Building Standards for Access to Oceanographic Observatory Instrumentation

A.R. Maffei[1)], A.D. Chave[1)], J. Graybeal[2)], J.B. Edson[1)], J.J. Fredericks[1)], J.W. Ammerman[3)], and R. Hanisch[4)]

1) Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA
2) Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA
3) Rutgers University, New Brunswick, NJ 08901, USA
4) Space Telescope Science Institute, Baltimore, MD 21218, USA

KEYWORDS: ocean observatory, interoperable portal, sensor network, grid

ABSTRACT

The methods and software interfaces used by ocean observatory operators and scientist users to access data generated by sensors and instruments attached to research observatories vary widely from facility to facility. The tools used to access instrument metadata and to control various instrument functions are also variable between observatories. An effort has begun in the US to define software standards that could be implemented at any ocean observatory to provide access to oceanographic instruments independent of the observatory on which they operate. Interoperability also enables the seamless combination of data streams from more than one observatory into a virtual ocean observatory. This paper describes the concepts and a proposed process to develop interoperability standards for ocean observatories.

1. INTRODUCTION

Since the 1800 s, oceanographers have explored and sampled across two-thirds of Earth using ships as observational platforms. This provides a series of snapshot views of the oceans, and has limited resolution in time. Measurements and models from this exploratory phase of oceanography have resulted in growing recognition of the diversity and complexity of processes that operate above, within and beneath the oceans. The questions posed from these efforts increasingly cannot be answered using only the tools of the present, in large part because of a limited ability to resolve episodicity and temporal change. For this reason, the ocean sciences are beginning a new phase in which scientists will enter the ocean environment and establish interactive networks for adaptive observations of the earth-ocean system. The growing move to establish ocean observatories (hereafter simply OOs) reflects this trend.

With the advent of OOs, marine scientists will face a data avalanche as the limitations of battery power and internal data storage for autonomous instruments are bypassed and the bandwidth and power capabilities of submarine fiber optic/power cables are realized. Real-time oceanography will demand new information technology (IT) approaches to provide the tools and methodologies for instrument access and control on and in the oceans and data access and analysis on land. These technologies must be scalable and adaptable to a range of OOs and instrumentation. A key goal of access standardization is architecting and building common instrument and data interfaces, implementing them within existing research observatories, and providing them to new OOs as they are designed. By providing each OO with interoperability through common instrument and data interfaces, the user community will have the ability to readily move instruments between OOs and easily integrate data streams from multiple observatories into science, education, and public outreach programs.

In many respects, the IT challenges faced by marine scientists mirror those in other fields like astronomy. For example, the National Virtual Observatory (NVO; see http://www.us-vo.org) will link the archival data sets of space- and ground-based observatories and provide standard tools to mine and utilize these vast data sets. However, OO requirements differ from those of astronomy in several important respects. First, OO data are real-time or near real-time with a wide variety of sensor types having data rates ranging from a few baud to 20 Mb/s (for compressed HDTV); this upper limit is certain to rise as technology evolves. Second, OO data are inherently very diverse; in addition to a broad range of digital data streams, they include physical samples (e.g., rocks or animals) and their associated descriptions or analyses. Third, the mix of instruments at an OO is inherently dynamic; data streams are added or removed as the science requirement changes. Fourth, users will desire to merge real-time data from several OOs, effectively forming a Virtual Ocean Observatory (VOO). Interoperability standards must accommodate those requirements that are unique to OOs, yet utilize the products of efforts like NVO or various grid implementations wherever possible.

2. FUNCTIONAL REQUIREMENTS

2-1. A Generic Ocean Observatory

OO implementations vary greatly, and in an attempt to establish common terminology, a generic structure will be defined (Figure 1). A suite of *sensors* is the fundamental measurement device at an observatory, and is ultimately the source of a data stream. Sensors are part of, or attached to, *instruments*. Instruments are attached to *node instrument ports* on an *observatory node*. They are connected to the port via an *access layer datacomm connection* (typically RS232/RS422 or10/100BaseT Ethernet). *Custom instrument ports* may support unique instruments with special needs. *Standard instrument ports* are more generic. OO nodes are connected to each other and to the *ocean observatory shore station* via a *core layer datacomm link* (typically multiplexed channels over a high-speed serial link, Gigabit Ethernet via cable or a lower bandwidth radio/satellite link). An *ocean observatory operations center* monitors, maintains, controls, and manages the components of the observatory.

An observatory *instrument control process* is used by operators or guest scientists to control node instrument ports, instruments, and sensors. *Core instruments* are managed by the OO

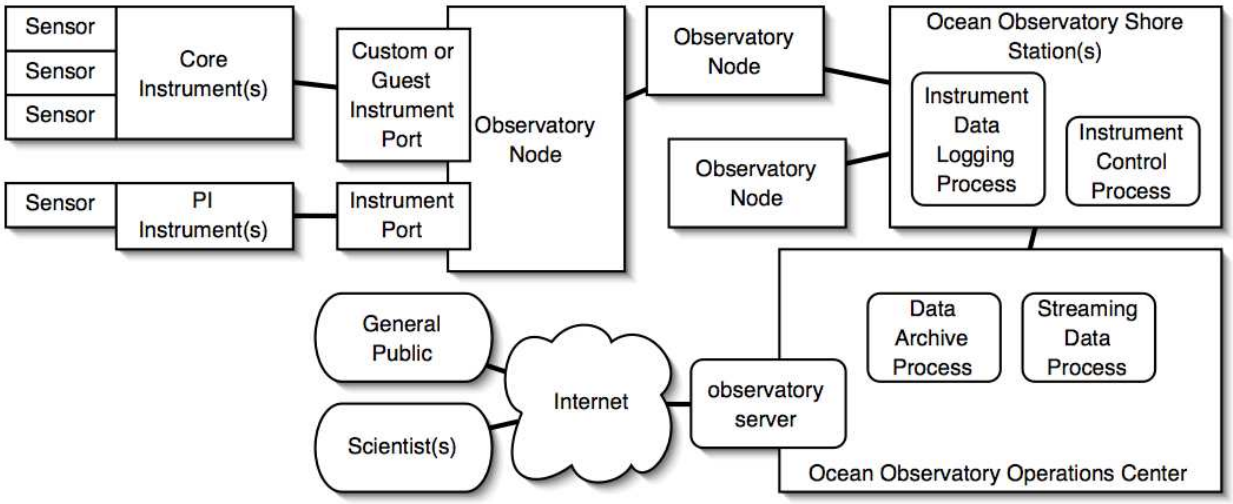

Figure 1. Components of a generic ocean observatory

operator or their designees. *PI instruments* may be one of a kind, and are deployed on behalf of a guest scientist. The *instrument data logging process* gathers real-time or near real-time data (and sometimes *instrument metadata*) from instruments and stores them temporarily. This step may happen in the water, on shore, or both. The data archive process gathers/receives data and/or metadata from the instrument logging process or, in some cases, directly from the instrument itself. The *data archive process* may extract instrument metadata from the data stream, post-process the data stream, or manage it in some other way. In some OOs, a *streaming data process* provides subscribers with a real-time stream of data from sensors/instruments.

In most cases, the OO is connected to the Internet and an *observatory server* provides scientists and the public with access to certain observatory services. Access standardization aims to establish a common set of external services on OOs and a common portal to access their components. The desired services include control of observatory instruments, access to observatory instrument data streams, and access to archival data from observatory instruments. We call the set of services (termed the *Common Ocean Observatory Interface*) together with the common access portal, the *Interoperable Ocean Observatory Portal* or IOOP.

2-2. Science Requirements

Establishing access standards requires architecting and building common instrument and data interfaces at the observatory level, and then implementing them within existing and future research observatories. Through providing each OO with interoperability through a common instrument interface and a common data interface, the user community will have the ability to readily move instruments between OOs and easily integrate data streams from multiple observatories into science programs. Investigators require two basic types of interaction with their instruments: instrument control and instrument data access. Once these are available at a number of OOs, the aggregate can be combined to form a Virtual Ocean Observatory (VOO).

A common data access interface will give a user access to data in real-time or near real-time from an OO or VOO (as well as to archived data). It will allow users to control and check the status of all instruments in the OO or VOO. Both capabilities enable immediate, broad-scale adjustments to environmental and instrumentation circumstances. These features facilitate process studies that have regional rather than local impact. The regional scale provided by a VOO will dramatically increase the ability to capture episodic events over a range of spatial and temporal scales. In addition, a VOO is a powerful tool for both education and outreach and as an aid to public safety agencies during extreme weather events.

A common instrument control interface, in conjunction with the common data access interface, must also be designed to selectively route some or all of the real-time data in an OO or VOO to any location. This might be of particular interest to a modeler who could take advantage of an OO or VOO to assimilate data streams into numerical models. The ability to pull in data that is distributed across space as well as time will enable major advances in prognostic modeling that have both scientific and public policy implications.

Finally, it is of particular interest to devise a system that can display selected data from multiple OOs (or related platforms such as ships or buoys, which can be treated as OOs in this model) during specific times in an easily interpretable form. For example integration of offshore data with more conventional land-based measurements will improve such everyday tools as weather forecasts in coastal areas, and support the development of air-sea interaction models.

2-3. Requirements for a Common Data Access Interface

The ability to access data from instruments at a wide variety of sites in a standard way will allow the user community to take full advantage of a VOO for research and education. Further, demonstration of the benefits provided by it will encourage new OOs to become participants in a growing virtual observatory organization. As such, a goal of standardization is to facilitate interoperability in data management and access. An incomplete list of functional requirements driving the interface might include:

- A common operational data nomenclature that uniquely defines an instrument, including its sensor parameters and component deployment variables
- Synoptic time-stamping of data with agreed upon accuracy
- Access to instrument data streams that can be controlled in accordance with user privileges
- Ability to determine what data are or will be available with sufficient detail to plan data strategies

- Notification when a data stream is or will be interrupted, either intentionally (e.g., for servicing) or unintentionally (e.g., through communication failure or power outage)
- A broadcast notification capability that can flag pre-defined events.

A variety of use scenarios need to be anticipated. While a client will ordinarily be a user or an application on the Internet, it might include a set of seafloor instruments (changing the sampling scheme in response to an event, for example) through the common instrument control interface described in the next subsection.

2-4. Requirements of a Common Instrument Control Interface

The interoperability provided by a common instrument control interface will encourage a broad range of investigators to become participants in a VOO. Existing research OOs have incorporated a range of often incompatible power and communications interfaces, and data management is typically rudimentary at best. As a result, instruments and their user interfaces have to be designed for use at a specific observatory, with varying degrees of change required to move them. This is not a desirable use scenario, and a common control interface is needed to change this through standardized access protocols and services.

An incomplete list of the functional requirements for the interface includes:

- The ability to reset the port interface protocol (Ethernet/RS232, baud rate, etc.)
- The ability to remotely power instruments up or down
- Telemetry of sensor status
- Provision for instrument reset or reboot
- Command to initiate data transfer in a default mode
- Command a pause in data transfer
- Integration of metadata with the data stream
- Selectively controlled access to the above (i.e., security)

This interface should provide a common look and feel to the basic functions required to control and query instrument packages. Other useful functionality should be incorporated as the interface is defined and implemented. Of course, not all features may be usable at all observatories due to site-specific capabilities, but optional features may include:

- Direct communications with an instrument via the Internet
- Initiation of sleep mode for packages with this capability
- User notification of data interruption

2-5. Security for a Virtual Ocean Observatory

A virtual ocean observatory will be created by establishing portals to the common instrument and data interfaces residing at each of the participating OOs. To provide seamless access to instruments and data, a level of security must be implemented that prevents access by unauthorized users, but maintains control by instrument owners and/or operators at each OO. Catalogs documenting instruments at each participating OO and core data sets must be controlled through user authentication and appropriate access controls. Once authenticated, users should have full access to their instruments or data streams, as well as any authorized access to other available resources across the entire VOO.

Individual observatories must control specific instrument access points (e.g., physical ports or IP addresses); creation of a super-agency controlling these access points, or other observatory-internal resources, is not advocated. An instrument package should only become a part of a VOO if it is properly integrated into the instrumentation infrastructure that is administered by each participating OO.

3. IT PERSPECTIVE

The primary goal of access standardization from an IT perspective is creation of a software infrastructure that supports and encourages the collaborative development and operation of independent OOs by implementing the conceptual interfaces of Section 2. To realize the scientific vision embodied in the OO concept, this infrastructure must encompass the complete range of the experimental process from sensor development and operation to data management, archiving, and access, as it is embodied in existing and planned OO systems.

The problems inherent to developing a federation of OOs have a lot in common with those faced by other recent projects that develop virtual observatories or large-scale data grids and real-time data management of sensor-nets (e.g., NVO, http://www.us-vo.org/ and NEESgrid, http://www.neesgrid.org/) and real-time data management of sensor-nets (e.g., VORBS, http://road-net.ucsd.edu/vorb.html). The general strategy for using grid, XML, and Web technologies as a foundation for a large-scale, multi-organizational cooperative has been well described and need not be restated in detail here. Instead, the remainder of this section focuses on the differentiating set of problems faced by real-time OOs that have not yet been fully addressed.

A dynamic instrument mix and the real-time aspects of OOs pose a challenging set of issues. OOs must be able to rapidly reconfigure their data collection priorities in response to often unpredictable natural events. Both core and PI instruments vary widely in complexity, autonomy, and bandwidth requirements, yet must share common resources providing power and communications. Data gathered at the OOs needs to be logged, distributed to users throughout the world in near real-time, and ultimately entered into a data archive. As OOs increase in complexity, providing both user utility and reliable operation become an ever increasing challenge. Real-time control and data management problems rapidly become more difficult than those associated with massive but homogeneous sensor nets or the fixed sensor suites of observational satellites. While both of these domains pose significant operational challenges, to a large extent they can be analyzed and solved at system design time. By contrast, OOs need an operational infrastructure that is dynamically reconfigurable for a wide range of sensor types and quantities, communication link configurations, and user demands. At the same time, it must remain easy for users to add new instruments and to observe and control experiments. Finally, the existing infrastructure at the OOs must remain functional until the new common software infrastructure is operational.

The architecture and implementation of the IOOP software infrastructure must take all these factors into consideration. Thus, it must provide a flexible, self-describing environment that facilitates interoperability, but allows innovation and uniqueness where required. There must be easy to use but well defined interfaces to the OO accessible to users at all layers of the architecture. Finally, given the trends in oceanography, the architecture must anticipate scaling to hundreds of OOs, tens of thousands of instruments and thousands of users.

Conceptually, the problem of operating an OO with its multitude of sensors and users is similar to monitoring and managing a distributed computing environment. An instrumented distributed computing environment has a wide variety of both operating system- and application-specific sensors to gather information about performance, health, and capacity of both physical subsystems and software components. At a given time, there may be one or more requestors for information from these sensors with different

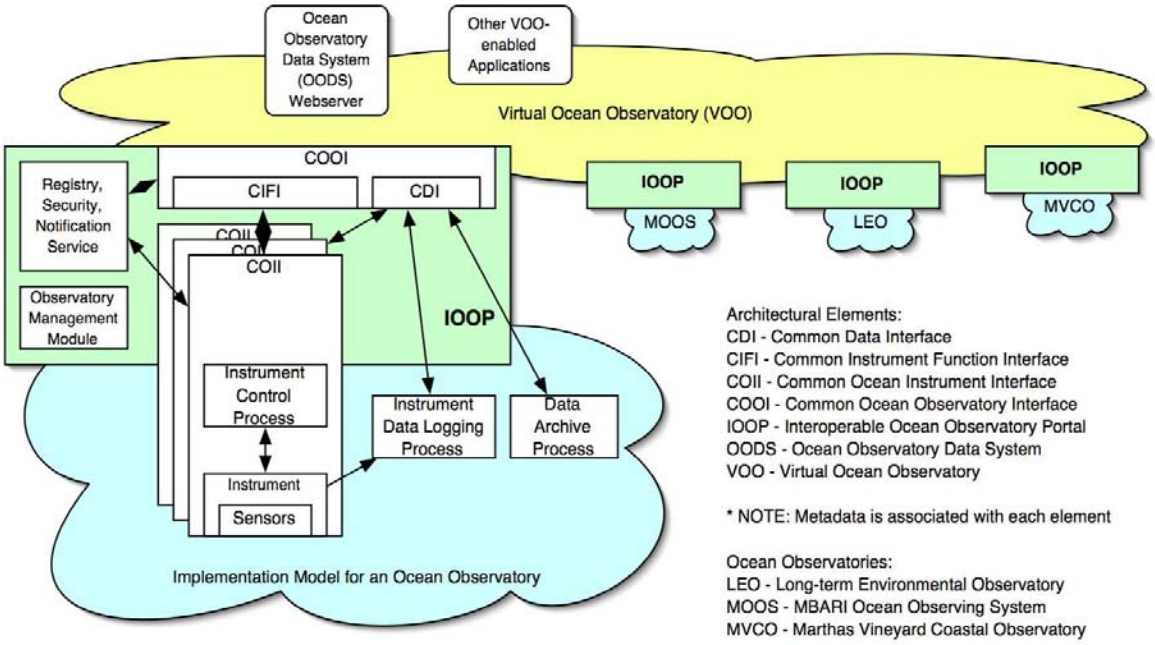

Figure 2. IOOP and VOO Architecture Elements

reporting requirements. To monitor large-scale distributed systems without imposing excessive measurement overhead, it is necessary to multiplex sensor data streams and otherwise optimize information distribution. Just as OOs need to be able to adjust observational priorities and resource usage to respond to natural events, computer monitoring systems must also rapidly adjust to unpredictable hardware and software events.

4. ARCHITECTURE

The architecture of the IOOP must balance many competing needs, including compatibility with a variety of OOs, provision of sophisticated and highly configurable services, and simple interfacing to the system. Figure 2 outlines an architecture that fulfills these needs, showing the critical Interoperable Ocean Observatory Portal (IOOP) components in the context of a Virtual Ocean Observatory (VOO). From the perspective of the VOO, the instrument-based services of each observatory (instrument control, online data stream access, and data archive access) are mediated by an IOOP. A user of the VOO thus sees each observatory through the same interface, and can access the services of instruments within all of the observatories as if they are part of the VOO. The Common Ocean Observatory Interface (COOI) is the publicly-exported interface which will be used by VOO customers and applications. The functionality of this interface is expected to grow over time, but should initially include the Common Instrument Functional Interface (CIFI) and the Common Data Interface (CDI) which will closely match control and data functions that observers are familiar with in current OOs. The COII is the IOOP internal interface to the services of sensors and instruments. Observatory users will interact with the COII primarily through the CIFI and CDI. The COOI will allow instrument developers and users to write programs which automatically find and connect to the desired instrument services. Note that this architecture supports the concept of a web front end for simple requests or exploration of the instruments attached to OOs that are part of a VOO.

4-1. Common Ocean Instrument Interface

A key aspect of access standardization is the development and promulgation of mechanisms to uniquely and dynamically identify real sensors and instrument packages, and to associate descriptive and calibration data with them.

Sensors and instruments interact with the OO infrastructure through a Common Ocean Instrument Interface (COII) and do not need to know anything about higher layers of control. The COII layer is the abstraction layer which couples sensors with their various properties to the OO infrastructure and ensures that descriptive metadata are associated with both the sensors and data streams. It hides physical connectivity and sensor-specific interfacing details from the higher levels of the infrastructure, and also hides the complexity of the infrastructure from the sensor. The COII provides a control structure for basic operations, with the following partial list of characteristics:

- Control of basic operations (start, stop, reset, status, data collection rate, etc)
- Provision of a pass-through mode for direct control of a sensor
- Transmission of data through the VOO data collection and control interfaces
- Registration/deregistration of sensors in the OO resource directory using grid services
- Implementation of security controls as needed

The realization of a COII could take many different forms, depending upon the sophistication of the sensor(s) it is managing. It could be totally or partially embedded in the sensor or hosted by the observatory. In the context of the Open Grid Services Architecture, there would be an instantiation of a COII interface object of each sensor (or group of sensors). This object would be sensor-controlling code. The controlling code could be a standalone program or a dedicated hardware controller running anywhere as long as it can be wrapped and made responsive to the basic control operations defined by the observatory.

4-2. Observatory Management Module

The Observatory Management Module is part of each IOOP and links to all of the other components. It provides a mechanism for intra-observatory coordination that might be required to accommodate various requests. This component provides mechanisms for overall control and management of an OO. While not strictly necessary in small OOs with tens of sensors and a few observers, it becomes increasingly important as OOs scale up. A partial list of characteristics includes:

- ¥ Manages data and control streams between sensor control interfaces and users
- ¥ Multiplexes communication links to sensors if required
- ¥ Provides support for optional data filtration and reduction
- ¥ Distributes data to multiple users (people and archives)
- ¥ Provides dynamic prioritization of OO resources to accommodate real-time events
- ¥ Maintains state information about the OO and its users in a resource directory
- ¥ Implements security control with normal/privileged access to sensor control interface
- ¥ Interface could be replicated and/or hierarchically-layered to provide robustness and scalability if required

4-3. Metadata Management

A key to providing automated data services over a wide range of system elements is proper definition and use of metadata. In science, consistent availability of descriptive metadata is critical to using and repurposing the data with minimal effort. For infrastructure development, XML is rapidly becoming the most prevalent mechanism to maintain and communicate information about software, systems, operations, and services. By defining the necessary capabilities or meta-knowledge using XML, developers can write general-purpose systems which can provide many unanticipated capabilities with minimal reprogramming or maintenance effort.

On the data side, metadata are especially critical for automating sophisticated data management services. For example, in MBARI's observatory data management system, the descriptions of data streams are captured before the stream is received, and are used to reprocess data into netCDF files, present them to users in multiple formats, and support user searches for data content. Further, any new data set may be submitted and processed without changes to the existing software. Extensive international efforts now focus on defining useful standards for marine metadata terminology, and MBARI is helping lead this effort with participation on the Marine XML Steering Committee.

Just as it is used to describe data, XML can be used to define available services. For example, instrument developers at NASA's Goddard Space Center developed an instrumentation markup language in XML (AIML), and have used it to define instrument services and their interaction with other observatories. Much of the grid computing effort depends on similar metadata descriptions of available services and protocols.

4-4. Role of Grid Technology

Grid software is a prime candidate for the foundation of the IOOP software infrastructure. By the time development starts, the Open Grid Services Architecture (OGSA) with the Globus Toolkit version 3 as a reference implementation (see http://www.globus.org/ogsa) will be generally available. It includes many new functions particularly appropriate to the IOOP design and to the multi-observatory environment. One of the most applicable grid technologies is the Globus security model. This is well suited for a federated environment and facilitates interactions with grid-based data repositories and computational grid projects. The meta-directory service (MDS), coupled with the Service Discovery and Notification services, provides a basis for a higher service to maintain configuration and status information about each observatory and its connected instrumentation.

While grid technologies constitute a useful set of building blocks for the IOOP infrastructure, they are in general too complex and burdensome to use in instruments and sensors. The scientists and engineers developing instrumentation and running experiments want a clean and simple set of interfaces, and individual OOs will not be inclined to integrate grid components deep in their architectures. Within the IOOP architecture, isolation is provided at the COII. From the grid perspective, this interface would be an encapsulated custom grid service instantiated as needed by a grid factory. Customers of an experiment could use grid services to locate and bind to the appropriate COII instance and then control the experiment and receive data through it. Each observatory would provide instrument services to the grid, which can be accessed and served by the COII instances.

4.5 Integrating Real-time and Archival Data

The portal providing a single interface for multi-observatory instrument control and data stream access presents a powerful and compelling user experience. However, many interesting research applications require integration of real-time with archival data. A typical scenario is the researcher or educator who wishes to review recent data, send commands to the source instruments, and evaluate newly arriving data streams in the context of the previously reviewed data. Other users will want to put up a running strip-chart of experimental data generated by instruments deployed on a number of different observatory platforms, but will want to interactively adjust the strip chart record to span some number of minutes, hours, or days before the 'live action' data begin.

Requests for historical data require integrated access to both data and metadata. At present, many distinct, physically separated data archives exist, and the users who design experiments and configure instruments for each OO are responsible for quality control. Such local responsibility may work well for single, small OOs, but haphazard archival of metadata often makes the use of data sets and especially the integration of data from many different observatories difficult.

Fortunately, there have been (and continue to be) significant developments in metadata catalog systems, and in corresponding data access protocols. The digital library community has established standards and working software that enable the discovery of resources across widely dispersed data holdings. Many scientific disciplines are evolving information management standards and software that enable discovery of data sets. One notable example is the collection of FGDC clearing houses that are interoperable with many kinds of software and data stores. The NVO is another increasingly successful implementation of highly distributed and heterogeneous data discovery. The electronic business community also develops standards and software for electronic business transactions; these are typically metadata-rich and concentrate on modeling business processes and transactions. Selected technologies from each community will apply to the IOOP.

Once data have been discovered, they must be accessed. This requires a common access protocol and a data archival architecture that supports that protocol. Many low- to mid-level technologies exist to support this need, with OPeNDAP

(http://dodsdev.gso.uri.edu/ODC/) perhaps the most widely applied in oceanographic systems.

Many existing projects in other scientific disciplines promise to demonstrate integrated archival data access across a range of data sets, and some projects, such as the NVO and the Rutgers OBIS project, have already achieved considerable success.

The MBARI Ocean Observing System (MOOS) software architecture addresses the entire life cycle of observatory operations issues, including those of archival data management. The Shore Side Data System, operating in a limited-functionality first release, uses many of the principles described here to systematize data management, minimizing ongoing development costs. It provides an excellent test bed to integrate "live" instrument data with archival data (whether from the same instrument, its predecessors, or an alternate data source) via the IOOP portals.