Sensors, Scientists and Streaming Data: Building a New Research Infrastructure for Community-Based Research Networks

The Cases of Global Lake Ecological Observatory Network (GLEON) and the Coral Reef Ecological Observatory Network (CREON)

December 5, 2007

Maureen C. Curran

North Temperate Lakes (NTL) LTER site: Paul Hanson on a remote sensing raft on Sparkling Lake, northern Wisconsin. A meteorological station above water and a thermistor chain and oxygen probe below water transmit near real-time data to the web; each buoy is solar-powered and hosts a datalogger. NTL is one of the first GLEON sites. *Photo by Tim Kratz*

A wide-ranging collaboration of researchers will develop, improve and deploy tools to support globally-encompassing, grass-roots, community-based research networks studying lake and coral reef ecosystems. They will build cyber- and social infrastructures of sensors, scientists, streaming data and sharing, developing tools with ease of deployment, robustness and stability to enable easily accessed, networked, collaborative ecological science on an international scale. In the process, innovative research will be conducted.

The University of California San Diego (UCSD), UC Santa Barbara (UCSB), the University of Wisconsin, State University of New York, Binghamton (SUNY) and the National Center for High-performance Computing (NCHC, in Taiwan) are the partners in this three-year, nearly $2 million project, which will be housed in the UCSD division of the California Institute for Telecommunications and Information Technology (Calit2).

"This project will not only help develop infrastructure for sensor networks on coral reefs, it will also play a big role in building the scientific community of users," said UCSB's Sally Holbrook, lead researcher of the Coral Reef Ecological Observatory Network (CREON). "The data that scientists will be able to share will be of critical importance to our studies of the effects of disturbances and climate forcing on coral reef ecosystems."

This multifaceted project builds on current and past work from several projects, in particular, the Global Lake Ecological Observatory Network (GLEON), CREON, the Open Source DataTurbine Initiative and a National Science Foundation (NSF) award to the Univ. of Wisconsin to build the research agenda around networked science. They will also be leveraging the infrastructure of two Long Term Ecological Research (LTER) Network sites (North Temperate Lakes, WI and Moorea Coral Reef, Polynesia) and the results of an earlier project which examined how to scale processes, as well as the Pacific Rim Applications and Grid Middleware Assembly (PRAGMA). These projects are/have been funded by NSF, and/or in some cases, the Gordon and Betty Moore Foundation, with additional funding to international sites from their country's domestic funding agencies.

Open Source DataTurbine Initiative

Gordon and Betty Moore Foundation

"This project allows us to glue together several efforts," noted UCSD's Peter Arzberger, the principal investigator (PI) of the new award. "It provides a concrete infrastructure for the first time to this distributed

community, and allows us to move the community forward in the way it conducts science." Arzberger is also the chair of the PRAGMA Steering Committee, a long-time participant in Calit2 and has leadership roles with GLEON and CREON.

Computer network researchers are joining with global lake and coral reef ecosystem scientists to create and test distributed network infrastructure and data; in parallel, they will also build a personal, social network infrastructure in order to conduct joint, shared research.

"We are delighted to be officially engaged in this project," said NCHC's Fang-Pang Lin, "The activities will help integrate technology and expertise from the sites involved across these two international applications (ecology of lakes and coral reefs)."

Lin continued: "Furthermore, this project will have impact well beyond these applications to other others, such as the flood grids and geogrid in Taiwan." Lin is the lead for Grid Applications at NCHC and co-lead of the PRAGMA Telescience Working Group. He is also a member of PRAGMA´s, GLEON´s and CREON´s Steering Committees.

There are several aspects to this project to create tools which will allow (and encourage) researchers to share data: development (including improved usability); user documentation; deployment and regular meetings of the community (similar to those held by PRAGMA which have proven the efficacy of this method for community building). Robustness and ease of use are paramount to success.

"From my perspective, the major promise of the project is that we will be able to have a robust infrastructure that will allow scientists to access high-frequency data from lakes around the world from a single portal," said GLEON's Tim Kratz. "This will make comparative analyses across a wide variety of lakes much easier to do." Kratz the director of the Trout Lake Station, Center for Limnology at the Univ. of Wisconsin-Madison

The overarching goal is to create the nucleus of a functioning network of distributed international sites streaming and sharing data to conduct networked (i.e., comparative) research. In parallel, the project will create social and professional interactions to build the community and encourage and enhance tool adoption and collaboration. And finally, to demonstrate the value of this approach by conducting novel research and testing new research ideas and models based on networked science.

To achieve this, multiple layers of activity will be ongoing simultaneously. Following is a brief overview of these activities with respect to the key partners and lead personnel.

The key leads at UC San Diego are Peter Arzberger, along with Tony Fountain and Sameer Tilak of the San Diego Supercomputer Center (SDSC) at UCSD. They will coordinate the network level activities, design, construct, deploy and document pieces of the infrastructure for the community. In particular, Fountain and Tilak will focus on improving usability and automating deployment of the SDSC-based Open Source DataTurbine Initiative tool for streaming data -- hardening the currently brittle "on and off ramps" to the cyberinfrastructure.

The University of Wisconsin and the North Temperate Lake Long Term Ecological Research Site (LTER) will provide the scientific lead for the driving research for lakes. Key lead personnel are Tim Kratz and Paul Hanson. In particular, they will drive the best practices involving metadata and common data nomenclature used in the project and will also help with guiding the deployments from U WI. Working closely with the UCSD team, they will be early adopters, testers, co-developers and deployers of the tools, as well as documenting the deployment operations. NTL is one of the first GLEON sites. U WI will be the long-term home of the GLEON data servers and portals.

UC Santa Barbara is the lead institution for the new Moorea Coral Reef LTER site located in Moorea, French Polynesia. They will provide the key driving science for coral reefs, and define the driving question. Key personnel leads are Sally Holbrook, Russ Schmitt and Andrew Brooks. They will assist in deploying codes to CREON, first

in Moorea. To do this, they will also work closely with the team in San Diego. They will develop key variables and common nomenclature, and ensure deployment of tools in the coral reef community. UCSB will be the long-term home of the CREON data servers.

The National Center for High-performance Computing (in Hsinchu) is the networking lead in Taiwan, and this group has been instrumental in developing the wireless networking to Yuan Yang Lake as well as wired monitoring to the Kenting Coral Reef ILTER site. The key leads are Fang-Pang Lin, Hsiu-Mei Chou and Jazz Wang. They will assist sites in setting up wireless networks and they will document these steps for others to follow. Yun-Te Lin, NCHC's key architect for tiled-display walls (TDW) and streaming data technology, will be involved as the project develops.

Ken Chiu of the State University of New York, Binghamton (SUNY) is an unofficial partner, who has been working closely with the team on data ingestion, sensor control, and scientific interfaces. He will interact with and advise the group on enhancements to and extensions of the deployed infrastructure, for use by scientists. He will be working closely with researchers affiliated with and managers at the Lake Sunapee Protective Association.

"What I hope will come out of this is a concrete model of using a deployed sensor network to conduct networked science," concluded Arzberger, "and to discover new understandings of limnological or coral reef significance." Arzberger is also the director of the National Biomedical Computation Resource (NBCR) at UCSD.

Funding for this project is being provided by the Gordon and Betty Moore Foundation. Both GLEON and CREON will in turn be leveraging this award to work towards their long term goal of attracting and adding new scientists and sites, building their networks further.

About Calit2 The California Institute for Telecommunications and Information Technology, a partnership between UC San Diego and UC Irvine, houses over 1,000 researchers organized around more than 50 projects on the future of telecommunications and information technology and how these technologies will transform a range of applications important to the economy and citizens' quality of life. http://www.calit2.net

About CREON The Coral Reef Environmental Observatory Network is a collaborating association of scientists and engineers from around the world striving to design and build marine sensor networks. A coordinated network involving coral reef research and information technology efforts such as envisioned for CREON, the Coral Reef Ecological Observatory Network, will enable key physical variables that drive coral reef ecosystems and their biotic responses to be quantified, studied in real time, and compared among coral reef ecosystems. At the current time there are three coral reef sites actively involved: the Moorea Coral Reef (MCR) LTER site in Moorea, French Polynesia, Davies Reef on the Great Barrier Reef, Australia, and the Kenting Coral Reef ILTER site in Taiwan. In addition to the Moore Foundation support, CREON activities are supported by NSF and its member institutions. http://www.coralreefeon.org/

About Open Source DataTurbine Initiative The Open Source DataTurbine Initiative is a robust open-source streaming data middleware system that satisfies the core requirements for sensor-based environmental observing systems. It is relatively mature and has been tested in a variety of real-world streaming data applications. It facilitates the development of complex distributed streaming data applications, including real-time virtual observatories and telepresence collaboratories. The Open Source Data Turbine Initiative is supported by NSF. http://www.dataturbine.org/

About GLEON The Global Lake Ecological Observatory Network is a grass-roots network of limnologists, ecologists, information technology experts, and engineers who have a common goal of building a scalable, persistent network of lake ecology observatories. The observatories consist of instrumented platforms on lakes around the world capable of sensing key limnological variables and moving the data in near-real time to web-accessible databases. Several lakes in Wisconsin and Yuan Yang Lake in Taiwan have been instrumented to date. GLEON builds on a successful pilot project involving the North Temperate Lakes LTER project, as well as the other projects and research teams in this story. In addition to support by the Moore Foundation, GLEON activities are supported by NSF and GLEON member institutions. http://gleon.org/

About Gordon and Betty Moore Foundation The Gordon and Betty Moore Foundation, established in 2000, seeks to advance environmental conservation and cutting-edge scientific research around the world and improve the quality of life in the San Francisco Bay Area. http://www.moore.org

About Long Term Ecological Research (LTER) Network LTER is a collaborative effort involving more than 1800 scientists and students investigating ecological processes over long temporal and broad spatial scales. The Network promotes synthesis and comparative research across sites and ecosystems and among other related national and international research programs. The National Science Foundation established the LTER program in 1980 to support research on long-term ecological phenomena in the United States. http://www.lternet.edu/

About National Biomedical Computation Resource (NBCR) The mission of the NBCR is to conduct, catalyze, and enable biomedical research by harnessing, developing, and deploying forefront computational, information and grid technologies, and disseminating results through work with and training of the biomedical research community. NBCR is supported by the National Center of Research Resources of the National Institutes of Health. http://nbcr.net/

About PRAGMA The Pacific Rim Application and Grid Middleware Assembly (PRAGMA) was formed in 2002 to establish sustained collaborations and advance the use of grid technologies in applications among a community of investigators working with leading institutions around the Pacific Rim. Applications are the key to PRAGMA, integrating focus that brings together the necessary infrastructure and middleware to advance the application's goals. PRAGMA is supported by the National Science Foundation and by PRAGMA member institutions. http://www.pragma-grid.net/

Media Contact: Maureen C. Curran, 858-822-4084