OOI CYBERINFRASTRUCTURE
NSF OCEAN OBSERVATORIES INITIATIVE

OCEAN OBSERVATORIES INITIATIVE

OOI – CyberInfrastructure

Requirements Workshop
Data Product Generation
University of California, San Diego,
La Jolla, CA
May 20-21, 2008

Workshop Report

Elizabeth Rosenzweig, Alan Chave, Michael Meisinger

CANDIDATE

August 2008

NSF
CONSORTIUM FOR Ocean Leadership
DISCOVERY • UNDERSTANDING • ACTION
UCSD

Document information

Project	Ocean Observatories Infrastructure (OOI) CyberInfrastructure (CI)
Document Custodian	OOI CI Architecture & Design Team (ADT)
Approval	Report: Alan Chave (Woods Hole Oceanographic Institution), OOI CI IO System Engineer User Requirements: Frank Vernon (UCSD), OOI CI IO Deputy Director
Created on	January 24, 2008
Last Changed on	August 28, 2008
Document Status	CANDIDATE

Document History

Date	Version	By	Description of Changes
2008-05-15	0.1	M. Meisinger	Initial workshop report from notes and previous workshop report structure
2008-07-25	0.2	E. Rosenzweig	Added content for presentations, scenarios and updated all information to reflect an accurate report of the workshop.
2008-08-07	0.3	E. Rosenzweig	Integrated review comments from A. Chave, added exported requirements list from DOORS
2008-08-12	0.4	A. Chanve	Candidate review
2008-08-12	0.5	E. Rosenzweig	Integrated review comments
2008-08-13	0.6	M. Meisinger	Polishing and formatting; changed status to CANDIDATE.
2008-08-28	0.7	E. Rosenzweig	Updated section 3.6 (Antelope technology) with input from K. Lindquist

Table of Contents

OOI - CyberInfrastructure

Data Product Generation
Requirements Workshop at UCSD, May 20-21 2008
Outcome and Summary

1 Executive Summary

In an effort to further the understanding of community user requirements and constraints for the planned Ocean Observatories Initiative (OOI) CyberInfrastructure (CI), the OOI CyberInfrastructure Implementing Organization (IO) has held a series of topic-oriented workshops with scientists and other future users of the CI. These workshops refine and complement the results of prior workshops and requirements efforts that led to the successful completion of PDR in December 2007.

The workshop covered in this report was targeted towards Data Product Generation and was held May 20-21, 2008 at the University of California, San Diego (UCSD) in La Jolla, CA. This meeting was the fourth in the series and succeeded three prior requirements workshops held between July 2007 and May 2008 (see [CI-RWS1], [CI-RWS2], [CI-OOP-WEB]).

Oceanographers and data producers from the community as well as from the Regional, Coastal, and Global Observatories of the OOI were invited to the workshop, as well as the system engineers from the OOI observatories. The workshop goals were CyberInfrastructure science user requirements elicitation and documentation, the validation of existing requirements, as well as a continuing outreach measure to future CI user communities. Calit2 at UCSD provided the scientific environment for a two day workshop that covered introductions to the planned CI and the OOI program, oceanographic science presentations, CI requirements elicitation and validation sessions, domain modeling and use case scenario development sessions.

The workshop outcome and results include

- CI user requirements elicited from data product generating scientists and engineers
- Refinement and validation of existing user requirements
- Domain models elaborated during the workshop
- CI use case scenarios for ocean observing elaborated during the workshop
- Collection of workshop presentation materials on the OOI CI Confluence web site [OOP-WEB]
- Science user questionnaires for requirements elicitation (extended and short versions)
- Completed participant questionnaires

2 Introduction

2.1 Goals and Background

In order to provide the U.S. ocean sciences research community with access to the basic infrastructure required to make sustained, long-term and adaptive measurements in the oceans, the National Science Foundation (NSF) Ocean Sciences Division has initiated the Ocean Observatories Initiative (OOI). The OOI is the outgrowth of over a decade of national and international scientific planning. As these efforts mature, the research-focused observatories enabled by the OOI will be networked, becoming an integral partner to the proposed Integrated and Sustained Ocean Observing System (IOOS; www.ocean.us). IOOS is an operationally-focused national system, and in turn will be the enabling U.S. contribution to the international Global Ocean Observing System (GOOS; http://www.ioc-goos.org) and the Global Earth Observing System of Systems (GEOSS; www.earthobservations.org). Additionally, the OOI will provide an ocean technology development pathway for other proposed net-centric ocean observing networks such as the Navy's proposed Littoral Battlespace and Fusion Integration program (LBSFI). Additionally, the global community spanning Canada, Asia, and Europe are also developing new ocean networks which all contribute to the GEOSS. Developing a robust capability to aggregate these distributed but highly linked efforts is key for their success.

The OOI comprises three distributed yet interconnected observatories spanning global, regional and coastal scales that, when their data are combined, will allow scientists to study a range of high priority processes. The OOI CyberInfrastructure (CI) constitutes the integrating element that links and binds the three types of marine observatories and associated sensors into a coherent system-of-systems. The objective of the OOI CI is provision of a comprehensive federated system of observatories, laboratories, classrooms, and facilities that realize the OOI mission. The infrastructure provided to research scientists through the OOI will include everything from seafloor cables to water column fixed and mobile systems. Junction boxes that provide power and two-way data communication to a wide variety of sensors at the sea surface, in the water column, and at or beneath the seafloor are central to these observational platforms. The initiative also includes components such as unified project management, data dissemination and archiving, and education and outreach activities essential to the long-term success of ocean observatory science. The vision of the OOI CI is to provide the OOI user, beginning at the science community, with a system that enables simple and direct use of OOI resources to accomplish their scientific objectives. This vision includes direct access to instrument data, control of facility resources, and operational activities, along with the opportunity to seamlessly collaborate with other scientists, institutions, projects, and disciplines.

A conceptual architecture for the OOI CyberInfrastructure was developed and published by a committee established by JOI in 2006 (see http://www.orionprogram.org/organization/committees/ciarch) [CI-CARCH]. It describes the core capabilities of such a system. Initial requirements were derived from similar cyber-infrastructure projects.

In May 2007, a consortium led by SIO/UCSD, including JPL/NASA, MIT, MBARI, NCSA, NCSU, Rutgers, Univ Chicago, USC/ISI and WHOI, was awarded a contract to be the Implementing Organization (IO) for the development of the OOI CI. The first six months of the design phase has focused on architecture and design refinement and consolidation, and an initial science user requirements analysis and community involvement effort. In December 2007, the preliminary CI design [CI-PAD] was successfully reviewed in a PDR (Preliminary Design Review) by a panel of independent experts appointed by NSF, who provided very positive review comments.

Current activities are targeting the Final Design Review (FDR) in November 2008, where all requirements and design documentation, operations management plans together with cost estimates and feasibility analyses will be reviewed. Major activities towards FDR focus on completing a baseline set requirements at all levels of the OOI and CI, covering user, system and subsystem requirements, with clear tracing to OOI science objectives [SCIPROSP] and user community expectations. Further activities target advancing the CI design and that of its subsystems to the next level to be ready for the start of OOI MREF construction. During all OOI design and construction activities, the validation of any previously elicited and documented user and system requirements through the community will remain a primary concern. Direct involvement of prospective CI user communities is of paramount importance to the success of the program. The requirements elicitation and management process is planned to be an ongoing activity in close collaboration with the user communities involved throughout the design and construction phases.

Earlier science user involvement occurred during the first CI requirements workshop (RWS1), July 23-24, 2007, at Rutgers University and the second CI requirements workshop (RWS2), January 23-24, 2008, at UC San Diego. The third requirements workshop was held on May 13-14, 2008, at WHOI. For each of these workshops, the outcomes were summarized in the form of publicly available reports [CI-RWS1, CI-RWS2, CI-ROOP].

This report covers the outcome of the fourth requirements workshop on Data Product Generation (DPG). The workshop took place May 20-21, 2008, at UCSD's California Institute for Telecommunications and Information Technology (Calit2) in La Jolla, CA. It was the fourth in a series of CI architecture and design team organized workshops to identify and elicit requirements from domain users. It was focused on user requirements for the generation of data products from raw data acquisition through data product quality control to their preservation and distribution. In this context, users may include instrument providers, observatory operators and scientific/educational users, among others. The workshop concentrated on three aspects of data product generation:

- Instrument/actuator data acquisition/transmission using direct access, data streaming or other means as well as their command, control and monitoring. Intertwined issues include authorization, authentication and policy as they affect data acquisition/transmission.
- Instrument lifecycle management ranging from initial configuration to data segmentation/streaming to data product generation through processing services. As for data acquisition, cross-cutting issues of security and access, and of policy, must be considered.
- The requirements for data preservation and distribution, including models for data and metadata as they relate to syntactic and semantic access, data archival and data discovery, publication and subscription. In addition, association and citation of attributes with data resources will be considered.

2.2 Science Background

Over the last decade the community has identified high priority science needs, and the OOI has been designed to quantitatively address these issues. Addressing the high priority questions is especially critical as the oceans are changing in our lifetimes, and developing a quantitative understanding of this behavior is crucial to understanding the possible trajectories of these changes along with their potential impacts on human society. The OOI will provide scientists with a sustained presence in extreme ocean environments that will enable fundamental discoveries. Given the need to develop a quantitative picture of the ocean, scientists require spatial time series spanning many scales across a range of marine biomes. The OOI will accomplish this by deploying a distributed but linked infrastructure in regions that are disproportionately important relative to their geographic size. This distributed infrastructure will collect data that will allow fundamental processes to be parameterized across a range of marine systems. The full spatially-distributed OOI network will be required to quantitatively test our understanding of the high priority science questions. Additionally, the network will provide a rich time series database that will enable hind-cast scientific analysis for decades to come.

One challenge for the CI will be providing the capability to take raw sensor data streams and generate data products that will directly address the science priorities of the community. This process is complex, and encompasses processes spanning raw data acquisition, data quality control, data preservation, and distribution. As just one example, many of the high priority science questions are focused on the physical regulation of primary productivity in the ocean. The core sensors required to address these issues are bio-optical sensors (absorption/attenuation meters, backscatter sensors and fluorometers). These sensors nominally provide raw data in volts that, after instrument-specific calibration coefficients are applied, can provide estimates of specific radiometric units. Historical re-analysis will require detailed metadata on this process, allowing any data products to be changed when appropriate using improvements in the calibration coefficients over time. For example, it was only appreciated after five years of use that new calibration coefficients for absorption/attenuation meters needed to incorporate both salinity and temperature corrections. These required many in the community to go back to old data and reprocess them. Scientists will then use the radiometric quantities to derive proxy estimates for phytoplankton biomass, phytoplankton health, and occasionally, phytoplankton community composition. These derived products will be shared by the scientific community; however, there will always be debate on the most effective and useful transformations to apply, and this necessitates access to the various stages in the sensor data, calibration and transformation procedures. These new evolving derived *in situ* products will be compared and used to calibrate new optical parameters observed by NOAA/NASA satellites (Figure 1). This is a small example illustrating the range of issues that the CI will need to support.

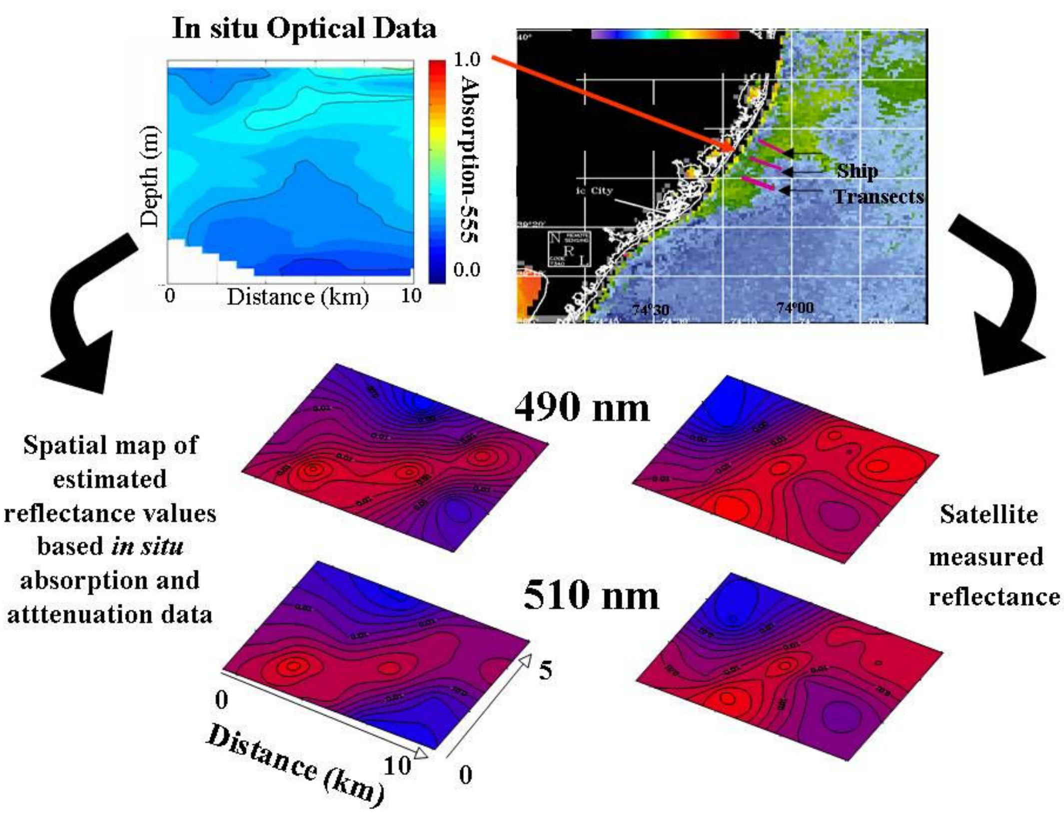

Figure 1: A typical data flow for optical measurements

Figure 1 describes a typical data flow for optical data. In situ optical profiling instruments during a cross-shore experiment collects a data transect used to derive surface ocean reflectance measurements using the radiative transfer equations. These derived reflectance measurements are then compared to satellite estimates of reflectance.

2.3 Outline

The remaining parts of this report are structured as follows: Section 3 summarizes the presentations given at the workshop and places them into the context of the scientific background. Section 4 documents the direct workshop outcomes, such as discussion summaries, domain models, elaborated scenarios, and prioritized requirements. Section 5 lists all CI user requirements and highlights the ones identified in this workshop. Section 6 documents participant feedback and conclusions from the organizers. The appendices contain further details about the workshop organization and background materials.

2.4 Preparation

The CI Architecture Design Team (ADT) has refined and adapted the previously existing questionnaire with relevant questions for user requirements elicitation that was structured into selected categories. A shortened and tailored version of the questionnaire was sent to the workshop participants. The scientists were asked to provide answers to the questions prior to the workshop. Appendix A of this report documents the participant questionnaire.

Each scientist was asked to prepare an overview presentation covering projects, research interests and relevant background information related to the OOI CI. The presentations were supposed to address the main topics covered by the questionnaire. The presentations covered approximately 15-20 minutes each, including questions.

2.5 Acknowledgements

This report was developed by the OOI CI ADT; it contains input from many sources, such as the workshop presentations by the organizers and invited science users, the completed participant questionnaires, the CI preliminary architecture and design, OOI science background information from the project scientists, and notes taken by Emilia Farcas, Michael Meisinger and Elizabeth Rosenzweig. This report contains summarizing and general statements extracted from meeting materials by the organizers.

We profoundly thank the participating scientists for their time and efforts during the workshop, and their valuable contributions to the OOI CI requirements elicitation process. Furthermore, we would like to thank them for their efforts in filling out the participant questionnaire and providing further materials after the workshop, and for reviewing and validating this report.

2.6 Disclaimer

The contents of this report reflect the understanding and analyses of the CI ADT based on written workshop notes and general background. Errors in transforming them into this report are the responsibility of the CI ADT. No statements in this report are verbatim quotations of participants; there were no audio recordings of the discussions taken during the workshop.

3 Presentations

3.1 OOI CI User Requirements Elicitation Process

Alan Chave (Woods Hole Oceanographic Institution), OOI CI System Engineer, welcomed the workshop participants and described the process for science user requirements elicitation. The OOI project is preparing for final design review in November 2008. A set of important activities covers completing and refining user requirements for the OOI integrated observatory with the cyber-infrastructure component as its “face”. This workshop’s goal is the collection of new requirements and validation of existing user requirements by science users involved with ocean observing programs. There will be other requirements workshops focusing on different topics.

Chave presented the requirements elicitation process (see Figure 2) and described the purpose of systematic and iterative requirements elicitation efforts involving multiple user communities over the course of the OOI CI project.

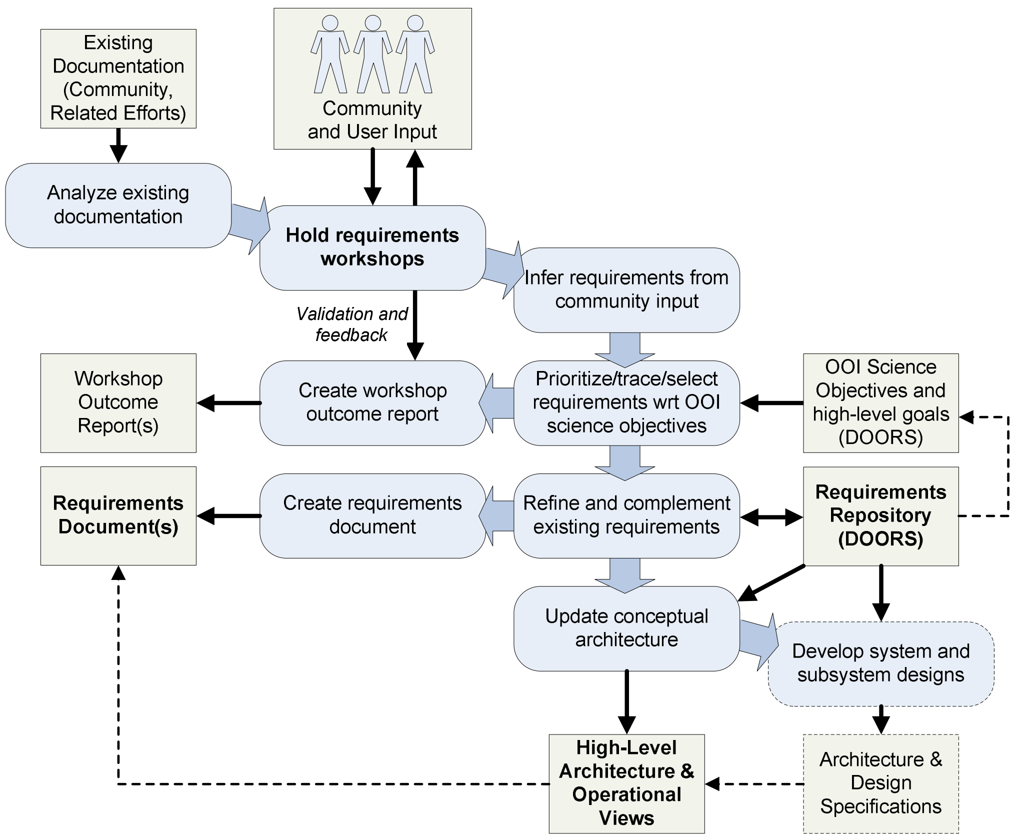

Figure 2: CI user requirements elicitation process

3.2 CI Overview, Requirements, Architecture

Matthew Arrott (UCSD/Calit2), OOI CI Project Manager, provided an overview of the OOI CyberInfrastructure project and the CI project organization. The main goal of the CI is supporting the three main research activities of observing, modeling, and exploiting knowledge through a set of well-rounded resources and services. The CI infrastructure will be distributed across the country and will have points of presence at the sites of the main OOI observatory components on the east and west coasts.

The design process involves several iterations that advance the understanding of requirements and design. Previous design cycles lead to the conceptual architecture, the proposal for the OOI CI and the refinements for Preliminary Design Review (PDR). The current iteration emphasizes refining requirements and design for FDR.

3.3 Project and Research Overview: Neptune Canada Infrastructure

Benoit Pirenne (University of Victoria) presented a description of the Data Management and Archive System (DMAS) for NEPTUNE Canada and VENUS projects. VENUS is a coastal cable system, with the first node operational in the Saanich inlet array since February 2006. The Saanich site was installed at 95-meter depth. The node includes half a dozen instruments on a 3 km cable. The instruments include cameras and hydrophones.

The second node came online in February 2008 in the Straits of Georgia. This node has similar instrumentation to the first node, but sits at the end of 40km of cable. Another node is planned at 300 meters depth to come online in September 2009, in the Fraser Delta.

NEPTUNE Canada is a regional-scope cabled observatory. The installation began in fall 2007 with the laying of 800 km of fiber optic cable. The shore station installation is currently ongoing. There are plans to install 5 nodes in May 2009 and to continue installation of about 120 instruments in July-August 2009.

DMAS is the cyber-infrastructure environment for NEPTUNE and VENUS that was started in 2005 with minimal requirements and a mandate to capture all data and make them available to end users. The directive was to keep data for 25 years in an archive. Finally, the mandate stated that the system must allow interactive control of the instruments.

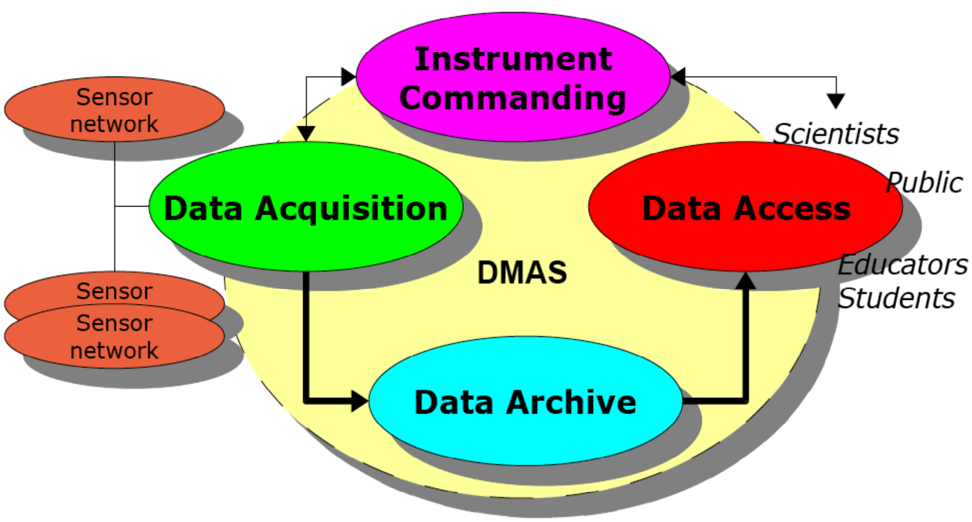

Figure 3: DMAS architecture

The Neptune infrastructure includes:

- Wetplant
- Shore station
- Archive center
- Users

The component representing users includes:

- Scientist
- Web services
- SQL
- Public

Data acquisition is done via instrument drivers, with parsing available for about 20 types of instructions. The software is hosted at the shore station. Data storage takes place at a data center on shore. This center has a database with 500 million individual scalar measurements from VENUS and is currently migrating to the Oracle database. The complex data and streams are stored as files, with metadata in the database that currently contains about 1 TB of data from VENUS.

Data access for Venus uses Web 1.0 forms to research and request VENUS data. Web 1.0 forms are the current baseline for Neptune; it provides geographic-based access and views of the seafloor topology (nice color and design, easy to understand). All DMAS functionality is shore-based with data collection via an IP channel, so that the device driver makes a TCP connection to an IP address using current technology.

The DMAS team uses agile development, creating new software versions every other week. Currently there are three development teams with regular iterations in place. This process starts by talking to users, and getting their feedback on the system. The necessary changes are made to external computing/storage resources. The Agile development process also includes collaborating with instrument manufacturers and interacting with USN/DND (US Navy, Department of National Defense in Canada).

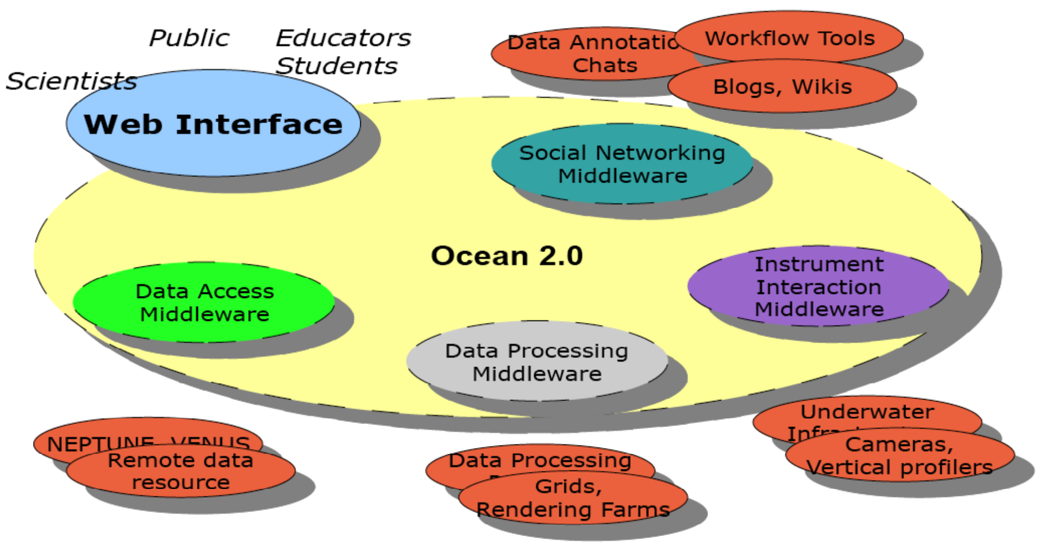

Figure 4: Ocean 2.0 architecture

<u>Specific Statements:</u>

- Since the nodes are expensive, the installation should be well planned.
- The NEPTUNE mandate is to capture data, keep it archived and allow for interactive control. One of Neptune's requirements is to keep the data accessible for 25 years. As of now, there are no comprehensive requirements documented for DMAS. This is due in part to the lack of maturity of the DMAS user community. The challenge is creating a working community that shares a facil-

ity. In the past, work was done more individually, so there is no roadmap in place to handle the task of collaboration..

- Scientists will have to work more closely together in the future. There will be a large volume of data and we need requirements and expertise in many areas. For example one might want to be connected to relevant data or people and the system should facilitate those connections.
- Need Web 2.0 system to get scientists to work together, get an information visualization process with customizable social networking (Linkedin, Facebook) and web services that allow access to diverse resources (see Figure 4).
- Applied for funding for Ocean 2.0 and will get grant to do a two year project
- The DMAS needs to be able to interoperate with other sources of data.
- OOI-CI requirements efforts to date confirm our findings, but priorities need to be set on which get implemented when.
- Systems have been developed in isolation, so it is challenging to get them to work together
- Scientists using our systems understand things that are happening at several/s, so it is very time based, the frequency is what is important, to track patterns, etc.
- Issue of US Navy data being classified and therefore it is problematic to analyze them in a timely fashion. As a result, data can come in later and be processed as historic data.
- Whidby Island keeps their data and can analyze them. The USN keeps the data classified and use them themselves (long term sequestering of data, over many years), so if you need data from this stream you have to wait.
- Need tools that are interactive to test ideas and algorithms.
- Need tools that can deal with 24/7 data and serendipitous events, as well as cross correlating them with other data such as from satellites.
- New sensors are much more capable than before and produce orders of magnitude more data. The sensor networks are multidisciplinary in nature and include biology and earth sciences. The data are available in real-time with access for everybody.
- There is a sense of urgency in exploiting the data, but there are so many data that one can't catch up with them. There is also competition from the rest of the world.
- Users know they need more access to expertise about data. There are too many data and they can't afford to work in isolation anymore. Primarily, these users need tools to find colleagues, exchange results, and prepare publications.
- Ocean science is being revolutionized by a new generation of data providers
- The science community is starting to realize a data "fire hose" and is ready to welcome new ways to deal with it.
- Technology has evolved and proposes solutions to these challenges
- The hard part will be to build interoperability and standards between data providers

3.4 Project and Research Overview: MBARI Data Architecture

Kevin Gomes (MBARI) presented the Monterey Bay Aquarium Research Institute (MBARI) Data Architecture. MBARI has several ocean observation systems that are deployed in the Pacific Ocean off the coast of California at Monterey.

Kevin described the MBARI architecture as having the following components:

- Sensing and Acquisition
    - Instruments, Moorings, AUV, Free Ocean CO2 Enrichment Project (FOCE), ALOHA (University of Hawaii), benthic rover
- Infrastructure

- Ocean Application System for Interdisciplinary Science (OASIS), Software Inrastructure and Applicatons for MOOS (SIAM), Shore Side Data System (SSDS), data processing applications
- Data
  - Formats, transformations, standards
- Interfaces
  - Infrastructure, user, application programming interface (API), service-oriented architecture (SOA)

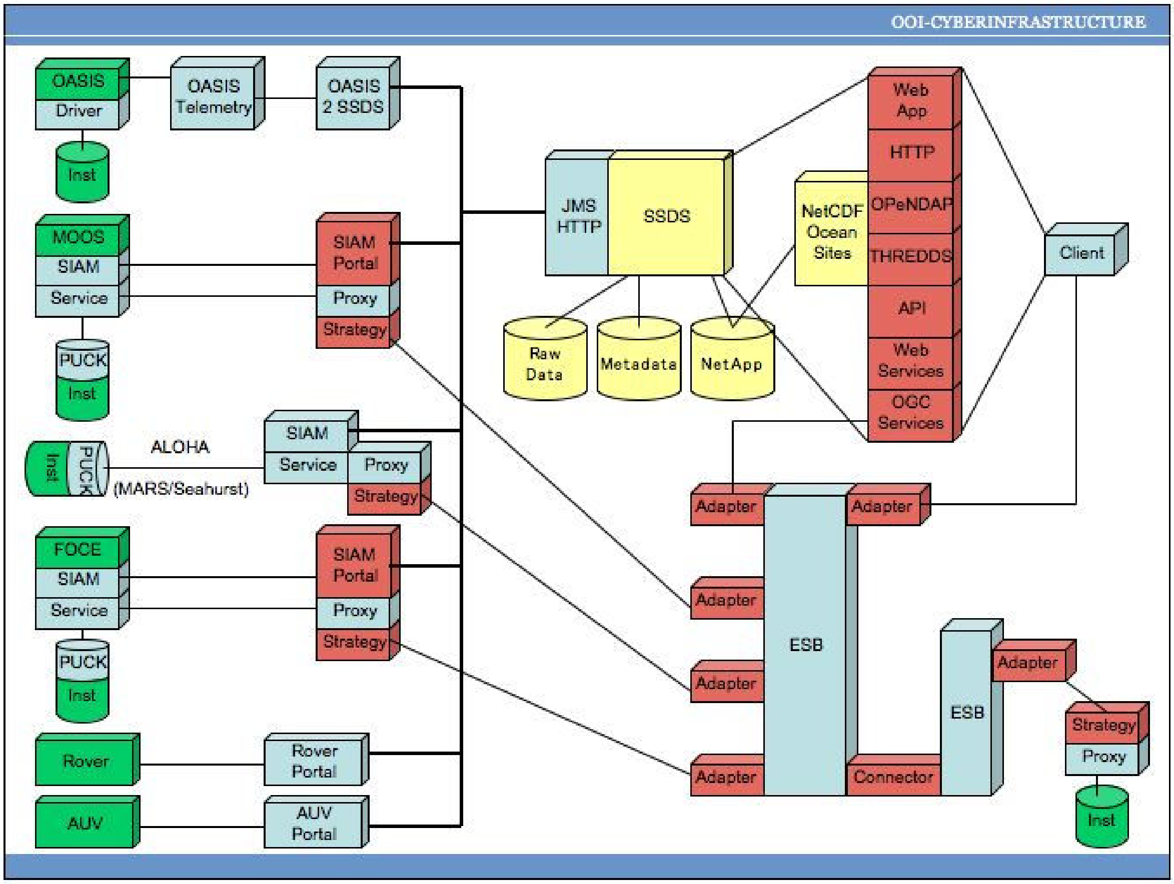

Figure 5: MBARI data architecture

The Ocean Acquisition System for Interdisciplinary Science (OASIS) is a legacy system that is pre-configured with everything needed to operate the system in one place. There nothing to configure once it is deployed. The binary compact data are processed shore side, converted to ASCII and metadata are attached using drivers that are part of the controller software. The instrument lifecycles are controlled by communication between the instrument and the shore station.. There are no metadata embedded in raw data; they are added later.

Monterey Ocean Observing System (MOOS) is a deep water oceanographic mooring system designed to deliver power, data and time signals to the seafloor in support of the subsea network of moorings.

Software Infrastructure and Applications for MOOS (SIAM) provides an environment that is a seamless integration of self-describing plug and play instruments into an observatory. SIAM uses drivers that travel with an instrument in a puck (plug and work) and provide the necessary data for integration into the observatory.

Shore Side Data Systems (SSDS) is a data management system that stores both raw and processed data (and associated metadata) for a wide variety of instruments.

MBARI performs some activities in an instrument's lifecycle automatically by intelligent commands sent to instruments. The higher level commands are still sent through human intervention in a direct communication session. Metadata are attached to the instruments and sent to the data archive when it is initialized. Rover stores its own data and then a portal processes data streaming packets to look like any other system in the infrastructure, thus seamlessly integrating with it. The AUV offloads data, then sends metadata with them into the archive.

The metadata schema was developed in house, and is not standards based.

Specific Statements:

- Make the CI so engaging that it works easily- as some examples seen in the recent Iron Man movie where a user interacts with large databases through touch screens, nice graphics, easy to find functions and voice technology. The current systems have bad user interfaces. They make users feel like they are fighting dragons and the CI is the dragon. INTERFACES ARE PARAMOUNT!
- "Give me your tired, huddled masses " (of data) yearning to be free (organized). The system can deal with them, put them in a format that you can use in a "balance between user interfaces and developer sanity"
- Need a balance between power, functionality, and ease of use. The systems have to do a lot, but it also has to be easier to get and use the data.
- Visualization and data acquisition perhaps are not balanced and that needs to be looked at- REQUIREMENT- visualization must render numerical data accurately
- Security policy not done, right now everything is open
- Data gaps are hard to locate due to bandwidth inconsistency, troubled connections, etc., so it is hard to differentiate between a gap and normal interrupted service
- Education and outreach needs work
- QC done in post processing
- Event detection is still up in the air, no definition of what an event is- work needs to be done here.
- Perceptivepixel.com is a good example of a power graphical user interface.

3.5 Project and Research Overview: Marine Metadata Interoperability Project

John Graybeal (MBARI) presented the Marine Metadata Interoperability (MMI) project. MMI is about building a digital community around metadata for observatory users who are interested in collaborating to solve science problems. The MMI team works to build community through workshops, training, presentations and posters. They have connected over 1000 people in the last 2 years. MMI has developed a 1500 page website and regularly sends out information on 10 different mailing lists.

MMI is a good snapshot of current solutions, but no one individual solution encompasses everything. New systems are being developed to help improve the way science is being done in the ocean. Systems are bringing together information and providing data in many forms. John suggested that highly visual user interfaces will enhance the performance and product generation of a system. He also presented some sample interfaces

John showed some novel ideas for user interfaces as part of his presentation. The first one looks at information from a cruise and arranges it graphically on a screen for an imaginary system. Figure 6 is a good example of clear, color guided information.

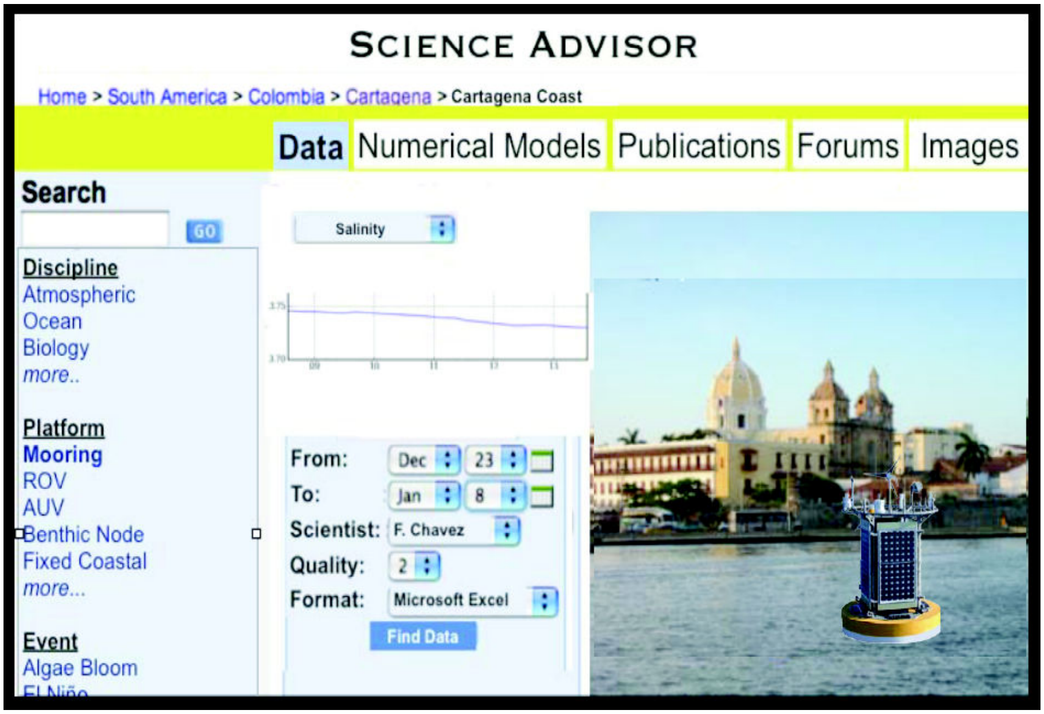

Figure 6: Design for Science Cruise Plans

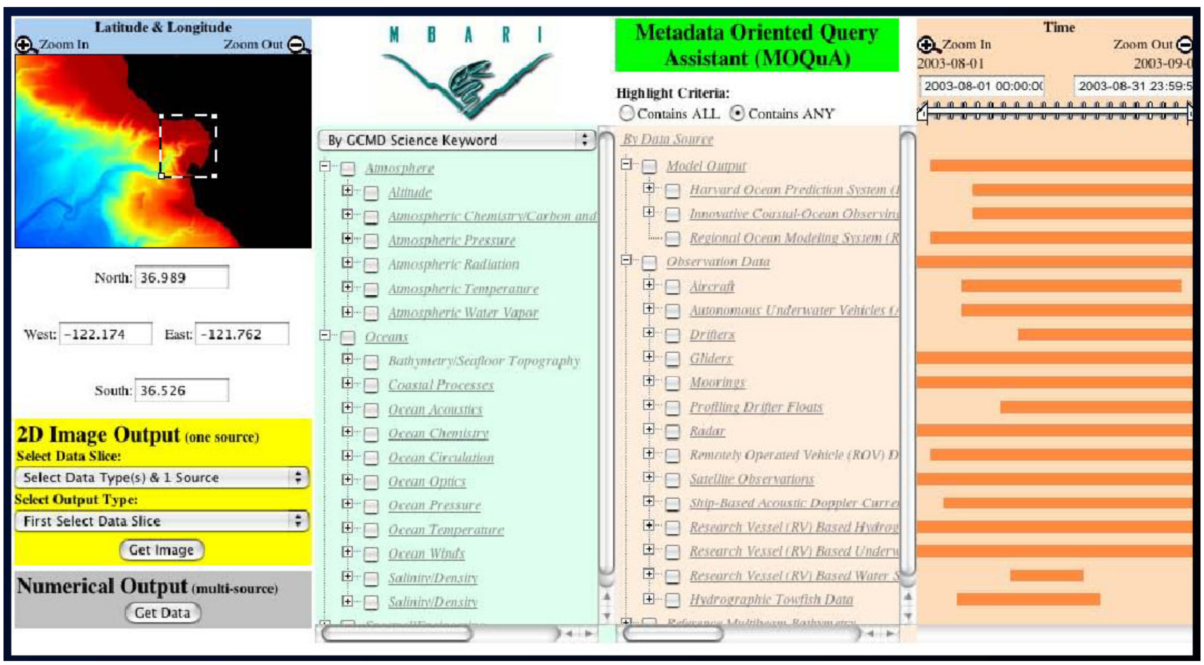

Figure 7: Doing Science: Another Option

John also showed another screen he entitled "Doing Science: Another Option"; see Figure 7. This demonstrates other ways to put together data by combining information from different sources on a single user interface screen. The figure represents John vision of how this new user interface might look.

MMI sees the need for interoperability at multiple levels: instruments and systems need to interact, which is very problematic and doesn't work well now. Across the world, there are many different platforms that operate together conceptually but utilize different formats. Furthermore, systems must make it easy for users to access data.

Quick Data Access Form (v1.0)

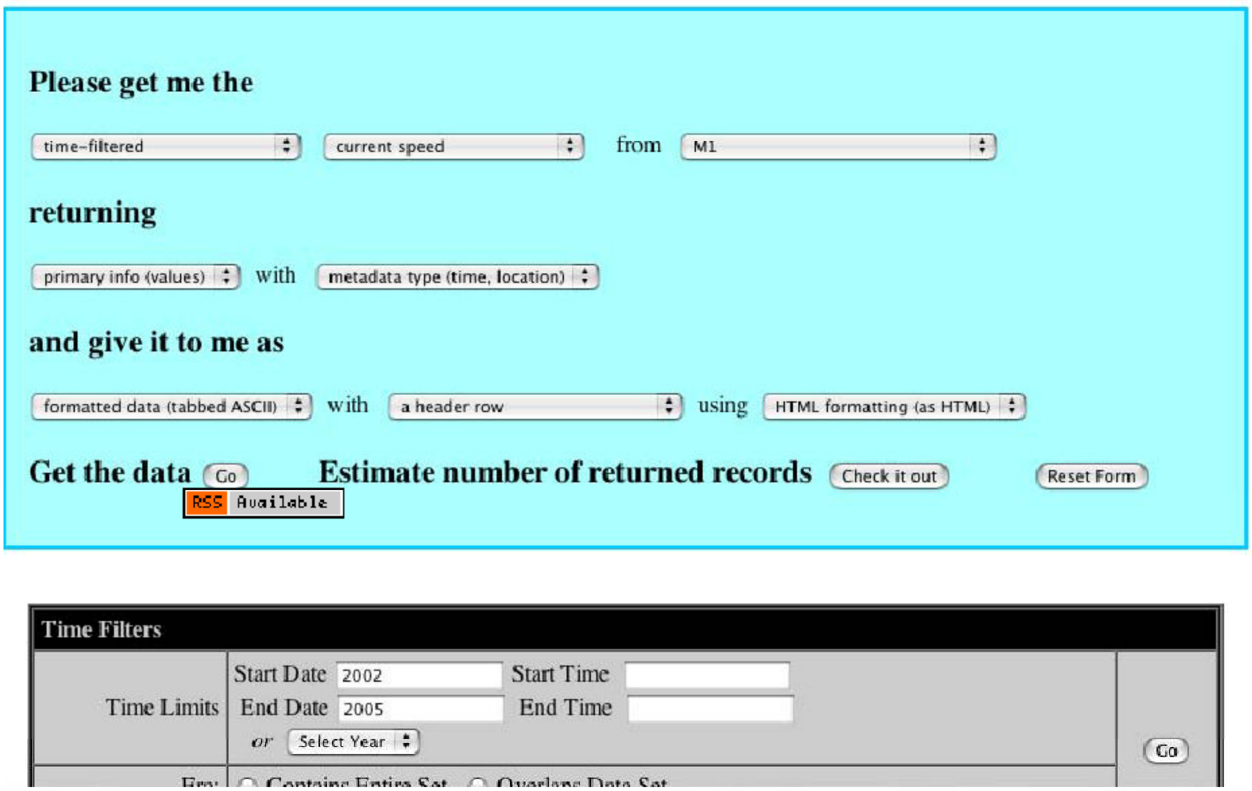

Figure 8: Doing Science: The Way It Is and Will Be

<u>Specific Statements:</u>
If you are asking users to enter metadata, the process has to be simple and not ask for too much information at the beginning. There are variable names and vocabulary that need to be captured such as "temperature, conductivity, pressure, data, time"; this vocabulary is usually mapped within the community of use. However, few formal marine vocabularies exist; most are not good, not common and there is no agreed upon format. Furthermore, there is currently no way to connect either the metadata or the architecture to nurture them.

The primary problem with science systems today is interoperability. There are many systems that collect data and are now coming together in so many different formats that it is a "Tower of Babel" that is causing confusion. The science community must be in agreement on several areas:

- The way to transfer the metadata: transport protocol
- What data to transfer: content standard

- What the contents mean: vocabulary

Semantic interoperability is accomplished when distributed and heterogeneous systems are able to solve issues related to controlled vocabulary.

The presentation discussed metadata as having critical importance and is the key to enabling secondary (reuse) of data. It is time consuming to complete metadata records, but it is essential to insuring that the system is up to date.

What is needed for marine metadata to work well is system interoperability. This means that subsystems must communicate, select, track, and tag the metadata as they go. One way to handle the issue is to determine how the infrastructure will control the metadata standards. The system has to set the acceptable bar for the standards, so the metadata will be useful.

People come looking for data but are not finding what they are looking for. This is a particular problem if the terminology is not the same. This causes many of the problems. Therefore, terminology is very important.

Quality assurance and control is on ongoing process. The person who benefits from metadata is not the user who creates them. Time will show that the people who are using the data are not scientists but people who are simply interested in learning about the ocean.

3.6 Technology Overview: Antelope

Kent Lindquist (Lindquist Consulting, Inc.) presented Antelope, which is an open architecture environmental monitoring system from Boulder Real-Time Technologies, Inc (http://www.brtt.com) as candidate technology for the OOI CI. Antelope is Unix-based, providing several hundred building-block executables that can be assembled into engineered near-real-time systems. Antelope also includes the documented toolkits used for the construction of these executables, allowing the creation of more tools and monitoring applications. The available libraries and tools have fostered the evolution of a well-established base of open-source contributed code that expands on the capabilities of the core system. The three main aspects of Antelope are an embedded near real-time relational database called "Datascope;" a store-and-forward data streaming system called "ORB" based on a client-server object ring buffer and open "ORB" protocol for reliable packet delivery; and finally a real-time systems framework for sophisticated field processing and central-observatory systems control.

The Antelope object ring buffer is a completely event-driven, message-oriented middleware layer used for both data-distribution and inter-process communication in distributed systems. The ring-buffer server itself is reliable and fault-tolerant, providing non-volatile storage and lossless multicast streaming of monitoring data and messages. The buffering technology itself is data-neutral, though transported packet streams can be addressed by content, allowing format-aware processing at other layers of the integrated monitoring system.

Real-time systems built on Antelope generally cover the whole spectrum of monitoring-system tasks, including:

- Data acquisition
  - Data communications
  - Data and information buffering
  - Data and information flow
  - Automated data processing

- Automated data and information archiving
- Real-time data and information integration
- Data and information-product distribution and sharing
- Real-time system monitoring and control
- Real-time graphical user interfaces

A standard Antelope real-time system is composed of field-interface modules that talk to the deployed field digitizers and instruments in the network; orb-based data communication elements such as orb-servers and transport modules; automatic processing programs where applicable, and data-archiving and redistribution utilities.

Antelope runs under Solaris, Linux, and Mac OSX, uses standard IP protocols (TCP) for wired and wire-less network communication, and wherever possible standard protocols (TCP/UDP/Serial) for communi-cation with deployed instruments. The network-transparent, asynchronous communication allows scalable support from low-power field processors to enterprise-class servers, including both low and high-bandwidth applications and either continuous or intermittent communications. Support is included for prioritization of data delivery, cron-style scheduling of periodic tasks, as well as some dynamic system reconfiguration.

Antelope currently interfaces to 30+ types of data acquisition systems and can interface with many other data servers. Programming interfaces include a well-developed C interface, script interfaces in Perl, TCL/Tk, Matlab, Python, PHP for Web support, and several others. The sensor types supported fall into two categories: those supported by the core Antelope applications and those added by the contributed-code community. The core sensor types include:

- Seismometers
- Accelerometers
- Displacement
- Infrasound
- Hydroacoustic
- Strain

All of these leverage a well-developed architecture inside Antelope for handling time-series data. Further dataset support has been added by the community of open-source contributors (many of these provided by the NSF-funded ROADNet project):

- Barometric pressure
- Temperature
- Wind speed
- Wind direction
- Differential pressure gauges
- Solar insolation
- pH
- Electric current
- Electric potential
- Dilution of oxygen
- Still camera images
- High-frequency coastal radar

3.7 Technology Overview: SRB/IRODS

Raja Rajasekar (UCSD/SDSC) presented an overview of the Storage Resource Broker (SRB) and iRODS projects. SRB is first generation Data Grid middleware developed at the San Diego Supercomputer Center (SDSC) The SRB includes a distributed file system based on a client-server architecture that allows users to access files seamlessly across a distributed environment based upon data attributes rather than just names or physical locations. The SRB replicates, syncs and archives data, connecting heterogeneous resources in a logical and abstracted manner.

The SRB has a diverse user base and a wide range of applications. The collections at SDSC include:

- One Petabyte, 200+ Million files
- Multidisciplinary scientific data
    - Astronomy, cosmology
    - Neuroscience, cell-signaling & other biomedical data
    - Informatics
    - Environmental & ecological data
    - Educational (web) & research data (chemical, physics, …)
    - Archival & library collections
    - Earthquake data, seismic simulations
    - Real-time sensor data
- Growing at 1TB a day
- Supporting large projects: TeraGrid, NVO, SCEC, SEEK/Kepler, GEON, ROADNet, JCSG, AfCS, SIO Explorer, SALK, PAT, UCSDLibrary

The iRODS project is a second generation data grid that performs data virtualization with the following attributes:

- A distributed file system based on a client-server architecture.
- Allows users to access files seamlessly across a distributed environment based upon data attributes rather than just their names or physical locations.
- It replicates, syncs and archives data, connecting heterogeneous resources in a logical and abstracted manner.

iRODS is a distributed workflow system – policy virtualization

- Policy is a first class "object" to be managed
- Long-term policy captures provenance
- Policy can be declarative/descriptive as well as procedural/normative
- Policy can be a process enforced on new "objects"
- Policy can be a constraint whose integrity can be checked at any time

3.8 OOI Infrastructure Design: RSN Instruments

Deb Kelley (UW) described the instruments and scientific inquiries posed by the Regional Scale Nodes (RSN). The presentation focused on the instruments and specific experiments and not on the RSN itself.

The biggest challenge and issue was in putting the sensors on the large infrastructure. Further issues include:

- A lot of stringent archive requirements because of the necessity for co-registration in time and space (e.g., precision to a second)
- Minimum response time of sensor determines maximum sample rate of sensor
- Availability of site characterization data
- Nested scales

- Environmental data
- People installing sensors off site need live video feeds to see how sensors are installed
- Site characterization in general and in detail for new sensors;
- How does the environment change over time for complex physical, biological, chemical processes
- Ex: 36 temp measurements in volts, then converted to physical units
- Physical samples include minerals
- Archiving data and metadata is a nightmare. No real formal handle on how to do this.
- There are x,y,z issues: No formal protocol for how to do coordinate transformations using standards
- For microbes, there are many different ways to process samples
- Petrological data base PET-DB

Specific statements:
- These meetings need to capture requirements for the OOI for more specific operations in collecting samples and accessing data.
- Is the CI going to provide capabilities attuned to different data types such as video and low bandwidth images.
- Make sure the results of the analysis get back into the system.
- "We are finicky about our testing. The overhead expense for using ships to deploy instruments is costly and it is expensive to make mistakes"
- Oceanographers don't like change much. They want baby steps; they are in the nuts and bolts and are still trying to solve them. They want help with the following:
  - Organize databases
  - Organize procedures
- "Oceanographers will have to understand and accept the ideas that a given data set is not 'final' but is subject to intermittent improvements over time (improvements can result from gaining more understanding about the context in which the data set was produced). With this idea comes the need to identify data set versions, access the 'best' or 'most recent" versions, be notified of the existence of a new version and identify the difference(s) between versions."

3.9 OOI Infrastructure Design: CGSN

Bob Weller (WHOI) presented information on the Coastal and Global Scale Nodes (CGSN). The presentation discussed the data collected by CGSN and how it can be incorporated with CI in support of CGSN nodes.

The CGSN includes the Pioneer and Endurance arrays on the coasts and the global nodes in the Irminger Sea, Southern Ocean, and Ocean Station Papa. Fixed assets include surface and subsurface moorings. Mobile assets include ocean gliders and AUVs.

CGSN data walkthrough for getting data from instruments to the database.
1. Data directly sensed at the node
2. Data collected by CI in support of node
3. Data fusion sets, including maps
4. Pre-deployment data include archives, used for field check and verification
5. Deploy sensor, deployment data hosted by CGSN to community
6. During deployment, operational data hosted by CGSN to community
   a. Real time data (QC/QA)
   b. Delayed mode data (QC/QA)

   c. User feedback interface. Tools
   d. Global updates and fixes
7. Recovery data hosted by CGSN to community
   a. Shipboard field cal/val
   b. Metadata
   c. Raw dump, real-time check
   d. Analysis
   e. Calibration
   f. Version control

<u>Specific Statements:</u>

- Always keep raw data before any transformation is done
- In lab, calibrate instruments, burning them against known standards to reveal problems before they leave for the ship
- Most instruments are not connected to anything so they rely on internal clocks, when you dump data you do lose clock accuracy. They found their worst case gives an 8 day drift which they find annoying. That didn't happen often.
- Metadata and data should all be searchable
- Need to contextualize data, can this be done automatically? The context becomes important in order to understand the data and use them for various end purposes
- Aggressive construction of a data set is what most people are doing, but we are moving to a world where most people want to communicate about the data set as part of a community dialog.

4 Workshop Outcome

4.1 Questionnaire Response Analysis

The CI ADT received substantial input from the participating scientists through the questionnaires that were handed to them prior to the workshop with the request to provide answers to as many questions as possible. The input from the questionnaires went directly into refining and validating the science user requirements. Selected statements are listed in various sections throughout this report, in particular the individual participant presentation sections and the general requirements discussion section.

4.2 General comments

- Good to hear scientists talk about their work and what they are struggling with. It would be good to sit down with a prototype of a user interface for a usability test. Either a low fidelity or high fidelity usability test should be included when developing the CI.
- Good to hear about the process of deploying instruments. Next stage would be to produce a template workflow, and check with scientists to see whether it is correct.
- Good to learn about CI tools and hear about the phases and work of CI
- As an engineer, it is good to hear input from the science community. Sometimes it has been hard to know how to contribute. Didn't see the goal of the two days and that would have been better to contextualize a bit.
- Got a lot out of discussion, even though they were sometimes parochial, it was good to hear the presentations about the systems. Virtual lab discussion was interesting, but didn't see it as part of the critical path - perhaps it was the engineering perspective
- See huge gains for the CI, and appreciated the data discussion. Thinks it could be a huge impact, historical tool for OI.
- Concern that although we were given a wish list of things to solve, can CI achieve the scope of issues that were presented today. Thinks spiral approach is going to be realistic.
- Encouraged to hear comments that point to a constructive path that brings salt-water people and CI community together.
- Happy to have brought in outside input and can learn a lot from process. Need to focus on getting it done and don't focus on the wish list. In the future, clearly define the CI interface, where does it stop and where do the scientists work begin.
- Helpful to see the definition of good and bad requirements
- Iterative process is good, but do it on the fly, it has been the way it has been done IDR projects and thought it worked.

4.3 Requirements Discussion Summary

The requirements discussion was structured along the extended questionnaire as documented in Appendix A. This section documents facts and statements made during this session. Stated requirements were added to the list of requirements in CI Science User Requirements section below.

4.4 CI Use Scenario Definition

In this session, the charge for the workshop participants was to brainstorm and discuss a hypothetical use scenario for a transformative community cyber-infrastructure.

4.5 Scenario 1: Physical Samples - from phenomenon to product

This scenario looked at how data is derived from a physical sample. The first step in any data-driven experiment is collecting data. In experiments that are run from ocean and land observatories, once the physical samples are collected, they are measured and classified by ingesting the data into a data management system. Then the data are analyzed and turned into a data product.

General Scenario Data are input to an information system that helps to support the process of identifying samples, turning samples into data and then performing data processing to create a data product. The assumption is that the physical samples are collected on a ship at sea, where there is a crew and a scientist on board the ship running an experiment that collects samples. The samples are collected, preserved, and analyzed on the ship. Decisions are made in situ regarding which samples to take back home to the lab and add to the collection.

WHO: User

- Science data user person
- PI and Chief Scientist
- Staff scientist- ODP world interface or watchdog keeps track of whole process
- Instrumentation technicians
- Resident technicians
- Engineer who collects data using instruments
- System engineer from IO
- CI
- Students

WHERE: Environment:

- Ship at sea
- ROV- Remotely operated vehicle
- Manned submersible

WHAT: Situation: Ship is out at sea collecting data, ship comes back to shore, and data go into database on shore. Tasks include:

- Verify recovered assets by matching them with record of deployed assets
- Verify assumptions about data
- Catch deviations
- Preserve collected data in recoverable representation
  - As DVDs in archive- probably created on ship
  - In archive as close as possible to the original data

WHY: Focus of scenario is the need to physically check instruments on a regular basis (annual, biannual, etc.) because the instruments are often left unattended for a year and these sensors could drift. If, as a rare occurrence, the sensor instruments are in a different location than where they were dropped, the data will have errors. The goal of the study is provide continuous data by doing sample collection
Physical samples are collected by lowering wire from a ship to open a bottle that collects 20 liters or so of water at 15-20 x, y, z locations. Right now measurements of physical samples, translated into data are collected on DVD made on ship.

Sampling activities will collect co-registered data of:

- Oxygen, salinity from sensors on same sampling device

- Activity log created during or after sample collection- need this information such as “time over, lat, long” ship is moving so that the info has to contain start and stop time, lat, long, etc. this gives region that has been studied
- Bridge log includes cruise information-who was on it

<u>Study Process:</u>
The data collection process is described below, including what actors are involved in each step.

1. Verify calibration sensors and make sure what was planned is what was deployed (Chief scientist and Marine IO-logistics lead, technical scientist)
   - Physical
   - CTD
   - With sensors on mooring
   - Technical document package (could a be 3 ring binder collecting all info)
2. This is where the CI provides contextual awareness
   - Weather, etc.
3. Review online checklist for standard operations (Marine IO)
   - Sensor calibration
   - Collect data
   - Catch deviations
4. Monitor system health (Operations group technician)
   - Checklist
5. Preserve data in a structured manner (Chief scientist)
   - As DVDs
   - In archive
   - With metadata
   - Calibrations
   - Structured
6. Verify calibration of sensors (Marine ops technician and Chief Scientist)
   - Physical
   - CTD
   - With sensors on mooring
   - Technical document
   - Perform data correction: e.g. say the data have been streaming for a year, people have been using them for a year, they will need to see version control (can you go back to previous version)
   - Measure two data sets as an alternative data product: e.g. surface mooring might have an old sensor there and you will put a new one in. They might put them both in at the same time, and then the time sync overlaps. Next, the new sensor data are collected. Some users will use only one set, some users might take only new data, and everyone might do this differently.
   - Replace uncompiled, low-resolution data with complete data that are available later (e.g., after physical retrieval)
7. Metadata added to enable transformation
   - Comments
8. CI provides version control of data sets
   - E.g., calibration corrections
9. Perform data corrections: community post processing of data (Marine ops technician and Chief Scientist)
10. Replace incomplete data-low res with complete data (contentious issue not always done the same, it depends on the data source and project and the nature of the connectivity)

- Summary
- Merge data sets, register them
- Capture citation, provenance

11. Feedback from user community: (OI scientist and users)
    - Email
    - Portal
    - Formalized user input
    - What: Flags, problem alerts
12. Community post processing of data
    - Register them, notify PI (Chief scientist)
    - Capture citation, provenance
13. Data processing, visualization
14. Preserve data in nonstandard way to uncover patterns, anomalies, etc.
    "Data nirvana" would be creating connections in data that a person might not see on their own.
15. Calibrate sensors at the end of the mission

Specific Statements:

- Chief scientist wants a tool that says: "I was on a ship that collected test files on this ship and I want to look at them now". He/she wants to see data from a ship while on the ship.
- Chief Scientist also wants to keep track of information such as weather, time, and location. He/she wants to correlate these with information from the bridge log on a ship. This is difficult to digitize because ship officers have to sign bridge log and it is a legal document.
- Contextual awareness information is already available on a computer on the ship. There is onboard information about environment, weather forecast, etc. If CI can catalog this and then make it available for cross-reference, it would be powerful.
- Users want the data to be reassembled in graphical form so they can quickly find the information they seek (e.g. detect a moving eddy in the vicinity of a mooring using ocean color data). Some of the information they look for in their studies include:
    - Currents
    - Surface waves
    - Salinity
    - Altimetry
    - Ocean color
- Users want "data nirvana"

4.6 Scenario 2: Part A Instrument Lifecycle

This scenario looked at the lifecycle of an instrument

Situation: How do you keep information for the entire lifecycle of the system?
This scenario looks at the lifecycle of a microbial sensor. This sensor looks at temperature, fluid, microbes and solids, and is considered an in situ sensor.

WHO User PI, Marine IO, etc., same as other

WHERE Environment: onboard ship

WHAT: The scenario looks at the sensor from conception to deployment to data collection to recovery:
The following steps describe the lifecycle of the microbial sensor deployed from a ship, including the users/actors that are involved in each step of:

1. Establish science driver (Principal Investigator)
2. Write a science driver (PI)
3. Collect science drivers (PI)
4. Develop an experimental design (testing, deployment, O&M) scheme which is included in PI proposal (Engineer and PI, involve Marine IO and CI)
5. Proposal gets reviewed
6. Iterate
7. Funding (at a reduced price)
8. Modify design (PI and Engineer)
9. Design, fabricate, test (PI, Engineer)
10. Constraints from CI and Marine IO
    - Estimate of costs
    - What is potential interference with other sensors?
    - Security issues
    - Data rate, power availability
    - Deployment cycle
    - Requirements on testing
    - Certification procedures
    - Education constraints
    - Hardware for hookup
        a. Connectors, dbase
        b. Cables
        c. ROV delays for deployment
    - Technical feasibility
11. Buy and fabricate all components (Engineer and PI)
12. Register instrument
13. Initial testing phase by PI and Marine IO
14. Write/adapt device drivers (PI and CI IO)
15. Testing by marine IO
16. Write and specify drivers, (PI, CI IO)
17. Integrate with junction box through shore based testing in lab (Marine IO)
18. End-to-end test by marine IO at test facility (Marine IO)
    - Dry
    - Wet (in a tank)
    - On simulator
    - Integration test
19. During development: Test deployment (optional) e.g. on MARS (PI and Engineer)
20. Register instrument with system (instrument provider to CI)
    - Obtain an ID
21. Final test before deployment on the ship using ship-based simulator ~ go/no go decision (PI and engineer, Chief Scientist, marine IO agent)
    - Turn on sensor to make sure everything works. Once there is assurance it does work then deploy. If it doesn't pass test, question is do you fix it or do you have to replace it
22. Have mission plan signoff by the following:
    - PI
    - Chief Sci
    - Ship Captain
    - IO Agent
    - Engineer
    - ROV operators
23. Deploy instruments (PI or delegate)

- Water samples taken
- Characterization
- Rock samples
- Video imagery
- Physical installation (ROV operator, ship operator, engineer)
- Power/communications connection

24. Commissioning instrument
    - Power on
    - Verify correct operation: Interrogating it-does it work? Bring up system
    - Data test period, QC
25. Instrument is operational
    - Data are flowing
    - Take raw data and perform calibration
26. Assumptions on resulting data
    - Standard labeling of physical samples
    - Standard format for data preservation
27. Preserve collected data in recoverable presentation (CI)
    - DVDs
    - In archive
    - Structures
    - Non-standard
    - With metadata for discovery
    - Calibrations
    - By CI
28. Verify calibrations of sensors (Marine ops and scientist)
    - Physical
    - CTD
    - With sensors on mooring
    - Technical document package
29. Add metadata to enable automatic transformation (development engineers and operations staff)
30. CI provides version control of data sets/versions
    - E.g. calibrations, QA/QC corrections
    - Perform data corrections (instrument providers, project members)
31. Merge two data sets as alternative data product (data processing/ scientist)
32. Replace incomplete data. Low resolution data with later complete data
33. Feedback from user community by (scientist, data analyst)
    - Email
    - Portal
    - Formalized user input
    - What
        a. Comments
        b. Problem alerts, flags
        c. Correspondence
34. Community post processing of data (public)
    - Register them
    - Capture citation/provenance
35. Data processing, visualization (data processor)
36. Define, develop, analyze, monitor, modeling process and tools (PI or designee)
    - Response capabilities in case of events
    - Prioritization
    - Event detection

    - Calibration verification
    - Procedures
    - Processing workflows
    - Record keeping requirements
    - Data visualization
    - Specific control interfaces
37. Develop technical documentation package
    - Published
    - In metadata
38. Define instrument use policy

4.7 Scenario 2: Part B Once the Instrument is Operational

The next piece of the process is to examine what happens when the instrument is operational and deployed. How do people interact with the sensor and what happens if there are problems while it is deployed.

<u>Focus for Part B</u>

- Scientific process
- Protocol
- Timing
- Checkpoints
- Staff (as it relates to process, who do you have involved, etc.)

<u>Process</u>:

1. Monitoring process
    a. Automatic failure detection (CI)
    b. Manual by PI
    c. Event response
        i. Set off other people's responses
2. Sensor down
    a. Automatic alarm that data have dropped out (or all 000s in data stream)
    b. Notification to marine IO, PI
    c. Auto restart hard failed
        i. Is there an auto restart in drivers, fault tree?
    d. If back to life:
        i. Investigation of cause
        ii. Make operational again
    e. If portal failure
        i. PI decision with marine IO
            1. Notify that it is down
3. Recovery cycle
    a. O&M plan
    b. Modification decision
4. Decommissioning (if sensor is still functional)
    a. Remove date stream from system
    b. Notification to interested users
5. Deprecation (no longer functional)
    a. Clean up as much as possible
    b. Instrument part of the seafloor/rock formation

c. Make record of data before decommission dates
d. Notify end of life

4.8 Scenario 3: Doing science "How do I use my data to do science and get my results out "

This scenario starts with a scientist who wants to use her data to do science and publish a paper. The assumption is that she is working in an integrated observatory that is integrating data from moorings and cabled infrastructure. This would be similar to the Endurance Array combined with RSN cabled infrastructure.

WHO: User: Principal investigator, chief scientist

WHERE: Integrated observatory, employs moorings and cabled observatory. An example would be the Endurance Array combined with RSN cabled infrastructure

WHAT: Situation: Science Question: What is the Variability of Ecosystem

Does the animal community structure change in the response to environmental events?

- Seismic event, Pacific northwest

Focus: Straw man hypothesis: A major earthquake event triggers a gas hydrate release that fundamentally disturbs the seafloor ecosystem and the water column.

Steps for science work:

1. Historical search, what data are available from the past, what has happened there before? Look at all the sources: chemistry, biology, geology baseline and literature search for stressors that cause change, what were responses, look for correlations
    - Data that exist, external sources
    - Need for ability to discover data-how would you do it? I.e. "Show me what data you have that correlate with this bug I am looking for and with the instrument that I am using."
    - Return data with time sync annotation correlated to activities for a sample population of animals in certain locations
    - Discover information in the following ways:
        - Google search
        - Visualization
        - Use search engine with standard vocabulary
    - Use personal contacts in the community, your personal network, to get information
2. Define baseline
    - Publish in system
3. Refine hypothesis:
    - Iterative process
4. Discover sampling capabilities of the integrated observatory
    - Check health of system: What instruments survived?
    - Look at Endurance array, Neptune Canada, other coastal nodes
        - What data are coming off, what is happening with the instruments
        - How do you know the sensors are there?
            - Find seismometers
            - Geodetic data, how much movement has there been
    - Look at sensors of a specific type in areas of interest
            - Physical
            - Chemical
    - Receive notification of events and status from sensors of interest

   - Visualization and evaluation of environment
   - Derived data products and applied processing
5. Evaluation of available data/status
   - Use available visualization, analysis
   - Define customized analysis and visualization
   - Assume that 10 resources provide standard visualization for standard measurements
   - Find ancillary data
     - Annotation
     - Calibrations
     - Associated metadata
   - Acquire data from a user's perspective
   - Start with visualization, map
   - Drill down to data sources-> quick look at the data and resources
6. Decide on data streams and subscribe
7. Request adaptation of existing resources
   - Change data request
   - Request high priority data transfer of past intervals
   - Notification of fulfillment of a request (or estimate)
8. Community process
   - Annotation of all existing resources
     - Flags
     - Unverified
     - Rating
   - Request for explanation
   - Assume resource-specific moderation process
   - Versioning
9. Select data representation formats for subscribed streams
   - Textual, numerical
   - Visual
     - Tool
     - Geometry
   - Files usable for specific tools
     - NetCDF
     - Map/Matlab
   - Binary/ASCII files
     - Get a sense of what to automate
   - Select 10 provided standard workflows for analysis and visualization
   - Snapshots of preprocessed data
     - E.g. flip book of data
10. Analyze
   - Standard navigation tools to browse data
   - Raw and processed data together
   - Summary statistics

NOTES:

- Steps 4-9 iterative process

Specific Statements:

- Change the way instrument providers do business so that they can supply the metadata and format that the scientist will need in the CI. Lots of ancillary data are valued by users. There needs to be a formal process that includes all they need. If a sensor is to capture it, then the specs need to be very clear at the outset of what we need.

- Annotations need to have a signature, know who did it. For various reasons it helps create trust in data if we know who is providing this information
- Users need to use data, what representation do I want data in? Visualization, numerical, etc.
- Users want help to find correlated data for research and CI should help them with that task

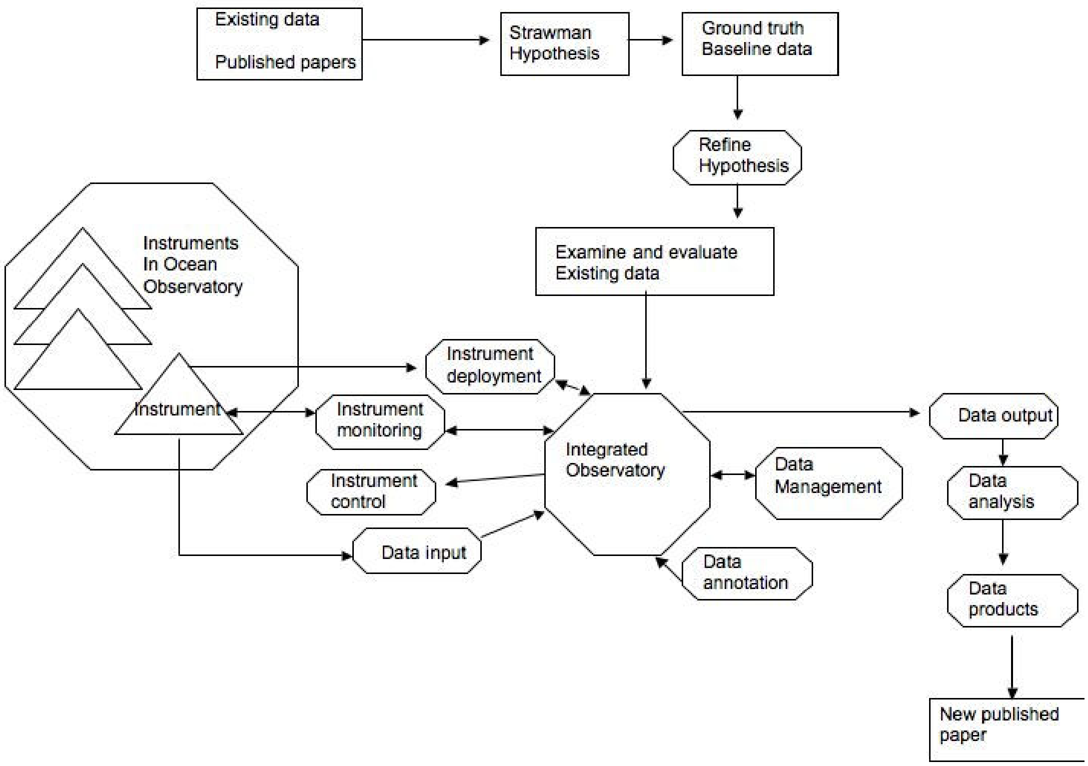

Figure 9: User and data interaction workflow for creating data products to creating publishable paper

4.9 Scenario 4: Virtual Observatory

The scenario focused in the future integrated virtual observatory. What would it be like?
Assumption: Scientists can share the resources between different groups that don't even know each other. There need to be clear issues of policy management

Keywords:

- Facebook style
- Tagging
- Emerging groups
- Dr. Chu
- Virtual research terms
- Shared resources
- Group ownership
- Policy, access control

WHAT Situation: The future Observatory is used by different groups in different labs coming together to create a new, a virtual lab.

WHO: User, principal investigator, chief scientist

All users need to see what is happening in the observatory in real time

- Own taxonomy/languages
- Rated priorities
- References

Process for purchase of instrument- PUCK

1. Receive an instrument- data management system gives you ID
2. Program your unique ID
3. Load payload- instrument driver or external puck and install metadata
4. Ready to go for ops people
5. Load on ship, go out to sea
6. It is at sea, "does what it does"
7. Code is on mooring
8. Protocol: strike up communication with instrument- gives it ID
9. Every communication includes ID

Data profile example:

- Come off cruise and scan all cruise notebooks
- Data input in Excel spreadsheet-
  - All instruments
  - Dates, lat, long
  - Notes
  - File names (when clock was on/off, when fixed clock,)
- Distribute to everyone on team
- Access network location

Anticipated future data products and data needs

- Ancillary data collection, standard presentation
- Establish credibility
- Tools to work with data
- Quality metrics for data
- Systems to formally capture data, preserve, report, link, evolve ancillary data
- Standards and mechanisms to record researchers knowledge
  - Data
  - Metadata
  - Linkage
- Standardized notion of information
- Framework for managing deployments
- Work done in-situ
- Immediate and automatic capturing of metadata and knowledge-perhaps an electronic lab book
- Progression towards less lossy data publication
- Automatic capturing of ancillary information supporting human workflow
- Serial number information available online

Specific Statements:

- CI needs to establish a mechanism to ensure rigor of data quality and interaction inside OOI
- Not too forceful
- Somehow the system should be disruptive

- CI to support but not define
- Marine IOs work on best production ops, workflows, policy manuals
- Support data migration
- Profiles exist in communities for workflow, data entry
- Profiles need to be prioritized

4.10 Scenario 5: Future Collaborative Process

The scenario is describing what a future collaborative process will be like using the cyber-infrastructure. The group provided a list of tools needed for this scenario to work.

Tools

- Resource tracking
- Calendar
- Timeline
- On paper, spreadsheet
- Some MS project
- Asset management, inventory database management
  - Includes history of every sensor
  - Deployment calendar- data repository, time, date, etc.
- Testing, recalibration for every sensor
- All the different categories for every test that you have to fill in
  - What was temperature
  - What was sensor it was calibrated against
  - What standard was it calibrated against
- Deployment map of sensors
  - Recovery times
  - Data analysis tools
- Step through data from raw to end product

- Every sensor has a page
  - Picture
  - Information about sensor
  - Calibration information that opens to a page that has history
    - Tables- Schedule of calibration, drives O&M schedule

Specific Statements:

- User will require information for the whole lifecycle of instrument. This includes all data from commissioning, deployment, tracking and end of lifecycle.

4.11 Scenario 6: Activities and Policies

The last working session of the workshop summarizes the issues and was a summary of the previous work. This session produced three prioritized lists that include the important processes for instrumentation, data and policies for the CI.

A brainstorm period started the session where the organizers recorded lists for Instruments, Data and Policies. Once the list was finalized, the participants were allowed to vote for the two most important issues on the Instrument and Data list. The group worked together to prioritize the policy list. The following is a picture of the white board list.

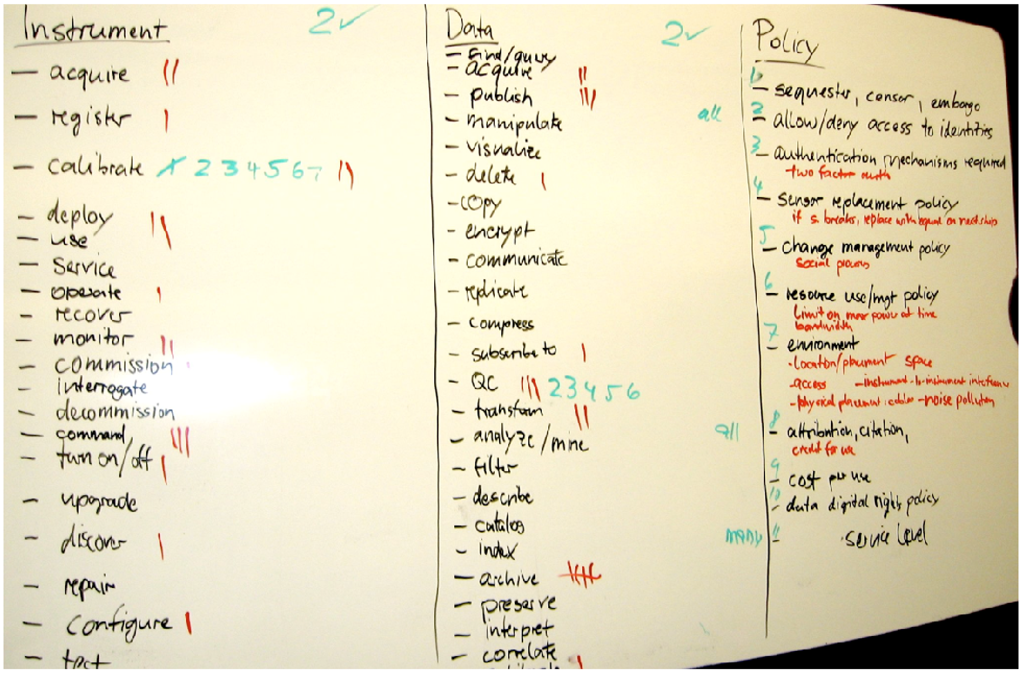

Figure 10: Group Voting on Important Activities

The following lists the issues on the Instrument list, sorting by the results of group voting.

Instruments	Data
Command (3)	Archive (5)
Acquire (2)	Publish (3)
Commission (2)	QC (3)
Calibrate -2 Allow/deny, 3 Authenticate, 4, Sensor replacement, 5 Change management (2)	Acquire (2)
Monitor (2)	Transform (2)
Deploy (2)	Delete (1)
Restore (1)	Subscribe to (1)
Operate (1)	Calibrate (1)
Turn on/off (1)	Delete
Discover (1)	Manipulate
Use	Communicate
Configure (1)	Copy
Service	Replicate
Recover	Encrypt
Interrogate	Compress
Decommission	Send, receive
Upgrade	Analyze
Repair	Filter
Test	Describe
	Characterize
	Index
	Preserve
	Interpret
	Correlate
	Find/navigate/query

Data Policy

1. Sequester, censor, embargo
2. Allow/deny access, authorization-applies to everything
3. Authentication mechanisms allowed (two factors)
4. Sensor replacement policy
    a. Replacement of sensor-has it misbehaved?
5. Change management policy
6. Resource use/management policy
    a. Limit on use: An event has happened and everyone wants to use the resource at the same time
    b. Bandwidth
    c. Frequency
7. Environment
    a. Space: Sensor locations, who determines the placement of the sensors
    b. Bandwidth
    c. Frequency
    d. Physical placement-cable management
    e. Access
    f. Instrument to instrument interference
    g. Noise in the water- probably a marine mammal issue
8. Attribution, citation- applies to everything
    a. Credit for use
9. Cost per use
10. Data digital use policy

5 Science User Requirements

5.1 Requirements Elicitation Process

The requirements listed in the next section represent the current collection of science user requirements for the OOI CI. Some of the requirements were identified in prior requirements workshops and partially validated by the participants. Further requirements originate from the analysis of related cyber-infrastructure efforts. The remaining requirements were identified through a thorough post-workshop analysis process. Requirements were either directly stated by the participants during the workshop discussions, called out in the participant questionnaires or inferred through a requirements analysis process by the CI architecture and design team. Requirements are grouped into categories and formatted according to a template as described below.

In order to uniquely identify the elicited requirements, each requirement in this report follows a standard template. Each requirement contains a unique identifier issued by the DOORS requirements management system. Furthermore, each requirement contains a label and an explanation. Requirement labels are constructed in a schematic way.

The listed requirements strive to be atomic (i.e., they express one idea only and do not contain sub-requirements). However, requirements might be related and one requirement might be influenced by another requirement. Further, the explanation might contain additional details about the requirement.

5.2 OOI Cyber User Requirements

This section contains a list of science user requirements as exported from the OOI cyber user requirements DOORS module on 7/31/08. It shows the identifiers and requirements labels and omits explanations and further attributes, such as priority. Please refer to [OOI-CU-REQ] for a full generated view containing all attributes. Requirements are grouped into categories, as indicated by the bold labels in the table. The numbering reflects the structure of the DOORS module. The requirements list contains all CI user requirements to date. Requirements that are traceable to the Data Product Generation requirements workshop are marked in italics.

ID	Requirement / Category Heading
	4.1 Resource Management
L2-CU-RQ-50	The CI shall support distributed resources, applications and actors
L2-CU-RQ-51	The CI shall provide the capability for a given resource to initiate change in another resource
L2-CU-RQ-52	All resources under CI governance shall be identifiable
L2-CU-RQ-53	All resources under CI governance shall be authenticatible
L2-CU-RQ-54	All resources under CI governance shall be authorizable
L2-CU-RQ-55	All resources under CI governance shall be auditable
L2-CU-RQ-56	The CI shall incorporate a policy-based decision system for the management of CI-governed resources
L2-CU-RQ-57	The CI shall ensure that resource utilization is governed by the rights and allocations of the initiating actor
L2-CU-RQ-58	The CI shall enable non-persistent connection of resources, users and applications
L2-CU-RQ-59	*The CI shall act as the facilitator and broker for resource usage*
L2-CU-RQ-60	The CI shall schedule resource usage based on capacity, capability and availability
L2-CU-RQ-61	The CI shall support the evolution of resources under CI governance
L2-CU-RQ-62	*The CI shall support the resource life cycle, providing notification to resource providers and*

ID	Requirement / Category Heading
	consumers when manual intervention is required
L2-CU-RQ-63	The CI shall provide a catalog listing all resources under CI governance
L2-CU-RQ-64	The CI catalog shall provide status information for all resources under CI governance
L2-CU-RQ-65	All resources under CI governance shall be discoverable, either directly, by content or through their associated metadata
L2-CU-RQ-66	Multiple actors shall be able to simultaneously discover the same resource
L2-CU-RQ-67	The CI shall integrate resource discovery with resource access subject to policy
L2-CU-RQ-68	The resource catalog shall link entries to the associated metadata
L2-CU-RQ-69	The resource catalog shall incorporate information about physical samples
L2-CU-RQ-70	The CI shall cross-reference CI-governed resource catalogs and external resource catalogs
L2-CU-RQ-71	The CI shall enable discovery of all information resources that are derived from a given original information resource
L2-CU-RQ-72	The CI shall provide resource subscribers automatic and manual fallback options with similar characteristics in case the original resource becomes unavailable
L2-CU-RQ-73	The CI shall provide services to group resources
L2-CU-RQ-74	The CI shall provide registration services for resource notification
L2-CU-RQ-75	The CI shall automatically register resources for notification to the observatory operator
L2-CU-RQ-76	*The CI shall provide notification of resource state change to all resource subscribers*
L2-CU-RQ-77	The CI shall bind metadata to all resources under CI governance throughout the resource life cycle
L2-CU-RQ-78	The CI shall support standard OOI-standard metadata content that includes, but is not limited to, a complete description of resource behavior, content, syntax, semantics, provenance, quality, context, citation, correspondence and lineage
L2-CU-RQ-79	The CI shall specify and utilize a standard vocabulary
L2-CU-RQ-80	The CI shall maintain the relationship between OOI standard metadata and the vocabulary
L2-CU-RQ-81	The CI shall allow resource discovery utilizing the standard vocabulary
L2-CU-RQ-82	*The standard vocabulary shall accommodate information on physical samples*
L2-CU-RQ-83	The CI shall provide data generating resources using proprietary metadata formats with a means to transform them to OOI standard metadata
L2-CU-RQ-84	*The CI shall support the provisioning of OOI standard metadata*
L2-CU-RQ-85	The CI shall verify compliance of metadata with the OOI standard
L2-CU-RQ-86	The CI shall update resource metadata within 5 seconds of resource reconfiguration
L2-CU-RQ-87	*The CI shall provide services for control and monitoring of observatory infrastructure resources*
L2-CU-RQ-88	*The CI shall provide services for pervasive resource monitoring and control*
	4.2 Data Management
L2-CU-RQ-90	The CI shall be capable of archiving all data and data products associated with an OOI observatory
L2-CU-RQ-91	The CI shall act as a broker for CI-managed data products
L2-CU-RQ-92	The CI shall ingest data with variable delivery order
L2-CU-RQ-93	*The CI shall support the delayed distribution of temporarily sequestered data*
L2-CU-RQ-94	The CI shall ensure the integrity and completeness of all archived data products throughout the OOI life cycle
L2-CU-RQ-95	The CI shall ensure that all archived data products can be restored to their most recent state
L2-CU-RQ-96	The CI shall provide a topic-based (publish/subscribe) data distribution infrastructure
L2-CU-RQ-97	The CI shall provide registration services for data subscriptions

ID	Requirement / Category Heading
L2-CU-RQ-98	The CI shall publish unprocessed raw sensor data
L2-CU-RQ-99	*The CI shall archive unprocessed raw sensor data*
L2-CU-RQ-100	The CI shall support the publication, distribution and archiving of different versions of the same data product or stream
L2-CU-RQ-101	The CI shall support real-time data delivery
L2-CU-RQ-102	The CI shall support guaranteed data delivery
L2-CU-RQ-103	The CI shall support store until requested (pull mode) data delivery
L2-CU-RQ-104	The CI shall support streaming data delivery
L2-CU-RQ-105	The CI shall integrate multiple data streams or data sets into a single stream or set, eliminating redundant entries
L2-CU-RQ-106	The CI shall support peer-to-peer communication between discoverable resources
L2-CU-RQ-107	The CI shall support secure data delivery
L2-CU-RQ-108	The CI shall adapt data delivery in the presence of limited available bandwidth according to policy
L2-CU-RQ-109	The CI shall notify registered resource users when data delivery cannot be achieved due to low available bandwidth
L2-CU-RQ-110	The CI shall adapt data delivery in the presence of high channel latency according to policy
L2-CU-RQ-111	The CI shall notify registered resource users when data delivery cannot be achieved due to high channel latency
L2-CU-RQ-112	The CI shall publish data from external data sources, databases, and data distribution networks from related scientific domains.
L2-CU-RQ-113	*The CI shall provide support for large volumes of data*
L2-CU-RQ-114	The CI shall archive and catalog text, images, pdf, .doc files and spreadsheets
L2-CU-RQ-115	The CI shall flag and notify data stream and data set state change
L2-CU-RQ-116	The CI shall flag and notify redundant data and metadata
L2-CU-RQ-117	The CI shall acknowledge requests for data with an estimate of delivery latency
L2-CU-RQ-118	The CI shall credit data publishers when data products are accessed
L2-CU-RQ-119	The CI shall provide services and interfaces for the acquisition of bulk data
L2-CU-RQ-120	*The CI shall associate bulk data with their metadata and related data products*
	4.2.1 Data Transformation
L2-CU-RQ-122	The CI shall support the moderation and auditing of published data
L2-CU-RQ-123	*The CI shall provide services for interactive and automated data quality control (QC)*
L2-CU-RQ-124	The CI shall perform automated quality control of observational data products in near real-time
L2-CU-RQ-125	The CI shall provide standard and user-defined methods to assess the quality of data
L2-CU-RQ-126	*The CI shall specify data models for resources based on characterization of structure (syntax)*
L2-CU-RQ-127	*The CI shall translate between standard syntactic data models without loss of information*
L2-CU-RQ-128	*The CI shall support translation between user-specified syntactic data models*
L2-CU-RQ-129	*The CI shall specify data models for resources based on characterization of meaning (semantics)*
L2-CU-RQ-130	*The CI shall support mapping between senders and receivers using the standard vocabulary without loss of information*
L2-CU-RQ-131	*The CI shall provide capabilities to define event detectors*
L2-CU-RQ-132	The CI shall provide event detection services
L2-CU-RQ-133	The CI shall provide registration services for event notification

ID	Requirement / Category Heading
L2-CU-RQ-134	The CI shall provide notification of detected events
L2-CU-RQ-135	The CI shall provide versioning for detected events
L2-CU-RQ-136	The CI shall update data sets as sensor calibrations become available
L2-CU-RQ-137	*The CI shall be able to accumulate knowledge about the scientific interpretation of observational data from manual mapping and linking of variables between different data sets*
L2-CU-RQ-138	*The CI shall be capable of co-registering data from different instruments in space and time*
	4.3 Research and Analysis
L2-CU-RQ-140	The CI shall suggest suitable data products, observation resources, analysis tools, visualization tools and other OOI resources based on user-specified research questions using the standard vocabulary
L2-CU-RQ-141	The CI shall support interactive data analysis and visualization through tools and user interfaces
L2-CU-RQ-142	The CI shall provide a standard, extensible set of data processing elements that provide data assimilation, alignment, consolidation, aggregation, transformation, filtering, subsetting, averaging and scaling
L2-CU-RQ-143	The CI shall provide capabilities for analysis and presentation of environmental data at specified sites
L2-CU-RQ-144	*The CI shall support the integration of external analysis tools*
L2-CU-RQ-145	*The CI shall provide capabilities to transform between coordinate systems*
L2-CU-RQ-146	*The CI shall provide capabilities to transform between map projections*
	4.4 Ocean Modeling
L2-CU-RQ-148	The CI shall enable the efficient configuration, execution, and debugging of numerical ocean models
L2-CU-RQ-149	The CI shall support the interaction of model developers and non-expert model users
L2-CU-RQ-150	The CI shall provide capabilities to tune numerical models
L2-CU-RQ-151	The CI shall provide a virtual model environment and simulator to determine optimal model inputs, parameterizations and outcome qualities
L2-CU-RQ-152	The CI shall enable the sharing of ocean modeling, data assimilation and visualization components, including the extension of models with new model components
L2-CU-RQ-153	The CI shall provide a repository and sharing capabilities for numerical model algorithms, model configurations, data processing tools and documentation
L2-CU-RQ-154	The CI shall archive numerical model workflows under configuration control
L2-CU-RQ-155	The CI shall recompute model data products using archived workflows
L2-CU-RQ-156	The CI shall enable the modification of archived workflows
L2-CU-RQ-157	The CI shall provide an environment for the development of community numerical models under community process support
L2-CU-RQ-158	The CI shall provide a non-restricted environment for the development of independent numerical models
L2-CU-RQ-159	The CI shall support the nesting of ocean models at different geographical scales
L2-CU-RQ-160	The CI shall provide a framework for the adaptation of model resolution to the available resources
L2-CU-RQ-161	The CI shall support model ensemble definition, execution and analysis
L2-CU-RQ-162	The CI shall publish both elements of and aggregated ensemble data products from ocean models
L2-CU-RQ-163	The CI shall support flexible high performance model execution
	4.5 Visualization
L2-CU-RQ-165	The CI shall provide interactive 2D, 3D and 4D visualization tools

ID	Requirement / Category Heading
L2-CU-RQ-166	*The CI shall provide 3D visualization of sensor locations and their environment*
L2-CU-RQ-167	The CI shall support the integration of external visualization tools
L2-CU-RQ-168	The CI shall provide extensible, configurable visualization capabilities for data streams
L2-CU-RQ-169	The CI shall provide a zooming interface for all visualizations with at least three levels of detail
L2-CU-RQ-170	The CI shall provide a user interface system that includes at least two different views of the data
	4.6 Computation and Process Execution
L2-CU-RQ-172	The CI shall support the execution of large scale numerical ocean models across different locations on the network
L2-CU-RQ-173	The CI shall support workflows for automated numerical model execution, including just-in-time input data preparation, model computation, output post-processing, and publication of results
L2-CU-RQ-174	The CI shall enable the one-time and recurring execution of numerical models on any net-worked computational resource with quality-of-service guarantees based on contracts and policy.
L2-CU-RQ-175	*The CI shall provide interfaces to compose workflows*
L2-CU-RQ-176	*The CI shall provide services to execute workflows on computational resources with varying characteristics*
L2-CU-RQ-177	The CI shall provide services to chain a plurality of workflows
L2-CU-RQ-178	The CI shall provide services to monitor and control instantiated processes
L2-CU-RQ-179	The CI shall provide actors with estimated performance/turnaround for instantiated proc-esses
L2-CU-RQ-180	The CI shall provide event-triggered workflow execution services
L2-CU-RQ-181	The CI shall provide real-time access to high performance computation resources
L2-CU-RQ-182	The CI shall provide process support for the planning and operation of observational pro-grams
L2-CU-RQ-183	The CI shall provide process support for the coordination of instrument recovery, mainte-nance and replacement
L2-CU-RQ-184	The CI shall support, automate and combine workflows of shipboard observers
	4.7 Sensors and Instrument Interfaces
L2-CU-RQ-186	The CI shall provide a real-time communication interface for remote resources
L2-CU-RQ-187	The CI shall support robust instrument development, operation and maintenance processes
L2-CU-RQ-188	The CI shall support discovery of the characteristics of sensors deployed on an instrument platform
L2-CU-RQ-189	The CI shall support adaptive observation resource control
L2-CU-RQ-190	The CI time standard shall be NIST traceable
L2-CU-RQ-191	The CI shall provide a synoptic time service with an accuracy of 1 microsecond to all re-sources connected to the OOI observatories
L2-CU-RQ-192	The CI shall serve synoptic time throughout the observatory using Network Time Protocol
L2-CU-RQ-193	The CI shall provide services to correct remote clocks to a synoptic standard
L2-CU-RQ-194	The CI shall provide services to synchronize remote clocks relative to each other with an accuracy of 1 microsecond
L2-CU-RQ-195	Upon receipt, the CI shall synoptically timestamp message headers with an accuracy of 1 millisecond
L2-CU-RQ-196	*The CI shall provide robust instrument access protocols*
L2-CU-RQ-197	*The CI shall provide direct bidirectional communications to resources that preserves their*

ID	Requirement / Category Heading
	native functionality
L2-CU-RQ-198	*The CI shall provide remote desktop access to resources that preserves their native functionality*
L2-CU-RQ-199	*The CI shall automatically close down inactive direct access sessions*
L2-CU-RQ-200	The CI shall provide interactive web-based configuration of instrument platforms, instruments and sensors
L2-CU-RQ-201	The CI shall provide capabilities and interfaces for monitoring of resource-specific operational and environmental parameters
L2-CU-RQ-202	The CI shall provide services for positioning of mobile assets with a precision commensurate with the location technology
L2-CU-RQ-203	The CI shall support automated docking of mobile resources, including power management and high speed data down and up load
L2-CU-RQ-204	The CI shall be capable of triggering instrument measurements
	4.8 Mission Planning and Control
L2-CU-RQ-206	The CI shall support swarm-based deployment patterns for mobile instruments
L2-CU-RQ-207	The CI shall provide a repository for instrument behaviors
L2-CU-RQ-208	The CI shall provide a repository for observation plans
L2-CU-RQ-209	The CI shall provide shore-side and on-vehicle control capabilities for autonomous observational resources
L2-CU-RQ-210	The CI shall support observational resource control at different user-selected levels
L2-CU-RQ-211	The CI shall integrate environment and vehicle behavior models for event detection, coordinated control and adaptive sampling
L2-CU-RQ-212	The CI shall provide capabilities and interfaces for planning longitudinal observations
L2-CU-RQ-213	The CI shall provide capabilities and interfaces for planning objective-driven observations
L2-CU-RQ-214	The CI shall provide capabilities and interfaces for ad-hoc interactive and automated modification of ongoing observations
L2-CU-RQ-215	The CI shall provide capabilities and interfaces for simulating and verifying observation plans
L2-CU-RQ-216	The CI shall provide resource provisioning calculations from observation plans
L2-CU-RQ-217	The CI shall support observation planning and scheduling decisions based on the opportunity cost of observations and resource provisioning
L2-CU-RQ-218	The CI shall provide graphical user interfaces for planning observations and missions with spatial and temporal visualization of observation parameters
L2-CU-RQ-219	The CI shall provide spatial visualization of observation data overlaid with observation plans
L2-CU-RQ-220	The CI shall support tasking, deployment, mission control and retrieval of mobile and fixed instruments
L2-CU-RQ-221	The CI shall provide capabilities and interfaces for the simulation of observational infrastructure
	4.9 Application Integration and External Interfaces
L2-CU-RQ-223	The CI shall provide documented resource-data connectors for all services
L2-CU-RQ-224	Conditional on OOI policy, the CI shall not impose specific processes, tools and formats on resource providers for the operation and control of their OOI-connected resources
L2-CU-RQ-225	The CI shall interface with external resource monitoring, operation and control systems
L2-CU-RQ-226	*The CI shall provide a Web 2.0 environment*
L2-CU-RQ-227	*The CI shall support interfacing with web service-accessible resources*
L2-CU-RQ-228	*The CI shall interface to live video feeds during instrument operation and maintenance*
L2-CU-RQ-229	*The CI shall provide interface support for Java-based tools and scripting languages*

ID	Requirement / Category Heading
L2-CU-RQ-230	*The CI shall provide standalone installations that may have no or intermittent connection to the OOI network*
	4.10 Presentation and User Interfaces
L2-CU-RQ-232	The CI shall provide annotation, commenting, ranking and rating services for CI-managed resources
L2-CU-RQ-233	*The CI shall provide user and group workspace capabilities*
L2-CU-RQ-234	*The CI shall provide capabilities to personalize user and group workspaces*
L2-CU-RQ-235	*The CI shall provide social networking capabilities*
L2-CU-RQ-236	The CI shall provide an intuitive interface to access the functionality of all CI services and resources
L2-CU-RQ-237	The CI shall present the full CI functionality at a single access point with a single dashboard
L2-CU-RQ-238	The CI shall provide services to make OOI-standard metadata human readable
L2-CU-RQ-239	The CI shall provide a resource monitoring and control interface
L2-CU-RQ-240	*The CI shall provide an adaptive, simple-to-use interface for data access*
L2-CU-RQ-241	*The CI shall provide transparent access to heterogeneous, large-scale computational resources*
L2-CU-RQ-242	*The CI shall provide transparent access to heterogeneous, large-scale storage resources*
L2-CU-RQ-243	The CI shall provide a single user interface that supports observatory operators, science and engineering users, the education community and the general public
L2-CU-RQ-244	The CI shall provide dialog box interaction for operations requiring the input of more than two parameters
L2-CU-RQ-245	The CI shall provide input screens that include tabs for any process that requires users to input more than five parameters
L2-CU-RQ-246	The CI shall provide a common font set for all screens
L2-CU-RQ-247	The CI shall employ a common look and feel based on a standard screen design
L2-CU-RQ-248	The CI shall employ a standard set of colors for use in all user interface presentation screens
L2-CU-RQ-249	The CI shall employ a standard workflow for all user interface screens
L2-CU-RQ-250	The CI shall employ a common navigation scheme that is consistent from application to application
L2-CU-RQ-251	*The CI shall provide visualization and metadata browsing of the processing pipeline*
L2-CU-RQ-252	*The CI shall provide checklists for standard instrument operations*
L2-CU-RQ-253	*The CI shall provide capabilities and interfaces to capture structured input, feedback and results from analysis processes on data*
	4.11 Security, Safety and Privacy Properties
L2-CU-RQ-255	*The CI shall authenticate and authorize all resources connected to an OOI observatory*
L2-CU-RQ-256	*The CI shall authenticate all observatory actors*
L2-CU-RQ-257	*The CI shall provide different levels of access to actors with different levels of authorization*
L2-CU-RQ-258	The CI shall enforce user privacy policies
L2-CU-RQ-259	*The CI shall be capable of auditing all services and resources under CI governance*
L2-CU-RQ-260	The CI shall trace resource utilization to the initiating actor
L2-CU-RQ-261	The CI shall support different levels of access for resources and their metadata
L2-CU-RQ-262	The CI shall protect physical resources from damage and misuse by enforcing resource use policies
L2-CU-RQ-263	The CI shall provide interfaces to define security and policy for information managers at participating institutions
L2-CU-RQ-264	The CI shall support the diversion, filtering and sequestering of raw data streams at the ac-

ID	Requirement / Category Heading
	quisition point
	4.12 Quality Properties
L2-CU-RQ-266	The CI infrastructure shall deliver messages with reliability that is comparable to that of the Internet
L2-CU-RQ-267	The CI shall provide robust, reliable remotely deployed components
L2-CU-RQ-268	The CI shall provide services with reliability and accuracy that is comparable to those of distributed Internet applications
	4.13 Education and Outreach
L2-CU-RQ-270	The CI shall provide numerical ocean models with a limited set of modifiable parameters for educational purposes
L2-CU-RQ-271	The CI access point shall provide educators with instructions about data usage
L2-CU-RQ-272	The CI access point shall provide the educator with a list of projects and their attributes
L2-CU-RQ-273	The CI access point shall provide the educator with a means for social networking.
L2-CU-RQ-274	The CI shall provide a discoverable repository for educator-provided tools
L2-CU-RQ-275	The CI shall provide versioning and citation for educator assets
	4.14 Documentation
L2-CU-RQ-277	*The CI IO shall make all source code for the OOI Cyberinfrastructure implementation and drivers publicly available, subject to applicable licenses*
L2-CU-RQ-278	*The CI IO shall document all external interfaces*
L2-CU-RQ-279	The CI IO shall document all device drivers
L2-CU-RQ-280	The CI shall provide discoverable web-based documentation for all services
L2-CU-RQ-281	The CI shall utilize a naming scheme that is compliant with OOI naming conventions
L2-CU-RQ-282	**4.15 Development Process**
L2-CU-RQ-283	The CI IO shall seek to influence the direction of CI standards to effectively meet the needs of OOI users
L2-CU-RQ-284	The CI shall utilize open standards and open source software to the maximum possible extent
L2-CU-RQ-285	The CI IO shall accommodate local innovation that can be scaled to the community level
L2-CU-RQ-286	The CI IO shall support the verification of hardware and software components that will be deployed on OOI infrastructure
L2-CU-RQ-287	The CI shall support modular components
L2-CU-RQ-288	The CI implementation shall be platform-independent
L2-CU-RQ-289	CI service interfaces and capabilities shall maintain backward compatibility as the services evolve
L2-CU-RQ-290	The CI architecture shall be scalable to accommodate an increasing range of actors, resources, and services
L2-CU-RQ-291	The CI shall be extensible to allow the addition of new resources, services and applications to the OOI infrastructure
	4.15.1 Other
L2-CU-RQ-293	*The CI shall provide process support for "dry" observational infrastructure development, verification and simulation*
L2-CU-RQ-294	The CI IO shall provide technically-qualified user care support and assistance through a human actor
L2-CU-RQ-295	The CI shall provide capabilities to maintain contact between users and user care
L2-CU-RQ-296	The CI shall provide capabilities to initiate and track trouble tickets
L2-CU-RQ-297	*The CI shall provide tools for observatory operators to communicate with users*

6 Workshop Conclusions

6.1 Feedback from the Participants

- Most useful were sessions where scientists described their interactions
- Good exercise would be to play with a prototype and discuss about interfaces, data exchange etc.
- Take away: there is hope, improved coherence, barriers dropped
- Best session was the use case of the process of deploying instruments
- Next step produce a template workflow
- Take away ideas and hope
- Best was the definition of the individual process
- Learning about the different science tools, acronyms
- Somehow encouraged about discussions progressing
- Struggled with specific contribution
- Not fully clear where to end up at the end of the workshop
- Outcomes very valuable
- Very small piece of the picture. Don't fully understand how to get to the goal from here
- Today a lot of detailed descriptions of Antelope were useful.
- Discussion about virtual lab. Trouble of thinking this is on the critical path of the project
- Educated about what the CI effort is about
- This size group is workable
- Appreciated the interest of CI on CGSN challenges.
- Growing appreciation of future support capabilities of the CI
- Real pilot in water and in programs will mature interrelationships.
- Excited
- Thanks for CI vision into the future
- Given a wish list and dirty laundry of community
- Some concern about if everything can be realized.
- Great group to work with for next 20 months
- Can learn a lot from CI processes.
- Stay focused on getting it done and not only the wish list
- Define clearly the CI interface, which is still blurry
- A lot of raw material created
- Broad coverage
- Would like to see a clear bulleted list of good requirements
- First time he has experienced in projects for information gathering. Typically done on the fly.
- Happy to get requirements to work on designing the capabilities.
- Plan towards ordering the deliverables
- In different MRE project without requirements elicitation efforts created big problems
- Inefficiencies built into the system
- Provided the opportunity to get things right
- Very dynamic days
- Good workshop process
- Interplay inside CI that was not really relevant
- At times scientists were quiet, but at other very valuable input
- Arguments provided to justify the program and its transformative nature
- Excellent program

- Base investment occurring here. The better the program can be scoped, the better an accurate baseline for the program can be set in terms of scope, cost, etc to manage external expectations.\
- Particularly liked last exercise:
- Focused on creating a list
- Voting is a create way to achieve something
- This way we get towards an equal participation of people; everybody is equally represented
- The variety of tools available in the CI was interesting to learn about, what's available and how the dependencies are.
- Getting net information, confirmation of what we really need
- Need different interfaces for different stakeholders
- Different policies
- Important to get the meta-data right
- Results will be solicited
- Feedback and prioritization are important
- Expect direct participation in the authoring of use cases and requirements from people of the CI team.
- Thank participants
- Thanks CI team
- Learned a lot about the subsystems.
- Good to walk through scenarios.

6.2 Next Steps and Action Items

Next steps include

- Consolidate requirements from all user requirements workshop into a consistent list of CI user requirements.
- Prioritize and rank all user requirements leading to a selection of baseline requirements for the construction of the OOI to be reviewed during FDR.
- The user community will be asked to validate the requirements
- Requirements validation and prioritization will continue past FDR

6.3 Conclusions from the Organizers

The fourth OOI CyberInfrastructure Requirements Workshop, hosted by the University of California, San Diego was a stimulating meeting and a success in providing material for further CI requirements definition. Furthermore, the workshop helped refine and complement the CI architecture and design. This workshop helped to foster the mutual understanding of prospective CI user communities and the CI design team

Direct outcomes such as use cases, identifying and validating requirements, and jointly developed domain models will be valuable assets in the subsequent CI design efforts. Further results include validations of requirements previously collected for the Conceptual Architecture, and initial outreach measures to CI user communities.

Appendices

A Workshop Participant Questionnaire

The CI ADT refined the OOI CI requirements questionnaire from the previous workshop (see [CI-RWS2]) with specific adaptations for the data product generation topics. The invited scientists were asked to provide answers to these questions prior to the workshop.

Intent of this template

- This slide set is a template for participants at the OOI CI requirements workshop
  - For presentations during the workshop
  - To capture relevant information in a structured way
- Goals of this exercise are
  - To capture as many CI-relevant details as possible before the workshop
  - To capture structured, relevant information for use during and after the workshop
  - To enable quick information access for domain modeling during the workshop
  - To provide you some ideas about the expected outcome and materials covered during the workshop from the perspective of the CI design team
- We ask you to please fill it out to the degree possible/applicable. Please try to provide answers to as many (relevant) questions as you can
- You can use this template as you like. You can modify it, take only parts of it, add your own slides, copy/paste from it, use it to structure your own text/spreadsheed/slideset documents …

General Goals for the Requirements Analysis

- Analyze the current situation
  - Definition of basic terms: instrument, platform, data, etc.
  - Tools, technologies, processes, data used and/or available
  - Organizational details (e.g. responsibilities, roles in team, workflows, policies)
  - Current shortcomings for whatever reason
- Determine short-term improvements
  - What would make every-day observation tasks easier and more effective?
  - What shortcomings should be eliminated most urgently?
- Identify CI transformative vision and requirements
  - Assuming there is a transformative community CI in place, what are the expectations for an "ideal CI"?
  - Capabilities, interfaces, necessary guarantees, resources provided, etc.
- Scope
  - As relevant to the OOI CyberInfrastructure
  - From the viewpoint of your community

Current situation and Expected changes

- What capabilities and properties do you require from a cyber-infrastructure that supports your current work? Please rank.
- What capabilities and properties do you expect from a transformative cyber-infrastructure in the oceanographic domain that would benefit you and the community in the next decade? Please rank.

- What works particularly well in your domain? Exemplary standards, tools, platforms, portals, technologies, etc?
- Please list the biggest impediments that currently exist for your work and/or the community. Please rank and explain.

Instruments and Instrument Platforms

- What instruments/platforms do you currently use and/or develop? Please explain some important specifics of these and any related tools.
- Please describe a typical everyday scenario acquiring data using your instruments/platforms. Example pictures, configurations, documentation etc. are always helpful. Please attach, if available.
- How do you manage the instrument life cycle?

Data

- What sequence of data processing, format transformations, event detection, quality control, other processing, and data archiving do you apply?
- What data formats and metadata formats do you use?
- How do you perform data quality control?
- How do you perform identification, characterization, and event detection on data (current data, historic data)?
- How do you handle incoming data delays and temporary data gaps?
- How do you store, archive, visualize, and publish your data product results?

Interfaces

- What application interfaces, user interfaces, and visualization support do you envision and/or require of an effective and easy to use community cyber-infrastructure?
- What instrument interfaces (both sensor and actuator) do you envision and/or require of an effective and easy to use community cyber-infrastructure?

Security, Privacy, Policy

- Please explain the relevant security and policy guarantees that you and/or your organization require. This includes authentication mechanisms, authorization (access control) and resource access policy strategies, privacy needs, intellectual property issues, etc.
- How do you manage access to instruments and data products? This includes preventing unauthorized access and preventing damage, over-use and abuse through usage policies.

Operation and Maintenance

- How do operation and maintenance requirements affect the design of your instruments/platforms and your daily work?
- What importance does this topic have in your overall work?
- How do you manage changes to instruments/platforms, data sources, technology platforms, standards etc.?

Education and Outreach

- How do education and outreach concerns affect your observation programs and the presentation of the results?
- How do you make data available for education and outreach purposes?
- What would make these tasks easier?

Comments, Expectations, Suggestions

- What do you expect from the upcoming OOI CI requirements workshop?
- What topics do you think are relevant and should not be missed by the organizers?

Additional reading materials, References

- Are there any similar projects/communities that you like and/or that are exemplary technology-wise?
- Are there standards, other national or international efforts that the OOI design team should consider/evaluate?
- Anything you think is relevant that you want to add to this questionnaire?
- Further reading materials
- References

B List of Previous User Requirements

Cat.	Req-ID	Requirement
Resource Management		
	RWS2-R1	The CI shall notify registered users and applications when new resources are added to the system.
	RWS1-R3	The CI shall be extensible to allow the addition of new resources and applications to the OOI infrastructure.
	RWS1-R9	The CI shall provide a catalog listing all resources under CI governance.
	RWS1-R9A	The CI shall enable users to discover observatory resources together with their meta-data based on resource characteristics and user-defined search criteria.
	RWS1-R11	The CI shall catalog physical samples in the CI resource catalog.
	RWS1-R12	The CI shall support cross-referencing from CI governed resources to external re-source catalogs and metadata.
	RWS1-R16	The CI shall bind metadata to all resources under CI governance throughout the re-source life cycle.
	RWS1-R18	The CI shall provide standard OOI metadata descriptions that include, but are not limited to, a complete description of resource behavior, content, syntax, semantics, provenance, quality, context and lineage.
	RWS1-R19	The CI shall allow the discovery of all information resources that are based on a given original information resource.
	RWS1-R20	The CI shall provide information resource subscribers automatic and manual fallback options with similar characteristics in case the original resource becomes unavailable.
	RWS1-R26	The CI shall provide notification of resource state change to all resource subscribers.
	RWS1-R33	The CI shall collect and provide resource access statistics.
Data Management		
	RWS1-R21	The CI shall be capable of archiving all data and data products associated with an OOI observatory or other CI-governed information resource.
	RWS1-R22	The CI shall support the publication, distribution and archiving of different versions of the same data product.
	RWS1-R23	The CI shall ensure the integrity and completeness of all data products throughout the OOI life cycle.
	RWS1-R24	The CI shall ensure that all archived data products can be restored in their complete and most recent state.
	RWS1-R30	The CI shall publish new data products resulting from processing of existing data products.
	RWS1-R31	The CI shall enable users and applications to subscribe to information resources in the form of data streams.
	RWS1-R47	The CI shall provide a topic-based (publish-subscribe) data distribution infrastructure that supports real-time and near real-time delivery, guaranteed delivery, buffering and data streaming subject to resource availability.
Science Data Management		
	RWS2-R2	The CI shall interface with, ingest and distribute data from external data sources, databases, and data distribution networks of related scientific domains.
	RWS2-R3	The CI shall provide interactive and automated data quality control (QC) tools.
	RWS2-R4	The CI shall provide standard and user-defined methods to assess the quality of data.

Cat.	Req-ID	Requirement
	RWS2-R5	The CI shall facilitate the moderation and auditing of published data.
	RWS2-R6	The CI shall act as a broker for CI-managed data products.
	RWS2-R7	The CI shall provide access to CI-manage data products in standard formats and subsets.
	RWS2-R8	The CI shall act as a broker between information and processing resources.
	RWS2-R9	The CI shall make unprocessed raw sensor data available on request.
	RWS2-R10	The CI shall track data provenance and correspondence.
	RWS2-R11	The CI shall credit data publishers when data products are accessed.
	RWS2-R12	The CI shall create and distribute related data products from a given source data product that have different characteristics, such as resolution, level of detail, real-timeform and quality,.
	RWS2-R13	The CI shall flag data stream state change.
	RWS2-R14	The CI shall support the provision of complete metadata by users.
	RWS1-R4	The CI shall support a standard set of data exchange formats.
	RWS1-R4a	The CI shall translate between the standard data exchange formats without loss of information.
	RWS1-R5	The CI shall allow the addition of user-defined data exchange formats and translators.
Research and Analysis		
	RWS2-R15	The CI shall provide capabilities and user/application interfaces for researching scientific materials and OOI-governed resources across disciplines.
	RWS2-R16	The CI shall suggest suitable data products, data transformations, observation resources, analysis tools, visualization tools and other OOI resources based on user-specified research questions in domain language.
	RWS2-R17	The CI shall support interactive and iterative analysis and visualization through infrastructure, tools and user interfaces.
	RWS2-R18	The CI shall provide tools, user interfaces and visualization for the analysis, combination and comparison of disparate, heterogeneous data sets..
	RWS1-R25	The CI shall provide a standard, extensible set of data product processing elements that provide data assimilation, alignment, consolidation, aggregation, transformation, filtering and quality control tasks.
Ocean Modeling		
	RWS2-R19	The CI shall enable the efficient configuration, execution, debugging and tuning of numerical ocean models.
	RWS2-R20	The CI shall support the interaction of model developers and non-expert model users.
	RWS2-R21	The CI shall provide facilities to develop and tune numerical models and their parameters.
	RWS2-R22	The CI shall provide a virtual model environment and simulator to determine optimal model inputs, parameterizations and outcome qualities.
	RWS2-R23	The CI shall enable the sharing of ocean modeling, data assimilation and visualization components, including the extension of models with new model components.
	RWS2-R24	The CI shall provide a repository and sharing capabilities for numerical model algorithms, model configurations, data processing tools and documentation.
	RWS2-R25	The CI shall archive numerical models under configuration control.
	RWS2-R26	The CI shall recompute model data products using archived models and workflows.
	RWS2-R27	The CI shall enable the modification of archived numerical models and workflows.
	RWS2-R28	The CI shall provide an environment for the development of community numerical models under community process support.

Cat.	Req-ID	Requirement
	RWS2-R29	The CI shall provide a non-restricted environment for the development of independent numerical models.
	RWS2-R30	The CI shall support nesting of ocean models at different geographical scales.
	RWS2-R31	The CI shall provide a framework for the adaptation of model resolution to the available resources.
	RWS2-R32	The CI shall support model ensemble definition, execution and analysis.
	RWS2-R33	The CI shall publish both elements and aggregated ensemble data products.
	RWS2-R34	The CI shall support flexible high performance model execution.
Visualization		
	RWS2-R35	The CI shall provide a uniform and consistent for numerical model output visualization and analysis in 2D, 3D and 4D.
	RWS2-R36	The CI shall provide interactive visualization of the 3D and 4D ocean.
	RWS2-R37	The CI shall support the integration of external visualization and analysis tools.
Computation and Process Execution		
	RWS2-R38	The CI shall support the execution of large-scale numerical ocean models across different locations on the network.
	RWS2-R39	The CI shall support workflows for automated numerical model execution, including just-in-time input data preparation, model computation, output post-processing, and publication of results.
	RWS2-R40	The CI shall enable the one-time and recurring execution of numerical models on any networked computational resource with quality-of-service guarantees based on contracts and policy.
	RWS1-R27	The CI shall provide uniform and easy-to-use interfaces to computational resources with varying characteristics to define executable processes.
Sensors and Instrument Interfaces		
	RWS2-R41	The CI shall provide flexible and reliable access to remote resources.
	RWS2-R42	The CI shall provide real-time monitoring of remote sensors.
	RWS2-R43	The CI shall provide continuous collection of scientific data during extreme weather events.
	RWS2-R44	The CI shall provide discovery for the number and characteristics of sensors deployed on an instrument platform.
	RWS2-R45	The CI shall support adaptive observation.
Mission Planning and Control		
	RWS2-R46	The CI shall provide capabilities and user/application interfaces for mission planning and control.
Application Integration and External Interfaces		
	RWS1-R1	The CI shall provision an integrated network comprised of distributed resources, applications and users.
	RWS1-R2	The CI shall enable non-persistent connection of resources, users and applications.
	RWS1-R6	The CI shall provide application program interfaces (APIs) to all CI services.
	RWS1-R7	The CI shall provide a synoptic time service with an accuracy of TBD to all resources connected to the OOI observatories.
Presentation and User Interfaces		
	RWS2-R47	The CI shall provide "one stop shopping" interfaces that provide and collocate relevant information regarding scientific research using OOI resources.
	RWS2-R48	The CI shall provide annotation, commenting, ranking and rating services for resources.

Cat.	Req-ID	Requirement
	RWS2-R49	The CI shall provide project and user workspace capabilities and user interfaces.
	RWS2-R50	The CI shall provide long-term and ad hoc social networking and collaboration capabilities.
	RWS1-R34	The CI shall provide homogeneous, intuitive, easy-to-use web-based interfaces to all CI services and resources.
	RWS1-R35	The CI shall provide the capability to make OOI-standard metadata human readable.
	RWS1-R38	The CI shall provide extensible configurable visualization capabilities for selected types of data streams.
	RWS1-R49	The CI shall provide real-time analysis and visualization for data resources.
Security, Safety and Privacy Properties		
	RWS2-R51	The CI shall provide interfaces to define security and policy for information managers at participating institutions.
	RWS2-R52	The CI shall provide secure operations.
	RWS2-R53	The CI shall only permit authenticated and authorized users to access OOI resources.
	RWS1-R43	The CI shall provide mechanisms to enforce user privacy policies.
	RWS1-R44	The CI shall enable any authenticated party to share their resources.
	RWS1-R44A	The CI shall grant or restrict resource access subject to use policy.
Quality Properties		
	RWS1-R46	The CI infrastructure shall provide services and deliver messages with reliability and accuracy that is comparable to that of distributed Internet applications.
Education and Outreach		
	RWS2-R54	The CI shall facilitate the creation of publicly available idealized numerical ocean models with a limited choice of modifiable parameters for educational purposes.
Documentation		
	RWS1-R41	The CI IO shall make all source code for the OOI CyberInfrastructure implementation and drivers publicly available, subject to applicable licenses.
	RWS1-R42	The CI shall provide documentation for all components of the CI, including all application program interfaces (APIs) to CI services.
	RWS1-R39	The CI IO shall provide all documentation in web-based formats.
Development Process		
	RWS2-R55	The CI IO shall circulate CI requirements and designs within and outside the OOI community so that comparable infrastructures can adopt them.
	RWS1-R8	The CI shall utilize open standards and open source software to the maximum possible extent.
	RWS1-R40	The CI IO shall provide a process for submitting and incorporating user-suggested changes to the CI.
	RWS1-R48	The CI shall provide for the flexible and transparent extension of CI services and interfaces to incorporate user-provided processes, user and application interfaces, applications and resources.

C Workshop Agenda

Day 1, May 20, 2008 - Context and current state

Time	Presenter(s)	Topics
09:30 AM	Alan Chave	Welcome from the organizers, Introductions
09:40 AM	Alan Chave	Requirements elicitation process; Overview of existing CI user requirements
10:10 AM	Benoit Pirenne	Neptune Canada cyber-infrastructure
10:55 AM	Kevin Gomes	MBARI Data Architecture
11:25 AM	John Graybeal	Marine Metadata Initiative
01:00 PM	CI ADT	Instrument data acquisition & processing; instrument lifecycle management use cases - vision for OOI.
03:15 PM	CI ADT	Data preservation & distribution use cases - past and present.
05:00 PM		Adjourn for day

Day 2, May 21, 2008 - Vision for the future

Time	Presenter(s)	Topics
08:30 AM	Mathew Arrott	Proposed CI infrastructure for the OOI
09:00 AM	Kent Lindquist	Antelope Overview
10:00 AM	Raja Rajasekar	SRB/iRODS Overview
10:15 AM	CI ADT	Instrument data acquisition & processing; instrument lifecycle management use cases - vision for OOI. Development of a transformative OOI usage scenario and requirements, covering science user, instrument provider and administrative operations, core capabilities, policy, governance and management aspects.
01:00 PM	CI ADT	Data preservation & distribution use cases - vision for OOI. Development of a transformative OOI usage scenario and requirements, covering science user, instrument provider and administrative operations, core capabilities, policy, governance and management aspects.
02:45 PM	CI ADT	Domain modeling session. Development of domain models based on the previously identified use cases, scenarios and workflows in direct collaboration of the scientists (providing input and clarification) and the CI architects (modeling of the concepts).
04:45	Alan Chave	Wrap up and Final feedback session
05:00 PM		Workshop Adjourns

D List of Participants

Name	Organization	Project Role
Matthew Arrott	UCSD/Calit2	CI Project Manager
Lorraine Brasseur	Ocean Leadership	
Alan Chave	WHOI	CI System Engineer
Emilia Farcas	UCSD/Calit2	CI System Modeler
Anthony Ferlaino	Ocean Leadership	
Kevin Gomes	MBARI	
John Graybeal	MBARI	
Mike Harrington	UW/APL	RSN System Engineer
Deborah Kelley	UW	
Igor Klacansky	UCSD/Calit2	CI System Modeler
Steve Lerner	WHOI	
Kent Lindquist	Lindquist Consulting	CI S&A Subsystem Lead
Michael Meisinger	UCSD/Calit2	CI Requirements Analyst
Matthew Moldovan	UCSD/SIO	
Benoit Pirenne	Univ of Victoria	
Arcot Rajasekar	UCSD/SDSC	
Lloyd Regier	UCSD/SIO	
Elizabeth Rosenzweig	Bubble Mountain Consulting	CI Information Architect, Consultant
Alex Talalayevsky	Ocean Leadership	
Keith von de Height	WHOI	
Bob Weller	WHOI	CGSN PI

E Abbreviations

Abbreviation	Meaning
CI	OOI CyberInfrastructure
CI ADT	OOI CyberInfrastructure Architecture and Design Team
CI IO	OOI CyberInfrastructure Implementing Organization
IOOS	Integrated Ocean Observing System
NetCDF	Network Common Data Form
OOI	Ocean Observatories Initiative
PDR	Preliminary Design Review
SCCOOS	Southern California Coastal Ocean Observing System

F References

Reference	Citation
[CI-CARCH]	CI conceptual architecture and initial requirements, available at http://www.orionprogram.org/organization/committees/ciarch
[CI-DPG-WEB]	OOI CI Requirements Elicitation Workshop, Data Product Generation. Website available at: http://www.ooici.ucsd.edu/spaces/display/WS/RWS-DPG
[CI-IOM-WEB]	OOI CI Requirements Elicitation Workshop, Integrated Observatory Management. Website available at: http://www.ooici.ucsd.edu/spaces/display/WS/RWS-IOM
[CI-OOP-WEB]	OOI CI Requirements Elicitation Workshop, Ocean Observing Programs. Website available at: http://www.ooici.ucsd.edu/spaces/display/WS/RWS-OOP
[CI-PAD]	ORI CI Architecture Document, PDR Final version, 16-Nov-2007
[CI-RWS1]	OOI CI First Science User Requirements Elicitation Workshop Report, OOI CI, Final version 1.0, 08-Nov-2007, available at: http://www.ooici.ucsd.edu/spaces/download/attachments/10453181/OOI-CI-ReqWS1-Report-FINAL.pdf?version=1
[CI-RWS2]	OOI CI Second Science User Requirements Elicitation Workshop Report, OOI CI, Final version 1.0, 09-May-2008, available at: http://www.ooici.ucsd.edu/spaces/download/attachments/10453181/OOI-CI-ReqWS2-Report.pdf?version=2
[CI-WEBSITE]	OOI CI Website, available at http://www.ooici.uscd.edu
[NORIA]	Network for Ocean Research, Interaction and Application (NORIA) Proposal, 22-Dec-2006
[OOI-CU-REQ]	OOI Cyber User Requirements, exported from OOI DOORS requirements database. Version of 7/31/08. Available at: http://www.ooici.ucsd.edu/spaces/display/WS
[SCIPROSP]	OOI Science Prospectus, Dec 2007, available at: http://www.oceanleadership.org/files/Science_Prospectus_2007-10-10_lowres_0.pdf