Development of a Marine XML
Prepared by Ben Searle, IODE Chairman

1. Introduction

The necessity to access greater quantities of marine data to support a broad range of maritime activities has been increasing rapidly over the past few years. Larger quantities of data combined with a greater need to make existing data more accessible is making the task of the marine data manager increasingly more difficult.

Today the information systems that are used to support marine data management are generally developed specifically for each individual agency. The systems are developed to support an agencies specific need and more importantly their specific data formats. This leads to inefficiencies in system development and promotes proprietary data structures further complicating data exchange and data fusion activities.

The IODE community is only too aware that we have not successfully been able to standardise data formats. A lack of standard data formats complicates data exchange and also makes the fusion of multiple data sets more difficult. It also results in our inability to share software for managing or processing data. This leads to a tremendous amount of duplication of effort with data managers all around the world developing systems to undertake mostly similar functions. There is a tremendous waste of effort and cost associated with this duplication.

Today the marine community is comprised of an increasing number of organisations and scientists collecting larger volumes of data for an ever-increasing number of purposes. The imperative to create national, regional and global databases is also increasing which is further highlighting the lack of an open, integrating marine data framework.

2. Internet Based Data Standards

It is fortunate that Information Technology has now reached a point where an environment has been created that can be readily adapted to meet the specific needs of the ocean community. In particular, the World Wide Web Consortium (W3C) has created a specification for the Extensible Markup Language (XML). The potential of XML to support multi-formatted data and improve electronic data exchange has been widely recognised by a number of industries, scientific disciplines, government agencies and the broad IT industry. A large number of groups are adopting XML and developing industry specific XML frameworks and XML is ideally suited to the marine community.

3. What is XML

XML is an Extensible Markup Language (extensible because it is not a fixed format like HTML). XML is not a single, predefined markup language: it is a meta-language “a language for
describing other languages” It is a set of rules for creating semantic tags used to describe data. XML is fast becoming the standard for data representation and exchange on the Internet. The basic ideas underlying XML are very simple: tags on data elements identify the meaning of the data, rather than, as with HTML, specifying how the data should be formatted, and relationships between data elements are provided via simple nesting and references. Web servers and applications encoding data in XML can quickly make the information available in a simple and useable format. As the information content is separated from information rendering, it is easy to provide multiple views of the same data.

XML provides many advantages in relation to the exchange of marine data. It can provide a platform for data centres to share data in a common syntax. Data can be stored in the originators format and yet still be available for exchange by using pre-defined tags to describe the data.

XML is already in use within the IODE community. The MEDI Pilot Project software provides the capability of producing XML output from the metadata database.

4. Benefits of XML

There are a number of reasons for using a marine XML:

8.1. Exchange of data. A major strength and source of potential of XML is that it facilitates the exchange of data between different applications and operating systems. One of XML's strongest points is its ability to do data interchange. Because different organisations (or even different parts of the same organisation) rarely standardise on a single set of tools, it takes a significant amount of work for two groups to communicate. XML makes it easy to send structured data across the web so that nothing gets lost in translation. XML is potentially the answer for oceanographic data exchange, as long as all sides agree on the markup to use.

8.2. Extensibility. Extensible means that it is not a fixed format like HTML. While HTML tags must follow pre-set standards, new XML tags can be created by anyone at any time. XML will allow groups of people or organisations to create their own customised markup languages for exchanging information in their domain. Examples of existing industry-specific XML include music, chemistry, electronics, linguistics, engineering and mathematics.

8.3. Plain Text. Since XML is not a binary format, files can be created and edited with a standard text making it useful for storing small amounts of data. At the other end of the spectrum, an XML front end to a database makes it possible to efficiently store large amounts of XML data. XML provides scalability for anything from small configuration files to an industry-wide data repository.

8.4. Data Identification. The XML standard specifies how to identify data, not how to display it. HTML, on the other hand, describes how things should be displayed without identifying the content. Because the different parts of the information have been identified, they can be used in different ways by different applications.

8.5. Stylability. When display is important, the style sheet standard, XSL, can dictate how to portray the data. Since XML is inherently style-free, different style sheets can be used to produce output in postscript, PDF, or any other format.
8.6. Hierarchical. XML documents are hierarchical in structure. Hierarchical document structures are, in general, faster to access because you can drill down to the part you need, like stepping through a table of contents.

5. Establishing An XML Based Marine Data Framework

One of the strengths of XML, its extensibility, is also a weakness since it is possible to ‘relatively easily’ to create XML structures. If a multiple groups attempt to develop their own XML specification, then we end up in a similar position to that of today, with numerous data structures and formats. The establishment of a number of ‘versions’ of marine XML would significantly negate the tremendous potential of this technology and only amplify the existing problems. There are already some marine groups around the world who are looking at the potential of XML and some of these intend to develop an XML framework to suit their own internal data management needs.

The marine community now has available an opportunity to create a truly universal and open marine data framework. The creation of a Marine XML data framework specification will provide a wide range of benefits to the user community. There will also be considerable benefits to the commercial sector, especially the producers of instruments and marine focussed application software. A Marine XML can provide a standardised data framework that can then support the creation of generic software, useable by everybody, rather than the present approach of building ‘one-off’ applications for each user or agency that results in an ‘isolated island of information’. A marine version of XML will simplify the process of ‘end to end’ data management, enabling a clean (electronic) flow of data from the instrument to the database, to the application and to the ‘client’ in a simple and common form. The beauty of XML is that it will support legacy (existing) data formats and information systems while providing maximum benefits when included in the development of new systems. Existing investments in instruments and computing systems are protected, and the user is able to easily benefit from the XML framework.

6. Developing a Marine XML Specification

6.1 Introduction

The process of developing a standard and a specification is difficult from two perspectives.

1. there is a technical complexity of developing a structure to support the wide range of marine data types that are often very complex in structure, and
2. political issues.

6.2 Technical Issues

The technical issues can be overcome with some research and collaboration between computing scientists with XML knowledge and domain specialists from the marine community. In addition to this, there are already a number of other XML based specifications being developed that can be used as a basis for the creation of the marine specification. A benefit of this is that the marine community need not ‘lead’ the way in this process.

One of the key technical issues that need to be resolved is the definition of field names or parameters (semantics). Given the crucial role of ‘semantics’ in the development of specific XML versions, the use of existing ‘key word or parameter lists will greatly assist with the
development process. There are three parameter keyword lists or data dictionaries that could provide the basis of the semantic structure of the Marine XML. These are:

1. NASA Global Change Master Directory (GCMD) key words list
2. British Oceanographic Data Centre (BODC) ocean data parameter list
3. Texas A&M University ocean data parameter list and description

It is likely that agreement will not be reached on each parameter name. For that reason, Marine XML Framework will rely on ontologies to determine the semantics of the parameters in order to make them easily convertible.

The three existing parameter keyword lists and data dictionaries provide the basis of the semantic structure of the Marine XML. XML tags and namespaces will be semantically related to the existing parameter lists through ontologies. This will ensure that standard tools can be created to provide translation between different marine data description conventions, even those that are still being maintained and changing. In this way, the Marine XML will remain open in its ability to integrate various standards.

### 6.3 Political Issues

The second issue (political) is more complex to resolve. There are a number of ways to develop a specification including using a consensus approach or getting a small group of specialists to define the specification and then releasing this to the ‘world’. The success of ‘standards that are simply ‘released’ to the community is very limited. IODE has some knowledge of this already with General Format Three (GF3). While GF3 was an excellent format, and ahead of its time in many ways, it has not gained a wide acceptance in the marine community.

The method that is proposed for the creation of the Marine XML specification is to use a consensus approach. However, rather than undertake this only within the IODE community, it is proposed that representatives from a much broader section of the marine community be invited to participate. This would include groups from the commercial sector with representatives from instrument manufacturers, software developers and consulting companies. Major government agencies and international programs both intergovernmental (eg. IODE) and scientific should also be included.

A relatively large number of ‘contributors’ to the specification development process brings its own problems. However, a distinct advantage is that by having the ‘major’ groups in the marine community contributing they will have a level of ownership of the concept and are therefore more likely to use the specification when it is developed. This is especially true of instrument manufacturers and applications developers. If the scientific instruments output data already tagged with XML tags and applications can automatically read and understand this data in a meaningful way, it is highly likely that the user community will use the XML data structures and capabilities.

### 7. A Marine XML Consortium

The proposed mechanism for developing a Marine XML Specification is through the creation of an international consortium. In a similar fashion to the development of other XML specifications, it is suggested that there be a membership fee and a number of different levels of participation. The membership fee will assist in defraying the costs of supporting a consortium
and the different levels of participation will support the participation of a broader range of interested groups.

It is also essential that a Marine XML Consortium liaise and cooperate with other standards and XML related bodies. It is more appropriate that a Consortium undertake this linkage with other 'consortia'. It is also important that a marine XML specification does not duplicate existing XML structures. Contact and linkages with similar groups such as the Open Geographic Information System Consortium (OGC) will assist in avoiding duplication of effort and improve the interoperability of a marine XML specification.

The Marine XML specification will also be build on relevant work already undertaken by a number of groups including the work of related sub groups of the ISO Technical Committee 211 and the Open Geographic Information System Consortium with its development of the "Geospatial Markup Language" (GML). GML is being developed in support of the spatial data industry and includes companies that are developing geographic information systems (GIS). It is anticipated that common areas with GML and other existing 'standards' will be utilised as much as possible within the marine XML.

A Marine XML Consortium will need to interact closely with W3C Consortium in order to ensure that the latest XML technology is incorporated into the Marine XML Framework in a timely, controlled and systematic fashion. It is anticipated that this approach will assist in reducing the number of 'varieties of marine XML' developed and provide the maximum benefit of this technology to the maximum number of users. The Marine XML Consortium will attempt to address this problem through the development of an XML based, open framework which could 'mediate' between existing standards and protocols.

8. Proposal

It must be accepted that the development of one or more versions of XML to support the management and exchange of marine data will occur. The IODE Committee needs to decide if IODE: -

- wishes to assist or participate in this process
- should attempt to develop a version of Marine XML alone or
- attempt to develop Marine XML as part of a Consortium of interested groups.

(end of document)