Cyberinfrastructure and Software Middleware Efforts Applied to Ocean Observing Systems

[1]K. J. Gomes, [2]M. Arrott, [3]R. Butler, [1]D. R. Edgington, [3]M. Freemon, [1]R. Herlien, [4]B. Howe, [3]Y. Liu, [1]T. O'Reilly, [3]V. Welch, [2]F. Vernon, [1]K. A. Salamy

[1]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Rd., Moss Landing, CA 95039; [2]Scripps Institute of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093;
[3]National Center for Supercomputing Applications, University of Illinois, 1205 W. Clark, Urbana, IL 61801; [4]Applied Physics Laboratory, University of Washington, 1013 NE 40th St, Seattle, WA 98105

ABSTRACT

Engineers and scientists at the Monterey Bay Aquarium Research Institute (MBARI), Scripps Institute of Oceanography (SIO), National Center for Supercomputing Applications (NCSA), and the Applied Physics Lab (APL) at the University of Washington (UW) have spent time building technologies to enable science on ocean observatories. From our experience building observing technologies we have learned many valuable lessons about supporting long-term observations consisting of large numbers of heterogeneous instruments. This collaboration is now applying those technologies and "lessons learned" to the next generation of cyberinfrastructure. We will discuss the various cyberinfrastructure technologies, how we applied them to deployed instrumentation and how we are using that existing infrastructure to develop and test other technologies to meet the scalability issues associated with large- scale ocean observatories.

INTRODUCTION

The oceanographic research community faces several challenges in data management: discovering, accessing, and using the data can be a burdensome, obstacle-laden task. Ocean scientists and their staff spend considerable research time simply getting to the stage where they can begin analysis of relevant data. This analysis often involves correlating data readings from multiple sensors in dispersed geographic locations and establishing relationships among them. This often must be done in near real-time to ensure rapid access to observational parameters which enable better science during key observation periods. Timely data access also enables the operations, science, and engineering teams to rapidly change observing strategies in response to observed conditions or experimental needs. Systematic and consistent instrument control will be necessary for this type of flexible observatory. Some examples of the necessary instrument control might be:

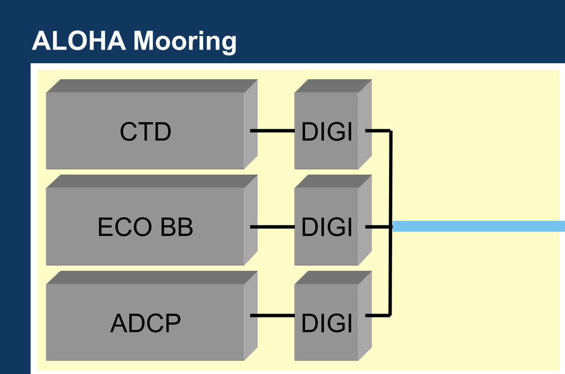
- Increasing or altering data sampling or reporting rates
- Taking operational actions to ensure robust data collection
- Allocating additional sensors to take extra observations to characterize an anomaly.

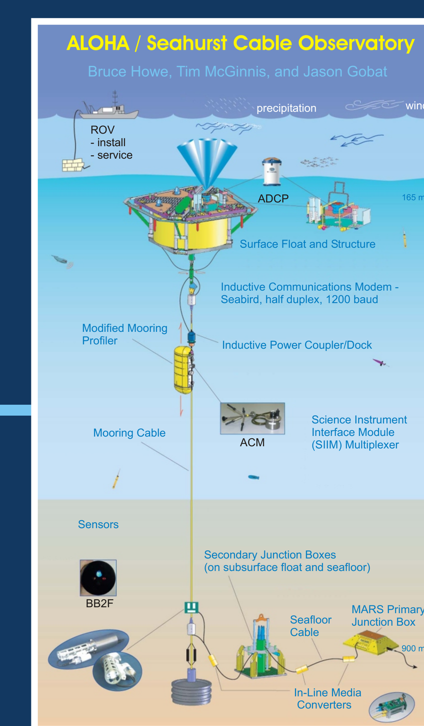
Through the Ocean Observing Initiative (OOI) program, new ocean observatory technologies are being designed to meet these needs. In addition, several other large projects that incorporate ground-based instruments, including the Large Synoptic Sky Telescope, the National Ecological Observatory Network ecological observatory, and the Water and Environmental Research System are being designed and constructed. These large, observatory projects are poised to build independent national scale in-situ and remote sensing cyberinfrastructures to gather and publish "community"-sensed data, and will generate synthesized products based on current needs for a specific targeted community of researchers. Each of these community-owned observatories is defining its own ways to collect information addressing a broad set of issues, and each is building customized mechanisms to generate and publish both derived and raw data to their own constituents. These independent approaches will be inefficient and present challenges both for inter-observatory coordination and for how researchers might efficiently aggregate sensor data from different observatories. However, individual approaches are being pursued because common observatory management middleware is needed that does not yet exist.

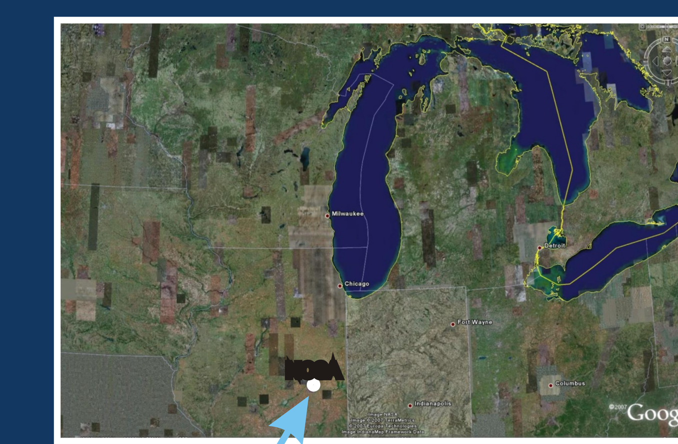
THE APPROACH

Our approach is driven by the belief that the existence of a single, overarching infrastructure to take care of all these common issues of data and observatory usage will accelerate the pace of scientific discovery and enhance our ability to do good science. The approach taken is geared towards the goal of building a working prototype of an Observatory Middleware Framework, designed and implemented to scale to a national level and interoperate with other large environmental observatories for the immediate needs of the ocean research community and the use of all earth observatories.

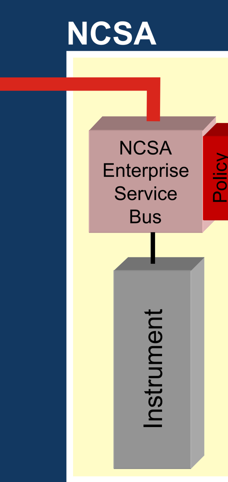
In our approach, we focus on two areas; firstly the evaluation of Enterprise Service Bus (ESB) technology for the federation of instruments deployed within different observatories, and secondly the design and development of crosscutting/ubiquitous observatory functionalities including governance and policy enforcement and their application on the ESB. In such a federated system, researchers may readily integrate existing and new services such as analysis engines, event detectors, and workflow engines for their own unique applications. For the federation of instruments we envision a framework that supports the federation of data streams from multiple observatories, creates common instrument access and management, provides fundamental mechanisms to enforce policies, captures and forwards rich metadata including data provenance, and provides mechanisms to handle the processing of data and events in a uniform way. This is being done through a largely integrated effort consisting of a large collection of existing middleware, sensor infrastructures, and software technologies in novel ways to produce a usable research environment that can be utilized by the driving environmental research applications in oceanography, ecology, and environmental engineering.

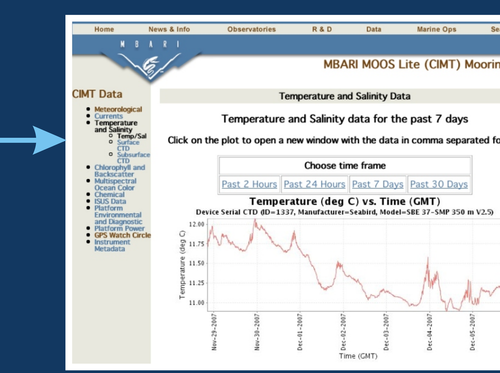




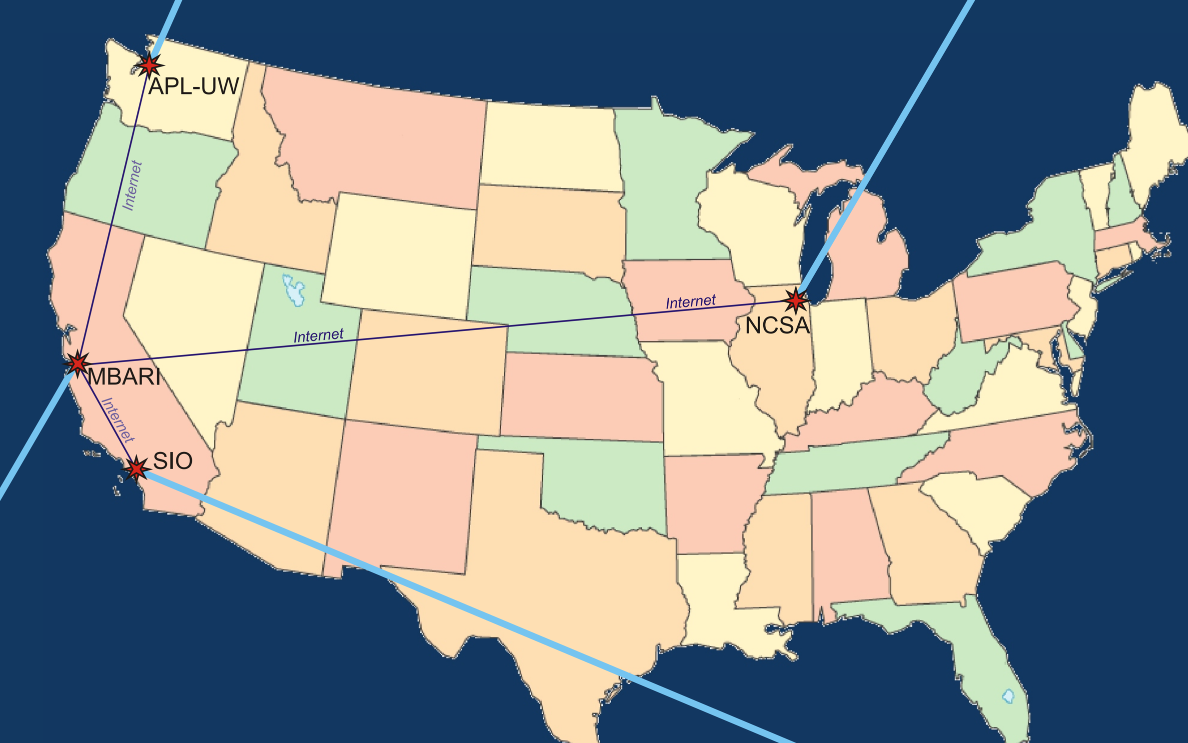


Enterprise Service Bus





Client



OBSERVATORIES

In order to effectively test our approach, the pilot deployment of the OMF will occur across five instrument installations (some of these being observatories) and three Enterprise Service Bus locations as examples for the broader observatory community. The instrument locations and observatories consist of the MOOS observatory at MBARI, the NSF MARS Observatory, the ALOHA / Seahurst Cable, a directly connected instrument at NCSA, and an instance of a ROADNet instrument at the California Institute for Telecommunications and Information Technology (CaliT2) at UCSD. MOOS is MBARI's mooring-based observatory that consists of three instrument nodes on a surface expression (mooring) and two benthic instrument nodes connected to the surface expression via optical electro-mechanical cable. The MARS cable is an NSF-sponsored cabled test bed that will be used to test various instrument deployments on cabled observatories. Serving as a test bed, the ALOHA / Seahurst cable is a proving ground for Bruce Howe's (University of Washington) ALOHA mooring that will eventually be deployed on the MARS cable. NCSA is developing an instrument that will be used to test instrument deployments directly on the ESB and in the future we will be integrating an instance of UCSD's ROADNet system.

FRAMEWORK (Enterprise Service Bus)

For the framework, we envision software which builds on an ESB and allows researchers to subscribe to data of interest and issue commands to instruments across multiple observatories and middleware. The ESB technology then provides the framework with the capability to apply interceptors that inject cross-cutting functionality such as ubiquitous services that enforce security and policy on raw data streams, events, and derived data. An ESB provides foundational services for complex architectures that can rapidly link sensors and processes regardless of their existing messaging protocol and can apply relevant policies across the messages being transported. ESBs are well-suited to integrate these systems because of the performance and scalability characteristics of ESBs, and their ability to interconnect a variety of message systems in a distributed topology, thereby easing the integration of legacy middleware. They have been widely accepted in industry and proven to readily integrate web service, Grid, HTTP, Java Message Service, and other well known message-based technologies.

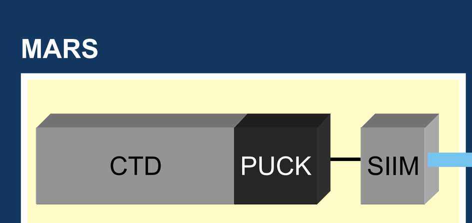
For this project, we plan to leverage the open source Mule ESB implementation which enables:

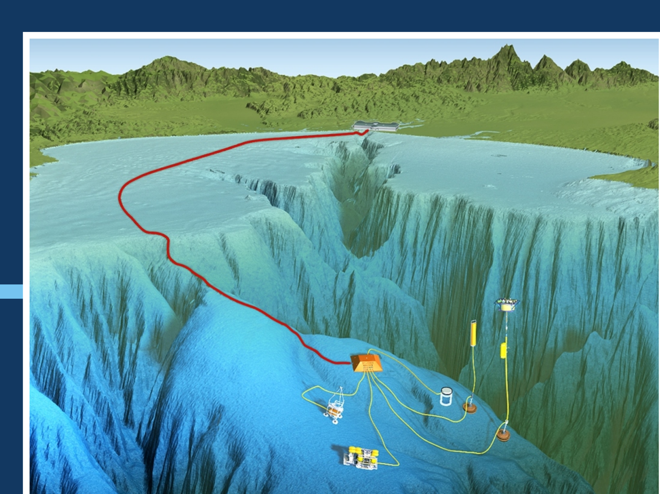
- Pluggable connectivity such as JMS, VM (embedded), JDBC, TCP, UDP, multicast, http, servlet, SMTP, POP3, file, XMPP
- Orchestration of services using Business Process Management (BPM), Mule components and routers
- Support for asynchronous, synchronous, and request-response event processing over any transport
- End-to-End support for routing, transport and transformation of events
- A highly scalable enterprise server using the Staged Event-Driven Architecture (SEDA)
- A REST API to provide technology-agnostic and language-neutral web based access to Mule Events
- Powerful event routing based on well known industry enterprise integration patterns
- Dynamic, declarative, content-based and rule-based routing options
- Improved system security, scalability, availability and robustness

Using the ESB, OMF governance capabilities will manage the flow of control messages and data streams to prevent access by unauthorized parties. The governance framework will support the following functionality: 1) Provisioning, in which instruments will join the OMF with access control policies attached; 2) Delegation, in which owners of instruments will want to delegate their privileges to other users of the network; and,
3) Identity Management, in which, to apply policy, users and instruments will be authenticated and identified.

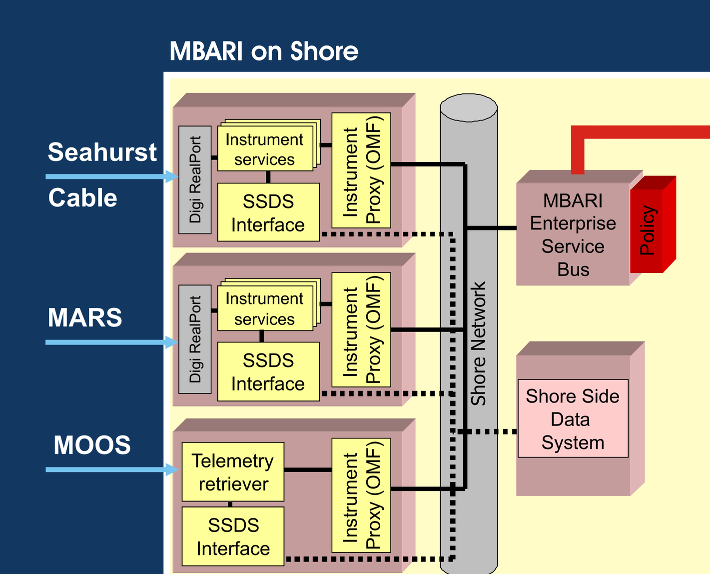
To achieve scalability, we adopt an attribute-based authorization approach, which allows policies to be defined based on the attributes of users and instruments rather than on individual identities only. For authentication and identity management we will integrate proven tools from the Grid and higher education communities, including the Globus Toolkit Grid Security Infrastructure, Shibboleth, GridShib, and GridGrouper. For provisioning and delegation we will leverage Internet2's Signet and SUN's XACML implementation. All of these technologies are based on open standards (SAML, XACML, and X.509) and have open source licenses.

Enterprise Service Bus

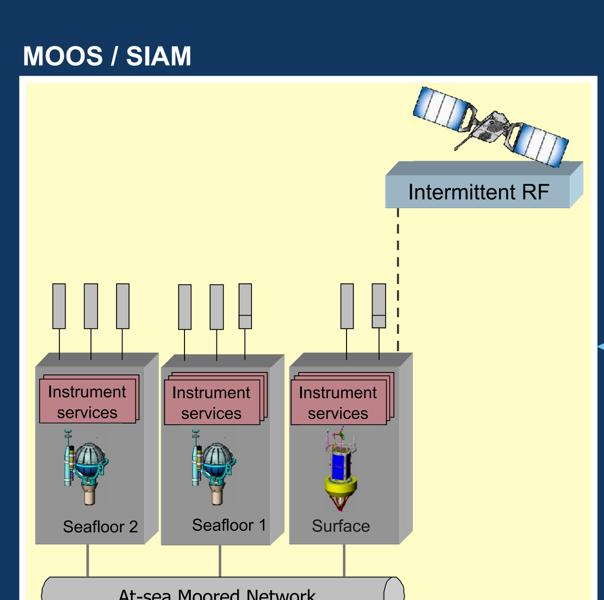




MARS



Enterprise Service Bus



MOOS / SIAM

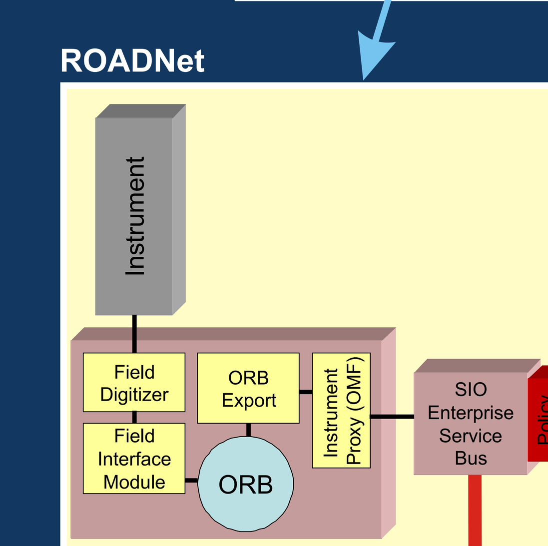
MOOS / SIAM

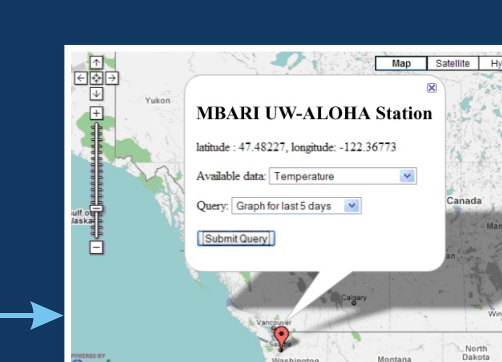
MIDDLEWARE

With these various instrument locations comes the task of integrating multiple middleware technologies that support that instrumentation. These instrument networks utilize different connection strategies and middleware that includes ROADnet, Software Infrastructure and Applications for MOOS (SIAM), and direct, native sensor integration without pre-existing observatory middleware. We will demonstrate the ability to manage multiple instruments on these different networks, capturing raw data streams, governed by a policy engine, from a single graphical user interface.

To accomplish this, a common instrument model (proxy) is being developed to represent the various instruments and middleware on the ESB. These models enable access, control, and semantic interoperability across observing systems, define a working interface specification to the instrument proxy, implement a set of generic instrument command services such as instrument on/off/reset, and enable common data access across heterogeneous middleware.

Enterprise Service Bus





Client