LOOKING: Cyberinfrastructure for Ocean Observatories

A.D. Chave
Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

M. Arrott and L. Smarr
Calit2, University of California at San Diego, La Jolla, CA 92093, USA

J.A. Orcutt
Scripps Institution of Oceanography, La Jolla, CA 92093, USA

E. Lazowska
Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA

J.R. Delaney
School of Oceanography, University of Washington, Seattle, WA 98195, USA

M. Abbott
College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA

Abstract – The Laboratory for the Ocean Observatory Knowledge INtegration Grid (LOOKING) is developing a design for ocean observatory cyberinfrastructure comprising a robust framework that supports dynamic harnessing of resources ranging from physical assets (e.g., sensors, actuators, network segments, power, compute/storage/visualization Grids), through retrospective informatics and analytic services, to coupling of real-time interactive sensing networks with predictive modeling services. This paper summarizes the current activities in LOOKING, and then focusses on the development of a sense and response framework that operates on a continuous basis with real-time data. The sense and response approach is further defined through a series of use case scenarios from which requirements can be drawn, ending with a discussion of the cyberinfrastructure representaton of an instrument.

I. INTRODUCTION

The Laboratory for the Ocean Observatory Knowledge INtegration Grid (LOOKING) is identifying, synthesizing, and assembling existing and emerging information technology concepts into a durable cyberinfrastructure (CI) architecture for ocean observatories. The goal of the effort is federation of ocean observatories into an integrated knowledge Grid.

From a scientific perspective, LOOKING is developing a CI design that supports dynamic harnessing of resources ranging from physical assets (e.g., instruments, network segments, power, compute/storage/visualization Grids), through retrospective informatics and analytic services, to coupling of real-time sensing networks with predictive modeling services. The design addresses scalability, extensibility, and reconfigurability. Fulfilling this mission enables LOOKING to achieve the ultimate vision of a fully autonomous sensor network capable of evolving and adapting to changes in user requirements, available technologies, software, and middleware, or environmental changes during the life cycle of the ocean observatory paradigm.

From a system engineering perspective, LOOKING is defining the system-level CI architecture requirements, developing working prototypes that enable the seamless operation of a federated ocean observatory, and implementing selected elements on test-beds of opportunity. Once the CI is defined and prototyped, revisiting existing and emerging hardware designs under a strong system-engineering framework will be required to ensure compatibility and produce a state-of-the-art ocean observatory system.

The product of LOOKING is an example of a service-oriented architecture [1,2] which allows users and developers to dynamically link instruments, infrastructure, data archives, modeling programs, and visualization systems by defining standard protocols and interfaces without concern for internal functionality. Some example use cases and one possible architecture based on web services is described in [3].

II. FEDERATED SYSTEM ARCHITECTURE

Agreement on the high level system architecture (Fig. 1) for LOOKING was essential to the prioritization and planning process. It separates the effort into three major subsystems which, together with an overarching system architecture definition process, forms the basis for categorizing development activities, prioritizing risks, and identifying the core architectural components that make up the foundation of a federated ocean observatory architecture.

A. Related Efforts

Existing observatory initiatives can be mapped into two categories: 1) data Grids focused on externalizing data, and 2) service Grids which extend that functionality to externalize the behavior of the resources producing the data (i.e., the instruments). Data Grids include BIRN, NVO, SEEK and GEON. LEAD, NEON and LOOKING are service Grids as well as data Grids. However, four salient differences emerge when comparing the LOOKING objectives with those of other observatories:

1. Expanding the application of a federated system of physical observatories from use by one community of interest to simultaneous use by many independent, diverse communities;
2. Shifting emphasis from retrospective inquiry into archived data to real-time inquiry to respond to what is being observed;
3. Significantly increasing the number and diversity of instruments;
4. Extending the use of observatories to include

interaction and experimentation with the environment.

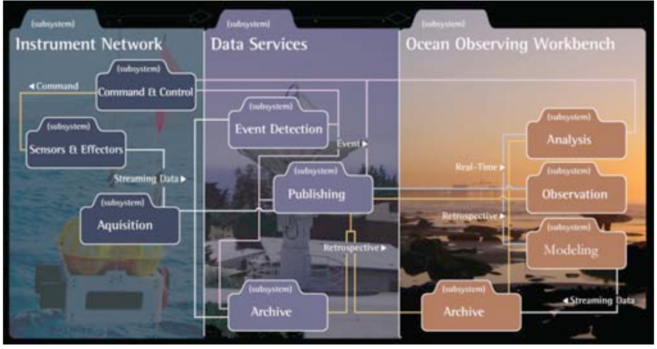

Fig. 1. UML diagram summarizing the three major subsystems in the LOOKING architecture. Together with an overarching system architecture effort, work in these areas comprises the major LOOKING activities.

LOOKING is using these differences to define where existing efforts can be leveraged and where investment in innovation is required. In particular, LOOKING is utilizing federated data management and archive system designs established by other projects. Delegation of such key elements frees LOOKING to concentrate on the high-risk aspects of federated ocean observatory CI.

B. Design of the Core Architecture Services

LOOKING is focused on characterizing (i.e., designing and prototyping) the software components required to support continuous real-time interactivity with a system of heterogeneous instrument networks (including the supporting infrastructure for the instruments) for an open set of communities of interest. This multi-dimensional requirement has been broken down into four architectural concerns: 1) a sensing framework for the continuous analysis of streaming data to identify events and patterns of significance; 2) a response framework for the real-time evaluation and enactment of protocols based on emerging events and patterns of significance; 3) a coordination framework to establish a scalable relationship between resource and application components across loosely coupled, distributed environments that are resilient to changes and advancements in the application, technology and context; and 4) a governance framework to facilitate sharing of resources based on agreements between the constituent members of an observatory. In this context, resources (where services are a class of resources) provide capability, while applications utilize resources to achieve an outcome.

Fig. 2 decomposes the first two of these frameworks into three major elements and 12 actions. The first (sense) element centers on the acquisition of knowledge, and spans acquisition, validation, assimilation and modeling actions. As a result, it is part of the sense framework. The sense actions may be distributed across an ocean observatory from seafloor instruments to land-based data management systems to geographically-dispersed data assimilation models. The second (deliberate) element focusses on decision analysis, occuring when an event is detected within the sense element. It includes identification, hypothesis formulation, evaluation, and decision actions, and hence links the sense and response frameworks. The third element (act) occurs when some action is required based on the outcome of the deliberate element. Encompassed actions include planning, scheduling, execution, and verification, and hence are located within the response framework. The outcome of the act element may require modification of the sense element, so that the three components operate in closed rather than open loop form.

The remaining LOOKING frameworks cross-cut with the sense and response areas, and are further described in Section V. For example, the coordination framework must reconcile resource availability with application requirements. The governance framework must ensure that different applications do not interfere with one another. The importance of this is only beginning to be appreciated, and will have a profound impact on ocean observatory CI architecture.

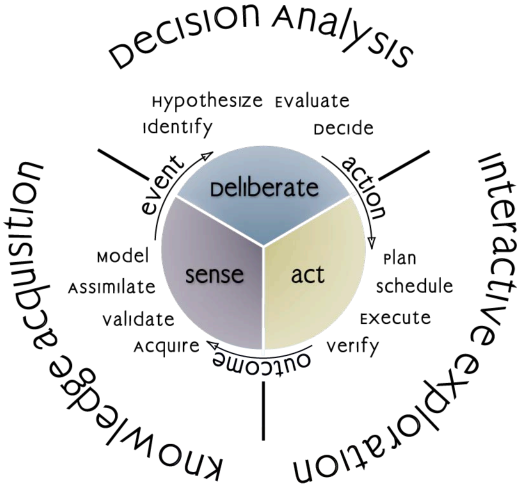

Fig. 2. Diagram defining the three major elements (sense, deliberate and act) and twelve actions in a combined sense and response framework.

III. LOOKING ACTIVITIES

Establishing the requirements and outcomes which define the major frameworks is being accomplished through three main thrusts combined with an overarching use case scenario-driven system requirements effort. The primary initial emphasis is on the sense and response frameworks decomposed in Fig. 2 onto which LOOKING activities may be mapped. The activities also map in a different way onto the subsystems in Fig. 1.

A. Ocean Observing Workbench

The objectives for this activity are: 1) develop retrospective analytics that support sophisticated scientific data query, retrieval, and analysis with support for spatial and temporal queries, 2) demonstrate the value of workflow systems which allow scientists to design scientific experiments and execute them efficiently using emerging Grid-based approaches to distributed computation, and 3) develop real-time visual tools based on integrated streaming data analytics and visualization services as part of workflow environments

The first activity is defining how oceanographic science can be advanced by means of interactive 3D data visualization and animation tools. The current state-of-the-art in oceanography is relatively primitive: data are typically stored in flat files or simple databases, visualization tends to be 2D, and even the ability to easily answer questions such as "what measurements are available at this place and time" are lacking. To change this, an iterative approach is being taken: adapt certain tools and techniques that have proven valuable in other science domains (e.g., astronomy) to oceanography, work with oceanographers as they utilize these tools and techniques, and evolve them based on feedback obtained from use cases. The initial focus is on bathymetry using the NASA World Wind client.

The remaining two objectives are being met through advanced information and scientific visualization techniques tightly coupled to data analytics software. The effort is centered on requirements definition and the adaptation of visual metaphors and software systems to examine subflows in the simulation of small-scale ocean circulation. Appropriate rendering of these high-resolution flows will allow comparison with ocean observatory instrument data, thus driving ocean data assimilation procedures. The second step is investigation of a user interface to facilitate interactive 2D and 3D exploration, aggregation, and analysis of data. The project is focusing on the design of integrated data analytics and visualization services as part of workflow environments (e.g., NCSA's "Data to Knowledge" (D2K) or Kepler/PtolemyII). The ultimate goals are incorporation of visual metaphors from the prototype visualization projects, and integration of them into a workflow environment.

B. Interactive Instrument Cluster

As an early demonstration of remote use of high definition television (HDTV), real time imagery of seafloor hydrothermal systems was transmitted in compressed form from the remotely operated vehicle Jason to the University of Washington in September 2005. The uncompressed imagery was mixed with another HDTV stream depicting a group of land-based students, researchers and teachers interacting with scientists at sea, and transmitted to the iGrid Conference at UCSD. This effort required significant technology push to simultaneously connect a high bandwidth link from ship to shore and a high bandwidth, high quality of service cross-country link.

One of the challenges of incorporating real-time HDTV into ocean observatories is processing the vast amounts of data to detect events of interest. The clear potential exists to collect more data than can be handled with present procedures. A second activity is demonstrating an algorithm to process HDTV automatically using an autonomous system to detect events of potential interest in the scene, tracking those events if moving and classifying them into categories (such as type of animal).

C. Ocean Observatory Services

The overarching focus of the ocean observatory services activity is the development of a sense and response framework (Fig. 2) that operates continuously in real-time. To this end, initial designs and prototypes for basic components of core subsystems of the framework are being defined, including data stream representation, real-time data stream processing, and instrument management.

The representation of data sets (e.g., data files) are often the focus of data management and archive systems, and only recently has attention turned to the unique characteristics and requirements of streaming data (Fig. 3). For broad and interoperable adoption of this data type by federated systems of ocean observatories, the characteristics of data streams must be defined and modeled. The primary goal is development of standard models and interfaces that represent the key behaviors and characteristics of data streams. The models must be based on the requirements of the data stream users (both the observatory infrastructure and the end user), and validated using actual data streams.

A data stream processing capability that expands and contracts to meet the demands of dynamic real-time monitoring and analysis is also being developed. Initial investigations are based on the hypothesis that workflow statements are the clearest expression of scientific inquiry in a continuous real-time observatory context. To this end, the project is focused on the integration of data stream, workflow and Grid technologies into a data stream processing (DSP) Grid. This is conceptualized as a workflow engine with the capability to dynamically marshal data stream resources, plan the decomposition of the process graph, and dispatch the component work across Grid resources. Anticipated uses of a DSP Grid include analysis and archiving of real-time data in a structured and semantic fashion, autonomous control of the source instrument network(s) through statistical analysis of data, enabling scientific and statistical analysis and control of realtime data via customizable systems, guiding the scientist through the discovery and binding process, and allowing simulations as testbeds for possible observatory networks.

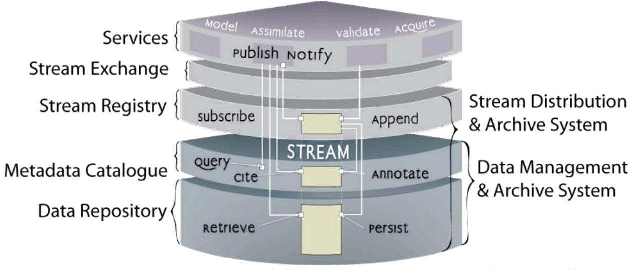

Fig. 3. A layered architecture capable of providing real-time streaming data. Note that the top layer is the sense framework of Fig. 2.

The concept of an instrument is central to a federated system of ocean observatories, and spans all four LOOKING frameworks. Instruments participate in relationships crossing numerous application domains, and must accomplish a complex set of interactions in a dynamic context. Providing a simulation environment for the definition, testing and verification of instrument interactions (both characteristics and behaviors) is an essential task in the architectural design of observing systems. The practical outcomes will be a model for the coordination of the computational components and artifacts in a loosely-coupled distributed application environment, and the identification and evaluation of a set of techniques to define the unified notion of an instrument. The approach must be iterative and evolving. The initial focus is on a restricted set of relationships, including instrument management in the application contexts of monitoring, lifecycle management, and access administration. Instrument management encompasses deployment, configuration, calibration, and integration into the infrastructure, monitoring (e.g., determining status) and remote interaction (e.g., requesting data). This includes both coupled computational components (e.g., the bidirectional symbiotic relationship between simulation models and instrument operations) and traditional operator-directed instrument control.

IV. USE CASE SCENARIOS

Definition of an ocean observatory CI architecture must be driven by the needs of the scientific community who will use it. Extracting requirements from the science community presents a challenge to the design team responsible for implementation, as the typical science user cannot readily quantify present and future needs that will lead to a formal design, and may not be familiar with the relevant information technologies.

To resolve this dilemma, the design team has to construct a wide range of use scenarios incorporating representative suites of sensors and platforms in close collaboration with a broad group of potential science users, and derive from these a set of initial requirements for CI. The initial requirements are used to define an initial architecture which is compared to additional use scenarios. Initial implementation of selected elements is then carried out for evaluation by scientists. This process continues iteratively through several cycles in an evolving spiral design process [4] carried out by an interactive team of scientists and CI specialists.

The LOOKING requirements were initalized using those from existing ocean observatory projects (such as NEPTUNE; http://www.neptune.washington.edu) and ongoing activities of the US ORION program (http://www.orionprogram.org). This approach allowed the LOOKING activities of Section III to be started while more elaborate use case scenarios were constructed and vetted by representatives of the science community. A brief overview of four developing use scenarios are described in rank order of complexity. The top level requirements are then extracted from them.

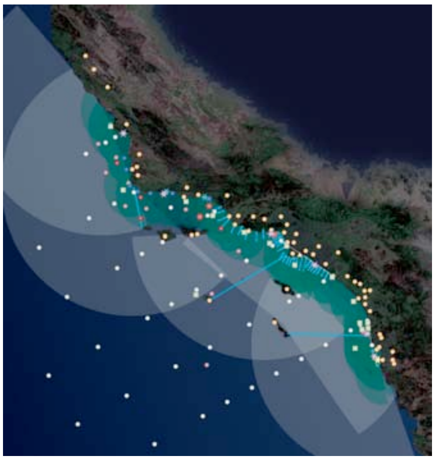

Fig.4. A large coastal ocean observatory consisting of short and long range CODARs and a diverse mix of buoys and short cabled observatories.

A. Observatory Operations and Resource Management

Fig. 4 shows a large coastal observatory comprised of long and short range coastal radar (CODAR) nodes and a mix of buoys and glider tracks covering most of offshore southern California. This constitutes a regional sense and response framework for coastal sciences processes and events composed of semiautonomous resource nexuses (e.g., discrete buoys). At the node level, resource allocation (e.g., power or bandwidth) is comparatively simple and can be implemented in local hardware or autonomous software. However, coordinating large numbers of nodes into a coherent scientific whole which is larger than the sum of the individual parts is a significant challenge. For example, linking the functionality of CODARs up and down the coast without human intervention is a major science requirement. Management of diverse types of data and their associated metadata is another. CI is needed to provide automatic control of these and other aspects of the overall observatory sense and response framework.

In a very real way, the concept of a sense and response framework is important at the operational as well as the scientific level. One of the major operations and maintenance challenges for a distributed ocean observatory such as that in Fig. 4 is tracking and coordinating the state of observatory resources. Thus, the science use case is also the operations use case.

B. Adaptive Operation of Instruments

Traditional data assimilation models operate in open loop form, with data being incorporated into the model run either retrospectively or in real-time. Dynamic data-driven application systems (DDDAS; see [5]) close the loop by allowing modification of sampling by the assimilation model. In a simple scenario, the assimilation model may change sample rates for selected instruments in response to an event. A more complex scenario has the assimilation model steering instruments on a mobile platform (such as a ship) to locations where property gradients are largest in the simulation. An even more complex scenario (Fig. 5) might incorporate the addition or removal of fixed or mobile instruments from the domain of interest in response to model output.

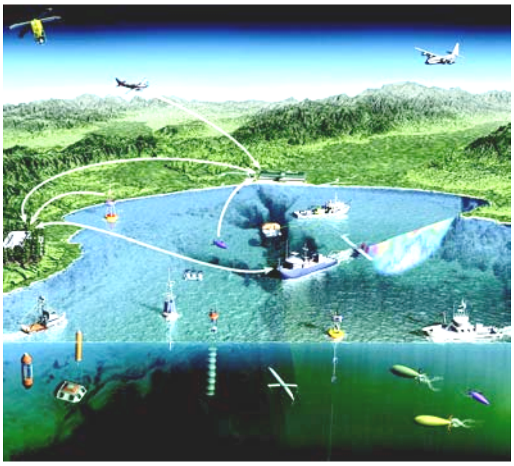

Fig. 5. Cartoon showing fixed and mobile instruments being added to or removed from the domain of interest and linked to shore-based DDDAS models (gray arrows).

Accomplishing a DDDAS scenario with fixed instruments requires a wide range of resource allocation, instrument control, and instrument communication services to coordinate the functionality of the assimilation model, the instrument suite, and the ocean observatory infrastructure. If some of the instruments are mobile or the sensor mix changes with time, then additional services for discovery and localization or tracking may be needed. Cross-cutting requirements for time synchronization and security services also exist. However, the primary communication path in this scenario is between dispersed instruments and terrestrial assimilation models, resulting in a comparatively simple network topology.

C. Remote Multi-Mission Laboratory

A more elaborate use case encompasses many heavily instrumented sites distributed around a regional cabled observatory (e.g., ten or more multidisciplinary moorings extending through the water column). This adds additional complexity through shared use of instruments and resources by multiple users and the difficulty of remote coordination of resources over large distances.

Fig. 6 depicts a single science site in this use case, where a diverse suite of sensors and actuators are deployed over a small area (for example, on the scale of a hydrothermal vent field) to accomplish multidisciplinary science. The sensor suite may include physical, chemical, and biological types, and the science mission may require frequent changes in their location or mix. Heavy use of stereo HDTV and high resolution acoustic imaging are anticipated, with concomitant demands on bandwith and power resources. Acquisition and storage of physical samples for later retrieval and onshore analysis may be needed. Accurate repeat positioning of actuators for sampling may also be required, which imposes closed loop control constraints on the hardware and software infrastructure.

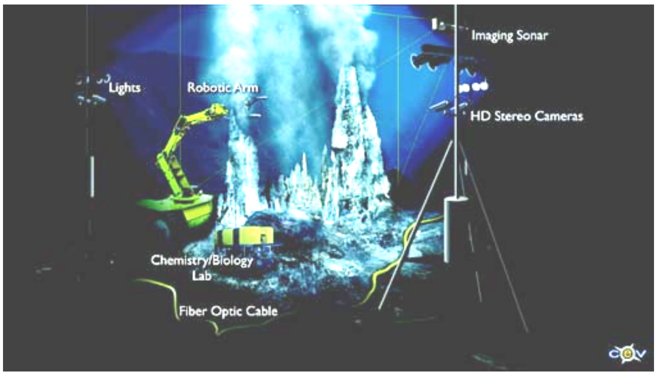

Fig. 6. A remote multi-mission laboratory for intensive physical, chemical and biological studies of a rapidly changing hydrothermal vent feature. See text for discussion.

This use case imposes stringent demands on the shared use of instruments and resources by many users. Quality of service, latency, and jitter requirements implied by real-time stereo HDTV and closed loop control of sampling actuators are stringent. From a CI perspective, a diverse set of services for resource allocation, time synchronization, instrument monitoring and control, bi-directional instrument communication, cross-calibration, coordination of sensing regimes (e.g., optical or acoustic), localization, tracking, and security are required. Closed loop control may not be feasible in the presence of high seafloor-to-shore latency without CI assistance, such as that used in remote surgery applications.

As an example of resource allocation complexity, consider an application which requires a substantial increase in either power (for example, because a light or pump needs to be energized) or bandwidth (for example, because an HDTV camera is turned on). Due to the distributed nature of a regional observatory, any increase in resources devoted to a given application must be coordinated, or else interference with other applications (in either instruments or the infrastructure) could ensue. For example, on a power system distributed along a submarine fiber optic cable, power delivery will be limited by line properties and load locations [6], and will have to be managed dynamically to avoid wholesale collapse when loads change. Dynamic resource allocation services are required to broker requests for changes in resources and to manage the available resources across the entire ocean observatory.

D. Mobile Instrument Platforms and Sensor Networks

Looking a decade into the future, the sensor suite at ocean observatory sites of interest may consist of a mix of large numbers of low capability, low cost fixed sensors (e.g., for the measurement of temperature over an area) and small numbers of high capability, high cost sensors (e.g., in situ spectrometers) in mobile platforms (Fig. 7). This combination simultaneously accomplishes continuous areal-scale, high resolution and directed, local-scale resolution measurements in an economical fashion. The enabling technology which makes this approach feasible is a network of high bandwidth optical modems [7] which provide a wireless extension of the observatory infrastructure, both making it possible to accommodate large numbers of sensors without physically attaching them to the observatory and allowing real-time access to fixed sensors and mobile platforms. The mobile platforms may operate continuously to accomplish pre-programmed sampling missions or under human control for exploratory sampling. Arrays of sensors which fuse into coherent sensor networks are a rapidly evolving application in terrestrial monitoring [8]. This can be accomplished by either linking all sensors to an optical modem network or through pervasive, direct peer-to-peer interconnection Since the characteristics of the terrestrial wireless and seafloor optical environments are similar, it is reasonable to expect both methods to be widely utilized on the seafloor in the future.

Fig. 7. Cartoon illustrating several autonomous underwater vehicles carrying out coordinated operations within a network of optical modems.

This use case aggregates all of the requirements of the previous three scenarios, involving both resource intensive applications and an ever-changing mix of mobile sensors which are complex in their own right and whose operation must be coordinated in real-time. Additional services to provide for discovery of topology and location-aware routing in a time-varying network may be necessary. Sensor networks may also require group management and collaborative information processing applications. A cross-cutting requirement is one of simplicity; for example, low cost sensors with wireless links may not have the capability to process complex time services.

V. INTEROPERABILITY OF RESOURCES AND APPLICATIONS

The requirement for a sense and response model coupled with increased resource capabilities enabled by cabled observatories necessitates a CI design that goes well beyond the traditional oceanographic linear approach to observation of “deploy, measure, retrieve, and analyze”. The preceding use cases are progressively more complex in terms of composition and coordination between applications and resources. Distribution of tasks is inextricably tied to the delegation of risk and responsibility. In addition, the requirement for dynamic federation of resources across multiple physical ocean observatories results in virtual ocean observatories having further specialized integrated capabilities. For the requirements of these classes of use cases to by fulfilled in a scalable, reliable and secure manner, the CI must address the interoperability and governance issues imposed on the solution by working across technical, operational and jurisdictional boundaries.

The service-oriented architecture (SOA) concept has received considerable attention in recent years. While it must be acknowledged that SOA is an important design principle and approach, it is not sufficient by itself to serve the needs of federated ocean observatories. LOOKING is working from the premise that there are three core interacting architectural elements of a federated ocean observatory on which the other system components are built:

> A Resource Framework that addresses the means by which resources are shared, including enforcement of terms and the management of resource usage;
>
> A Governance Framework that establishes and manages the identities and roles of observatory participants, as well as their relationships with the resources executing on their behalf or at their request;
>
> A Data Communication Framework that recognizes that the interoperability of data is realized in the communication processes between system components.

These three frameworks must tie together present and future projection of requirements and capabilities, and map onto a three layer representation of an instrument. The top layer is a prototypical autonomous instrument that is imbued with models of operation so that it can carry out a series of observation assignments as tasked by a mission planning service. A local instrument agent controls sensors and actuators to execute the assignments, and provides feedback to the mission planning service. The second layer constitutes the programmable instrument infrastructure which enables the functionality of an autonomous instrument in a loosely-coupled, distributed environment. The bottom layer contains the resource management services which link diverse instrument applications to the observatory resources.

The overall approach taken in LOOKING is abstraction of the concept of an instrument and prototyping of the frameworks needed to support multiple instruments as a federated managed resource. The ultimate objective is provision of the tools and infrastructure components that are required for continuous, automated investigation so that scientists can focus on science, engineers can focus on scaling the observatory elements, and operators can effectively react to system change.

REFERENCES

[1] Foster, I., “Service-oriented science”, *Science*, vol. 308, pp. 814-817, 2005.

[2] Hey, T., and A.E. Trefethen, “Cyberinfrastructure for e-science”, *Science*, vol. 308, pp. 817-821, 2005.

[3] St. Arnaud, B., A.D. Chave, A. Maffei, E. Lazowska, L. Smarr, and G. Gopalan, “An integrated approach to ocean observatory data acquisition/management and infrastructure control using web services”, *Mar. Tech. Soc. J.*, vol. 38, pp. 155-163, 2004.

[4] Boehm, B., A. Egyded, J. Kwan, D. Port, A. Shah, and R. Madachy, “Using the WinWin spiral model: A case study”, *IEEE Comp.*, vol. 32, pp. 33-44, 1998.

[5] Darema, F., “Grid computing and beyond: The context of dynamic data-driven applications systems”, *Proc. IEEE*, vol. 93, pp. 692-697, 2005.

[6] Howe, B.M., H. Kirkham, and V. Vorperian, “Power system considerations for undersea observatories”, *IEEE J. Ocean Eng.*, vol. 27, pp. 267-274, 2002.

[7] Farr, N., A. Chave, L. Freitag, J. Preisig, S. White, D. Yoerger, and P. Titterton, “Optical modem technology for seafloor observatories”, *Proc. SSC06*, Dublin, Ireland, 2006.

[8] Chong, C.-Y., and S.P. Kumar, “Sensor networks: Evolution, opportunities and challenges”, *Proc. IEEE*, vol. 91, pp. 1247-1255, 2003.