PRAGMA

PACIFIC RIM APPLICATIONS AND GRID MIDDLEWARE ASSEMBLY **2013-2014 COLLABORATION OVERVIEW**

PRAGMA began in 2002 as a workshop series.

INTRODUCTION

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

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 necessary information technologies 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 new researchers** with experience in cross-border science and by continuing to engage strategic partners;
- **Building and enhancing the essential people-to-people trust** and organization developed through regular face-to-face meetings—a core component 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 applications, and scientific expeditions and data-sharing require organizational support to succeed.

PRAGMA's Past Successes

PRAGMA began 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, responded to various disasters or emergencies (including SARS in 2003 and the Japanese tsunami of 2011), created a persistent experimental test bed for international cyberinfrastructure, and enhanced multiple software packages.

At the Mutienyu section of the Great Wall—courtesy Teri Simas

In addition, PRAGMA has incubated new scientific networks such as GLEON (Global Lakes Ecological Observation Network) and international projects such as the GEO Grid. It has also played a critical role in the formation of programs that help train the next generation of cyber-scientists through research exchanges, e.g., PRIME (Pacific RIM Experiences for Undergraduates, University of California San Diego), PRIUS (Pacific Rim International UniverSities, Osaka University) and MURPA (Monash University Research Project Abroad).

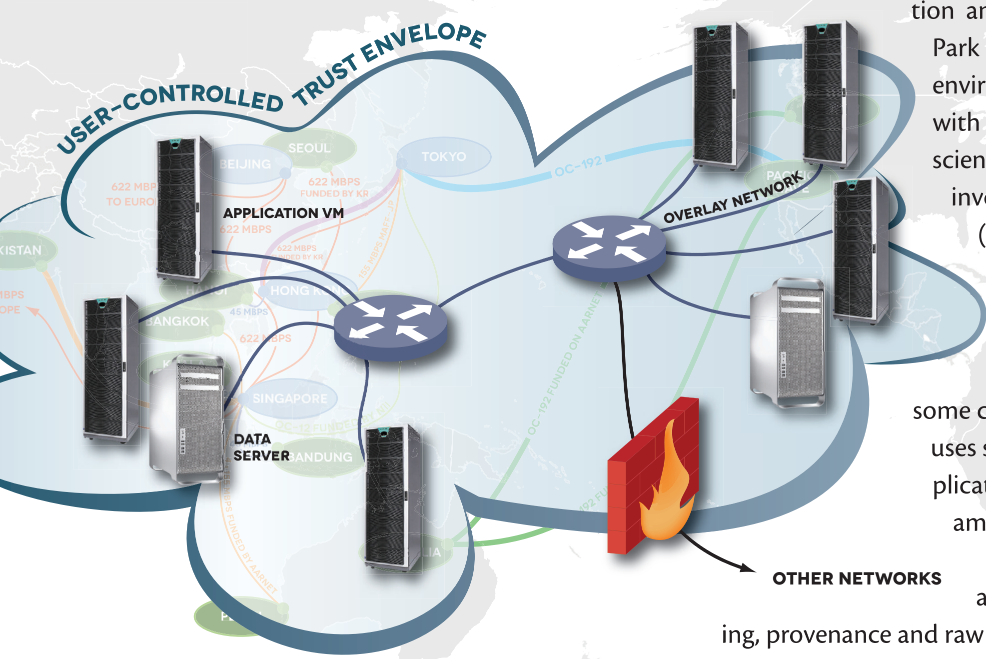
The Future of PRAGMA

Over the next several years, PRAGMA will advance the goals of several scientific expeditions that are forged from multidisciplinary teams to solve scientific problems through developing cyberinfrastructure. One of these sustained scientific expeditions will address the societal issues inherent to **predicting lake eutrophication** (i.e., defined as excessive plant, algal and bacterial growth due to nutrient enrichment)—one of the greatest water quality challenges facing freshwater ecosystems throughout the world. A second project seeks to **understand biological adaption 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. PRAGMA researchers will also **develop approaches for computer-aided drug discovery in infectious diseases** while focusing on neglected tropical diseases that affect about one billion people on this planet.

Each of these expeditions will advance the specific research goals of these groups and develop experimental infrastructure. For example, the lake and biodiversity expeditions require developing tools that track usage of data and protecting sensitive data by using virtual private networks. The drug discovery expedition integrates tools to create useful workflows for virtual drug screening. In all cases, we will work with existing scientific networks (e.g., GLEON) or grow nascent networks of researchers to ensure our developments have a broader impact.

In addition to these three scientific expeditions, PRAGMA will explore many other short-term expeditions, as well as possible expeditions in disaster recovery, building on the work of National Institute of Advanced Industrial Science and Technology

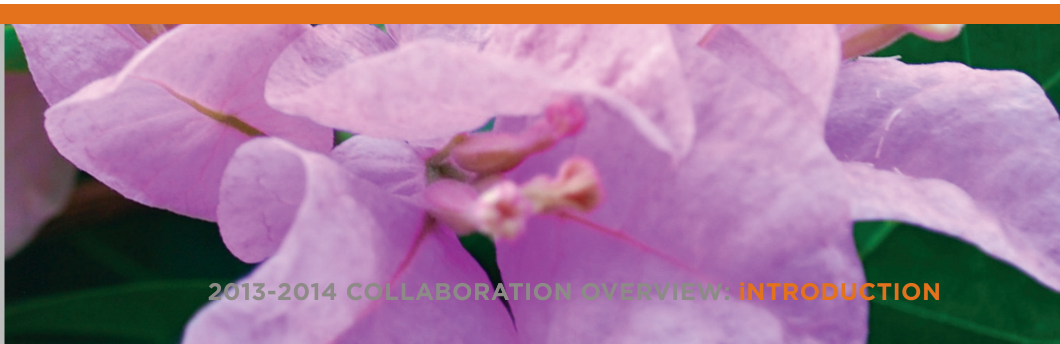
(AIST), the National Center for High-performance Computing (NCHC), and the University of California San Diego (UCSD). These expeditions will also be made possible by the network developed by the GEO Grid Project, led by AIST, and involving PRAGMA partners to include NCHC, Vietnam Academy of Science and Technology, and National Electronics Computer Technology Center (NECTEC). Other planned expeditions include cultural heritage sharing involving Osaka University, National Institute for Information and Communication Technology (NiCT), Balboa Park Online Collaboratory (BPOC), and UCSD; and environmental observatory deployments coupled with other scientific expeditions in biodiversity, lake science, coral reef, forestry, and disaster mitigation involving UCSD, Taiwan Forestry Research Institute (TFRI), Walailak University, and GLEON.



PRAGMA also intends to build on its successes in sharing virtual machines to become a **multi-provider cloud** (or multi cloud) with some 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 address fundamental issues in selective data sharing, provenance and raw data acquisition in a practical way. 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.

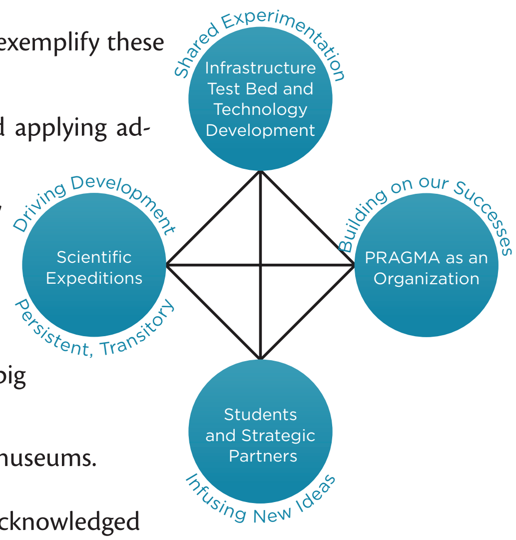
To ensure the long-term vibrancy of PRAGMA, we actively seek to engage new participants through developing student activities and forming strategic partnerships. Building on the activities of PRIME, PRIUS and MURPA—and motivated by the GLEON Student Association's (http://www.gleon.org/Students.php) success—we have created *PRAGMA Students,* a group that provides research and leadership opportunities for graduate students associated with PRAGMA. In addition, we have implemented plans to develop strategic partnerships with a focus on three key geographic areas: India, China and Southeast Asia. Through this activity, PRAGMA members will engage new partners in targeted workshops and in interactions of mutual benefit. These special workshops augment our ongoing meetings, which are open to members and others interested in participating, contributing to, and learning from PRAGMA activities. The regular meetings rotate among member institutions to engage researchers and students who are new to PRAGMA and to educate members about specific applications and technologies developed by the member institution and region. In addition, we explore new interests of members, such as Cyber-learning.

Images: (top) Hiroshima Peace Memorial—courtesy of Jesus Rios; Yangmingshan, Taiwan—courtesy of Sharon Grubner; (all others) Scenes from Bangkok—courtesty of Teri Simas



In this year's PRAGMA Collaborative Overview (2012-2013) we highlight activities that exemplify these directions:

- Advancing scientific expeditions in lake ecology and biodiversity by developing and applying advanced cyberinfrastructure approaches;
- Building toward a trust envelope via automating the use of virtual cluster technology and software defined networks;
- Developing PRAGMA Students as a group to promote research and leadership opportunities for students in PRAGMA;
- Expanding community building in areas of biodiversity, environmental observing, and big data, among others;
- Migrating technology, including that developed by PRIME students, to the public via museums.



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

Over the next five years, our success will be measured by the extent 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** who are able to participate in and lead networked groups; **expand participation** in PRAGMA and more generally in networked science; and ultimately **enable small-to-medium size international groups to make rapid progress in conducting research and education** by providing and developing international, experimental cyberinfrastructure.

Dedication

We wish to acknowledge the major contribution of Cindy Zheng (in adjacent picture, on left, with Teri Simas at PRAGMA 24), who was instrumental in building PRAGMA and who was the "glue" that helped keep the infrastructure together. While she enjoyed the PRAGMA community very much, she reminded us that there is a time for everything. She has retired, but she will stay in our hearts and minds as part of the PRAGMA family. This year's Collaborative Overview is dedicated to her hard work over many years.

PRAGMA HIGHLIGHTS

PRAGMA BIODIVERSITY EXPEDITION

PRAGMA Biodiversity Expedition and Trust Envelopes

The PRAGMA Biodiversity Expedition is designed to address compelling biological science questions related to the distribution of and adaptation to extreme environments. Mount Kinabalu (4095 m), in Sabah, Malaysia, is a biodiversity rich, yet extreme environment located in the northern part of Borneo. The mountain is marked by numerous ultramafic (serpentine) outcrops where the soil and substrates are high in iron, magnesium, nickel, and other metals, which creates an environment that is toxic to many plant and animal species. Studying biodiversity on Mount Kinabalu allows researchers to tackle what has been a broad scientific challenge—understanding how plants, animals, and microbes adapt to extreme environments, a changing climate, and toxic conditions. This year, the PRAGMA Biodiversity Expedition continued to address the technical issues posed by data and application-sharing among collaborators, began acquiring satellite imaginary from Mount Kinabalu, and engaged a broader group of researchers in a dialogue about cyberinfrastructure in biodiversity research.

How do you share data and other elements of cyberinfrastructure with a virtual community when there are limitations or bottlenecks in moving the data? Such roadblocks might include sensitive data, licensing, and national and international restrictions limiting data distribution. PRAGMA Biodiversity Expedition members are collaborating with PRAGMA's Resource Working Group to examine and address some of these issues. A network trust envelope is one solution that increases data access security, facilitates data and software access across network firewalls, and provides a mechanism where those within the network have the same access to data, software, and other resources, even though they may be at distant nodes.

Over the last year, PRAGMA collaborators have focused on virtualizing the Lifemapper software application and cluster environment, developed by the University of Kansas (KU) Biodiversity Institute. This software is used to model and predict the spatial distribution of species based on known occurrences of organisms, environmental data, and satellite imagery. The expedition extended Lifemapper's computational ability from serving a dedicated cluster to enabling deployment on any high-performance computing platform. In addition to making computation more efficient and portable, the collaborators have expanded provenance information generated by Lifemapper and the compute environment and have used the Karma Provenance Tool from Indiana University (IU) to more fully document the remote computer process (see Figure 1).

The expedition team has completed an effort to port the extended application to a

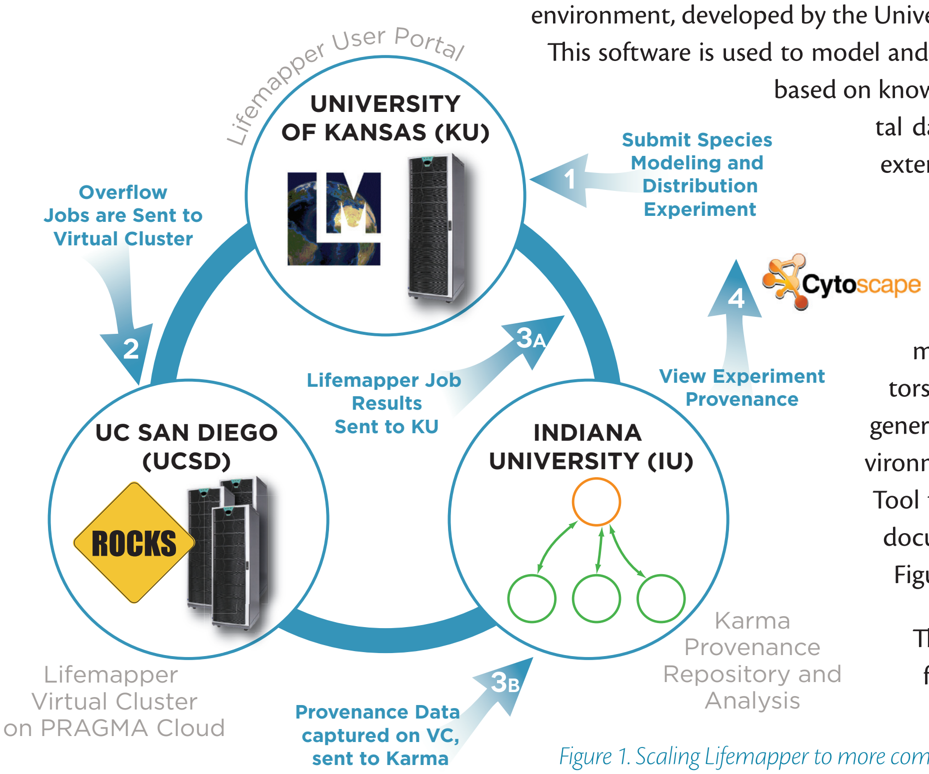


Figure 1. Scaling Lifemapper to more computational resources and capturing provenance

How do you share data and other elements of cyberinfrastructure with a virtual community when there are limitations or bottlenecks in moving the data?

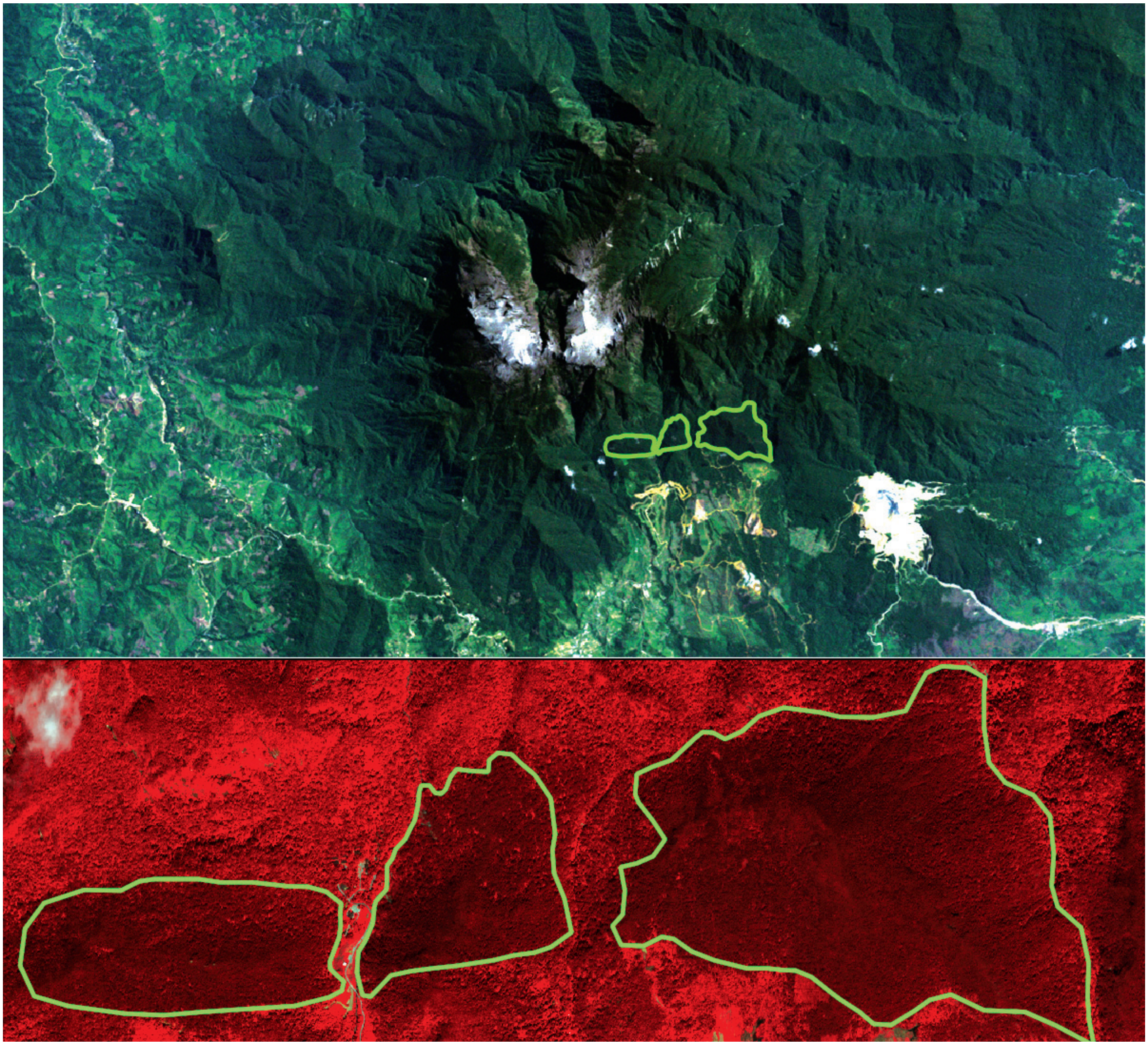


Figure 2. Top: Landsat 5 image of Mount Kinabalu (June 1991) and surrounding landscape with three of the many identified ultramafic outcrops, shown as a true-color composite of visible bands 3, 2, and 1. Ground resolution is ca. 28.5 m per pixel. The image is a rare cloud-free data capture. Below: The same three ultramafic outcrops are shown on a high resolution Quickbird (2008) image combining panchromatic (60 cm ground resolution) and multispectral bands 4, 3, and 2 (2 m ground resolution). This view is a false-color infrared composite where the ultramafic areas are easily distinguished to the eye as darker red with a smoother texture because of the smaller stature of cloud forest vegetation.

virtual environment and tested quick-and-easy spin-ups of the virtual cluster in the U.S. and Japan. These virtual clusters now join computational provenance information (e.g., hardware, software, time, error and other information) with experiment metadata (e.g., inputs, outputs, and processes for the simulations) to further document biodiversity research and enable scientific accountability. Our new ability to move the entire compute cluster to big geospatial data sets within a trust envelope (documenting processes along the way) is a big win. These efforts will allow collaborators greater ease in sharing information and resources in a controlled environment, which is essential for certain studies. Moreover, the provenance information will allow researchers to easily reconstruct the steps of a computational experiment.

In a parallel effort to build the data set, the PRAGMA Biodiversity Expedition team and its partners are working with recently acquired satellite imagery of ultramafic areas on Mount Kinabalu. The new acquisitions include partial scenes from two commercial satellite platforms, QuickBird and WorldView-2. The QuickBird imagery (Fig. 2) provides a high-resolution, 60 cm panchromatic (a single grayscale pixel represents ca. 60cm x 60cm on the ground) and four bands of 2 m data. In addition, two recently captured (June 2013) WorldView 2 scenes provide small, but cloud-free views of key areas of interest with eight multispectral bands at 2 m resolution. (Cloud free satellite data is notoriously difficult to obtain in the tropics.)

With these data, ultramafic vegetation and individual tree crowns are visible using a false color composite. Trees on ultramafic outcrops often have stunted growth forms, and these differences in canopy texture are readily detected with high resolution imagery. The team is employing quantitative analytical methods using Lifemapper, eCognition, and other GIS and remote sensing software—along with plant specimen and ecological plot data—to model species distributions, as well as organism and ecosystem characteristics. These data will be linked to environmental factors, including elevation, climate, and soil chemistry. This represents an initial effort to build a body of biodiversity knowledge in a global biodiversity hotspot through detailed analysis of which plants grow where, particularly those that are able to adapt to extreme conditions. Ultimately, we wish to know how and why some species are able to adapt and evolve over time, leading to a better understanding of multiple species, genetics, and ecosystem conservation, as well as potential societal benefits.

Finally, the PRAGMA Biodiversity Expedition is building out its network of collaborators and partners. The Mini-PRAGMA, hosted by the University of Indonesia in June 2013, included a one-day workshop on biodiversity research in Southeast Asia (see **Community Building**). The timing and location of the Mini-PRAGMA also afforded a small group of PRAGMA members (Fig. 3) to travel to Mount Kinabalu, where they met with Sabah Parks representatives and local scientists to develop plans for further collaboration. In June 2014, Sabah Parks will host the International Conference of Serpentine Ecologists, and we are developing a plan for another Mini-PRAGMA as a Special Symposium of the conference.

PARTICIPANTS: *U Florida:* Reed Beaman (lead), Andreina Weichselbaumer; *University of Kansas Biodiversity Institute:* Aimee Stewart (lead developer of Lifemapper); *Indiana University:* Beth Plale, Yuan Luo, and Quan Zhou; *UCSD:* Nadya Williams; *UCSD/SDSC:* Philip Papadopoulos; *AIST:* Yoshio Tanaka

EXPEDITION PARTNERS: *University of Queensland:* Antony van der Ent, Kasper Johansen, Peter Erskine; *Kyoto University:* Kanehiro Kitayama; *Sabah Parks:* Rimi Repin

Figure 3. Yoshio Tanaka, José Fortes, and Fang-Pang Lin on Mount Kinabalu— courtesy of Reed Beama

LAKE EXPEDITION

Learning the Phytoplankton 'Rules' to Predict Water Quality in Lakes

Predicting freshwater quality in lakes is a global challenge. Many indicators of water quality (including transparency, chlorophyll, and cyanobacterial toxins) are governed by phytoplankton species composition and biomass. At a fundamental level, phytoplankton species composition and biomass are determined by 'rules,' or mechanistic relationships that link phytoplankton species performance to abiotic (light, nutrients) and biotic (zooplankton) factors. These rules control phytoplankton dynamics and are necessary for predicting water quality. However, due to the complexity of abiotic-biotic coupling, phytoplankton communities exhibit scale-dependent responses that are challenging to predict. Even with decades of monitoring data and increased modeling capacity, predicting the identity of dominant species remains problematic because the environment in which phytoplankton exist is dynamic, and processes governing phytoplankton occur at minute to seasonal scales. The urgency to make progress in this area is great, as phytoplankton communities and water quality are rapidly changing due to altered climate and land use.

Addressing these questions requires networks—networks of scientists accessing networks of data, and recreating virtual lake ecosystems using a network of computer resources. No two lakes are alike, and so a generalized modeling framework needs calibration to the local lake, which requires local knowledge and data as well as model tuning through parameter estimation. While recreating the physical environment in lakes is challenging, simulating the phytoplankton community with enough realism to understand and predict toxic species has proven to be extremely challenging. In fact, this is so challenging that it leads researchers to question whether the 'rules' governing phytoplankton are well-enough understood.

In the PRAGMA lake expedition, scientists Hanson and Carey lead a team of limnologists who are using numerical simulation to examine the fundamental rules governing phytoplankton community dynamics. Similar to weather forecasts, lake numerical simulations recreate the physical and chemical environment in multiple dimensions, but with the added complexity of biology. For example, explosive growth of the toxin-producing cyanobacterium, *Microcystis*, requires just the right combination of physical environment, reduced resource competition from other phytoplankton species, and low abundance of zooplankton that would normally keep *Microcystis* populations in check. However, the sequence of physical-biological events that lead to blooms of harmful species remains a mystery—in this case, to be solved *in silico*. Scientists seed a virtual lake with a virtual phytoplankton community comprised of the physiological characteristics drawn from the established library of such traits. The community is allowed to develop and evolve in a virtual physical environment calibrated to observed data. Whether the evolved community does, or does not, match the observed may be a matter of chance, based on the traits initially selected. Thus, the procedure must be repeated thousands of times. Realizing this vision requires lots of computers, because each simulation can take hours to days to complete.

The researchers have traditionally used HTCondor (High Throughput Computing Condor), which manages the workload required for these simulations. However, one key challenge that has been difficult to overcome is lack of computational capacity: the software required for the researchers' Windows-based simulation environments is not available in typical HTCondor pools. Rather than investing significant effort in development, porting, and testing their workflows to

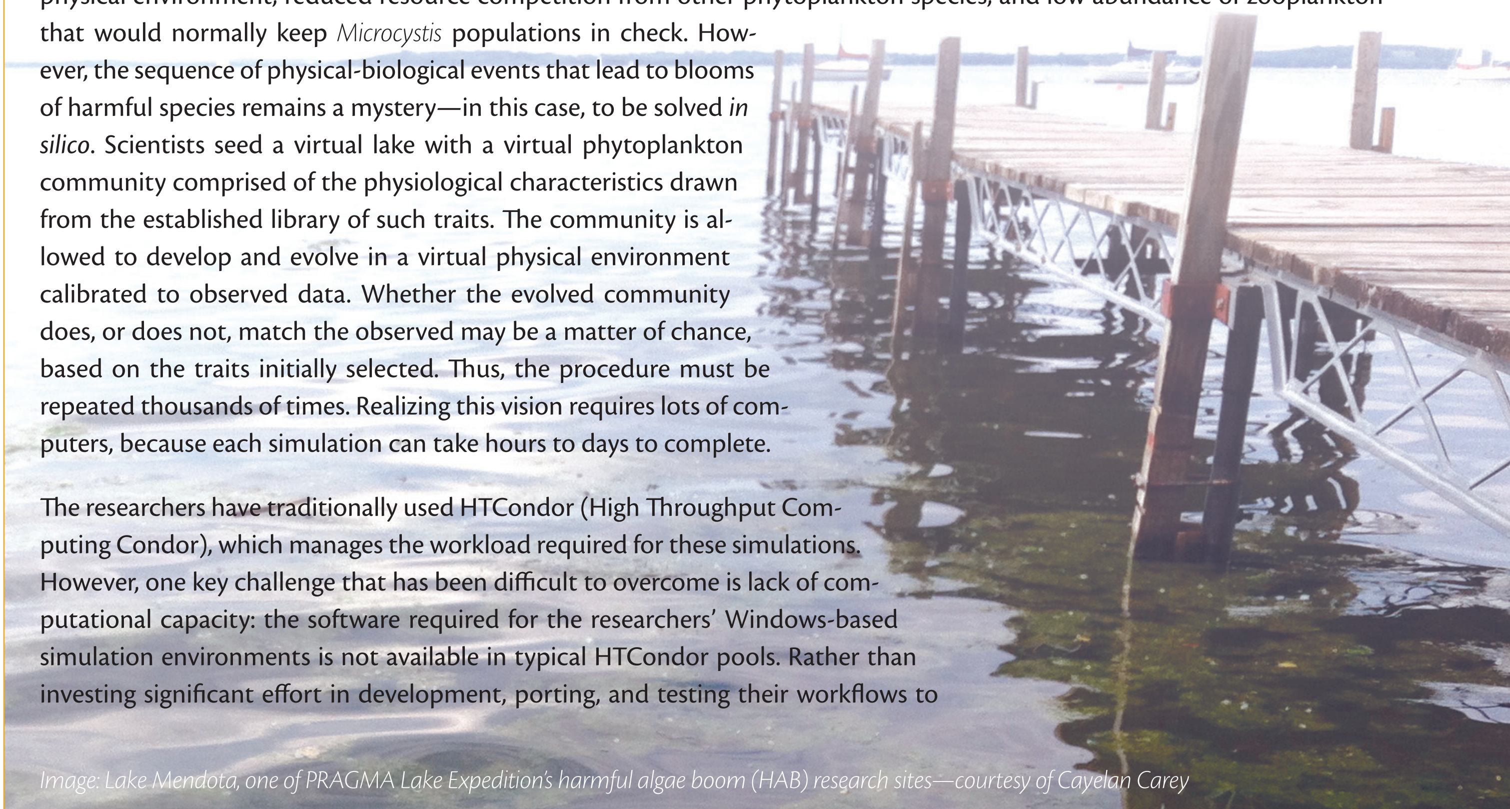


Image: Lake Mendota, one of PRAGMA Lake Expedition's harmful algae boom (HAB) research sites—courtesy of Cayelan Carey

map to a different software environment, the researchers have focused on encapsulating their simulation environment in virtual machine (VM) images. Their preferred environment can be deployed, on-demand, on cloud infrastructures.

While VMs provide a basis for customizing software environments, another challenge remains how to interconnect VMs deployed across multiple institutions and possibly commercial clouds so that HTCondor and the simulations run seamlessly? The approach to address this problem is also to apply virtualization at the network layer, with the IP-over-P2P (IPOP) software-defined network developed at the University of Florida PRAGMA site.

The open-source IPOP overlay virtual network (www.ipop-project.org) allows the researchers to define and deploy their own virtual private network (VPN) that can span physical and virtual machines in their home institution, collaborating institutions, and commercial clouds. In a pilot implementation in the summer of 2013, students demonstrated the ability for Windows VMs configured with IPOP and deployed on Amazon's EC2 cloud to automatically join a HTCondor pool that runs at the researchers' lab. This is possible because IPOP provides an isolated virtual private address space, and self-configures overlay table routes in a manner that transparently deals with firewalls and network address translators (NATs) (see adjacent Figure 4).

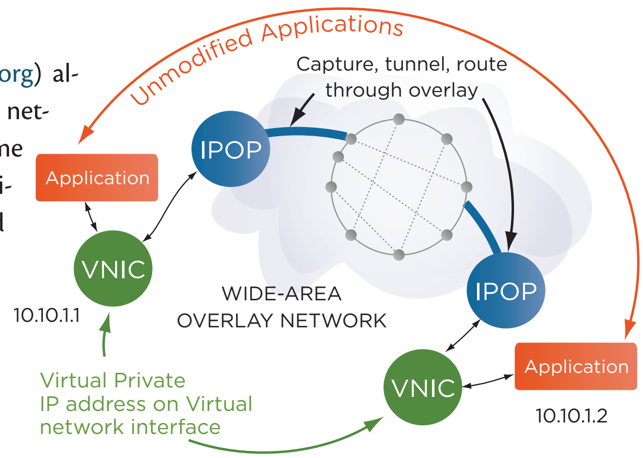


Figure 4. IPOP enables unmodified applications to connect over a self-configuring virtual private in a seamless manner. The key mechanisms used are: a virtual network interface to capture/inject packets, peer-to-peer overlay links through which virtual IP packets are tunneled, and overlay routing across wide-area networks.

How is it working? The power of networks—human, data, and computer—is paying off. Through the Global Lake Ecological Observatory Network (GLEON), Hanson, Carey and colleagues have amassed data from exemplar lakes, and the GLEON network is queuing up more. Numerical simulation software, provided by GLEON colleagues in Brazil, is being configured and calibrated to lakes. At the same time, GLEON scientists with experience in HTCondor are working with PRAGMA computer scientists from Florida to learn how overlay virtual networks can provide access to hundreds, and eventually thousands, of virtual machines to run simulations. The payoff may be a vastly improved capacity to understand and predict the diversity and dynamics of life in lakes, and thus better predict water quality in lakes.

PARTICIPANTS: *U Wisconsin:* Paul Hanson; *Virginia Tech:* Cayelan Carey; *U Florida:* Renato Figueiredo

STUDENTS: *U. Wisconsin:* Grant Langlois, Luke Winslow; *Virginia Tech:* Jonathan Doubek; *U. Florida:* Pierre St. Juste, Kyuho Jeong

Images: (above left) Cayelan Carey taking a sample from Lake Mendota; (above right) Paul Hanson lowering a light meter into Lake Mendota while Carey records the temperature profile during the GLEON 'Spring Blitz—courtesy of Adam Hinterthuer

MULTICLOUD

Progress in PRAGMA Cloud Developments

As we have seen in the two scientific expeditions, certain applications requirements (e.g., ease of scalability of resources, or ability to share data and software in a defined manner) drive technology developments. In this section, we give an overview of PRAGMA's move toward virtualization of computational resources, in particular automating the deployment of virtual clusters, and developing and using different approaches to software-defined networks. Our goal is to create user-defined trust envelopes that will allow distributed communities to work together by sharing data and simulations using their own resources.

From Grid to Virtual Clusters

In order to make PRAGMA's distributed infrastructure more practical, we have been migrating from traditional grid infrastructure to an easier-to-use distributed infrastructure over the last four years, beginning with PRAGMA 17 (Hanoi, October 2009). Our motto, "build once, run everywhere," is driving our developments, which are also based on key user needs.

We have been investigating virtualization as a practical mechanism for implementing this new approach. Our challenge was to share a virtual machine image among multiple hypervisors (i.e., virtual machine monitors) including Xen and KVM, and among Virtual Machine (VM) hosting middleware including Rocks, Ezilla, OpenNebula, and Amazon EC2. Over the past few years, our progress in sharing VM images has been shown at the PRAGMA Workshops. In the PRAGMA 20 Workshop (Hong Kong, March 2011), the usage of PRAGMA Cloud was first demonstrated. That prototype implementation shared virtual machine images on different cloud-hosting environments such as Rocks/Xen and OpenNebula/KVM. Although the demo was successful, everything was done manually, with the expectation that future implementations would automate this process.

For automatic deployment of VM images, we developed a set of scripts which were demonstrated at PRAGMA 21 (Sapporo, October 2011). At PRAGMA 21, we demonstrated that VM images could be booted at four sites by the provided scripts: UCSD, AIST, NCHC, and Amazon EC2. For more details of this early work see PRAGMA Collaborative Overview 2012-2013 (goc.pragma-grid.net/pragma-doc/overview/2012.pdf).

From Virtual Machine to Virtual Cluster

The automation of VM deployments contributed greatly to growing the PRAGMA Cloud. Within a few weeks, the number of PRAGMA Cloud sites grew from three to 13 and formed a rich variety of VM hosting environments. As the next step, we shifted our focus from sharing a single VM image to sharing a Virtual Cluster (VC) image, which consists of VM images of a front-end node and a compute node. Sharing virtual clusters among different cloud systems is a complex problem that requires retaining both the relationship between front-end and compute nodes and many cluster functions.

Three pilot sites (UCSD, AIST, NCHC) had developed an automated implementation of VC porting. At PRAGMA 24 (Bangkok, March 2013), the pilot teams demonstrated that a VC was booted at different Cloud hosting environments, including Rocks/Xen, OpenNebula/KVM, and Amazon EC2, using the same VC image. Although the demo at PRAGMA 24 was successful, we decided to redesign the *pragma VC boot* script for better portability to other hosting environments. By reviewing script files developed for Rocks and OpenNebula, we re-designed and re-implemented *pragma_boot* to be driven by three subscripts, *fix_images*, *allocate*, and *boot* subscripts, in this order. We can then adapt these subscripts to each VM hosting environment while retaining the overall *pragma_boot* workflow. We have working examples for Rocks and OpenNebula.

The *fix_images* script prepares the given VC images to be run on the current system (fix kernel, drivers, boot options, for current platform, etc.). The *allocate* script takes care of verifying that there are enough resources to satisfy the user request. If so, it will also allocate public IP, private IPs, MAC addresses, and computing resources. If the system can create SMP (symmetric multiprocessing) nodes it can allocate less compute nodes with multiple CPUs in each node. If the *allocate* script is successful it will write a /root/vc-out.xml file inside

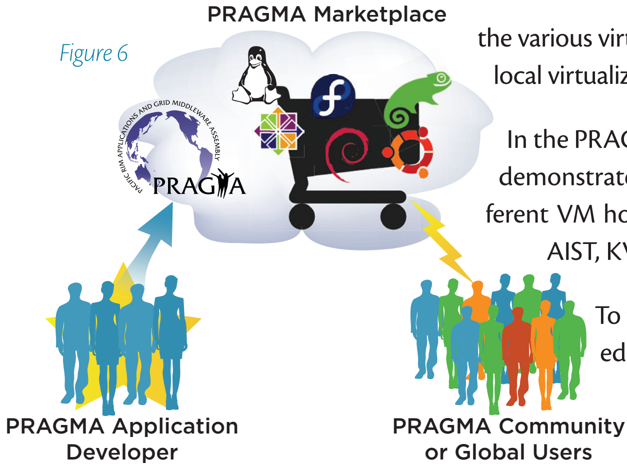


Figure 6

the various virtual machines images. The *boot* script takes care of starting the VM on the local virtualization engine.

In the PRAGMA 25 Workshop (Beijing, October 2013), VC sharing automation was demonstrated. In particular, we showed that VC images could be booted in different VM hosting environments such as Xen/KVM at UCSD, KVM/OpenNebula at AIST, KVM/Ezilla/OpenNebula at NCHC.

To further enhance the automatic feature of PRAGMA Cloud, NCHC devoted its efforts on the development of application Marketplace. The Marketplace provides a metadata service to bridge Cloud users and the VM images available. The VM images registered on the Marketplace can be stored in the physical storage preferred by the owner of the image, and various VM repositories can be federated together with flexibility. In addition, a small tool is available for the community to convert the VM format before or after the registration of the VM onto the Marketplace to help the interoperability of clouds.

PARTICIPANTS: *NAIST:* Pongsakorn U-chupala, Kohei Ichikawa; *UCSD:* Nadya Williams, Luca Clementi; *UCSD/SDSC:* Philip Papadopoulos; *AIST:* Yoshio Tanaka, Akihiko Ota; *NCHC:* Chi-Ming Chen, Kuo-Yang Cheng, Weicheng Huang

Scientific Applications on Virtual Machines and Clusters

Our technology development is intertwined with application needs. This year we highlight two areas of application for which there have been good collaborations between developers and application users.

As noted in the PRAGMA Biodiversity Expedition, Lifemapper, developed at the University of Kansas, is a software package that is used to help predict species distribution. Within the PRAGMA context, parts of the Lifemapper code are being prepared and will be available as a virtual cluster (also demonstrated at PRAGMA 25). For more about this activity, see the PRAGMA Biodiversity Expedition highlight.

We also have been working on issues associated with docking compounds to active sites of proteins. Large-scale molecular docking experimentation is a necessary part of drug discovery, and requires large computing resources. However, using heterogeneous computational environments characteristic of a Grid can yield inconsistent results. This inconsistency arises from the fact that the docking program produces different results depending on the version of operating systems, architecture, and C libraries (e.g., glibc). These discrepancies, which occur over sequentially repeated molecular docking experiments, will significantly alter the results and affect the outcome of future work , if used as the primary means for identifying potential drugs. Virtualization helps address this issue of inconsistency. More about this work can be seen in the PRIME Virtual Cluster Docking project in the Students section.

Network Overlay

As noted above, PRAGMA and its partners in the community have made great process in developing approaches to virtualizing computational resources, allowing for scalability and flexibility. Moreover, PRAGMA activities have shown how to easily port codes to other platforms. This allows us to perform large scale computation—such as large data analyses and simulations—in a dynamic manner.

Despite the development of virtualization technologies in computing resources, how best to build virtual networks is still an open issue. PRAGMA has pursued several approaches to software-defined networking:

- IP-over-P2P, IPOP, as described in the Lake Expedition
- ViNe, developed at U Florida
- OpenFlow, for controlling network routing

We describe the latter two in the subsequent sections.

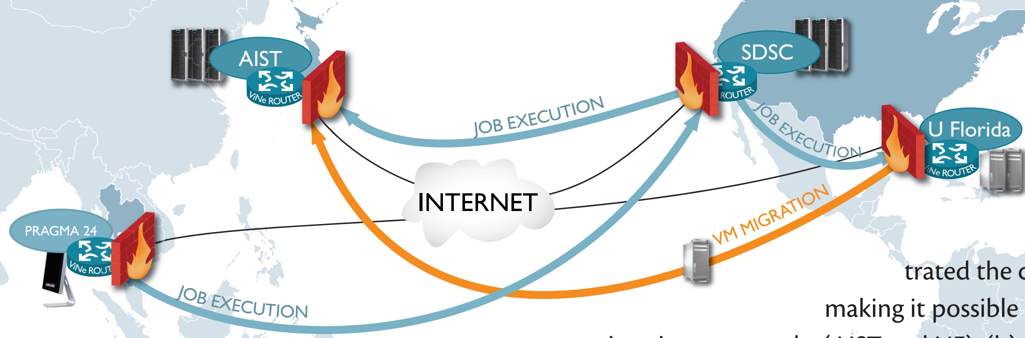
VIRTUAL NETWORKS: ViNe

The ViNe Approach to Software-Defined IP Overlays

ViNe is a project developed at the University of Florida that implements routing and other communication mechanisms needed to establish wide-area IP (internet protocol) overlays. Not all PRAGMA resources (physical or virtual) can be configured with publicly accessible IP addresses, and ViNe can offer an overlay network solution to enable communication among resources on public and private networks without the need for reconfiguring the physical network infrastructure.

ViNe has two main components: the ViNe Infrastructure (consisting of ViNe routers or VRs, with a focus on the ability to re- establish connections if a disruption occurs, as well as to establish means to traverse firewalls (i.e., connectivity recovery/tunnel establishment and fast transport of overlay packets) and the ViNe Management (consisting of an overlay management system that is responsible for the operation and reconfiguration of VRs). Both components provide application programming interfaces (APIs) that enable self-management of virtual networks (e.g., ViNe Management invokes VR APIs to configure new overlay routes for the deployment of a newly defined virtual network). They also provide software defined IP overlays (e.g., changes in overlay topology can be initiated by end users, cloud middleware, and/or application software by invoking ViNe management APIs).

With instructions prepared by the UF team, the UCSD and IU teams were able to quickly deploy ViNe technology in their clusters managed by Rocks cluster software. An additional cluster at AIST was also "ViNe-enabled" to run initial experiments. VM live-migration and the HTCondor workload management system were selected as sample applications to run on ViNe overlays.



VM live-migration demonstrations illustrated the connectivity enabled by ViNe overlays by making it possible to (a) remotely access physical machines in private networks (AIST and UF); (b) establish file sharing (NFS) through ViNe overlays; and (c) move VMs between AIST and UF without stopping them.

Figure 7. ViNe established the communication among physical and virtual resources on private networks at UCSD, UF, AIST, and PRAGMA 24 Workshop site (Thailand). This setup allowed live wide-area VM migration experiments and the deployment of a Condor pool across multiple countries.

HTCondor was set up using VMs hosted at UCSD, AIST, UF, and IU and connected through ViNe overlays. Demonstrations during PRAGMA-24 highlighted the "software definition" aspects of ViNe: (a) ViNe software was quickly deployed on a laptop in Thailand; (b) a VM running on the laptop was connected to the overlay and joined the Condor pool; (c) through a command-line interface, it was shown how quickly and easily a VM can move from one ViNe overlay to another isolated ViNe overlay; and (d) it was shown that it is possible to easily change the set of ViNe management commands/interfaces to support different overlay programming needs.

We are planning to expand ViNe deployment to reach more PRAGMA sites and offer user-controlled isolated overlays for different applications.

PARTICIPANTS: *UF:* José Fortes, Maurício Tsugawa; *IU:* Yuan Luo; *UCSD/SDSC:* Philip Papadopoulos; *UCSD:* Cindy Zheng, Nadya Williams, Luca Clementi; *AIST:* Yoshio Tanaka

VIRTUAL NETWORKS OPENFLOW

A Network Performance-Aware Routing for PRAGMA Cloud

As noted above, there is a great deal of emphasis in PRAGMA on virtualizing computing, and in particular building virtual computing clusters. To build a virtual computing cluster from aggregated computing resources requires network virtualization technologies that provide a virtual private network for each user. We have therefore focused on the recently emerged concept of Software Defined Networks (SDN) and the open standard OpenFlow technology as a possible solution that allows us to control network resources in a dynamic manner with software-based technologies.

Virtual network technologies are helpful in establishing network connectivity among computing resources by hiding the topologies and policies of the physical networks. However, hiding topologies and policies often causes performance problems in the virtual network. Since mapping between the actual network topologies and the virtual topologies is hidden, it becomes difficult to optimize network routing to increase network performance. This poses a significant problem, especially when a virtual network is established over widely distributed computing environments like the PRAGMA Cloud.

We have therefore introduced an OpenFlow-based, network performance-aware routing method, which controls network routing automatically by taking into consideration the observed inter-domain network throughput and latency, i.e., how fast or slow messages pass between networks, a key measurement for any real or software defined network. The premise behind this method is to allow us to keep track of dynamic network topologies, network throughput and latency, and then allocate appropriate network resources based on the demand of user applications for virtual computing environments.

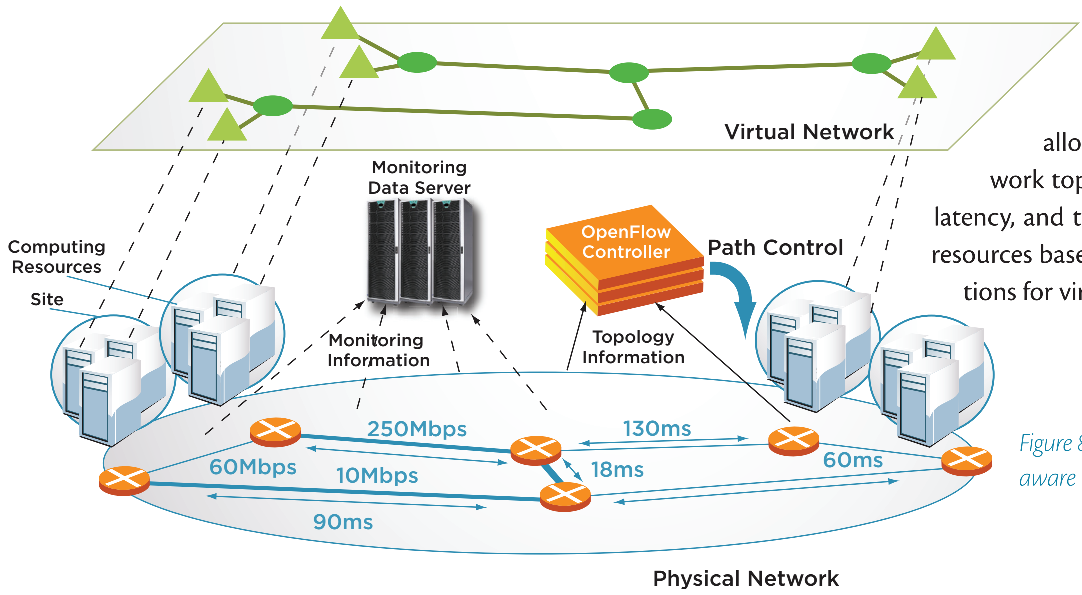


Figure 8. Network performance-aware routing system

Images: (top, left to right) One in the collection of 728 ancient murals in The Long Corridor along Kunming Lake in the Summer Palace—courtesy of Teri Simas; Downtown Osaka—courtesy of Jesus Rios; (bottom) Working session at PRAGMA 24—courtesy of Teri Simas

As a prototype, we have developed an OpenFlow controller that controls the network routing based on the observed network throughput and latency (see Figure 8). We have expanded the OpenFlow network environment deployed at PRAGMA 22 and constructed an experiment environment consisting of five organizations in PRAGMA Cloud as illustrated in Figure 2. Network throughput and latency observed by Netperf (software that provides network performance between sites) and ping (a tool to determine if a specific address is reachable) at a certain point in time are also indicated in Figure 9. As shown, the performance of each link differs. At PRAGMA 24, we successfully demonstrated the network performance-aware routing controller taking optimal paths for virtual networks in this experiment environment.

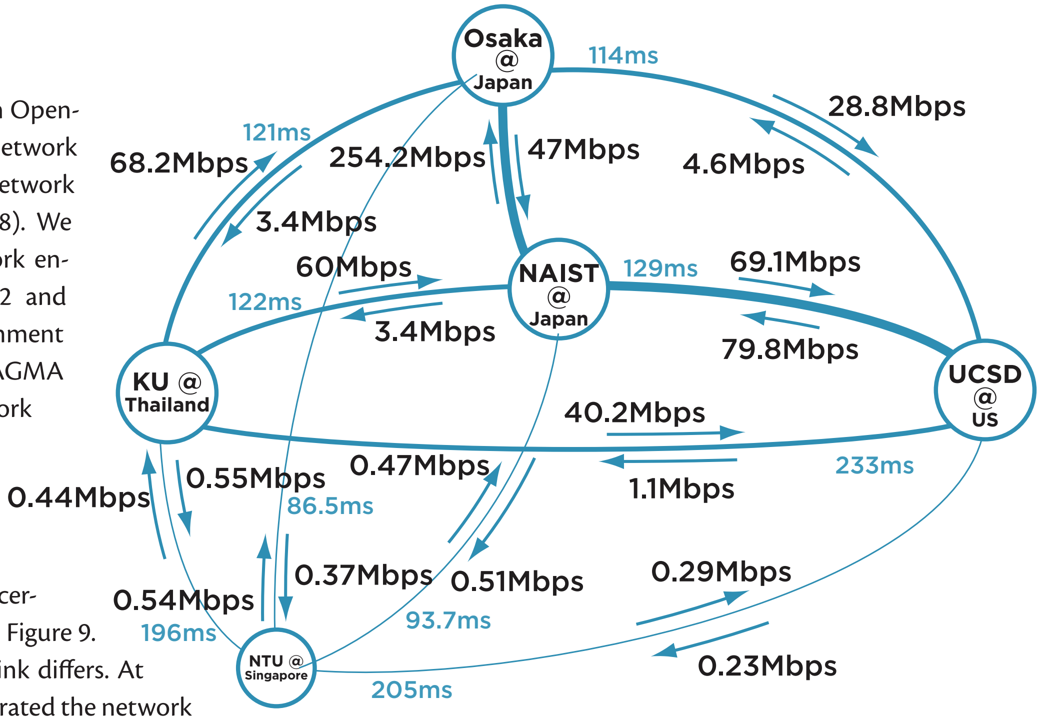


Figure 9. Network performance of the experiment environment

Following PRAGMA 24, we have improved the implementation of this OpenFlow controller. For the network performance-aware routing, the measurement method of network performance is an important key factor. However, the prototype implementation of the network measurement at the time of PRAGMA 24 was not sophisticated enough. We therefore have decided to include a distributed network monitoring system, "OverLoad," developed by the Kasetsart University team, into our OpenFlow controller. For this development, we have formed an international development team.

PARTICIPANTS: *NAIST:* Kohei Ichikawa; *Osaka University:* Taiki Tada, Susumu Date, Shinji Shimojo; *Tsukuba University:* Hirotake Abe; *NICT:* Hiroaki Yamanaka, Eiji Kawai; *AIST:* Yoshio Tanaka, Akihito Ota, Tomohiro Kudoh; *UCSD/SDSC:* Philip Papadopoulos; *UCSD:* Cindy Zheng; *Kasetsart University:* Nawawit Kes, Pongsakorn U-chupala, Putchong Uthayopas; *Nanyang Technological University:* Zoebir Bong, Derrick Lim, Bu Sung Francis Lee

EXPEDITIONS SENSOR OBSERVING

The OSDT Android SensorPod: A Smart Controller for Wireless Sensor Networks

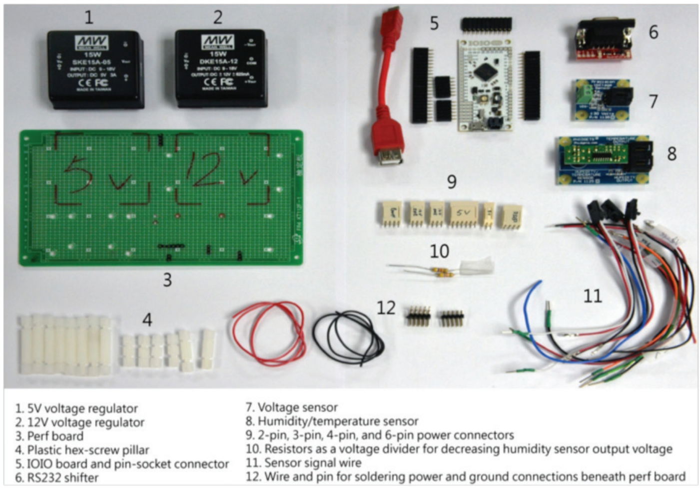
The integration of mobile and cloud-computing infrastructure represents a disruptive technology in the field of environmental science, since basic assumptions about technology requirements are now open to revision. It allows established facilities to plan graceful migrations as they adapt to inevitable technological change, and the immediate effects may be more apparent in communities and countries where existing infrastructure is limited, as is the case in many developing countrie s. Wireless sensor networks and virtualization of data and network services is the future of environmental science infrastructure. The OSDT Android SensorPod was designed with these considerations in mind.

The OSDT SensorPod is a custom-designed mobile computing platform for assembling wireless sensor networks as part of environmental monitoring applications. Funded by an award from the Gordon and Betty Moore Foundation, the UCSD's OSDT SensorPod represents a significant technological advance in the application of mobile and cloud computing technologies for near-real-time applications in environmental science, natural resources management, and disaster response and recovery. It provides a modular architecture based on open standards and open-source software that allows system developers to align their projects with industry best practices and technology trends. It also allows researchers to avoid being locked into expensive proprietary software and hardware systems. The OSDT Android SensorPod is managed as an open source project within the UCSD Open Source DataTurbine Initiative (www.dataturbine.org). As such, it depends on the generous contributions of a broad international community of users and developers.

SensorPod Assembly and Applications

The OSDT Android SensorPod was designed to be relatively quick and easy to assemble from common and inexpensive commodity products.

The OSDT Android SensorPod has been connected with various sensors and deployed in a variety of environments for diverse environmental monitoring applications, including lakes, forests, and oceans. It has demonstrated its stability with a six-month run at the Scripps Pier in UCSD and a four-month run at Sparkling Lake in Wisconsin (pictured below). Furthermore, as noted in the section on **Community Building**, it has been designed for ease-of-construction, and others have assembled it and deployed it at partner sites in Thailand and Taiwan.

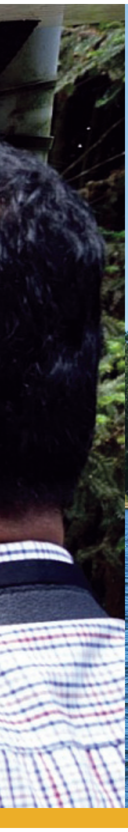


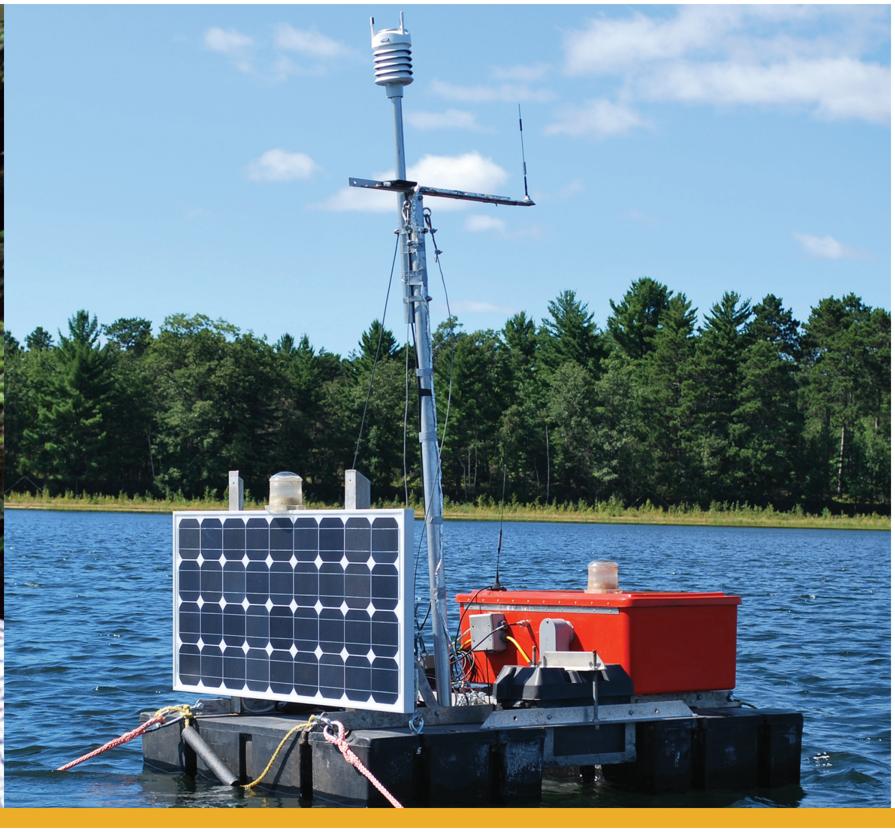
Components in the OSDT Android SensorPod—courtesy of Yu-Huang Wang

The SensorPod deployment at the Medical Plant Garden in Lienhuachih Research Station in Taiwan is part of a larger Taiwan Forestry Research Institute (TFRI)-observing system for studying biodiversity, climate change, and phenology in a forest environment. TFRI staff and UCSD students installed this OSDT Android SensorPod deployment on August 13, 2012. This deployment included a Vaisala WXT 520 meteorological station with sensors for air pressure, humidity, temperature, rainfall, wind direction, and wind speed.

For more activities related to OSTD SensorPod, see both **Community Building** and **PRIME Environmental Observing** in the **Students** section.

PARTICIPANTS: *UCSD:* Tony Fountain, Peter Shin, Thaddeus Trinh (PRIME 2013), Sameer Tilak, with Ariana Tsai and Sara Taghizadeh (both PRIME 2012); *U Texas El Paso:* Gesuri Ramirez; *Erigo Technologies:* John Wilson; *Cycronix:* Matt Miller; *TFRI:* Chau Chin Lin, Yu-Huang Wang, Sheng-Shan Lu; *U Wisconsin:* Tim Kratz, Tim Meinke, Ken Morrison, Jordan Reed; *SIO:* Jennifer Smith, Nichole Price, Susan Kram





Images: (left to right) OSDT Android SensorPod mounted inside the open white box at the Medical Plant Garden at Lienhuachih Research Station in Taiwan. It is powered by solar panels and communicates via Wi-Fi to a DataTurbine server at the TFRI station—courtesy of Wei Cheng

OSDT Android SensorPod deployed on a lake buoy in Wisconsin, USA. The OSDT Android SensorPod is inside the orange box. There are underwater sensors beneath the buoy and additional sensors above water (the weather station at the top of the pole). Data is sent via cell signal to a DataTurbine server running in the Amazon.com EC2 cloud. This system was deployed in 2012 and 2013, and used in an experiment to manage invasive species in a lake in Wisconsin—courtesy of Grace Hong

JAPANESE GARDEN

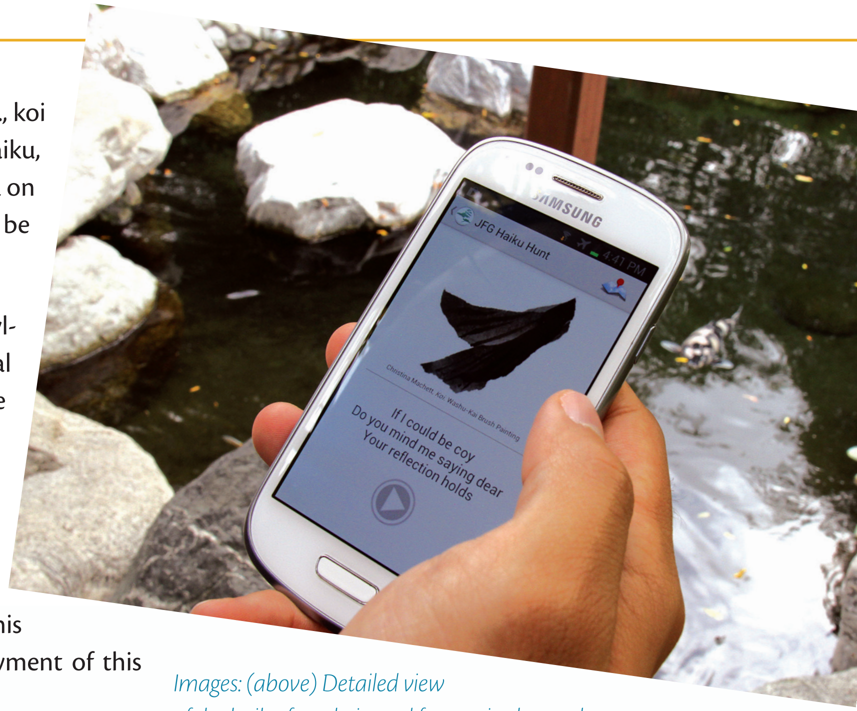
San Kei En Project (Japanese Friendship Garden Haiku Hunt)

Deployment of location-based apps in outdoor spaces can be difficult, especially in gardens with varied terrain that can obstruct Wi-Fi signals. Additionally, signage alerting users to these services can be problematic because it interferes with the aesthetic goals of Asian gardens. This project created the first outdoor, location-aware mobile app using the Place-Sticker technology developed by Information Services International-Dentsu LTD, Tokyo, Japan (ISID). It continues the productive collaboration between Balboa Park Online Collaborative (BPOC), the National Institute of Information and Communications Technology (NICT) in Japan, and the Pacific Rim Undergraduate Experience (PRIME) program at UCSD. Two PRIME students from 2012, Michael Yao and Shaocong (Scott) Mo, developed the first iteration of this application, which was subsequently refined and launched at Balboa Park's Timken Museum of Art in April 2013. The Timken was the first museum in the United States to feature an Android-based mobile gallery guide using this type of experimental location-based technology. The project was highlighted in The Center for the Future of Museums (CFM) dispatches, a nationally recognized forum that assists museums in shaping a better tomorrow by exploring cultural, political and economic challenges.

PRIME 2013 student Jesus Rios extended this Android-based application by developing an interactive game that would lead visitors through the San Diego Japanese Friendship Garden (JFG) while providing additional metadata content not readily available to the public. Haikus, crowd-sourced through an online contest, provided the "riddles" visitors had to solve to find specific locations. Visitors to the garden begin by downloading the app to their mobile device and are presented with abstract Sumi-e drawings representing the ten locations featured in the app. The visitor taps on an image and a haiku appears with a detail of the Sumi-e drawing, as well as sound clues (when appropriate). The visitor must solve the clues and haiku to find the location. A Place-Sticker device is assigned to each garden feature and will send low-power radio signals to visitors' mobile devices signaling their loca-

tion. When the visitor finds the particular garden feature (e.g., koi pond, cherry tree grove, dry waterfall, etc.) described in the haiku, supplementary information (text, audio or video) is displayed on their device. Once the visitor leaves this location, the tour will be updated and the visitor can proceed to the next haiku.

This prototype system was presented at The Lab in the Knowledge Capital (Umeda, Osaka, Japan). This is a large intellectual entertainment space where people are connected and where the future is conceived. Demonstrating in this space allowed us to obtain initial feedback from users on the functionality of the application. Based on this feedback, the application is undergoing minor adjustments and will be deployed at JFG in Fall 2013. The public launch will also include a new solar powered prototype version of the Place-Sticker device. This project will be the first successful development and deployment of this type of technology in an outdoor environment.



Images: (above) Detailed view of the haiku for a koi pond feature in the garden—courtesy of BPOC; (background) Wesley Hsu (PRIME 2011) using the app at the Dragon Bridge in the Japanese Friendship Garden, San Diego, California—courtesy of BPOC

PARTICIPANTS: *NICT:* Shinji Shimojo; *BPOC:* Vivian Kung Haga, Wesley Hsu (PRIME 2011); *Information Services International-Dentsu (ISID):* Kazuhiro Toda; *Ritsumeikan University:* Nobuhiko Nishio; *UCSD:* Jason H. Haga, Jesus Rios (PRIME 2013)

BUILDING COMMUNITY

New and innovative ideas are critical for the health and growth of an organization. PRAGMA works to foster an environment where thoughts can flourish through a mixture of open workshop discussions and by promoting active engagement of researchers and participants. To this end, PRAGMA focuses on developing outstanding talent and building collaborative and strategic partnerships. Expanding the PRAGMA community is a key component of our strategy to remain vibrant, attract ambitious minds and focus on new challenges. In this section we discuss several activities we have undertaken in the past year to build community.

EXPANDING THE BIODIVERSITY EXPEDITION FOCUS

One of PRAGMA's key activities is fostering international scientific expeditions by forging teams of domain scientists and cyber-infrastructure researchers, who develop and test necessary technologies to solve specific scientific questions and create usable, international-scale, cyber environments. Since PRAGMA 22 (Melbourne, Australia—April 2011), PRAGMA has been developing an expedition in the biodiversity (see **Highlights**) of ultramafic (high magnesium- and iron-oxide concentrations) outcroppings to gain a better understanding how plants, animals and microbes adapt to extreme and/or changing environments and toxic conditions. With an initial focus on Mount Kinabalu, PRAGMA's longer-term goals will include enlarging the participation of this expedition as well as expanding the scope of the expedition's focus, both in geography and scientific breadth.

The PRAGMA expedition members have organized or co-organized four workshops over the past year to increase the interest in biodiversity and to expand the focus. The four activities include:

- Workshop on Biodiversity, Ecosystems Services, and CI, hosted by Konkuk University—October 9, 2012—as part of PRAGMA 23;
- Bridging Big Data Infrastructure Workshop: Expedition on the Network Science Landscape, hosted by NCHC—December 3-6, 2012—(see below);
- Earth and Biodiversity Observing Workshop, hosted by Kasetsart University—March 23 and 24, 2013—as part of PRAGMA 24;
- Mini-PRAGMA Biodiversity, hosted by the University of Indonesia Computer Science Department on—June 5, 2013—(see **Mini-PRAGMA** below).

In each of these instances other organizations were made aware of the PRAGMA expedition. We became aware of other potential partners (e.g., Asia Pacific Biodiversity Observation Network (AP-BON); iDigBio Cyberinfrastructure; AsiaFlux), and new connections were made. Most notable has been the connection between GEOGrid and Green Network of Excellence (GRENE) on Environmental Information. Two interesting projects were discussed at the March meeting: Linking vegetation plots and satellite data to safeguard the biodiversity in tropical rain forests in Borneo, and documenting plant species distributions in tropical Asia.

Images: (left to right) Bangkok at night—courtesy of asia-trip.info; Silk lantern—courtesy of Teri Simas

Group photo from Mini-PRAGMA Indonesia

MINI-PRAGMA WORKSHOP IN INDONESIA

With the intention of engaging more researchers from Indonesia in PRAGMA activities, Professors Heru Suhartanto and Arry Yanuar of Universitas Indonesia (UI) agreed to organize a Mini-PRAGMA Workshop to share PRAGMA activities and engage researchers from Indonesia into collaborations.

The Mini-PRAGMA Workshop spanned three days—June 3-5, 2013—and was held on the UI campus in Depok. The first-day consisted of a Tutorial and Training on Cloud Computing and Software Defined Networks (SDN), led by José Fortes (U Florida) and Yoshio Tanaka (AIST), and included 37 students and researchers from UI. The second day, PRAGMA participants and researchers from Indonesia held a series of presentations and discussions. In total there were more than 50 participants. Finally, on the third day there was Mini-PRAGMA Biodiversity Workshop, which was attended by researchers from UI, the Bogar Agricultural University, Indonesian Institute of Sciences (LIPI)—Bogor, and BIOTROP SEAME (Southeast Asian Regional Centre for Tropical Biology)—Bogor. There were 47 participants. Participants represented an active biodiversity research community in Indonesia and demonstrated their own advances in IT and a willingness to partner, adopt and develop infrastructure to facilitate research and collaboration. Participants discussed possible future collaborations. Following this mini-workshop, Professor Yanuar chaired an International Conference on Medicinal Chemistry and Timmerman Award 2013, on October 29-30, 2013, http://icmcta2013.com.

In building community efforts, it is clearly important that we retain a long-term perspective. Professor Suhartanto was first introduced to PRAGMA two and a half years earlier, in December 2010, at the Southeast Asia Institute Program, hosted and organized by the National Center of High-performance Computing (NCHC), Taiwan. The first Mini-PRAGMA was held in July 2012 and hosted by the University of Hong Kong, with the goal of engaging more researchers from China in PRAGMA. We were able to realize part of that goal at the recent PRAGMA 25 Workshop with engagement from participants of the 2012 Mini-PRAGMA.

BRIDGING BIG DATA INFRASTRUCTURE

The Bridging Big Data Infrastructure (BBDI) Workshop reflects the growing interest among many disciplines in addressing data needs. The BBDI Workshop was organized by NCHC, GLEON, and PRAGMA, and hosted by NCHC on December 3-6, 2012. It brought together, perhaps for the first time, an extensive diversity of projects covering three topical areas (limnology, biodiversity, and disaster mitigation and recovery), across many international organizations, in a single meeting room (see http://event.nchc.org.tw/2012/datainfrastructure/index.php).

The setting of Huisun Forest Station allowed for outstanding collaborative interactions. Several overarching themes emerged that are relevant to community efforts to bridge big data infrastructure:

- Workflows (distinct form workflow software) are increasingly important in the conduct of science in a data-rich world.
- Training in areas of data and use of data infrastructure (including software and workflows) is an immediate need.
- Three interwoven components of science practice are 1) scientific questions, 2) the underlying infrastructure (including software) and technology development, and 3) training.

As a result of these discussions, two significant action items were implemented:

- The Taiwan Forestry Research Institute (TFRI) established a DataONE node, which allows for the discovery and persistent storage of data.
- GLEON is collaborating with DataONE on a new award, which involves building international data sharing capacity in lake sciences in the GLEON network.

EXTENDING SENSORPODS FOR ENVIRONMENTAL APPLICATIONS

Over the last decade, scientists have begun to harness the advances in wireless networking and sensor technology through deployments of sensors throughout the world. The sensors are being used to observe and understand environmental processes, structures of buildings under stress, and even the processes of the human body. To make this technology routinely usable by researchers, there are several challenges that need to be overcome, including ease of deployment of both sensor networks and sensors into existing networks, streaming data to data analysis facilities, and developing new algorithms for interpreting data, as well as models to help understand and predict processes.

In PRAGMA there has been an interest in using issues of environmental observing to drive the development of sensor deployment approaches and data streaming technologies. Among the disciplines driving applications for environmental observing are forestry research, coral reef science, lake ecology, and flood monitoring and preparation. During the April 2013 PRAGMA 24 meeting in

The Summer Palace, Bejing, China—courtesy of Teri Simas

Thailand, the cyberinfrastructure researchers from UCSD and scientists from Taiwan and Thailand recognized the need for an in-depth technical training workshop. One disruptive technology that combines mobile and cloud technology and has proven to be stable is the Open Source DataTurbine (OSDT) Android SensorPod (see **Highlights**).

The First OSDT Android SensorPod Workshop was held from July 1-5, 2013 at Lienhuachih Research Center of Taiwan Forestry Research Institute (TFRI), Central Taiwan. An overarching goal of the workshop was to enhance existing collaborations and to create new opportunities for technology sharing and scientific research in the area of environmental sciences (which are broadly interpreted as ecology, marine science, agriculture, hydrology and natural resource management). The workshop attendees focused on advanced environmental observing technology with an emphasis on using the OSDT Android SensorPod. The workshop participants shared experiences with wireless sensor networking applications and participated in a hands-on technical training course on the OSDT Android SensorPod. Furthermore, subsequent to the workshop, new sensors were deployed using these technologies, and in new areas of application (e.g., monitoring oyster farms or soils in medicinal forests). A report from the workshop is available: http://www.dataturbine.org/content/first-osdt-android-sensorpod-workshop-report-now-available.

This type of training workshop, followed by deployment and continued interaction, is a critical step to building a multidisciplinary community in environmental observing.

Future Activities Include

Lower Mekong Initiative

PRAGMA is partnering with Indiana University, the Network Startup Resource Center and National Center for Atmospheric Research (NCAR) in organizing the Building Network-Enabled International Science and Engineering Collaborations in the Lower Mekong Region Workshop, to be held in Vietnam in March 2014.

The workshop will bring together operators and managers of Research and Education Networks (RENs) at the campus and national levels; researchers from institutions within the region and from the U.S. who collaborate in three main areas of science and engineering: disaster management, climate change and forestry; and policy-makers and leaders from government and academia involved in supporting the development of cyberinfrastructure and collaborative science in their countries and in the region.

SEAIP 2013

The South East Asia Institute Program has been a forum for researchers, technologists and administrators throughout South East Asia to share ideas and look for ways to collaborate. This year's workshop will focus on Big Data, particularly in the areas of biodiversity, environmental observing, and disaster mitigation. The workshop will be held December 1-6, 2013 in Taichung Taiwan (see http://seaip.narlabs.org.tw).


STUDENTS
VISA

PRAGMA STUDENTS

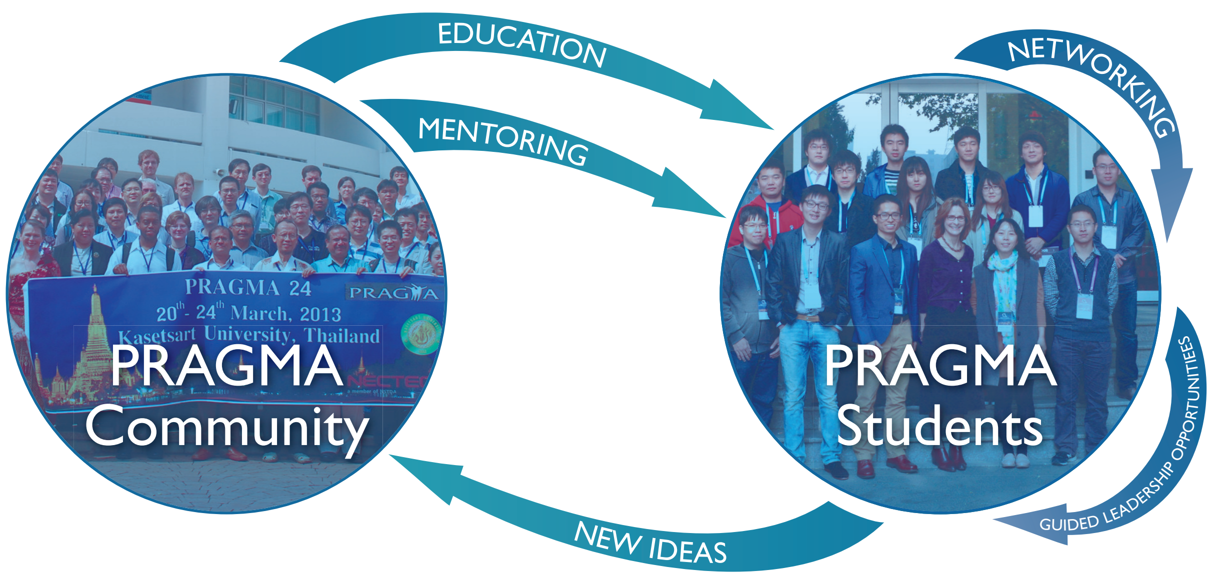
DEVELOPING THE INTERNATIONAL INTERDISCIPLINARY RESEARCHER

One of PRAGMA's key activities includes expanding opportunities for early-career researchers with experience in cross-border science. Encouraging talent and broadening participation is a strategy that ensures additional sources of new ideas and projects, which are critical to PRAGMA's long term success. In this section we describe three activities, aimed at developing international interdisciplinary researchers and future leaders—PRAGMA Students, PRIME, and MURPA/QURPA.

PRAGMA Students

PRAGMA Students was formed in 2012 to enhance opportunities for students to gain professionally from PRAGMA's trusted social and technical networks. PRAGMA Students is a student organization within PRAGMA led by a governing student body and advised by a PRAGMA researchers. Activities of the PRAGMA Students group include but are not limited to: organizing student workshops and poster sessions as part of the bi-annual PRAGMA Workshops, hosting online seminars for the broader PRAGMA community, and working together to develop a unique model to involve and lend outstanding opportunities for graduate students participating in PRAGMA's collaboratory scientific research.

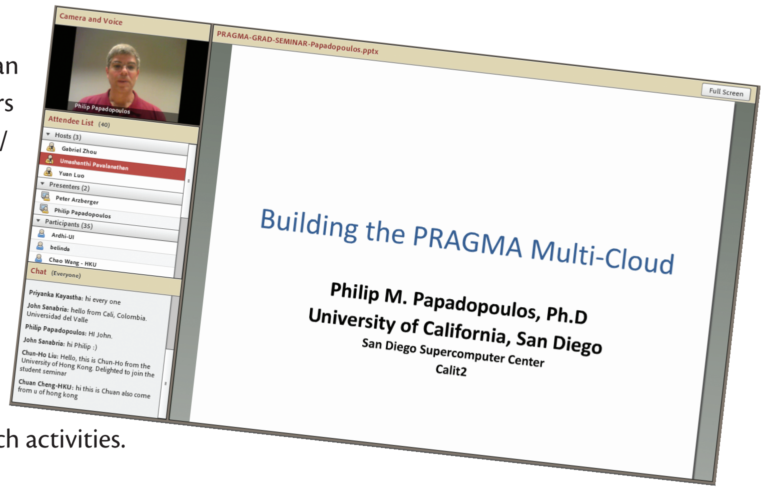
Since PRAGMA is focused on research and experimentation, engaging with the PRAGMA Student group provides students with experience and training that they cannot find at their home institutions. Experience in PRAGMA Students will yield outcomes that are on par—both in term of experience and prestige—of an internship at a big technology company or a large research lab.




PRAGMA has an extensive history as a trusted network, and fosters leadership training that helps students gain valuable experience in professional settings. Taking this into account, there are several advantages of being a member of the PRAGMA Student group. Benefits include:

- Strengthening of research and goals through collaborations and engagement with mentors and advisors, both domestic and international, outside of the students' home institutions.
- Collaboratory experiences include participation in short-term residential research opportunities in countries within the Pacific Rim.
- Participation in professional workshops and conferences in an international setting that include information sharing, innovation, collaborations and leadership.

In Spring 2013, PRAGMA Students hosted a remote lecture to an audience of more than 50 attendees, including PRAGMA members and those outside the PRAGMA community. The speaker, UCSD/SDSC's Philip Papadopoulos, discussed "Building the PRAGMA Multi-Cloud" (see http://connect.iu.edu/p1un2k6gqy6). PRAGMA Students have additionally hosted their first and second PRAGMA Student Workshop at PRAGMA 24 and 25. Finally, PRAGMA Students were responsible for organizing poster sessions for the PRAGMA 24 and 25 Workshops, which provided additional leadership opportunities while engaging them in a wide array of topics associated with international scientific research activities.



"Building the PRAGMA Multi-Cloud" webcast

PRAGMA STUDENTS LEADERSHIP: *Indiana U:* Yuan Luo, Gabriel Zhou; *Konkuk U:* Meilan Jiang

PRAGMA STUDENTS ADVISORS: *Indiana U:* Beth Plale; *Konkuk U:* Karpjoo Jeong; *Kasetsart U:* Putchong Uthayopas

Images: (above, left to right) PRAGMA Students—courtesy of Yuan Luo; PRAGMA Student—courtesy of Yuan Luo; Erickson Nguon and Anthony Adams ice climbing Fox Glacier—courtesy of Anthony Adams

PRIME

Hands-on Research Internship and Cultural Awareness Experience

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 UCSD. 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 180 students have participated in PRIME. This year, the 10th year of the program, PRIME sent 12 students to the seven sites listed at right. Two of these sites are new host sites. The first, the Nara Institute of Science and Technology (NAIST) is home to a key technical contact, Dr. Kohei Ichikawa, who had been a graduate student at Osaka University when PRIME started and enjoyed hosting students there. The second is the University of Queensland, with key contact Dr. David Abramson, who recently moved from Monash University and who has been hosting PRIME students since PRIME's inception.

PRIME added more training pre-departure for students this year as well, helping provide both a brief overview of culture through history in Japan, as well as instruction on how to prepare a scientific poster. In addition, we will host a PRIME seminar in Fall 2013 to have PRIME students present the results of their work to prospective students. The updated curriculum allows students to showcase a more detailed review of their research.

Finally, we also gave our website, prime.ucsd.edu, a much needed facelift in light of our main goal of making information easier to find for our audiences: 1) students interested in applying, 2) representatives from other programs who are interested in our curriculum, and 3) representatives from other organizations who want information regarding PRIME's goals and activities, and now they impact our PRIME students.

HOST SITES IN 2013

Nara Institute of Science and Technology (NAIST)

National Institute for Information and Communication Technology (NICT)

National Taiwan University (NTU)

Osaka University

Taiwan Forest Research Institute (TFRI)

University of Auckland

University of Queensland

We want to acknowledge last year's unexpected and generous gift from one PRIME's former alumni, Haley Hunter-Zinck (PRIME 2008, Monash University). She is currently a graduate student at Cornell University, working on a Ph.D. in computational biology and medicine. Her unrestricted gift to PRIME is a tribute to her mother, who was passionate about promoting the success of women in science. Hunter-Zinck hopes that her donation to PRIME will inspire others to follow her lead in supporting the international program. This gift is gratefully accepted for future students, and allows the program to continue to evolve to improve the experience. We also acknowledge the support of the US National Science Foundation, NICT, an award from the Japan Society for the Promotion of Science to NASIT, and of the the host sites.

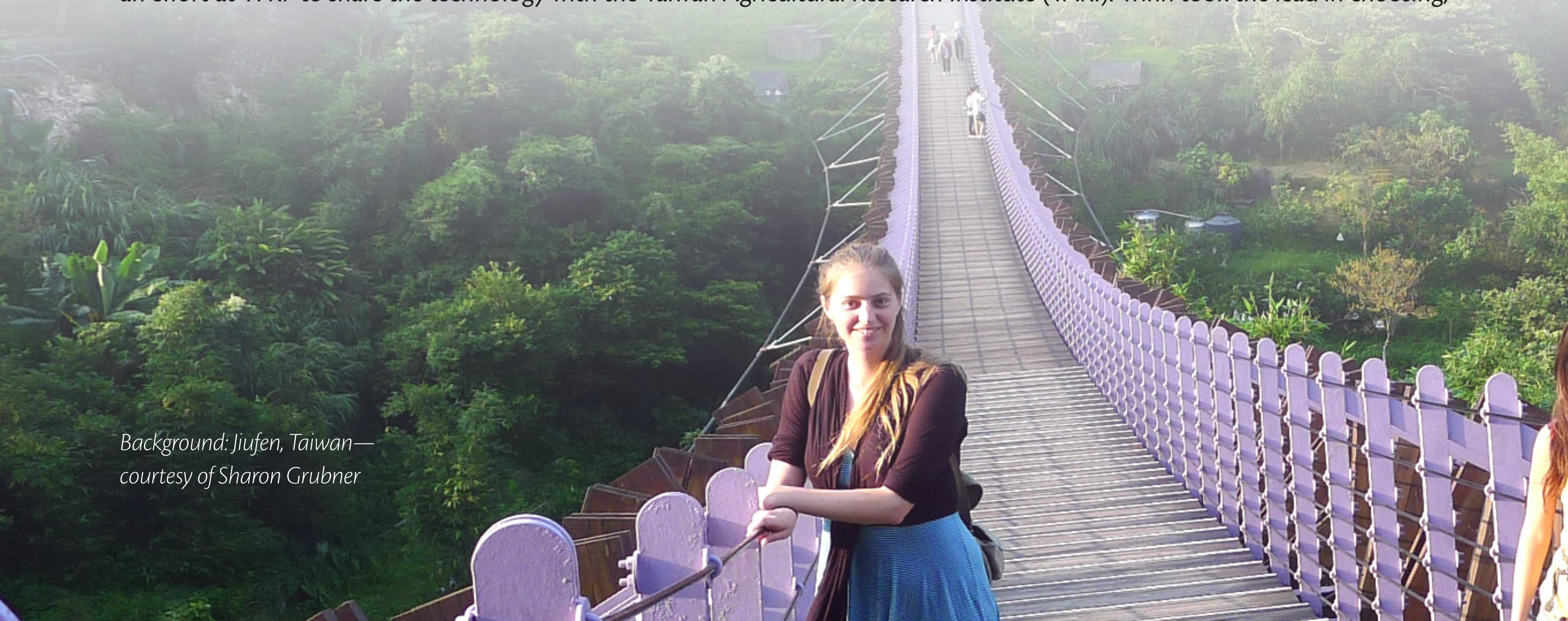
In the subsequent pages and in the San Kei En Project (in **Highlights**) we share projects and results from a selection of this year's PRIME students. These projects range from application of technology to drug discovery to the development of technology for applications in environmental observing or public display of culture.

PRIME ENVIRONMENTAL OBSERVING

Extending Environmental Observing Technology OSDT Android SensorPod to New Communities

Remote environmental monitoring systems are enabling environmental scientists and other field scientists to conduct their research in high resolution in near real-time. Although these technologies allow new scientific ideas to develop, they are challenging to adopt because of their associated high costs and the low accessibility of proprietary systems and custom radios. The Open Source DataTurbine (OSDT) Android SensorPod solves these problems by employing relatively inexpensive Android smart phones for network communication, open-source software components for extensibility, and off-the-shelf hardware components for easy repair and upgrade (see **Expedition Sensor Observing** in **Highlights**).

This technology is facilitating the growth of a grassroots international collaboration. Since 2010, the Taiwan Forestry Research Institute (TFRI) has hosted seven UCSD undergraduates through the PRIME program to strengthen the collaboration between ecologists in Taiwan and cyberinfrastructure researchers at UCSD who are focusing on environmental monitoring applications. Among the students, Ariana Tsai and Sara Taghizadeh (both PRIME 2012), helped TFRI adopt the OSDT Android SensorPod at the Lienhauchih Research Center. This year, PRIME student Thaddeus Trinh furthered the grassroots movement by assisting with an effort at TFRI to share the technology with the Taiwan Agricultural Research Institute (TARI). Trinh took the lead in choosing,



Background: Jiufen, Taiwan—courtesy of Sharon Grubner integrating, testing, and deploying new types of sensors for the TFRI site, Chia-Yi Litchi Orchard. He successfully integrated several sensors into the OSDT Android SensorPod, including Decagon 10HS Soil Moisture Sensor, Decagon CTD Water Depth Sensor, and Kipp & Zonnen SMP3 Solar Pyranometer. Although integrating the FTS FS-3 Fuel Stick Sensor was not completed during his stay, he proactively sought out advice from an electrical engineer at Linear Technology, thereby providing solutions and demonstrating his growth in engineering leadership.

Trinh also assisted in leading the hands-on technical session, and led a session on how to integrate new types of sensors into the system in the First OSDT Android SensorPod Workshop. (see also **Expanding SensorPods for Environmental Applications** in the section of **Community Building**).

PARTICIPANTS: *PRIME 2013 student:* Thaddeus Trinh; *UCSD:* Tony Fountain, Peter Shin; *TFRI:* Chau Chin Lin, Sheng-Shan Lu, Yu-Huang Wang

MONITORING NETWORKS OF SENSORS

Monitoring Networks of Sensors: Applying Cloud Tools to Sensor Networks

Wireless sensor networks (WSNs) act as gateways between the physical and the digital worlds. The increasing computing power, memory, storage and multi-modal network connectivity (Wi-Fi, Bluetooth, cellular) of WSNs, as well as decreasing prices of smart portable devices (tablets and smart phones) make them an excellent choice as a data acquisition platform for field-deployed sensors. However, due to the remote and often harsh environments in which these sensors are embedded and the large number of services (data acquisition, storage, processing, and transmission) running on these smart phones, a manual inspection of performance and service failures is not possible. To that end, the goal of this project was to monitor state-of-the-health of a deployed hardware platform (smart phone and sensors) as well as critical services in an automated and scalable fashion.

We employed the INCA (http://inca.sdsc.edu) framework to monitor failures and performance issues with services deployed on a cellphone platform running Android OS. INCA is framework developed at UCSD that detects Grid infrastructure problems by executing periodic automated, user-level testing of Grid software and services. For this project, we adapted the INCA framework to monitor services on a resource-constrained, field-deployed cellphone running Android OS that acts as a data acquisition platform. This will allow domain scientists and cyberinfrastructure developers to monitor the failures and performance degradations in services, which in turn can then be used to appropriately restart or upgrade services in a scalable manner.

PARTICIPANTS: *PRIME 2013 student:* Fabian Lema; *UCSD:* Sameer Tilak and Shava Smallen; *University of Queensland:* David Abramson

Life-sized Gundam model in Odaibo, Tokyo—courtesy of Kittinan Ponkaew

PRIME BIODIVERSITY

Improving Natural History Museum Data Collections: Informatics Tools for Biodiversity Data Conversion

Maintaining and accessing biodiversity data is an extremely important issue for many institutions. Such data is usually stored in different formats, making access difficult. Additionally, the multiple databases where the data are stored have non-uniform entry tags that make data provenance problematic.

Creating a centralized database that contains all data in a unified format addresses these problems. This project was a collaboration between the San Diego Natural History Museum (SDNHM), Balboa Park Online Collaborative (BPOC), the National Institute of Information and Communications Technology (NiCT), and the UCSD PRIME program.

SDNHM has multiple collections of taxonomy, each stored in its own database. Therefore, creating a central database containing all the information that can be updated periodically in an efficient manner is critical for the future use of SDNHM collections.

This project used Specify 6.5 (University of Kansas) as a centralized database that aggregates all the biodiversity data from various departments at SDNHM. Specify was chosen because of its ability to store more information than Darwin Core Archive (DCA), its ability to export to DCA, and the availability of premade tools and a support team to help in the process. The aggregated data can then be used to create a central database, allowing easy access to all biodiversity data from SDNHM, without having the data in different formats. Furthermore, this allows the museum to create a customized Web portal to display its data to the public, making it user friendly to all members of the public.

Python-based tools were used to convert the different data formats to a Microsoft Excel format. The data were then inspected and reformatted so that any unneeded information was removed so that the data would conform to Specify requirements. The data were then successfully imported into a single Specify database. A simple web portal was also created allowing access to this database.

In summary, this project created a centralized database for biodiversity information. The tools developed will continue to be used as a model for data aggregation and unification for other museums. SDNHM plans to expand the functionality of its Web portal and use the centralized database to provide metadata for future interactive exhibits.

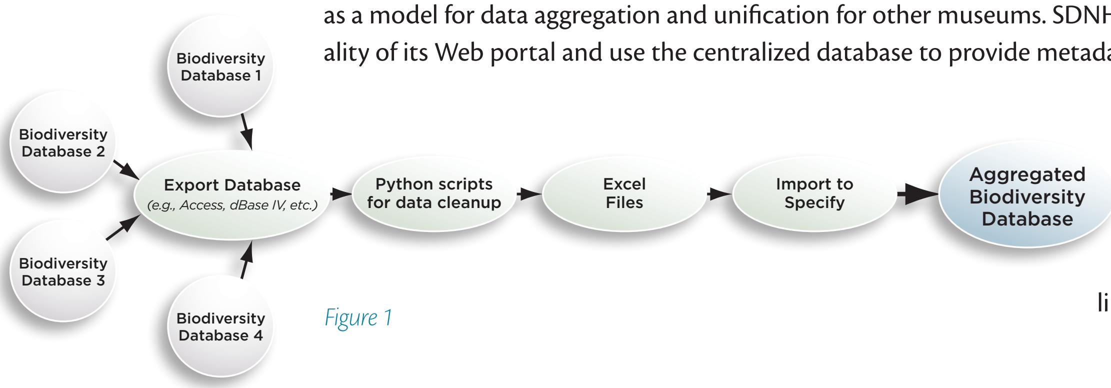


Figure 1

PARTICIPANTS: *PRIME 2013 student:* Kittinan Ponkaew; *UCSD:* Jason H. Haga; *BPOC:* Vivian Kung Haga; *NICT:* Shinji Shimojo; *SDNHM:* Michael Wall, Brad Hollingsworth

PRIME VIRTUAL CLUSTER DOCKING

Drug Discovery: Configuration and Deployment of Virtual Machines for Molecular Docking Experiments

Molecular docking experiments that simulate the interaction between a ligand and a protein receptor (i.e., virtual screening) are a popular tool used among biomedical researchers for drug discovery. Previous projects using docking simulations have been done on a physical grid-computing set-up, where each node is comprised of physical hardware located at different PRAGMA sites (each with its own environment and infrastructure).

Physical grid computing can create inconsistent results in virtual docking simulations because of the heterogeneous hardware components of the compute clusters. These inconsistent results can be resolved, however, by using virtual machines in a cloud computing environment. Thus, the goal of this project was to create virtual machines with identical configurations and to deploy them in PRAGMA's cloud testbed to emulate a completely homogeneous physical grid, thereby creating a computing environment that can yield consistent virtual screening results.

Various KVM-based virtual machines were created with different operating system configurations, compliers, and architecture platforms. Each configuration was installed with DOCK (a popular software for docking, i.e., aligning optimally, a ligand with a protein receptor), and our team screened a set of standard ligands provided with the DOCK package. We compared the results to the results provided in the DOCK package and to the findings from a previous PRIME student's project (Yim, et al. 2010). The configuration that yielded the same results was identified as the best configuration.

The virtual machine configuration selected had a 32-bit architecture equipped with Cent OS 5.9 and gcc (GNU Compiler Collection) complier version 4.1.2. This image was then moved onto the PRAGMA Cloud testbed using an AIST-prepared OpenNebula-based virtual infrastructure and UCSD-prepared Rocks-based virtual infrastructure. Both environments had the new *pragma_boot* tool available (see **Multicloud** in the **Highlights** section). Preliminary testing demonstrates that this virtual machine is functional, but requires some optimization. Once the virtual cluster is established on the cloud, future virtual screening research will be simplified, will produce more consistent results, and will streamline the entire virtual screening workflow.

PARTICIPANTS: *PRIME 2013 student:* Kevin Lam, Karen Rodriguez; *UCSD:* Jason H. Haga, Nadya Williams; *NAIST:* Kohei Ichikawa

Yim, W-W, S Chien, S Date, **JH Haga.** Grid Heterogeneity in *In-silico* Experiments: An Exploration of Drug Screening Using DOCK on Cloud Environments. *8th HealthGrid Conference, Paris, France,* DOI: 10.3233/978-1-60750-583-9-181, 2010.

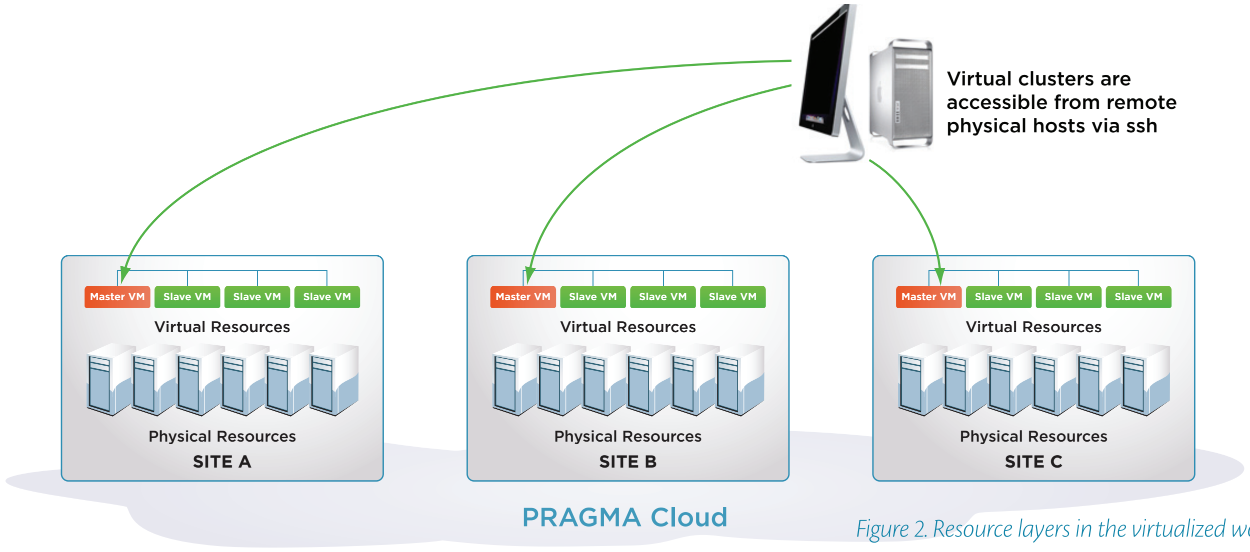


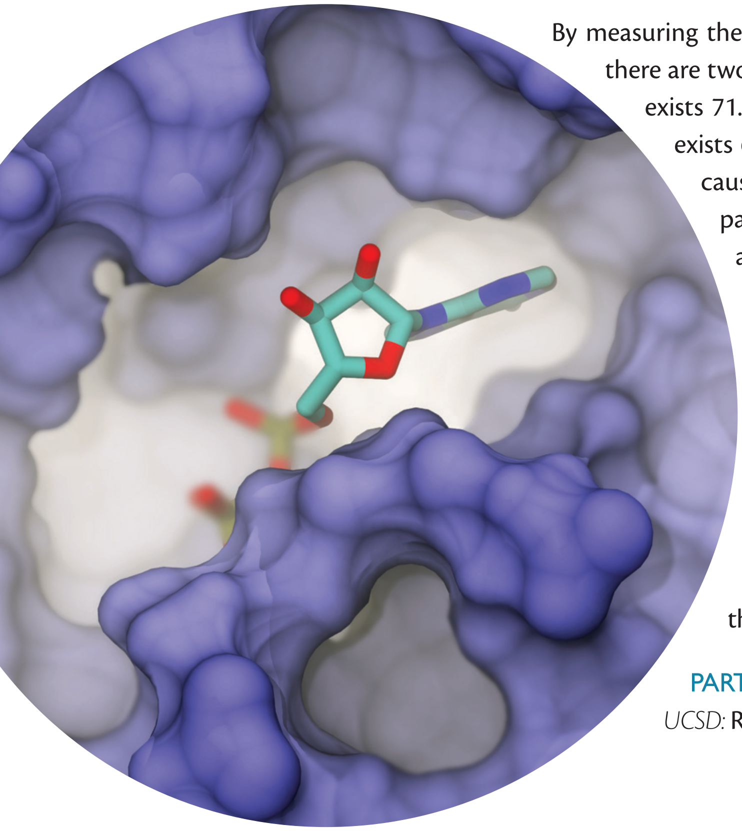
Figure 2. Resource layers in the virtualized workflow

STRUCTURAL INVESTIGATION

A Step Toward Understanding Raine Syndrome: Structure Investigation of Fam20C, a Golgi Casein Kinase

Raine syndrome is a rare and devastating disease that involves osteoclerotic bone dysplasia in humans. This syndrome is characterized by ectopic calcification, which increases bone density and often causes death in neonatal stages, because growing bones are too brittle to form properly. The broad aim of this research project was to investigate the structure and function of Fam20C, the first kinase that was found to be secreted to the exterior of the cell. Fam20C is important because mutations of this protein can lead to Raine syndrome. Studying the protein's structure, and specifically its ATP binding pocket, will allow researchers to better understand its function, better control it for experimental purposes, and use that information for drug discovery efforts in the future.

While crystal structures provide a static "snapshot" of the molecular machinery in the cell, proteins are known to be highly dynamic entities. The approach of the project is to conduct molecular dynamics simulations of Fam20C in order to reveal how this critical protein moves *in vivo*. Understanding the dynamics of Fam20C could be important not only for understanding how the protein works, but also for potential drug discovery applications. The simulation of the unbound form was carried for 200 ns and carried out with FF99SB Amber force field. Several tools were used to analyze the structure of Fam20C once the simulation was complete: root mean square deviation and fluctuation (RMSD and RMSF) to study certain regions of the protein, bond lengths throughout the simulation to measure salt bridges in the binding pocket, VolMap VMD plug-in to calculate isovalue and water occupancy, GROMOS clustering, and FTMap and FTProd to analyze clusters of binding sites inside and outside of the pocket. All of these metrics allow us to quantify the dynamical motion exhibited by Fam20C.



By measuring the fluctuation of bond lengths throughout the simulation, it was found there are two dominant salt bridges throughout the simulation: Lys192-Glu218 that exists 71.8% of the simulation time, and Lys178-Glu213 salt bridge interactions exists 66.1% of the time. The results of the above analysis are interesting because they indicate that alternate conformation is possible between these pairs—this critical information could be exploited for drug design. The above observation also means that the binding pocket will have fluctuations in its electrostatic properties, depending on which conformation the salt bridge is.

The hinge region, involving the Phe297 side chain (i.e., an aromatic ring), acts as a gating mechanism for the binding pocket. Our analysis indicated frequent movement of the hinge's aromatic ring, which allows observation of how flexible it is for drug design purposes, allows estimation of the size of the pocket, illustrates its changes and suggests what can potentially bind to it, and shows the free energy contributions to binding that would need to be compensated by a future drug molecule.

PARTICIPANTS: *PRIME 2013 student:* Sharon Grubner; *NTU:* Jung-Hsin Lin; *UCSD:* Rommie Amaro, Victoria Feher, Lane Votapka

Figure 3. ADP-bound Golgi casein kinase, Fam20C. The kinase is shown in surface representation with ADP bound at its ATP binding pocket—using the visualization program Visual Molecular Dynamics—courtesy of Jacob Durrant and Sharon Grubner

CHAGAS DISEASE

Trans-Sialidase's Role in Chronic Chagas Disease, and its Potential for Infection Inhibition by Employing Natural Products

Currently there are 8 to 10 million people worldwide infected by Chagas Disease, a parasitic infection caused by the protozoan Trypanosoma Cruzi (*T. Cruzi*). The vast majority of infected individuals live in rural areas of Latin America with limited healthcare access. If untreated they will suffer from a lifelong infection, which could ultimately lead to severe cardiac muscle and gastrointestinal damage and eventually result in death. The malady can be successfully treated in acute phase; however, this phase often goes unnoticed. Once the malady enters chronic phase (about 90 days after infection), we lack the ideal drugs for treatment. The available drugs, which are administered in both acute and chronic phases, can cause severe negative side effects and have limited efficiency in eliminating the parasite from the host, especially once the disease has entered chronic phase. Accordingly, there has been a growing interest in developing drugs that target the protozoan directly on a biochemical level.

Several novel drug targets have been suggested, and among those is the surface protein, Trans-sialidase (TcTS), found on the membrane of the protozoan. This protein promotes host cell invasion, and allows the parasite to evade cell defenses, such as lytic processes. Furthermore, it has been suggested that it might play a role in immunomodulation, which permits the protozoan to elude immune responses and live in the host for decades, allowing it to become a chronic infection. Thus, TcTS has an important role for *T. Cruzi's* survival success.

For this study an *in-silico* approach was utilized, since the protein structure of TcTS had already been identified. The structure was docked, using both AutoDock Vina and AutoDock 4.2, to compounds in the Taiwan Pharmaceutical Databank (http://tpd.mc.ntu.edu.tw/index.php), where several candidates were identified. Dihalenaquinolide A, derived from a Taiwanese marine sponge, Petrosia elastica (Figure 4), had (among all screened natural products) a consistent low free binding energy at around -15.2 kcal/mol to TcTS's active site. Moreover, by running a perl script, it was determined that key residue interactions, such as Asp59 (which is essential for catalysis), were established between the two species. Other noteworthy compounds with consistently promising results were, (+)-Ovigeridimerin, derived from the trunk bark of Hernandia nymphaeifolia, with -13.5 kcal/mol; and Bisisodiospyrin, identified in the stem of diospyros maritima, with -15.7 kcal/mol.

Chagas Disease is a malady affecting a highly vulnerable population and great efforts have been made globally to find a cure. Some potential drug candidates have been synthetic, but more often natural products have shown to be good templates and are novel in their drug design. The identified candidates, however, are just the first step in a long process in determining their actual efficiency as an inhibitor to TcTS processes. Ultimately, in-vitro assays will have to be performed to verify the *in-silico* predictions.

PARTICIPANTS: *PRIME 2013 student:* Hanne Inez Wolff; *UCSD:* Philip Bourne, Chirag Krishna; *NTU:* Jung-Hsin Lin, Nan-Lan Huang, Pei-Ying Chu

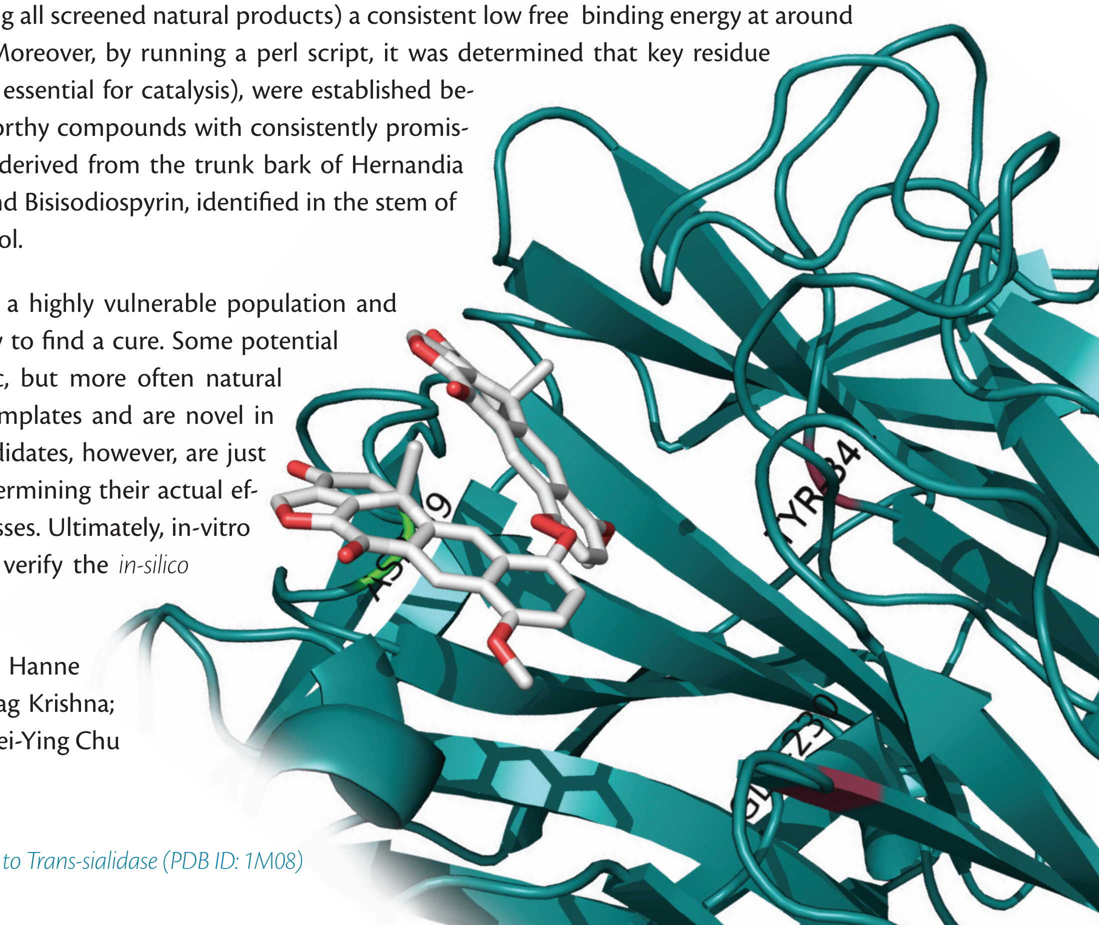


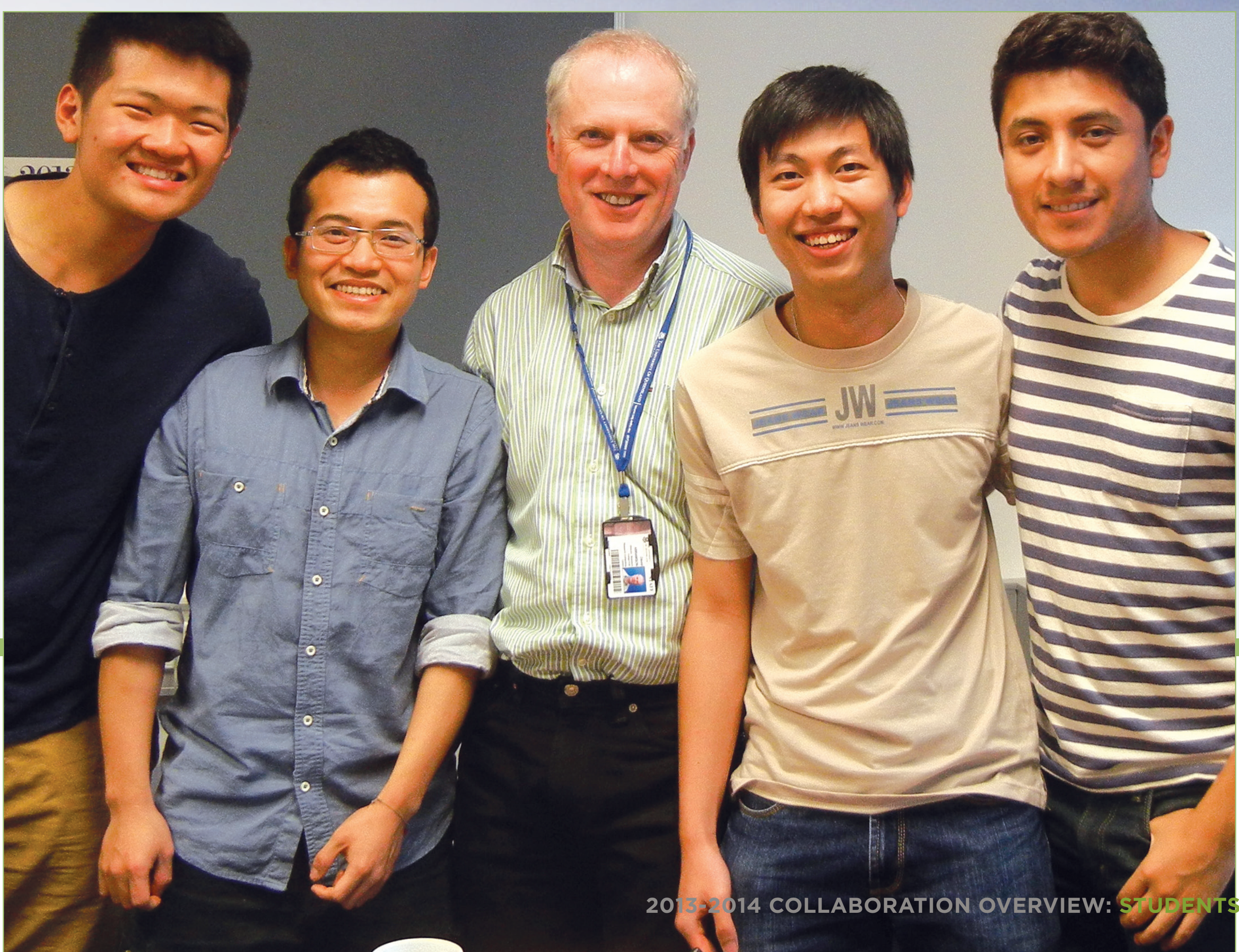
Figure 4. Dihalenaquinolide A docked to Trans-sialidase (PDB ID: 1M08)

MURPA/QURPA

International Research Internships come to Research Commuting Centre, University of Queensland

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 five years, 26 Monash University students have travelled to international partner institutions under the Monash Undergraduate Research Projects Abroad program (MURPA). They have been placed at UCSD, The National Center for Supercomputing Research in Illinois, The Technion Institute in Israel and the University of Warwick. Students are placed for a period of eight weeks, allowing them to integrate into the research groups as team members. Students have a local mentor in Australia as well as one at the remote site, and often bridge international research projects. In 2012/'13, three students travelled to UCSD and two to Warwick University in the UK, specifically working in Bayesian networks, distributed computing, wireless sensor networks, machine learning, and scientific workflows.



This year the University of Queensland (UQ) has launched the Queensland Undergraduate Research Projects Abroad Program (QURPA), and UQ students will join MURPA students for the first time. This not only allows UQ students to engage in some fascinating advanced computing projects, but opens the door to trans-national student-led collaboration in undergraduate research. We have also added collaborators at the Institute for Infocomm Research (I2R) in Singapore in addition to projects at UCSD. Projects span a wide range of advanced computing technologies and leading edge applications, and serve as incubators for research collaborations at UQ.

MURPA and QURPA involve an advanced seminar scheme, in which students can attend virtual seminars given by leading world experts. These seminars also allow students to "meet" potential UCSD mentors and gather information about potential projects. In 2013, seminars were sourced from faculty at UCSD, the University of Indiana, I2R in Singapore, the National Centre for Supercomputing Applications (NCSA) and Rensselaer Polytechnic Institute. That same year, seminars were broadcast simultaneously to Monash (in Melbourne) and UQ (in Brisbane), and audiences were able to ask questions from either venue. The seminar infrastructure supports a wide range of video conference technologies (both open source and commercial), and is displayed on a 20 MPixel OptiPortal.

Images: (background) Skydiving Lake Taupo—courtesy of Erickson Nguon; (left) Raymond Tran (PRIME 2013), Hoang Nguyen (MURPA 2009), David Abramson; Minh Huynh (MURPA 2012), and Fabian Lema (PRIME 2013) at the University of Queensland Australia—courtesy of Fabian Lema

WORKSHOPS AND WORKING GROUPS

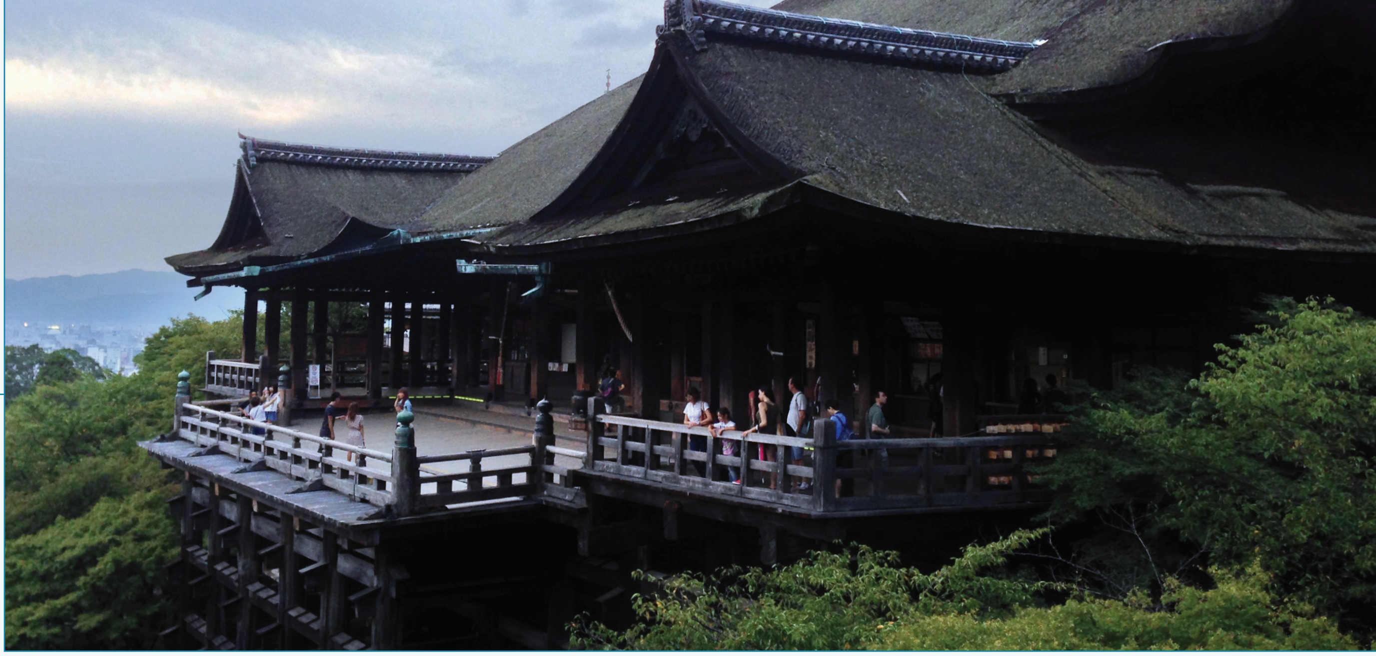
PRAGMA Workshops are meetings of all members of the PRAGMA community. They are the major vehicle for information exchange between working groups, researchers, and institutions; they also provide excellent opportunities to engage new researchers and students at the host sites. New participants bring new perspective, applications, technologies, students, and resources. These 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 develop new collaborations with members of the hosting institutions.

The workshops are organized to allow for the activities of the four working groups in PRAGMA, which are as follows:

- Resources Working Group: Working to make the distributed resources of PRAGMA useful to diverse applications. Co-leaders: Yoshio Tanaka (AIST) and Philip Papadopoulos (UCSD/SDSC).
- Telescience Working Group: Focusing on a variety of activities that require access to, or use of, remote equipment, such as tiled-display walls (TDW) and sensors. Co-leaders: Shinji Shimojo (NICT and Osaka U) and Fang-Pang Lin (NCHC).

Images: (above right) Kyomizu-dera ji (Water Temple in Kyoto)—courtesy of Jesus Rios; (background) The Great Wall of China—courtesy of Teri Simas



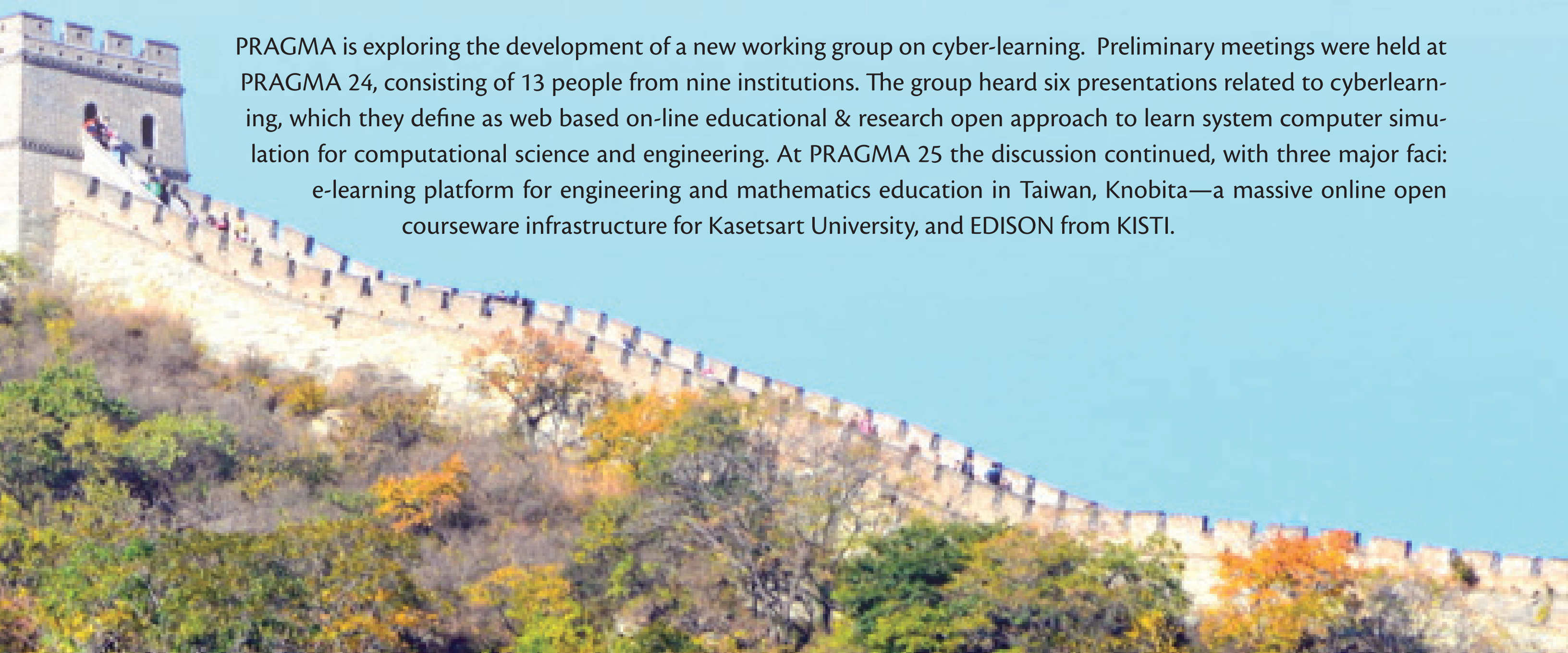
- GEO Working Group: Creating an infrastructure to share and integrate data on global earth observations, including remote sensing data and data from land-, lake-, and ocean-based sensors. Co-leaders: Sornthep Vannarat (NECTEC), Ryosuke Nakamura (AIST), and Franz Cheng (NARL).
- Biosciences Working Group: Focusing much of its efforts over the last several years on integrating technologies to create an infrastructure to advance the screening of potential compounds to combat infectious diseases. Co-leaders: Wilfred Li (UCSD/SDSC) and Habibah Wahab (USM).

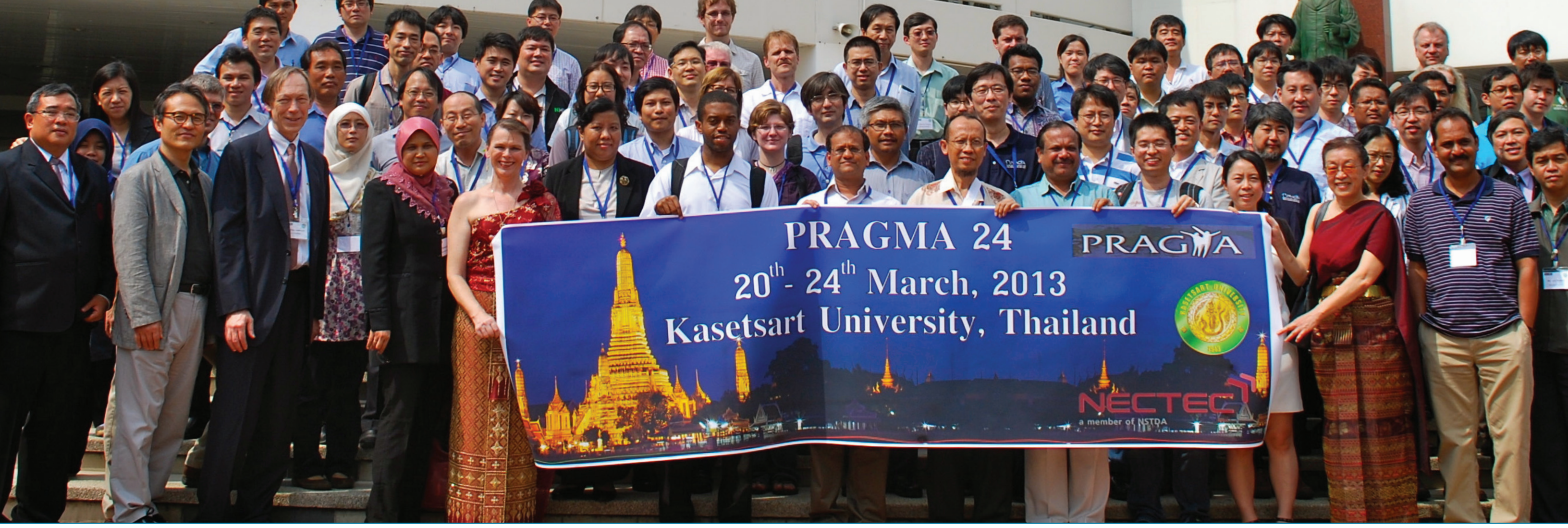
In addition, PRAGMA structures allows for new working groups to be formed. One such nascent group is on Cyber-Learning.

Cyber-Learning in PRAGMA, with EDISON as an example

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

PRAGMA is exploring the development of a new working group on cyber-learning. Preliminary meetings were held at PRAGMA 24, consisting of 13 people from nine institutions. The group heard six presentations related to cyberlearning, which they define as web based on-line educational & research open approach to learn system computer simulation for computational science and engineering. At PRAGMA 25 the discussion continued, with three major faci: e-learning platform for engineering and mathematics education in Taiwan, Knobita—a massive online open courseware infrastructure for Kasetsart University, and EDISON from KISTI.





One simulation-based cyber-learning system is called EDISON, "Education-research Integration through Simulation on the Net" (http://www.edison.re.kr). Anyone can learn and understand a theory or system through computer simulation anywhere and can predict operations of the system through easily changing parameters of the simulation model. It was created in July 2011 by KISTI (Korea Institute of Science and Technology Information) and currently covers three areas: Computational Fluid Dynamics, Nano Physics, and Computational Chemistry. It also has participants at major universities in Korea and has acquired more than 10,000 users since inception. The web portals for overseas users in nanophysics (http://nano.edison-project.org) and computational chemistry (http://chem.edison-project.org) were recently developed and introduced at PRAGMA 25.

INTERIM LEADERSHIP: Lead: Ruth Lee (KISTI); Co-leads: Hsi-ching Lin (NCHC), Putchong Uthayopas (Kasetsart U).

PRAGMA Workshops

In 2013, two PRAGMA Workshops were held:

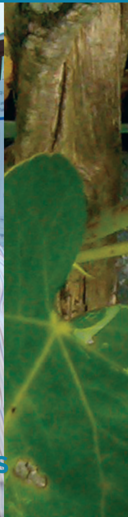
- PRAGMA 24—March 20-22, 2013—Bangkok, hosted by Kasetsart University
- PRAGMA 25—October 16-18, 2013—Beijing, hosted by the Computer Network Information Center, Chinese Academy of Sciences

PRAGMA Workshops are also used to host auxiliary, targeted activities. One set of activities at these two workshops were workshops organized by and for PRAGMA Students (see section on **Students**). In addition, in the last year we held two targeted workshops in association with PRAGMA Workshops (See **Section 3: Building Community**) which included:

- Workshop on Biodiversity, Ecosystems Services, and CI, hosted by Konkuk University on October 9, 2012, as part of PRAGMA 23
- Earth and Biodiversity Observing Workshop, hosted by Kasetsart University on March 23 and 24, 2013, as part of PRAGMA 24

In addition, the 2013 IEEE 9th International Conference on e-Science was held October 22-25 in Beijing, following the PRAGMA 25 Workshop.

Images: (above, left to right) Group photo from PRAGMA 24; Presenter at PRAGMA 24; Summer Palace, Bejing—all courtesy of Teri Simas; (below, left to right) Tony Fountain and Putchong Uthayopas at PRAGMA 24—courtesy of Teri Simas; Hibiscus—courtesy of Hanne Wolff; Philip Papadopoulos in Bangkok—courtesy of Teri Simas



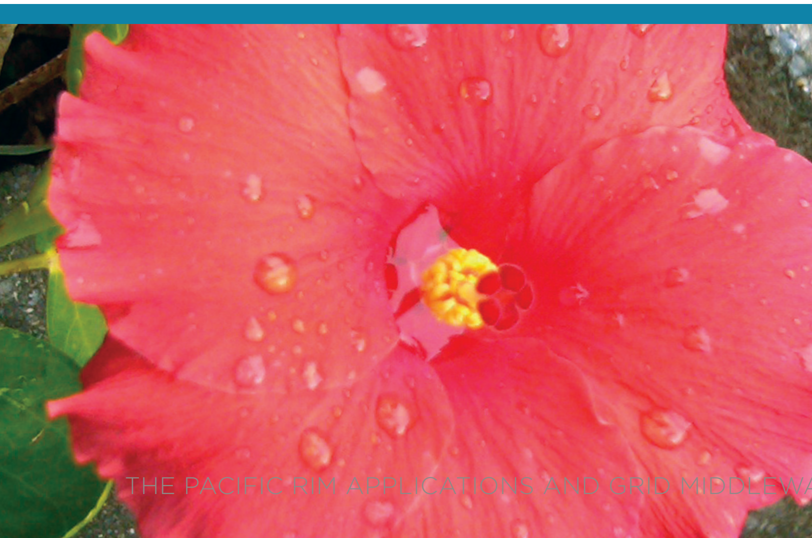
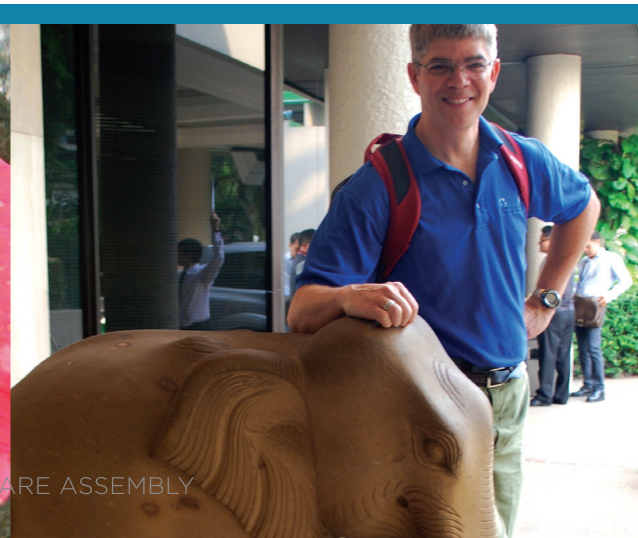


In addition, training programs such as SEAIP (Southeast Asia International Joint Research and Training Program (http://seaip.narlabs.org.tw) provided PRAGMA with new members.

This year we also held a Mini-PRAGMA in Indonesia June 3-5, 2013, held on the Universitas Indonesia (UI) campus in Depok, and hosted by the Faculty of Computer Science. See the section on **Community Building** for details.

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 26—April 9-11, 2014—Tainan, Taiwan, hosted by the National Center for High-performance Computing and National Cheng Kung University
- Lower Mekong Initiative, Hanoi, Vietnam, March 2014 (see **Community Building**)
- PRAGMA 27—October 15-17, 2014—Bloomington, Indiana, USA, hosted by Indiana University's Data to Insight Center

INSTITUTIONS, PARTNERS,

*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/about/committee.*

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.

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*

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@ais.cmc.osaka-u.ac.jp*

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

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

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

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

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): Piyawut Srichaikul, *piyawut.srichaikul@nectec.or.th*; 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*

NATIONAL INSTITUTE OF SUPERCOMPUTING AND NETWORKING (NISN), KOREA INSTITUTE OF SCIENCE AND TECHNOLOGY INFORMATION (KISTI): Kum Won Cho*, *ckw@kisti.re.kr*

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

Sun Moon Lake—courtesy of Sharon Grubner

AND SPONSORS

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;* Tom DeFanti, *tdefanti@ucsd.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*

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

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: Jacqueline Brown, *jbrown@ms.uw.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, *sclin@gate.sinica.edu.tw;* Eric Yen, *eric@sinica.edu.tw*

ADVANCED SCIENCE AND TECHNOLOGY INSTITUTE (ASTI): Denis Villorente, *denis@asti.dost.gov.ph;* Peter Antonio B. Banzon, *peterb@asti.dost.gov.ph*

BeSTGRID NEW ZEALAND (BeSTGRID): Nick Jones, *n.jones@auckland.ac.nz*

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;* Luke Jing Yuan, *jyluke@mimos.my*

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

Additional Active PRAGMA Organizations

BALBOA PARK ONLINE COLLABORATIVE (BPOC; www.bpoc.org), is a nonprofit organization that provides technical support to more than 20 art, science and cultural organizations in San Diego. It has 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 **Highlight** on the San Kei En Project).

BIODIVERSITY INSTITUTE, UNIVERSITY OF KANSAS (UK; 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, in the use of Lifemapper software and deployment of its computational engine on PRAGMA Virtual Clusters.

LAN ZHOU UNIVERSITY (LZU; www.lzu.edu.cn) has contributed resources to the PRAGMA Grid, and representatives from the university have attended the PRAGMA 12, 13, 17, 18, 19, and 22 Workshops. They are currently providing resources and are participating in the PRAGMA Cloud, and have authored a VM, demonstrated at PRAGMA 23.

NARA INSTITUTE OF SCIENCE AND TECHNOLOGY (NAIST; www.naist.jp/en) has been participating in PRAGMA activities through the Resources Working Group, and has been experimenting with OpenFlow technology. Furthermore, NAIST has hosted PRIME students (see section on **Students**).

NATIONAL APPLIED RESEARCH LABORATORY (NARL; www.narl.org.tw/en) was established in 2003 to consolidate nine national laboratories into a single nonprofit organization. Its goal is 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 as determined by national policy. 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; http://www.nict.go.jp/index.html), 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'S RESEARCH COMPUTING CENTRE (UQ; www.rcc.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 UCSD students from June to August 2013 (see section on **PRIME**) and sending students to PRAGMA sites (see **MURPA QURPA** section).

Universiti Teknologi Malaysia (UTM; www.utm.my) researchers have been involved in the "PRAGMA Biodiversity Expedition" and in the workshops on biodiversity.

PARTNERS

GLEON, the Global Lakes Ecological Observatory Network, 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, has grown to a network of more than 400 members and 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 the 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 RESEARCH CENTER (NSRC; www.nsrc.org) has longstanding experience in running hands-on networking training workshops and providing engineering assistance at both the campus and national network levels. They have worked in more than one hundred countries throughout the world over the past 20+ years. NSRC has been working with PRAGMA recently in the area of Southeast Asia, supporting researchers from Myanmar to attend PRAGMA 24 and collaborating with PRAGMA and IU on the Lower Mekong Initiative to help enable more international science education.

SPONSORS

PRAGMA is supported by its member institutions and the U.S. National Science Foundation (NSF ACI 1234983, PI:Papadopoulos), involving researchers at UCSD, U Florida, Indiana U, and U Wisconsin. This involves support from NSF Division of Advance Cyberinfrastructure, Office of International and Integrative Activities, 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 grid activities are funded by 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.

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 (60703024) and the Chinese Ministry of Education (NCET-09-0428).

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

AIST's sponsors include the Special Coordination Funds for Promoting Science and Technology (MEXT, Japan) and the Core Research of Evolutional Science and Technology Program (JST, Japan).

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 MOST through the K e-Science and Ministry of Information and Communication (MIC) through Korean Grid Infrastructure Implementation and Middleware Development Project (K*Grid).

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 National Science Council, Taiwan.

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

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, the National Biomedical Computation Resource (NBCR: NIH, NCRR P 41 RR08605) the Gordon and Betty Moore Foundation, and the Qualcomm Institute, UCSD.

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

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

PRAGMA

www.pragma-grid.net

Image from Huisun Forest Station, Taiwan—courtesy of Peter Arzberger