2014

OVERVIEW

The Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) is a robust, international network of research scientists from more than 30 institutions who address science and cyberinfrastructure challenges of common interest.

PRAGMA pursues activities in four broad interdependent areas:

- **Fostering international "scientific expeditions" by forging teams of domain scientists and cyberinfrastructure researchers** who develop and test information technologies that are needed to solve specific scientific questions and create usable, international-scale, cyber environments;
- **Developing and improving a grassroots, international cyberinfrastructure** for testing, computer science insight and advancing scientific applications by sharing resources, expertise and software;
- **Infusing new ideas by developing young researchers** who gain experience in cross-border science and by extending engagements with strategic partners; and
- **Building and enhancing the essential people-to-people trust and organization** developed through regular, face-to-face meetings—a core foundation of PRAGMA's success.

None of these activities can stand alone. Ideas become stale without new people. Infrastructure that has impact must be built to meet the specific needs of users and applications, while scientific expeditions and sharing of data require organizational support to succeed.

PRAGMA's Past Successes

PRAGMA was launched in 2002 as a workshop series to explore the technical, organizational and trust elements needed to enable small- to medium-sized international networks of research scientists. Since its inception PRAGMA has expanded its efforts and membership, responding to various disasters and emergencies (including SARS in 2003 and the Japanese tsunami of 2011), and creating a persistent, experimental test bed for international cyberinfrastructure, as well as enhanced and multiple software packages.

Over the next several years, PRAGMA will advance the goals of scientific expeditions...

PRAGMA has also incubated new scientific networks such as the Global Lakes Ecological Observation Network (GLEON) and GEO Grid. The organization's members have played a critical role in the formation of programs that help train the next generation of cyber-scientists through research exchanges, including the Pacific Rim Experiences for Undergraduates (PRIME) at the University of California San Diego, the Monash University Research Project Abroad (MURPA) program and its extension to the University of Queensland, both in Australia.

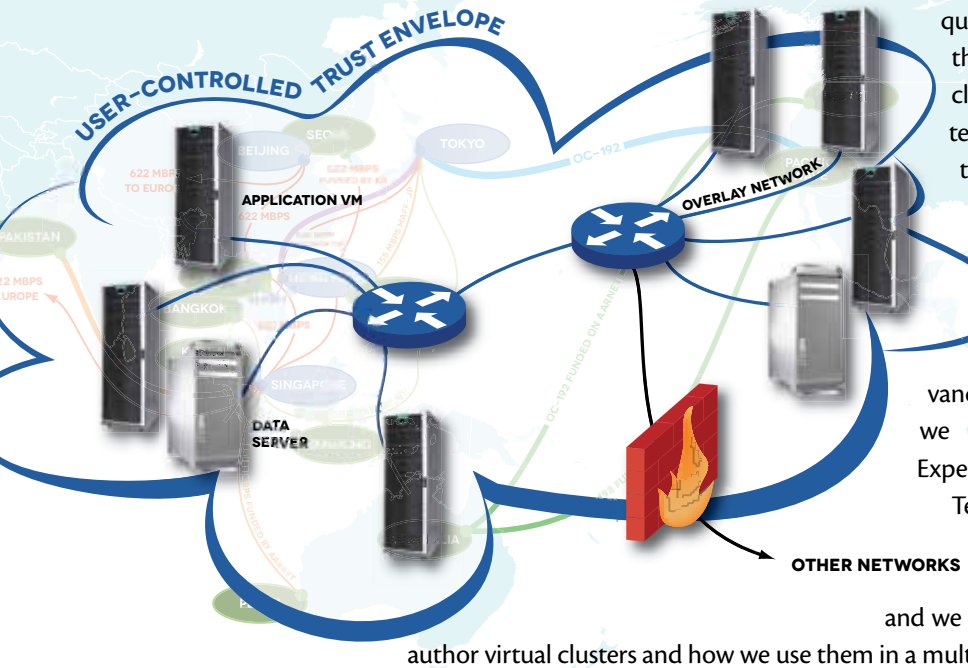
The Future of PRAGMA

Over the next several years, PRAGMA will advance the goals of scientific expeditions that are forged from multidisciplinary teams to solve scientific problems through developing cyberinfrastructure. One such expedition seeks to **understand biological adaptation in extreme environments,** while researching the specific biotic, abiotic and evolutionary factors that affect patterns of diversity, distribution and endemism in ultramafic (high magnesium- and iron-oxide concentration) regions in Southeast Asia. A second expedition will address the societal issues inherent in **predicting lake eutrophication** (i.e., excessive plant, algal and bacterial growth in lakes due to nutrient enrichment)—one of the greatest water-quality challenges facing freshwater ecosystems throughout the world.

These expeditions will develop experimental infrastructure in order to advance the specific research goals of their participants. For example, the biodiversity expedition is seeking to share a software platform via virtualization technologies, while the lake expedition seeks to scale the use of resources through software-defined networks to meet their scientific needs. Both expeditions require developing tools that track usage of data and protect sensitive data by using virtual private networks. In all cases, we will work with existing scientific networks (e.g., GLEON) or grow nascent networks of researchers to ensure that our projects have a broad impact.

Images: (images are top to bottom, those captured are listed left to right) PRAGMA 26, Tainan, on the campus of National Cheng Kung University; Hsiu-Mei Chou (NARL/NCHC), Cayelan Carey (Virginia Tech), Teri Simas (UCSD), PRAGMA 26, Tainan; Allen Nguyen beneath the last torii gate of the Yoshida trail en route to the summit of Mt. Fuji; Peter Arzberger (UCSD), Li-Der Chou, Ce-Kuen Shieh, and Fang-Pang Lin (NARL/NCHC), William Chang (NSF) at PRAGMA 26

PRAGMA also intends to build on its successes in migrating virtual machines to become a **multi-provider cloud** (or multi-cloud) with some of the characteristics of federated clouds. Our approach uses software-defined networks (SDN) to simplify application design by solving various connectivity issues among cooperating, virtualized resources. The SDNs also form trust envelopes so that PRAGMA can practically address fundamental issues in selective data sharing, provenance and raw data acquisition. Furthermore, the PRAGMA multi-cloud can be used as a test bed and resource to evaluate cyberinfrastructure components, gain insights into long-term data use, and advance science. This year we also established an Experimental Network Testbed (ENT) to allow testing of different SDN approaches, and we standardized how we author virtual clusters and how we use them in a multi-cloud environment.



For the long-term vibrancy of PRAGMA, we actively seek to engage new participants through student activities and forming strategic partnerships. Building on the activities of PRIME and MURPA, and motivated by the GLEON Student Association's (www.gleon.org/students) success, we have created PRAGMA Students as a student organization that provides graduate students in PRAGMA with collaborative research and leadership opportunities. Students will also participate in our continued efforts to develop strategic partnerships with both regional (e.g., Southeast Asia) as well as topical areas of interest (biodiversity, disaster mitigation, and such). Through this activity, PRAGMA members will engage new partners in targeted workshops and in interactions of mutual benefit. These special workshops augment our ongoing workshops that are open to members and others interested in participating, contributing to, and learning from PRAGMA activities. The regular workshops rotate among member institutions to engage researchers and students who may be relatively new to PRAGMA, and to educate members about specific applications and technologies developed at other member institutions and in other regions. In addition, we are in the process of activating new areas in PRAGMA that have arisen directly from conversations among the organization's members (for example, related to cyber-learning).

Images: (top) On June 27, 2014 Derek Song, Anthony Nguyen and Katy Pham visiting the Todaiji Temple in Nara, Japan.; (left) Tainan, Taiwan; (right) PRAGMA Students leadership and members, PRAGMA 26, Tainan, left to right: Yuan Luo (IU), Pongsakorn U-chupala (AIST), Beth Plale (IU), Meilan Jiang (Konkuk U), Hiroki Ohtsuji (Osaka U), Chawanat Nakasan (AIST), Quan (Gabriel) Zhou (IU)

In this year's PRAGMA Collaborative Overview for 2014-2015, we highlight activities that exemplify these directions:

- Advancing scientific expeditions in biodiversity and lake ecology by developing and applying advanced cyberinfrastructure approaches;
- Building toward a trust envelope via automating the use of virtual-cluster technology and software-defined networks, as well as building an experimental network testbed;
- Developing PRAGMA Students as a group to promote research and leadership opportunities for students in PRAGMA; and
- Expanding our efforts to build communities in around the areas of biodiversity, water disaster mitigation, and big data, among others, with a particular focus on Southeast Asia.

We are pleased to note that a project developed by a PRIME student not only is in use at a museum, but the app was awarded the Honeysett & Din Student Award in May 2014.

Our advances and accomplishments are possible because of the support we have received, which is acknowledged in the section of this report on **Sponsors**.

Looking to the future, our success will be measured in coming years by the extent to which we:

- **Help advance the scientific goals** of our expeditions;
- **Provide tools and infrastructure** to these groups and the broader community;
- **Harness technology** trends in mobile sensing, smart software, advanced data capture and storage, ubiquitous computing and expanded networks;
- **Create a new generation of scientists and engineers** able to participate in and lead networked groups;
- **Expand participation** in PRAGMA and, more generally, in networked science; and ultimately
- **Enable small-to medium-sized international groups to make rapid progress in conducting research and education** by providing and developing international, experimental cyberinfrastructure.

We look forward to welcoming new participants to PRAGMA and invite interested researchers to our workshops in 2015 and beyond.

EXPEDITIONS AND HIGHLIGHTS

BIODIVERSITY EXPEDITION

Biodiversity research examines the variation among living things and systems, ranging in scale from molecules, genes, cells, and individuals to species through ecosystems. The Pacific Rim, and Southeast Asia in particular, encompass globally significant areas of biodiversity, and Wallace's Line marks a biogeographic transition zone between Asia and Australia. The region offers unique opportunities to study island effects and geological and tectonic complexity as drivers in adaptation and evolution. The high human and economic growth, with co-occurring biodiversity loss and effects from climate change, increase the urgency to use efficient and effective infrastructure and methods in research. One of PRAGMA's interests involving Southeast Asia is to bring a focus on the area that is of great importance in understanding biodiversity.

The PRAGMA Virtual Biodiversity Expedition (VBE) is designed to use advanced technology to address pressing biological questions about species distributions and adaptations in extreme environments. In order to provide a compelling domain science challenge and leverage PRAGMA expertise, we have emphasized biodiversity research regarding ultramafic outcrops on Mount Kinabalu (the highest peak in Borneo at 4095m) in Sabah, Malaysia. These outcrops form extreme environments of "edaphic islands" in a "sea" of more typical soils. Ultramafic soils are deficient in essential macronutrients and have high concentrations of phytotoxic elements and hence pose a challenge for plants to adapt and survive. Ecosystems on ultramafics are widely known to host highly species-rich and endemic vegetation; they make excellent natural laboratories to address topics in evolution, phylogeography, ecology and physiology, including questions of adaptation, and biotic response to climate. Some ultramafic areas have been well studied through comparative ecological studies involving long-term plot studies in Kinabalu Park, involving environmental monitoring sensors. Geospatial analysis allows for understanding complex datasets of various kinds of spatial information, such as spectral data, geological maps and vegetation field data.

The sharing of computational capability, software and data across international networks in a trusted environment is a compelling opportunity for biodiversity researchers. Technologies must be adaptable and responsive to international agreements, collection permitting, and data restrictions on access for use and distribution. As a case in point, the Nagoya Protocol covering genetic resources under the Convention on Biological Diversity was recently ratified by the 50th country and became international law on October 12, 2014 (90 days after the 50th signatory). The implementation of Nagoya is likely to carry implications for museum specimens (a type of preserved genetic resource). The sharing of related biodiversity data from the molecular to ecosystem scales may also be impacted by Nagoya. Information systems will need to provide sufficient metadata, usage tracking, and security to support international collaborative research under a new, more restrictive paradigm. PRAGMA researchers provide the kind of cyberinfrastructure that enables valid research use of biodiversity data in an international context and the international collaborations and partnerships lead to mutually productive outcomes.

This year, the VBE continued to address the technical issues of data and application sharing among collaborators through virtualization. We acquired Unmanned Aerial Vehicle (UAV) data in a series of flights around Mount Kinabalu, along with additional commercial satellite imagery from Kinabalu. PRAGMA participants in Malaysia developed a prototype mobile app for capturing biodiversity data, and Sabah Parks hosted a mini-PRAGMA workshop at the International Conference on Serpentine Ecology (ICSE), engaging a broader group of researchers with the VBE. This section describes the first three of these activities; ICSE is described in the section on **Building Community**.

Virtualizing Lifemapper Software Infrastructure and Sharing Data

The PRAGMA VBE is using member expertise in application deployment to address limitations and bottlenecks surrounding data movement, such as the restricted-use Kinabalu satellite data. Over the last year, PRAGMA collaborators have continued working on virtualization of the Lifemapper software infrastructure, adding the primary data-management component to the computational component.

Lifemapper is a computer infrastructure developed by the KU Biodiversity Institute at The University of Kansas; it includes a suite of data and tools for biodiversity researchers and an archive of species distribution maps calculated from public specimen data. Lifemapper analysis and modeling tools include *Species Distribution Modeling* (LmSDM), which predicts the potential geographic distribution of individual species based on where they have been found (and environmental conditions in those locations), and *Range and Diversity Modeling* (LmRAD), which supports macro-ecological modeling of multiple species over large-scale landscapes.

The infrastructure consists of four independent sub-systems (see figure on the next page), which communicate with each other to process biological data: 1) LmDbServer for data management; 2) LmCompute for computations; 3) LmWebServer for system communications between LmDbServer and LmCompute, and public communications between LmDbServer and users; and 4) client applications. To simplify the initial work, we combined LmDbServer and LmWebServer into a single component, LmServer.

One of the goals of this project is to create a reproducible, platform-independent and configurable infrastructure, capable of fast deployment. These improvements will allow individual projects, groups and scientists to run independent virtual instances of one or more components, configured for their location and data, which are platform independent. Lifemapper is able to scale for greater computational and storage needs. It allows scientists to control how and where their data is deployed, and simplifies end-to-end processing of multi-part datasets. Additionally, the virtual environment allows deployment of the entire infrastructure on a laptop for occasions when hardware is limited or network connectivity varies.

For PRAGMA 25, we streamlined and automated the build process for LmCompute to allow easy deployment of a virtual cluster at SDSC (see section on **Virtual Cluster Sharing**), and we extended the code to produce process-specific provenance logging using Indiana University's KARMA provenance collection tool.

Images: (left to right) Falling Creek Reservoir, Virginia, part of the PRAGMA-GLEON lake expedition—courtesy of Cayelan Carey; Koxinga Ancestral Shrine, Tainan, Taiwan—courtesy Teri Simas; Peter Erskine (U Queensland) holding UAV, in Mount Kinabalu National Park—courtesy of Antony van der Ent

LIFEMAPPER:

A TOOL FOR UNDERSTANDING BIODIVERSITY DISTRIBUTION

Lifemapper (lifemapper.org) provides online biodiversity single- and multi-species research tools and an archive of species distribution maps from geospatial occurrence data aggregated by the Global Biodiversity Information Facility (often based on records going back 300 years). Notably, Lifemapper goes one step further to predict where an individual species could exist now and in the future, based on where it is documented to live. The system does this by combining data on the occurrence of species with global climate, terrain and land cover information to identify environmental correlates of species ranges. Support: U.S. National Science Foundation (several awards)

For PRAGMA 26, we decoupled LmWebServer and LmDbServer from the KU-specific implementation and created a software build process for configurable, deployable virtual servers, running on the same host and enabling a fast turnaround from software update to server availability. The broader purpose was to increase the scalability, availability and flexibility of Lifemapper to enable scientists to assemble multi-species experiments or perform other LM-facilitated processing on unique, restricted-use, very large, or other special-needs datasets.

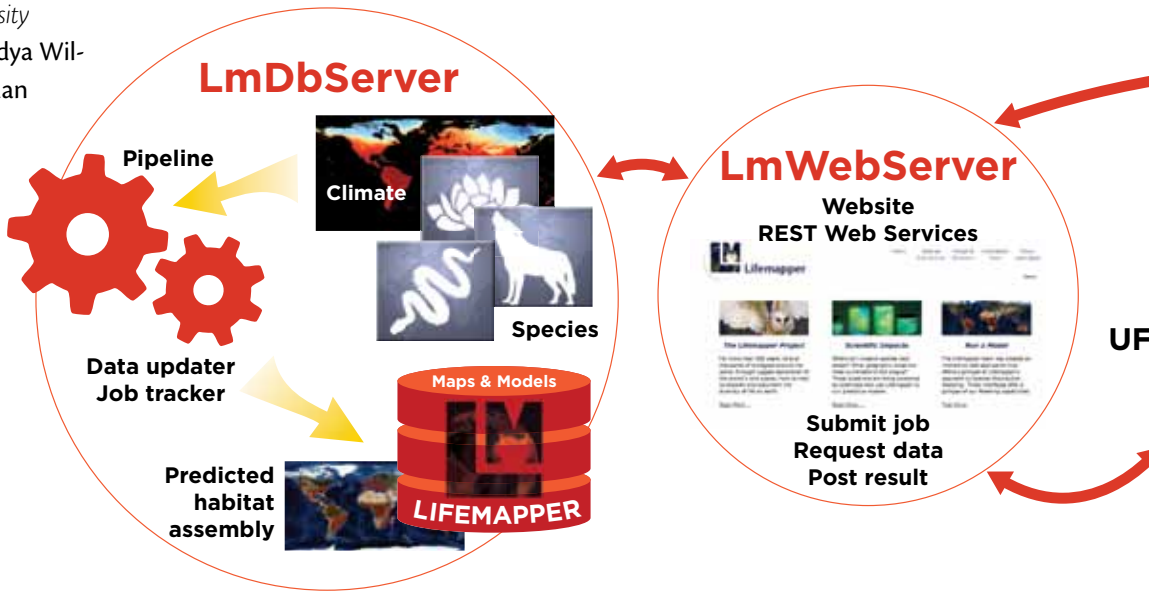
For PRAGMA 27, we completed deployment work by automating data and metadata seeding for a test dataset, then initiated data computation and confirmed that LmWebServer and LmDbServer served both jobs to LmCompute and data to end users. We successfully deployed the entire Lifemapper infrastructure in two unique environments using:

1. KVM virtualization, a solution for Linux on an x86 hardware platform. We have two virtual clusters at SDSC, one running LmCompute and another running LmServer.
2. VirtualBox, a cross-platform x86 and AMD64/Intel64 virtualization product. We have created a cluster in VirtualBox on an OSX laptop, with the front-end serving as LmServer and one node acting as LmCompute.

In future work, we intend to:

- Simplify data integration, allowing users to configure and populate their Lifemapper installation with local or remote data;
- Leverage cluster deployment at different sites using pragma_boot (see **Virtual Cluster Sharing**);
- Expand the provenance work with Indiana University, using metrics to estimate needed job resources for individual jobs and job queues;
- Test Lifemapper for specialized experiments by creating a Mt. Kinabalu-Lifemapper virtual infrastructure; and
- Use software-defined networks to create a VPN (aka Trust Envelope) that can connect virtual components at different sites into a global overlay researching species communities of the ultramafic regions of Mt. Kinabalu.

PARTICIPANTS: *U. Kansas Biodiversity Institute:* Aimee Stewart; *UCSD:* Nadya Williams; *IU:* Beth Plale, Yuan Luo, Quan (Gabriel) Zhou; *U Florida* and *NSF:* Reed Beaman



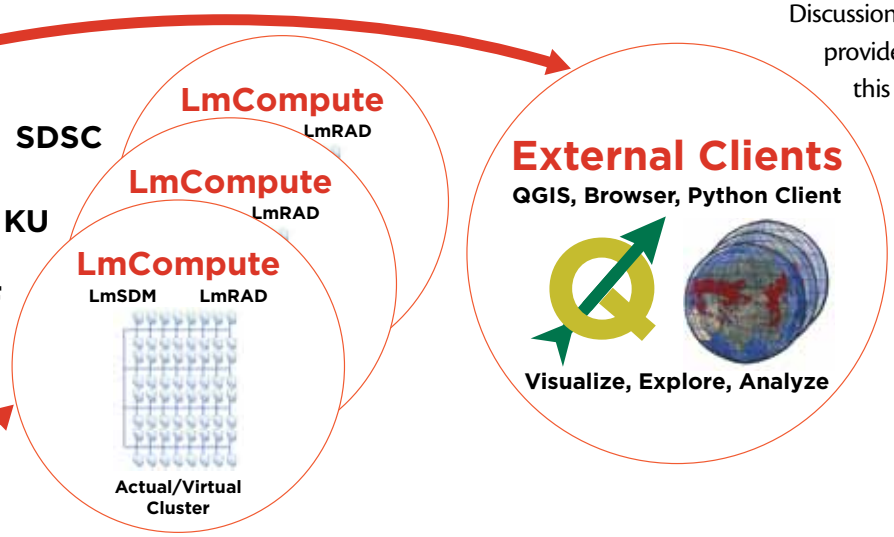
UAVs Provide Unique Data for Expedition

Figure 1: UAV flying near Donkey Ear—courtesy of Reed Beaman.

Ultramafic soils, deficient in the essential macronutrients and containing high concentrations of phytotoxic elements, pose a challenge for plants to adapt and survive. Paradoxically, these soils are also widely known to be highly species-rich and host high percentages of locally endemic plants and animals. Mount Kinabalu thus makes an excellent natural laboratory to address topics in evolution, phylogeography, ecology and physiology, including questions of adaptation, and biotic response to climate. While some ultramafic areas have been well studied, there remains a lack of knowledge about the plant distributions in Kinabalu Park, largely due to the difficulty and remoteness of the terrain, its inherent complexity as well as lack of taxonomic data.

Figure 2: Mt. Kinabalu, southern slope. The red lines indicate the area covered by the UAV flyovers. Source is Antony van der Ent, using Google Earth imagery.

Unmanned Aerial Vehicles (UAVs) can be used to capture sub-meter, ultra-high spatial resolution multispectral imagery. As an emerging technology for remote sensing, UAVs provide an unparalleled source of data for fine-scaled ecological studies, including vegetation classification and mapping. The specific characteristics of ultramafic vegetation—such as lower stature, smaller crown size, and markedly different species composition—makes it possible to map the extent using analysis of spectral response and texture. The south slope of Mt. Kinabalu (along the main summit trail) is of particular scientific interest because of the elevation gradient and concomitant changes in underlying geology, causing the occurrence of distinct vegetation types. This forms an ideal case study for UAV deployment aimed at mapping distributions of different vegetation types. In June 2014, prior to the International Conference on Serpentine Ecology (ICSE), expedition researchers conducted five days of initial UAV flights on the south and southwest slopes of Mt. Kinabalu (see Figure 2 for flyover area). The University of Queensland (UQ) provided two GPS-guided UAVs with the capacity to obtain very high spatial resolution (< 10 cm), four-band (RGB plus mid-infrared) multi-spectral imagery. The UAV is a small, electric-powered vehicle with a two-meter wingspan and it weighs approximately 4.9 kilograms (see Figure 1). It is controlled by a small ground station made up of a laptop and transmitter that allow for communication and control of the UAV. The team represented Sabah Parks, UQ, the University of Florida, and Skycam UAV in New Zealand. Over five days, more than 10,000 still images were captured, along with GoPro videos. The still images are being processed into a set of georectified image products and a point cloud using Pix4D software. Discussions are underway with UQ's Research Computing Centre to provide technical support and computational resources to speed this process.



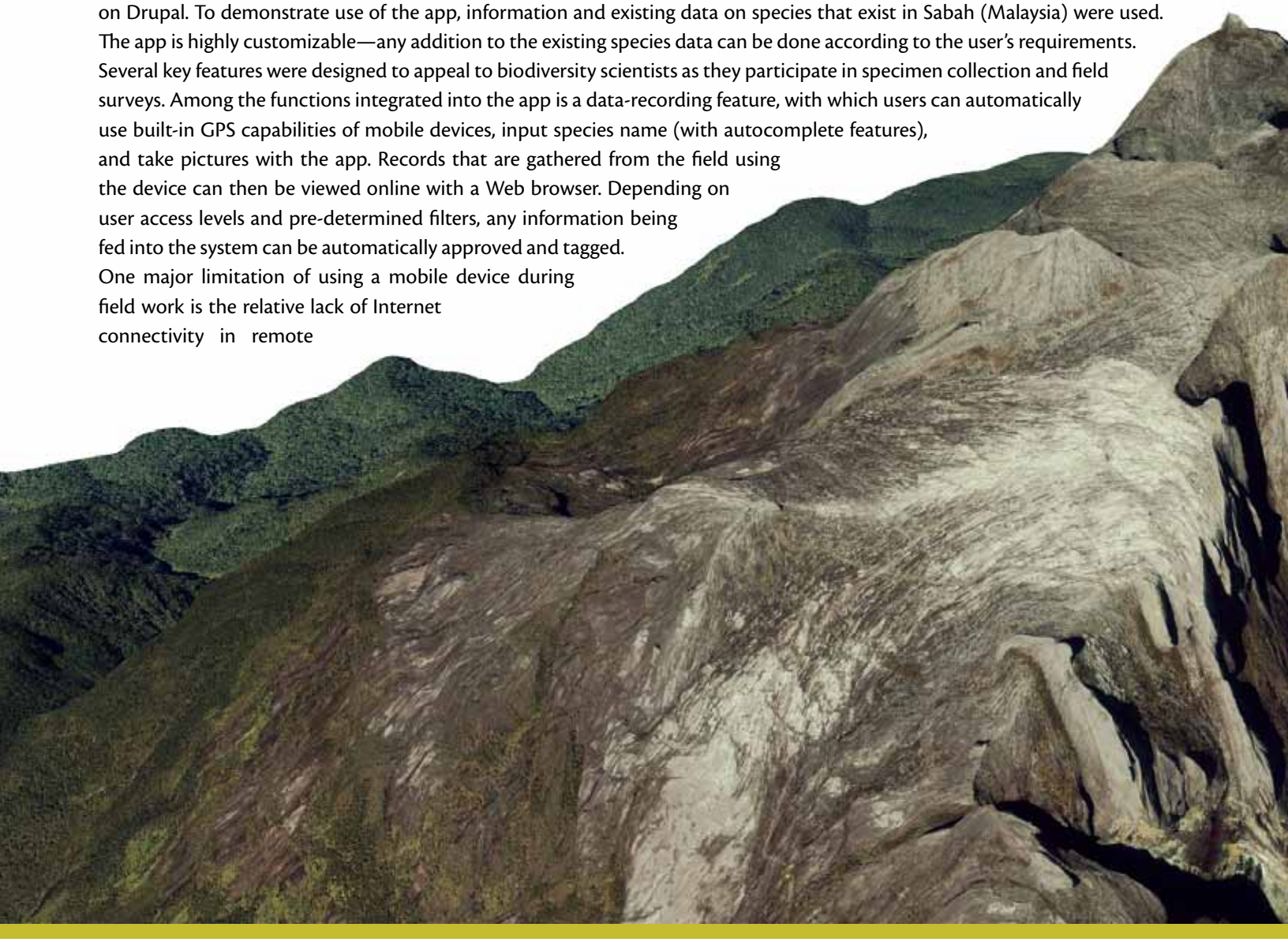
PARTICIPANTS: *Sabah Parks:* Jamili Nais, Rimi Repin, Martin Mogurin, Alim Biun, Sukaibin Sumail; *Skycam UAV:* Lew Woods; *U Queensland:* Antony van der Ent, Peter Erskine, Andrew Fletcher; *U Florida* and *NSF:* Reed Beaman

EcoSabah: On-the-Fly Data Recording for Species Occurrence in Biodiversity Field Work

Data preservation is one of the most challenging tasks in the field of biodiversity. A massive amount of heterogeneous information is gathered during field work—physical specimens, metadata, and other biological details. It is difficult to organize many different types of information, leading to missing links between specimens and ancillary data. Manual record-keeping is time-consuming and leads to additional errors. Pervasive computing, utilizing devices such as smartphones, may provide tools to aid in species identification, field data collection, and information dissemination. A digitization pipeline can assist biodiversity researchers in every step of data collection and archiving.

The Bioinformatics Research Group at the Universiti Teknologi Malaysia (BIRG UTM) has developed a prototype EcoSabah mobile application system to assist biodiversity researchers in these tasks. The central characteristics that guided the development of EcoSabah were that the system be low-cost, intuitive, and easy to obtain. Smartphones with open-source development platforms (e.g., Android) are particularly economical and available to many people.

The EcoSabah mobile Android app was developed as a proof-of-concept implementation of mobile technologies in the field of biodiversity. Mobile components were developed using the Phonegap framework, and web server components were built mainly on Drupal. To demonstrate use of the app, information and existing data on species that exist in Sabah (Malaysia) were used. The app is highly customizable—any addition to the existing species data can be done according to the user's requirements. Several key features were designed to appeal to biodiversity scientists as they participate in specimen collection and field surveys. Among the functions integrated into the app is a data-recording feature, with which users can automatically use built-in GPS capabilities of mobile devices, input species name (with autocomplete features), and take pictures with the app. Records that are gathered from the field using the device can then be viewed online with a Web browser. Depending on user access levels and pre-determined filters, any information being fed into the system can be automatically approved and tagged. One major limitation of using a mobile device during field work is the relative lack of Internet connectivity in remote

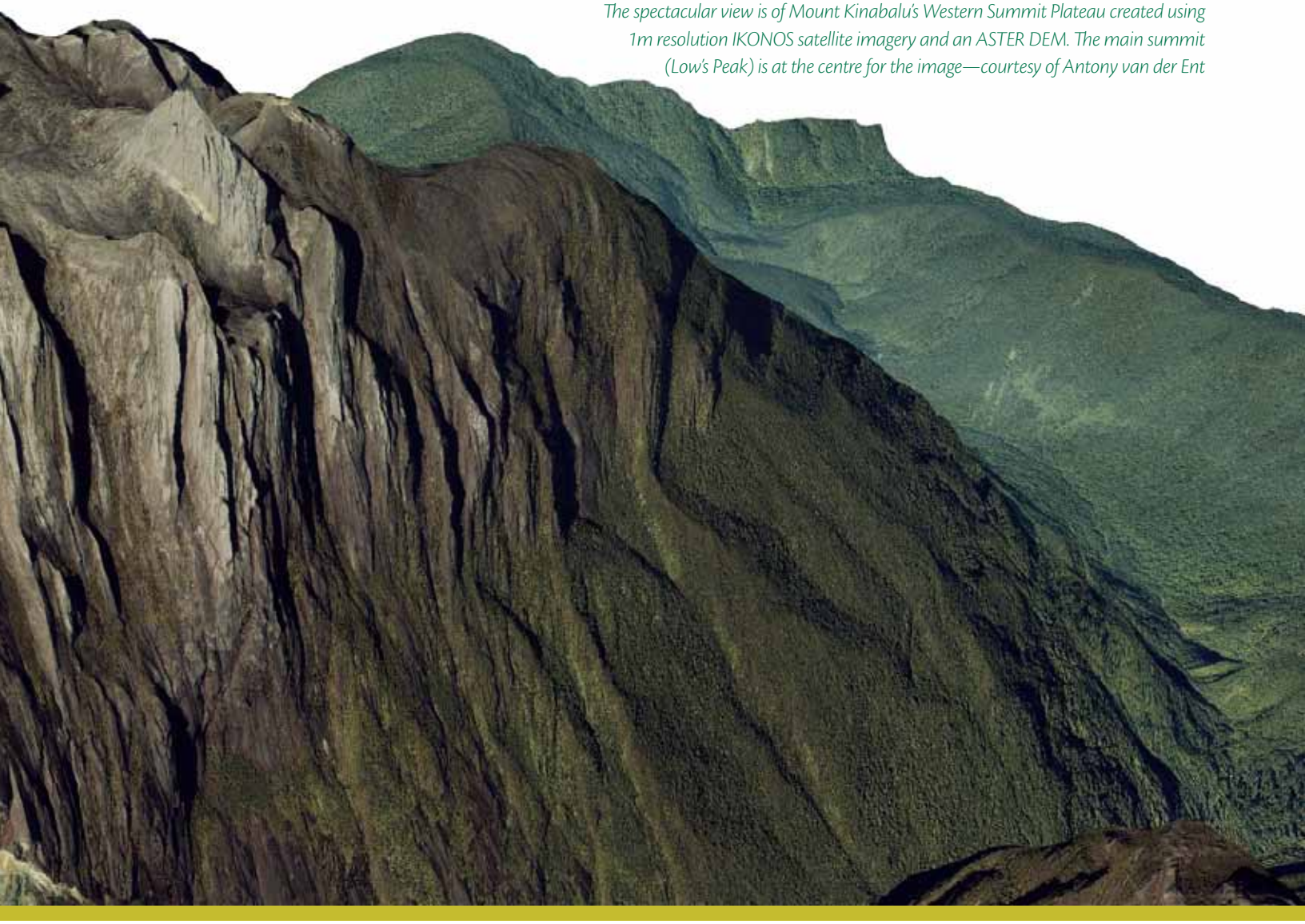


areas. To overcome the bandwidth problem, the app comes with an offline data-recording function that can be used without connecting to the Internet. All offline records are then synched to the server once sufficient connectivity is achieved.

Currently, the system is undergoing a major revamp to incorporate features to enable citizen science. Instead of creating yet another app or tool, the system will utilize existing platforms used by the public to encourage their participation in biodiversity conservation efforts. A data-mining feature is currently being set up to mine social media such as Twitter, where tweets with predetermined keywords (or hashtags) will be gathered and mapped by the system. Novice users or enthusiasts can now participate in biodiversity conservation simply by posting tweets with certain hashtags, while enabling location services on their device. In the future, we plan to expand the app's interoperability for multi-platform usage on iOS, Blackberry and other major mobile operating systems. We also plan to incorporate new features such as a function to recognize species, and even barcode reading for specimens.

PARTICIPANTS: *Bioinformatics Research Group, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia (UTM), Johor:* Nor Arlina Amirah Ahmad Ghani, Nurul Shakina Talkah, Shaiful Rahim, Iylia Zulkifli, Parveen Bal, Mohd Shahir Shamsir Omar

Note: Researchers from UTM were at the SEAIP Meeting (see ***Building Community****) in December 2013 where this component of the PRAGMA Biodiversity Expedition was first discussed. UTM researchers were already starting work on their mobile app, and the interaction at SEAIP encouraged the ongoing collaboration.*



The spectacular view is of Mount Kinabalu's Western Summit Plateau created using 1m resolution IKONOS satellite imagery and an ASTER DEM. The main summit (Low's Peak) is at the centre for the image—courtesy of Antony van der Ent

PRAGMA-GLEON EXPEDITION

Improving Our Ability to Predict Extreme Events

Harmful algal blooms are extreme events in lakes and reservoirs—extreme, both in terms of the health and safety risks from algal toxins, as well as the degradation of the aesthetics and general usability of the water. In the PRAGMA-GLEON expedition, we are just beginning to learn how this extreme ecosystem phenomenon is expressed in the observational data from new sensor networks and just how challenging reproducing those events in model simulations can be.

As part of the expedition, we are mining information from buoys deployed in lakes around the world to learn more about the factors that cause harmful algal blooms. Sensors with high sampling rates are able to detect the rapid transitions that occur when an algal bloom forms, and characterize their severity and duration within the context of the full distribution of phytoplankton concentrations through a season. The data reveal consistently low to moderate concentrations of phytoplankton, periodically interrupted by extreme events (Fig. 3). This "long-tailed" distribution is a key emergent property of lakes that defines a traditionally elusive lake feature—the probability of a bad surprise.

Understanding the processes that lead to surprise is a fundamental pursuit of the PRAGMA-GLEON expedition. Testing this understanding requires recreating the ecosystem feature in our numerical simulations. If we have the rules right and the model well-calibrated, then we recreate the surprise. So far, surprise in silico has been as elusive to produce as it once was to observe in the days before sensor networks (Fig. 4). To test thoroughly whether the current rules (i.e., model structure) might recreate surprises, we randomly sampled the trait-space of phytoplankton to see whether some combination of traits leads to a phytoplankton community that has explosive growth. The many parameters representative of biological traits can have a

GLEON'S NETWORK SCIENCE

The Global Lakes Ecological Observatory Network (GLEON, gleon.org) conducts innovative science by sharing and interpreting high-resolution sensor data to understand, predict and communicate the role and response of lakes in a changing global environment. GLEON is a network with more than 400 researchers worldwide, which includes more than 100 student members. The network runs a graduate student training program focusing on team science and the use of big data and cutting-edge analytical tools. GLEON receives funding from the National Science Foundation, and member sites have funding from local and regional sources.

Image: Falling Creek Reservoir, Virginia, part of the PRAGMA-GLEON lake expedition—courtesy of Cayelan Carey

IPOP: A VIRTUAL NETWORK ASSISTING SCIENTISTS

IP-over-P2P (IPOP) is a virtual network at the University of Florida's Advanced Computing and Information Systems Laboratory (ipop-project.org). IPOP is a self-configuring IP-over-P2P virtual network overlay which provides the capability for nodes behind Network Address Translators (NATs) and some firewalls to appear all in the same network space. IPOP is transparent to applications and can be deployed on a variety of operating systems, thus seamlessly integrating with unmodified, off-the-shelf software. It is supported by funding from the NSF, in part from Grant no. 1339737.

range of values, the expression of which differ under varying lake physical-chemical conditions (e.g., lake size, thermal dynamics, or nutrient loads). As a result, tens of thousands of simulations must be run to sample adequately the space of possibilities.

Relying on a single computer to run these simulations has proven to be inadequate. Distributed computing systems that support concurrent execution of large numbers of jobs exist—in particular, the HTCondor project (research.cs.wisc.edu/htcondor). However, there are barriers to using such systems effectively; in particular, the lack of computational capacity available at a researcher's institution, and the difficulty in ensuring that simulations are properly ported and configured to execute at typical computing facilities. For instance, researchers are accustomed to running their simulations interactively on their desktop using the Windows OS and the R environment for statistical computing, whereas shared computing facilities typically provide a Linux-based environment for batch-job submissions.

In this expedition, we apply technologies developed by PRAGMA partners to tackle the problem of distributed resources in an innovative way—by enabling the community to pool together resources across institutional boundaries in a straightforward manner, and by allowing researchers to continue to use a computer environment they are used to.

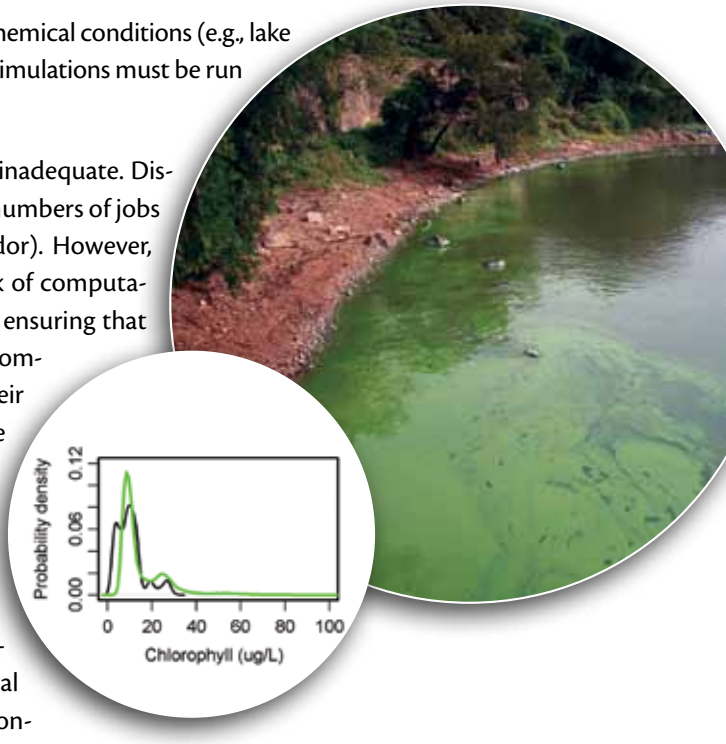


Figure 3. The real ecosystem (right) has extreme events, such as the bloom shown in the photograph. In the distribution of values, these are labeled in the 'long tail' of the distribution as high blooms. The modeled system is much more evenly distributed.

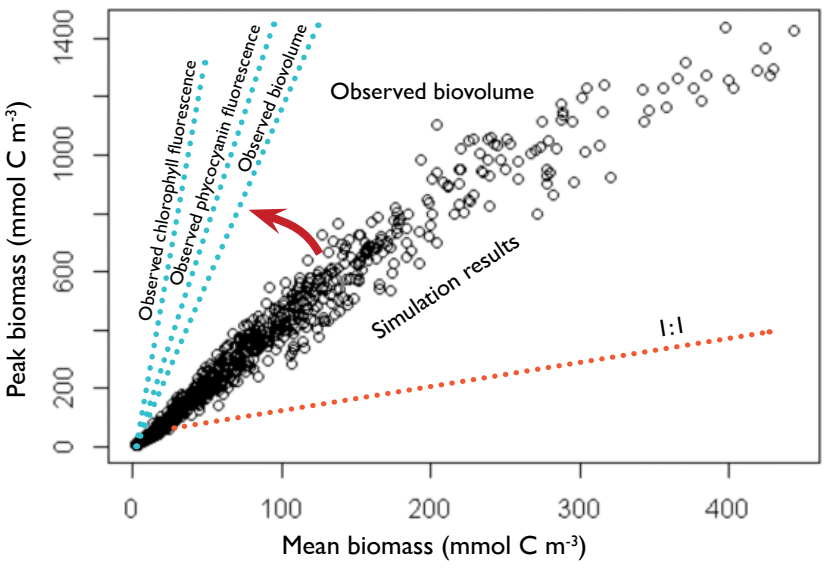


Figure 4. The lake simulations (circles) can never predict both the mean and extreme values in the observed data (dashed lines). To do so would require that the simulations reproduce more extreme values for any mean biomass concentration, effectively rotating the results along the trajectory of the red arrow.

Figure 5. The team. Clockwise from top: Paul Hanson, Cayelan Carey, Ken Subratie, Youna Jung, Craig Snortheim, and Renato Figueiredo.

The approach combines two open-source software systems: the IP-over-P2P (IPOP) virtual network project led by the University of Florida, and the HTCondor job scheduler project from the University of Wisconsin. The IPOP overlay software allows collaborators at different institutions to join a virtual private cluster of computing resources, while not requiring significant IT expertise or local support. On the other hand, the HTCondor scheduling software allows simulations to harvest the shared resources brought together in the virtual private cluster. On the job submission side, researchers can initiate simulation batches from their own local desktops—continuing to use a computer and software environment to which they are accustomed. On the job execution side, each computing resource that joins the overlay increases its aggregate computing capacity—including opportunistic, idle cycles from end-user desktops, as well as dedicated cycles from servers deployed within the participating institutions, or in pay-as-you-go cloud computing services. The expedition team has deployed a prototype overlay currently with 24 processor cores deployed at U. Florida, U. Wisconsin, and Virginia Tech, with plans to expand the computational capacity from resources across PRAGMA and GLEON sites.

Improving our predictions of harmful algal blooms requires a team of scientists (Fig. 5)—modelers, computer scientists, and phytoplankton ecologists—working together to apply diverse knowledge and skills to solve the complex technical and intellectual problems. While there remains much to learn, the nature of the prediction problem has been isolated. While the high productivity and algal biomass of eutrophic lakes and reservoirs undoubtedly relates to the nutrient loads from the watershed, the frequency and timing of harmful blooms have underlying biotic interactions not represented in our models today. As we continue the search for solutions, our reliance on virtual networks and distributed computing will grow, and the expanded human and technological resources will lead to exciting new discoveries.

PARTICIPANTS: *U Wisconsin:* Paul Hanson, Craig Snortheim; *Virginia Tech:* Cayelan Carey, Jonathan Doubek; *U Florida:* Renato Figueiredo, Ken Subratie, Youna Jung

PRAGMA EXPERIMENTAL NETWORK TESTBED (PRAGMA-ENT)

The PRAGMA Experimental Network Testbed (PRAGMA-ENT) expedition was established at PRAGMA 25 in October 2013, hosted by CNIC in Beijing. Its goal: to construct a breakable, international software-defined network (SDN) testbed for use by PRAGMA researchers and collaborators. PRAGMA-ENT is breakable in the sense that it offers complete freedom for researchers to access network resources to develop, experiment, and evaluate new ideas without being concerned about interfering with a production network. PRAGMA-ENT is architected to offer the necessary networking support to the PRAGMA multi-cloud and user-defined trust envelopes.

The PRAGMA-ENT team worked with Internet2, Florida Lambda Rail (FLR), National Institute of Information and Communications Technology (NICT), and Pacific Wave engineers to establish an international Layer-2 (reliable direct point-to-point data connection) backbone. Virtual Local Area Networks (VLANs) were allocated to create logical direct paths between (1) the University of Florida (UF) and the University of California, San Diego (UCSD); (2) UF and Nara Institute of Science and Technology (NAIST) in Japan; and (3) between UCSD and NAIST. Both paths between the U.S. and Japan will be expanded on the Japanese end to reach the National Institute of Advanced Industrial Science and Technology (AIST) and Osaka University. VLANs and associated technologies were used to create seamless connections between OpenFlow-enabled switches, deployed at participating sites. Initial testing of the network achieved 1 Gigabit-per-second (Gbps) throughput between all established direct paths, i.e., UF/UCSD, UF/NAIST, and UCSD/NAIST.

Since all OpenFlow switches are interconnected, independent of the geographical location, it is possible to develop SDN controllers to manage trust envelopes. PRAGMA-ENT will provide a virtual network slice for each application, user, and/or project in order to enable independent and isolated development and evaluation of software-defined network functions. We are currently implementing AutoVFlow to create virtual network slices. AutoVFlow is developed by Yamanaka et. al. at NICT to address shortcomings of existing SDN slicing approaches such as FlowVisor[1]. Preliminary application ideas include: deployment of global-scale virtual clusters; evaluation of overlay network technologies; optimal routing control; multipath routing control; and a network resource-aware job management system. In particular, future plans include using ViNe overlays to expand ENT to sites without direct connection to the backbone (e.g., commercial clouds, such as Azure). As noted in the section on **Virtual Cluster Sharing** (next section), there have been successful experiments where a Rocks cluster front-end at UCSD was able to provision a Rocks node at UF through ENT. Finally NAIST is currently working on experiments with SDN multi-path routing.

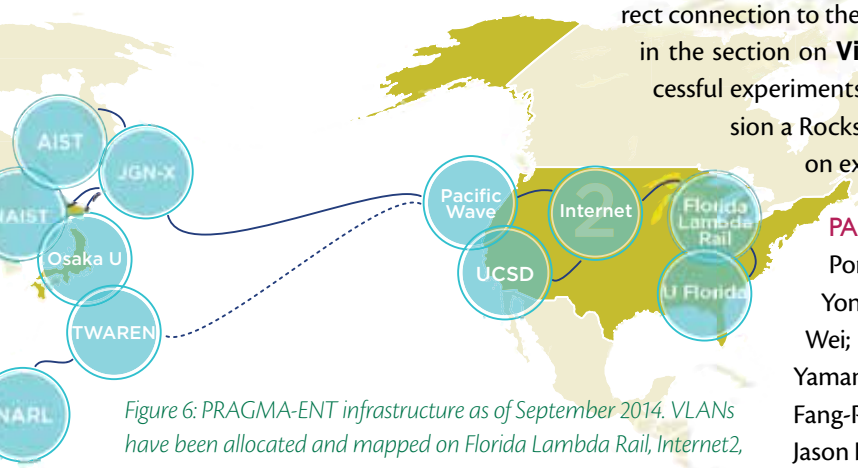
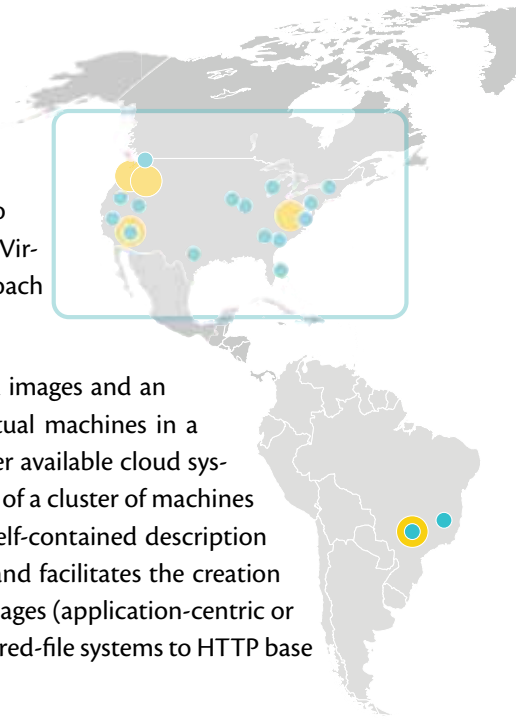


Figure 6: PRAGMA-ENT infrastructure as of September 2014. VLANs have been allocated and mapped on Florida Lambda Rail, Internet2, JGN-X and Pacific Wave. The following mappings are active: UF to UCSD; UF to JGNX; and UCSD to JGNX.

PARTICIPANTS: *UF:* Maurício Tsugawa; *NAIST:* Kohei Ichikawa, Pongsakorn U-chupala, Chawanat Nakasan; *CNIC:* Kevin Dong, Yongmao Ren; *IU:* James Williams, Jennifer Schopf; *JLU:* Xiaohui Wei; *Kasetsart U:* Putchong Uthayopas; *NICT:* Jin Tanaka, Hiroaki Yamanaka; *NARLabs:* Li-Chi Ku, JenWei Hu, TeLung Liu; *NARL/NCHC:* Fang-Pang Lin; *AIST:* Atsuko Takefusa, Yoshio Tanaka; *AIST* and *UCSD:* Jason Haga; *Osaka U:* Shinji Shimojo, Susumu Date, Yasuhiro Watashiba, Yoshiyuki Kido; *UCSD:* Luca Clementi, Philip Papadopoulos

[1]Hiroaki Yamanaka, Eiji Kawai, Shuji Ishii and Shinji Shimojo, "AutoVFlow: Autonomous Virtualization for Wide-area OpenFlow Networks," The Third European Workshop on Software Defined Networks (EWSDN 2014), Sep. 2014.

VIRTUAL CLUSTER SHARING



Scientific applications and the software platforms that enable the application often need to be migrated to or shared with others, (e.g., moving the application to data proximity, or to access additional capability). To enable sharing of software platforms and to run computational experiments across different sites, PRAGMA members have defined a common way to publish virtual clusters. This activity has led to the development of the Virtual Cluster Sharing technology and the creation of a standard approach to authoring PRAGMA Virtual Clusters.

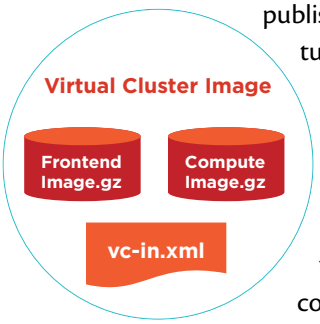


Figure 7: Virtual Cluster Image Representation

A PRAGMA Virtual Cluster is composed of virtual-machine disk images and an XML file (vc-in.xml) describing how to connect the booted virtual machines in a computing cluster (Figure 7). This approach is different from other available cloud systems because its smallest deployment unit is based on the concept of a cluster of machines connected together and not just on a single virtual machine. This self-contained description mechanism simplifies the creation of computational infrastructure and facilitates the creation of cluster appliance libraries. Libraries can contain different cluster images (application-centric or general) and can be hosted on various storage systems (from local shared-file systems to HTTP base websites).

Before a virtual machine is booted, the hosting environment places a vc-out.xml file in its root file system. The file contains all the needed information to configure the network of each host in the cluster.

PRAGMA members have developed several software tools in order to verify the soundness of the design and to use the Virtual Cluster Sharing for real simulations:

- PRAGMA Bootstrap is an extensible framework that can bootstrap PRAGMA Virtual Cluster on different cloud infrastructures. All the project source code and detailed documentation of the vc-in.xml and the vc-out.xml file format is available at github.com/pragmagrid/pragma_boot. Currently, PRAGMA Bootstrap supports booting Virtual Clusters on physical infrastructure with Rocks Cluster 6.1 distribution and with Open Nebula;
- DynIP is a software application which can be used to enable a generic Rocks Cluster to be packaged and executed as a PRAGMA Virtual Cluster. (github.com/rocksclusters/dynip). DynIP can parse the vc-out.xml produced by the PRAGMA Bootstrap and reconfigure a Rocks Cluster accordingly, achieving a seamless integration with the Virtual Cluster sharing environment; and
- VC Out Parser is an application which can enable a generic Red Hat Enterprise Linux and derivatives (e.g., Centos, Scientific Linux) to

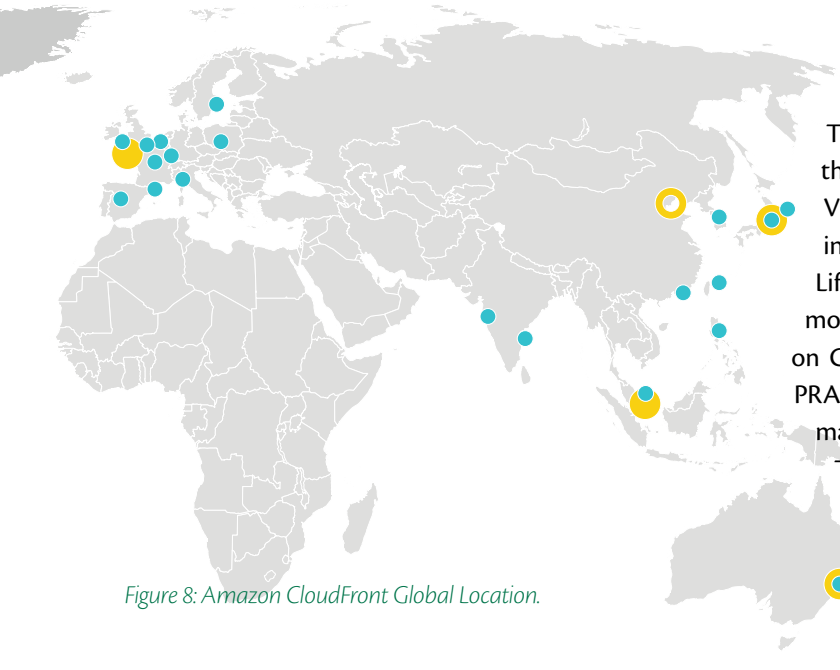


Figure 8: Amazon CloudFront Global Location.

properly bootstrap and configure the network inside a PRAGMA Virtual Cluster (github.com/pragmagrid/vc-out-parser).

These software tools were presented at PRAGMA 25 during the demo session. It involved the deployment of a Rocks-based Virtual Cluster running Lifemapper on AIST Open Nebula cloud infrastructure (see the **Biodiversity Expedition** for more on Lifemapper). At the same meeting, a different application of molecular docking and virtual screening, a docking cluster based on Centos, was deployed on UCSD Rocks hosting infrastructure. PRAGMA Bootstrap, VC Out Parser and DynIP were all used to manage the deployment of the VM both at UCSD and at AIST. The docking cluster deployed at UCSD was also used in the months following PRAGMA 25 by PRIME undergraduates to run docking simulations.

In 2014, NAIST's Pongsakorn U-chupala spent several weeks at UCSD. During his visit he enhanced PRAGMA Bootstrap to support downloading Virtual Cluster Images from Amazon CloudFront and Amazon S3. This functionality was demonstrated at PRAGMA 26, where a sustained download bandwidth of more than 50MB/s both from Japan and from the U.S. was achieved (Figure 8).

Via PRAGMA-ENT (see previous section), a virtual cluster running on different sites (e.g., 3 nodes at UCSD, 2 nodes at NAIST, and 2 nodes at UF) was instantiated transparently.

PARTICIPANTS: *Virtual Cluster Sharing:* *UCSD:* Luca Clementi, Philip Papadopoulos; *NAIST:* Pongsakorn U-chupala; *Lifemapper Application:* *UCSD:* Nadya Williams; *U Kansas:* Aimee Stewart; *Docking Application:* *AIST* and *UCSD:* Jason Haga; *NAIST:* Kohei Ichikawa; *UCSD PRIME 2013 students at NAIST:* Kevin Lam, Karen Rodriguez

Images: (left to right) Osamu Tatebe (U Tsukuba) at PRAGMA 26; Tainan Confucian Temple, PRAGMA 26; Philip Papadopoulos (UCSD), Peter Arzberger (UCSD), Ce-Kuen Shieh (NARL/NCHC) at PRAGMA 26; Kirkwood Observatory, IU , PRAGMA 27

VIRTUAL CLUSTER MANAGEMENT ON MULTI-CLOUD PRAGMA

To provide users with an easy-to-use interface for managing the life cycle of virtual clusters—from startup to status monitoring and shutdown—a higher-level control architecture activity was started in 2014. This activity led to the design of a Personal Cloud Controller (PCC) tool, a lightweight management tool that integrates various PRAGMA tools including PRAGMA Bootstrap and IPOP with a well-known resource management tool called HTCondor. An illustration of this is shown in Figure 9, where a PCC-enhanced HTCondor (PCC-HTCondor) is used to allocate and manage two virtual clusters on physical cluster resources from four PRAGMA institutions: UCSD, IU, NCHC, and AIST. Scientists can launch and view the status of their virtual clusters using a Web interface or directly with the Condor command-line tools. The Web interface provides a listing of available virtual-cluster images and it allows the user to select the number of cores for a new virtual cluster from the available resources. The request is then forwarded to PCC-HTCondor to invoke PRAGMA Bootstrap to create the virtual cluster on one or more clusters via Rocks or OpenNebula. From there, it sets up virtual networking (i.e., iPOP, ViNE, or OpenFlow) to connect the nodes together as a single cluster.

An initial prototype implementation of PCC was created by Ph.D. students Yuan Luo (Indiana University) and Shava Smallen (UCSD) and it was presented as a demo and poster during PRAGMA 26 in April 2014. Luo developed the PCC-HTCondor component that invoked PRAGMA Bootstrap to create a couple of virtual clusters on a single small Rocks cluster at UCSD, while Smallen created the Web interface front-end. PCC- HTCondor started and controlled the virtual clusters from its Virtual Machine universe using extended resource specifications that were advertised and matched in HTCondor's ClassAd mechanism. A custom Grid Ascii Helper Protocol (GAHP) module then invoked PRAGMA Bootstrap.

For PRAGMA 27, PCC was enhanced to perform cross-institute virtual cluster management, as shown in Figure 9. It utilizes multiple physical clusters at difference institutes and links them together with a virtual network overlay built with IPOP, an open-source, user-centric software virtual network allowing end users to define and create their own virtual private networks (VPNs). As part of this work, a new IPOP Rocks roll was developed by Nadya Williams (UCSD). It automates the installation and configuration of IPOP and its software dependencies (ejabberd, erlang, and so on).

Future work includes further automation, additional SDN support (i.e., ViNE and OpenFlow), and a workload assessment tool that would enable the LifeMapper project to supplement their cluster resources with additional virtual-resources that grow and shrink based on historical information on workload rates and current demand.

PARTICIPANTS: *IU:* Yuan Luo, Beth Plale; *UCSD:* Shava Smallen, Nadya Williams, Philip Papadopoulos; *AIST:* Atsuko Takefusa; *NCHC:* Serena Pan

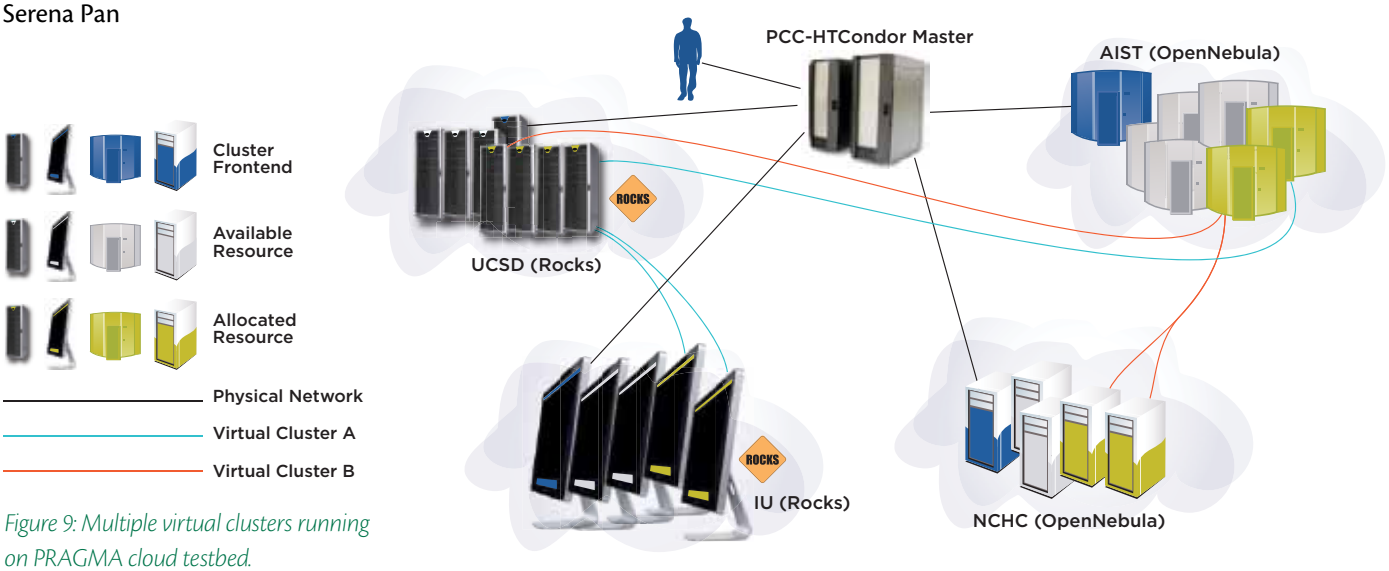


Figure 9: Multiple virtual clusters running on PRAGMA cloud testbed.

EDISON PAVES THE WAY FOR CYBERLEARNING IN PRAGMA

Simulation-based cyberlearning is a new opportunity for the next generation to change education and research paradigms. It provides new approaches for gaining knowledge, especially in advanced computational science and engineering applications. Simulation-based cyberlearning environments are connected with the Internet and offer a learners' paradise: the ability to learn anywhere and at any time.

PRAGMA began exploring and then development of a new working group on cyberlearning. Preliminary meetings were held at PRAGMA 24 and PRAGMA 25. In advance of the PRAGMA 26 workshop in Tainan, there was a focused workshop for researchers from Korea and Taiwan, looking at fields where cyberlearning would have a major impact. The same workshop also explored how to expand the user base for cyberlearning by identifying research themes for simulation-based cyberlearning that would lead to joint research.

A cornerstone of PRAGMA's efforts on cyberlearning is "Education-research Integration through Simulation On the Net" (EDISON), which has been developed over several years by the Korea Institute for Science and Technology Information. Through the EDISON system (www.edison.re.kr), anyone can learn and understand a theory or system through computer simulation anywhere. They can also predict operations of the system through easily changing parameters of the simulation model. Currently EDISON supports simulation activities in five areas of computational science and engineering: Computational Fluid Dynamics, NanoPhysics, Computational Chemistry, Computational Structural Dynamics, and Multi-disciplinary Optimization. In the past year, new features have been added to EDISON, including active Web portals that serve three communities: Computational Fluid Dynamics, NanoPhysics, and Computational Chemistry. The EDISON team has also added content development and simulation software deployment. The net result: EDISON now reaches more than 150 universities and nearly 20,000 students in Korea alone.

In November 2013, the team behind EDISON received the HPC Innovation Excellence Award from the International Data Corporation (IDC). This was the first time the award was presented to a Korean institution.

The EDISON team also looked to build collaborations with PRAGMA (notably with new PRAGMA member, Vietnam's Institute of Computational Science and Technology, based in Ho Chi Minh City) as well as with research institutes and universities throughout the world.

For more information about specific portals, see Computational Fluid Dynamics (cfd.edison-project.org), NanoPhysics (nano.edison-project.org) and/or Computational Chemistry (chem.edison-project.org). In addition to the portals, there are other components to the system. Science Appstore technology manages the metadata of analysis tools such as simulation programs and content. Scientific Workflow implementation technology and Simulation Management allow users to operate and manage simulation jobs by using large-scale, heterogeneous computational resources (e.g., supercomputers, cloud computing, etc.).

PARTICIPANTS: *Leadership: KISTI:* Ruth Lee (lead); *NCHC:* Hsi-ching Lin (co-lead); *Kasetsart U:* Putchong Uthayopas (co-lead)

Image: Kamakura, Japan, in Hachiman Shrine—courtesy of Peter Arzberger.


PRAG
MAT
STUDENTS
PRIDE
MUSEUM
Triton Day 2012

STUDENTS

ENGAGING CURRENT AND FUTURE GENERATIONS OF RESEARCHERS

PRAGMA is committed to engaging new communities of researchers. We focus on two broad efforts. The first is to engage students in PRAGMA activities; the second is to work with communities of researchers to broaden the impact of the work we do, and to leverage their interests. Both of these activities enrich PRAGMA by infusing ideas, questions, and energy into our workshops, and into activities between workshops.

In this section we will talk about three specific activities to engage students: PRAGMA Students, PRIME, and MURPA/QURPA.

PRAGMA STUDENTS

Stimulating International Cross-Disciplinary Collaboration among Young Researchers

PRAGMA Students, formed in 2012, aims to help students gain opportunities for professional experiences with PRAGMA's trusted social and technical networks. A student organization inside PRAGMA, the group is led by a student committee and advised by senior PRAGMA researchers. To date, activities of PRAGMA Students have included: organizing PRAGMA-affiliated student workshop and poster sessions as part of the biannual PRAGMA Workshops; hosting online seminars with lecturers drawn from the broad PRAGMA community; and developing a unique model to provide multiple opportunities for students to participate in PRAGMA's collaborative scientific research.

PRAGMA provides a trusted people network and opportunities in leadership that helps students gain valuable professional experience. There are several advantages to joining PRAGMA Students and the PRAGMA community. These opportunities include:

- Participating in professional workshops and conferences in an international setting, which offer information sharing, inspiration and collaboration;
- Strengthening both research and careers through collaborations and engagement with mentors and advisors, both domestic and international;
- Fostering cross-disciplinary collaborations which stimulate research and enhance people networking;
- Gaining collaboration experiences through short-term residential research opportunities at other PRAGMA sites.

At the April 2014 PRAGMA 26 workshop in Tainan, Taiwan, PRAGMA Students organized a student workshop. One of the leadership opportunities at PRAGMA 26 included the ability for students to organize poster sessions. Members of PRAGMA Students reviewed and selected posters from all workshop participants, and in many cases presented posters. More than 20 posters were carefully selected and then presented.

To showcase more student research, PRAGMA Students also hosted a Lightning Talks session during the workshop. They invited seven students with accepted posters to give short talks about their respective posters. During the Lightning session, an active Twitter stream allowed the PRAGMA community to share comments and questions in real time.

In addition to the Lightning and poster sessions, PRAGMA Students hosted a panel discussion to enhance engagement between students and senior researchers across the PRAGMA community. Students also engaged in discussion about a wide array of topics associated with mentors and advisors, offering creative perspectives about ways to improve the PRAGMA community.

At PRAGMA 27, the student organization mounted an Expedition Programming Challenge, organized by Quan (Gabriel) Zhou. The programming challenge was an opportunity for both undergraduate and graduate students (local and visiting) to participate in PRAGMA Expedition challenges. This two-and-a-half-day activity allowed students to learn some of the important computing-related problems faced by researchers in two PRAGMA Expeditions, giving them a chance to help solve specific problems posed by the expeditions. It was a great opportunity to enhance cross-disciplinary collaboration among PRAGMA members, and to strengthen research by addressing interesting and carefully selected programming challenges. In preparation for this activity, PRAGMA Students co-chair Meilan Jiang reached out to members of the Student Association of the Global Lake Ecological Observatory Network (GLEON), in particular Kohji Muraoka (U Waikato), ahead of the October 2014 PRAGMA 27, to organize jointly a hackathon sponsored by students in PRAGMA and GLEON (blog.gleon.org/pragma-gleon-expedition-2014). At the end of PRAGMA 27, the expeditions learned from the outcome of these challenges: in one case how to improve creation of the VirtualBox virtualization of Lifemapper; in the other how to analyze data from a GLEON lake.

In addition to the opportunities to help plan and execute activities at PRAGMA workshops, the student organization also hosted a distributed seminar on February 24-25, 2014. Professor Jose Fortes presented his ideas in a talk titled, "Towards Distributed Software Defined Systems." The Fortes seminar was attended by more than 20 students, and can be viewed online at connect.iu.edu/p5apeig31gn.

Two PRAGMA students this year were provided short-term internship opportunities in Calit2's Qualcomm Institute on the UCSD campus. Pongsakorn U-chupala is a graduate student from the Nara Institute of Science and Technology (NAIST), and he contributed to the Virtual Clusters Sharing project. Chawanat Nakasan (also a grad student at NAIST) became a visiting scholar at UC San Diego. The internship program serves as an example of collaboration between PRAGMA institutions and strengthening research by short-term internships. (Both of the students from NAIST in Japan did their undergraduate degrees in Thailand at Kasetsart University, another member of PRAGMA.)

After PRAGMA 26, the PRAGMA Students Association Committee welcomed NAIST's Pongsakorn U-chupala as a new member of the committee. He remains an active member of the PRAGMA community and will work on behalf of PRAGMA Students' activities. At the same time, Yuan Luo—who helped launch PRAGMA Students as an organization—stepped down from the committee. He rotated out of PRAGMA Student leadership, but fortunately for the association, he will remain active in PRAGMA.

CURRENT AND FORMER LEADERS OF PRAGMA STUDENTS: *IU:* Quan (Gabriel) Zhou (co-chair); Yuan Luo (former co-chair emeritus); *Konkuk U:* Meilan Jiang (co-chair); *NAIST:* Pongsakorn U-chupala (co-chair)

PRAGMA STUDENT ADVISORS: *IU:* Beth Plale; *Konkuk U:* Karpjoo Jeong; *Kasetsart U:* Putchong Uthayopas

Images: (top) Anthony Nguyen, Katy Pham, and Derek Song in Banpaku-koen Park—courtesy of Anthony Nguyen; (left) Meilan Jiang (Konkuk U); (right) Tainan Skyline—courtesy of Garrett Chan

2014 HOST SITES

Nara Institute of Science and Technology (NAIST)

National Institute for Information and Communication Technology (NICT)

National Center for High-performance Computing (NCHC)

National Taiwan University (NTU)

University of Queensland (UQ)

PRIME

The Pacific Rim Experiences for Undergraduates (PRIME) program was created in 2004 to provide a project-based, hands-on research internship program combined with a cultural awareness experience for science and engineering undergraduates at UC San Diego. PRIME grew out of the PRAGMA collaborative framework and people network. PRIME's projects are based on PRAGMA collaborations, as well as additional collaborations between UCSD and PRAGMA researchers.

To date, more than 190 students have participated in PRIME. This year, the 11th of the program, PRIME sent 10 students to the five sites listed at right. Projects this year helped develop workflows for neuroimaging and protein motion, virtualization approaches to docking chemical compounds to proteins, and interactive control of displays used by museums.

PRIME added other activities for students. In Fall 2013, PRIME hosted a seminar allowing PRIME 2014 students to present the results of their work to prospective PRIME students. We added this feature to provide students with a fuller experience of the research process—from the lab bench to the presentation.

We are also pleased to note that a PRIME 2013 student, Jesus Rios, continued to develop his project upon return from Japan. He created an app that is used in San Diego's Balboa Park, and it won the Honeysett & Din Student Award (part of the American Alliance of Museums MUSE awards) on May 18, 2014. The citation for the award to Rios' Japanese Friendship Garden (JFG) Haiku Hunt app reads:

> *Highly interactive, infusing the existing visitor experience with playfulness, participation and freshness, this project combines education and entertainment in an exciting way. The app leverages the geo-location functionality of a mobile device, traditional Japanese haiku and a scavenger hunt to encourage learning and discovery.*

For more information, see cse.ucsd.edu/node/2548.

The PRIME program continues to benefit from a generous gift from a PRIME alumna, Haley Hunter-Zinck (PRIME 2008, Monash University). She recently received her doctorate from Cornell University and is now a post-doc at the Federal University of Rio Grande do Sul in Porto Alegre, Brazil working in the area of medical informatics. Her unrestricted gift to PRIME is a tribute to her mother, who was passionate about promoting the success of women in sci-

ence. Hunter-Zinck hopes that her donation to PRIME will inspire others to follow her lead in supporting the international program. Her gift is gratefully accepted for future students, and allows the program to continue evolving and improving the PRIME experience. We also acknowledge the support of the U.S. National Science Foundation, and of all of our host sites, and in particular, support from the Japan Society for the Promotion of Science via NAIST, NICT, and the Balboa Park Online Collaborative. For more information about PRIME, visit the site at prime.ucsd.edu.

PARTICIPANTS: *UCSD PRIME Leads:* Gabriele Wienhausen, Teri Simas, James Galvin, Jason Haga; *UCSD Mentors:* Ilkay Altintas, Rommie Amaro, Jason Haga *(also AIST)*, Andrew McCulloch, Jurgen Schulze, Philip Papadopoulos; *Host Mentor Leads: NAIST:* Kohei Ichikawa; *NICT:* Shinji Shimojo; *NARL/NCHC:* Fang-Pang Lin; *NTU:* Jung-Hsin Lin; *UQ:* David Abramson

We wish to acknowledge the contribution and dedication of our colleague, Anushka Michailova, who died tragically in May 2014. She was committed to ensuring students performed their best, and she was enthusiastic about helping to bridge technology and modeling with deep biophysics. For more on an award in her memory, see jacobsschoolofengineering.blogspot.com/2014/06/anushka-michailova-memorial-award.html.

Background Image: Wulai, New Taipei—courtesy of Garrett Chan.

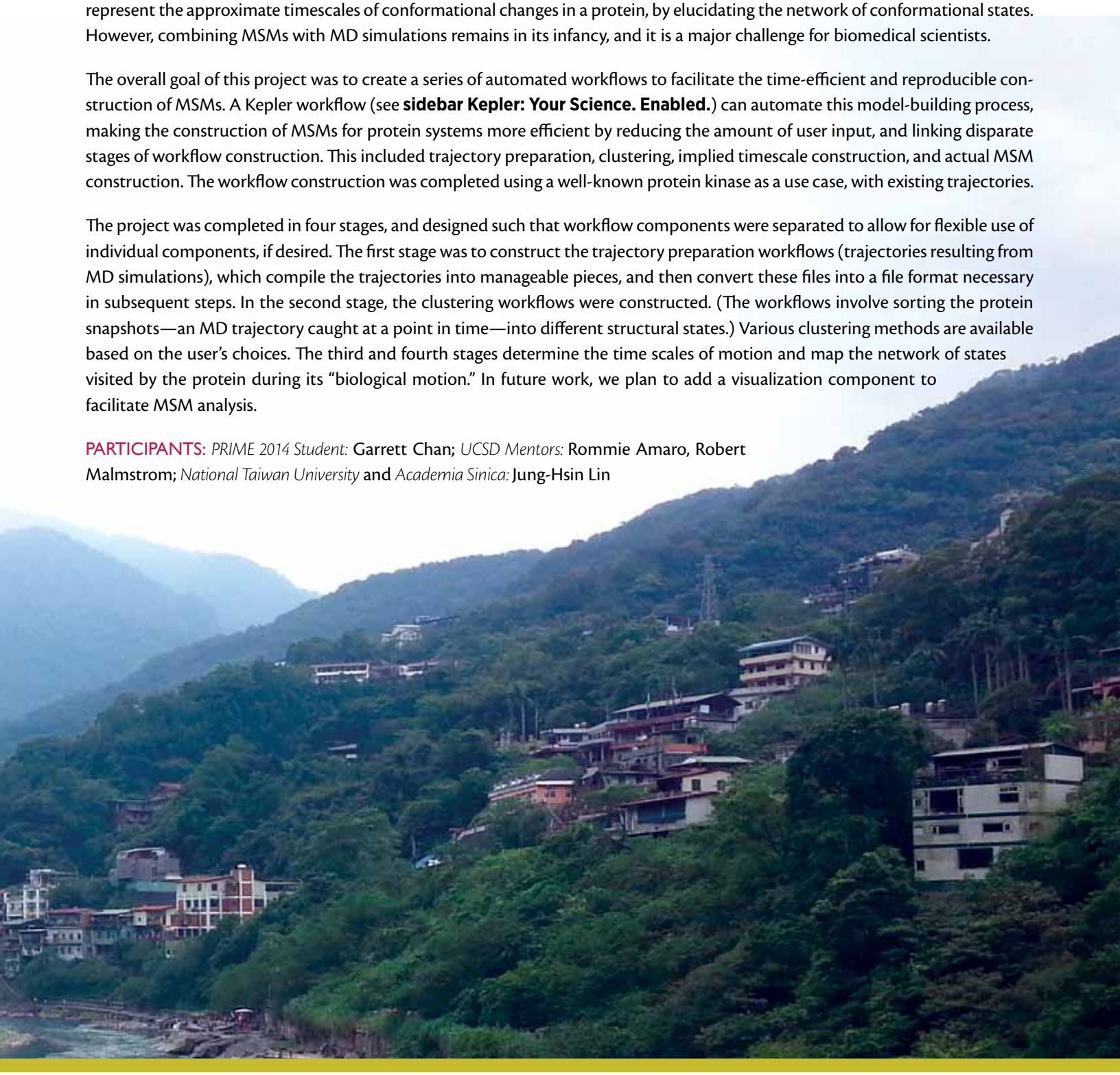
Markov State Models, Kepler Workflows and Understanding a Protein's Biological Motion

Understanding how proteins move and change conformations—particularly over long time scales—is a key challenge in biology. Moreover, protein shape and motion are critical components of its function. One way to determine their rates of motion is through the use of molecular dynamics (MD) simulations and Markov state models (MSMs). MSMs are probabilistic models that represent the approximate timescales of conformational changes in a protein, by elucidating the network of conformational states. However, combining MSMs with MD simulations remains in its infancy, and it is a major challenge for biomedical scientists.

The overall goal of this project was to create a series of automated workflows to facilitate the time-efficient and reproducible construction of MSMs. A Kepler workflow (see **sidebar Kepler: Your Science. Enabled.**) can automate this model-building process, making the construction of MSMs for protein systems more efficient by reducing the amount of user input, and linking disparate stages of workflow construction. This included trajectory preparation, clustering, implied timescale construction, and actual MSM construction. The workflow construction was completed using a well-known protein kinase as a use case, with existing trajectories.

The project was completed in four stages, and designed such that workflow components were separated to allow for flexible use of individual components, if desired. The first stage was to construct the trajectory preparation workflows (trajectories resulting from MD simulations), which compile the trajectories into manageable pieces, and then convert these files into a file format necessary in subsequent steps. In the second stage, the clustering workflows were constructed. (The workflows involve sorting the protein snapshots—an MD trajectory caught at a point in time—into different structural states.) Various clustering methods are available based on the user's choices. The third and fourth stages determine the time scales of motion and map the network of states visited by the protein during its "biological motion." In future work, we plan to add a visualization component to facilitate MSM analysis.

PARTICIPANTS: *PRIME 2014 Student:* Garrett Chan; *UCSD Mentors:* Rommie Amaro, Robert Malmstrom; *National Taiwan University* and *Academia Sinica:* Jung-Hsin Lin



The Internet of Things: Interactive Museum Models with Microcontrollers

The Internet of Things is a new popular term that describes the inter-connection of billions of sensors, devices, wearable technologies, and even household appliances to create a global network of systems and services that go beyond machine-to-machine communication. Museums are beginning to leverage this trend to create new means of engaging audiences at increased levels of interactivity. For example, the San Diego Railroad Museum will launch a new scale-model train exhibit that reflects the history of Balboa Park for the park's Centennial Celebration in 2015. The goal of this collaborative project is to make this exhibit more interactive and engaging through the use of microcontrollers, which are small, power-efficient computers on a single integrated circuit. This prototype microcontroller system would allow the public to control and operate various elements of the model using their own mobile devices. A second goal of this project is to test the feasibility of streaming live video from a train-mounted camera, allowing people everywhere to have a first-person view through a Web interface. Both of these goals, when implemented, will engage visitors more actively and make the exhibit more memorable overall.

We have created a working prototype that allows users to control the on- and off-state of electronics. We also created a prototype Web interface that would

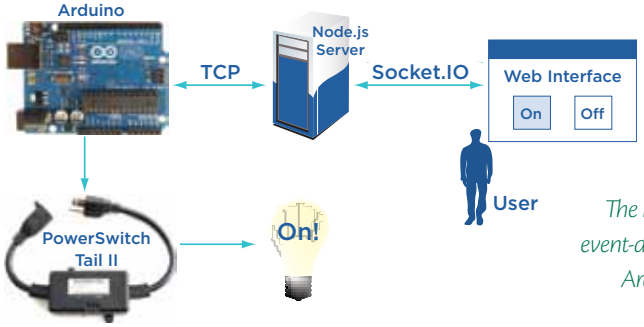


Figure 1: Diagram shows connection between Arduino and users via a Web interface. The Node.js server facilitates the connection, which is a runtime environment that uses an event-driven, non-blocking I/O model. It relays signals from the browser (via Socket.IO) to the Arduino (via TCP). The PowerSwitch Tail II controls the turning on/off of lights or sounds.

Images: (left to right) Lego train—courtesy of Jason Haga; Iwatyama Monkey Park, Arashiyama in Kyoto, Japan and Banpaku Kinen Koen, Suita, Japan—both courtesy of Anthony Nguyen

VISITING THE VISLAB

VisLab OSAKA, located in the Kita-Umeda district of Osaka, Japan, is an outreach activity of the Visualization Lab at PRAGMA member Osaka University. The exhibition is part of Knowledge Capital, a public space that features global prototypes and world-leading technologies. Visitors can experience cutting-edge inventions and participate in the creative process by providing comments and reactions to the inventions on display. Within the lab are "communicators," specialist staffers who approach visitors and encourage engagement with other visitors, as well as new encounters and experiences. The communicators gather public feedback and relay the information to companies, researchers and other event organizers. The main goal of the VisLab is to make difficult–to-understand information more accessible to everyone through art and good design.

Video Demo: drive.google.com/file/d/0B6_Ayl7YnEAIZGRfcFJUWTVqSVk/edit?usp=sharing

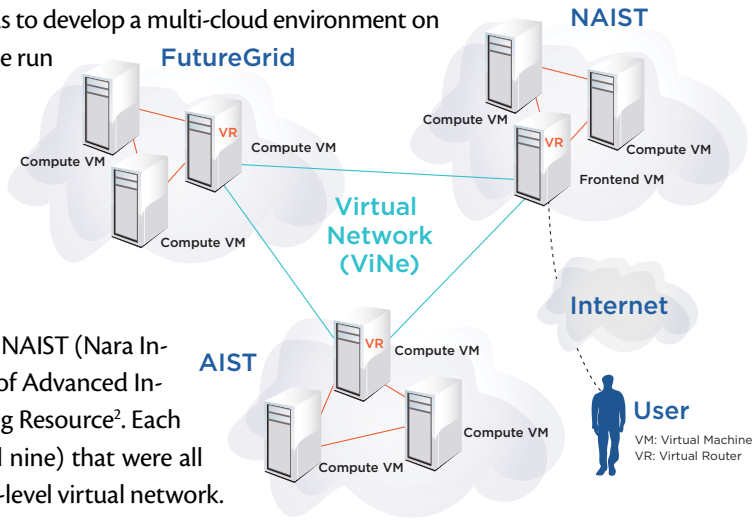
facilitate the interactions between the users and the microcontrollers. In order to control lights and sounds, we used an Arduino (an open-source electronics hardware and software platform), combined with the PowerSwitch Tail II to control the voltage and receive signals from the Web interface. As visitors switch lights and sounds on and off using the interface, the server sends instructions to the Arduino and the corresponding action is executed. The video-streaming prototype was implemented using a Raspberry Pi and a webcam mounted on the model train. The Raspberry Pi, a credit card-sized computer, served as a recording device and server so that users can view the streaming of the model through a separate Web interface. The quality of the video stream is optimized for the best streaming performance depending on the type of mobile device used to view the video.

The prototype system was showcased to the public at the VisLab in Osaka's Grand Front Building. Future work will include refining the user interface and integrating the system with the scale model being built in Balboa Park for its centennial.

PARTICIPANTS: *PRIME 2014 Student at NICT:* Nicole Wong; *UCSD* and *AIST:* Jason Haga (Mentor); *Balboa Park Online Collaborative:* Chad Weinard; *Osaka U* and *NICT:* Shinji Shimojo (Host Mentor); *Kobe Institute of Computing:* Teruaki Yokoyama; *Osaka U:* Yuichi Teranishi and Yoshimasa Ishi

Deployment of a Multi-Cloud Environment for Virtual Screening; Implementing Hadoop in a Virtual Screening Workflow

Unlike with grid computing, which requires users to be mindful of the software-hardware interface, large-scale computational research increasingly uses cloud computing. The goal of this project was to develop a multi-cloud environment on which molecular docking tests (i.e., virtual screenings) could be run via the software program DOCK[1]. This environment provides access to essentially unlimited resources and is more flexible than a hardware-based grid system. The long-term goal is to establish a multi-cloud computing paradigm for virtual screenings that would aid in the drug development/discovery process (see also section on **Virtual Cluster Sharing**).



The multi-cloud environment consisted of clouds located at NAIST (Nara Institute of Science and Technology), AIST (National Institute of Advanced Industrial Science and Technology), and FutureGrid Computing Resource[2]. Each cloud contained a cluster of three virtual machines (in total nine) that were all connected by ViNe (University of Florida, see sidebar), a user-level virtual network.

Figure 2: Conceptual diagram illustrating the complete multi-cloud set-up. Blue lines indicate connections between virtual routers (VRs) using ViNe to create a virtual network between virtual machine (VM) clusters.

VIRTUAL NETWORK (VINE) ENABLES SCIENCE

The Virtual Network (ViNe) Architecture for grid computing is a project developed at the University of Florida that implements routing and other communication mechanisms needed to deploy, operate and manage user-level, software-defined IP overlay networks. ViNe is particularly appealing for cloud computing because it allows the establishment of wide-area virtual networks supporting symmetric communication among public and private network resources (even when they are behind firewalls). It does so without requiring changes to either the physical network or the operating system of machines, and its virtualization overheads are low.

See: vine.acis.ufl.edu/wiki/index.php/Main_Page

Advantages of ViNe are that it does not require any changes to the virtual machine specifications and that it allows communication past boundaries such as firewalls. Thus, full connectivity can be established between virtual machines across clouds through ViNe.

Overall, the performance of the multi-cloud environment produced job-completion times on the same order of magnitude as the previously established grid computing-based virtual screening workflow. This is significant because it suggests that the overhead from executing jobs on multiple clouds is negligible. Furthermore, the added flexibility of a multi-cloud environment system makes deployment and use of these systems preferable. To specifically assess the performance and flexibility of the multi-cloud environment, DOCK tests were run with various cloud cluster sizes and workloads. Increasing the number of virtual machines on one cloud did not skew the job distribution to any particular cloud. When the workload was increased by adjusting the number of processes run simultaneously, job completion times were about 2.5 times greater, but remained comparable to the completion times for the grid-enabled virtual screening workflow. More time was needed, perhaps due to the relatively small number of cloud resources; thus when using larger clouds, it is anticipated that increasing the number of jobs will not have a large effect on completion rates.

In parallel to the multicloud testing using Dock, another PRIME project created virtual machine images that included a complete Hadoop framework, with a longer-term goal of facilitating the use of extremely large chemical databases and allowing easy use of multiple resources. We confirmed that the fully-distributed mode of Hadoop was functional using a cluster of virtual machines at NAIST. This mode facilitates the distribution of ligand data to datanodes, DOCK6 job execution, and fault tolerance of the system.

Future work will include adding fault tolerance to the multiclould system by incorporating Hadoop YARN (for job scheduling) and Map/reduce (for data input/handling) as the job distributer. Subsequently, a large-scale (>15 million ligands) virtual screening will be performed to fully assess the multi-cloud environment's performance and provide new information for drug discovery research.

PARTICIPANTS: *PRIME 2014 Students at NAIST:* Anthony Nguyen (multicloud); Katy Pham and Derek Song (Hadoop); *UCSD* and *AIST:* Jason Haga (Mentor); *NAIST:* Kohei Ichikawa (Host Mentor); *U Florida:* Maurício Tsugawa; *PRIME 2013 Students at NAIST:* Karen Rodriguez and Kevin Lam

[1]dock.compbio.ucsf.edu/Overview_of_DOCK

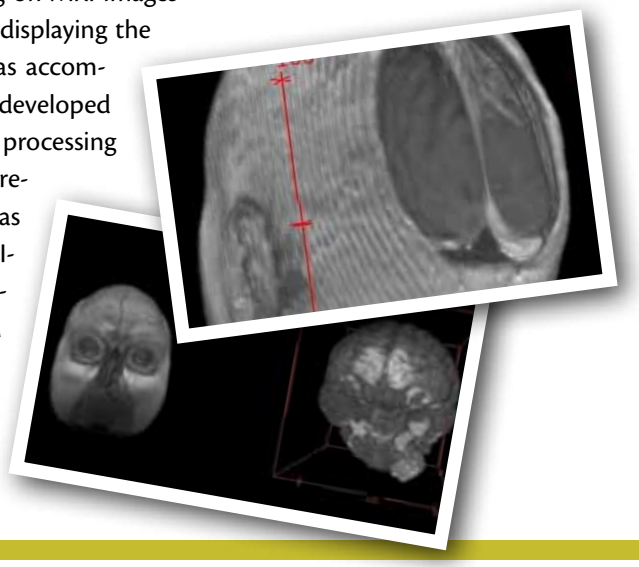
[2]portal.futuregrid.org/about

Images: (left to right) Nara Tokae festival—courtesy of Anthony Nguyen; Holding a koala and scuba diving off of Lady Elliot Island—courtesy of Matthew Schwegler

Kepler Workflows for Neuroimaging Tools and 3D Image Generation

Advances in imaging technologies offer the promise of understanding the functions of healthy, diseased and abnormal brains. However, analyzing these images requires the use of multiple tools often in an increasing number of steps (i.e., a workflow). The goal of this project is to extend the functionality of the Kepler Workflow system by integrating the workflow with tools used within the neuroimaging community. To accomplish this, we worked with researchers at the neuroimaging labs on the University of Queensland (UQ) campus to understand their work and research needs better. We then mapped the user needs into the constraints of the Kepler environment by integrating existing neuroimaging tools. We also developed a new 3D module to render images natively within the Kepler Workflow system.

Many of the imaging tools integrated into Kepler came from FSL, a comprehensive library of neuroimaging tools, which includes the Brain Extraction Tool (BET) and Linear Image Registration Tool (FLIRT). Additionally, we developed tools for converting between the two key MRI file types, NIFTI and DICOM. With these tools functioning on MRI images within Kepler, the next step was to create a means for displaying the manipulated images in a meaningful way. This task was accomplished using ImageJ, the open-source 3D imaging API developed by the NIH. Using this API, we created a native actor (i.e., processing component) within Kepler that displays 3D images representing a new functionality inside of Kepler. The actor was extended to work on a variety of file types such as DICOM, NIFTI, TIFF, etc. Thus the newly created, functioning Kepler workflow was able to take a series of DICOM files, extract the brain, conduct a linear regression to the norm, and display the final product to a user.



KEPLER: YOUR SCIENCE. ENABLED.

The Kepler Project is dedicated to furthering and supporting the capabilities, use and awareness of the free and open-source scientific workflow application known as Kepler. Kepler is designed to help scientists, analysts and computer programmers create, execute and share models and analyses across a broad range of scientific and engineering disciplines.

The Kepler Project is supported by award #0722079 from the U.S. National Science Foundation and other sources. For more information, see kepler-project.org.

Hoang Nguyen, a former MURPA Student who is now a UQ PhD student, developed a Web portal version of Kepler, extending it from its local running environment to a more accessible Web interface. A virtual machine of the new workflow was developed and executed in the Kepler portal. Extending the code converter to include converting NIFTI to MINC file types allowed the Web portal to display the output images of the new Kepler tools.

PARTICIPANTS: *PRIME 2014 Student at U Queensland:* Matthew Schwegler; *U Queensland:* David Abramson (Host Mentor) and Hoang Nguyen (MURPA 2009 Student); *U Queensland, UQ Center for Advanced Imaging:* Graham Galloway and Andrew Janke; *UCSD:* Ilkay Altintas (Mentor)

Smartphone App and GPS Data to Help Social Groups Manage Their Meet-ups in Real Time

Social event planning programs are becoming increasingly popular among smartphone users and we envision improving location-sharing services specifically through our own event and location application called "Rendezvous." The goal of this project is to develop an app that will allow users to create events and, for a period of time, it will provide access to a map displaying invited friends' locations. This new capability can curtail texting and driving since members of an event can see each friend's estimated arrival time and drivers will not need to send updates continually. With these arrival times available, the app can also allow users to run a few errands comfortably—rather than idly waiting for guests.

The Rendezvous app differs from other location-sharing apps in that each member of an event can specify how long to share their location, and each member chooses who they want to share their location with, as opposed to broadcasting their location to everybody using the app. Additionally, one of the project's biggest goals is to protect members' privacy. Rather than storing their phone numbers on a server, the app will assign each member an ID number to be stored instead. As much as possible,

Images: Screen shots of the Rendezvous app created by Katerina Zorko

the project aims to limit and monitor the amount of personal information that is sent to the server and is then shared with other members—specially for features such as one where the user can stop location sharing altogether (at which point it is crucial that the user's location stored on the server is not sent to other group members).

To date, basic design of the app has been completed, including transitioning from various screens of the app, creating new events through the user interface as well as on the server, choosing location-sharing settings, and sending notifications. The next step is to establish the connection of the front end with the back end. This involves sending requests to the server we have built to implement certain actions, such as creating or deleting an event, or saving the user-specified settings associated with an assigned ID number. After these server calls are set up, a user will be able to view another user's location on his or her own map, which will more or less complete our first prototype of Rendezvous.

PARTICIPANTS: *PRIME 2014 Student at U Queensland:* Katerina Zorko; *U Queensland:* David Abramson (Host Mentor); *UCSD:* Jurgen Schulze (Mentor); *UCSD Alumnus:* Vinay Mavram

Image: (left to right) Peng-Lien Cook, Rob Cook (Director of QCIF), David Abramson, Kyle-David Suico, Matthew Schwegler, Heather Abramson, Katerina Zorko

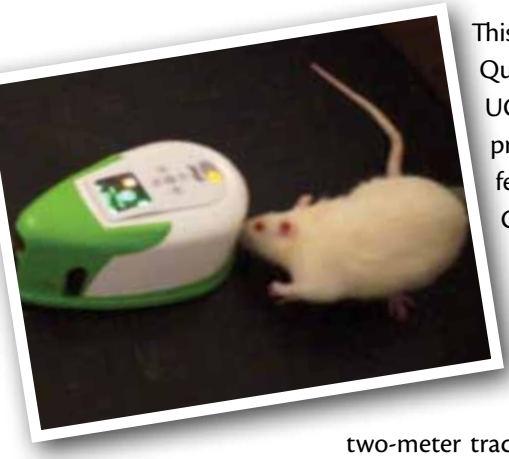
Background: Mount Luxmore, South Island, New Zealand—courtesy of Katerina Zorko

MURPA/QURPA

International Research Internships at the University of Queensland and Monash University

In today's educational arena, universities must provide students with opportunities to work and study abroad to prepare them for global citizenship and professional competence in a multi-cultural workplace. Numerous reports have challenged universities to develop educational programs that provide an integrated academic basis for developing students' cultural/global competencies.

Over the past six years, 32 students from Australia's Monash University have travelled to international partners under the Monash Undergraduate Research Projects Abroad (MURPA) program. They travelled to the University of California, San Diego, the National Center for Supercomputing Applications in Illinois, the Israel Institute of Technology-Technion in Israel, the Institute for Infocomm Research (I2R) in Singapore, and the University of Warwick in the United Kingdom. Students are placed for a period of eight weeks, allowing them time to integrate into research groups as team members. Students have a local mentor in Australia as well as one at the remote site, and the team members often bridge international research projects. In 2013-14, three Monash students travelled to UCSD, 2 to I2R in Singapore, and 1 to Warwick University Their projects are described at messagelab.monash.edu.au/MURPA/MURPAStudentPres2014.



This year for the first time, a University of Queensland (UQ) student, Brian Song, travelled to UCSD and joined students from Monash. Song's project concerned iRat-Rat experiments in Professor Andrea Chiba's Rat Lab, located in UCSD's Cognitive Science department. The overall purpose was to see if they could elicit or verify altruistic motives made by (real) rats. Song also assisted in a neural recording experiment in Professor Doug Nitz's lab, where they raced the iRat (pictured) and a rat down a straight, two-meter track. The purpose of the experiment was to observe the neural activity of mirror place cells in the rat as it races against the iRat.

The UQ student also met with key researchers at Calit2 (Qualcomm Institute), the Salk Institute and Brain Corporation (all in San Diego). Calit2 is a significant partner for UQ's Research Computing Centre.

MURPA (at Monash) and QURPA (at UQ) make it possible for students to attend virtual (remote) seminars given by world-renowned experts before they leave Australia for their foreign destinations. These seminars also allow students to "meet"

potential UCSD mentors to seek out information about potential research projects. In 2014, seminars originated with faculty experts from the University of Southern California, UCSD, and Purdue University (all in the U.S.); I2R in Singapore; and both Monash and UQ in Australia—making the seminar series truly global. As in the past, the seminars were broadcast simultaneously to Monash (in Melbourne) and UQ (in Brisbane), with audiences able to ask questions from either venue. The seminar infrastructure supports a wide range of videoconferencing technologies (both open source and commercial), and is shown on a 20 megapixel OptIPortal display system.

MURPA students have also become resources for future MURPA, QURPA as well as PRIME students, helping mentor the next generation of students. For an example of this, read the story in this publication about PRIME 2014 student Matthew Schwegler and MURPA 2009 student Hoang Nguyen, who worked jointly on Kepler Workflows for Neuroimaging Tools and 3D Image Generation.

Images: (left to right) MURPA Lecture. Ilkay Altintas (UCSD) on screen, taken in U Queensland—courtesy of David Abramson; View in the Blue Mountains—courtesy of PRIME 2014 Student Matthew Schwegler

BUILDING COMMUNITY

OUTREACH AND ENGAGEMENT

A fundamental component of PRAGMA activities involve engagement with new ideas and people. PRAGMA Workshops are one form of such engagement. The PRAGMA Students group is another venue for encouraging new ideas and bringing more students into the PRAGMA community. In addition, we take intentional steps to co-organize other meetings that focus on topics of keen interest to PRAGMA and to individuals and communities we seek to engage. This year we highlight activities, all in Southeast Asia in areas of regional and global importance, including e-science, big data, water disaster management and water quality, and biodiversity in extreme environments. These activities were on the agenda at PRAGMA-related meetings in 2013 and 2014:

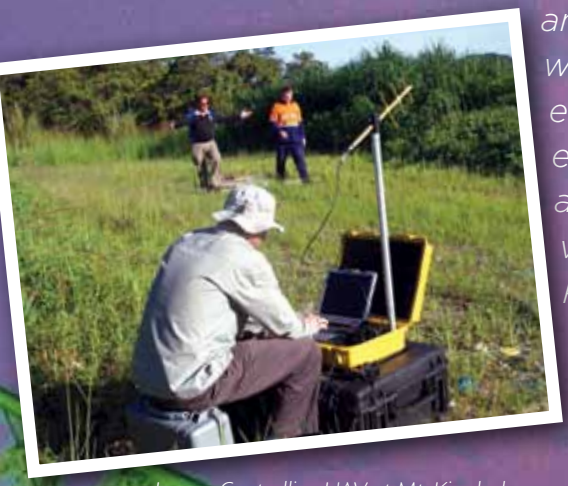


Image: Controlling UAV at Mt. Kinabalu National Park – courtesy of Antony van der Ent (see ***Biodiversity Expedition****)*

Southeast Asia International Research and Training Program (SEAIP) and PRAGMA Institute

TAICHUNG AND TAINAN, TAIWAN; DECEMBER 2–6, 2013

The SEAIP's ongoing series of workshops (seaip.narlabs.org.tw) have opened doors for collaborations between researchers in Southeast Asia and the rest of the world, forming the basis for growing PRAGMA collaborations in region. The theme of the 8th SEAIP was "E-Science for Scientific Expedition." The informal atmosphere was conducive to learning and to developing collaborations. The workshop was organized by Taiwan's National Center for High-performance Computing (NCHC), with funding from the country's Ministry of Science and Technology as well as from the National Applied Research Laboratory. In addition to sessions at NCHC's Taichung and Tainan branches, there was a session held at the National Cheng Kung University in historic Tainan. The 2013 workshop also was used to plan other outreach activities (ICSE and Shonan, mentioned below) and helped spur development of a mobile app of use to the biodiversity expedition.

The next SEAIP meeting is scheduled to be held December 1–5, 2014 in Tainan and Kaohsiung.

International Conference on Serpentine Ecology (ICSE) and Mini-PRAGMA Workshop

KOTA KINABALU AND MT KINABALU NATIONAL PARK, MALAYSIA; JUNE 10–12, 2014

Building on the activities of the Biodiversity Expedition and the associated outreach activities at PRAGMA 23 and 24, along with activities in Taiwan in December 2012 at the Bridging Big Data Infrastructure workshop in Taiwan, and later at the Mini-PRAGMA in Indonesia in June 2013, another mini-PRAGMA workshop was planned and held in association with the 8th International Conference on Serpentine Ecology (ICSE, icse2014.com). The workshop was the premier gathering of researchers from around the world who are interested in biodiversity and ecology in ultramafic (high magnesium- and iron-oxide concentrations) outcroppings. Prior to the workshop, there were several days for initial fly-overs with unmanned aerial vehicles (UAVs) of transects of Mount Kinabalu to provide input to PRAGMA's Biodiversity Expedition. At the mini-workshop there was an opportunity to exchange views on the needs of the serpentine ecology community and PRAGMA's technologies. The talks included one titled, "EcoSabah: Android-Based Reporting Solution for Species Occurrences in Sabah," by Nor Arlina Amirah bt Ahmad Ghani from the Universiti Teknologi Malaysia (a presentation encouraged at the SEAIP 2013 meeting). A follow-up from this meeting was the coupling of the leads of the UAV project from the University of Queensland (UQ) with a key PRAGMA partner at UQ on how to process the data from the UAV. (See **Biodiversity Expedition**)

Background: Kobe Port Tower at night—courtesy of Anthony Nguyen

Water Disaster Management and Big Data

SHONAN VILLAGE, JAPAN; JULY 7–10, 2014

Water disasters include flooding from typhoon, seasonal storms and other natural or man-made causes, usually coupled with mudslides that may occur without warning. Water disaster management is a challenge in many parts of the world, and big data is emerging as a new opportunity for prediction and decision support.

During presentations at PRAGMA 21 (Sapporo, October 2011), PRAGMA members recognized the need to focus efforts on understanding how technologies being developed by PRAGMA and the wider community could be applied to disaster preparation, mitigation and recovery. The organizers proposed and succeeded in developing a workshop on Water Disaster Management under the Shonan Meeting organization (see www.nii.ac.jp/shonan/blog/2013/01/27/water-disaster-management-and-big-data)

Using water disaster management as a specific disaster management application, the goal of this meeting was to bring together an interdisciplinary community of researchers, practitioners and developers. They explored current and future challenges identified by planners and responders that require research, development or deployment of big data and cyberinfrastructure solutions.

The group focused on three types of water disasters: flooding, typhoon, and tsunami. They also discussed the unique as well as common technology needs in areas such as monitoring, prediction and infrastructure. These types of disasters require integration of data from various sources, real-time monitoring and reporting (from sensors or social media), and computational resources.

Several concrete actions were suggested, for subgroups to pursue, including continuing discussions at upcoming workshops, including PRAGMA 27 and 28, as well as at the SEAIP meeting slated for December 2014 in Taiwan. Concrete actions would include the development of a prototype system for migrating data centers and sharing phased-array radar data between Japan and Taiwan to give more information about the volume and location of rainfall during a typhoon.

A copy of the report can be found at seaip.narlabs.org.tw, at the Shonan Meeting link.

Lower Mekong Initiative

HANOI, VIETNAM: AUGUST 18–22, 2014

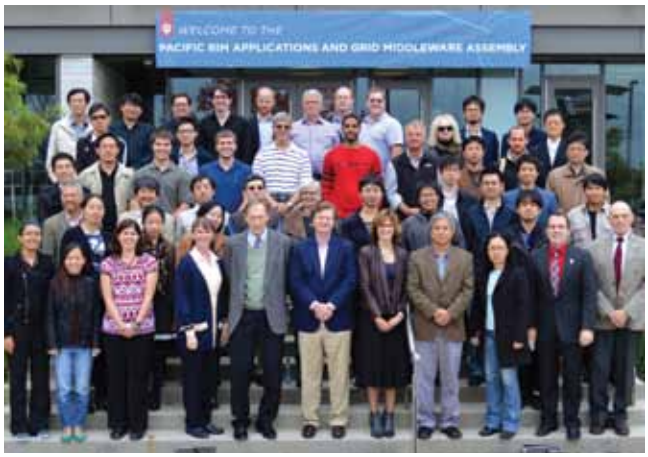
The Lower Mekong Region (LMR) includes Cambodia, Laos, Myanmar, Thailand and Vietnam. The region is defined by the trans-boundary river that includes the LMR countries and the People's Republic of China, and the Mekong River is of key economic importance to the people who live along it in the region.

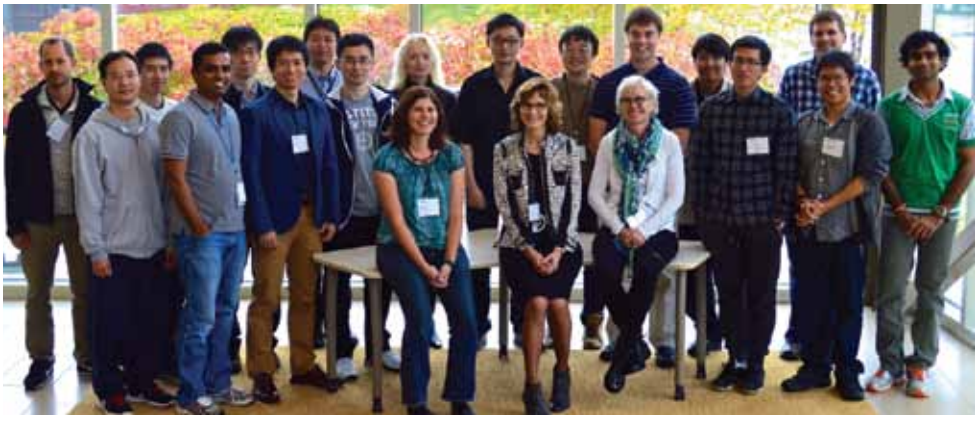
The Lower Mekong Initiative (LMI) workshop was conceived in 2012 to build networking connectivity in many countries in the region. It coincided with the U.S. State Department's interest in the Lower Mekong. The LMI workshop (internationalnet-

working.iu.edu/initiatives/LMI) brought together network experts and water resource managers and scientists from the LMR countries and the United States.

PRAGMA was one of the co-organizers of the workshop, motivated by its partnership with TransPac (a member of PRAGMA) and the Network Startup Resource Center (www.nsrc.org). The workshop also aligned with PRAGMA's goals of focusing regionally on Southeast Asia and topically on issues relating to water management.

Preliminary feedback from the workshop was captured by co-organizer in the form of video interviews available at: youtu.be/oJHeUYiu-Y8. One follow-up was to engage workshop attendees with the broader PRAGMA community at the December 2014 SEAIP workshop in Taiwan (seaip.narlabs.org.tw). A report from the LMI Workshop will be available in the near future and will be posted at the workshop website: internationalnetworking.iu.edu/archives/LMI.





Images: (clockwise from the top) Weicheng Huang and Fang-Pang Lin (NARL/NCHC) with Nadya Williams (UCSD) at PRAGMA 27 in McCormick's Creek State Park—courtesy of Peter Arzberger; Group photo of PRAGMA Students Workshop participants at PRAGMA 27—courtesy of Teri Simas; Group photo for PRAGMA 27—courtesy of Teri Simas; Shava Smallen (UCSD) and Beth Plale (IU) at PRAGMA Students workshop at PRAGMA 27—courtesy of Teri Simas

WORKSHOPS AND WORKING GROUPS

PRAGMA workshops are meetings of all members of the PRAGMA community. They are the major vehicle for information exchange between working groups, researchers, and institutions. They also provide excellent opportunities to engage new researchers and students at the host sites. New participants bring new perspectives, applications, technologies, students and resources. The twice-yearly workshops are critical for demonstrating progress between meetings and to plan for actions prior to the subsequent workshop.

Workshops are hosted by different organizations to provide a platform for PRAGMA members to meet and discuss research interests and, ideally, to develop new collaborations with members of the host institution.

The workshops are organized to allow for updates from PRAGMA's four working group:

- Resources Working Group: Working to make the distributed resources of PRAGMA useful to diverse applications. Co-leaders: Yoshio Tanaka (AIST) and Philip Papadopoulos (UCSD);
- Telescience Working Group: Focusing on activities that require access to, or use of, remote equipment, such as tiled-display walls and sensors. Co-leaders: Shinji Shimojo (NICT and Osaka U) and Fang-Pang Lin (NARL/NCHC);
- GEO Working Group: Creating an infrastructure to share and integrate data on global earth observations, including remote-sensing data and data from land-, lake-, and ocean-based sensors. Co-leaders: Sornthep Vannarat (NECTEC), Ryosuke Nakamura (AIST), and Franz Cheng (NARL); and
- Biosciences Working Group: Focusing its efforts on integrating technologies to create an infrastructure to advance the screening of potential compounds to combat infectious diseases. Co-leaders: Jason Haga (AIST and UCSD) and Habibah Wahab (USM).

In addition, PRAGMA allow for new working groups to be formed, and one group now in its early stages is focused on Cyber-Learning (see **Edison Paves the Way for Cyberlearning in PRAGMA**), co-lead by Ruth Lee (KISTI), Hsi-ching Lin (NCHC), and Putchong Uthayopas (Kasetsart U)

Images: (left to right) Voice of Namasia Singers, Tainan, Taiwan at PRAGMA 27; PRAGMA 26 Participants—courtesy of Teri Simas

PRAGMA WORKSHOPS

In 2014, two PRAGMA Workshops were held:

- PRAGMA 26, 9–11 April, 2014, in Tainan, Taiwan, hosted by the National Center of High-performance Computing and National Cheng Kung University.
- PRAGMA 27, 15–17 October, 2014, at Bloomington, Indiana, hosted by Indiana University's Data to Insight Center.

We are grateful to the hosts and thank them for holds successful workshops.

PRAGMA Workshops also host auxiliary, targeted activities. Two such activities were organized by and for the PRAGMA Students organization (see section on **Students**). This year there were special sessions associated with each workshop. They included the:

- Workshop on Cyber-Learning, which focused on building ties between Korean and Taiwanese researchers. It was hosted by NCHC on April 9, 2014, as part of PRAGMA 26 in Tainan; and the
- Networking Workshop, hosted by IU's Data to Insight Center and Pervasive Technology Institute on October 15, 2014, as part of PRAGMA 26 in Bloomington.

In addition, training programs such as the Southeast Asia International Joint Research and Training Program (seaip.narlabs.org.tw) exposed more researchers to PRAGMA, many of whom became new members.

The section on **Building Community** describes several workshops during the last year that PRAGMA helped organize or co-organize. They included the Southeast Asia International Research and Training Program (SEAIP) and PRAGMA Institute in Taiwan; the International Conference on Serpentine Ecology (ICSE) and Mini-PRAGMA Workshop, both in Malaysia; the Water Disaster Management and Big Data meeting in Shonan Village, Japan; and the Lower Mekong Initiative in Vietnam.

Looking to the future, we will continue to employ these strategies to engage new researchers. In addition, we will work with our members to identify strategic partners and engage them through focused scientific or technical workshops. We will focus in particular on India, China, and Southeast Asia as regions where there is a growing investment in infrastructure and where there are natural partnerships of mutual interest. Listed below are our planned workshops:

- PRAGMA 28, April 8–10, 2015, Nara, Japan, hosted by the Nara Institute of Science and Technology.

MEMBERS

*PRAGMA is an institution- and people-based organization governed by a Steering Committee that invites new members, determines locations of workshops, and sets overall direction. More information about Steering Committee members (denoted with an asterisk * in the listing below) may be found at www.pragma-grid.net/people.php*

ACTIVE MEMBERS

A key component of PRAGMA is active involvement, by participation in workshops, contributing resources, hosting workshops, and/or promoting and supporting student and researcher exchanges. The following institutions have contributed to PRAGMA activities in the past year. This year we welcome two new institutional members, Nara Institute of Science and Technology (NAIST) in Nara, Japan, and the Institute for Computational Science and Technology (ICST) in Ho Chi Minh City Vietnam.

CENTER FOR COMPUTATIONAL SCIENCES (CCS), UNIVERSITY OF TSUKUBA: Osamu Tatebe, *tatebe@cs.tsukuba.ac.jp;* Taisuke Boku, *taisuke@cs.tsukuba.ac.jp;* Mitsuhisa Sato, *msato@cs.tsukuba.ac.jp*

COLLEGE OF COMPUTER SCIENCE AND TECHNOLOGY (CCST), JILIN UNIVERSITY (JLU): Xiaohui Wei*, *weixh@jlu.edu.cn,* Hongliang Li, *lihongliang@jlu.edu.cn*

COMPUTER NETWORK INFORMATION CENTER (CNIC), CHINESE ACADEMY OF SCIENCES (CAS): Baoping Yan*, *ybp@cnic.ac.cn;* Kai Nan*, *nankai@cnic.ac.cn*

CYBERMEDIA CENTER (CMC), OSAKA UNIVERSITY: Shinji Shimojo*, *shimojo@cmc.osaka-u.ac.jp;* Susumu Date*, *date@cmc.osaka-u.ac.jp; Yoshiyuki Kido, kido@cmc.osaka-u.ac.jp*

DATA TO INSIGHT CENTER, INDIANA UNIVERSITY (IU): Beth Plale*, *plale@indiana.edu*

INSTITUTE FOR COMPUTATIONAL SCIENCE AND TECHNOLOGY (ICST): Thanh N. Truong, *Thanh.Truong@ICST.org.vn, TNTruong.ICST@gmail.com;* Lam K. Huynh, *Lamhuynh.us@gmail.com;* Tuan Huynh, Cyber Lab, *vantuanspkt@gmail.com*

KASETSART UNIVERSITY (KU): Putchong Uthayopas*, *pu@ku.ac.th,* and *putchong@gmail.com*

KONKUK UNIVERSITY (Konkuk): Karpjoo Jeong*, *jeongk@konkuk.ac.kr*

NARA INSTITUTE OF SCIENCE AND TECHNOLOGY (NAIST): Kazutoshi Fujikawa, *fujikawa@itc.naist.jp,* Kohei Ichikawa, *ichikawa@is.naist.jp*

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING (NCHC), NATIONAL APPLIED RESEARCH LABORATORIES (NARL): Whey-Fone Tsai*, *wftsai@nchc.narl.org.tw;* Fang-Pang Lin*, *fplin@nchc.narl.org.tw*

NATIONAL ELECTRONICS AND COMPUTER TECHNOLOGY CENTER (NECTEC): Sornthep Vannarat, *sornthep.vannarat@nectec.or.th*

NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST): Satoshi Sekiguchi*, *s.sekiguchi@aist.go.jp;* Yoshio Tanaka*, *yoshio.tanaka@aist.go.jp;* Jason Haga, *jh.haga@aist.go.jp*

NATIONAL INSTITUTE OF SUPERCOMPUTING AND NETWORKING (NISN), KOREA INSTITUTE OF SCIENCE AND TECHNOLOGY INFORMATION (KISTI): Kum Won Cho*, *ckw@kisti.re.kr;* Jong-Suk Ruth Lee, *jsruthlee@kisti.re.kr*

THAMMASAT UNIVERSITY: Prapaporn Rattanatamrong, *rattanat@gmail.com;* Worawan Marurngsith Diaz Carballo, *wdc@cs.tu.ac.th;* Kasidit Chanchio, *kasiditchanchio@gmail.com*

UNIVERSITI SAINS MALAYSIA (USM): Habibah A. Wahab*, *habibahw@usm.my;* Chan Huah Yong, *hychan@cs.usm.my;* Mohd Azam Osman; *azam@cs.usm.my*

UNIVERSITY OF CALIFORNIA, SAN DIEGO (UCSD): including Calit2's Qualcomm Institute, San Diego Supercomputer Center (SDSC), Center for Research in Biological Systems (CRBS), National Center for Microscopy and Imaging Research (NCMIR), National Biomedical Computation Resource (NBCR): Peter Arzberger*, *parzberg@ucsd.edu;* Philip Papadopoulos*, *phil@sdsc.edu;* Teri Simas, *simast@sdsc.edu,* Rommie Amaro, *ramaro@ucsd.edu*

UNIVERSITY OF FLORIDA (UF), in particular the Advanced Computing and Information Systems Laboratory and the Florida Museum of Natural History: José Fortes, *fortes@acis.ufl.edu;* Renato Figueiredo, *renato@acis.ufl.edu;* Reed Beaman, *rbeaman@flmnh.ufl.edu*

UNIVERSITY OF HONG KONG (HKU): W.K. Kwan*, *hcxckwk@hku.hk;* P.T. Ho, *hcxchpt@hku.hk;* Lilian Y.L. Chan, *lilianyl@hku.hk*

UNIVERSITY OF WISCONSIN (UW), in particular the Center for Limnology: Paul Hanson, *pchanson@wisc.edu*

Networking Members

Networking partners provide access to expertise to improve the efficiency of the resources groups in running distributed experiments and applications.

ASIA-PACIFIC ADVANCED NETWORK (APAN): Seishi Ninomiya, *snino@isas.a.u-tokyo.ac.jp;* Kento Aida, *aida@nii.ac.jp*

PACIFIC WAVE: John Silvester, *jsilvest@usc.edu*

STARLIGHT AND TRANSLIGHT/STARLIGHT INITIATIVES: Maxine Brown, *maxine@uic.edu*

TransPAC3, INDIANA UNIVERSITY: Jennifer Schopf, *jmschopf@indiana.edu;* George McLaughlin, *george@mclaughlin.net*

Other Members

ACADEMIA SINICA GRID COMPUTING CENTRE (ASGC): Simon Lin, *Simon.Lin@twgrid.org;* Eric Yen, *eric.yen@twgrid.org*

ADVANCED SCIENCE AND TECHNOLOGY INSTITUTE (ASTI): Denis Villorente, *denis@asti.dost.gov.ph;* Jelina Tetangco, *jeng@asti.dost.gov.ph*

CENTER FOR HIGH PERFORMANCE COMPUTING, HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY (HUT): Huu-Duc Nguyen

GLOBAL SCIENTIFIC INFORMATION AND COMPUTING CENTER (GSIC), TOKYO INSTITUTE OF TECHNOLOGY (Titech): Satoshi Matsuoka*, *matsu@is.titech.ac.jp;* Hidemoto Nakada, *hidenakada@aist.go.jp*

INSTITUTE OF INFORMATION TECHNOLOGY-VIETNAM (IOIT-VN): Thai Quang Vinh*, *qvthai@ioit.ac.vn*

MIMOS: Thillai Raj T. Ramanathan, Ong Hong-Hoe, *hh.ong@mimos.my;* Ng Kwang-Ming, *kwang.ming@mimos.my;* Luke Jing Yuan, *jyluke@mimos.my*

MONASH UNIVERSITY (Monash): David Abramson*, *david.abramson@monash.edu,* Paul Bonnington, *paul.bonnington@monash.edu*

NZ eSCIENCE INFRASTRUCTURE: Nick Jones, *n.jones@auckland.ac.nz*

UNIVERSITY OF HYDERABAD (UoH): Rajeev Wankar, *wankarcs@uohyd.ernet.in, rajeev.wankar@gmail.com*

More information about each of the PRAGMA Institutional Members can be found at www.pragma-grid.net/members-partners.php

Additional Organizations Active in PRAGMA

BALBOA PARK ONLINE COLLABORATIVE (www.bpoc.org), is a nonprofit organization that provides technical support to more than 20 art, science, and cultural organizations in San Diego. They have contributed resources to the PRIME program to foster a culture of innovation and collaboration among nonprofit organizations through technology, with a focus on Balboa Park. (See page 26).

BIODIVERSITY INSTITUTE, UNIVERSITY OF KANSAS (biodiversity.ku.edu), and its researchers and students conduct research on seven continents in areas such as biodiversity informatics, systematics and ecology and evolutionary biology. They have contributed to the biodiversity expedition through participation in workshops and in the use and extension of the Lifemapper software.

NATIONAL APPLIED RESEARCH LABORATORY (NARL; www.narl.org.tw/en) was established in 2003 to consolidate nine national laboratories into a single nonprofit organization to construct, operate, and maintain the large-scale R&D facility and platform in support of academic research and foster the necessary manpower in various advanced fields focused by the nation. NCHC is one of the laboratories in NARL. NARL has provided leadership in the GEO Working Group, and can bring to bear several other laboratories at NARL for PRAGMA collaborations.

NATIONAL INSTITUTE FOR INFORMATION AND COMMUNICATION TECHNOLOGY (NICT; www.nict.go.jp), is an incorporated administrative agency that conducts general research and development on information technology supporting the ubiquitous society of the future. NICT supported students in the PRIME program in from 2009 through 2013 and has participated in the activities of the Telescience Working Group through support of the high-definition video conferencing testing.

UNIVERSITY OF QUEENSLAND (www.uq.edu.au) has recently become involved in PRAGMA through David Abramson's move there. David remains actively involved in PRAGMA and PRIME, supporting two students from UCSD from June to August 2013 (see Section on PRIME) and sending students to PRAGMA sites (see MURPA QURPA section).

UNIVERSITI TEKNOLOGI MALAYSIA (www.utm.my) researchers have been involved in the "**Biodiversity Expedition**" and in the workshops on biodiversity.

Background: Minoh Park, Osaka Prefecture, Japan—courtesy of Anthony Nguyen

Partners

GLOBAL LAKES ECOLOGICAL OBSERVATORY NETWORK, (GLEON, gleon.org) is a grassroots network of limnologists, ecologists, information technology experts, and engineers who uses the network of people, sensors, and data to understand issues such as eutrophication or climate change at regional to global scales. GLEON, established based on an early PRAGMA expedition to place sensors on a lake in Taiwan in 2004, GLEON has grown to a network of more than 300 members, has developed new knowledge and insights, created new data products and developed a very successful Graduate Student Association. There are several ties between GLEON and PRAGMA, including shared personnel, learning from the GLEON GSA to develop PRAGMA Student group, the shared Scientific Expedition on Lake Eutrophication, and the joint hosting of a workshop on Big Data in Taiwan in December 2012. For more about GLEON, see gleon.org.

Network Startup Resource Center (www.nsrc.org) develops network communications infrastructure and local engineering capacity in areas of the world where inadequate research and education network connectivity poses a significant barrier to international science collaborations. NSRC has worked in more than one hundred countries around the world over the past 20+ years to enhance cyberinfrastructure, both physical (network connectivity) and human (technical capacity), to enable and improve international scientific research. NSRC's recent work with PRAGMA to support researchers from Myanmar to attend PRAGMA 24, and teaming up with PRAGMA and IU on the Lower Mekong Initiative, facilitates more international science education in Southeast Asia. Through hands-on, lab-based curricula and a train-the-trainers approach, NSRC provides technical capacity development to augment networking expertise across the PRAGMA community.

SPONSORS

PRAGMA is supported by its member institutions and the U.S. National Science Foundation (NSF OCI 1234983, PI: Papadopoulos), involving researchers at UCSD, U Florida, Indiana U, and U Wisconsin. This involves support from NSF Office of Cyberinfrastructure, Office of International Science and Engineering, Division of Information and Intelligent Systems in the Directorate for Computer and Information Science and Engineering, and the Emerging Frontiers Office of the Directorate for Biological Science. This funding provides partial support for involvement of Indiana University, University of Florida and University of Wisconsin researchers. In addition, previous support came from NSF OCI-0627026, PI: Papadopoulos, and involved support from NSF's Office of Shared Cyberinfrastructure, Office of International Science and Engineering, Division of Information and Intelligent Systems, and Division of Biological Infrastructure.

ASGC is an e-Science center funded by both the Academia Sinica and National Science Council of Taiwan.

ASTI's e-science activities receives funding support from the Department of Science and Technology (DOST), Philippines.

CMC/Osaka University is supported by JGN-X of the National Institute of Information and Communications Technology (NiCT), Japan. Software defined networking work is supported in part by the collaborative research of the National Institute of Information and Communications Technology (NICT) and Osaka University (Research on High Functional Network Platform Technology for Large-scale Distributed Computing).

CCS's (at U Tsukuba) PRAGMA participation is partially supported by the JST CREST "Development of System Software Technologies for post-Peta Scale High Performance Computing."

CCST at Jilin University receives funding support from the Chinese Natural Science Foundation (61170004) and the Chinese Ministry of Education (985-project).

CNIC's sponsors include the Chinese Academy of Sciences, the Ministry of Science and Technology of China and the Natural Science Foundation of China.

HUT receives support from the Hanoi University of Science and Technology and Ministry of Science and Technology, Vietnam.

KU's PRAGMA participation has been partly funded by an SRU Grant, Kasetsart University Research and Development Institute (KURDI) and the National Research Council of Thailand.

KISTI receives major funding from the Ministry of Science, ICT and Future Planning (MSIP) through the 'EDucation-research Integration through Simulation On the Net (EDISON)' project, and from the National Research Council of Science and Technology (NST) through 'Construction and Collaboration of Simulation-based Cyber-Learning between Korea and Taiwan' project.

Monash University PRAGMA activities (including MURPA) are supported by a range of grants from the Australian Research Council and Monash University internal funding. In particular, the MURPA program wishes to acknowledge the financial support of the Faculty of Information Technology and the Monash e-Research Centre. We are extremely grateful to a number of donations from the Cybec Foundation, the Gringlas family and Gapgemini.

NCHC is one of the institutes monitored by the National Applied Research Laboratories that receives major funding support from the Ministry of Science and Technology, Taiwan.

NECTEC receives its funding through Thailand's National Science and Technology Development Agency (NSTDA).

Thammasat University (TU) Faculty of Science and Technology supports the participation of TU researchers in PRAGMA.

TransLight/StarLight receives major funding from NSF (OCI-0962997).

TransPAC3 receives major funding from NSF and Japan's National Institute of Information and Communications Technology.

USM's grid activities in Malaysia are funded mainly through E-science and USM Central Funding.

PRIME has support from a variety of sources: PRIME Alumni Haley Hunter-Zinck, given in tribute to her mother; the NSF Office of Cyberinfrastructure; National Biomedical Computation Resource (NIH/NIGMS 2 P41 GM103426-21); the Balboa Park Online Collaborative; the National Institute for Information and Communication Technology (NICT); and the Nara Institute of Science and Technology (NAIST). In addition, several students received scholarships to help support their participation (URS Ledell Scholarship; Revelli Scholarship, and Celniker Scholarship).

Work to build GLEON is supported in part by an award from the Gordon and Betty Moore Foundation (PI: Tony Fountain, UCSD) and the NSF Research Coordination Network award (PI: Paul Hanson, U. Wisconsin-Madison).

Images: (left to right) PRIME 2014 Student Katy Pham in Yoyogi Park poses in front of a sign noting that Tokyo is hers to discover!—courtesy of Katy Pham; National Palace Museum in Taipei—courtesy of Garrett Chan

PRAGMA would like to acknowledge those who helped prepare this report: Jennifer Matthews for graphic design; Doug Ramsey and Tiffany Fox for editing; Teri Simas for copy editing, and the section authors and photographic contributors!

Elliot Island, Australia—courtesy of Matthew Schwegler

PRAGMA 26 Group Photo, Tainan, Taiwan

Hachiman Shrine, Oinari, Fox Spirit—courtesy of Peter Arzberger

Mt. Kinabalu, UAV launch—courtesy of Antony van der Ent

Kobe Port, Japan—courtesy Anthony Nguyen

Tainan Confucius Temple (est. 1665)—courtesy of Teri Simas

PRAGMA 26, Voice of Namasia—courtesy of Teri Simas

Kepler Workflow applied to neuroimaging—courtesy of Matthew Schwegler

Lego train—courtesy of Jason Haga

View of Taipei—courtesy of Garrett Chan

Mt. Kinabalu's Western Summit Plateau—courtesy of Antony van der Ent

Sydney Opera House—courtesy of Katerina Zorko

Beneath the last Tori on Mt. Fuji—courtesy of Allen Nguyen

Tainan, Taiwan, during SEAIP 2013

Tower of Sun, Banpaku Kinen Koen, Suita Japan—courtesy of Anthony Nguyen

Fall Creek Reservoir, Virginia—courtesy Cayelan Carey

PRAGMA

www.pragma-grid.net