
THE PACIFIC RIM APPLICATIONS AND GRID MIDDLEWARE ASSEMBLY
PRAGMA 2017

CONTENTS

2017 COLLABORATIVE OVERVIEW

INTRODUCTION

The Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) is an open, international organization that makes cyberinfrastructure accessible, easy to use, and useful for long tail of science communities who lack access to computational and data resources to advance their science.. These communities are addressing societally important problems and they have yet to take full advantage of the new cyberinfrastructure technologies and approaches. PRAGMA operates by bringing together cyberinfrastructure (CI) experts and researchers to understand the opportunities in technologies and challenges faced by scientists, and then to co-design CI solutions. Initial solutions are tested on PRAGMA's shared, experimental infrastructure, and later moved onto production systems at several PRAGMA member sites.

In 2017 PRAGMA celebrated its 15th birthday. Birthdays are a cause for celebration, a time to reflect about past accomplishments and experiences, and to plan for the future.

SINCE ITS INCEPTION IN 2002, PRAGMA has built a trusted network of collaborators around the Pacific Rim region. That trust allows PRAGMA to react quickly to large events that impact its members, such as mobilizing the U.S. academic research community to assist PRAGMA member National Center for High-performance Computing (NCHC) in its response to SARS (tele-visual-connectivity among patient, family and medical care professions), and in response to the 2011 tsunami, assuming responsibility by NCHC and the San Diego Supercomputer Center (SDSC) for providing some of the National Institute of Advanced Industrial Science and Technology (AIST) services. The trust and collaborations in PRAGMA led to scientific insights by explorations between scientists and CI experts in areas of molecular chemistry, climate and environmental interactions from savannah burns, and understanding of volcanic ash distribution around Costa Rica's capital, San José. PRAGMA evolved with and led changes in accessing and provisioning CI via virtualization and the transport of virtual clusters that have migrated into the U.S. national (e.g., SDSC's Comet) and international infrastructure. It has also led to the development of PRAGMA Experimental Network Testbed (ENT), enabling infrastructure research by researchers, including students, about software-defined networks. PRAGMA has enthusiastically engaged students within its normal workshops to provide professional leadership experiences, and in research projects and student exchanges. Other activities initiated by PRAGMA members were the Southeast Asia International Joint Research and Training Program (SEAIP) by NCHC in 2005 and the Collaborations to Enable Transnational Cyberinfrastructure Applications (CENTRA) by NCHC, National Institute of Information and Communication Technology (NICT) and the University of Florida in 2015.

In addition, PRAGMA has sought to engage new research areas, often in collaboration with others. They have included disaster mitigation (via a Shonan Meeting), biodiversity (through multiple venues), big data (through SEAIP), and smart and connected communities (via CENTRA). Starting in 2003, these accomplishments and others have been documented in the annual PRAGMA Collaborative Overviews. See also (Arzberger 2017), and a presentation given at the PRAGMA 32 Workshop in April 2017[1].

At the PRAGMA 32 Workshop, there was a celebration of PRAGMA's first 15 years. The workshop also marked the transition in leadership. Philip Papadopoulos, a co-founder of PRAGMA and the key architect for a number of the structures that still serve PRAGMA well, was unanimously elected Chair of the PRAGMA Steering Committee by the Committee. He replaced founding chair Peter Arzberger, who remains on the Steering Committee.

IN THE PAST YEAR PRAGMA has accomplished several collaborative successes. Among them:

- LAKE EXPEDITION: GRAPLEr, an open-source, distributed computing system that allows submission, efficient computation, and retrieval output of hundreds or thousands of General Lake Model (GLM) simulations, can now run on the NSF

XSEDE "Comet" virtual cluster resource as well as the PRAGMA cloud resource at NAIST. It is now being used by undergraduate and graduate students at Virginia Tech to understand the effects of eutrophication and climate change on harmful algal blooms, and training modules have been developed and taught to several hundred students, with clear impact on students' understanding of the effects of climate change on lakes. In particular, the GRAPLEr framework allows researchers to study an unprecedented number of climate- and land-use scenarios on water quality in GLEON lakes. The goal of this effort is to predict how lakes around the globe will respond in the future to different forms of anthropogenic stress.

- BIODIVERSITY EXPEDITION: Lifemapper, an open-source modeling environment, allows researchers to model species distribution or habitat suitability using species occurrence data from biological collections around the world, together with climatological and geological data. Most recently, it was extended to cover phylogenetic (evolutionary) data derived from DNA sequencing studies. Lifemapper has been redesigned to make it portable and easy to deploy. In addition to completely refactoring the code base for more flexible workflows this year, the team also set up one virtual Lifemapper instance at NCHC in Taiwan, which will host a computational pipeline for Taiwanese-focused data and research, as well as another on XSEDE's Comet computing resource at the San Diego Supercomputer Center (SDSC) at UC San Diego.

- PRAGMA EXPERIMENTAL NETWORK TESTBED (ENT): PRAGMA-ENT is designed to offer complete freedom for researchers to access network resources to develop, experiment with, and evaluate new ideas without concerns about interfering with a production network. PRAGMA-ENT has expanded its resources this year to MIMOS Berhad and National Institute of Information and Communication Technology (NICT), and it is deploying a distributed software-defined storage system to determine research challenges resulting from distributed storage systems over virtual network and computing resources. In addition, researchers have enhanced their previous multipath data-transfer experiments on PRAGMA-ENT, while other researchers have used PRAGMA-ENT to conduct experiments that reduce the load on end-point hypervisors and improve the throughput (compared to tunneling).
- MULTI-CLOUD: To enable use of PRAGMA's multi-cloud infrastructure, several challenges were addressed this year. UC San Diego and NCHC collaborated to develop a process to create a universal cluster image format. In addition, an authoritative central virtual cluster image repository was setup; this central repository is used on local sites to synchronize virtual images automatically. Finally, undergraduate students from Thammasat University developed a new GUI, which was added to PRAGMA Cloud Scheduler.
- CYBERLEARNING: EDucation-research-industry Integration through Simulation On the Net (EDISON) is an innovative cyberlearning platform developed by the Korea Institute of Science and Technology Information (KISTI). This year, new domains of learning were added and more users took advantage of the infrastructure. Furthermore, new functionality was added for data-driven simulations. Finally, in October 2017, KISTI and NCHC launched a joint research lab for close collaboration at the annual Korea SuperComputing Conference (KSC).
- DATA MANAGEMENT: Through a multi-lateral collaboration between Indiana University, AIST, the International Rice Research Institute (IRRI) and the Advanced Science and Technology Institute (ASTI), PRAGMA demonstrated that it is possible to organize data products in a scenario where the number of scientists simultaneously using the IRRI tools is vastly scaled out. The next step is to put in place protections on a per-scientist basis so that results are not made immediately public.
- IMMERSIVE VISUALIZATION: In a collaboration initiated by CENTRA with the University of Hawaii, AIST is building a disaster management platform developed in SAGE2 (Scalable Amplified Group Environment 2: developed at the Laboratory for Advanced Visualization and Applications (LAVA) at University of Hawaii, Manoa and Electronic Visualization Laboratory (EVL) at University of Illinois, Chicago), which would allow users to simultaneously interact with a variety of content in a shared collaborative environment.

[1] *https://tinyurl.com/y92knsh4*

- DISSEMINATION: In July 2017 a special issue of *Concurrency and Computation: Practice and Experimentation* was released with seven research articles about PRAGMA (Plale and Chen 2017). PRAGMA has produced, revised and/or enhanced, and published software for others to test or use. PRAGMA also organizes special sessions at its workshops to focus on new areas and new communities that could benefit from PRAGMA's approaches. For example, at PRAGMA 32, there was a one-day meeting on Smart UNiversity TOWNS (SUNTOWNS), organized by CENTRA under joint auspices of PRAGMA, to look at potential opportunities and challenges in matching technology opportunities with community and city-wide challenges.
- STUDENTS: In PRAGMA, students continue to play leadership roles in the organization. They serve on the Steering Committee and organize poster sessions. They are also involved in research via research visits, publications and software development such as PRAGMA's Cloud Scheduler (as noted above) or an immersive visualization environment involving sensor data from rivers in connection with disaster management.

AFTER 15 YEARS, WHAT DOES PRAGMA HOPE TO ACCOMPLISH IN THE NEXT FIVE YEARS? PRAGMA's long-term goal remains the same: make cyberinfrastructure accessible and easy to use for targeted communities. We will actively:

- continue co-designing CI and bringing together CI experts with members of scientific communities with CI needs;
- increase training in communities with whom we have co-designed solutions in the past;
- engage new communities through a sustained and systematic approach to identify them;
- enhance activities for students by adding activities for undergraduates as well as professional development opportunities for all students; and
- create opportunities via special sessions and meetings to engage new communities.

PRAGMA also remains committed to making its software available and to publishing its findings in journals or in conference proceedings.

Our specific accomplishments in the last year along with concrete plans for the next year are highlighted in the following pages.

REFERENCES

Arzberger, P. 2017. Reflection on the origins, evolution, and future of PRAGMA. *Concurrency and Computation: Practice and Experience.* Wiley, 29(13). doi:10.1002/cpe.4136

Plale, B. and Chen, M. 2017. Eds. Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) special issue on international clouds for data science, *Concurrency and Computation: Practice and Experience.* Wiley, 29(13). doi: 10.1002/cpe.4140

PRAGMA Collaborative Overview: *http://www.pragma-grid.net/overview/*

PRAGMA Software: *http://www.pragma-grid.net/products/*

AT A GLANCE

PRAGMA was established in March 2002, with the mission to enable the long tail of science through scientific expeditions and infrastructure experimentation for Pacific Rim institutions and researchers.

MEMBERS: 28 Institutional Members. See *Institutions and Sponsors* for institutions and abbreviations used in this Collaborative Overview.

WEBSITE: *http://www.pragma-grid.net*

GOVERNANCE: Steering Committee and PRAGMA's Operating Principles and Procedures PRAGMA Steering Committee Chair: Philip Papadopoulos

WORKSHOPS OF PARTICIPANTS: Twice a year to share progress and plan future activities. Open to all, hosted by PRAGMA members, often in conjunction with other activities.

2017 WORKSHOPS

- PRAGMA 32: April 12–14, 2017, University of Florida, Gainesville, Florida, USA. Held in conjunction with the 2nd annual CENTRA meeting, and co-hosting the first Smart UNiversity TOWNS (SUNTOWNS) conference, and a Lake Modeling Workshop.
- PRAGMA 33: October 16–17, 2017, University of Queensland, Brisbane, Australia. Held in conjunction with the eResearch Australasia Conference. After PRAGMA 33, Lake Modeling Workshop held at Griffith University.

2018 WORKSHOPS

- PRAGMA 34: May 9–12, 2018. Co-hosted by AIST and Osaka University. Location in Tokyo, Japan. Held in conjunction with the 3rd annual CENTRA meeting, May 14–16, 2018.
- PRAGMA 35 and beyond: Discussions are underway to identify hosts, locations and dates.

SCIENTIFIC EXPEDITIONS

- LIMNOLOGY: Predicting lake eutrophication and training the next generation of lake scientists
- BIODIVERSITY: Understanding biological adaptation in regions around the Pacific Rim
- ENT: Developing an experimental network testbed for experimenting with software-defined networks and monitoring impacts of choices

WORKING GROUPS

- RESOURCES: Making the distributed resources of PRAGMA useful to diverse applications
- TELESCIENCE: Making and improving access to or use of remote equipment (e.g. tiled-display walls or sensors)
- BIOSCIENCES: Creating stable infrastructure to perform computational genomics analyses with a focus on rice breeding and integrating technologies to create an infrastructure to advance the screening of potential compounds to combat infectious diseases
- CYBERLEARNING: Developing simulation-based learning technology

STUDENTS

- PRAGMA STUDENTS: Stimulating international, cross-disciplinary collaborations among students and junior researchers

PARTNER ACTIVITIES

- GLEON: The Global Lake Ecological Observatory Network's mission is to understand, predict and communicate the role and response of lakes in a changing global environment. *http://gleon.org*
- CENTRA: The long-term goal of Collaborations to Enable Transnational Cyberinfrastructure Applications is to advance the scientific understanding of distributed, software-defined cyberinfrastructure. CENTRA refers to a partnership and evolving framework for collaborations among research centers, institutes and laboratories across the world. It engages junior researchers with software-defined infrastructure, and initial collaborations in three application domains: environmental modeling, disaster management, as well as smart and connected communities. *http://www.globalcentra.org*
- SEAIP: The Southeast Asia International Research and Training Program is hosted annually by Taiwan's National Center for High-performance Computing (NCHC). The program aims to promote collaborations in cyberinfrastructure among researchers in Southeast Asia, and between researchers in Southeast Asia and their counterparts in other parts of the world. *https://event.nchc.org.tw/seaip*

SPONSORS

- Multiple, often associated with members and funded through many different national science foundations. See *Institutions and Sponsors* for a list of this year's sponsors by institution.

& EXPEDITIONS
WORKING GROUP HIGHLIGHTS

THE GLEON RESEARCH AND PRAGMA LAKE EXPEDITION COLLABORATION – GRAPLE

Improving simulation modeling to advance the state-of-the-art water quality prediction

The PRAGMA Lake Expedition is an inter-disciplinary collaboration between computer scientists and lake modelers affiliated with GLEON (the Global Lake Ecological Observatory Network). This collaborative effort is advancing the current understanding of the effects of eutrophication and climate change on harmful algal blooms through batch simulation of lake hydrodynamics, while also advancing the state-of-the-art in water quality prediction through the use of models. Its main software product is GRAPLEr, an R-based open-source software that brings the power of distributed computing to the fingertips of lake ecology modelers, allowing them to submit hundreds or thousands of General Lake Model (GLM[2]) simulations, run these lake model simulations efficiently, and retrieve and visualize model output (Subratie et al. 2017).

The lake team, led by Cayelan Carey, Paul Hanson and Renato Figueiredo, has continued to make progress in the development of GRAPLEr, using it in several teaching and research activities throughout the year. In particular, the team is using the advanced capacity of the GRAPLEr's distributed computing framework to analyze the effects of an unprecedented number of climate- and land-use scenarios on water quality in GLEON lakes. The goal of this effort is to predict how lakes around the globe will respond in the future to different forms of anthropogenic stress. In terms of GRAPLEr development and improvements, the team has enhanced the cyber-infrastructure with HTCondor nodes connected via the PRAGMA IPOP overlay (IPOP) from Nara Institute of Science and Technology (NAIST), a PRAGMA Cloud site in Nara, Japan, as well as from the NSF XSEDE "Comet" virtual cluster resource. These resources have allowed GRAPLEr to tap into hundreds of compute cores, and have allowed graduate and undergraduate-level researchers to execute climate change scenarios with hundreds of thousands of simulations per batch. As part of this effort, the team organized and led a hands-on Lake Modeling Workshop co-located with PRAGMA-32 in Gainesville, Florida, USA in April 2017.

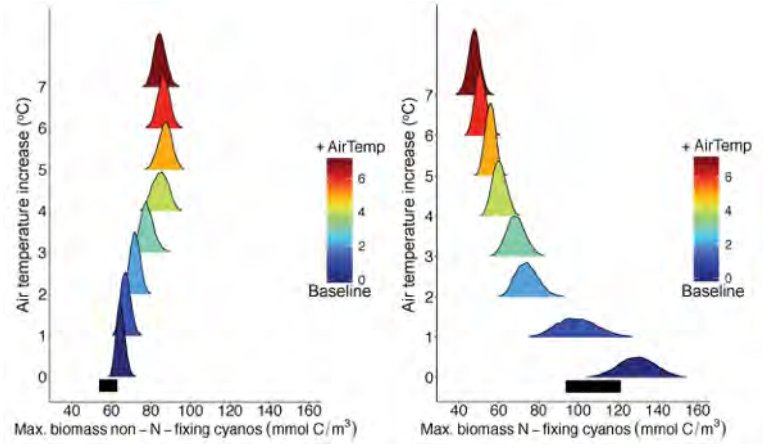


Figure 1: More than 100,000 lake modeling simulations were run, one each for a unique climate—land-use combined scenario. These simulations indicate that blooms for non-nitrogen-fixing cyanobacteria will increase as air temperatures warm, but blooms of nitrogen-fixing cyanobacteria will decrease in Lake Mendota, Wisconsin, USA. This switch in dominance may result in increased bloom toxicity in the lake in the future.

[2] *http://aed.see.uwa.edu.au/research/models/GLM*

Image: Lake Modelers and CI Experts at Lake Modeling Workshop—courtesy of Richard Garand UF

As an outcome of the workshop, GRAPLE students are developing two manuscripts: the first one comparing the responses of different GLEON lakes to climate change, and the second examining how climate and land use change will interact to predict cyanobacterial blooms.

Limnology has not traditionally leveraged national-scale high-performance computing / high-throughput computing (HPC/HTC) resources to address computational needs, and most limnologists have limited computational resources and are therefore limited to running a few simulations at a time on their personal computers. It is therefore a major leap to be able to dispatch the execution of hundreds of thousands of simulations for an experiment and obtain summary results through a convenient desktop interface (R-Studio), which can change how researchers in the field approach the use of modeling.

Ecosystem modeling is a critically important tool for environmental scientists, yet is rarely taught in undergraduate and graduate classrooms. To address this gap, the team collaborated on a lake-modeling teaching module that exposes students to a suite of modeling skills and tools—including computer programming, numerical simulation modeling, and distributed computing—that the students apply to study how lakes around the globe are experiencing the effects of climate change. To date, this module has been taught to several hundred students in universities across the U.S. Our assessment indicates that participation in the module significantly increased both undergraduate and graduate students' understanding about climate change effects on lakes. (Carey and Gougis, 2017)

PARTICIPANTS

Virginia Tech: Cayelan Carey, Arianna Krinos, Kait Farrell, Nicole Ward, Mary Lofton, Ryan McClure, Jon Doubek; *University of Wisconsin:* Paul Hanson; *University of Florida:* Renato Figueiredo, Jaikrishna Tanjore, Ken Subratie, Vahid Daneshmand; *Nara Institute of Science and Technology (PRAGMA Cloud resources added to GRAPLEr):* Kohei Ichikawa; *UC San Diego (Collaborating on PRAGMA Cloud addition):* Shava Smallen

REFERENCES

Carey, C.C. and Gougis, R.D. 2017. Simulation modeling of lakes in undergraduate and graduate classrooms increases comprehension of climate change concepts and experience with computational tools. *Journal of Science Education and Technology*. 26: 1-11. doi: 10.1007/s10956-016-9644-2

Subratie, K.C., Aditya, S., Mahesula, S., Figueiredo, R., Carey, C.C., Hanson, P.C. 2017. GRAPLEr: A distributed collaborative environment for lake ecosystem modeling that integrates overlay networks, high-throughput computing, and WEB services. *Concurrency and Computation: Practice and Experience.* Wiley, 29(13). doi: 10.1002/cpe.4139

SOFTWARE

GRAPLEr. See *http://graple.org* to download and learn more about the software.

PRAGMA IPOP overlay: *http://ipop-project.org*

VIRTUAL BIODIVERSITY EXPEDITION AND LIFEMAPPER

Allowing greater use of Lifemapper through refined software, improved data infrastructure and new features

Analyzing and modeling patterns of how species diversity emerges and how they are sustained on local, regional, continental, and global scales is at the heart of biogeographic study. Understanding these processes and the resulting structure of natural communities is central to sustaining and conserving biological diversity.

With open-source tools developed by the Lifemapper Project, the Virtual Biodiversity Expedition performs species distribution or habitat suitability modeling with species occurrence data from biological collections around the world. Data may come from individual researchers or projects, or aggregators such as the Integrated Digitized Biocollections (iDigBio[3]), Biodiversity Information Serving Our Nation (BISON[4]), and the Global Biodiversity Information Facility (GBIF[5]). Lifemapper integrates those data with climatological and geological global data, and most recently with phylogenetic (evolutionary) data derived from DNA sequencing studies of plants and animals.

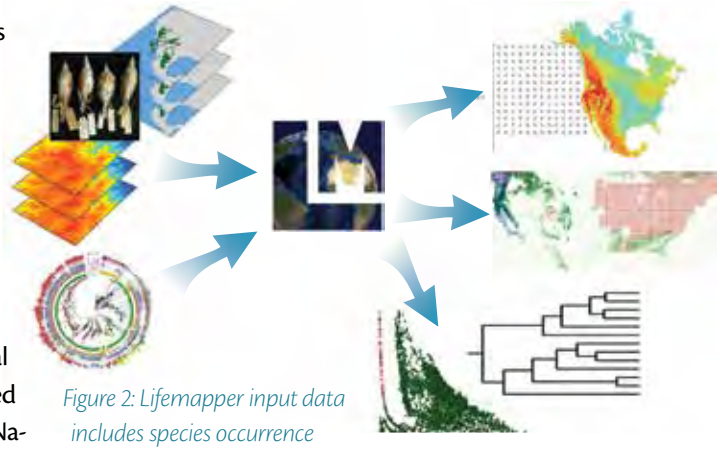


Figure 2: Lifemapper input data includes species occurrence records, climate and other environmental layers, as well as phylogenetic trees. Outputs include data and linked displays of biogeographic and phylogenetic diversity, with site-based and species-based plots, maps, and trees.

Over the last five years, Lifemapper (LM), with the help of PRAGMA, was streamlined and hardened to make it portable and simplify its deployment within the PRAGMA community and beyond. Key changes include: 1) a re-packaged software infrastructure for Rocks Linux clusters, optimized for local hardware; 2) a new bootstrapping process to quickly build and initialize the database for default datasets and workflows; 3) well-defined data input requirements, suitable for researcher initialization; and 4) an improved, efficient method for software updates. The software development cycle is shown in Figure 3.

In the last year, Lifemapper has completely redesigned its database and refactored the code to more easily integrate the inputs, outputs, and analyses of the previous single-species workflow with evolutionary data and other inputs for new multi-species, phylogenetic, and geographic analyses. The refactored code also supports flexible workflows with different data pre- and post-processing steps, inputs, and analyses, including the initiation of large multi-part analyses with a single configuration file and data input packages. Lifemapper has also created a new web client, suitable for any user-level. This new interface will enable us to develop for the collaborative visualization software SAGE2[6] in a collaboration led by CENTRA (see *CENTRA* section).

These improvements, made with collateral funding from three NSF research awards, have transformed the Lifemapper tools into a highly portable and deployable system (Williams et al. 2017). Lifemapper now can be easily installed on physical or virtual hardware ranging from NSF XSEDE resources (Comet) to laptops, as a self-contained system on a single host, or as a linked set of servers across resources in a multi-national testbed infrastructure. For the XSEDE Comet Lifemapper cluster, we enabled a use of high-throughput, low-latency InfiniBand network to connect to a data server intended for use with a high resolution North American species data. In August 2017 we took a first step to set up a virtual Lifemapper instance at NCHC in Taiwan by setting up virtual clusters intended for use with higher-resolution Taiwanese data, a result envisioned via CENTRA discussions.

[3] *https://www.idigbio.org/*
[4] *https://bison.usgs.gov/*
[5] *http://www.gbif.org/*
[6] *http://sage2.sagecommons.org/*

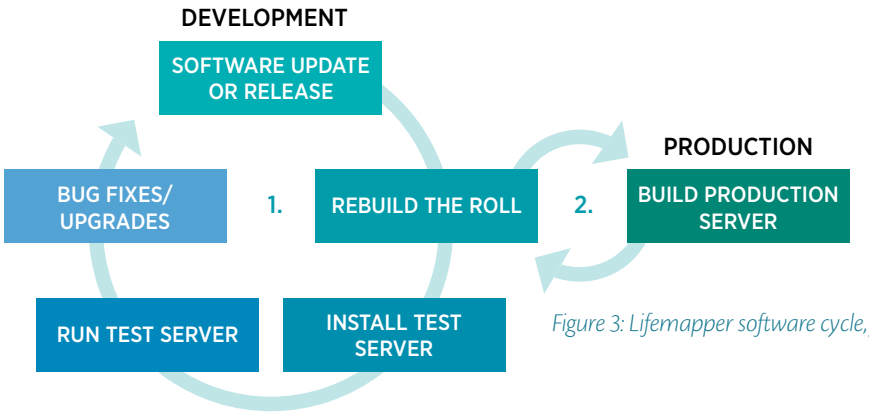


Figure 3: Lifemapper software cycle, from development to production.

Key objectives for the Biodiversity Expedition in the next five years include:

BROADER PRAGMA PARTICIPATION: We will cultivate researchers around the Pacific Rim and among PRAGMA members, building on previous workshops with researchers in Indonesia, Taiwan, and Vietnam.

EASIER ACCESS AND COLLABORATION: We will enhance LM tool usability by deploying Jupyter notebook interfaces to the most used LM workflows, and by developing a front-end client to interface with visualization environments; for example the Scalable Amplified Group Environment (SAGE 2), building on collaborations led by CENTRA.

ENHANCED BIODIVERSITY MODELING FUNCTIONS: We will bring additional phylogenetic data into the spatial and biogeographical analysis already supported by Lifemapper for a PRAGMA project comparing the evolutionary changes of different regions across the Pacific Rim. We will also introduce a new dimension into geospatial analyses of biological diversity. This direction is being driven by new collaborations that span the U.S. and the Pacific Rim.

PARTICIPANTS

University of Kansas: James Beach, Aimee Stewart; *UC San Diego:* Philip Papadopoulos, Nadya Williams; *National Center for High-performance Computing (NCHC):* Hsiu-Mei Chou, Weicheng Huang, Fang-Pang Lin, Serena Pan

REFERENCE

Williams, N., Stewart, A., Papadopoulos, P. 2017. Virtualizing Lifemapper software infrastructure for biodiversity expedition. *Concurrency and Computation: Practice and Experience. Wiley,* 29(13). doi: 10.1002/cpe.4137

SOFTWARE

Lifemapper website *http://lifemapper.org*

The core Lifemapper code is available at *https://github.com/lifemapper*

Lifemapper software now includes tools for installing and configuring Lifemapper third party dependencies, configuring the Lifemapper installation for local hardware, and bootstrapping the database with default or user data. Lifemapper GitHub repositories *https://github.com/pragmagrid/lifemapper-server* and *https://github.com/pragmagrid/lifemapper-compute* contain code for installation of the two primary components (server and compute) of the Lifemapper; these components may be installed on the same or separate Rocks clusters.

PRAGMA EXPERIMENTAL NETWORKING TESTBED (ENT)

Deploying and conducting experiments on software-defined networks

The PRAGMA Experimental Network Testbed (PRAGMA-ENT) expedition aims to construct a breakable international Software-Defined Networking (SDN)/OpenFlow testbed for use by PRAGMA researchers and collaborators. PRAGMA-ENT is breakable in the sense that it will offer complete freedom for researchers to access network resources and develop, experiment with, and evaluate new ideas without concern they are interfering with a production network. PRAGMA-ENT also will provide networking support to the PRAGMA multi-cloud and user-defined trust envelopes. This will expose SDN to the broader PRAGMA community and facilitate the long tail of science by creating new collaborations and infrastructure among institutes in the Pacific Rim area (Ichikawa et al. 2017).

The PRAGMA-ENT team has been connecting resources in the U.S. (University of Florida, UC San Diego and Indiana University), Japan (Nara Institute of Science and Technology, National Institute of Advanced Industrial Science and Technology, and Osaka University), Taiwan (National Applied Research Laboratories) and Thailand (Thammasat University). Since all OpenFlow switches at each site are interconnected through dedicated virtual local area network (VLAN) links and overlay virtual links, PRAGMA users can develop their own controllers to manage the entire network testbed and/or perform or measure large scale network experiments. Researchers used PRAGMA-ENT to enhance their previous multipath data transfer experiment so that it can allocate optimal number of TCP streams on each path (Huang et al. 2017). Another researcher led experiments to reduce the load on end-point hypervisors and improve the throughput (see next section, *PARES*).

We are currently working with researchers at MIMOS to expand the backbone of ENT to Malaysia and the Japan-wide Orchestrated Smart/Sensor Environment (JOSE), which is a virtual research testbed provided by NICT. We are in the process of deploying a distributed software-defined storage system using this environment to determine the research challenges using distributed storage systems over virtual network and computing resources.

In addition, we have been working to provide an alternative international path between the U.S. and Japan by using Science Information NETwork (SINET), another national research and education network in Japan. Using multiple international paths will improve the aggregated network bandwidth of the ENT backbone. This will help us to perform more large-scale and high-performance network experiments.

PARTICIPANTS

Nara Institute of Science and Technology (NAIST): Kohei Ichikawa, Pongsakorn U-chupala, Chawanat Nakasan, Che Huang, Yasuhiro Watashiba; *University of Florida:* Matthew Collins, Kyuho Jeong, Renato Figueiredo, José Fortes; *National Institute of Information and Communication Technology (NICT):* Hiroaki Yamanaka, Jin Tanaka; *UC San Diego:* Nadya Williams, Shava Smallen, Philip Papadopoulos; *Osaka University:* Yoshiyuki Kido, Susumu Date, Shinji Shimojo; *National Institute of Advanced Industrial Science and Technology (AIST):* Jason Haga, Ryosei Takano, Yoshio Tanaka; *National Institute of Informatics (NII):* Atsuko Takefusa; *National Applied Research Laboratory (NARLabs)*

and the National Center for High-performance Computing (NCHC): Li-Chi Ku, Fang-Pang Lin, JenWei Hu, TeLung Liu; *Indiana University, Bloomington:* Quan (Gabriel) Zhou, Beth Plale, Jennifer Schopf; *Internet2:* John Hicks, Rick McMullen *(now at Texas A&M)*; *Kasetsart University:* Putchong Uthayopas; *Thammasat University:* Wanida Putthividhya, Prapaporn Rattanatamrong; *MIMOS:* Luke Jing Yuan

REFERENCES

Huang, C., Nakasan, C., Ichikawa, K., Watashiba, Y., Iida, H. 2017. A Multipath OpenFlow Controller for Multiple TCP Stream Applications, *Journal of Information Processing*. 25. 924-933.

Ichikawa K. et al., 2017. PRAGMA-ENT: An International SDN testbed for cyberinfrastructure in the Pacific Rim. *Concurrency and Computation: Practice and Experience.* Wiley, 29(13). doi:10.1002.cpe.4138

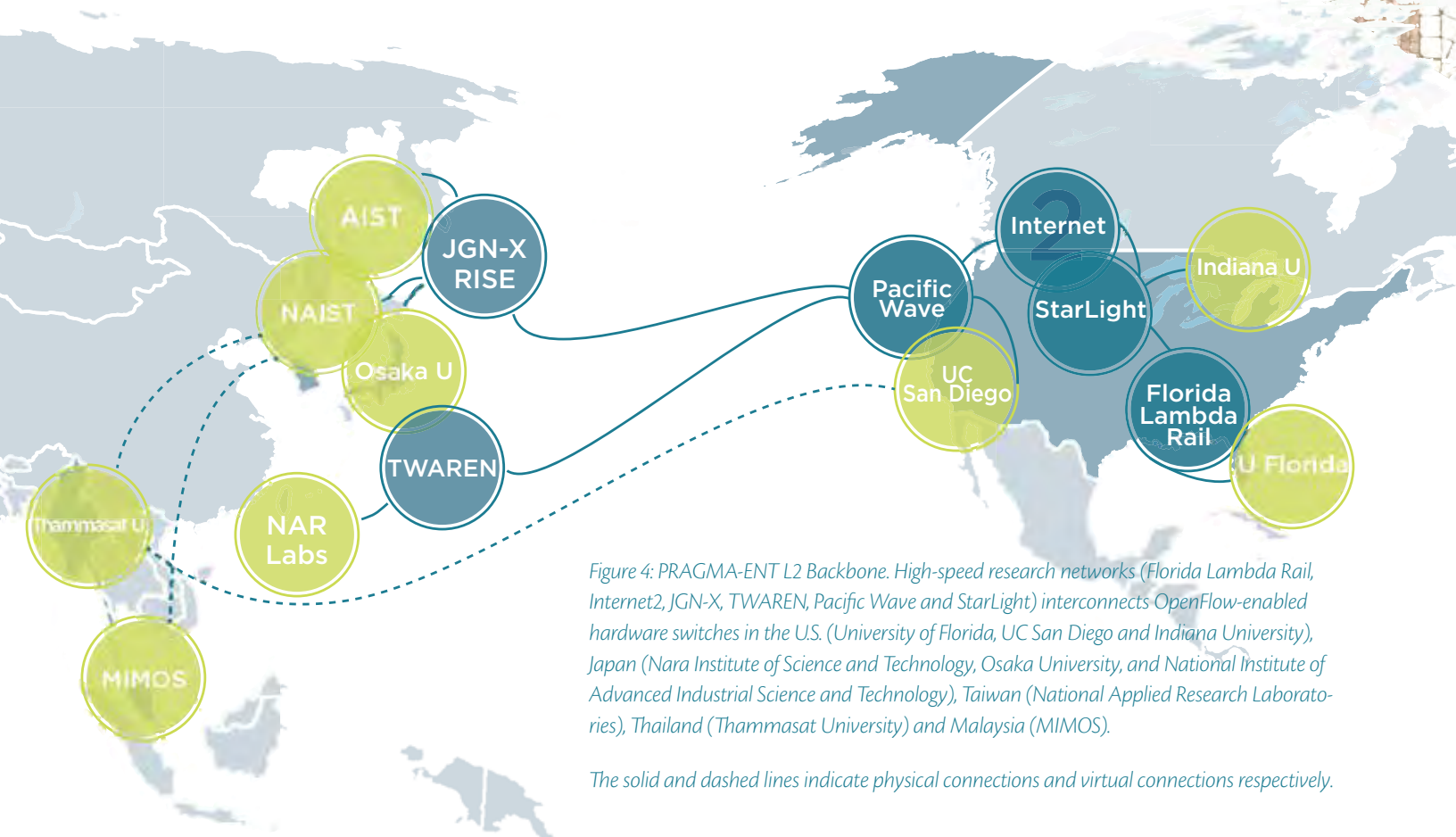


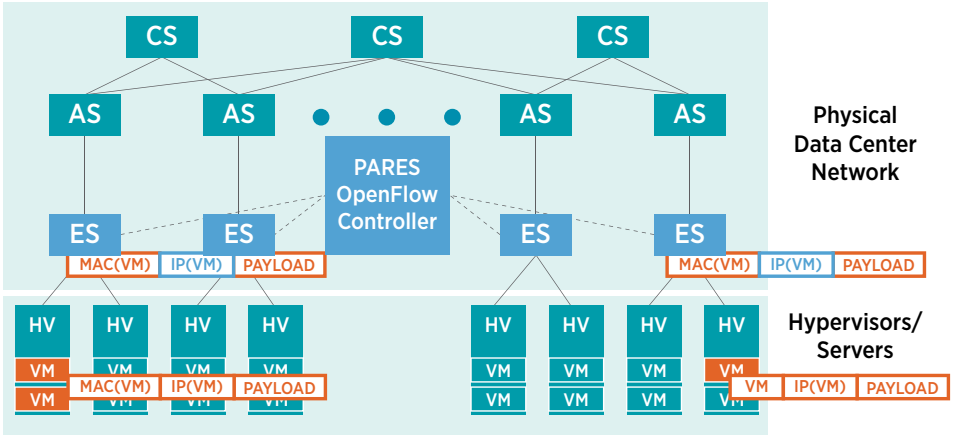
Figure 4: PRAGMA-ENT L2 Backbone. High-speed research networks (Florida Lambda Rail, Internet2, JGN-X, TWAREN, Pacific Wave and StarLight) interconnects OpenFlow-enabled hardware switches in the U.S. (University of Florida, UC San Diego and Indiana University), Japan (Nara Institute of Science and Technology, Osaka University, and National Institute of Advanced Industrial Science and Technology), Taiwan (National Applied Research Laboratories), Thailand (Thammasat University) and Malaysia (MIMOS).

The solid and dashed lines indicate physical connections and virtual connections respectively.

PACKET REWRITING ON SDN-ENABLED EDGE SWITCHES (PARES)

Using the PRAGMA-ENT to conduct research

Multi-tenant data centers for cloud computing require the deployment of virtual private networks in an on-demand manner to enforce isolation between tenants and privacy in communication. To tackle the challenge of deploying isolated, per-tenant networks on-demand, network virtualization techniques such as encapsulation and tunneling have been widely used. However, inherent to these approaches is process overhead on end-points (such as the host hypervisor), which reduces the effective throughput between tenants as compared to the native network. This problem is exacerbated by increases in line rates, now exceeding 10Gbps.



CS = Core Switch, AS = Aggregate Switch, ES = Edge Switch, HV = Hypervisor, VM = Virtual Machines

Figure 5: PARES uses SDN primitives to program edge switches to rewrite packets sent by a VM with hypervisor source/destination addresses, and to perform the reverse mapping from hypervisor to VM address at the destination, without requiring modifications to the layer-3 physical data center network core/aggregate switches.

In this collaborative effort, PRAGMA researchers developed and characterized the performance of PARES (PAcket REwriting on SDN), a technique which uses the packet rewriting feature of SDN switches to provide multi-tenancy in data center networks at edge switches, thereby reducing the load on end-point hypervisors and improving the throughput, as compared to tunneling (Jeong et al. 2017).

We deployed PARES in two testbeds by leveraging PRAGMA-ENT end hosts and switches (see previous section, *PRAGMA-ENT*). The first testbed uses a single hardware OpenFlow switch of model PICA8 3295 and two physical machines (Intel(R) Xeon(R) CPU E5530 with 24GB memory) attached to it. The second testbed uses three physical compute nodes (with the same specification as above) and 10 Gbps Intel X520-DA2 NIC. The proposed PARES data center architecture achieved near line-rate multi-tenancy virtualization at 10Gbps link without incurring process overhead at end-point hypervisors or guest servers, which is about 4 times higher than conventional techniques.

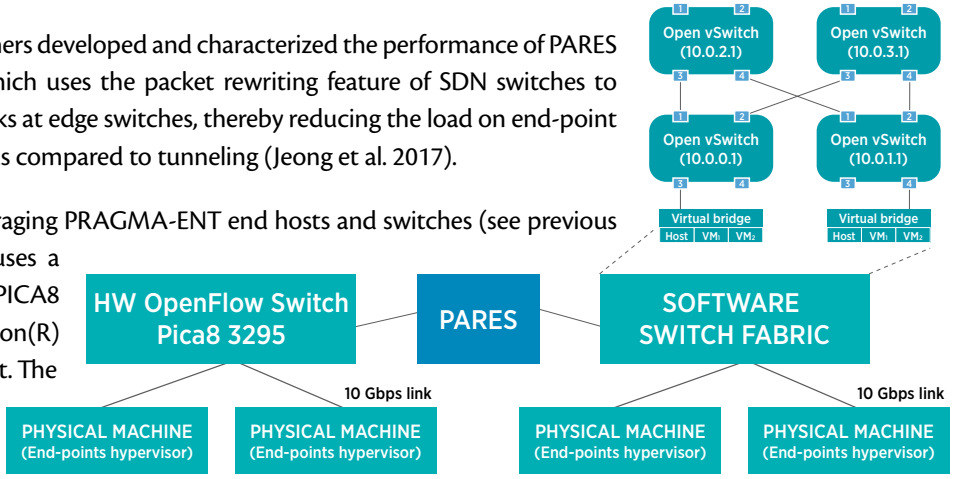


Figure 6: To evaluate the performance benefit of PARES with experiments on PRAGMA-ENT, we used hardware and software SDN switches to implement PARES packet rewriting features, as well as physical-to-host end-point hypervisors.

PARTICIPANTS

University of Florida: Renato Figueiredo, Kyuho Jeong; *Nara Institute of Science and Technology (NAIST):* Kohei Ichikawa

REFERENCE

Jeong, K., Figueiredo, R.J., and Ichikawa, K. 2017. PARES: Packet Rewriting on SDN-Enabled Edge Switches for Network Virtualization in Multi-Tenant Cloud Data Centers, *The IEEE International Conference on Cloud Computing (CLOUD)*. doi: 10.1109/CLOUD.2017.11

SHARING VIRTUAL CLUSTER IMAGES ON THE PRAGMA CLOUD TESTBED

Integrating tools from international partners to improve Cloud Testbed usability

The PRAGMA Cloud Testbed provides users with a persistent distributed international infrastructure where they can run application experiments. Different application environments are available to PRAGMA users as virtual cluster images. Currently, there are virtual cluster images for BioLinux, GRAPLEr, and DOCK applications and base virtual cluster images for Rocks and CentOS 7. These virtual cluster images can then be deployed on a PRAGMA cloud site and run locally using the pragma_boot tool.

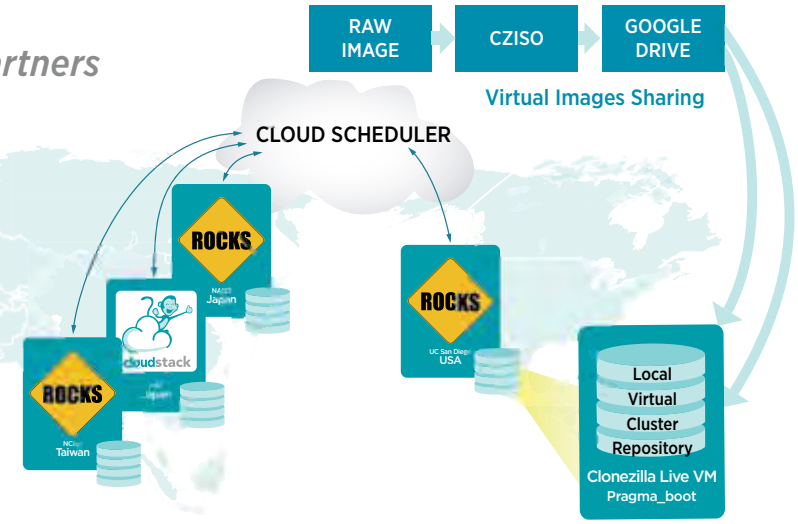


Figure 7: The PRAGMA Cloud Testbed (see section on PRAGMA Cloud Scheduler Evolution*) with Google Drive repository for storing universal cziso-formatted virtual cluster images.*

Initially, the virtual cluster images were stored in different formats depending on where they were created (e.g., ZFS volume or QCOW2) and were manually copied to the other PRAGMA sites where the virtual cluster images would be used. This created some inconsistency because not every virtual cluster image could be run at every site. In addition, there was no automated method to distribute updates for the virtual cluster images. The solution to this problem was to enable a universal format for the virtual cluster images and then convert images as needed at local sites to the site-supported format. The universal virtual cluster images are stored in an authoritative central repository (see *Software* subsection below).

To create a universal virtual cluster image format, we leveraged a partition and disk imaging/cloning tool, Clonezilla, from the National Center for High-performance Computing (NCHC) in Taiwan. NCHC enhanced the Clonezilla tool to output a virtual cluster image in a universal ISO file format. From this universal ISO format, Clonezilla converts the virtual cluster image into any local format that is supported at a PRAGMA site (e.g., RAW format). The pragma_boot developers at the University of California, San Diego (UC San Diego) created a cziso tool that provides a command-line interface to enable the conversions. The tool instantiates a Clonezilla Live Virtual Machine, and configures and runs Clonezilla to create images. It also contains a command to easily test the restored images.

For the authoritative central virtual cluster image repository, we leveraged UC San Diego's unlimited storage access on Google Drive. We created a repository in Google Drive and uploaded the cziso-generated virtual cluster image files to the repository. In order to use these universal images, we added a new cziso repository driver and a new sync repository command to the pragma_boot tool. The driver downloads the cziso images from Google Drive and runs the cziso tool to restore the images in the site-supported local image format. The sync repository command is scheduled to run nightly to compare the timestamps of images stored in the Google repository with local images timestamps. If the local images are older, pragma_boot will download and convert any updated virtual cluster images. This method of sharing cluster images is currently in use at UC San Diego, Indiana University, and NAIST and there are plans to deploy it at AIST and NCHC by the end of 2017.

PARTICIPANTS

UC San Diego: Shava Smallen, Nadya Williams, Philip Papadopoulos; *National Center for High-performance Computing (NCHC):* Weicheng Huang, Steven Shiau, Ceasar Sun

SOFTWARE

Cziso is available at *https://github.com/pragmagrid/cziso*

Clonezilla, developed by the National Center for High-performance Computing, can be downloaded from *http://clonezilla.org*

Pragma_boot is available at *https://github.com/pragmagrid/pragma_boot*

Pragma boot authoritative central repository of virtual images *https://drive.google.com/drive/folders/0B3cw7uKWQ3fXcmdfRHBCTV9KaUU*

PRAGMA CLOUD SCHEDULER EVOLUTION

Bringing international student talent to improve the Cloud Scheduler user interface

The PRAGMA Cloud Scheduler enables users to run application experiments using virtual clusters on a persistent distributed international testbed infrastructure. It is designed to be lightweight and easy to use so that sites can join the testbed and users can run virtual clusters. In our initial implementation, we leveraged a calendar reservation system called Booked (developed by Twinkle Toes Software) as our Web-based front-end GUI (Smallen et al. 2017). Booked was easy to set up, had a simple look and feel, and was customizable. It had some nice additional features such as usage reporting, support for different time zones for users, and a REST API interface. However, once we started working with Booked to adapt it for our needs, we found that any needed code changes were very time-consuming due to its heavily object-oriented structure. The code did not have the structure to handle all our use cases. In addition, the user web interface for multiple resource reservation is more complex that the original room reservation scenario.

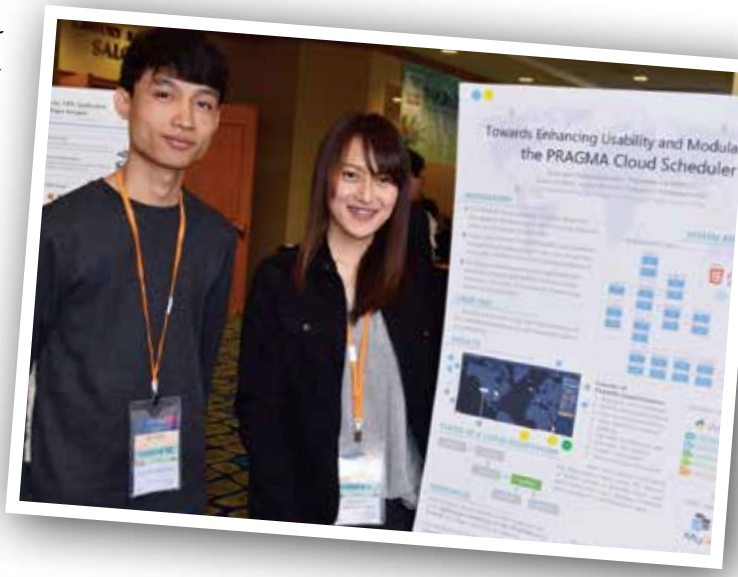


Figure 8: Undergraduate students Nannapas Banluesombatkul and Prapansak Kaewlamul from Thammasat University presenting their posters on the new PRAGMA Scheduler Web-interface they developed at PRAGMA 32.

During the PRAGMA 31 meeting, a student hackathon was held for undergraduate students at Thammasat University. One of the hackathon projects was to build a Google Map web page of the PRAGMA Cloud scheduler testbed. During the two-day hackathon, a team of three students developed an elegant Google Map web page that was presented during the main PRAGMA 31 meeting and was well received by the PRAGMA audience. The project had a great potential to be used as a new GUI interface for the Cloud Scheduler and after the discussion with the Thammasat Computer Science Department and the students' advisor, it became a senior project for the students (a requirement for graduation). We spent the next eight months advising two students, Nannapas Banluesombatkul and Prapansak Kaewlamul, while they designed and developed the new Cloud Scheduler using Nose.js, React, SASS, and the Google Maps API. The

students applied human-computer interaction principles to design a user interface in a dashboard style, along with an interactive map that allows users to create and manage reservations, as shown in the figures below.

The new web-based Cloud Scheduler system is divided into three main layers. The presentation layer provides the user interface and was designed using on a MVC (Model, View and Controller) model and developed with the ReactJS framework. The business layer is based on the SOLID principle to ensure that each component has its own clear single function and is extensible in the future without excessive code modification. It was developed using an object-oriented programming paradigm in Python. The data layer is a redesigned database that is more concise and efficient for storing site, reservation, and user information, and was developed using MySQL. The students presented their new Cloud Scheduler web interface during PRAGMA 32 (poster and presentation, see Figure 8) and conducted usability studies with workshop participants. Feedback from the usability studies was incorporated back into the code and the students worked with the Cloud Scheduler developers at UC San Diego for a few weeks to finish the integration. The new Cloud Scheduler GUI was ready for users before the PRAGMA 33 workshop, in October, 2017.

FIgure 9: Cloud Scheduler interface main dashboard view.

Figure 10: Cloud Scheduler interface: details of creating a reservation.

PARTICIPANTS

UC San Diego: Shava Smallen, Nadya Williams, Philip Papadopoulos; *Thammasat University:* Nannapas Banluesombatkul, Prapansak Kaewlamul, Prapaporn Rattanatamrong

REFERENCES

Banluesombatkul, N., Kaewlamul, P., Rattanatamrong, P., Williams, N. Smallen, S. 2017. PRAGMA Cloud Scheduler: Improving Usability of the PRAGMA Cloud Testbed, 21st International Computer Science and Engineering Conference.

Smallen, S., Williams, N., Papadopoulos, P. 2017. Lightweight scheduling for the PRAGMA cloud testbed, *Concurrency and Computation: Practice and Experience.* Wiley, 29(13). doi: 10.1002/cpe.4132

SOFTWARE

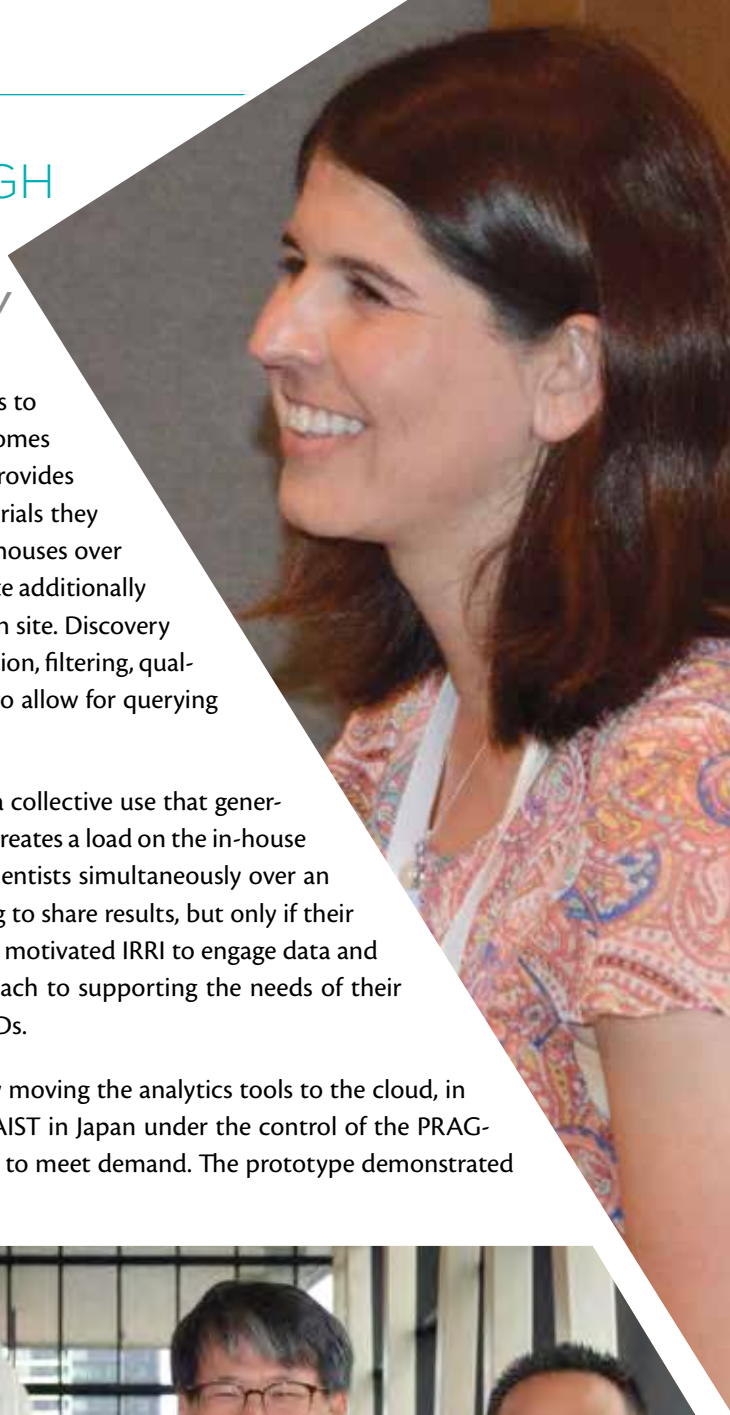
PRAGMA Cloud Scheduler GUI is available at *https://cloud.pragma-grid.net*

PRAGMA Cloud Scheduler GUI code is available at *https://github.com/pragmagrid/cloud-scheduler-gui*

PRAGMA Cloud Scheduler roll for installing on Rocks clusters is available at *https://github.com/pragmagrid/cloud-scheduler*

IMPROVING RICE YIELDS THROUGH DATA TECHNOLOGIES

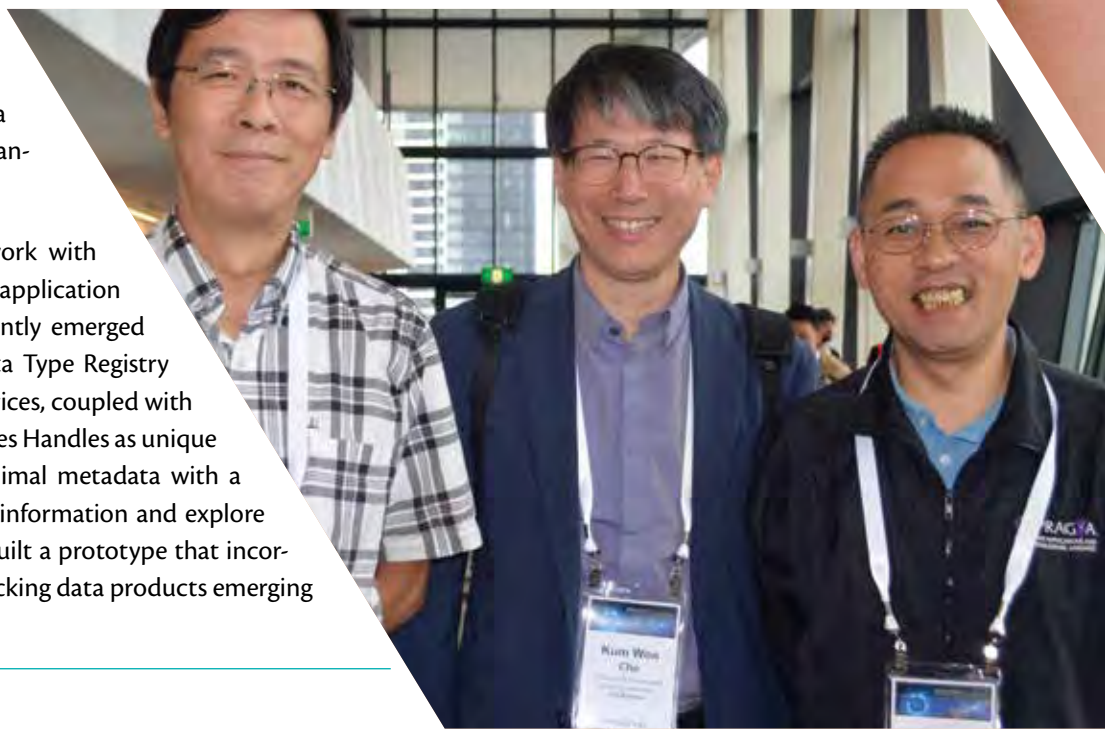
Collaborating to prototype new data functionality for International Rice Research Institute



Gene sequencing can improve rice yields by making it possible for scientists to rapidly examine individual genomes and compare them across multiple genomes to detect variants. The International Rice Research Institute[7] (IRRI), which provides scientists and rice producers in Southeast Asia with information and materials they need to develop and improve technologies that enhance rice production, houses over 3,000 rice genome sequence variants for the *indica* rice genome. The institute additionally makes it possible for researchers to use a genome analysis toolkit housed on site. Discovery through the toolkit involves identification of sites of possible genomic variation, filtering, quality evaluation, and annotation. The outputs are stored in a binary format to allow for querying the large matrices during the downstream analysis.

Researchers come to IRRI to carry out genome analysis activities, creating a collective use that generates considerable data and resource utilization. This demand subsequently creates a load on the in-house computers at IRRI, and the results of analyses carried out by numerous scientists simultaneously over an extended period of months can be difficult to manage. Scientists are willing to share results, but only if their ability to publish novel results is not hindered in the process. This dilemma motivated IRRI to engage data and computer science researchers at PRAGMA to prototype a different approach to supporting the needs of their genome analysis community that utilizes cloud computing and persistent IDs.

The shortage of compute cycles that IRRI was experiencing is addressed by moving the analytics tools to the cloud, in this case, the toolkit is deployed on virtual machines that are running at AIST in Japan under the control of the PRAGMA Cloud Scheduler (Smallen et al. 2017) which gives a scale-out capacity to meet demand. The prototype demonstrated that new VMs can be successfully instantiated. The next step is to design a scale-out that involves multiple users simultaneously using the system, and doing so in such a way that a copy of data results returns to a single managed location for subsequent curation.



PRAGMA's exploratory data science work with persistent IDs in the context of the IRRI application draws on two products that have recently emerged from the Research Data Alliance: a Data Type Registry and a Persistent ID Types API. These services, coupled with the Handle system that issues and resolves Handles as unique identifiers, is a means to associate minimal metadata with a persistent identifier. We call this kernel information and explore storing minimal data provenance. We built a prototype that incorporated the Persistent ID scheme for tracking data products emerging

[7] *http://irri.org/*

from the IRRI genomic analysis toolkit, and use provenance relationships to connect the analysis results to the input configuration parameters that generated the result. This final attribute addresses the critical need to understand the result both immediately and at various points in the future. While we demonstrated that it is possible to organize data products in a scenario where the number of scientists simultaneously using the IRRI tools is vastly scaled out, the next step is to put in place protections on a per-scientist basis so that results are not made immediately public.

The prototyping exercise gave a PhD student at Indiana University an opportunity to spend five weeks at AIST in Japan, exposure which has enhanced the student's own PhD studies.

The work carried out here resulted in the successful funding of the RPID (Robust PID) testbed (*https://rpidproject.github.io/rpid*), under NSF grant 1659310 CC* Storage: Robust Persistent Identification of Data (RPID). The RPID testbed hosts the PID assignment and resolution services used in this prototyping effort, and additionally extends the services and their use to the broader research community.

PARTICIPANTS

Indiana University: Beth Plale, Quan Zhou; *National Institute for Advanced Industrial Science and Technology (AIST):* Jason Haga; *International Rice Research Institute (IRRI):* Ramil Mauleon, Venice Juanillas

REFERENCE

Smallen, S., Williams, N., Papadopoulos, P., 2017. Lightweight scheduling for the PRAGMA cloud testbed, *Concurrency and Computation: Practice and Experience.* Wiley, 29(13). doi: 10.1002/cpe.4132

SOFTWARE AND SERVICES

Rice Genomics PID Data Service provides data management to analysis on rice genomes carried out on Virtual Clusters running Galaxy. This service is a partnership between Indiana University, National Institute of Advanced Industrial Science (AIST) in Japan) and The International Rice Research Institute (IRRI) in the Philippines.

URL: PID Data Service *https://github.com/Data-to-Insight-Center/RDA-PRAGMA-Data-Service*

URL: PID PRAGMA Extension *https://github.com/Data-to-Insight-Center/RDA-PRAGMA-Data-Service/tree/master/pragmapit-ext*

PRAGMA Data Repository is an interface to MongoDB data repository that is clearly-defined and for hosting both long-tail data objects or large data sets.

URL: *https://github.com/Data-to-Insight-Center/PRAGMA-Data-Repository*

PROV-Scaffold leverages the Komadu (Suriarachhi et al. 2015) provenance capture system to capture provenance from the Galaxy system used by IRRI.

URL: *https://github.com/Data-to-Insight-Center/Prov-scaffold*

Suriarachchi I., Zhou Q., Plale B., 2015. Komadu: A Capture and Visualization System for Scientific Data Provenance. *Journal for Open Research Software.* *https://openresearchsoftware.metajnl.com/articles/10.5334/jors.bq*. doi: 10.5334/jors.bq

CYBERLEARNING AND DISTANCE LEARNING

Continuing to Advance Computational Science and Engineering Learning Environments through the Cyber-Learning Platform EDISON

EDISON (EDucation-research-industry Integration through Simulation On the Net) is an innovative cyber-learning platform developed by KISTI (Korea Institute of Science and Technology Information) in 2011. Through the EDISON platform, a variety of high-performance computing (HPC) simulation programs (called science apps) are distributed and accessed online by researchers and students both in Korea and in the PRAGMA community. EDISON—which is funded by Ministry of Science, Technology, and Future Planning of Korea—is a multi-year joint research project between KISTI and several computational science and engineering (CSE) domain-specific research clusters, each consisting of 5–7 academic institutions. KISTI plays a crucial role and is responsible for developing and managing the platform; meanwhile the specialized groups develop, supply and use the science apps for lectures and research. Currently, the groups represent seven different CSE domains: Computational Fluid Dynamics (CFD), Nano Physics (Nano), Computational Chemistry (Chem), Computational Structural Dynamics (CSD), Computer-aided Optimal Design (Design), Computational Medicine (CMED), and Computational Electronic Magnetic & Radio Frequency (CEM&RF). EDISON began servicing the one area of CFD in 2011 and current services seven areas (listed above), with plans to continue to expand by hosting a new area every year.

The EDISON platform has been widely used and is technically robust enough to serve the chosen CSE areas in Korea and around the globe. Many simulation jobs are time-consuming and require the ability to store data for later use and for feeding to another simulation's input. To support these data-driven simulations, we have developed number of functionalities such as a simulation data transform-loader, a simulation result predictor, a simulation query interface, and a metadata database.

The system has provided 340 science apps and 568 educational contents in total and has serviced 42,852 domestic and international users (cumulative) for the past six years (as of January 2017). The growing popularity and technical excellence of the platform has enhanced the international recognition of the EDISON system, resulting in several awards and breakthroughs:

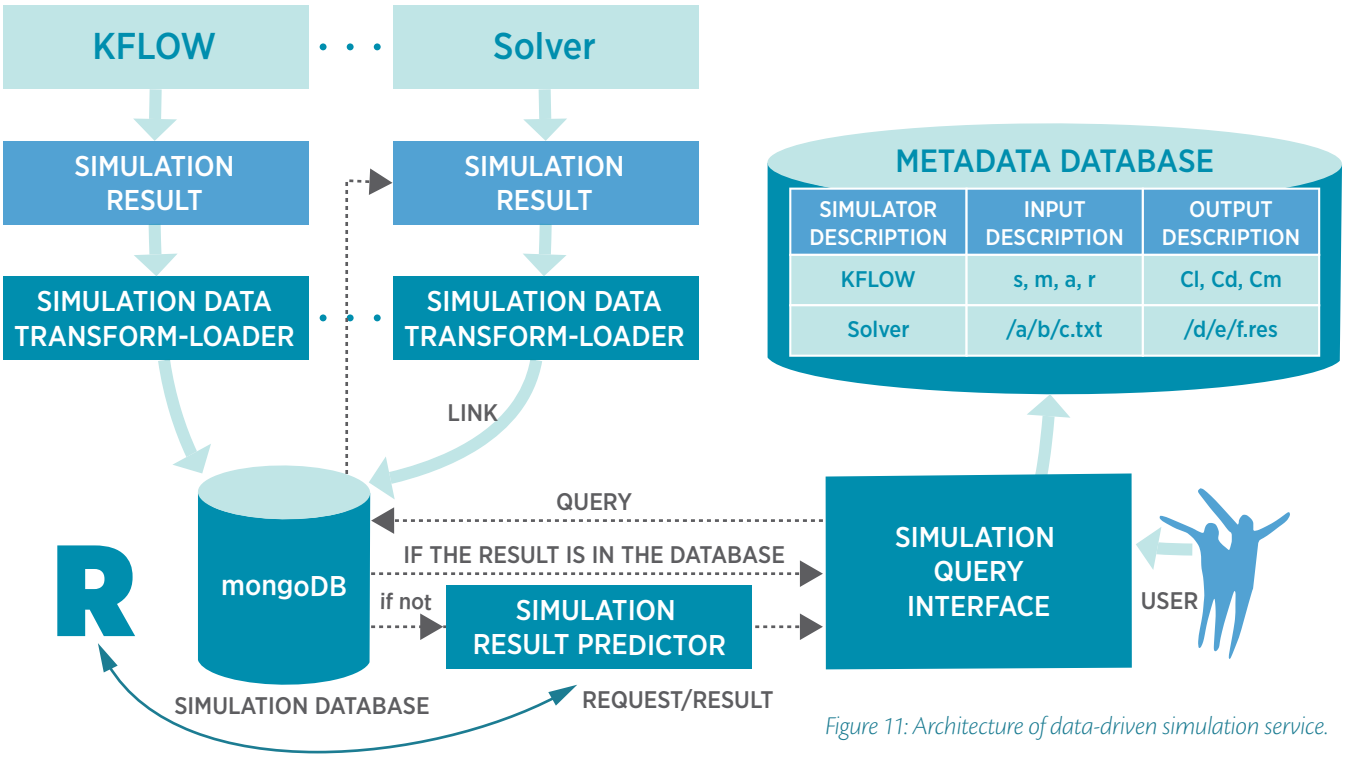


Figure 11: Architecture of data-driven simulation service.

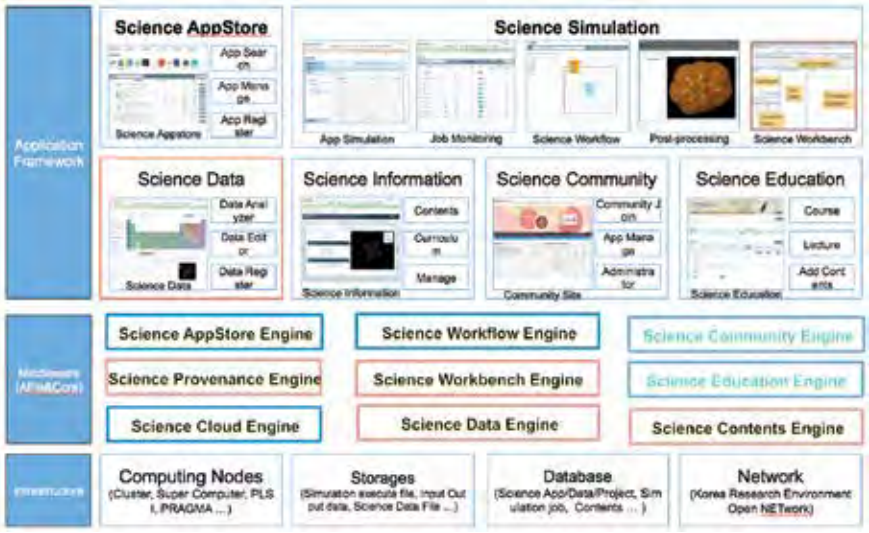


Figure 12: EDISON Portal Framework and 7 service areas.

i) 2013 IDC HPC Innovation Excellence Award at Supercomputing Conference held in Denver (first-ever for Korea);

ii) One of 50 best achievements in basic research selected by Korea National Science Foundation in 2014;

iii) Deployed copy site for one of the PRAGMA participants, National Center for High-performance Computing (NCHC) in Taiwan, January 2016;

iv) Best Poster Award at ACM/IEEE CCGrid, May 2016, held at Cartagena, Colombia, and a number of Best Paper Awards at several other conferences;

v) Internet Eco Award Korea on innovation of society contribution, July 2016; and

vi) One of 100 best national projects selected by Ministry of Science and Information and Communication Technology of Korea in 2017.

Of note, KISTI and NCHC will launch a joint research lab for close collaboration at the annual Korea SuperComputing conference (KSC) to be held in October 2017.

EDISON is a free, open-source software. KISTI has officially released the EDISON framework on GitHub (*https://github.com/sp-edison/edison-2016*), making a total of 21 portlet project directories available. The portlet directories include all the technical components needed to run and operate the EDISON platform. In addition, the site provides users with a variety of user guides for platform installation, general use, Science App development, and portal management. These guides will assist users to quickly get familiar with the system. The source code of the platform is also available at the EDISON project website (*https://www.edison.re.kr/project-download*). EDISON is licensed under the GNU Lesser General Public License. Please contact the EDISON development team at KISTI for possible collaboration opportunities.

PARTICIPANTS

Korea Institute for Science and Technology Information: Kum Won Cho, Jongsuk Ruth Lee; *NCHC KISTI Joint Research Lab:* Shyi-Ching Lin

REFERENCES

Ma, J., Lee, J.R., Cho, K. and Park, M. 2017. Design and Implementation of Information Management Tools for the EDISON Open Platform. *KSII Transactions on Internet and Information Systems,* 11(2):1089-1104. doi: 10.3837/tiis.2017.02.026

Ma, J., Lee, J-S., Cho, K. and Park, M. 2015. Analysis system design for knowledge visualization based on manufacturing data. *ICIC Express Letters.* ISSN 1881-803X, 10(4):871-876.

SOFTWARE

EDISON: *https://www.edison.re.kr/project-download*, and *https://github.com/sp-edison/edison-2016*

TOWARD AN IMMERSIVE VISUALIZATION ENVIRONMENT FOR DISASTER MANAGEMENT

Engaging Students and Leveraging the Pan-Pacific Visualization Alliance Initiated in CENTRA

Past research efforts in PRAGMA focused on natural disaster management, investigating technologies for the migration of data centers and simple information portals for disaster management professionals. This continues to be an area of active interest and collaboration among PRAGMA members. Recently, in light of the increasing amount of urban sensor data, AIST has undertaken a new research area to develop a next-generation information visualization environment using novel immersive display and interaction technologies, such as virtual reality (VR). At PRAGMA 32, AIST demonstrated a new VR application for viewing sensor data from a river in Japan—the result of a collaboration between AIST and Monash University as part of the MURPA program (Figure 13) (see section on *MURPA/QURPA*). This application leveraged publicly available data that was collected from 17,000 different sensors and rendered via an immersive 3D virtual representation of Japan. This allows users to have a country-wide view, while also allowing the user to query individual sensors to provide a local view. Preliminary user-testing has shown it to be very effective in providing a "global view" of sensor data.

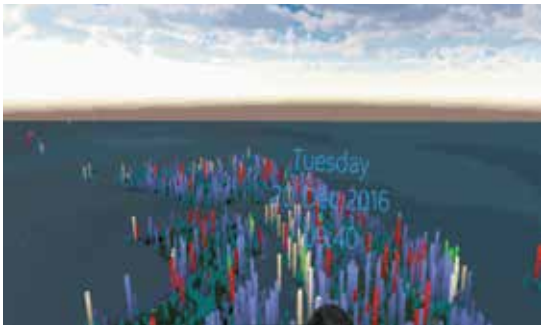


Figure 13: Screenshot of the VR river sensor data application showing the global view of the different sensors in Japan.

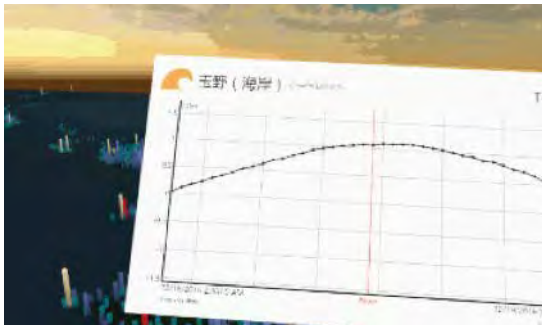


Figure 14: Screenshot of a chart from a single sensor in the VR environment showing data from a local area in Japan.

In a complementary activity, AIST is also continuing efforts to utilize large-format tiled-display walls to create a different type of immersive environment for viewing and analyzing big data during natural disasters. The focus of this activity is on a disaster management platform developed in SAGE2 (Scalable Amplified Group Environment 2; LAVA at University of Hawaii, Manoa and EVL at University of Illinois, Chicago), which would allow users to simultaneously interact with a variety of content in a shared collaborative environment. This platform could then relay relevant information to decision-makers at multiple locations during a disaster. This, combined with the VR application, has potential to transform the way disaster management professionals view heterogeneous data sets. The longer-term vision is to create an immersive platform that provides relevant data in 2D or 3D formats that can be quickly understood by decision-makers during critical situations. Future work will include more collaboration with city governments to provide user feedback of the application and the overall usability of the system. The stable PRAGMA infrastructure will become a key test-bed for these activities and enhance the Pan-Pacific Visualization Alliance established in CENTRA.

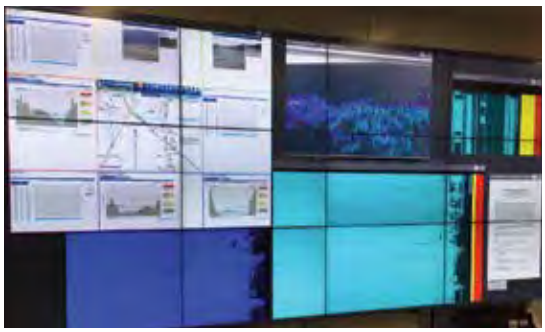
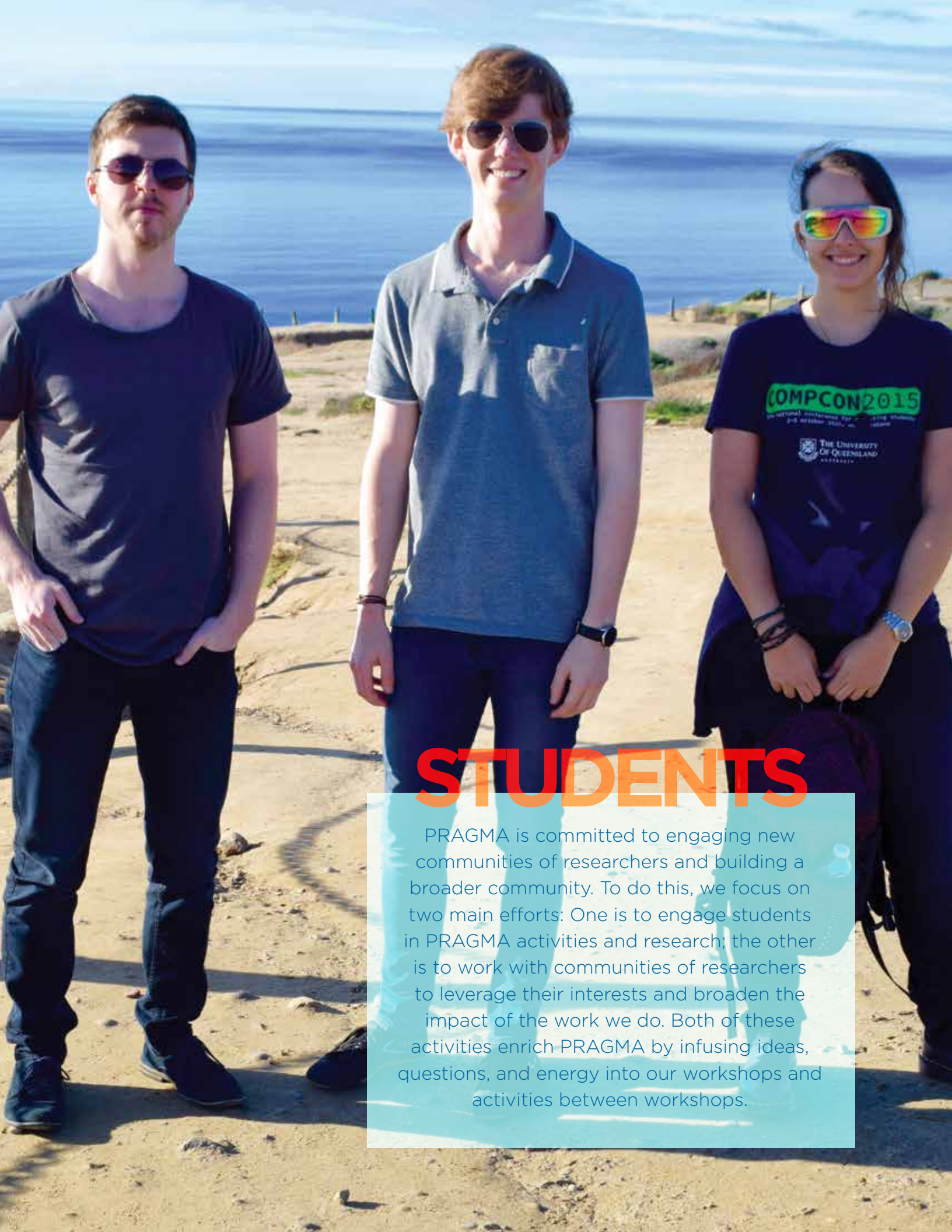


Figure 15: Image of the tiled display wall at AIST, illustrating the disaster management platform in SAGE2. A different view of the VR application can be seen in the top, center portion of the visualization wall.

PARTICIPANTS

National Institute for Advanced Industrial Science and Technology (AIST): Jason Haga; *Monash U:* Matthew Ready (MURPA Student), Tim Dwyer (MURPA Mentor)

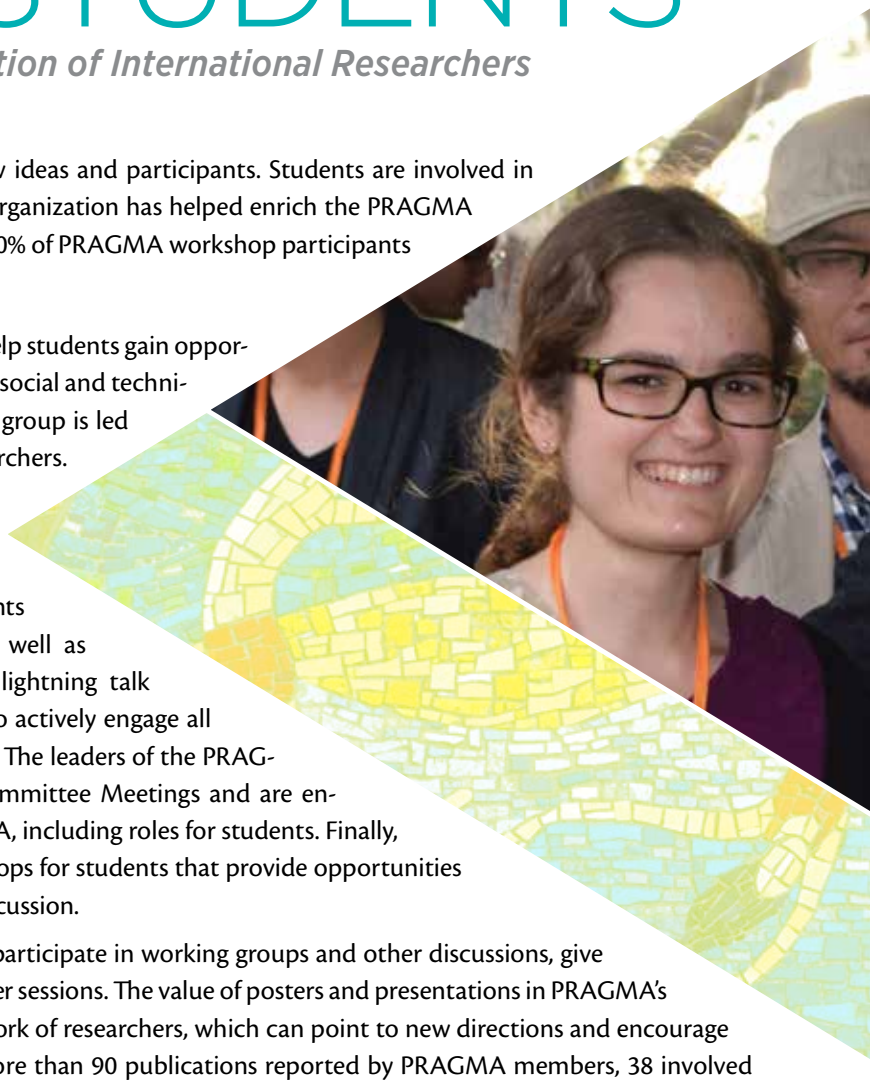


STUDENTS

PRAGMA is committed to engaging new communities of researchers and building a broader community. To do this, we focus on two main efforts: One is to engage students in PRAGMA activities and research; the other is to work with communities of researchers to leverage their interests and broaden the impact of the work we do. Both of these activities enrich PRAGMA by infusing ideas, questions, and energy into our workshops and activities between workshops.

STUDENTS IN PRAGMA & PRAGMA STUDENTS

Building a Future Generation of International Researchers



Engaging students in PRAGMA is a critical way to infuse new ideas and participants. Students are involved in all aspects of PRAGMA, and, in turn, the PRAGMA Student organization has helped enrich the PRAGMA experience for all participants. Over the last five years, about 20% of PRAGMA workshop participants have been students.

PRAGMA Students, an organization formed in 2012, aims to help students gain opportunities for professional experiences within PRAGMA's trusted social and technical networks. As a student organization inside PRAGMA, the group is led by a student committee and advised by senior PRAGMA researchers.

Current activities for students in PRAGMA include the following:

- **LEADERSHIP OPPORTUNITIES:** PRAGMA Students organize the poster presentations, involving students as well as PRAGMA participants. In addition, they organized new lightning talk sessions at the PRAGMA 32 and 33 Workshops as a way to actively engage all PRAGMA Workshop participants in the poster discussions. The leaders of the PRAGMA Student group also attend the PRAGMA Steering Committee Meetings and are engaged in discussions about the future directions of PRAGMA, including roles for students. Finally, PRAGMA Students has been involved in organizing workshops for students that provide opportunities for networking, professional development, and research discussion.
- **RESEARCH INTERACTIONS:** Students in PRAGMA participate in working groups and other discussions, give presentations on work completed or planned, and give poster sessions. The value of posters and presentations in PRAGMA's trusted environment is the feedback from PRAGMA's network of researchers, which can point to new directions and encourage students' accomplishments. In the last five years, of the more than 90 publications reported by PRAGMA members, 38 involved students as co-authors. Similarly, 12 of the 24 reported software components or improvements involved students. Specifically, at the PRAGMA 32 meeting, of the 19 posters, more than 15 were from students, with one of the top three poster awards going to an undergraduate. For the list of lightning talks and student winners, see *Workshops and Working Groups* at *http://www.pragma-grid.net/meetings/pragma-32/*.
- **RESEARCH VISITS AND PROGRAMS:** Research visits by students provide opportunities to advance project goals, provide new interactions with researchers at the host institutions, and build ties between host and home institutions and researchers. For example, Nannapas Banluesombatkul and Prapansak Kaewlamul, students from Thammasat University, visited both the University of Florida and UC San Diego as part of a project to revise a prototype graphical user interface for PRAGMA's Cloud Scheduler (See *PRAGMA Cloud Scheduler Evolution*). In addition, AIST increased the number of students it hosted in summer 2017 to approximately seven students.

Figure 16 (left to right): Camilo Vélez Cuervo (U Florida), Arianna Krinos (Virginia Tech), Gil Jae Lee (U Florida)

We would like to note that two of the members of the PRAGMA Student Steering Committee, Meilan Jiang and Quan Zhou, have received their Ph.D.s. Congratulations!

PRAGMA STUDENT STEERING COMMITTEE

Pongsakorn U-chupala (NAIST), Chawanat Nakasan (NAIST), Meilan Jiang (Konkuk), Quan Zhou (Indiana U)

PRAGMA STUDENT MENTORS

Beth Plale (IU), Karpjoo Jiang (Konkuk), Putchong Uthayopas (Kasetsart)

PRAGMA 32 BEST POSTER AWARD—1ST (TIE):

Angeline Alfred, Chris Falck, Luis Fernandes de Pina, Dylan Girard, Alexandro Gonzalez, Derya Tansel, Daniel Tola, Aleksander Weber, Jordan Williams, Camilo Vélez, Glen Walters, Yong-kyu Yoon, and Toshikazu Nishida. "*Smart Lake—A Modular Internet-connected Water Quality Monitoring System.*"

Arianna Krinos, Renato J. Figueiredo, Paul C. Hanson, Amy L. Hetherington, Kensworth Subratie, Jaikrishna T. Sukumar, Cayelan C. Carey. "*Numerical Simulation Modeling Coupled to the GRAPLEr Distributed Computing Platform Provides Insight into Lake Water Quality Responses to Climate and Land Use Change.*"

PRAGMA 32 BEST POSTER AWARD—3RD

Gil Jae Lee and José A.B. Fortes. "*Self-Tuning of Job Concurrency for Hadoop Performance Improvement.*"

PRAGMA 33 BEST POSTER AWARD—1ST

Che Huang, Chawanat Nakasan, Kohei Ichikawa,Yasuhiro Watashiba and Hajimu Iida. *"An Optimal Multipath Assignment Technique for OpenFlow Network."*

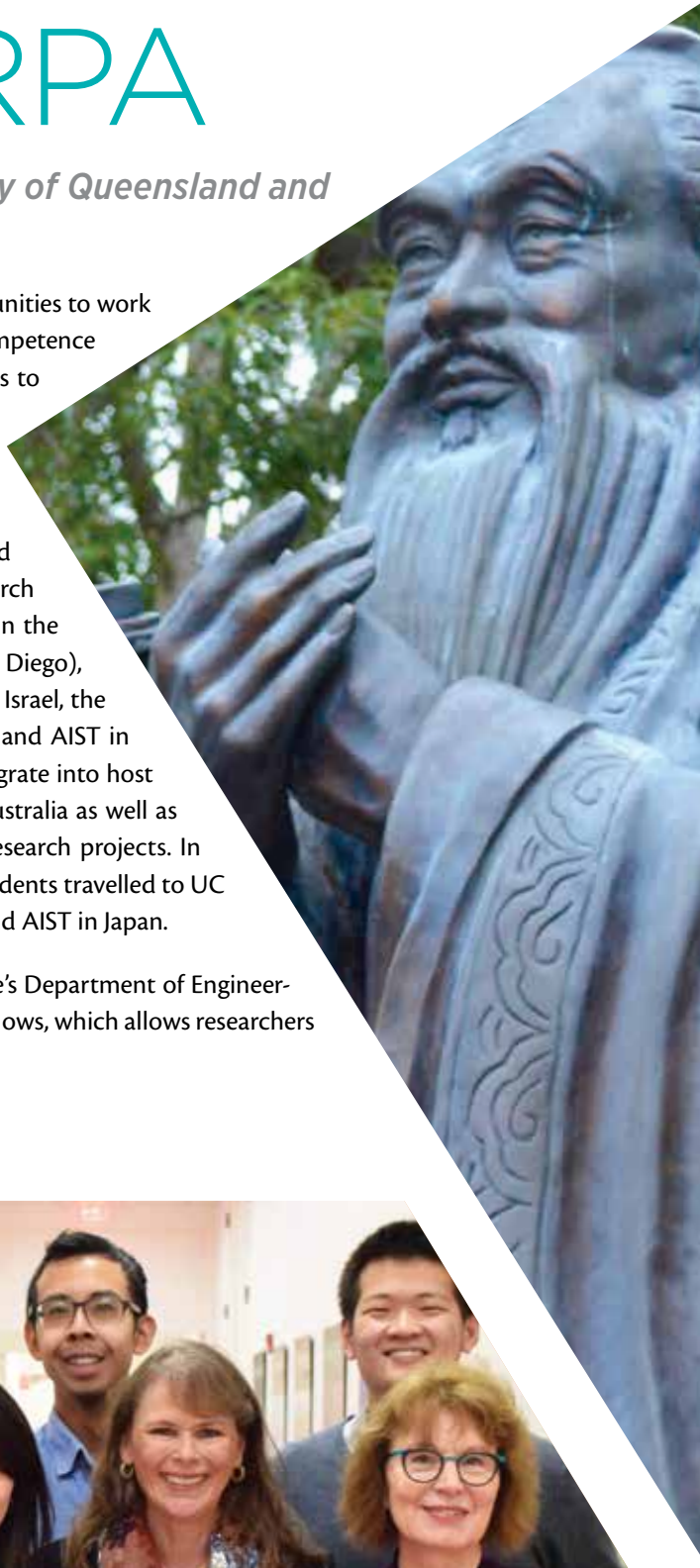
PRAGMA 33 BEST POSTER—RUNNER-UP AWARDS

Husen Rusdiansyah and Heru Suhartanto. "*Network and Storage Failure Tolerance on High Performance Computing System Using Application Migration Approach.*"

Renato Figueiredo, Ken Subratie and Saumitra Aditya. "*Dynamically Aggregating Smart Community Sensors, Edge and Cloud Resources with Overlay VPNs.*"

MURPA/QURPA

International Research Internships at the University of Queensland and Monash University



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

Since 2008 Monash University students have been travelling to international partners under the Monash Undergraduate Research Projects Abroad (MURPA) program and since 2014 under the University of Queensland Research Projects Abroad (QURPA) program. To date, 49 students have participated in the programs and have travelled to the University of California, San Diego (UC San Diego), the National Center for Supercomputing Research in Illinois, The Technion in Israel, the Institute for Infocomm Research (I2R) in Singapore, University of Warwick, and AIST in Japan. Students are placed for a period of eight weeks, allowing them to integrate into host research groups as team members. Students have both a local mentor in Australia as well as one at the remote site and often serve as a bridge between international research projects. In 2016/2017, five Monash students and two University of Queensland (UQ) students travelled to UC San Diego, the University of Warwick and Cambridge University in the UK, and AIST in Japan.

Zane Van Iperen from UQ worked with Dr. Timoleon Kipouros of Cambridge's Department of Engineering. His project concerned multi-objective optimization using scientific workflows, which allows researchers

Figure 17 (left to right): Professor David Abramson and Daya Kern, with MURPA 2016/2017 students: Sheng Long Lee, James Lee, Matthew Ready, Ben Fleming, and Ben Steer.

Figure 18: The view of a MURPA/QURPA seminar at the University of Queensland. The presenter is Professor Taisuke Boku, University of Tsukuba.

to design new products that are in some way optimal. At Cambridge, which works with a number of aerospace companies, this could mean redesigning jet engines.

Daya Kern from UQ worked with Prof. Andrew McCulloch in the Department of Bioengineering and Dr. Ilkay Altintas of the San Diego Supercomputer Center, both at UC San Diego, on another project concerning scientific workflows, but as applied to electro-physiological cardiac models. These models allow researchers to simulate the functions of the heart and to explore the effects of disease and drugs on the organ.

Both projects were performed under the supervision of Prof. David Abramson, Director of UQ's Research Computing Centre (RCC), with the help of Dr. Hoang Nguyen, an RCC expert on scientific workflows.

Monash University Computer Science student Ben Fleming travelled to the San Diego Supercomputer Center at UC San Diego and also worked with Dr. Ilkay Altintas on the WIFIRE project. WIFIRE is an integrated system for wildfire analysis, with specific regard to changing urban dynamics and climate. Ben's project concerned API (application program interface) and UI (user interface) design and machine learning techniques. The result was a program that monitors live weather data and classifies the wildfire risk as either high or low.

See section on *Toward and Immersive Visualization Environment for Disaster Management* for the work of MURPA Student Matthew Ready at the National Institute for Advanced Industrial Science and Technology (AIST).

Both the MURPA and QURPA programs include an advanced seminar scheme in which students can attend virtual seminars given by world reknown experts before they leave. These seminars allow students to "meet" potential mentors and get information about potential projects. In 2017, the seminar series focused on presenters from the Asia Pacific Region. Seminars were sourced from China, Japan, Malaysia, and Taiwan. As done in the past, seminars were broadcast simultaneously to Monash University (in Melbourne) and UQ (in Brisbane), with audiences 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.

BUILDING COMMUNITY

PRAGMA is interested in building and broadening its community, thus infusing ideas into PRAGMA and broadening the impact of PRAGMA's approach, accomplishments, and products.

To address PRAGMA's goal of making cyberinfrastructure (CI) accessible, easy to use, and useful to communities, we continue to seek out new areas and develop new and existing communities. We do this through workshops around our biannual PRAGMA workshop (see *Workshops and Working Groups*), as well as through other mechanisms. Sometimes the effort to engage new participants focuses on geography and location, and sometimes it focuses on research topics. In this section, we highlight key activities during the last year to engage more participants in PRAGMA and to disseminate results broadly.

REACHING A BROAD AUDIENCE

Dissemination of Results

In addition to face-to-face meetings (described below) to engage new participants, we disseminate our practices, experiences, and results via publication, presentations, and posters. Over the last five years, PRAGMA collectively published more than 90 articles, nearly 40 of which involved students. Most recently, a special issue of *Concurrency and Computation: Practice and Experience* (Plale and Chen, 2017) was released, with seven research papers and an overview. This special issue resulted from the first International Clouds for Data Science (PRAGMA ICDS'15) conference held in conjunction with PRAGMA 29 in October 2015, at the Universitas Indonesia. The papers in the special issue, and in general all of our publications, fall into three broad categories: research that takes advantage of the research networks and connected local clusters that connect the PRAGMA institutions; research inspired by the science expeditions defined as part of PRAGMA; and location-specific solutions that came about because of the long standing and nurturing collaborative community that PRAGMA has built and fostered over a long period of time.

In addition, PRAGMA-developed software is typically open source and available on our website (see *Reference* below). Over the last five years, PRAGMA members have created, improved, or added components to 21 published software products: seven to enable the PRAGMA Multi-Cloud (PRAGMA Cloud Scheduler, Cziso, DynIP, PRAGMA Boot, PRAGMA Boot Google Drive Virtual Cluster, Vc-out-parser, Clonzilla), one for the Lake Expedition (GRAPLEr) and one for the Virtual Biodiversity Expedition (Lifemapper); three for PRAGMA-ENT (Opimon, Overseer, smoc); two SDN (iPop, ViNe); three for data provenance (Rice Genomic PID Data Service, PRAGMA Data Repository, PROV-Scaffold); two in Biosciences (Hydra, Virtual Image Biolinux genomics tools); one in Cyberlearning (EDISON); and one in Telescience (JFG Haiku Hunt).

Finally, PRAGMA produces the PRAGMA Collaborative Overview (PCO) on an annual basis, which highlights accomplishments across PRAGMA activities.

REFERENCE

Plale, B. and Chen, M. 2017. Eds. Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) special issue on international clouds for data science, *Concurrency and Computation: Practice and Experience*. Wiley, 29(13). doi: 10.1002/cpe.4140

SOFTWARE

http://www.pragma-grid.net/products/ and individual sections in this year's Collaborative Overview

PRAGMA COLLABORATIVE OVERVIEW

http://www.pragma-grid.net/overview/

ENGAGING NEW SOUTHEAST ASIA MEMBERS

Southeast Asia International Research and Training Program (SEAIP) and PRAGMA Institute: Tainan and Kenting, December 5-9, 2016

The SEAIP's ongoing series of workshops (see *http://seaip.narlabs.org.tw/*) are organized by the National Center for High-performance Computing and are supported by the National Applied Research Laboratories and the Ministry of Science and Technology. SEAIP meetings have opened doors for collaborations between researchers in Southeast Asia and the rest of the world and have formed the basis for growing PRAGMA and CENTRA collaborations in Southeast Asia.

Figure 19: SEAIP 2016, National Cheng-Kung University, Green Magic School.

The theme of the 12th SEAIP workshop, held December 5-9, 2016, was "Cloud Computing, Big Data, and Internet of Things." Organized by the National Center for High-performance Computing (NCHC) with funding from the Ministry of Science and Technology in Taiwan, as well as from the National Applied Research Laboratory, the atmosphere of the workshop was informal and conducive to learning and developing collaborations. The first sessions were held at the Green Magic School at the National Cheng Kung University (NCKU) in Tainan, Taiwan. The NCKU Green Magic School was designated by the Architecture and Building Research Institute of the Ministry of the Interior as a "Diamond Level Candidate for Green Building Labeling" and will be awarded "Platinum Certification for Green Building" from the Leadership in Energy and Environmental Design of the U.S. Green Building Council. A second session was held at the National Museum of Marine Biology and Aquarium in Kenting. Part of this year's SEAIP meeting focused on smart and connected communities, with perspectives from several different cities and regions around the Pacific Rim, the United States, and Europe. Another part of the meeting focused on research activities from participants from Southeast Asia. *The next SEAIP meeting will be held December 4-8, 2017 in Hsinchu, Taiwan.*

CHAIR OF SEAIP: Fang-Pang Lin, NCHC

CENTRA: YEAR 2

Diversifying Opportunities for the CENTRA and PRAGMA communities

The Collaborations to Enable Transnational Cyberinfrastructure Applications (CENTRA) is creating a framework for collaboration. The CENTRA concept and partnerships are, in part, the result of PRAGMA activities in building collaboration networks and identifying opportunities to advance cyberinfrastructure research in the Pacific Rim. The founding members of CENTRA are the Advanced Computing and Information Systems (ACIS) Laboratory of the University of Florida with support from the National Science Foundation (NSF) of the United States, the Center of Excellence for Cyber-Enablement of Applications (CECEA) of the National Center for High-performance Computing (NCHC) of Taiwan with support from the Ministry of Science and Technology (MOST) of Taiwan, and the ASEAN International Virtual Organization (IVO) funded by the National Institute of Information and Communications Technologies (NICT), Japan.

This last year, CENTRA organized and facilitated multiple international meetings and scientific exchange visits that advanced collaborations and team building. They include the CENTRA 2 All-Hands meeting held April 10-12, 2017, the SUNTOWNS Workshop, under joint auspices of PRAGMA, held April 12, 2017, and the Smart Cities Student Hackathon held April 1-2, 2017, all in Gainesville, Florida. At these meetings, CENTRA researchers reviewed ongoing collaboration projects, as well as initiated projects that require contributions and resources from researchers and/or cyberinfrastructure from different countries. CENTRA further promotes these collaborations by supporting visits among researchers.

Several CENTRA-initiated projects synergize with PRAGMA activities. For example, the CENTRA-initiated and led Visual Alliance project (discussed below) has motivated planning in the PRAGMA Virtual Biodiversity Expedition to visualize data; the Airbox project was initiated under CENTRA, and its grassroots network was extended by PRAGMA members (see *highlight from PCO 2016*); and in other cases, PRAGMA activities form an infrastructure to test the ideas of CENTRA projects (e.g. using PRAGMA-ENT by the Visualization Alliance), thereby strengthening both activities. PRAGMA and CENTRA are committed to a free flow of ideas, people, and technologies across each group's activities, acknowledging diversity in applications, institutional involvement, and organizational goals.

The following are some examples of CENTRA-initiated projects, their leading investigators, and some of the research challenges:

The Visualization Alliance project, led by Jason Leigh (U. Hawaii at Manoa), considers research questions that arise when building and managing cyberinfrastructure for visualization of data and cooperative work across geographically distributed sites. Each CENTRA site has a visualization wall and associated computers and storage resources to execute a tool (SAGE2, Scalable Amplified Group Environment) that allows different walls to be shared and synchronized for ancillary data analytics and simulations. SAGE2 was developed by J. Leigh's group at Hawaii at Manoa and the Electronic Visualization Laboratory of the University of Illinois at Chicago. Current sites include institutions in the U.S. (University of Hawaii at Manoa and University of Florida), Taiwan (NCHC), and Japan (AIST and Osaka University). The domain contexts for the research questions addressed by this project include decision support and outreach in energy, coral reef, water, and disaster management. The middleware research context includes the leveraging of software defined networks to provide network quality of service, security, and integration for critical decision support. Additional IT issues considered include the use of persistent data identifiers to better enable the joining of data sets to be visualized, support for merging big-data analysis with machine learning, and the development of a tool for distributed cooperative computer-enhanced brainstorming.

The IT for Natural Disaster Management project, led by Jason Haga (National Institute for Advanced Industrial Science and Technology (AIST)), considers the use of distributed immersive visualization and analysis systems (developed under the Visualization Alliance) to support activities needed to respond to river-related disasters. See *Toward an Immersive Visualization Environment for Disaster Management* section for details.

The Persistent Identifier (PID) Kernel Information project, led by Beth Plale (Indiana University), considers the problem of enabling persistent IDs (PIDs) in environmental and biodiversity data collections, using a Robust Persistent IDentifier (RPID) testbed to evaluate PID services. In part through engagement with the Research Data Alliance, the project is defining a minimal set of metadata (called PID kernel information) that travels with PIDs and can be processed at Internet speeds by data discovery services. The novelty of the work emerges when a small amount of data provenance is included amongst the kernel information. This "universal" provenance for research data, where all data objects created as a result of scientific research have data provenance associated with them, would move us to a world where assessments of fitness and trust for secondary use of research data can be done with far higher degrees of success than can be achieved today.

The Airbox project, led by Ling-Jyh Chen (Academia Sinica, Taiwan), looks at the problem of designing and deploying affordable accurate sensors worldwide to measure particulate matter (PM2.5), aggregating data from these sensors, opening data to access by the public, and analyzing the data to explain causes and detect sources of pollution across multiple countries.

The Edge Sensor and Resources Overlay VPN project, led by Renato Figueiredo (University of Florida), looks at dynamically connecting groups of trusted distributed Internet-of-Things (IoT) devices with edge and cloud resources in the same logical network using P2P overlay virtual private networks. It studies techniques for the dynamic instantiation of stream processing workflows across edge/cloud resources. Target applications include smart/connected community scenarios where users need to collaborate or cooperate in dynamic, time-limited ways and/or form ad-hoc groups, volunteering resources to help respond to emergencies.

The SDN-IP Peering for IoT Data Transmission project, led by researchers from NICT, NCHC, and NECTEC (Thailand), studies architectures and mechanisms to federate IP Networks with software-defined IP networking for resilient and effective infrastructure for disaster mitigation and smart city applications.

Figure 20: CENTRA 2 All-Hands Meeting, Gainesville Florida 10–12 April 2017

The Distributed Lifemapper project, led by James Beach and Aimee Stewart (University of Kansas), is pursuing the use of species distribution and macroecological modeling for biogeographic, phylogenetic, and biodiversity analyses of plant and animal species in North America and Asia. The effort utilizes unique Lifemapper installations for CENTRA sites using researcher-specified or online species occurrence data along with observed climate and predicted climate data based on the Intergovernmental Panel on Climate Change (IPCC) scenarios, as well as topographic and soils data, to explain the ecological and evolutionary contributions to observed spatial patterns of species diversity. The Lifemapper software and its functionality, as well as this year's deployment at the National Center for High-performance Computing, NARLabs, Taiwan on OpenStack with Taiwanese data, is described in the Virtual Biodiversity Expedition section. In collaboration with the CENTRA Visualization Alliance, the Lifemapper team is planning to develop a compatible front-end Lifemapper application for the SAGE2 visualization wall for biogeographical pattern discovery and cooperative work across CENTRA countries.

Exchange visits supported by CENTRA help strengthen collaborations, team building, and technology transfer. This year, CENTRA supported several such visits, including James Beach and Aimee Stewart's (University of Kansas) visit to NCHC to discuss the deployment of Lifemapper and the use of high-definition data from Taiwan; the visit of Dylan Kobayashi (University of Hawaii) to NCHC to help establish a node of the Visualization Alliance; and a visit by Hui Ping Tsai's (National Chung Hsing University) visit to the University of Florida, as a follow-up to CENTRA 2, to participate in an experiment linking visualization walls in Japan, Taiwan, Hawaii and Florida.

Finally, CENTRA was designed to grow its membership as opportunities and alignment with projects arise. This year, CENTRA welcomed INESC Tec, the Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal, as a member.

EXPLORING NEW RESEARCH TOPICS AND BROADENING ENGAGEMENT

EXPLORING THE ROLE OF UNIVERSITIES IN THEIR SMART TOWNS AND COMMUNITIES

Smart UNiversity TOWNS (SUNTOWNS) Workshop, Gainesville, Florida, April 12, 2017

This workshop explored the role of universities in the area of smart and connected communities. In particular, this international workshop was developed to contribute to the acceleration of the University of Florida, Gainesville synergy by providing a forum to exchange ideas and lessons from smart and connected community initiatives in university towns from the U.S. and other countries. This workshop attracted more than 50 participants from around the United States, across the Pacific Rim, and from Europe.

Figure 21: Lake Modeling Workshop.

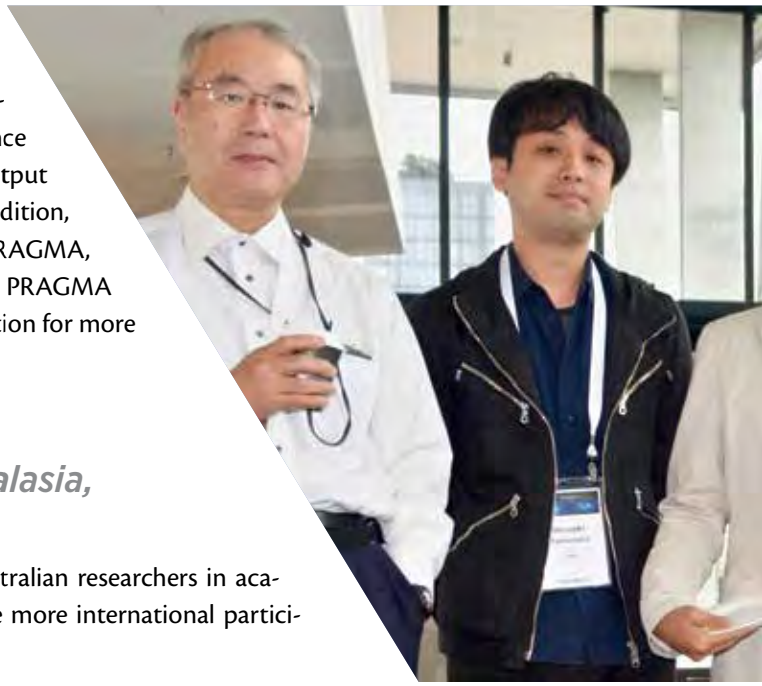
The SUNTOWNS workshop took place under the auspices of CENTRA and PRAGMA to explore both the needs of towns with universities and well as technologies to address those needs. It was clear that this dialog needs to be ongoing with commitment from all members of the community, including academia, government and industry.

DEEPENING CONNECTIONS WITH THE LAKE ECOLOGY COMMUNITY

Lake Modeling Workshop, Gainesville, Florida, April 12, 2017

Figure 22: Lake ecology researchers discuss new opportunities for developing sensor hardware at the Lake Modeling Workshop.

In conjunction with PRAGMA 32 held in Gainesville, a small group of lake ecologists and computer scientists came together in a special workshop, working collaboratively to apply lake models and the GRAPLEr cyber-infrastructure (www.graple.org) to address a specific science question focused on changing water quality in GLEON lakes. The output of the workshop was focused on developing a joint publication. In addition, for the students involved, it was an opportunity to learn more about PRAGMA, participate in working group sessions, and to share their science with PRAGMA members. See the *GLEON Research and PRAGMA Lake Expedition* section for more details as well as Figure 21 for participants.



BUILDING TIES IN AUSTRALIA (1)

Hosting of PRAGMA 33 with e-Research Australasia, Brisbane, Australia, October 18–20, 2017

The e-Research Australasia Conference is a major conference for Australian researchers in academics and industry. This year, conference organizers wanted to have more international partici-

pants at the conference. This desire overlapped with PRAGMA's interest in engaging new collaborators and participants. As a result, PRAGMA helped co-organize special sessions in both experimental and software defined networking as well as visualization for collaborative data science. PRAGMA also organized a booth for this workshop that involved demos about PRAGMA-ENT, Visualization, and Data.

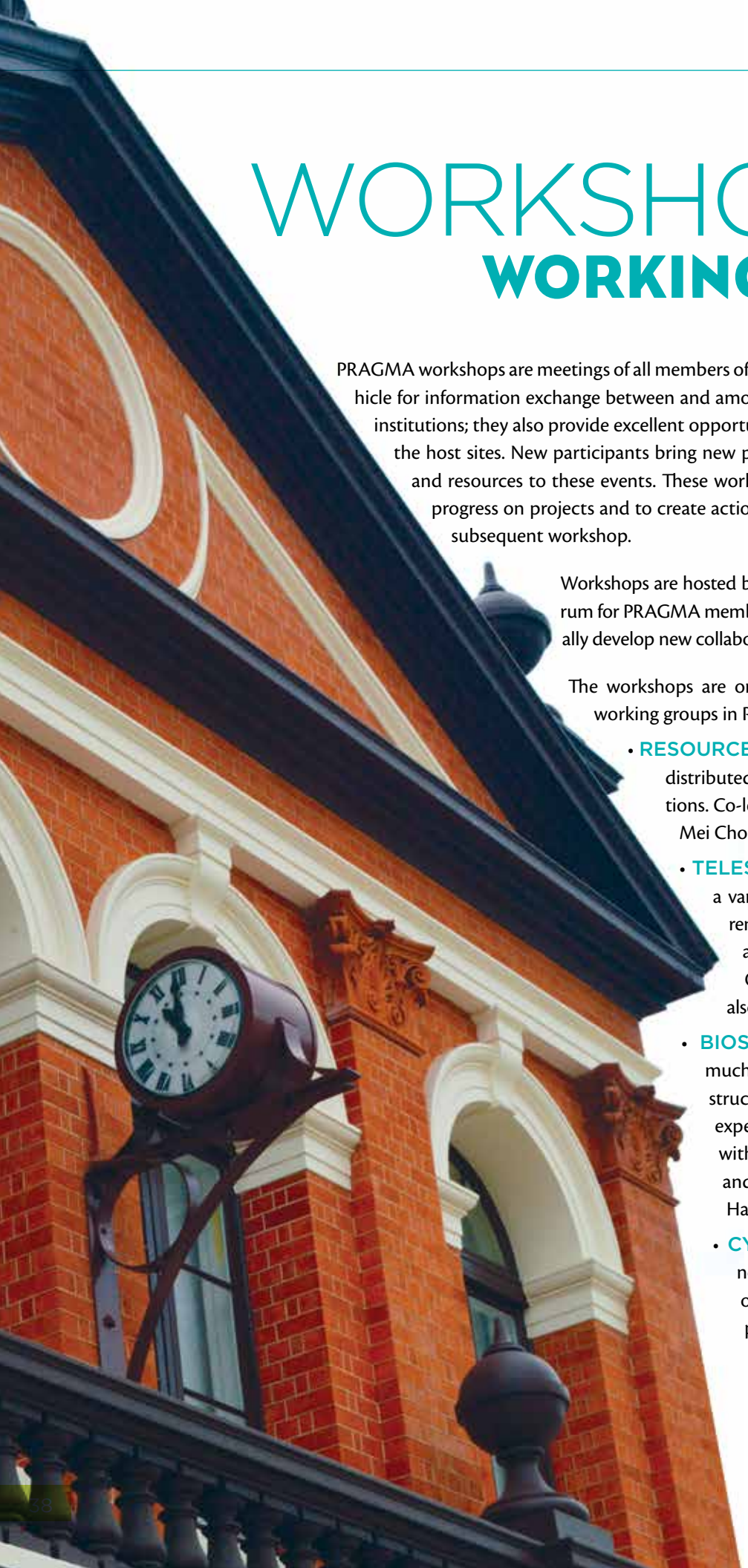
BUILDING TIES IN AUSTRALIA (2):

Participating in a Lake Modeling Workshop at Griffith University, Brisbane, Australia, October 18–19, 2017

Two members of PRAGMA's GRAPLEr Team participated in a workshop organized at Griffith University.

Approximately ten years ago a workshop was held in New Zealand to bring together lake model users and developers, from multiple countries, that contributed valuable insights into new model developments. In the intervening ten years new opportunities in the areas of software, data assimilation, databases and sensor data hold the promise of similar advances. This workshop brought together more than thirty researchers, primarily from Australia and New Zealand to look at challenges and opportunities for future model developments. The workshop was hosted by the Australian Rivers Institute at Griffith University, and two members of the GRAPLEr team participated actively. One product will be a paper on the issues and opportunities raised at the workshop, to help inform future model developments.

Figure 23: Dr. Jason Haga, AIST, demonstrates a virtual reality application for river disaster management to Dr. José Fortes, University of Florida, Gainesville at the 2017 eResearch Australasia Conference in South Brisbane, Queensland, Australia. The VR application was developed by Matthew Ready, a 2017 MURPA student who performed his internship at AIST, Japan



WORKSHOPS & WORKING GROUPS

PRAGMA workshops are meetings of all members of the PRAGMA community. They are the major vehicle for information exchange between and among working groups, expeditions, researchers, and institutions; they also provide excellent opportunities to engage new researchers and students at the host sites. New participants bring new perspectives, applications, technologies, students, and resources to these events. These workshops are a critical opportunity to demonstrate progress on projects and to create action plans for accomplishing tasks prior to the next subsequent workshop.

Workshops are hosted by different member organizations to provide a forum for PRAGMA members to meet and discuss research interests and ideally develop new collaborations with members of the hosting institutions.

The workshops are organized according to 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: Nadya Williams (UC San Diego) and Hsiu-Mei Chou (NCHC).
- **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). This group also includes activities on disaster management.
- **BIOSCIENCES WORKING GROUP:** Focusing much of its efforts over the last several years on infrastructure development that enables virtual screening experiments and computational genomics analyses with an emphasis on combating infectious diseases and, more recently, on rice breeding. Leader: Jason Haga (AIST).
- **CYBER-LEARNING:** Focusing on use of technologies to improve understanding in several areas of computational science through the use and improvement of EDISON. Co-leaders: Jongsuk Ruth Lee (KISTI) and Hsi-ching Lin (NCHC).

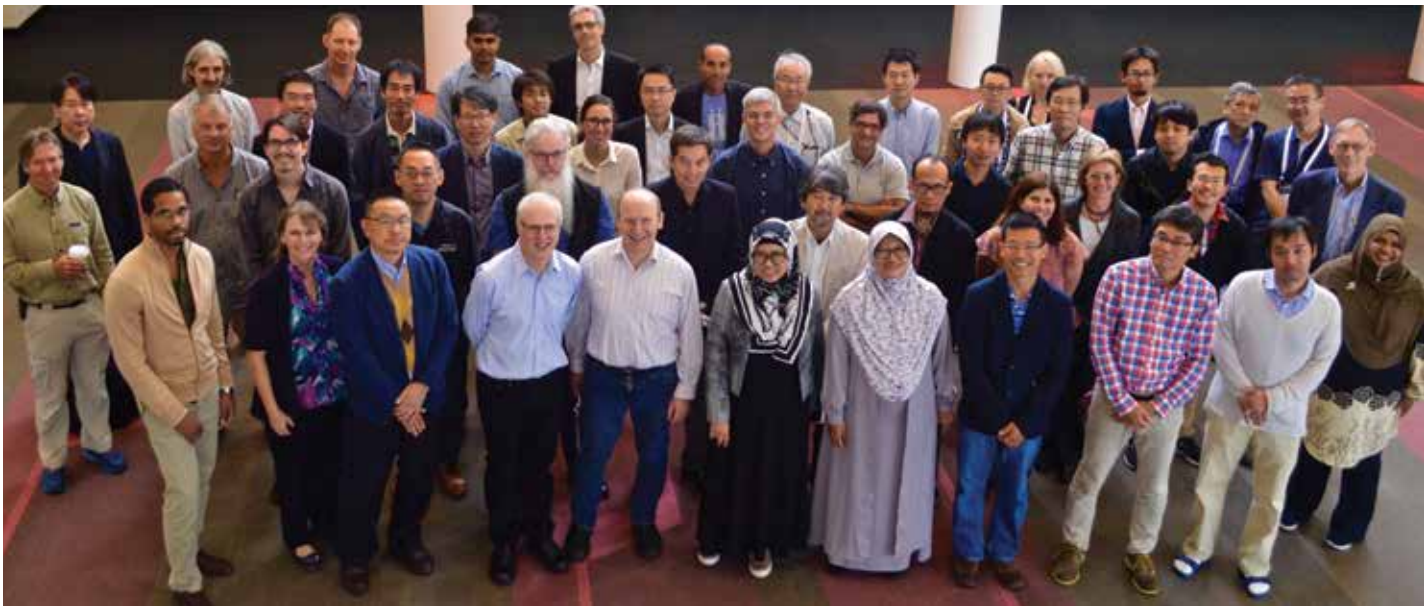


Figure 24: PRAGMA 33 attendees in Brisbane, Queensland, Australia

PRAGMA WORKSHOPS

In 2017, two PRAGMA Workshops were held:

- **PRAGMA 32:** April 12-14, 2017, University of Florida, Gainesville, Florida, United States. PRAGMA 32 was held in conjunction with the 2nd Annual CENTRA meeting and co-hosted the first Smart UNiversity TOWNS (SUNTOWNS) conference, as well as a Lake Modeling Workshop.
- **PRAGMA 33:** October 16-17, 2017, University of Queensland, Brisbane, Australia. This workshop was held in conjunction with the eResearch Australasia Conference. In addition, a Lake Modeling Workshop was held at Griffith University after PRAGMA 33.

More information about the PRAGMA 32 and 33 workshops can be found at *http://www.pragma-grid.net/meetings/pragma-32/* and *http://www.pragma-grid.net/meetings/pragma-33/*, respectively.

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

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. Listed below are our planned upcoming workshops:

- **PRAGMA 34:** May 9–12, 2018, to be co-hosted by AIST and Osaka University, Tokyo Japan. This workshop will be held in conjunction with the 3rd Annual CENTRA meeting, taking place May 14–16, in Tokyo.

We acknowledge the many contributions of members in hosting PRAGMA workshops and thank the organizers and host institutions for their efforts to ensure PRAGMA's continued success.

MEMBERS, PARTNERS, SPONSORS

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 for PRAGMA initiatives. More information about steering committee members (denoted with an asterisk * in the member list below) can be found at www.pragma-grid.net/people.php.

Active Members

A key component of PRAGMA is active involvement by participation in workshops, contributing resources, hosting workshops and/or promoting and supporting student and researcher exchanges. The following institutions have contributed to PRAGMA activities in the past year..

ADVANCED SCIENCE AND TECHNOLOGY INSTITUTE (ASTI): Jelina Tanya H. Tetangco*, *jeng@asti.dost.gov.ph*

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

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

KASETSART UNIVERSITY (KU): Putchong Uthayopas*, *putchong@ku.th*

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

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

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

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

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

THAMMASAT UNIVERSITY: Prapaporn Rattanatamrong*, *rattanat@gmail.com;* Wanida Putthividhya*, *wputthividhya@gmail.com*

UNIVERSITAS INDONESIA: Heru Suhartanto*, *heru@cs.ui.ac.id;* Wisnu Jatmiko, *wisnuj@cs.ui.ac.id*

UNIVERSITY OF CALIFORNIA, SAN DIEGO (UC San Diego): including the California Institute for Telecommunications and Information Technology (Calit2), San Diego Supercomputer Center (SDSC), Center for Research in Biological Systems (CRBS): Philip Papadopoulos*, *phil@sdsc.edu;* Peter Arzberger*, *parzberg@ucsd.edu;* Teri Simas, *simast@sdsc.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*

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

UNIVERSITY OF WISCONSIN-MADISON (UW-Madison), 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): Markus Buchhorn, *markus@apan.net*

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

STARLIGHT IRNC STARLIGHT SDX, AND GLOBAL RESEARCH PLATFORM INITIATIVES: Maxine Brown, *maxine@uic.edu*

TransPAC, INDIANA UNIVERSITY: Jennifer Schopf, *jmschopf@indiana.edu;* Andrew Lee, *leea@indiana.edu*

Other Members

CENTER FOR COMPUTATIONAL SCIENCES (CCS), UNIVERSITY OF TSUKUBA: Osamu Tatebe, *tatebe@cs.tsukuba.ac.jp;* Taisuke Boku, *taisuke@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): Ze Luo, *luoze@cnic.cn;* Yu Chen, *cy@cnic.cn*

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

MIMOS: Hong-Hoe Ong, *hh.ong@mimos.my;* Jing-Yuan Luke, *jyluke@mimos.my*

MONASH UNIVERSITY (Monash): David Abramson*, *david.abramson@monash.edu,* Paul Bonnington, *paul.bonnington@monash.edu;* Maria Indrawan, *maria.indrawan@infotech.monash.edu.au*

NATIONAL ELECTRONICS AND COMPUTER TECHNOLOGY CENTER (NECTEC): Piyawut Srichaikul, *piyawut.srichaikul@gmail.com, piyawut.srichaikul@nectec.or.th*

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

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

For more information about each of the PRAGMA Institutional Members, visit www.pragma-grid.net/members-partners.php

Additional Organizations Active in PRAGMA

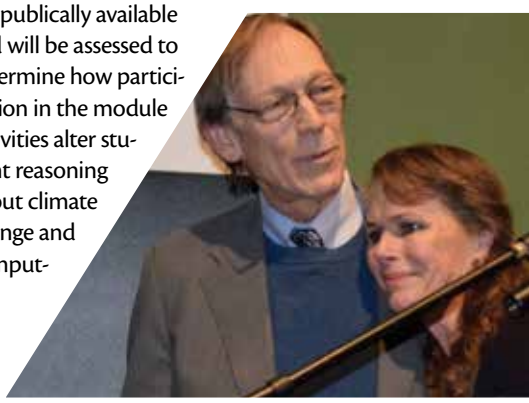
BIODIVERSITY INSTITUTE, UNIVERSITY OF KANSAS (biodiversity.ku.edu), and its researchers and students conduct research on seven continents in areas such as biodiversity informatics, systematics and ecology and evolutionary biology. They have contributed to the biodiversity expedition through participation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

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

NATIONAL INSTITUTE FOR INFORMATION AND COMMUNICATION TECHNOLOGY (NICT; www.nict.go.jp) is an incorporated administrative agency that conducts general research and development on information technology supporting the ubiquitous society of the future. NICT supported PRIME students 2009–2015 and has participated in the activities of the Telescience Working Group through support of the high-definition video conferencing testing.

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

VIRGINIA TECH Cayelan Carey and colleagues in Project EDDIE (Environmental Data-Driven Inquiry and Exploration; projecteddie.org, an NSF-funded project) have developed sensor-based and time series data analysis activities that can be integrated into classrooms to improve quantitative skills, reasoning, and increase student engagement. Prof. Carey is expanding this effort by developing additional teaching modules that use the overlay network developed as part of PRAGMA to run lake simulations of climate change scenarios. The modules will be piloted at ten universities in the 2017–2018 academic year before becoming publically available and will be assessed to determine how participation in the module activities alter student reasoning about climate change and computing.



Partners

GLEON, the Global Lakes Ecological Observatory Network, is a grassroots network of limnologists, ecologists, information technology experts and engineers who use GLEON's network of people, sensors and data to understand issues such as eutrophication or climate change at regional to global scales. GLEON, which was 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 500 members. It has developed new knowledge and insights, created new data products and developed a very successful Graduate Student Association (GSA). There are several ties between GLEON and PRAGMA, including shared personnel, shared 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 Resource Center (NSRC, 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 recently working with PRAGMA in the Southeast Asia region to support researchers from Myanmar to attend PRAGMA 24. They have also been collaborating with PRAGMA and IU on the Lower Mekong Initiative to help enable more international science education. In addition, NSRC has been able to encourage participation in PRAGMA workshops.

SPONSORS

PRAGMA is supported by its member institutions and the U.S. National Science Foundation (NSF OCI 1234983, PI: P. Papadopoulos), involving researchers at UC San Diego, University of Florida-Gainesville, Indiana University and the University of Wisconsin-Madison. This involves support from the NSF Office of Cyberinfrastructure, Office of International Science and Engineering, Division of Information and Intelligent Systems in the Directorate for Computer and Information Science and Engineering, and the Emerging Frontiers Office of the Directorate for Biological Science. This funding provides partial support for involvement of Indiana University, University of Florida and University of Wisconsin researchers. In addition, previous support came from NSF OCI-0627026, PI: P. 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. A reallocation of funds has engaged researchers at the University of Kansas for the Virtual Biodiversity Expedition, and a recent NSF-funded supplement has been able to engage researchers at Virginia Tech for development and dissemination of tools developed by the Lake Expedition.

AIST's research activities are partly supported by Japan Science and Technology Agency (JST) and New Energy and Industrial Technology Development Organization (NEDO), Japan.

ASTI's grid activities are funded by the Department of Science and Technology (DOST), Philippines.

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

CCST at Jilin University receives funding support from the Chinese Natural Science Foundation (61170004) and the Chinese Ministry of Education (20130061110052).

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

Kasetsart University'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 Ministry of Science and ICT (MSIT) through EDISON (Education-industry Integration through computational SimulatiOn on Net) Project and KISTI through its Supercomputing R&D Program.

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.

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

NECTEC a member of National Science and Technology Development Agency (NSTDA), receives major funding support from Ministry Of Science and Technology, Thailand.

Osaka University and National Institute of Information and Communications Technology (NICT) research was supported in part by institutional funding for their collaborative research project title "Research on High Functional Network Platform Technology for Large-scale Distributed Computing.

PRAGMA-ENT is supported in part by JGN of the National Institute of Information and Communications Technology (NICT), Japan.

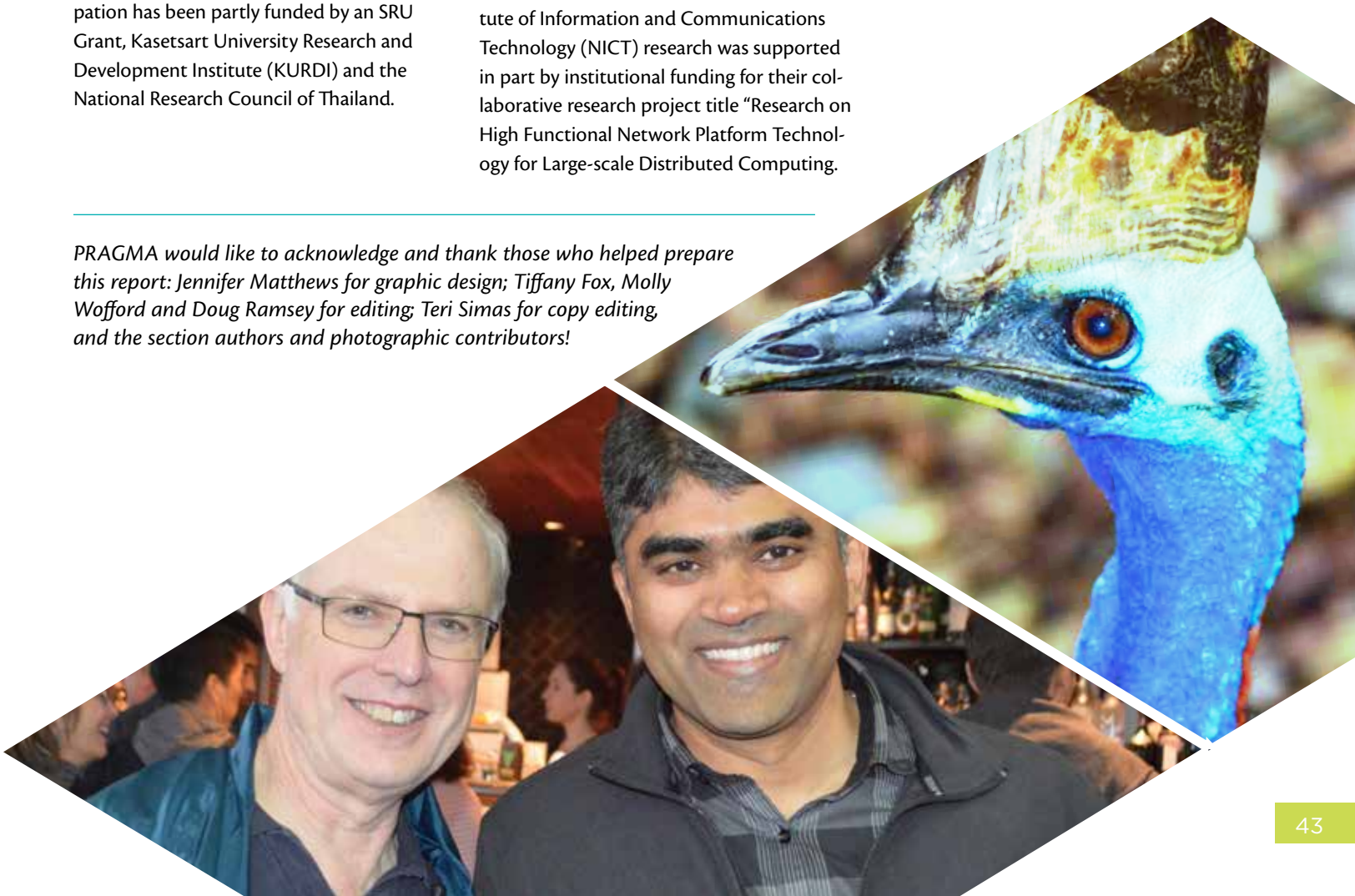
StarLight receives major funding from NSF (ACI-1450871) for "IRNC:RXP: StarLight SDX A Software Defined Networking Exchange for Global Science Research and Education."

TransPAC receives major funding from NSF and collaborates closely with PacificWave, Japan's National Institute of Information and Communications Technology, APAN, TEIN, and other Asian networking groups.

Universitas Indonesia's PRAGMA participation has been partly funded by the Faculty of Computer Science, and research grants from The Ministry of Research Technology and Higher Education, Republic Indonesia.

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

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



PRAGMA
www.pragma-grid.net