - GF3 Formatting System Pre-defined Subsets. See http://octopus.ices.dk/ocean/formats/gf3.htm (February 2005).

9. The International Council for the Exploration of the Sea. 1999. Report of the Working Group On Marine Data Management. MEDS, Ottawa, Canada. 3-6 May 1999.

10. The International Council for the Exploration of the Sea. 2000. Report of the Working Group On Marine Data Management. Hamburg, Germany 10–13 April 2000.

11. IOC/IODE GETADE. 2005. IODE Expert Info Expertise GETADE. See http://ioc.unesco.org/iode/categories.php?category_no=21 (February 2005).

12. Intergovernmental Oceanographic Commission of UNESCO (IOC). March 2000. IODE Group of Experts on Technical Aspects of Data Exchange. Eighth Session. Greenbelt, MD, USA, 13-17 March 2000. Intergovernmental Oceanographic Commission, Reports of Meetings of Experts and Equivalent Bodies, 17 pp + annexes.

13. Reed, G. IOC/IODE-TADE-VIII/13add, Paris, 3 March 2000.

14. Intergovernmental Oceanographic Commission of UNESCO (IOC), April 2002. IODE Group of Experts on Technical Aspects of Data Exchange. Ninth Session. Finnish Institute of Marine Research, Helsinki, Finland. 20-22 April 2002. Intergovernmental Oceanographic Commission, Reports of Meetings of Experts and Equivalent Bodies, 13 pp + annexes

15. JCOMM/IODE Expert Team on Data Management Practices (ETDMP) First Session. 2003. Hosted by Flanders Marine Institute Oostende, Belgium 15-18 September 2003.

16. Informal Session of the JCOMM Expert Team on Data Management Practices (ETDMP). 2002. Hosted by Research and Development Department, Scientific Technical and Cultural Affairs (OSTC), Belgium Brussels, Belgium 28 November 2002

17. The International Council for the Exploration of the Sea. 2001. Report of the Working Group On Marine Data Management. Birkenhead, United Kingdom. 2–5 April 2001.

18. ICES Annual Report 2001. Statutory Meeting Report. ISSN 0906-0596. February 2002.

19. The International Council for the Exploration of the Sea. 2002. Report of the Report of the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML. Helsinki, Finland 15–16 April 2002.

20. The International Council for the Exploration of the Sea. 2003. Report of the Report of the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML. Gothenburg, Sweden 26–27 May 2003.

21. The International Council for the Exploration of the Sea. 2004. Report of the Report of the ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML. Oostende, Belgium, 6-7 May 2004.

22. ICES Annual Report 2003. Statutory Meeting Report. ISSN 0906-0596. May 2004.

23. Chile National Oceanographic Data Centre. 2005. Centro Nacional de Datos Oceanográficos de Chile. See http://www.shoa.cl/cendoc/ (February 2005).

24. PICES Technical Committee for Data Exchange. 2005. See http://tcode.tinro.ru/# (February 2005).

25. Inter-Agency Committee on Marine Science And Technology (IACMST). National Marine Environmental Data (MED) Coordinator. Annual Report 2003 –

Summary. See http://www.oceannet.org/medag/reports/medrep2003.pdf (July 2004).

26. IOC-WMO-UNEP-ICSU Steering Committee of the Global Ocean Observing System (GOOS). Sixth Session 26 - 28 February 2003. Cape Town, South Africa. See http://unesdoc.unesco.org/images/0013/001310/131055e.pdf (July 2004).

27. Marine XML An Interoperability Framework for Global Ocean Observation Systems. 2001. Part B. 18 December 2001. 31pp.

28. The Annotated XML Specification. 2005. See http://www.xml.com/axml/testaxml.htm (July 2004).

29. Walsh, Norman. A Technical Introduction to XML. 2005. See http://www.xml.com/lpt/a/98/10/guide0.html (July 2004).

30. IOC (UNESCO). 1987. GF3: A general formatting system for geo-referenced data. IOC Manuals and Guides 17, Volume 2.

31. Executive Order 12906. Published in the April 13, 1994, edition of the Federal Register, Volume 59, Number 71, pp. 17671-17674.

32. Marine Metadata Interoperability. 2005. MMI (marinemetadata.org) - Welcome!. See http://marinemetadata.org/ (February 2005).

33. ISO. 2000. Geographic information – Metadata, ISO 19115-3.

34. FGDC, 2000. Federal Geographic Data Committee. Content Standard for Digital Geospatial Metadata Workbook Version 2.0, Federal Geographic Data Committee. Washington, D.C.

35. European Directory of Marine Environmental Data (EDMED). 2005. British Oceanographic Data Centre (BODC). See http://www.bodc.ac.uk/cgi-bin/framer?http://www.bodc.ac.uk/services/edmed/ (February 2005).

36. United Kingdom Directory of Marine Environmental Data Sets. 2005. EDMED online search, See http://www.oceannet.org/medag/catalogues/edmed/edmed_frontpage.html (February 2005).

37. Goddard Space Flight Center. 2005. What is DIF? See http://gcmd.gsfc.nasa.gov/User/difguide/whatisadif.html (February 2005).

38. Goddard Space Flight Center. 2005. Directory Interchange Format (DIF) Writer's Guide, Version 9. See http://gcmd.gsfc.nasa.gov/User/difguide/difman.html (February 2005).

39. UNESCO. 2005. MEDI User Manual. See http://ioc.unesco.org/oceanteacher/resourcekit/Module1/DataCentre/medi.html (February 2005).

40. Goddard Space Flight Center. 2005. GCMD Homepage. See http://gcmd.gsfc.nasa.gov (February 2005).

41. Library of Congress. 2005. MAchine Readable Cataloguing (MARC) record. . http://lcweb.loc.gov/marc/ (February 2005).

42. The British Library. 2005. United Kingdom MARC Implementation Manual. See http://www.bl.uk/services/bibliographic/marc/marcintro.html (February 2005).

43. Dublin Core Metadata Initiative. 2005. See www.dublincore.org (February 2005).

44. International Organization for Standardization. 2005. Information and Documentation - The Dublin Core metadata element set. 2003. ISO 15836. See also http://www.iso.org/iso/en/ISOOnline.frontpage (February 2005).

45. Isenor, Anthony W. 2003. The Importance of Metadata in System Development and IKM. DRDC Atlantic TM 2003-011. Defence R&D Canada – Atlantic.

46. Teng, Ying. 2000. Use of XML for Web-Based Query Processing of Geospatial Data, Thesis TR00_135, Master of Computer Science, University of New Brunswick.

47. National Environmental Satellite, Data, and Information Service. 2005. FGDC to DIF Mapping. See http://www.eis.noaa.gov/fgdc/fgdcdif.html (February 2005).

48. Goddard Space Flight Center. 2005. FDGC to DIF Mapping. See http://gcmd.nasa.gov/Aboutus/standards/fgdc_to_dif_7.0.html (February 2005).

49. Library of Congress. 2005. MARC 21 to Dublin Core Mapping. See http://www.loc.gov/marc/dccross.html (February 2005).

50. Goddard Space Flight Center. 2005. DIF to Dublin Core Mapping. See http://gcmd.nasa.gov/Aboutus/standards/dublin_to_dif_8.0.html (February 2005).

51. Goddard Space Flight Center. 2005. DIF to ISO 19115 Mapping. See http://gcmd.nasa.gov/Aboutus/standards/difiso.html (February 2005).

52. Federal Geographic Data Committee. 2005. FGDC/ISO 19115 Harmonization. See http://www.fgdc.gov/metadata/whatsnew/fgdciso.html (February 2005).

53. Mikhailov, Nicolay and Sergey Belov. 2004. The Reference Distributed Data Model and Testing the Distributed Information Resource Systsem Construction. Presented at the Third meeting of the ICES/IOC SGXML, 6-7 May 2004.

54. Mikhailov, Nicolay, Evgeny Vyazilov and Sergey Belov. 2004. The Model of the Distributed Marine Information Resources – Approaches and Decisions. Presented at the Third meeting of the ICES/IOC SGXML, 6-7 May 2004.

55. The NERC DataGrid Phase I Description. 2004. The NDG Team, February 2004. Document Number 1.0 (BNL 04/02/04).

56. Isenor, Anthony W. 2003. XML Based Manipulation of Codes Exchanged Between Data Systems, DRDC Atlantic TM 2003-132, Defence R&D Canada – Atlantic.

57. 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.

58. World Wide Web Consortium. 2005. RDF Thesaurus (Thesaurus server model proposed for BODC On-line Dictionary). See http://www.w3.org/2001/sw/Europe/reports/thes/rdfthes.html (February 2005).

59. British Oceanographic Data Centre. 2005. On-line BODC Parameter Dictionary. See http://www.bodc.ac.uk/cgi-bin/framer?http://www.bodc.ac.uk/documents/bodc_params_download.html (February 2005).

60. The Netherlands Ministry of Transport, Public Works and Water Management. 2005. DONAR Data Model. See http://www.donarweb.nl (February 2005).

61. PANGAEA - Publishing Network for Geoscientific & Environmental Data. 2005. See http://www.pangaea.de (February 2005).

62. Integrated Taxonomic Information System. 2005. See Canadian site at http://www.cbif.gc.ca/pls/itisca/taxaget?p_ifx=cbif or US site at http://www.itis.usda.gov/ (January 2005).

63. Swedish Meteorological and Hydrological Institute. 2005. SMHI. See http://www.smhi.se/en/index.htm (January 2005).

64. Marine Environmental Data Service. 2005. Canadian on-line codes. See http://www.meds-sdmm.dfo-mpo.gc.ca/meds/About_MEDS/standards/ (October 2004).

65. Isenor, Anthony W. and J. Robert Keeley. 2004. Using Generic Data Objects in XML: An Oceanographic Application, submitted to IEEE Computing in Science and Engineering.

66. eXchangeable Faceted Metadata Language. See http://xfml.org/ (August 2004).

67. The International Council for the Exploration of the Sea. 2004. Report of the Working Group On Marine Data Management. Brussels, Belgium 3-5 May, 2004.

68. Isenor, Anthony W. 2004. Describing Generic Ocean Environmental Data Objects Using the Geography Markup Language, DRDC Atlantic TM 2004-087, Defence R&D Canada – Atlantic.

69. Tokyo Bay Environmental Information Center Project. 2004. See http://www.tbeic.go.jp/ (February 2005).

70. Ikegami, Masaharu, Yutaka Kameyama, Keita Furukawa and Tsuneki Sakakibara. 2003. Development of Coastal and Esturaine Environmental Data Sharing Using XML. The 3rd Internatinal Workshop on Scientific Use of Submarine Cables and Related Technologies, pp.293-297.

71. University Corporation for Atmospheric Research (UCAR) Climate and Global Dynamics Division. 2005. NetCDF Climate and Forecast (CF) Metadata Convention. See http://www.cgd.ucar.edu/cms/eaton/cf-metadata/ (February 2005).

Annex 1: SGXML Meeting Participants

The meeting participates represented 14 countries and two organizations, totalling 39 individuals. The chairs of other international groups are also recognized.

Country or International Organization	Participants
Australia	Greg Reed
Belgium	Christophe Bamelis Francisco Hernandez Angelino Meerhaeghe Edward Vanden Berghe (Chair GEBICH)
Canada	Jean Gagnon Anthony W. Isenor (Cochair SGXML)
Finland	Pekka Alenius Riikka Hietula Kati Manni Riitta Olsonen Kimmo Tikka
France	Thierry Carval Michele Fichaut (Cochair WGMDM)
Germany	Friedrich Nast Reinhard Schwabe
ICES	Harry Dooley
IOC	Peter Pissierssens Vladimir Vladymyrov
Japan	Keita Furukawa Tsuneki Sakakibara
Norway	Helge Sagen (Cochair WGMDM)
Poland	Slawomir Sagan Marcin Wichorowski

Russia	Sergey Belov Nicolay Mikhailov (Chair ETDMP) Georgiy Moiseenko
Sweden	Jan Szaron
The Netherlands	Pieter Haaring
United Kingdon	Laura Bird Ray Cramer Garry Dawson Michael Hughes Phil Knight Roy Lowry Lesley Rickards (Chair IODE)
United States	Donald W. Collins Robert D. Gelfeld (Cochair SGXML) Robert Starek

Annex 2: Code Mapping Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
  <xs:annotation>
    <xs:documentation>This is an annotated version of the code schema developed by the
"ICES/IOC Study Group on the Development of Marine Data Exchange Systems Using
XML" (2003-2004).  Version 1.0</xs:documentation>
  </xs:annotation>
  <xs:element name="multiplier">
    <xs:annotation>
      <xs:documentation>This is a multiplication factor that is used to convert units
associated with a particular code.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:simpleContent>
        <xs:extension base="xs:float">
          <xs:attribute name="pt_link" type="xs:int" use="required"/>
        </xs:extension>
      </xs:simpleContent>
    </xs:complexType>
  </xs:element>
  <xs:element name="codeset">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="codeset_name" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the name given to the set of codes to which the
following code belongs.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="code" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is a particular code.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="codeset_owner" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the responsible owner of the code
set.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element ref="multiplier" minOccurs="0" maxOccurs="unbounded">
          <xs:annotation>
            <xs:documentation>This is a multiplication factor associated with the
```

```
tranformation from one code unit to another code unit.</xs:documentation>
          </xs:annotation>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="synonym">
    <xs:annotation>
      <xs:documentation>Synonym is used to describe alternate words that may be used for
the dictionary_term.  The synonym allows multi-lingual use of the
structure.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element name="synonym_instance" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the synonym owner's description of the synonym.
Including the owner's description allows others to compare the synonym
descriptions.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="synonym_term" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the actual term that is used for the synonym.  It would
be common for this to be represented as a code from the synonym owner's
system.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="synonym_owner" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the owner of the synonym
description.</xs:documentation>
          </xs:annotation>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="definition">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="instance" type="xs:nonNegativeInteger">
          <xs:annotation>
            <xs:documentation>As with a common language dictionary, there may be
multiple definitions for a dictionary term.  The instance is a numeric that counts these
definitions.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="definition_owner" type="xs:string">
```

```
          <xs:annotation>
            <xs:documentation>This is organization that owns and is responsible for the
definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="short_name" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is an abbreviated name for the particular
definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="creation_date" type="xs:date" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is the date the definition was
created.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="change_date" type="xs:date" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is the last date the definition was
modified.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="methodology" type="xs:string" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is a description of the method used to obtain the data
value described by the definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="unit_of_measurement" type="xs:string" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is the unit associated with the particular definition.  Not
all definitions will have units.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="min_value" type="xs:float" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is minimum value associated with the
definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="max_value" type="xs:float" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is maximum value associated with the
definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="null_representation" type="xs:string" minOccurs="0">
```

```
          <xs:annotation>
            <xs:documentation>This is the representation that a NULL value may take for the
definition.  Not that often, NULL values are not XML friendly.  If the schema provides
range checking based on min and max values, and the NULL value is used for content,
then the schema check will result in errors when the NULL representation is outside the
common range of the definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="accuracy" type="xs:float">
          <xs:annotation>
            <xs:documentation>This is any associated accuracy with measurement of the
definition.  This may not apply to all definitions.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="authority_citation" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the publication style citation for this particular
definition.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element ref="codeset" minOccurs="0" maxOccurs="unbounded">
          <xs:annotation>
            <xs:documentation>This is an element that contains the set of information related
to the code that corresponds to the definition.  A single definition can have more than one
code associated with it.  The codes may span different organizations or systems within a
single organisation.</xs:documentation>
          </xs:annotation>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="dictionary_entry">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="dictionary_term" type="xs:string">
          <xs:annotation>
            <xs:documentation>A single listed item in the dictionary.  This element contains
the basic term for the dictionary.  In a common language dictionary, this element would
contain a word as listed in the dictionary.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="role" type="xs:string">
          <xs:annotation>
            <xs:documentation>This represents the role of the dictionary entry.  Role is a
higher level categorization of the entry.  For example, for code dictionaries roles may
include country codes, ship codes, parameter codes, etc.</xs:documentation>
          </xs:annotation>
        </xs:element>
```

```
        <xs:element ref="definition" maxOccurs="unbounded">
          <xs:annotation>
            <xs:documentation>The definition element allows multiple definitions of a single
dictionary term.  This is similar to a common language dictionary, where a single word
(the dictionary term) is allowed to have multiple definitions.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element ref="synonym" minOccurs="0" maxOccurs="unbounded">
          <xs:annotation>
            <xs:documentation>A synonym to the dictionary term.</xs:documentation>
          </xs:annotation>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="dictionary">
    <xs:annotation>
      <xs:documentation>A dictionary is a document that lists and explains the words of a
language.  In this dictionary strcuture, the dictionary entry represents a single item listed
within the dictionary.  The dictionary term represents the word that would be listed in the
dictionary.</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element name="dictionary_owner" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is a group or organisation that is recognized as
possessing ownership over the dictionary.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="dictionary_citation" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is the publication citation for the
dictionary.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="dictionary_description" type="xs:string">
          <xs:annotation>
            <xs:documentation>This is a description of the dictionary.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element name="date_structure" type="xs:string" minOccurs="0">
          <xs:annotation>
            <xs:documentation>This is the structure used for the date format within this
dictionary.</xs:documentation>
          </xs:annotation>
        </xs:element>
        <xs:element ref="dictionary_entry" maxOccurs="unbounded">
```

```
          <xs:annotation>
            <xs:documentation>This is a single listed item within the
dictionary.</xs:documentation>
          </xs:annotation>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Annex 3: XML Schema for Keeley Brick Profile Structure

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">

  <!--  Name: Keeley Brick Represention in XML
        Version: 2.1
        Date:  April 23, 2004

This is the working draft of the schema associated with the
  Canadian XML      efforts to implement 'Keeley Bricks' into an
  XML structure.
  The schema isdivided into five basic sections:
                                1) The actual top level structure
                                2) All Compound bricks
                                3) All Pure bricks
                                4) All Attribute Groups
                                  5) Misc. groups

  Present outstanding issues include:
                                  a) typing L to indicate lat/long format to be used
                                  b) the local_tag is not yet defined

Dec. 13, 2002 - Revised to set attribute occurence.
Dec. 16, 2002 - Revised to remove typing from latitude and
          longitude, and set the same date format for all
          date elements
Jan. 7, 2003  - Added 'name' attribute to the coefficient
          element within calibration brick.  Rearranged the
          XML types in the five groups defined above.
Jan. 20, 2003 - Removed local_tag from schema.  Removed
          order_number attribute            from comment element.
Feb. 10, 2003 - Corrected error.  instrument was suppose to
          be mandatory in variable_set.
Feb. 18, 2003 - Removed pt_code from history brick and added
                set_code.
Feb. 26, 2003 - Removed mandatory requirement on instrument
                brick inside variable_set.  Removed set_code
                in history and replaced it with an optional
                pt_code.
Mar. 18, 2003 - Added typing categories for Date, and Date/Time.
  Apr. 23, 2004 - Removed incorrect comment for typing_qualifiers
                and typing_qualifiers_mandatory.  Placed other
                valid comments in annotations.
```

```
  Anthony W. Isenor  -->

  <xsd:element name="data_collection" type="collection"/>

  <!-- This is the top level of the brick structure for point data.  -->

  <xsd:complexType name="collection">
    <xsd:sequence>
      <xsd:element name="comment" type="comment_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="data_dictionary" type="data_dictionary_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="provenance" type="provenance_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="location_set" type="location_set_cbrick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="data_set" type="data_set_cbrick" minOccurs="0"
maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:complexType>

  <!-- ***** Compound bricks in this section
**************************************************** -->

  <xsd:complexType name="data_set_base">
    <xsd:sequence>
      <xsd:element name="availability" type="availability_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="comment" type="comment_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="data_point" type="data_point_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="data_set_id" type="data_set_id_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="provenance" type="provenance_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="quality" type="quality_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="quality_testing" type="quality_testing_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="variable_set" type="variable_set_cbrick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="location_set" type="location_set_cbrick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="history_set" type="history_set_cbrick" minOccurs="0"
maxOccurs="unbounded"/>
```

```
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="data_set_cbrick">
    <xsd:complexContent>
      <xsd:extension base="data_set_base">
        <xsd:sequence>
          <xsd:element name="data_set" type="data_set_cbrick" minOccurs="0"
maxOccurs="unbounded"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>

  <xsd:complexType name="history_set_cbrick">
    <xsd:complexContent>
      <xsd:extension base="history_set_cbrick_1">
        <xsd:sequence>
          <xsd:element name="history" type="history_brick" minOccurs="0"
maxOccurs="1"/>
          <xsd:element name="previous_value" type="previous_value_brick"
minOccurs="0" maxOccurs="1"/>
          <xsd:element name="location_set" type="location_set_cbrick" minOccurs="0"
maxOccurs="1"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>

  <xsd:complexType name="history_set_cbrick_1">
    <xsd:sequence>
      <xsd:element name="comment" type="comment_brick" minOccurs="0"
maxOccurs="unbounded"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="location_set_cbrick">
    <xsd:sequence>
      <xsd:element name="comment" type="comment_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="depth_pressure" type="depth_pressure_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="latitude" type="latitude_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="ldate" type="ldate_brick" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="longitude" type="longitude_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="quality" type="quality_brick" minOccurs="0" maxOccurs="1"/>
```

```
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="variable_set_cbrick">
    <xsd:sequence>
      <xsd:element name="analysis_method" type="analysis_method_brick"
minOccurs="0" maxOccurs="1"/>
      <xsd:element name="calibration" type="calibration_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="comment" type="comment_brick" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="instrument" type="instrument_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="sampling" type="sampling_brick" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="sensor" type="sensor_brick" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="units" type="units_brick" minOccurs="1" maxOccurs="1"/>
      <xsd:element name="variable" type="variable_brick" minOccurs="1"
maxOccurs="1"/>
    </xsd:sequence>
  </xsd:complexType>


  <!-- ***** Pure bricks in this section
****************************************************** -->

  <xsd:complexType name="analysis_method_brick">
    <xsd:sequence>
      <xsd:element name="analysis_date" type="date_format" minOccurs="1"
maxOccurs="1"/>
      <xsd:element name="analysis_id" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="analyst_name" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="method" type="xsd:string" minOccurs="1" maxOccurs="1"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="availability_brick">
    <xsd:annotation>
      <xsd:documentation>The availability brick declares the possible release of the dataset
                                    in the community.</xsd:documentation>
    </xsd:annotation>
    <xsd:sequence>
      <xsd:element name="avail_date" type="date_format" minOccurs="1"
maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attributeGroup ref="indicator_qualifiers"/>
```

```
  </xsd:complexType>

  <xsd:complexType name="calibration_brick">
    <xsd:sequence>
      <xsd:element name="algorithm_type" type="xsd:string" minOccurs="1"
maxOccurs="1"/>
      <xsd:element name="application_date" type="date_format" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="calibration_date" type="date_format" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="coefficients" type="coefficient_set" minOccurs="0"
maxOccurs="unbounded"/>
      <xsd:element name="number_of_coefficients" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="process" type="xsd:string" minOccurs="0" maxOccurs="1"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="comment_brick">
    <xsd:simpleContent>
      <xsd:extension base="xsd:string"/>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:complexType name="data_dictionary_brick">
    <xsd:sequence>
      <xsd:element name="dictionary_name" type="xsd:string" minOccurs="1"
maxOccurs="1"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="data_point_brick">
    <xsd:simpleContent>
      <xsd:extension base="xsd:string">
        <xsd:attributeGroup ref="pt_qualifiers"/>
        <xsd:attributeGroup ref="stat_qualifiers"/>
        <xsd:attributeGroup ref="typing_qualifiers"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:complexType name="data_set_id_brick">
    <xsd:simpleContent>
      <xsd:extension base="xsd:string">
        <xsd:attributeGroup ref="level_qualifiers"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>
```

```
  <xsd:complexType name="depth_pressure_brick">
    <xsd:simpleContent>
      <xsd:extension base="xsd:string">
        <xsd:attributeGroup ref="kind_qualifiers"/>
        <xsd:attributeGroup ref="pt_qualifiers"/>
        <xsd:attributeGroup ref="stat_qualifiers"/>
        <xsd:attributeGroup ref="typing_qualifiers"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:complexType name="history_brick">
    <xsd:sequence>
      <xsd:element name="application_date" type="date_format" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="executor" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="process_identifier" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="version" type="xsd:string" minOccurs="0" maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attribute name="action" type="xsd:string"/>
    <xsd:attributeGroup ref="optional_pt_qualifiers"/>
  </xsd:complexType>

  <xsd:complexType name="instrument_brick">
    <xsd:sequence>
      <xsd:element name="description" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="manufacturer" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="model" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="serial_number" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attributeGroup ref="type_qualifiers"/>
  </xsd:complexType>

  <xsd:complexType name="latitude_brick">
    <xsd:simpleContent>
      <xsd:extension base="lat_restriction">
        <xsd:attributeGroup ref="position_qualifiers"/>
        <xsd:attributeGroup ref="stat_qualifiers"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:complexType name="ldate_brick">
```

```
    <xsd:choice>
      <xsd:group ref="date_choice"/>
      <xsd:group ref="time_choice"/>
    </xsd:choice>
    <xsd:attributeGroup ref="position_qualifiers"/>
    <xsd:attributeGroup ref="stat_qualifiers"/>
  </xsd:complexType>

  <xsd:complexType name="longitude_brick">
    <xsd:simpleContent>
      <xsd:extension base="long_restriction">
        <xsd:attributeGroup ref="position_qualifiers"/>
        <xsd:attributeGroup ref="stat_qualifiers"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:complexType name="previous_value_brick">
    <xsd:simpleContent>
      <xsd:extension base="xsd:string">
        <xsd:attributeGroup ref="pt_qualifiers"/>
        <xsd:attributeGroup ref="typing_qualifiers"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:complexType name="provenance_brick">
    <xsd:sequence>
      <xsd:element name="agency" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <xsd:element name="country" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="data_grouping" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="date_created" type="date_format" minOccurs="1"
maxOccurs="1"/>
      <xsd:element name="description" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="institute_code" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="originator" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="originator_identifier" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="platform_name" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="project" type="xsd:string" minOccurs="0" maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attributeGroup ref="platform_qualifiers"/>
  </xsd:complexType>
```

```
  <xsd:complexType name="quality_brick">
    <xsd:sequence>
      <xsd:element name="qt_date" type="date_format" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="tests_failed" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="tests_performed" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attribute name="justification_code" type="xsd:string"/>
    <xsd:attributeGroup ref="pt_qualifiers"/>
    <xsd:attributeGroup ref="reliability_qualifiers"/>
    <xsd:attributeGroup ref="use_qualifiers"/>
  </xsd:complexType>

  <xsd:complexType name="quality_testing_brick">
    <xsd:sequence>
      <xsd:element name="test_description" type="xsd:string" minOccurs="1"
maxOccurs="1"/>
      <xsd:element name="test_id" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <xsd:element name="test_name" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <xsd:element name="test_version" type="xsd:string" minOccurs="1"
maxOccurs="1"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="sampling_brick">
    <xsd:sequence>
      <xsd:element name="id" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="interval" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <xsd:element name="method" type="xsd:string" minOccurs="0" maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attributeGroup ref="pt_qualifiers"/>
  </xsd:complexType>

  <xsd:complexType name="sensor_brick">
    <xsd:sequence>
      <xsd:element name="manufacturer" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="model" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="serial_number" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="type" type="xsd:string" minOccurs="1" maxOccurs="1"/>
    </xsd:sequence>
  </xsd:complexType>

  <xsd:complexType name="units_brick">
    <xsd:sequence>
      <xsd:element name="conversion" type="xsd:string" minOccurs="0"
```

```
maxOccurs="1"/>
      <xsd:element name="reference" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="variable_name" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attributeGroup ref="pt_qualifiers"/>
    <xsd:attribute name="received_units" type="xsd:string"/>
    <xsd:attribute name="stored_units" type="xsd:string" use="required"/>
  </xsd:complexType>

  <xsd:complexType name="variable_brick">
    <xsd:sequence>
      <xsd:element name="accuracy" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="below_detection" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="decimal_places" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="maximum_value" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="minimum_value" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="null_value" type="xsd:string" minOccurs="0"
maxOccurs="1"/>
      <xsd:element name="precision" type="xsd:string" minOccurs="0" maxOccurs="1"/>
      <xsd:element name="variable_name" type="xsd:string" minOccurs="1"
maxOccurs="1"/>
    </xsd:sequence>
    <xsd:attributeGroup ref="duplicate_qualifiers"/>
    <xsd:attributeGroup ref="kind_qualifiers"/>
    <xsd:attributeGroup ref="pt_qualifiers"/>
    <xsd:attributeGroup ref="typing_qualifiers_mandatory"/>
  </xsd:complexType>


  <!-- ***** Attribute Groups in this section
********************************************* -->

  <xsd:attributeGroup name="duplicate_qualifiers">
    <xsd:attribute name="duplicate_indicator">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="N"/>
          <xsd:enumeration value="D"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:attribute>
  </xsd:attributeGroup>
```

```
  <xsd:attributeGroup name="indicator_qualifiers">
    <xsd:annotation>
      <xsd:documentation>The following lists the allowed content for the attribute and
definition associated with the content.</xsd:documentation>
      <xsd:documentation>R - Restricted</xsd:documentation>
      <xsd:documentation>O - Open</xsd:documentation>
      <xsd:documentation>C - Consultation required.</xsd:documentation>
    </xsd:annotation>
    <xsd:attribute name="indicator" use="required">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="R"/>
          <xsd:enumeration value="O"/>
          <xsd:enumeration value="C"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:attribute>
  </xsd:attributeGroup>

  <xsd:attributeGroup name="kind_qualifiers">
    <xsd:annotation>
      <xsd:documentation>The following lists the allowed content for the attribute and
definition associated with the content.</xsd:documentation>
      <xsd:documentation>I - Independent</xsd:documentation>
      <xsd:documentation>D - Dependent</xsd:documentation>
    </xsd:annotation>
    <xsd:attribute name="kind">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="I"/>
          <xsd:enumeration value="D"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:attribute>
  </xsd:attributeGroup>

  <xsd:attributeGroup name="level_qualifiers">
    <xsd:attribute name="level" use="required">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="cruise"/>
          <xsd:enumeration value="station"/>
          <xsd:enumeration value="profile"/>
          <xsd:enumeration value="record"/>
          <xsd:enumeration value="related"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:attribute>
```

```
</xsd:attributeGroup>

<xsd:attributeGroup name="platform_qualifiers">
  <xsd:attribute name="platform_type">
    <xsd:simpleType>
      <xsd:restriction base="xsd:string">
        <xsd:enumeration value="profiling float"/>
        <xsd:enumeration value="ship"/>
        <xsd:enumeration value="moored buoy"/>
        <xsd:enumeration value="drifting buoy"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:attribute>
</xsd:attributeGroup>

<xsd:attributeGroup name="position_qualifiers">
  <xsd:attributeGroup ref="kind_qualifiers"/>
  <xsd:attribute name="property">
    <xsd:simpleType>
      <xsd:restriction base="xsd:string">
        <xsd:enumeration value="start"/>
        <xsd:enumeration value="bottom"/>
        <xsd:enumeration value="end"/>
        <xsd:enumeration value="creation"/>
        <xsd:enumeration value="original"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:attribute>
</xsd:attributeGroup>

<xsd:attributeGroup name="pt_qualifiers">
  <xsd:attribute name="pt_code" type="xsd:string" use="required"/>
  <xsd:attribute name="pt_link" type="xsd:string"/>
</xsd:attributeGroup>

<xsd:attributeGroup name="reliability_qualifiers">
  <xsd:attribute name="reliability_code">
    <xsd:simpleType>
      <xsd:restriction base="xsd:unsignedShort">
        <xsd:enumeration value="0"/>
        <xsd:enumeration value="1"/>
        <xsd:enumeration value="2"/>
        <xsd:enumeration value="3"/>
        <xsd:enumeration value="4"/>
        <xsd:enumeration value="5"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:attribute>
```

```
  </xsd:attributeGroup>

  <xsd:attributeGroup name="optional_pt_qualifiers">
    <xsd:attribute name="pt_code" type="xsd:string"/>
    <xsd:attribute name="pt_link" type="xsd:string"/>
  </xsd:attributeGroup>

  <xsd:attributeGroup name="stat_qualifiers">
    <xsd:attribute name="statistic" type="xsd:string"/>
  </xsd:attributeGroup>

  <xsd:attributeGroup name="type_qualifiers">
    <xsd:annotation>
      <xsd:documentation>The following lists the allowed content for the attribute and
definition associated with the content.</xsd:documentation>
      <xsd:documentation>adcp - Acoustic Doppler Current Profilier</xsd:documentation>
      <xsd:documentation>bottle - water sampling bottle</xsd:documentation>
      <xsd:documentation>cm - current meter</xsd:documentation>
      <xsd:documentation>CTD - Conductivity, Temperature, Depth
instrument</xsd:documentation>
      <xsd:documentation>dbt -</xsd:documentation>
      <xsd:documentation>float - any surface, subsurface, or oscillating
float</xsd:documentation>
      <xsd:documentation>model -</xsd:documentation>
      <xsd:documentation>radar -</xsd:documentation>
      <xsd:documentation>staff -</xsd:documentation>
      <xsd:documentation>staff_gauge -</xsd:documentation>
      <xsd:documentation>sounder - Any device for obtaining acoustic depth
measurements</xsd:documentation>
      <xsd:documentation>thermistor -</xsd:documentation>
      <xsd:documentation>uway - underway</xsd:documentation>
      <xsd:documentation>unknown -</xsd:documentation>
      <xsd:documentation>water_level_gauge -</xsd:documentation>
      <xsd:documentation>wave_buoy -</xsd:documentation>
      <xsd:documentation>wave_directional_buoy -</xsd:documentation>
      <xsd:documentation>wave_pressure_gauge -</xsd:documentation>
      <xsd:documentation>wave_recorder -</xsd:documentation>
      <xsd:documentation>XBT - eXpendible bathythermograph</xsd:documentation>
    </xsd:annotation>

    <xsd:attribute name="type" use="required">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="adcp"/>
          <xsd:enumeration value="bottle"/>
          <xsd:enumeration value="cm"/>
          <xsd:enumeration value="CTD"/>
          <xsd:enumeration value="dbt"/>
```

```
        <xsd:enumeration value="float"/>
        <xsd:enumeration value="model"/>
        <xsd:enumeration value="radar"/>
        <xsd:enumeration value="staff"/>
        <xsd:enumeration value="staff_gauge"/>
        <xsd:enumeration value="sounder"/>
        <xsd:enumeration value="thermistor"/>
        <xsd:enumeration value="uway"/>
        <xsd:enumeration value="unknown"/>
        <xsd:enumeration value="water_level_gauge"/>
        <xsd:enumeration value="wave_buoy"/>
        <xsd:enumeration value="wave_directional_buoy"/>
        <xsd:enumeration value="wave_pressure_gauge"/>
        <xsd:enumeration value="wave_recorder"/>
        <xsd:enumeration value="XBT"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:attribute>
</xsd:attributeGroup>

<xsd:attributeGroup name="typing_qualifiers">
  <xsd:annotation>
    <xsd:documentation>The following lists the allowed content for the attribute and
definition associated with the content.</xsd:documentation>
    <xsd:documentation>T - Time</xsd:documentation>
    <xsd:documentation>D - Date</xsd:documentation>
    <xsd:documentation>DT - Date and time</xsd:documentation>
    <xsd:documentation>R - Number with a decimal</xsd:documentation>
    <xsd:documentation>I - Integer</xsd:documentation>
    <xsd:documentation>C - Character</xsd:documentation>
  </xsd:annotation>
  <xsd:attribute name="typing">
    <xsd:simpleType>
      <xsd:restriction base="xsd:string">
        <xsd:enumeration value="T"/>
        <xsd:enumeration value="D"/>
        <xsd:enumeration value="DT"/>
        <xsd:enumeration value="R"/>
        <xsd:enumeration value="I"/>
        <xsd:enumeration value="C"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:attribute>
</xsd:attributeGroup>

<xsd:attributeGroup name="typing_qualifiers_mandatory">
  <xsd:annotation>
    <xsd:documentation>The following lists the allowed content for the attribute and
```

```
definition associated with the content.</xsd:documentation>
      <xsd:documentation>T - Time</xsd:documentation>
      <xsd:documentation>D - Date</xsd:documentation>
      <xsd:documentation>DT - Date and time</xsd:documentation>
      <xsd:documentation>R - Number with a decimal</xsd:documentation>
      <xsd:documentation>I - Integer</xsd:documentation>
      <xsd:documentation>C - Character</xsd:documentation>
    </xsd:annotation>
    <xsd:attribute name="typing" use="required">
      <xsd:simpleType>
        <xsd:restriction base="xsd:string">
          <xsd:enumeration value="T"/>
          <xsd:enumeration value="D"/>
          <xsd:enumeration value="DT"/>
          <xsd:enumeration value="R"/>
          <xsd:enumeration value="I"/>
          <xsd:enumeration value="C"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:attribute>
  </xsd:attributeGroup>

  <xsd:attributeGroup name="use_qualifiers">
    <xsd:attribute name="use_code" type="xsd:string"/>
  </xsd:attributeGroup>


  <!-- ***** Mics. groups in this section
********************************************** -->

  <xsd:complexType name="coefficient_set">
    <xsd:simpleContent>
      <xsd:extension base="xsd:string">
        <xsd:attribute name="name" type="xsd:string"/>
      </xsd:extension>
    </xsd:simpleContent>
  </xsd:complexType>

  <xsd:group name="date_choice">
    <xsd:sequence>
      <xsd:element name="pdate" type="date_format" minOccurs="1" maxOccurs="1"/>
      <xsd:element name="ptime" type="time_restriction" minOccurs="0"
maxOccurs="1"/>
    </xsd:sequence>
  </xsd:group>

  <xsd:complexType name="date_format">
    <xsd:simpleContent>
```

```
    <xsd:extension base="date_restriction"/>
  </xsd:simpleContent>
</xsd:complexType>

<xsd:simpleType name="lat_restriction">
  <xsd:restriction base="xsd:decimal">
    <xsd:minInclusive value="-90.0"/>
    <xsd:maxInclusive value="90.0"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="long_restriction">
  <xsd:restriction base="xsd:decimal">
    <xsd:minInclusive value="-180.0"/>
    <xsd:maxInclusive value="180.0"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:group name="time_choice">
  <xsd:sequence>
    <xsd:element name="ptime" type="time_restriction"/>
  </xsd:sequence>
</xsd:group>

<!--Note:  This restriction is required because I have
discovered that the validator I am using does not correctly
implement the date or time xsd datatypes.  The following
restrictions help ensure the proper checking of the date and
time datatypes.  Note that the restrictions are in addition
to the datatype defined by date and time, and so do not
restrict the exact form of the date or time.  (Example:
The pattern for hours implies that 88 is a valid value.
Hovever, the time type properly restricts the values to 23
or less.

Note also that the restrictions force Zulu time to be
specified using the capital Z character.  Also, no time
zome specification is allowed.

A.W.Isenor (Dec. 2002)-->
<xsd:simpleType name="date_restriction">
  <xsd:restriction base="xsd:date">
    <xsd:pattern value="([0-9]{4}-[0-9]{2}-[0-9]{2}Z)"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="time_restriction">
  <xsd:restriction base="xsd:time">
```

```
      <xsd:pattern value="([0-9]{2}):([0-9]{2}):(([0-9]{2})|([0-9]{2})\.[0-9]*)Z"/>
    </xsd:restriction>
  </xsd:simpleType>
</xsd:schema>
```

Annex 4: Biological Net Tow Data in Generic XML Structure


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<data_collection xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="bricks_v2.xsd">
 <data_dictionary>
  <dictionary_name>EVB Standard Taxa</dictionary_name>
 </data_dictionary>
 <provenance>
  <agency>Flanders Marine Data and Information Centre</agency>
  <date_created>2004-04-22Z</date_created>
  <description>Biological dataset</description>
  <institute_code>****</institute_code>
  <originator_identifier>Belgica 94/21</originator_identifier>
  <platform_name>Belgica</platform_name>
 </provenance>
 <data_set>
  <data_set_id level="cruise">4</data_set_id>
  <data_set>
   <data_set_id level="station">9</data_set_id>
   <data_set>
    <data_point pt_code="TRLG">200</data_point>
    <data_point pt_code="TRVL">156.5</data_point>
    <data_set_id level="profile">900</data_set_id>
    <variable_set>
     <units pt_code="TRLG" stored_units="m"/>
     <variable pt_code="TRLG" typing="R">
      <variable_name>Trawl length</variable_name>
     </variable>
    </variable_set>
    <variable_set>
     <units pt_code="TRVL" stored_units="m**3"/>
     <variable pt_code="TRVL" typing="R">
      <variable_name>Trawl volume</variable_name>
     </variable>
    </variable_set>
    <variable_set>
     <instrument type="unknown">
      <description>Sorbe sledge</description>
     </instrument>
     <sensor>
      <type>net 200</type>
     </sensor>
     <units pt_code="NPOS" pt_link="3" stored_units=""/>
```

```
  <variable pt_code="NPOS" pt_link="3" typing="C">
    <variable_name>Net Position</variable_name>
  </variable>
</variable_set>
<variable_set>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
  <sensor>
    <type>net 50</type>
  </sensor>
  <units pt_code="NPOS" pt_link="4" stored_units=""/>
  <variable pt_code="NPOS" pt_link="4" typing="C">
    <variable_name>Net Position</variable_name>
  </variable>
</variable_set>
<variable_set>
  <analysis_method>
    <analysis_date>2004-04-29Z</analysis_date>
    <analyst_name>Ann Dewicke</analyst_name>
    <method>Some book on Hydrozoa</method>
  </analysis_method>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
  <units pt_code="Genus" pt_link="1" stored_units=""/>
  <variable pt_code="Genus" pt_link="1" typing="C">
    <variable_name>Genus of the beast</variable_name>
  </variable>
</variable_set>
<variable_set>
  <analysis_method>
    <analysis_date>2004-04-29Z</analysis_date>
    <analyst_name>Jan Wittoeck</analyst_name>
    <method>Information Guide 2</method>
  </analysis_method>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
  <units pt_code="Genus" pt_link="2" stored_units=""/>
  <variable pt_code="Genus" pt_link="2" typing="C">
    <variable_name>Genus of the beast</variable_name>
  </variable>
</variable_set>
<variable_set>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
```

```
  <units pt_code="Stage" stored_units=""/>
  <variable pt_code="Stage" typing="C">
    <variable_name>Stage of development</variable_name>
  </variable>
</variable_set>
<variable_set>
  <analysis_method>
    <analysis_date>2004-04-23Z</analysis_date>
    <method>A very good eye</method>
  </analysis_method>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
  <units pt_code="Gender" stored_units=""/>
  <variable pt_code="Gender" typing="C">
    <variable_name>Gender of the beast</variable_name>
  </variable>
</variable_set>
<variable_set>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
  <units pt_code="number" stored_units=""/>
  <variable pt_code="number" typing="I">
    <variable_name>The number of counts of the beast</variable_name>
  </variable>
</variable_set>
<variable_set>
  <analysis_method>
    <analysis_date>2004-04-23Z</analysis_date>
    <method>weighing scale</method>
  </analysis_method>
  <instrument type="unknown">
    <description>Sorbe sledge</description>
  </instrument>
  <units pt_code="biomass" stored_units="g"/>
  <variable pt_code="biomass" typing="D">
    <variable_name>The biomass of the beasts</variable_name>
  </variable>
</variable_set>
<location_set>
  <latitude property="start">51.1832</latitude>
  <ldate property="start">
    <pdate>1994-09-06Z</pdate>
    <ptime>13:40:00Z</ptime>
  </ldate>
  <longitude property="start">2.7017</longitude>
</location_set>
```

```
        <history_set>
          <comment>These data were supplied by Edward Vanden Berghe and transformed
into XML by Anthony W. Isenor</comment>
          <history>
            <application_date>2004-04-22Z</application_date>
          </history>
        </history_set>
        <data_set>
          <data_set_id level="record">29</data_set_id>
          <location_set>
            <depth_pressure pt_code="DEPT">14.141</depth_pressure>
          </location_set>
          <data_set>
            <data_point pt_code="Genus" pt_link="1">Crangon</data_point>
            <data_point pt_code="Genus" pt_link="2">Crangon</data_point>
            <data_point pt_code="Stage">Postlarva</data_point>
            <data_point pt_code="number">15</data_point>
            <data_point pt_code="biomass">99</data_point>
            <data_point pt_code="NPOS" pt_link="3">top</data_point>
          </data_set>
          <data_set>
            <data_point pt_code="Genus" pt_link="1">Crangon</data_point>
            <data_point pt_code="Stage">Zoea</data_point>
            <data_point pt_code="number">46</data_point>
            <data_point pt_code="biomass">99</data_point>
            <data_point pt_code="NPOS" pt_link="4">bottom</data_point>
          </data_set>
        </data_set>
      </data_set>
    </data_set>
  </data_set>
</data_collection>
```

List of symbols/abbreviations/acronyms/initialisms

A	Archival (metadata type)
ACE	Advisory Committee on Ecosystems
AODC	Australian Oceanographic Data Centre
B	Browse (metadata type)
BIO	Bedford Institute of Oceanography
BODC	British Oceanographic Data Centre
C	Collection (metadata type)
CSDGM	Content Standard for Digital Geospatial Metadata
CSR	Cruise Summary Report
CTD	Conductivity-Temperature-Depth
D	Discovery (metadata type)
DDF	Data Documentation Form (US NODC)
DFO	Fisheries and Oceans
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.
DNA	Designated National Agencies
DND	Department of National Defence (Canada)
DRDC	Defence R&D Canada
E	Extra (metadata type)
EBCDIC	Extended Binary Coded Decimal Interchange Code

EC	European Commission
EDMED	European Directory of Marine Environmental Data
EnParDis	Enabling Parameter Discovery
ESADS	Earth Science and Applications Data Systems
ETDMP	Expert Team on Data Management Practices (JCOMM)
EU	European Union
FGDC	Federal Geographic Data Committee (USA)
FIMR	Finnish Institute of Marine Research
FRS	Fisheries Research Services
G8	Group of Eight
GCMD	Global Change Master Directory
GEBICH	Group of Experts on Biological and Chemical Data Management and Exchange Practices (IODE)
GETADE	IOC/IODE Group of Experts on Technical Aspects of Data Exchange
GF3	General Format 3
GILS	Global Information Locator Service
GIS	Geographic Information System
GML	Geography Markup Language
GOOS	Global Ocean Observing System
ICES	International Council for the Exploration of the Sea
ICSU	International Council for Science
IFREMER	Institut Francais pour le Recherche et l'Exploitation de la Mer
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange

IOS	Institute of Ocean Sciences
ISO	International Organisation for Standardization
ITIS	Integrated Taxonomic Information System
JCOMM	Joint WMO/IOC Commission on Oceanography and Marine Meteorology
JGOFS	Joint Global Ocean Flux Study
JGSI	Japan Geographical Survey Institute
JODC	Japanese Oceanographic Data Center
MARC	MAchine Readable Cataloguing
MAST	Marine Science and Technology
MEDI	Marine Environmental Data Information Referral Catalogue system (IOC)
MEDS	Marine Environmental Data Service (Canada)
MMI	Marine Metadata Interoperability
NASA	National Aeronautics and Space Administration
NDG	NERC DataGrid
NERC	Natural Environment Research Council (UK)
CF	Climate and Forecast
NODC	National Oceanographic Data Centre
OAI	Open Archive Initiative
OBIS	Ocean Biogeographic Information System
OGC	Open GIS Consortium
OWS	OGC Web Services
PC	Personal Computer
PICES	North Pacific Marine Science Organization

RNODC	Responsible National Oceanographic Data Centre
S	Summary (metadata type)
SEERAD	Scottish Executive Environment and Rural Affairs Department
SGXML	ICES/IOC Study Group on the Development of Marine Data Exchange Systems using XML
SMHI	Swedish Meteorological and Hydrological Institute
SSF	Science Strategic Funds
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
TBEIC	Tokyo Bay Environmental Information Center
TC	Technical Committee
TCODE	Technical Committee for Data Exchange
TOR	Terms of Reference
UML	Unified Modelling Language
UN	United Nations
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UPD	Universal Parameter Dictionary
US	United States
VLIZ	Vlaams Instituut voor de Zee (Flanders Marine Institute)
W3C	World Wide Web Consortium
WGMDM	Working Group on Marine Data Management (ICES)
WMO	World Meteorological Organisation
XBT	eXpendable Bathythermograph
XML	eXtensible Markup Language

XSD	XML Schema Definition
XSL	eXtensible Stylesheet Language
XSLT	eXtensible Stylesheet Language Transformation

Distribution list

Document No.: DRDC Atlantic ECR 2005-005

LIST PART 1:

2	DRDC Atlantic LIBRARY FILE COPIES (1 cd and 1 hard copy with CD insert)
4	DRDC Atlantic LIBRARY (SPARES) (2 cds and 2 hard copies with CD insert)
1	W. A. ROGER
1	M. E. LEFRANÇOIS
1	G. R. MELLEMA
1	D. G. HAZEN
1	P. HINES
4	A. W. ISENOR (3 CD copies, 1 hard copy with CD insert)
15	TOTAL LIST PART 1

LIST PART 2: DISTRIBUTED BY DRDKIM 2-2-5

1 NDHQ/DRDKIM 2-2-5

1 Pekka Alenius
Finnish Institute of Marine Research,
P.O. Box 33, (Lyypekinkuja 3),
00931 Helsinki,
Finland

1 Christophe Bamelis
Belgian Sea Fisheries Department (CLO-DVZ)
Ankerstraat 1
8400 Oostende, Belgium

1 Sergey Belov
Russian National Oceanographic Data Centre
6 Korolev St., Obninsk, Kaluga Region,
Russian Federation, 249020

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

1 Thierry Carval
Institut Francais pour le Recherche et
l'Exploitation de la Mer (IFREMER),
Center de Brest,
Departement IDM,
BP 70,
29280 Plouzane
France

1 Donald W. Collins
U.S. National Oceanographic Data Center
1315 East West Highway, 4th Floor,
Silver Spring, MD, 20910, U.S.A.

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

1 Garry Dawson
Maritime Environment Information Center,
UK Hydrographic Office,
Admiralty Way, Taunton,
Somerset TA1 2DN,
U.K.

1 Julie Gillin
International Council for the Exploration of the Sea (ICES),
Palaegade 2-4,
1261 Copenhagen K,
Denmark

1 David Griffith
International Council for the Exploration of the Sea (ICES),
Palaegade 2-4,
1261 Copenhagen K,
Denmark

1 Michele Fichaut (Cochair WGMDM)
Institut Francais pour le Recherche et
l'Exploitation de la Mer (IFREMER),
Centre de Brest,
TMSI./IDM/SISMER,
BP 70,
29280 Plouzane
France

1 Keita Furukawa
National Institute for Land and Infrastructure Management
Ministry of Land, Infrastructure and Transport
3-1-1, Nagase, Yokosuka 239-0826, Japan

1 Jean Gagnon
Marine Environmental Data Service (MEDS),
Department of Fisheries and Oceans,
200 Kent Street, 12th Floor,
Ottawa, Ontario K1A OE6
Canada

1 Robert D. Gelfeld (Cochair SGXML)
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 Pieter Haaring
National Institute for Coastal and Marine Management (RIKZ)
P.O. Box 20907
2500 EX The Hague
The Netherlands

1 Francisco Hernandez
Flanders Marine Data and Information Centre
Flanders Marine Institute
Vismijn, Pakhuizen 45-52, B-8400 Oostende, Belgium

1 Riikka Hietula
Finnish Institute of Marine Research
P.O. Box 33, Fin-00931,
Helsinki, Finland

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

1 J. Robert Keeley
Marine Environmental Data Service
Department of Fisheries and Oceans
W12082 - 200 Kent Street
Ottawa, Ontario Canada
K1A 0E6

1 Phil Knight
Proudman Oceanographic Laboratory,
Bidston Observatory, Prenton,
Merseyside, CH43 7RA,
United Kingdom

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

1 Kati Manni
Finnish Environment Institute,
Data and Information Centre
P.O. Box 140, FIN-00251,
Helsinki, Finland

1 Angelino Meerhaeghe
Management Unit of the Mathematical Model North Sea (MUMM)
Royal Belgian Institute for Natural Sciences
Gulledelle 100
B-1200 Brussel
Belgium

1 Nicolay Mikhailov (Chair ETDMP)
Russian National Oceanographic Data Centre
6 Korolev St., Obninsk, Kaluga Region,
Russian Federation, 249020

1 Georgiy Moiseenko
Russian Federal Research
Institute of Fisheries and Oceanography (VNIRO)
17, V. Krasnoselskaya
Moscow, 107140,
Russia

1 Friedrich Nast
Deutsches Ozeanographisches Datenzentrum (DOD),
Bundesamt für Seeschiffahrt und Hydrographie
Bernhard-Nocht-Str. 78
D-20359 Hamburg, Germany

1 Riitta Olsonen
Finnish Institute of Marine Research,
P.O. Box 33, (Lyypekinkuja 3),
00931 Helsinki, Finland

1 Peter Pissierssens
Head, Ocean Services
UNESCO-IOC
1, rue Miollis
75732 Paris Cedex 15
France

1 Greg Reed
RAN METOC
Maritime Headquarters
Wylde Street
Potts Point, NSW 2011
Australia

1 Lesley Rickards (Chair IODE)
British Oceanographic Data Centre,
Proudman Oceanographic Laboratory,
6 Brownlow Street,
Liverpool L3 5DA,
United Kingdom.

1 Helge Sagen (Cochair WGMDM)
Institute of Marine Research
Norwegian Marine Data Centre
Nordnesgt 50, 5817, Bergen
Norway

1 Slawomir Sagan
Institute of Oceanology PAS,
Powstancow Warszawy 55,
81-712 SOPOT, PL

1 Tsuneki Sakakibara
University of Tokyo
Center for Spatial Information Science
Cw-502, 4-6-1, Komaba, Meguro-ku
Tokyo 153-8505, Japan

1 Reinhard Schwabe
Deutsches Ozeanographisches Datenzentrum (DOD),
Bundesamt für Seeschiffahrt und Hydrographie
Bernhard-Nocht-Str. 78
D-20359 Hamburg, Germany

1 Tobias Spears
Informatics
Bedford Institute of Oceanography
PO Box 1006
Dartmouth, NS B2Y 4A2
Canada

1 Robert Starek
Commander, Naval Meteorology and Oceanography Command
1100 Balch Blvd., Code N522
Stennis Space Center, MS 39529-5005

1 Einar Svendsen
Chair ICES Oceanography Committee
Oceanography and Climate
Institute of Marine Research
P.O.Box 1870 Nordnes
N-5817 Bergen, Norway

1 Jan Szaron
Swedish Meteorological and Hydrological Institute,
Oceanographic Unit,
Nya Varvet 31,
SE - 426 71 Vastra Frolunda,
Sweden

1 Kimmo Tikka
Finnish Institute of Marine Research,
P.O. Box 33, (Lyypekinkuja 3),
00931 Helsinki, Finland

1 Edward Vanden Berghe (Chair GEBICH)
Manager, Flanders Marine Data and Information Centre
Flanders Marine Institute
Vismijn, Pakhuizen 45-52, B-8400 Ostend, Belgium

1 Vladimir Vladymyrov
Programme Specialist
Ocean Services
UNESCO-IOC
1, rue Miollis
75732 Paris Cedex 15
France

1 Marcin Wichorowski
Institute of Oceanology PAS,
Powstancow Warszawy 55,
81-712 SOPOT, PL

44 TOTAL LIST PART 2

58 TOTAL COPIES REQUIRED

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

   Defence R&D Canada – Atlantic
   PO Box 1012
   Dartmouth, Nova Scotia
   B2Y 3Z7

2. SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable).

   UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title).

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

4. AUTHORS (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

   Anthony W. Isenor and Roy K. Lowry

5. DATE OF PUBLICATION (month and year of publication of document)

   March 2005

6a. NO. OF PAGES (total containing information Include Annexes, Appendices, etc).

   92

6b. NO. OF REFS (total cited in document)

   71

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered).

   External Client Report

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include address).

   Defence R&D Canada – Atlantic
   PO Box 1012
   Dartmouth, NS, Canada B2Y 3Z7

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant).

   11cg01

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written).

10a ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

   DRDC Atlantic ECR 2005-005

10b OTHER DOCUMENT NOs. (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

   DRDC Atlantic TR 2005-005

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

   (x) Unlimited distribution
   ( ) Defence departments and defence contractors; further distribution only as approved
   ( ) Defence departments and Canadian defence contractors; further distribution only as approved
   ( ) Government departments and agencies; further distribution only as approved
   ( ) Defence departments; further distribution only as approved
   ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected).

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

In 2001 the Intergovernmental Oceanographic Commission (IOC) and the International Council for the Exploration of the Sea (ICES) cooperatively formed a Study Group to examine the application of the eXtensible Markup Language (XML) to marine data exchange systems. The Study Group first met in April 2002 to address issues around the transfer of oceanographic data. The Group has met three times, with the final meeting in May 2004. This document represents the final report of the Group.

The Study Group concentrated its efforts on metadata standards, parameter dictionaries and generic data structures for use in an XML-based language. The Group evaluated several international metadata structures and produced mappings between some structures. In terms of the parameter dictionaries, the Group conducted mappings between several international parameter dictionaries, made structural advances to some dictionaries and attempted to account for dictionary issues imposed by units. The generic data structure development produced about 20 data objects that were then used to create an XML data structure for the transport of ocean environmental data. The structure was applied to one and three-dimensional data sets. The Group has also made numerous recommendations to continue the development of international data exchange systems.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title).

Extensible Markup Language; XML; data model; data structure; metadata; parameter dictionary;

This page intentionally left blank.

Defence R&D Canada

Canada's leader in defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE DÉFENSE

www.drdc-rddc.gc.ca