COL/NW/07

Expanding the TAO Portal with OPeNDAP access and enhanced Web Services

FY 2005 Proposal to the NOAA HPCC Program

September 13, 2004

| Title Page | Proposed Project | Budget Page |

Principal Investigator: **Michael J. McPhaden**

Line Organization: OAR
Routing Code: R/PMEL
Address:

NOAA/PMEL
7600 Sandpoint Way
Seattle, WA 98115

Phone: (206) 526-6783
Fax: (206) 526-6700
E-mail Address: Nancy.N.Soreide@noaa.gov

Co-investigators:

Eugene F. Burger	Joe Sirott	Nancy N. Soreide
Eugene.Burger@noaa.gov	Joe.Sirott@noaa.gov	Nancy.N.Soreide@noaa.gov

Proposal Theme: **Technologies for collaboration, visualization or analysis**

Michael J. McPhaden	Dennis W. Moore	Cynthia L. Loitsch	Eddie N. Bernard
Oceanographer	OC Division Leader	Program Support Officer	Director
NOAA/PMEL	NOAA/PMEL	NOAA/PMEL	NOAA/PMEL

Expanding the TAO Portal with OPeNDAP access and enhanced Web Services

Proposal for FY 2005 HPCC Funding

Prepared by: Eugene F. Burger, Joe Sirott and Nancy N Soreide
and Michael J. McPhaden

Executive Summary:

The TAO network of moored buoys in the Pacific Ocean forms the observational cornerstone of the El Nino Observing System, while the PIRATA network of moored buoys in the Atlantic Ocean is used to study ocean-atmosphere interactions in the tropical Atlantic. The TAO and PIRATA Web Service, funded by HPCC in FY03[1] allows users to automate their data extraction processes through unattended machine to machine interaction. During development and testing of this service, the TAO/PIRATA project identified significant new requirements for the TAO data access portal. First, we will enhance the TAO Web Services to improve collection of TAO performance metrics by adding user authentication, allowing data managers to contact users, maintain a database of portal usage and add an additional layer of security. We will expand the data logging capabilities of the Web Services server to include information such as which data variables were selected, the geospatial region selected, and the user who requested the data. Secondly, we will data from the TRITON buoys in the western part of the TAO/TRITON Array into the Web Services database. Thirdly, we will interface the "dapper" in-situ data server to the TAO Web Services database, making TAO data available for the first time through the increasingly important OPeNDAP protocol to clients such as ncbrowse and LAS. This effort integrates components from two previously funded HPCC projects[1] [2] into an operational system for important NOAA datasets.

Unlike most in-situ OPeNDAP servers, the "dapper" server provides comprehensive metadata, including space/time distribution of the data collection and of the individual station files within the collection, as well as information about which variables are present in each of the station files. Although this proposal is completely independent, the installation of a "dapper" in-stu data server is an important component of providing the in-situ dataset on the Grid with OGSA Grid services. For details see the pending FY05 HPCC proposal "GridDapper: Putting NOAA in-situ Data on the Grid".

[1] FY03 funded proposal "Utilizing XML/SOAP to provide a machine-to-machine interface to TAO Data" (Soreide, et. al.)
[2] FY01 funded proposal "A Climate Data Portal" (Soreide et. al.)

Problem Statement:

Problem details: The TAO Web Services system provides access to TAO and PIRATA data by exposing an API that can be interrogated by a user created web service client. For these web services to be fully implemented in a production environment, data managers need additional data usage logging capability as well as a user authentication capability. User authentication would require TAO Web Service users to pre-register their clients and embed a unique identifier in their client. In tandem with the additional usage logging capability, this would allow data retrieval activity to be associated with a web service client identifier. One advantage of this added capability is that it would allow users, as is now the case with the TAO data Access web page, to be notified of data revisions which affect data they have previously downloaded.

Although web services do not allow any more access to a web server than a web page does, a web service is different in that a software interface is exposed. By using the web service, a client will invoke a method (or function) on the server. In the TAO web service context this method will deliver data in XML format to the client. Since this action is automated it a possible that a process could place an undue burden on a server by rapidly repeating large data requests, either through intentional (malicious) or unintentional actions. By funding the addition of the additional logging will allow system administrators to track high usage clients and alert users.

The TAO Array became the TAO/TRITON Array to recognize the introduction of twelve TRITON buoys in the western Pacific by the Japanese agency JAMSTEC (covered by a PMEL-JAMSTEC Memorandum of Understanding). The basis of that agreement includes seamless delivery of data from the TAO/TRITON Array as a unified and integrated dataset, but the TAO Web Services do not include TRITON buoy data.

Finally, the OPeNDAP protocol has been selected as a "pilot" data transport component in the Data Management and Communications (DMAC) plan for the U.S. Integrated Ocean Observing Systems (IOOS). Serving data in the "dapper" OPeNDAP protocol from the TAO Web Services database will allow near real-time TAO data to be part of the emerging IOOS data management system. It also provides a significant component of the effort required to put real-time TAO data on the Grid with OGSA Grid Services.

Relationship to HPCC program objectives: The OPeNDAP component of this proposal addresses the HPCC theme of "**Technologies for Collaboration, Visualization, and Analysis"** by enhancing **distributed data integration** and incorporating **advanced communications protocols**. The logging and user authentication component is an augmentation of an existing application and due to the enhanced logging capabilities, this effort will address the primary goal of the NOAA HPCC to "*support information technology research and development that enables NOAA to provide greater access to its vast holdings of real-time and historical to users in a more complete, more usable form and much more timely manner through the increased use of advanced technologies associated with the Internet*."

Proposed Solution:

Synopsis: We will develop enhanced logging and user authentication capability for the TAO Web service as well as an OPeNDAP interface. This involves the following tasks:

The first task is user authentication. Before users can connect their client to the TAO Web Service, they will have to get an identifier key from a TAO user authentication site. The user will get the key from an authentication web site where the user will enter information that will allow the data managers to contact this user should they post any data revisions or find a particular client to be placing undue load on the web services server. This key will be embedded as a parameter in all method calls to the TAO Web Service and will be logged with the data retrieval activities. The parameter field for the identifier has already been added to the TAO web service code, although the authentication has to be added. The web service client information will be saved to a database created for this purpose. Clients that do not pass a valid key in a method call will be denied access to the TAO Web Service.

The second task is the data logging capability that will be added to the existing TAO Web Service code. This will enable logging of the data parameters requested, the period for which the data was requested along with the date, time and the client key. If no client key is associated with this, only the client's IP address will be logged. All this information will be written to a MySQL database.

The third task is an internal web site that will provide access to the logging database for TAO data managers and system administrators. The website will have a reporting capability that can be used to view TAO web service usage.

Next, we will integrate TRITON data into the TAO Web Services database, assuring that the Web Services and OPeNDAP servers serve the unified, integrated TAO/TRITON dataset.

Finally, we will interface the Web Services database to the Dapper OPeNDAP in-situ data server. Dapper , an OPeNDAP interface to the Climate Data Portal (CDP) (Soreide, Kilonsky, and Sun, FY00 HPCC Funded Project titled with *"A Climate Data Portal"*), will be modified to serve data in the OPeNDAP protocol from the TAO Web Services database.

Technologies applied: The TAO Web Services code and Dapper are written in Java, and all enhancements will be written in JAVA. Database connectivity to the MySQL database will be though JDBC.

Analysis:

There is no way of determining who is using the current system and what resources are being used. There is no way of notifying users if data that they have downloaded have changed. This leaves the system vulnerable to abuse and also makes it difficult to customize the system to best suit the interests of our users. Our proposed enhancements will improve performance tracking for the TAO project by correcting these problems by giving the TAO project the capability to track data usage more closely, alert users to data changes, and stop malicious users from abusing the system

Installing the “dapper” OPeNDAP interface provides an alternative data transport that will make the TAO Web Services portal compatible with the standards emerging from the DMAC component of the IOOS. It also provides a significant component of the effort required to put TAO data on the Grid with OGSA Grid Services.

Performance Measures:

What factors should be considered in determining the success of this project

Milestones

Month 3 – Create user authentication website and backend
Month 6 – Enable data logging and user web service authentication
Month 6 – Add TRITON buoy data to TAO Web Services database
Month 9 – Add “dapper” OPeNDAP interface
Month 12 – Complete testing and move code to production

Deliverables

Provide a list of the final products from this project

- TAO Web Service User Authentication web site
- Enhanced usage logging and authentication for TAO Web Service
- OPeNDAP interface to TAO Web Service