COL/NW/06

A Testbed for Marine XML

FY 2004 Proposal to the NOAA HPCC Program

August 11, 2003

Title Page | Proposed Project | Budget Page |

Principal Investigator: Gregory C. Johnson
Line Organization: OAR
Routing Code: R/PMEL
Address: Pacific Marine Environmental Laboratory
7600 Sand Point Way NE
Seattle, WA 98115

Phone: (206) 526-6806
Fax: (206) 526-6744
E-mail Address: Gregory.C.Johnson@noaa.gov

Charles Sun	Tony LaVoi	Nazila Merati	Nancy N. Soreide
NESDIS/NODC	NOS/CSC	OAR/PMEL	OAR/PMEL
Charles.Sun@noaa.gov	Tony.Lavoi@noaa.gov	Nazila.Merati@noaa.gov	Nancy.N.Soreide@noaa.gov

Proposal Theme: **Technologies for Collaboration, Visualization, or Analysis**

_____	_____	_____
Gregory C. Johnson	Cynthia Loitsch	Eddie Bernard
Oceanographer	Program Support Oficer	Director
NOAA/PMEL	NOAA/PMEL	NOAA/PMEL

A testbed for Marine XML

Proposal for FY 2004 HPCC Funding

Prepared by: Nazila Merati and Gregory C. Johnson

Executive Summary:

Marine XML is an emerging technology for encapsulation of marine data, with the potential of providing an efficient means to store, transfer and display marine data.[1] In the proposed effort, a three Line Office OAR/NESDIS/NOS team will partner with Environmental Systems Research Institute (ESRI) to develop a NOAA testbed for Marine XML utilizing data from the Argo global ocean observing network of profiling floats.

Pacific Marine Environmental Laboratory (PMEL) has a US Argo float group, the National Oceanographic Data Center (NODC) is the official Argo archive center, and Coastal Services Center (CSC) is widely recognized as a NOAA center of excellence for metadata. ESRI is the largest commercial provider of GIS software. This NOAA team will work together with ESRI to develop Marine XML standards appropriate for Argo data exchange and ingest into GIS systems. The PMEL and NODC experience with ocean observations data will assure that our implementation is suitable for other ocean observing system data sets. Argo is a newer ocean observing system with well developed metadata, and the CSC metadata specialist (who is the NOAA metadata consultant) will work with the team to assure that the project develops a standards-based data dictionary and schema that is appropriate for Argo data and extensible for use with other oceanographic data type. The ESRI core development team will work with this project (at no cost) on XSLT conversion routines to the geodatabase and to modify existing data schema to incorporate NOAA observational data into ArcGIS. This is a very strong effort that will be well integrated with the NOAA Enterprise GIS, with two of the investigators (Merati and LaVoi) representing their Line Offices on the working group.

Problem Statement:

Although XML has been used by many industries for data exchange and publishing, and in the marine realm, the Australian Oceanographic Data Center (AODC), Canada's Department of Fisheries and Oceans (DFO) and UNESCO have both published standards on the use of Marine XML for data integration, data processing, data exchange and data publishing, NOAA has not developed Marine XML standards for use for observational platforms. Such standards have the potential to enable the transfer of observational data between data users and into commercial off the shelf (COTS) technology. However, there are numerous unresolved questions about performance related to data file size, network bandwidth and compression requirements.

---

[1]Marine XML is supported by the Marine XML Consortium established by the Intergovernmental Oceanographic Commission (IOC). The European Union very recently initiated a Marine XML project.

Like NOAA, ESRI has observed Marine XML activities with interest, but for them too, this potential has not yet been explored.

Relationship to HPCC objectives: This project addresses the **HPCC goal** to "*improve technology for access to critical data, information and unique resources in a manner that increases mission effectiveness and furthers NOAA's service to the nation*".

As part of the "**Technologies for Collaboration, Visualization, or Analysis**" theme, the proposal addresses the goal that "*Applications and tools proposed under this theme should be designed using a framework approach that enables the software to be integrated with other related packages and interfaced with major off-the shelf software products.*".

Further goals that this proposal addresses in the "**Technologies for Collaboration, Visualization, or Analysis**" theme are to "*explore the use of leading-edge E-government framework technologies such as Extensible Markup Language (XML) and standard schemas such as Electronic Business using eXtensible Markup Language (ebXML)*", and "*demonstrate new techniques for working with NOAA data and information*," and "*develop tools that are extensible, scalable, and available for easy deployment throughout NOAA*.".

Proposed Solution:

We propose a NOAA-ESRI testbed for Marine XML, an important emerging technology, to be implemented by experienced and capable partners with an important and preeminent NOAA data set. This effort will be an important step towards addressing the unresolved issues around the use of Marine XML, and one which will be watched with great interest by the NOAA Enterprise GIS group.

Argo float data are stored in netCDF, the Argo standard for file storage. NetCDF files are used by many oceanographic projects and scientific applications for display and analysis, many COTS software products are unable to read in netCDF files. Although some ad-hoc methods have been used to bring some data sets into a format compatible with GIS, the format barrier poses as an obstacle for many wishing to bring in observational oceanographic data, but do not have sophisticated programming skills to write file translators. An example of this is standard GIS software packages, where most use a proprietary format. However, the newer versions of ArcGIS and other GIS software are able to handle XML as an input format, and ESRI is looking seriously at Marine XML. The advent of XML will make data transfer from the data sources to the COTS software easier.

We plan to investigate the use of Marine XML frameworks for the description of Argo Float data. The float data serves as a good testbed since the data set is of a modest size, consists of both profile and drift data and is well documented. We will analyze the current Marine XML schemas that have been developed by AODC, UNESCO and DFO. We will determine if we can implement the existing schemas or have to create new schemas for the Argo float data. When a consensus has been reached, NODC and PMEL scientists will work with NOS/CSC's metadata specialist to build a data dictionary and schema that will describe the Argo Float data. In the process, we plan to create a schema that can be extended to work with other NOAA observational data types (XBTs, SeaCAT, CTD).

Once the schema and data dictionary are created, the Argo Float data will be converted from the existing netCDF format to XML using a netCDF to XML translator and then to Marine XML using XSLT. The Marine XML encoded Argo Float data will be validated with standard XML syntax checkers and refinements will be made to the Marine XML schema.

We will then investigate incorporating Marine XML wrapped Argo float data into ArcGIS 9. This may require the ESRI core development team to work with NOAA personnel to look at XSLT conversion routines to the geodatabase and to modify existing data schema to incorporate observational data into ArcGIS. ESRI has offered the technical services of their core development team at no cost to NOAA to work on this issue.

While XML serves as an excellent method to transfer data between systems and software packages, we will be handling quite large data sets. We will also assess the best methods for packaging the XML data dictionary, schemas and files to optimize file transfer times over the NOAA network using the connections between PMEL, NODC and NOS. We will test traditional and Internet 2 connections to compare transfer times.

When a final schema has been established, the project will make the schema, documentation, any code developed and style sheets available to interested NOAA users. The investigators will also provide the HPCC panel and NOAA with a best practices guide to writing Marine XML schema for NOAA observational data systems. This project will provide a new opportunity for spatial data sharing through standards technology and foster collaboration within the NOAA user community.

Give a list of the major activities being considered in this project

1. Work with NODC and PMEL scientists to identify key Argo data sets to be used for framework

2. Using the data specifications from example libraries, determine optimal parameters for a Marine XML dictionary to represent Argo data

3. Build schema for Argo XML, have NODC and PMEL scientists look over schema and Marine XML dictionary definitions

4. Test schema and Marine XML with Argo data set, assess bandwidth versus file size issues for file transfers, test feasibility of web transfers of files over the network to NESDIS and NOS

5. Load XML encapsulated Argo float data into ArcGIS 9 and personal geodatabase, work with the ESRI ArcGIS development team to refine and enhance XSLT code to make loading of Argo float data seamless

6. Test performance of XML-Argo float data in a geodatabase for analysis and mapping

7. Based on performance of XML handling, file size and file transfer quality and speed, fine tune schema and file transfer options

8. Test metadata creator in ArcGIS 9 for compatibility with Argo Marine XML schema, incorporate changes to document dictionary

9. Provide schema to NOS and NODC to modify with other marine observation platforms, work with developers to determine what changes need to made to the schema to support other NOAA observational projects

Analysis:

This project will allow NOAA to take advantage of the efficiencies of Marine XML as a data transfer standard. Working with ESRI at the beginning of their integration of Marine XML into ArcGIS will ensure that their implementation serves NOAA's data needs.

Appropriateness: This proposal addresses more than one HPCC networking goal as well as some of the larger goals in the HPCC program. The Marine XML schema and framework developed under this proposal can be extended for other NOAA observational platforms. The ability to output observational data in an XML format will allow NOAA data to be incorporated into COTS software for use by NOAA personnel as well as provide data centers a method of distributing the data. The proposal joins links HPCC program with ESRI and with the NOAA Enterprise GIS efforts by making geospatial data accessible. Additionally, this proposal supports NOAA's efforts in the Geospatial One Stop efforts.

Technology: This proposal will utilize the current technologies to create XML schemas and Dictionaries. It will extend and enhance existing Marine XML dictionaries and will build specific dictionaries for Argo float data. Incorporation of XML encoded data into ArcGIS and the ArcGIS suite, will require ESRI developers to assist NOAA to determine the best translation methods. The proposed solution will also test existing NOAA networks to test data transfer times and data quality between the line offices involved in the project.

Scope: The scope is wide, given that the partners in the proposal span three NOAA line offices (OAR, NESDIS, NOS) working with ESRI on a prominent NOAA data set. The effort is well integrated with the NOAA Enterprise GIS effort initiated recently by the NOAA CIO. If successfully implemented, the XML schema and framework can be applied to other NOAA data sets and used to transfer data to NOAA users without the overhead of data translation. As the Marine XML framework and schema matures, we propose using it as early as equipment calibration and deployment, giving the scientist a better way of tracking the observational instrument throughout its life history. NOS and NCDDC have the GIS and data base expertise within NOAA to optimize the use of the Marine XML framework to provide GIS users best ideas for data implementation and serving via web portals and applications.

Leverage: This project leverages off of results that other Marine XML projects that have developed. AODC and DFO have built Marine XML schema that we will use as a starting place, though we will have to customize the schema to work with NOAA's observational platforms. Unidata has written a netCDF to XML translator that we will be able to use for the creation of the XML files. PMEL developers have extensive background with XML technology and standards. Currently PMEL is implementing a FY2003 HPCC proposal "Utilizing XML/SOAP to provide machine-to-machine interface with TAO data". This project could utilize the XML/SOAP protocols tested in this project to send the Marine XML schemas via the internet. This would be a good test of the performance of the XML encoded Argo float data. As NOAA

adopts GIS as a method to manage, display and serve NOAA's geospatial data holdings, adoption of Marine XML as a data transfer method as well as a standard, will make the incorporation of observational data into NOAA's geospatial framework easier. Additionally, NOS/CSC and PMEL have expertise with ESRI's data model method for GIS data storage and distribution.

Cost Benefit: The benefit to cost of this project is high. The development and adoption of a Marine XML framework for NOAA observational data will make the management and transfer of the data between users easier. The ability to incorporate XML encoded observational data into ESRI GIS software will reduce the amount of developer time to create translators will decrease the amount of effort and time to make data available in a GIS format. The availability of observational system data in a GIS formats will enhance the sharing and delivery of spatial data between NOAA offices, the public at large, this is one of the tenets of the NOAA Enterprise GIS effort.

Compare the selected plan with other alternatives:
The alternative to this proposal would be to continue to distribute the Argo float data as netCDF files. If NOAA observational data are to be incorporated into GIS software for display and analysis, NOAA researchers will have to create file translation software for data integration into these systems. We will also be unable to take advantage of the efficiencies provided by XML.

Performance Measures:

This project will be successful if it creates a NOAA Marine XML schema and data dictionary and demonstrates the utility of Marine XML for use with ocean observation data exchange and ingest into GIS. Another measure of success is the integration of the Argo float XML files into ArcGIS 9 and the geodatabase for display.

Milestones

Month 01 - Identify data set to use for XML schema
Month 01- Review AODC's and DFO's XML schemas and data dictionaries, determine their suitability for use
Month 02- Create new schema, framework and data dictionary, work with CSC metadata specialist for syntax
Month 03- Install netCDF to XML translator
Month 04- Test XSLT translator from XML to marine XML s
Month 05- Test with Argo Float data set, run validating software
Month 06- Test performance of XML encoding, optimize file size if necessary
Month 08- Incorporation of Argo Float data into ArcGIS 9 , work with ESRI core developers
Month 12- Final schema delivered to users with data dictionary and recommendations for implementation

Deliverables

Provide a list of the final products from this project

- Marine XML schema, data dictionary for Argo Float data
- Processing pathway to take Argo data from netCDF to Marine XML
- Tools to ingest Marine XML to ArcGIS