# CYBERINFRASTRUCTURE FOR THE U.S. NSF OCEAN OBSERVATORIES INITIATIVE: A MODERN VIRTUAL OBSERVATORY

***Orcutt, J.***[1,2]***, Vernon, F.***[1]***, Peach, C.***[1]***, Arrott, M.***[2]***, Farcas, C.***[2]***, Farcas, E.***[2]***, Krueger, I.***[2]***, Meisinger, M.***[2]***, Chave, A.***[3]***, Schofield, O.***[4]***, Kleinert, J.***[5]

[1] *Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, 92093-0225, USA*
[2] *California Institute of Telecommunications & Information Technology, University of California, San Diego, La Jolla, CA, 92093-0436, USA*
[3] *Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA*
[4] *Rutgers University, Princeton, NJ, 08901, USA*
[5] *Raytheon Intelligence and Information Systems, Aurora, CO, 80011, USA*

## 1. INTRODUCTION

The US National Science Foundation has initiated a transformation of ocean science with the *Ocean Observatories Initiative (OOI)*. The OOI is designed to provide new, persistent, interactive capabilities for ocean science, and has a global physical observatory footprint. The *OOI Integrated Observatory* comprises Regional Scale Nodes (RSN) and Coastal/Global Scale Nodes (CGSN) providing cabled and buoy observatories with mobile instrument platforms, respectively.

The *OOI Cyberinfrastructure (CI)* constitutes the integrating element of the OOI Integrated Observatory and is extensible to promot2 platform and sensor expansion in the 25-30 year future of the OOI. The CI links and binds the physical observatory, computation, storage and network infrastructure into a coherent system-of-systems. The core capabilities and the principal objectives of the OOI Integrated Observatory are collecting near-real-time data, analyzing data, modeling the ocean on multiple scales and enabling adaptive and interactive experimentation within the ocean. 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 accomplish the range of tasks ocean scientists will engage in when the OOI is implemented. Instead, highly distributed collections of capabilities are required that facilitate:

- End-to-end data preservation, provenance maintenance and access,
- End-to-end, human-to-machine and machine-to-machine control of how data are collected and analyzed,
- Direct, closed loop interaction of models with the data acquisition process,
- Virtual collaborations created on demand to drive data-model coupling and share ocean observatory resources (e.g., instruments, networks, computing, storage and workflows) – the *Virtual Observatory,*
- End-to-end preservation of the ocean observatory process and its outcomes, and
- Automation of the planning and prosecution of observational programs.

The OOI CI provides the software services, the user interfaces to support these applications; in addition it provides the underlying integration infrastructure comprising message-based communication, governance and security frameworks, similar to the role of the operating system on a computer. Users or system managers are able to *compile* new, specialized observatories through the CI.

## 2. SCIENTIFIC PROCESS VIEW OF THE OOI CI

From the outset, the OOI CI was designed to implement a set of activities derived from a process model that structures their interaction and compartmentalization (Fig. 1). The model supports the real-time coupling and aggregation of the main *observe, model* and *exploit* activities as well as their respective sub-activities.


*Figure 1. CI Process Model*

The *observe* activity constitutes the data collection strategy that lies at the core of scientific investigation. The outcome is a set of observational products such as data or quantities derived from them that serve as the input to a *model* activity. The result of modeling is a set of derived products that yield an interpretation of the data and the processes that determine them. Based on this improved understanding of the underlying physics, chemistry and biology, the *exploit* activity is used to plan, simulate and execute additional observation activities. The entire scientific process operates as a closed rather than an open loop, with or without human intervention. The integrated data, model output, data analysis and resource exploitation planning/execution all become elements of a knowl*edgebase,* a great step beyond the more familiar *database.*

## 3. FUNDAMENTAL STRATEGIES

The CI integration strategy is based on two core principles: messaging and service-orientation. A high-performance message exchange provides a communication conduit with dynamic routing and interception capabilities for all interacting elements of the system-of-systems. The message interface is isolated from any implementation technologies, and provides scalability, reliability and fault-tolerance. Service-orientation is the key to managing and maintaining applications in a heterogeneous distributed system. All functional capabilities and resources (e.g. sensors, storage, gliders, buoys or cables) represent themselves as services to the observatory network, with precisely defined service access protocols based on message exchange. Services are also defined independently of implementation technologies. Assembling and integrating proven technologies and tools using the integration infrastructure provide the functional capabilities of the CI.

The CI deployment strategy is based on the concept of a capability container (Fig. 2) that provides all essential infrastructure elements and selected deployment-specific application support. Capability containers can be deployed wherever CI-integrated resources are required across the observatory network, and can be adapted to available resources and their environment. This includes (for example) platform controllers on remote, intermittently connected global moorings, compute units placed in the payload bay of AUVs and gliders and the full range of terrestrial CI deployments (Cyber Points of Presence or CyberPoPs). The CI multi-facility strategy supports the collaboration of multiple independent domains of authority bringing together their own resources, with no central governance and policy authority, based on agreements and contracts. This enables the sharing and use of observatory resources across the network, governed by consistent policy.

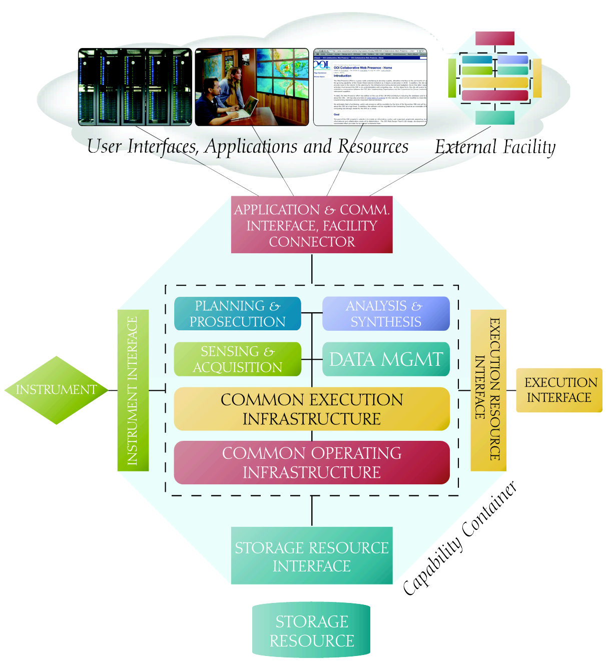


*Figure 2. Capability Container*

## 4. SERVICES NETWORKS AND SUBSYSTEMS

The CI's functional capabilities are structured into six services networks (SNs) that participate in operational activities, resulting in services that support user applications. The application-supporting services networks are Sensing and Acquisition, Data Management, Analysis and Synthesis, and Planning and Prosecution. The infrastructure services networks are Data Management, Common Execution Infrastructure (CEI), and Common Operating Infrastructure (COI). Data Management provides both application and infrastructure services.

The services networks (SN) and their constituent services are implemented and integrated into subsystems of the same name. Subsystems are the primary units of work breakdown and will be delivered by individual subsystem construction projects.

The Sensing and Acquisition SN performs instrument management, mission execution, and data acquisition tasks. The Instrument device model comprises one or more physical sensors or actuators, and is represented in the CI by a logical device, the Instrument Agent. The physical device provides sensor data and status information to the Instrument Agent, and receives configuration information and commands from it. Each Instrument Agent has an Instrument Supervisor that is responsible for monitoring the state of health of the Instrument. Each platform has a Platform Agent that is responsible for controlling all devices on that platform. Observatory Management is responsible for coordinating all observatory resources to ensure their safe operation. The Exchange service is a projection of COI capabilities, with the main purpose of establishing a publish-subscribe model of communication that ensures distributed data delivery to all end-points. It appears in all SN's.

The Instrument Agent is responsible for translating commands from the Platform Agent to the Instrument, event handling, acquiring data from the Instrument, managing the state of the Instrument, providing direct access (e.g., ssh) to the Instrument, and updating the Instrument clock from an external source. The Instrument Supervisor receives status information

## 5. THE OOI AS A SYSTEM OF SYSTEMS.

The primary goal of the Integrated Observatory Cyberinfrastructure functional design is to support the activities and applications of:

- Scientific Investigation, supporting researchers in the study of environmental processes though observations, simulation models and expressive analyses and visualizations, with results that directly feed back to improve future observations.
- Education and Participation, supporting education application developers, educators and the general public for accessing and understanding OOI resources in ways suitable for specific target audiences.
- Community Collaboration, enabling OOI users to share knowledge and resources, and to work together in project settings and ad hoc communities.

In support of these activities, a variety of Integrated Observatory resources of different type and purpose need to be administered, including:

- Observation Plans, providing activity sequences, service agreements and resource allocations for observational campaigns, and similar templates for event-response behaviors;
- Data Sets, representing observational and derived data and data products in the form of data archives and real-time continuous data streams;
- Processes, representing data collection and processing workflows that arrange multiple steps involving multiple actors and resources;
- Instruments and marine observatory infrastructure elements, such as telemetry systems, GPS and data loggers;
- Models, including numerical ocean forecast models and their configurations, as well as other analysis and event detection processes;
- Knowledge, representing all metadata, ancillary data, analysis results, association and correspondence links between resources, and knowledge captured in ontologies for semantic mediation purposes.

The support for these activities and resources rests on a collection of infrastructure services that provide resource management, interaction, communication and process execution. The *CI Capability Container* is the extensible, deployable base unit of CI capabilities. The Integrated Observatory's functional capabilities are structured into six services networks (i.e., subsystems) that were discussed earlier: four elements that address the ocean and Earth science- and education-driven operations of the OOI integrated observatory, and two elements that provide core infrastructure services for the distributed, message-based, service-oriented integration and communication infrastructure and the virtualization of computational and storage resources.

The *Sensing and Acquisition* services network provides capabilities to interface with and manage distributed seafloor instrument resources, as well as provide quality control services. The *Data Management* services network provides capabilities to distribute and archive data, including cataloging, versioning, metadata management, and attribution and association services. The *Analysis and Synthesis* services network provides a wide range of services to users, including control and archival of models, data analysis and visualization, event detection services and collaboration capabilities to enable the creation of virtual laboratories and classrooms. The *Planning and Prosecution* services network provides the ability to plan, simulate and execute observation missions using taskable instruments; it is the CI component that turns the OOI into an interactive observatory.

The remaining two services networks are the *Common Execution Infrastructure* (CEI) and the *Common Operating Infrastructure* (COI). The CEI provides an elastic computing framework to initiate, manage and store processes that may range from initial operations on data at a shore station to the execution of a complex numerical model on the national computing infrastructure and on compute clouds. The COI includes capabilities to manage identity and policy, manage any resource's life cycle, as well as catalog and repository services for observatory resources. An efficient messaging and service bus that incorporates security and governance, and provides guaranteed delivery, lies at its heart. Service-orientation and messaging realize loose coupling of components, resulting in the flexibility and scalability that are key in such a complex large-scale system-of-systems. All OOI functional capabilities and resources represent themselves as services to the observatory network, with precisely defined service access protocols based on message exchange. Fig.2 depicts a capability container, indicated by the octagon shape, with interfaces to local resources and to the network environment. Local resources include physical resources such as instruments (sensors) and marine observatory infrastructure, storage resources such as disks and network drives, and computing resources such as grid nodes, cloud computing instances, and CPUs on mobile platforms such as AUVs (Autonomous Underwater Vehicles). Capability container can also be connected to user interfaces, external applications and to capability containers in different, independent facilities that have their own domains of authority and operation. No matter where deployed, the capability container provides all of the infrastructure and application support required for an installation site within the OOI Integrated Observatory network. The capability container hosts the six services networks and their resource interfaces as depicted in the figure. The footprint of a capability container can vary depending on the resource constraints of its hosting environment. The selection of functional capabilities present in a specific capability container depends on the respective needs and resource availability at this specific location in the network. For instance, on an intermittently connected instrument platform, instrument access, data acquisition and data buffering capabilities provided by the Sensing and Acquisition services are required, while the Analysis and Synthesis capabilities are not. In contrast, at the core installation sites, data processing, numerical model integration and event response behaviors need to be present.

## 6. SUMMARY

The Ocean Observatories Initiative faces the enormous challenge of building a cohesive distributed system-of-systems that incorporates a large number of autonomous and heterogeneous systems, deals with instruments and computational resources of a wide range of capabilities, serves the needs of diverse stakeholders, and accommodates change over the timescale of decades. A carefully thought out architecture is key to addressing this challenge. We find that simplicity wins and a few core principles help us organize the OOI properly.