12.2 IMPLEMENTING MARINE XML FOR NOAA OBSERVING DATA AND GIS INTEGRATION

Nazila Merati and Eugene F. Burger
PMEL/JISAO Seattle, WA
Charles Sun
NOAA/NODC Silver Spring, MD

1. INTRODUCTION

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]
XML has been used by many industries for data exchange and publishing. In the marine realm, the Australian Oceanographic Data Center (AODC), Canada's Department of Fisheries and Oceans (DFO) and UNESCO have published standards on the use of Marine XML for data integration, data processing, data exchange and data publishing. These agencies have 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.

In 2004, NOAA's Pacific Marine Environmental Laboratory (PMEL) was allocated funds to test the feasibility of using MarineXML as a way to transfer Argo profiling float data from its native format into an ESRI geodatabase - using tools and software developed by ESRI and NOAA. This project is be the first time that NOAA/PMEL real time observation data would be fully incorporated into a geodatabase. The project also involves several groups at NOAA working to standardize, archive and build a MarineXML format that will work with real time observing data. The project also tests the ability of NOAA to distribute real time data using XML and SOAP protocols.

2. BACKGROUND

The Argo program is a broad-scale global observing program of temperature and salinity floats. Argo floats have been deployed throughout the world's oceans and compliment the Jason altimeter mission. The Argo floats are designed to drift at a fixed pressure (usually 1000 db) for ten days. After this period, the floats move to a profiling pressure (typically between 1000 and 2000 dbar) then rise, collecting profiles of pressure, temperature, and salinity data on their way to the surface. Once at the surface, the floats transmit the data collected via a satellite back to data centers while float drift data are at the same time calculated by satellite. After a day the float submerges again and repeat their mission. The floats have a nominal lifetime of five years, and will yield valuable information about large-scale ocean water property distributions and currents, including their variability over time scales from seasonal to the duration of the array. As of November 2004, there were 1454 active Argo Profiling floats transmitting information to the various global data centers.

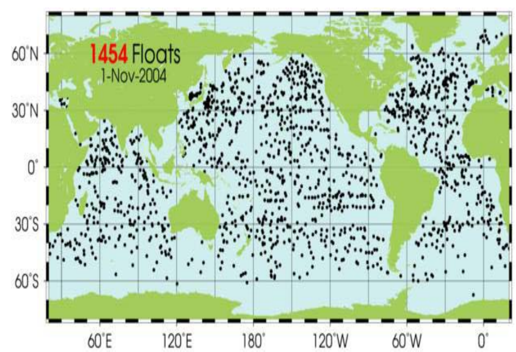

Fig. 1. Active Argo floats as of 1 November 2004. Map courtesy of the Official Argo Information Center, http://argo.jcommops.org/

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 but, unfortunately, 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 who wish 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

[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.

versions of ArcGIS and other GIS software are able to handle XML as an input format. The advent of widespread XML use will make data transfer from the data sources to the COTS software easier.

3. RESULTS

We investigated the use of Marine XML frameworks for the description of Argo Float data. The float data serves as a good test bed since the data set is of a modest size, consists of both profile and drift data and is well documented. During the course of this project we analyzed current Marine XML schemas developed by AODC, UNESCO and DFO. To date, we are using the AODC XBT schema as a starting point for the the Argo float data. At the same time, we are building 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 profiling data types including Expendable Bathythermographs (XBT), SeaCATs and Conductivity, Temperature and Depth recorders (CTDs). These are all profling instruments that store data in netCDF and are used by other NOAA research groups.

As the schema is tested and modified 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 are investigating incorporating Marine XML wrapped Argo float data into ESRI's ArcGIS 9. XSLT conversion routines currently are being developed at ESRI and PMEL to modify existing data schema to incorporate observational data into ArcGIS geodatabase. The software may need to be modified to work with datasets with multiple z and t values to make data transfer easier.

4. DISCUSSION

While XML serves as an excellent method to transfer data between systems and software packages, we will be handling quite large data sets. NOAA programmers are assessing 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 NOAA's laboratories. PMEL's Tropical Ocean Array (TAO) project investigated the use of MarineXML for their web services data access to TAO moored buoy data, but found limitation within Marine XML to describe interrupted or incomplete data sets within the Period tag. They decided to use their own unique XML to deliver the data.

Other community-based standards for XML that the researchers have investigated include Borehole XML used by the geo-technical community that addresses the multiple z values that occur with in-situ observational data types, Sensor Markup Language (SensorML), and Earth Sciences Markup Language (ESML).

5. CONCLUSION

When successfully implemented, the XML schema and framework will be applied to other NOAA data sets and used to transfer data to NOAA users without the overhead of data translation, which has slowed down data exchange between researchers. 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. The development of an XML standard for NOAA's observational data will simplify the translation of legacy data into existing and evolving GIS formats as well as simplifying data exchange between various programs.

We are currently looking at the newly emerging Sensor Markup Language as an option to MarineXML. SensorML compiles with the OpenGIS Consortium standards for web mapping and may be able better suited for longer duration observations. The development of XML standards to describe Argo float data will prove useful as data centers move towards web services for data distribution. Additionally, the ability to smoothly transfer in-situ data from native forms to ESRI geodatabases and into data models will allow scientists to collaborate on global observational data issues.

6. ACKNOWLEDGEMENTS

This publication was supported by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement #NA17RJ1232 , Contribution #1101 and PMEL contribution #2766. The views expressed herein are those of the author(s) and do not necessarily reflect the views of NOAA or any of its sub-agencies. This work was funded by NOAA's HPCC program. The authors would like to acknowledge the Argo Program and the TAO project at PMEL for the use of their data for testing.

Primary Author:

Nazila Merati
Research Scientist
NOAA/OAR/PMEL
7600 Sand Point Way NE
Seattle, WA 98115
Nazila.merati@noaa.gov