

Overview of NSF's Ocean Observatories Initiative's Cyberinfrastructure Component

Science User Requirements

Almost two centuries of ship-based expeditionary research and three decades of satellite observations have successfully revealed basic descriptions of the oceans and their interactions with terrestrial and atmospheric systems. As understanding has advanced, it has become increasingly apparent that many critical processes occur at temporal and spatial scales that cannot be effectively sampled or studied using these traditional tools. Ship-based studies, for example, are limited in their ability to investigate episodic events or to study dynamic systems that change over time periods longer than a few months. The signal levels associated with some phenomena (e.g., secular changes in temperature of the ocean or the motions of tectonic plates) are small compared to background fluctuations, and require measurements over months to years just to resolve a signal. Some processes occur only during certain times of the year or during energetic events, such as storms, when measurements from shipboard platforms are simply not possible. Therefore, the National Science Foundation has begun to implement the Ocean Observatories Initiative (OOI) to advance science, technology, education, and public awareness by building an interactive suite of ocean observatories.

The OOI will provide users with the means to explore the oceans for decades through three interconnected systems spanning global, regional and coastal scales. The global component addresses planetary-scale problems via a network of moored buoys linked to shore via satellite. A regional cabled observatory will 'wire' the Juan de Fuca plate in the Northeast Pacific Ocean with a high speed optical and power grid serving a wide range of tectonic, physical oceanography, hydrothermal vent and biogeochemistry issues. The coastal component of the OOI will expand the growing set of national coastal observing assets and provide extended opportunities to adaptively study high frequency forcing in coastal environments. These systems are to be bound together by a Cyberinfrastructure (CI) backbone that will integrate the OOI observatories, and associated sensors and models, into a coherent system-of-systems with full two-way communication capabilities to facilitate direct interaction with the oceans. Additionally, the CI allows the existing network to also access a diversity of other (i.e., non-OOI) data streams and provide a coherent 4-D view of the ocean for any researcher. The completed OOI will allow anyone to interactively view any element of the global system, and mirrors ongoing international efforts by Canada, Japan and Europe.

The science to be addressed by the proposed ocean observatories will span many themes including climate change and biogeochemical cycling; ecosystem dynamics, turbulent mixing and biophysical interactions; and fluids & life in continental margin sediments. The broad range of science questions will be addressed with a distributed network of deep ocean moorings, a plate-scale Regional Scale Nodes (RSN) element, a long term coastal time series Endurance array, and a coastal adaptive sampling Pioneer array. The distributed network of hardware will enable science that spans time and space scales beyond any of the individual components of the OOI through an advanced CI capability that allows any actor two-way interactivity, command & control, and data discovery through both real-time, and historical data archives. The user requirements of the CI are diverse, and vary with specific science

needs that have been identified in a series of NRC and community science documents. Specific science goals that will require CI capabilities include:

□ *Data discovery and acquisition from real-time data streams and historical distributed data archives.*

The ability to utilize real-time and historical data will be central to enabling hindcast and data assimilative nowcast and forecast modeling efforts focused on a range of science problems such as A) What are the temporal dynamics in ocean mixing processes? What is the dissolved and particulate distribution and speciation of trace elements, nutrients, and carbon species in the water column? What is the sequestration of carbon from the surface ocean to the deep sea? These questions will require aggregation of all archived data distributed between OOI components and national/international science agencies to provide sufficient spatial data for hindcast studies of the coupling between the physics, chemistry, and biology of the water column. Real-time data required by data-assimilation numerical models will be used to conduct predictive skill experiments utilizing the well-sampled ocean of the OOI environment to assess which models best predict ocean conditions. At the heart of these questions is the requirement for spatial time series to statistically characterize ocean processes over a range of spatial (millimeters to tens or hundreds of kilometers) and temporal scales (seconds to decades) to resolve everything from turbulence to seasonal and interannual dynamics. Examples of non-OOI data streams might include ocean data collected by NOAA's Integrated Ocean Observing System (IOOS) or ARGO's program, satellite data products collected by NOAA-NESDIS, NASA, and/or international science partners.

□ *Adaptive sampling:* One of the powerful capabilities of the OOI is the ability to adaptively sample the ocean based either on model forecasts or on real-time data. This will be a new capability for oceanography that will allow scientists to study episodic events and/or explore unexpected processes in depth. Questions that this OOI capability will be absolutely key for include: What are the air-sea exchange processes that occur during severe weather events? What processes determine the transport of carbon, nutrients, and planktonic organisms in coastal ocean ecosystems? What are the key processes that drive the sequestration of material from the coastal ocean to the deep sea? What is the shelf-slope exchange of heat, salt, and nutrients, and what is the corresponding impact on the marine food web? For example for shelf slope exchange studies the decorrelation length scales are equal in both the alongshore and cross-shore directions, traditional sampling cannot resolve the appropriate scales, and the available data hinder numerical studies of the potential key processes regulating the exchanges. Therefore, the array needs to adaptively sample the shelf/slope exchange of heat, salt, and biological material in order to identify and quantify the dominant driving mechanisms. Studies on the shelf-slope exchange processes will be a focus of the Pioneer Array consisting of moorings, autonomous underwater vehicles (AUVs) with docking stations, and gliders. Another example is use of the RSN to study earthquake and vent phenomena where mobile assets might be dispatched to collect spatial maps over vents fields triggered by tectonic activity. The CI network will allow researchers to adaptively sample the environment using a variety of mooring and AUV systems that will need to be operated collectively to provide the real-time ability to access data from the ocean, and adjust either the sampling from individual sensors on a mooring or to initiate the coordinated movement of multiple mobile assets. The CI provides an easy-to-use, real-time framework for scientist actors to adjust their initial sampling strategy, which is typically based on preconceived notions, when unexpected events or processes are encountered. These undiscovered processes are likely to be the norm during the OOI field campaigns, as the oceans remain chronically under sampled.

□ *Automated ocean surveillance for event detection and classification procedures.* Traditional attempts at adaptive sampling are driven with scientist actors initiating and guiding changes in the sampling. The CI will expand on this by providing a capability for the network to initiate sampling in an automated fashion based on real-time data. The CI will allow scientists to program behaviors that then drive the network responses to events. This requires a robust command and control capability. The network behaviors will be initialized by data or derived data products. For example, recent work by NASA is developing bioinformatics algorithms that combine all available data streams to objectively classify unique water masses that could conceivably initiate increased chemical and biological sampling by the OOI assets as new water masses pass into the study area. Science questions that will rely heavily on this capability include: What are the air-sea exchange processes that occur during severe weather events? How do storms drive benthic resuspension processes, and do these resuspension events play a critical role in initiating phytoplankton blooms? What physical and biological processes lead to hypoxic conditions in the coastal ocean? What regulates the transport and transformation of nutrients on continental shelves? For example, wave and optical data from OOI moored sensors might be used to detect the onset of storm-induced mixing and resuspension. This would automatically trigger both vertical profilers and mobile assets to study the time-dependent spatial evolution of the resuspension and corresponding transport in the water column. The automated activity may include single sensors or swarms of mobile assets, all of which have been provided appropriate behaviors for the mission at hand.

OOI Guiding Principles

The design goals stated in the National Research Council report, *Enabling Ocean Research in the 21st Century: Implementation of a Network of Ocean Observatories*, provide a general summary for the community vision of the OOI's capabilities:

- Continuous observations at time scales of seconds to decades
- Spatial measurements from millimeters to kilometers
- Sustained operation during storms and other severe conditions
- Real-time or near-real time data as appropriate
- Two-way data transmission of data and remote instrument control
- Power delivery to sensors between the sea surface and the sea floor
- Standard sensor interface protocols
- Autonomous underwater vehicle (AUV) docks for data download and battery recharge
- Access to deployment and maintenance vehicles that satisfy the needs of specific observatories
- Facilities for instrument maintenance and calibration
- A data management system that makes data publicly available
- An effective education and outreach program

Thousands of science questions can be investigated using OOI's powerful new infrastructure. It has been necessary, however, to make hard decisions regarding what capabilities would be included in the initial build to keep the project within the financial bounds. Within the total envisioned capital investment and

the planning figure for out-year operations and maintenance, the OOI must strike a balance between its role as a platform that will host separately funded experiments and campaigns, much like a research vessel, and as an infrastructure that will supply critical measurements from an initial investment in key sensors. To provide a framework for these decisions while retaining the flexible, expandable, and transformative nature of the facility, the program's advisory committee provided the following guiding principles:

- Power and communications capabilities exceeding those of traditional ocean observing platforms are the leading aspect of the transformative characteristics of the OOI Network
- The OOI should emphasize fewer, more-capable nodes over more numerous, less-capable nodes (i.e., with traditional capability)
- A mix of fixed platform and mobile assets is important for many science questions, and will help develop and maintain the user base of the OOI
- The “integration across three scales” (coastal, regional, global) should be exploited to the extent possible for appropriate science questions
- The OOI is a research platform that will enable future experiments/capabilities beyond those included in the initial configuration
- The OOI Network should achieve a balance between enabling science “out of the box,” and designing infrastructure to support separately funded PI experiments

OOI Cyberinfrastructure Capabilities

The OOI will be a globally distributed multiscale network of observing assets bound together by a cyberinfrastructure (CI) backbone that will link the infrastructure elements, sensors, and models into a coherent system-of-systems. The CI's full two-way communication capabilities will play a key role in transforming oceanography from a primarily ship-based expeditionary science to a distributed, observatory-based discipline in which scientists continuously interact with instruments, AUVs, actuators, facilities, and other scientists to remotely explore the Earth-ocean-atmosphere system. The CI will allow access to other (i.e., non-OOI) data streams to provide users with a coherent four-dimensional view of the ocean. Using the OOI CI, users will be able to combine OOI water-column data with NOAA and NASA satellite ocean-surface imagery and NOAA Integrated Ocean Observing System (IOOS) subsurface data. These interactive and dataaggregation capabilities will complement parallel international efforts by Canada, Japan, and Europe and have the potential to change fundamentally how ocean science is conducted.

The OOI CI will provide a range of capabilities. In principle, anyone—scientist, engineer, or educator—will have access to two-way interactivity, command and control, and data discovery through near-real-time data and historical data archives. An advanced CI capability will allow data delivery, analysis, and synthesis from across the OOI on a localized semantic frame of reference. The CI will permit mediation among different protocols, data streams, and derived products to test scientific understanding. In accordance with the OOI program's data policy, calibrated and quality-controlled data must be made publicly available with minimal delay. The delays comprise the time needed to fill a data packet (e.g. 1s) and communicate ashore through the Internet (>250ms for geosynchronous satellites). Generally data latencies will be on the order of seconds although they can be larger for low-power buoys or AUVs that

connect only occasionally to the Internet. These near-real-time data will require sophisticated system security, and protocols to control and prioritize access to sensors and data in accordance with agreed-upon policies in the program's operational framework. Near-real-time, reliable data delivery and interactivity will be an important tool, enabling shore-based scientists to adaptively adjust sampling strategies.

A traditional data-centric CI, in which a central data management system ingests data and serves them to users on a query basis, is not sufficient to support the ocean science community's OOI requirements. Therefore, throughout the observatory network, strategically placed computation and storage infrastructure will be provided at Cyberinfrastructure Points of Presence (CyberPoPs). These CyberPoPs include integrated, near-real-time data processing and archive sites located at a few central facilities and at marine observatory shore stations or control centers. In situ computation and storage resources will be located within science instrument interface modules and selectively within the marine networks. The external functions of a sensor (gain, sample rate, on/off, power supply level, etc) will be accessible through Web services.

Although each observatory scale (coastal, regional, global) will have a CyberPoP with a central machine for data archiving and quality assurance functions, the CI will enable virtual observatories to be built anywhere using different combinations of the physical assets. For example, investigators interested in specific problems would "subscribe" to real-time data from sensors that specifically address their research topics. This decentralized grid concept means that the CI is broadly distributed around the United States and the world, and thousands of custom views can be provided simultaneously. To complement the observational data stream, large computational models and simulations can be run on the national grid infrastructure (i.e., the TeraGrid and the Open Science Grid). Finally, observatory participants will be able to securely incorporate computation and storage capabilities within the instrument platforms (e.g., autonomous underwater vehicles and gliders) and research facilities. Thus, the CI blurs the conceptual boundaries of the coastal-, regional-, and global-scale observing assets, considering rather the OOI network as an illuminated grid of sensors, each with unique Internet presence.