Digital Moorea Cyberinfrastructure for Coral Reef Monitoring

Tony Fountain #1, Sameer Tilak #2, Peter Shin #3, Sally Holbrook #4, Russell J. Schmitt #5, Andrew Brooks #6, Libe Washburn #7, David Salazar #8

# California Institute of Telecommunications and Information Technology, UCSD
La Jolla, CA, 92093, USA

tfountain@ucsd.edu, stilak@ucsd.edu, pshinn@ucsd.edu

# Marine Science Institute, UCSB
Santa Barbara, CA, 93106, USA

holbrook@lifesci.ucsb.edu, schmitt@lifesci.ucsb.edu, brooks@msi.ucsb.edu, washburn@icess.ucsb.edu

salazar@msi.ucsb.edu

Abstract—Digital Moorea is a collaborative vision of a coral reef ecosystem instrumented with real-time sensors connected to high-performance backend resources and sophisticated client applications. It will be a living laboratory for long-term studies of marine ecology and a testbed for evolving technologies for environmental and biological sensing, communications, and analysis. A diverse team of ecologists, computer scientists, and engineers from the Marine Science Institute at the University of California Santa Barbara (MSI, http://www.msi.ucsb.edu/) and the California Institute of Telecommunications and Information Technology (CalIT2, http://www.calit2.net/) are collaborating to bring this vision to reality at the Moorea Coral Reef site (MCR LTER, mcr.lternet.edu) of the U.S. National Science Foundation’s Long Term Ecological Research (LTER) program.

I. INTRODUCTION

Digital Moorea is a collaborative vision of a coral reef ecosystem instrumented with real-time sensors connected to high-performance backend resources and sophisticated client applications. It is envisioned to be a living laboratory for long-term studies of marine ecology and a testbed for evolving technologies for environmental and biological sensing, communications, and analysis. A diverse team of ecologists, computer scientists, and engineers from the Marine Science Institute at the University of California Santa Barbara (MSI UCSD, http://www.msi.ucsb.edu/) and the California Institute of Telecommunications and Information Technology at the University of California San Diego (CalIT2 UCSD, http://www.calit2.net/) are collaborating to realize this vision.

Coral reefs and their associated communities are some of the most diverse ecosystems on the planet. It is estimated that one quarter to one third of all marine species inhabit coral reefs [9]. Despite occurring in nutrient-poor waters, coral reefs are among the most productive and species-rich ecosystems on earth. They also provide critical ecosystem services, e.g., 10% of world’s fishing harvest [10].

Coral reefs are vulnerable to disturbances that can occur on both local and global scales, e.g., pollution, non-sustainable harvesting, and development activities. They also are particularly sensitive to the environmental drivers associated with climate change, especially increases in temperature and ocean acidity. To understand the ecology of coral reefs and to inform both scientists and managers, researchers have identified key questions including: How resistant are coral reefs to degradation? After degradation, what affects recovery? Can rates of recovery keep up with rates of disturbance? What is the joint effect of disturbances and climate forcing? Addressing these questions will require detailed observations and controlled experiments.

The Moorea Coral Reef (MCR) Long Term Ecological Research (www.mcr.lternet.edu) site is the only member of the U.S. LTER network (www.lternet.edu) focused on coral reefs. The MCR site is the coral reef complex surrounding the island of Moorea, French Polynesia (17°30’S: 149°50’W). Moorea is a small, triangular volcanic island 20 km west of Tahiti in the Society Islands. An offshore barrier reef forms a system of shallow (mean depth ~ 5-7 m), narrow (~0.8-1.5 km wide) lagoons around the 60 km perimeter. All major coral reef types (e.g., fringing reef, lagoon patch reefs, back reef, barrier reef and fore reef) are easily accessible for study. Field research is conducted at the UC Berkeley Gump Research Station (http://moorea.berkeley.edu/).

Sensor deployments in the lagoons and on the fore reef surrounding Moorea began in 2005. Currently there are six instrumented sites (two per side of the island) and a seventh site located in one of the deep-water bays on the island’s north shore (Figure 1). All data collection and processing is performed by the sensors and data are stored using internal memory. SCUBA divers retrieve the sensors every few months to download data, replace batteries, clean housings, and then redeploy on the reef. These sensor deployments provide the foundation for building the Digital Moorea real-time sensing system.
In October 2007, we initiated the Digital Moorea project with funding from the National Science Foundation and the Gordon and Betty Moore Foundation. The overall goal was to develop an integrated cyberinfrastructure that supports the complete lifecycle of data products (from acquisition to publication) and the full range of scientific activities (including monitoring, experimentation, analysis, modelling, and collaboration). Our approach involves incrementally integrating the stand-alone sensor deployments into an end-to-end cyberinfrastructure that supports real-time science and engineering activities. From October 2007 until May 2009, our team developed design documents and fielded prototypes to test infrastructure ideas and components. These activities culminated with an operational field deployment in May 2009.

In May 2009, the first phase of Digital Moorea was realized with the deployment of conductivity, temperature, and depth sensors on the fringing reef and the installation of a weather station containing temperature, rainfall, wind, barometric pressure and photosynthetically active radiation sensors on shore at the Gump Research Station. These sensors stream observations to a Data Center located at UCSD that includes cyberinfrastructure for real-time streaming data acquisition, scalable event stream processing, and data publication services. Scientists at UCSB and other remote locations access the data and event streams via a suite of client applications for visualization, modelling, and analysis. The system is engineered to be scalable, robust, extensible, and secure. It is built using state-of-the-art open-source software tools. This paper describes the Digital Moorea cyberinfrastructure, the requirements and challenges that drove the architecture design, the current deployment status, and future plans.

II. DIGITAL MOOREA CYBERINFRASTRUCTURE

The Digital Moorea cyberinfrastructure was designed to be long-lived and evolve gracefully from a small-scale sensor deployment to a fully operational real-time environmental observatory. The first phase deployment was the result of an agile development process involving a cross-functional team of engineers, biologists, and computer scientists [13]. Our development process involved the following activities: (1) Analysis of system requirements, (2) Analysis of risk factors, (3) Adoption of design principles, (4) Architecture design, (5) Technology selection, (6) Bench testing, and (7) Field deployment. These activities were conducted in a highly collaborative iterative process with a focus on rapid software releases and frequent stakeholder reviews.

1. System Requirements

Digital Moorea is intended to serve a broad range of science applications and support both experimental and production cyberinfrastructure activities. It must satisfy a rigorous set of system requirements including:

1) Heterogeneous sensors/instruments: The system must support integration of numerous instruments from a variety of vendors. These include both commercial and experimental instrument packages.

2) Real-time processing: The system must support acquisition and processing of observations in real-time. The temporal scale of processing must be tuneable to adapt to varying application demands and system characteristics.

3) Automated event detection: The system must be able to detect and respond to events associated with environmental phenomena and system state of health in a timely manner.

4) Data acquisition over wired and wireless networks: The system must operate efficiently over a variety of network infrastructures, including local wireless networks and long-haul internet-scale infrastructure.

5) Support for evolving research programs: The cyberinfrastructure must be extensible and easily configurable to adapt to evolving changes in sensors, networks, and processing platforms.

6) Interoperability with other coral reef observatories: The system must support the sharing of resources, including data and analysis tools, with other research observatories, especially members of the Coral Reef Environmental Observatory Network (CREON) [4].

2. Risks and Challenges

The Digital Moorea site is remote and the local infrastructure is somewhat rudimentary, e.g., limited power, networking, and on-site technical staff. This introduces significant risk factors and challenges into system design, development, and operations including:

1) Remote location and long distance data acquisition are challenges. Moorea is approximately 6,600 km from the Data Center at UCSD. The distance between field operations and Data Center, especially when operations rely on heterogeneous (and in some places low-bandwidth) networks, challenges efficient real-time operations.
2) The remote management of software services and hardware components is difficult. This is further complicated by a lack of dedicated IT staff on the island. Field operations often require periodic site visits by personnel from UCSB/UCSD.

3) Unreliable power sources on Moorea are the norm. The island power infrastructure is subject to frequent blackouts, brownouts, and power surges. Engineering and operating the system to be robust under unreliable power sources is a challenge.

4) Hardware equipment failures have been common. Digital Moorea consists of numerous infrastructure components, all of which are subject to failure. For example, over the last year we have experienced failures of routers, the DSL line, compute servers, and hard-drives. The remoteness of the Moorea site and the lack of permanent on-site technical staff raise challenges in troubleshooting and repairing these types of problems.

5) Limited and evolving global-scale networking infrastructure presents challenges for application tuning and operations. The current network between Moorea and researchers in the USA is limited. However, this is expected to change dramatically in 2010 when a new fiber optic cable will be deployed between Hawaii and French Polynesia. The challenge is to engineer the system to operate efficiently under the current networking limits and to scale to handle increased loads as the new network becomes available.

3. Design Principles

Based on consideration of the state-of-the-art in observing systems cyberinfrastructure and the real-world requirements and challenges mentioned above, we generated a set of design principles to inform the architectural decisions. These include:

1) Design for real-time operations from the beginning. Real-time data acquisition, event detection, and instrument control are fundamental to the operational requirements of Digital Moorea. Therefore we start from the premise that streaming-data support should be built into the infrastructure from the beginning.

2) Utilize open-source software for major system components. Digital Moorea is a large, complex system. Open-source software provides the appropriate model for our system development.

3) Keep the island infrastructure as simple as possible. Given the limited resources available on Moorea and the challenges associated with managing remote infrastructure, we decided to minimize the complexity of the remote deployment. Whenever possible, systems services are located at the Data Center rather than at the island.

4) Employ scalable infrastructure components. All system components should be capable of handling high demands, including data acquisition, transmission, processing, and application servers. Even though the first deployment of Digital Moorea has low to moderate demands, i.e., few sensors, few processing workflows, and few clients, the system components were chosen to scale gracefully to orders-of-magnitude increases along each of these dimensions.

5) Plan for a phased deployment. Digital Moorea is a complex and long-term project. However, there is an immediate science agenda that requires basic data services. So, rather than a multi-year design process followed by construction, we chose to deploy the foundational services of Digital Moorea while additional components are still on the drawing board or under development in the laboratory.

The requirements, risks, and design principles presented above provided the basis for our architectural design decisions, technology selections, and system deployment plans.
III. CURRENT DEPLOYMENT

The following description of the current deployment is organized around the three sites over which the cyberinfrastructure is distributed (Figure 2). First, the island infrastructure at Moorea is described. This is followed by a description of the Data Center cyberinfrastructure at UCSD. Finally, the client applications, which are available to scientists at UCSB and other locations, are described.

A. Moorea Field Deployment

There are two main components to the island cyberinfrastructure: (1) a suite of sensors, and (2) an Uplink Node that transmits the data streams to the Data Center at UCSD.

Table 1 provides a list of the sensors deployed in the real-time Digital Moorea system. This is a representative sample of the types of aquatic and terrestrial sensors that are needed for coral reef ecology. The sensors were deployed as follows. The weather station is installed on the roof of a lab building, about 8 meters away from the water, facing the bay at a height of 6 meters above mean water level. It contains a Campbell CR1000 data logger that samples every five minutes from a temperature and humidity probe, a pyranometer, a rain gauge, and a wind monitor. A Seacat 16 plus is deployed in the water about 30 meters away from the lab in approximately 3.5 meter water depth. It measures conductivity, temperature, pressure and salinity every 2 minutes. Facing the shore where the Seacat 16 plus is deployed, an IP-enabled Axis camera is installed on the roof of the next building, about 10 meters away from the weather station at a height of 4 meters.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Type of Measurement</th>
<th>Sampling Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seacat CTD 16 plus</td>
<td>temperature, conductivity, salinity, pressure</td>
<td>2 minutes</td>
</tr>
<tr>
<td>HMP 45C temperature and relative humidity</td>
<td>temperature, relative humidity</td>
<td>5 minutes</td>
</tr>
<tr>
<td>LI200X Pyranometer</td>
<td>solar radiation</td>
<td>5 minutes</td>
</tr>
<tr>
<td>TB4-1 Rain Gauge</td>
<td>rain</td>
<td>5 minutes</td>
</tr>
<tr>
<td>CS100 Barometer</td>
<td>pressure</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Wind Monitor</td>
<td>wind direction and speed</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Axis 223M Camera</td>
<td>mjpeg video</td>
<td>5 minutes</td>
</tr>
</tbody>
</table>

TABLE I
DEPLOYED SENSORS IN REAL-TIME SYSTEM

As mentioned in Section II.3, one of the primary design principles for the island deployment was to minimize risks associated with the limited and unreliable infrastructure. These considerations led us to eliminate the local computer on the island, which previous experience had shown to be unreliable. Instead, we employed a pair of MOXA serial-to-Ethernet converters so that the serial interfaces of the deployed sensors/instruments are accessible from the Data Center (www.moxa.com). The MOXA boxes allow us to treat the data connection between Moorea and UCSD as one long serial line. Both the Campbell data logger and the Seacat CTD communicate using the serial RS-232 protocol. The serial lines from these instruments are connected to a serial-to-Ethernet MOXA 6250 box. Then, on the UCSD side, a paired MOXA box receives the data and feeds it to the data acquisition system in the Data Center. The MOXA 6250 box also contains a 1GB SD memory card, which buffers the sensor data during intermittent network failures. All instruments are on line power.

B. UCSD Data Center Cyberinfrastructure

The Data Center cyberinfrastructure is designed according to modern enterprise computing concepts adapted to the requirements, challenges, and design principles described earlier. It is an open architecture, with modular open-source software components. It provides well-defined services that are largely autonomous, employing concepts from Service-Oriented Architecture design and Cloud Computing [12]. The major modules in the Data Center architecture are implemented using mature open-source software packages and are managed in a Cloud-Computing environment. Cloud computing (i.e., the integration of virtualization, infrastructure as a service (IaaS), software as a service (SaaS), and utility computing) provides a scalable and cost-efficient computing paradigm [1, 7]. In the current Digital Moorea deployment we utilize the Xen hypervisor to provide virtualization in our cloud-computing environment.

Xen is a state-of-the-art open source virtual machine monitor (hypervisor) that allows one physical machine to run multiple virtual machines [2, 4, 6]. Virtualization offers several benefits, including reduction in machine idle time (better use of existing hardware), reduction in IT infrastructure costs, simplified system administration (less physical machines to manage), and increased uptime and faster failure recovery (since virtual machines are not bound to any physical machine, they are more portable and can be efficiently migrated across physical machines). Xen supports applications that require elasticity, i.e., the ability to acquire and release computer resources as demand waxes and wanes. For example, under periods of increased demand, virtual machines can be migrated from low-capacity nodes to high-capacity nodes. As the demand is reduced, the system can migrate back and free up the high-capacity nodes. This type of behaviour is a good match for operational scenarios where demand fluctuates in response to (1) environmental events, e.g., increased sampling or model activation based on observed phenomena, and (2) periodic increases in user demand due to collaborative experiments or training sessions.

1) Downlink Node:

The Downlink Node is responsible for receiving the data streams from Moorea and replicating and routing those data streams to various applications within the Data Center. The Downlink Node consists of 2 components: (1) a MOXA box...
model 5110, and (2) a DataTurbine server. The MOXA box captures the data streams from the MOXA box 6250 in Moorea. It is similar to the MOXA box 6250 that is used in the Uplink Node, however it does not have the 1GB external buffer (since this is not needed on the Data Center side). The MOXA 5110 box connects to a dedicated DataTurbine server.

DataTurbine is a real-time streaming data engine. It is an open-source middleware product supported by NSF, NASA, and private industry. It is managed by the NSF-sponsored Open Source DataTurbine Initiative at CalIT2 (www.dataturbine.org). The DataTurbine middleware satisfies a core set of infrastructure requirements that are common in environmental observing systems, including reliable data transport, a framework for integrating heterogeneous instruments, and a comprehensive suite of services for data management, routing, synchronization, monitoring, and visualization [6, 11]. From the perspective of distributed systems, the DataTurbine middleware is a "black box" to which applications and devices send and receive data. DataTurbine handles all data management operations between data sources and sinks, including reliable transport, routing, scheduling, and security. DataTurbine accomplishes this through the innovative use of flexible network bus objects combined with memory and file-based ring buffers. Network bus objects perform data stream multiplexing and routing. Ring buffers provide tunable persistent storage at key network nodes to facilitate reliable data transport. There are two DataTurbine servers running in the Xen-based cloud in the Digital Moorea Data Center at UCSD. One DataTurbine server is part of the DownLink node, where it replicates the sensor streams and sends one stream to the Event Detection Module and one stream to a second DataTurbine server. The second DataTurbine server publishes the data and event streams to external applications and clients and is described in Section III.3.

2) Event Detection Module:

In Digital Moorea, real-time event detection is needed. Events can be classified as either (1) environmental phenomena, or (2) system state of health conditions. The Digital Moorea Event Detection Module is designed to handle both types of events.

The detection of environmental events is accomplished by observing patterns in environmental sensor streams. Here are 3 examples of environmental events that are of interest to MCR scientists. (1) Coral bleaching: The onset of some types of coral bleaching can be detected by a rapid rise in temperature during periods of intense sunlight. (2) Nutrient runoffs: Freshwater runoffs from agricultural fields introduces nutrients into the bays and waters around Moorea. This can be detected by observing temporal correlations between salinity and rainfall measurements. (3) Internal waves: Characteristic periodicities in temperature caused by internal waves can be detected by correlations between temperature measurements at various depths and locations. In these three cases the response to the detection of an event is to initiate a field campaign to gather water samples or other biological data. However, more complex scenarios will involve automated responses, e.g., increasing sample rates or activating and coordinating additional sensors.

The detection of system state of health events is accomplished by observing patterns in cyberinfrastructure metric streams. Events can be indicators of (1) catastrophic failures, e.g., loss of a network link, or (2) calibration problems, e.g., sensor or clock drift, or (3) trends that threaten quality of service (QoS) agreements, e.g., buffer saturation, server loads, or transmission delays. As with the environmental events, the responses to state of health events can be simple, e.g., send a technician to replace a battery, or sophisticated, e.g., migrate a virtual machine to a more powerful compute platform.

Our approach to event detection is based on Event Stream Processing (ESP) and Complex Event Processing (CEP) [14]. For the Event Detection Module we adopted the open source Event Stream Processing and event correlation engine Esper (http://esper.codehaus.org/). Relational databases and the structured query language (SQL) are designed for applications in which most data are fairly static and complex queries are infrequent. Esper inverts the typical data-query relationship. Instead of storing the data and running queries against stored data, the Esper engine stores queries and runs the data through the queries. Esper event stream queries provide the windows, aggregation, joining and analysis functions needed for detecting complex events. The Event Detection Module is located at the UCSD Data Center, however we plan to migrate much of this functionality to the island in future deployments. The long-term goal is to embed event detection and adaptive control onto smart buoys and other field devices.

3) Real-Time Data Server:

The Data Center provides a publish-and-subscribe service for real-time data and event streams. This allows users to receive continuous updates and drive various real-time applications. In the current deployment a dedicated DataTurbine server provides the publish-and-subscribe service. Data streams from the Downlink Node are mirrored directly to this DataTurbine server. Similarly, the event streams from the Event Detection Module also feed directly into this server. While the current publish-and-subscribe service is based on native DataTurbine APIs, future plans call for exposing these services through interfaces compliant with standards established by the Open Geospatial Consortium’s (OGC) Sensor Web Enablement (SWE) initiative (www.opengeospatial.org/projects/groups/sensorweb).

4) Archival Database:

The current implementation of archival data storage for Digital Moorea consists of a PostgreSQL relational database that is linked to the Real-time Data Server. We employ a minimalist database schema and use a generic SQL-based query client. The next version of the Archival Database will include translators to netCDF and access through an
OPeNDAP server (opendap.org). As with the real-time data streams, the archived data will be published through an OGC-compliant Sensor Observations System (SOS) service (www.opengeospatial.org/standards/sos).

5) OptIPortal:

Situational awareness in large sensor-based systems is a challenge due to the large number of data streams that are collected and the limited real estate of most computer displays. An OptIPortal is a large-scale visualization system that can be deployed on a variety of hardware platforms (wiki.optiputer.net/optiportal/). OptIPortals drive tile-wall displays, i.e., numerous displays composed to yield one very large virtual display. Existing OptIPortals consist of between 4 to 55 displays, with the larger OptIPortals providing 200 megapixels. For Digital Moorea, the live data streams and the video camera feeds are integrated into the OptIPortal system at CalIT2, providing large theatre-type viewing opportunities for science, education, and policy sessions. The CalIT2 system lays a foundation for future development of a dedicated control room for managing the Digital Moorea system.

C. Client Applications

The system supports a number of client applications that can be run remotely. Some of these operate on real-time data streams; some operate on the archived data. These include the DataTurbine Real-time Data Viewer (RDV), a utility for creating embedded web page graphs, a database query interface, a MATLAB interface, and a GoogleEarth plug-in. There is also an interface to the Kepler workflow system (kepler-project.org/). These generic applications connect directly to the DataTurbine server or to the PostgreSQL database. Future clients will be customized for Digital Moorea applications and connect via OGC-compliant interfaces.

IV. RESULTS

A. Current Status

The system has been operational since coming on line in May 2009. The Data Center services have been very stable. The only interruptions were for scheduled system maintenance. The field data acquisition system has also been stable, except for problems with the DSL modem and the DSL phone line. These problems have been fixed and there have been no additional problems. The system has been robust to occasional power and network outages. We are currently gathering system metrics, e.g., network latency, and conducting various tests on system performance.

B. Plans for the Next Phase

For the next phase of Digital Moorea, we have extensions planned in 3 areas: (1) integrating additional sensors, adding buoys and wireless radios, (2) exposing data and system resources as OGC-compliant services, and (3) deploying an enterprise service bus (ESB) infrastructure in the Data Center [2].

V. CONCLUSION

A team of ecologists, computer scientists, and engineers from the Marine Science Institute at the University of California Santa Barbara and the California Institute of Telecommunications and Information Technology are collaborating to develop a real-time coral reef monitoring system at the Moorea Coral Reef LTER site. The system is designed to support a broad range of science and engineering activities. It includes cyberinfrastructure for data acquisition, event detection, data processing and publication. The first phase of this system was deployed in May 2009. Initial tests indicate that the system is robust and reliable. Future plans call for system extensions, including additional sensor deployments, OGC-compliant service interfaces, and a Data Center enterprise service bus.

ACKNOWLEDGMENTS

This work is supported by a grant from the Gordon and Betty Moore Foundation, also NSF OCI Award 0722067, and NSF MCR LTER Award OCE 04-17412.

REFERENCES