



The design of the NSF Ocean Observatories Initiative Cyberinfrastructure

John Graybeal (1), John Orcutt (2), Frank Vernon (2), Matthew Arrott (1), Oscar Schofield (3), Cheryl Peach (2), Alan Chave (4), Mark James (1), and Michael Meisinger (1)

(1) University of California, Calit2, La Jolla, United States, (2) University of California, San Diego, Scripps Institution of Oceanography, La Jolla, CA United States, (3) Coastal Ocean Observation Lab, Rutgers University, Rutgers, NJ, United States, (4) Woods Hole Oceanographic Institution, Woods Hole, MA, United States

The US NSF's Ocean Observatories Initiative has been funded to collect ocean data for a period of 25-30 years. The project began in October 2009 for a five year construction phase to design, test and build the initial infrastructure for a comprehensive observational system needed to support the collection, dissemination and preservation of data including data directly related to in situ ocean climate observations. The OOI Cyberinfrastructure is a departure from previous data efforts in the oceans. In particular, the data must be open and available in near-real-time with latencies as low as seconds. The original instrument data will be preserved as well as subsequent versions of data, which will include the presentation of data in engineering units (e.g. temperature) and data that have gone through a series of QA/QC tests and changes. The various versions of data will preserve metadata on the processes applied providing data provenance into the future. A scientist thirty years from now will be able to access any of the various versions of the data and, if necessary, apply new QC processes to the data. Users of the data will be able to construct their own versions of the OOI observatory concentrating on the data streams of greatest meaning and, in turn, will be able to publish data on the OOI CI relevant to the individual's or team's interests.

The OOI has constructed a Data Distribution Network (DDN) in collaboration with NOAA IOOS, which will provide data in the formats most typically used; the OOI is not creating yet another data standard, but adapting the available data to the oceanographic community's current and growing practices. The DDN is presently running on Amazon's Elastic Compute Cloud (EC2); it is possible that many future oceanographic computations and particularly numerical models will be running on this and similar cloud systems to provide needed elasticity in scaling computing needs for dealing with events. Metadata associated with all data will allow the discovery of data and instruments of interest. In addition to the data resulting from the deployment of new platforms and sensors, the CI requirements include the development of interfaces with other programs and repositories including NOAA and the WMO. The OOI Cyberinfrastructure has designed an architecture that minimizes a metric we refer to as Time to Science (TTS) to remove as many impediments as possible to ready data access. Release 1 of the software/middleware will be available in Summer 2011.

The OOI CI's Integrated Ocean Network allows not only data return from sensors, but provides the connectivity to control individual platforms (e.g. a cabled network) and sensors (e.g. surface pH). The connections, while made over the network (cables and satellites), employs message passing for data and commands mediated by the Advanced Message Queuing Protocol (AMQP). The ocean network includes a 10Gbps continental facility for network and data management. Connectivity to Europe through Starlight in Chicago and two ports in the western US to Asia can be activated in the future to support potential global collaboration in data, sensor, and modeling activities.