PACIFIC RIM APPLICATION AND GRID MIDDLEWARE ASSEMBLY

2018 COLLABORATIVE OVERVIEW

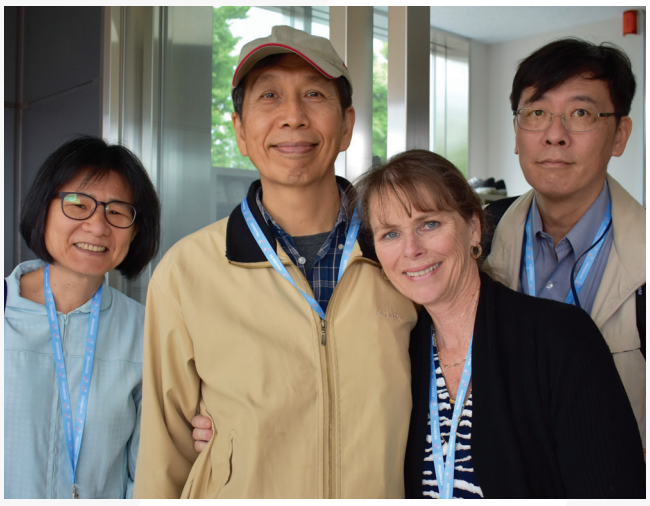
INTRODUCTION

The Pacific Rim Application and Grid Middleware Assembly, PRAGMA, is an open, grass-roots, international organization that makes cyberinfrastructure accessible, easy to use, and useful for long tail of science communities to advance their science. These communities, often international in composition, are addressing societally important problems and are frequently unable to access, fully utilize, or keep pace with rapidly evolving cyberinfrastructure (CI) technologies and approaches. Through active participation and contributions of all members, PRAGMA focuses on how to make these new and rapidly changing CI technologies usable by these communities of scientists, within a trusted envelope of shared, easy-to-use computer and data resources.

Since its inception in 2002, PRAGMA, through its workshops, projects, and students, acts as

- A **conduit of ideas** among members, for example by pushing the envelope of AI research through the recent deployment of the National Institute of Advanced Industrial Science and Technology's (AIST) state-of-the-art AI Bridging Cloud Infrastructure, and between members and the broader community of CI researchers;
- An **enabler of multi-disciplinary, multi-institutional, international collaborations**, exemplified by our expeditions in water quality, biodiversity, and experimental network research, resulting in new knowledge that often transforms fields of research;
- An **experimenter with new technologies and approaches,** often on shared, jointly-constructed testbeds, such as PRAGMA's Experimental Network Testbed (PRAGMA-ENT) and the PRAGMA cloud, for pre-production testing, measurement, and research, in some cases leading to incorporation of approaches in national infrastructure, e.g., NSF's production virtualization resource Comet;
- A **producer and refiner of community software** driven by specific applications, but generalizable to others, for example GRAPLEr, Lifemapper, and PRAGMA's cluster virtualization suite, including PRAGMA boot and Cloud Scheduler GUI; and
- A **forum for students and researchers** to build the trusted people networks of inter-disciplinary scientists for long-term successful collaborations and careers.

The PRAGMA organization actively marries application needs with technology capabilities to advance science and infrastructure, both in short-term experiments and long-term expeditions. It actively seeks and engages new participants, promotes opportunities for development of students, and works to build and maintain increasing levels of trust among its community. It promotes collaborative and individual partner accomplishments and, as appropriate, migrates tools experimented with (or developed by) PRAGMA to national and international infrastructures.



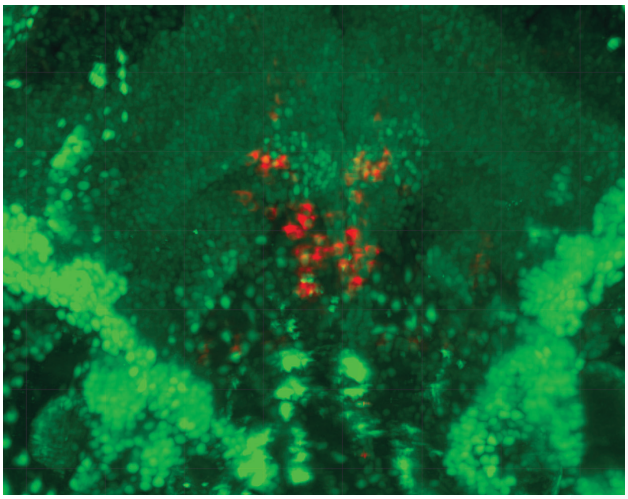
The annual PRAGMA Collaborative Overview highlights accomplishments by PRAGMA and its members. Collectively, this year's accomplishments advance PRAGMA's long-term goals of creating accessible, easy-to-use, and useful trusted computational and data resource environments (envelopes) for long tail science communities. This year's PRAGMA Collaborative Overview highlights the following specific key accomplishments:

Lake Expedition: Researchers have used GRAPLEr, an R-based open-source software, to gain new understanding of the impact of land use and climate change on water quality and have improved the usability of GRAPLEr for use in research and teaching in undergraduate and graduate classrooms. Furthermore, driven by the need for more accurate predictions of water quality, the multidisciplinary lake expedition team, through the Smart and Connected Reservoir project (which was motivated by technology advances achieved in PRAGMA projects), designed and prototyped a sensor gateway that will securely relay data from sensors in the field to data centers via hybrid software defined networking (SDN) overlays. These sensor data are to be incorporated into modeling applications, such as PRAGMA's GRAPLEr computing system, to achieve the goal of improving water quality forecasts.

Biodiversity Expedition: Understanding species distribution and conducting macroecological modeling for biogeographic and biodiversity analyses of plant and animal species are challenges for biodiversity researchers and conservationists. Lifemapper is an open-source modeling environment that supports those needs. PRAGMA supported upgrades to streamline workflows in Lifemapper and tune the software environment for optimal resource allocation. Both advances extend the usefulness of the tool for researchers by allowing faster deployment of Lifemapper at new research sites, creating software that can run efficiently on a range of PRAGMA-member hardware for local researchers, and facilitating more rapid development of new workshop training modules.

PRAGMA Experimental Network Testbed (PRAGMA-ENT): PRAGMA-ENT offers researchers complete freedom to access network resources to develop, experiment with, and evaluate new ideas without concerns of interfering with a production network. This year's efforts focused on developing an on-demand virtual local area network (VLAN) service to deploy an international path between the U.S. and Japan in an on-demand manner. In addition, researchers were able to evaluate their SDN monitoring tool in the wide-area environment as well as evaluate SDN-IP approaches using the PRAGMA-ENT environment.

Hybrid Overlay/SDN Virtual Network: Anticipating scientists' needs to securely connect sensors to data centers and networks, this activity is exploring the use of SDN in IoT/edge computing with hybrid solutions that utilize both SDN switches and overlay networks to create software-defined virtual networks.



Cloud and Virtualization: Virtualization has been an approach to allow for a flexible configuration of PRAGMA testbed resources that we refer to as the PRAGMA cloud. This year we 1) enabled multi-site virtual clusters and integration with PRAGMA-ENT, 2) experimented with GPUs and deep learning applications, and 3) explored the use of containers and a container orchestration engine, in particular Kubernetes, to allow for automating deployment, scaling, and management of containerized applications. Our goal is to provide an environment in the PRAGMA cloud where we can explore the use and measure the performance of containers for different applications.

AI Bridging Cloud Infrastructure: PRAGMA partner AIST has deployed the AI Bridging Cloud Infrastructure (ABCI) for the development and deployment of AI applications for industry and society. This unique facility opened in August 2018, and its staff have begun working with key users to facilitate their research and advance the usage of ABCI. To promote the value of ABCI and push the envelope of AI research, ABCI leadership has built partnerships with other groups, such as the Pacific Research Platform led by researchers at UC San Diego and is working with PRAGMA member the National Center for High-Performance Computing (NCHC) in Taiwan to help develop a new AI-focused infrastructure at NCHC. The ABCI team is also exploring how the PRAGMA community can build a large-scale distributed AI research platform by connecting ABCI and AI supercomputers operated by the other PRAGMA member institutions.

PRAGMA Data Infrastructure: Driven by specific needs in diverse and societally important areas of biodiversity, food security, and air quality, as well as opportunities provided by new data platforms such as the AI Bridging Cloud Infrastructure, PRAGMA is expanding its Cloud Infrastructure to an international shared data platform. Concretely, we have tests planned to mirror iDigBio biodiversity data at the NSF Comet facility, to move data from UC San Diego to the ABCI facility in Japan to support neuroscience research, and to integrate persistent identifiers (PIDs) with distributed data storage systems. Moreover, we are deploying tools to display Tstat monitoring data, are able to assign PIDs to dynamic data we work with, and have developed expertise around data storage systems and managing time series data.

Visualization: Harnessing the advent of large display walls for immersive visualization and analytics, this collaborative effort builds on Scalable Amplified Group Environment 2 (SAGE2) collaborative environments to provide visual prototypes of complex workflows in multiple areas, including eHealth and smart cities. Key technical advances include creating integrated data backends to multiple visual displays and adding voice commands to facilitate interaction with the data. This effort aims to work with diverse groups to diversify applications using this technology.

Cyberlearning: EDISON, EDucation-research Integration through Simulation On the Net, is the well-known e-science platform that is aimed at advancing learning through simulations. This year's technical advancements include implementing a new scientific workbench and new scientific workflows, which allow users to bring together various environments or modules. Its achievements have been recognized through the winning of the "Top 100 R&D for the Outstanding Achievement in 2017" award.

Community Building: PRAGMA is open to all who are willing to actively participate in PRAGMA activities, workshops, projects, experiments, and platforms. In addition, to stay aware of a broader community and seek new challenges, we make concerted efforts to reach new communities of individuals and applications via the Southeast

Asia International Research and Training Program (SEAIP) hosted by PRAGMA member NCHC in Taiwan, collocating meetings with our workshops, such as meetings on AI infrastructure and research, and by PRAGMA members attending conferences in new areas.

Students: PRAGMA is committed to creating globally-minded problem solvers for 21st century challenges. Students play a fundamental role in PRAGMA activities, engaging in both research and professional skill development at workshops. Students are also provided opportunities to gain hands-on experience working on projects and, in some cases, spending face-to-face time with researchers in a host lab. These have been mutually beneficial to the students and to PRAGMA. Some exchanges are highlighted this year, one of which has led to preliminary work on capturing time series data for measuring efficiency of virtualization approaches.

PRAGMA EVOLVES

Over the last 16 years, PRAGMA leadership has evolved gradually through rotation of members on the steering committee. Accompanying this year's selection of a new chair, we re-envisioned an organizational structure to more broadly engage members of the steering committee and PRAGMA. PRAGMA's strong people network and clear, flexible operating principles and procedures has enabled a smooth transition of leadership to a new generation of leaders around the Pacific Rim. These structural changes are summarized in PRAGMA at-a-Glance. This year, we welcome five new members to the steering committee: Jason Haga (AIST), Weicheng Huang (NCHC), Ruth Lee (KISTI), Nurul Malim (USM), and Shava Smallen (UC San Diego). We also would like to thank five individuals who have helped shape PRAGMA: Philip Papadopoulos, former PRAGMA Chair and a member since its inception, Whey-Fone Tsai, a member since inception, Karpjoo Jeong who hosted PRAGMA 23 and helped start programs specifically targeting PRAGMA students, Putchong Uthayopas, who co-hosted PRAGMA 12 and PRAGMA 24, and who joined PRAGMA very early, and Wing-Keung Kwan, who hosted PRAGMA 20 and has been an enthusiastic participant. Finally, we wish to thank Teri Simas for her efforts to help make PRAGMA the family it is today.

LOOKING FORWARD

In the coming years, we will build on our current activities and plans described in this Collaborative Overview. PRAGMA's success will be weighted by

- New science from the expeditions and new insights from our experimentation and measurements;
- Use of tools and other PRAGMA products by the broader community and incorporation of some into national and international infrastructures;
- New collaborations generated by PRAGMA and new communities able to use these new cyberinfrastructure technologies;
- Student opportunities for research and professional development; and
- Publications that share the results of our work with the broader community.

PRAGMA's future is bright thanks to its strong people-network, partners working together to make CI technologies useful to new scientific communities and their challenges, and its investment in people who will be future leaders in their fields.

REFERENCES

For more about PRAGMA, its past accomplishments, and software, please see:

Arzberger, P. (2017) A reflection on the origins, evolution, and future of PRAGMA. *Concurrency Computat: Pract. Exper.* e4136. doi.org/10.1002/cpe.4136.

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

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


台中
高美濕地日落

AT A GLANCE

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

Members: 27 Institutional Members. *See Institutions and Sponsors* for a list of institutions and abbreviations used in this Collaborative Overview.

Website: www.pragma-grid.net

Governance: Steering Committee and PRAGMA's Operating Principles and Procedures

Steering Committee Co-Chairs: Shinji Shimojo (Osaka University) and Shava Smallen (UC San Diego)

Chairs of other steering committee functional area roles:

- **Membership:** Fang-Pang Lin (National Center for High-performance Computing, NCHC)
- **Workshop Engagement:** Kohei Ichikawa (Nara Institute of Science and Technology, NAIST)
- **Communications:** Heru Suhartanto (U Indonesia)
- **Mentoring:** Jason Haga (National Institute of Advanced Industrial Science and Technology, AIST)

Collectively, the PRAGMA Steering Committee Co-Chairs and the Functional Area Role Chairs constitute an executive committee of the steering committee.

Workshops of Participants: Workshops are held twice a year to share progress and plan future activities. Open to all, workshops are hosted by PRAGMA members, often in conjunction with other activities.

2018 WORKSHOPS

- **PRAGMA 34:** May 9–12, 2018. Co-hosted by AIST and Osaka University. Location in Akihabara, Tokyo, Japan. Held in conjunction with the Workshop on High Performance Infrastructure for AI, in conjunction with the 3rd annual CENTRA meeting.
- **PRAGMA 35:** October 3–6, 2018. Hosted by Universiti Sains Malaysia, Penang, Malaysia and co-hosted by Malaysia Digital Economy Corporation. Held in conjunction with Big Data Week Asia (Kuala Lumpur) and the Big Data Summit 2 (Penang).

2019 WORKSHOPS

- **PRAGMA 36:** April 25–27, 2019. To be hosted by the Korea Institute of Science and Technology Information (KISTI). Held in Jeju, Korea in conjunction with the 4th annual CENTRA meeting (April 22–24, 2019).
- **PRAGMA 37:** September 11–13, 2019. To be hosted by the UC San Diego, La Jolla, California, USA.

SCIENTIFIC EXPEDITIONS

- **LIMNOLOGY:** Predicting lake eutrophication and training the next generation of lake scientists
- **BIODIVERSITY:** Understanding biological adaptation in extreme environments
- **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 for diverse applications in projects like PRAGMA Cloud Testbed, PRAGMA ENT, and Open Data Platform
- **TELESCIENCE:** Making and improving access to or use of remote equipment (e.g., tiled-display walls or sensors) in areas like environmental monitoring and traffic flow
- **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 with current focus on EDISON

STUDENTS

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

PARTNERS

- **GLEON:** The Global Lake Ecological Observatory Network's mission is to understand, predict, and communicate the role and responses of lakes in a changing global environment (gleon.org).
- **CENTRA:** The long-term goal of the 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 (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 (event.nchc.org.tw/2018/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.

HIGHLIGHTS

THE GLEON RESEARCH AND PRAGMA LAKE EXPEDITION (GRAPLE) COLLABORATION

Improving usability, securely integrating sensor data into modeling via sensor gateways, and advancing understanding of water quality globally

The PRAGMA Lake Expedition is an interdisciplinary collaboration between computer scientists and lake modelers affiliated with The Global Lake Ecological Observatory Network (GLEON). This collaborative effort is advancing the current understanding of the effects of climate change and eutrophication (i.e., increased nutrient pollution of nitrogen and phosphorus, leading to increased plant growth) on harmful algal blooms in lakes. The GRAPLE team is advancing state-of-the-art 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 (see graple.org).

The lake team (led by Cayelan Carey, Renato Figueiredo, and Paul Hanson) has continued to make progress in the development of GRAPLEr, using it in several teaching and research activities throughout the year. Development efforts focused on improving the usability of GRAPLEr, particularly in regards to the handling of user accounts and access control, to facilitate use of GRAPLEr by instructors and students in classes. Assessments of undergraduate students using the GRAPLEr platform in ecology undergraduate curricula at 12 universities showed that participation in our training activities significantly increases students' computational literacy (Farrell and Carey 2018).

Beyond the classroom, recent GRAPLEr applications highlight how the distributed computing platform (spanning PRAGMA resources at the University of Florida (UF) and the Nara Institute of Science and Technology (NAIST) in Japan, as well as the San Diego Supercomputer Center's dedicated XSEDE cluster, Comet) is enabling new discoveries on how climate change can affect lake water quality (Figure 1). Students and postdocs on the lake expedition team are using GRAPLEr to model nutrient dynamics of lakes with contrasting levels of initial water quality under a range of likely future climate warming scenarios. Our temperature warming simulations reveal substantial year-to-year and among-lake variability in the responses of different water quality metrics. While concentrations of total nitrogen in the surface waters (epilimnion) consistently decrease under the highest warming scenario for all lakes, regardless of their initial water quality conditions,

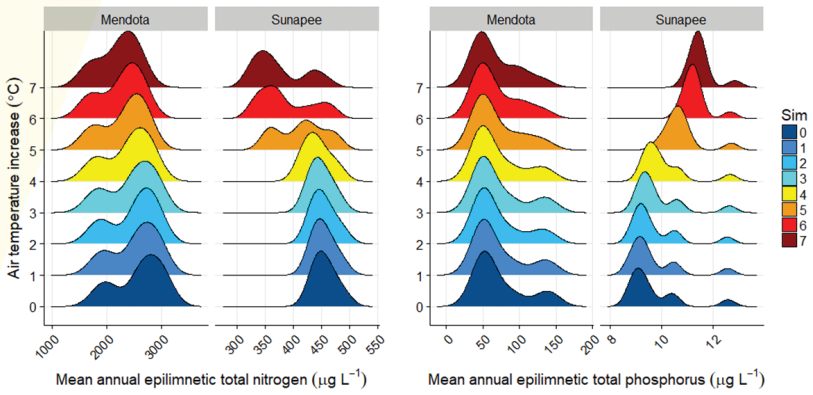


Figure 1. As air temperatures increase (from 0 to 7°C, shown by the blue to red scale), simulation modeling indicates that while epilimnetic (surface water) total nitrogen concentrations will generally not change, phosphorus concentrations will increase, but only in lakes with initially low levels of nutrients. Lake Mendota (Madison, Wisconsin, USA) has high initial levels of nitrogen and phosphorus, whereas Lake Sunapee (Sunapee, New Hampshire, USA) has low levels of nitrogen and phosphorus. These two lakes and GLEON sites provide exemplar contrasting conditions for testing warming scenarios using the GRAPLEr platform (adapted from Farrell et al. 2018).

surface total phosphorus concentrations consistently increase, but only in lakes with good initial water quality that occurs with low phosphorus concentrations (Farrell et al. 2018). These differential responses of nitrogen and phosphorus may promote greater harmful algal blooms in lakes that previously had not exhibited degraded water quality. Our observed differences in nitrogen and phosphorus concentrations in response to warming highlight the power of GRAPLEr for capturing both intra- and inter-annual variability across a range of lakes, which in turn allows us to more effectively predict ecological responses to climate change and adaptively prioritize management efforts.

Through an NSF-supported Smart and Connected Communities (SCC) project called Smart and Connected Water Systems (see SmartReservoir.org) that emerged synergistically from a PRAGMA lake expedition collaboration, a main focus of activities in the past year has been on the design and prototyping of "sensor gateways"—computers that run a full-fledged operating system, middleware, and applications, yet are small and robust enough to be deployed in the field—that connect to the Internet (typically through cellular 4G/LTE links) and to sensors on lakes/reservoirs. Sensor gateways run a full-fledged Linux operating system and connect to a "trusted envelope," allowing significantly more flexibility and programmability compared to traditional approaches and providing a layer of access control and privacy through virtual private networking (VPN) techniques that prevent access and eavesdropping from untrusted parties. The computer science prototyping and deployment of sensor gateways connected to IPOP, an open-source user-centric software virtual network, within the SCC project directly resulted from the PRAGMA lake expedition. The SCC project further expands and integrates the IPOP overlay network with modules to communicate with sensors attached to Campbell data-loggers, Git-based modules to manage committing sensor data from the gateway to servers within the VPN, management modules, and a workflow that automates data curation and publication. The SCC project's aim is to create an end-to-end workflow for ecological forecasting, where data from sensor gateways seamlessly drive the execution of models in the cloud (through GRAPLEr), within the trust envelope of a distributed overlay VPN. PRAGMA, in turn, will benefit from this new application of PRAGMA technology, which feeds back to the lake expedition's next steps. Our next goals within the PRAGMA lake expedition are to apply the lessons learned and new technology developed in the SCC project to additional GLEON lake sites, and potentially other PRAGMA expeditions.

PARTICIPANTS

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

REFERENCES

Farrell, K. J. and Carey, C. C. 2018. Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. *Ecology and Evolution.* In press. DOI: 10.1002/ece3.4363.

Farrell, K. J., Carey, C. C., Krinos, A. I., Ward, N. K., Hanson, P. C., Figueiredo, R. J., Daneshmand, V., and Subratie, K. Increasing air temperatures differentially alter intra- and inter-annual nitrogen and phosphorus cycling in a eutrophic and an oligotrophic lake. Oral presentation, Ecological Society of America, New Orleans, August 2018.

Subratie, K., Aditya, S., Mahesula, S., Figueiredo, R., Carey, C. and Hanson, P. (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,* [online] 29(13), p.e4139. Available at: dx.doi.org/10.1002/cpe.4139.

SOFTWARE

GRAPLEr. See graple.org to download and learn more about the software.

PRAGMA IPOP overlay: ipop-project.org

SMART AND CONNECTED WATER SYSTEMS

NSF 1737424 funded the Smart and Connected Water Systems as part of its Smart and Connected Communities program.

The Smart and Connected Water Systems has multiple aims, including creating an embedded network of high-frequency water quality sensors, developing and evaluating a new model-data fusion, and working directly with the community on models of trust and training materials. See smartreservoir.org for more information.

VIRTUAL BIODIVERSITY EXPEDITION AND LIFEMAPPER

Adapting Lifemapper software to varying installation environments, user backgrounds, data size, and computational complexity via upgrades, optimization tuning, and training workshops

Figure 2: Lifemapper workshop at NCHC, May 2018.

Lifemapper, an open-source modeling environment created at the University of Kansas, allows researchers to pursue species distribution and macroecological modeling for biogeographic and biodiversity analyses of terrestrial species. The Virtual Biodiversity Expedition (VBE) is utilizing unique Lifemapper installations for PRAGMA sites using species occurrence data, specified by a researcher, and phylogenetic (evolutionary) data derived from DNA sequencing studies to explain the ecological and evolutionary contributions to observed spatial patterns of species diversity.

PRAGMA work includes enabling Lifemapper to more efficiently utilize a range of available resources from small virtual cluster installations for individual researchers to the NSF XSEDE Comet supercomputer installation for high throughput projects. As part of this effort, code was streamlined and refactored to create more flexible workflows via Makeflow and Work Queue, a framework developed by Cooperative Computing Lab at The University of Notre Dame. As part of a separate project, Lifemapper also added a new meta-community phylogenetic analyses (MCPA) capability to link evolutionary speciation and geologic and graphic hypotheses to analyses of species distributions. We increased the utility of these MCPA results by creating a new browser-based client for Lifemapper servers and output packages. The new client enables browsing through output packages created for multi-species experiments and utilizes multiple linked views, including a phylogenetic tree, a map, and scatterplots to visualize different dimensions of the data simultaneously (Figure 3). Selections in one visualization window trigger selections in the other windows. While this work was primarily

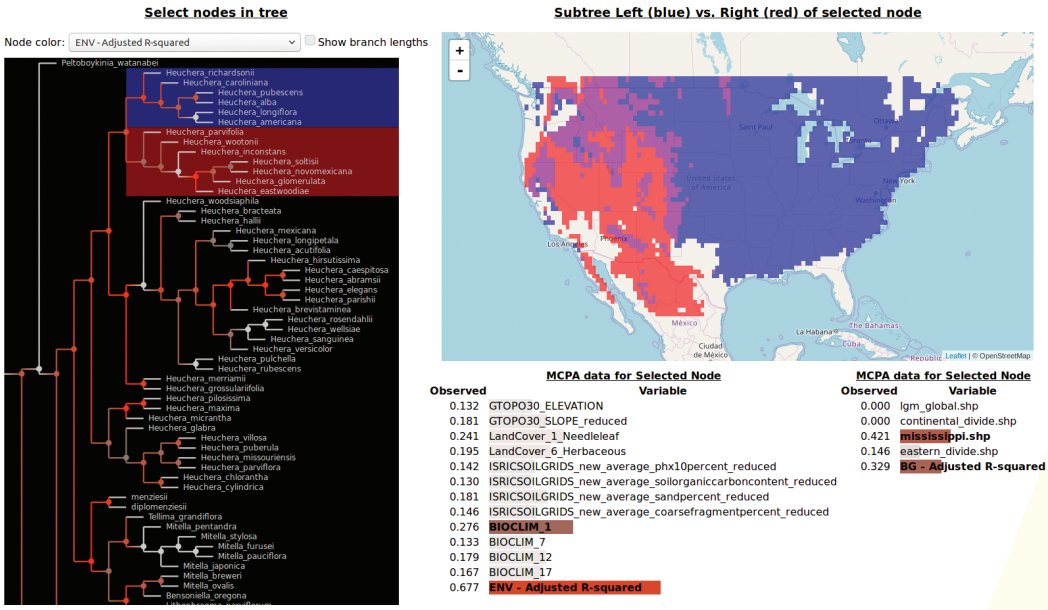


Figure 3. Linked view of phylogenetic tree, map of highlighted species distributions, and variable correlations displayed in Lifemapper.

funded by other sources, the new client was built with the intention of facilitating user interaction, data exploration, and training in PRAGMA's "Distributable Lifemapper" installations and workshops. In collaboration with CENTRA, the new client is browser-based to enable easy porting to a SAGE2 (Scalable Amplified Group Environment). SAGE2 is a software and collaborative virtual environment that allows local and remote participants to visualize data on large, high-resolution displays. It is increasingly available at PRAGMA sites. While we are working towards a native SAGE2 Lifemapper client, multiple users running Chrome or the standalone SAGE2 user interface client on their own machines can currently interact with a Lifemapper client running on a SAGE2 server.

The Lifemapper toolkit is designed to work with Rocks clusters using the Rocks roll packaging system, allowing it to be deployed and run in different computational environments. Rolls contain RPM (Package Manager) format packages that install and configure the Lifemapper software for either virtual or physical hardware. This spring, we deployed Lifemapper at the National Center for High-performance Computing (NCHC), a member of National Applied Research Laboratory in Taiwan, with GBIF (Global Biodiversity Information Facility)-provided species occurrences and high-resolution environmental data for Taiwan. We developed new training materials, and in May 2018, PRAGMA, CENTRA, and NCHC jointly sponsored a workshop for 30 students and researchers at NCHC (Figure 2) with the new Lifemapper client where participants were introduced to using Lifemapper for predicting species distributions, macroecological modeling, and biodiversity analyses, highlighting phylogenetic tree integration. Participants have joined a Lifemapper user group mailing list to share information and strategies for Lifemapper-enabled research.

Lifemapper software improvements over the last five years have allowed researchers to consider increasing the scale of problems addressed. Deploying Lifemapper to high-end resources facilitates those studies. In particular, Lifemapper is deployed in the USA using virtual clusters on the XSEDE program's Comet supercomputer resource, enabling agile and rapid scaling to match computational load. We populated and computed North American plant distribution data (using data from iDigBio[1]) on this virtual Comet cluster and also identified problems of scaling to available resources. With the computation and analysis of distribution for all of North America plants, researchers can address major questions of ecological and evolutionary importance with significant conservation implications. For example, we can ask: Where are endemics, species unique to a certain area, located? And we can formulate important hypotheses based on biogeographic features and test them. For example, does the Mississippi River represent a major differences in phylogenetic diversity to the east and

[1] www.idigbio.org

west of the river? In short, the resultant computation is a resource for researchers of North American plants, conservationists, as well as ecologists from diverse disciplines.

We are currently focusing on optimizing available resources for the size and complexity of a dataset, and on enabling scientists to use their own datasets. During the summer of 2018, we identified areas of improvement for workflow optimizations that will lead to the higher throughput of Lifemapper analyses. A major goal is to redesign computational jobs in order to streamline and simplify individual job submissions. In the process, we will develop a more accurate picture of CPU and memory requirements for individual job types so we can better allocate and use resources. Another exciting area of our current work is streamlining the data ingestion process, which will allow multiple users to work with multiple data packages while maximizing resource utilization.

PARTICIPANTS

University of Kansas: James Beach, Aimee Stewart, C.J. Grady; *UC San Diego:* Philip Papadopoulos, Nadya Williams; *National Center for High-performance Computing (NCHC):* Fang-Pang Lin, Hsiu-Mei Chou, Jih-Sheng Chang

REFERENCES

Grady, C.J., Stewart, A.M (22 July 2018) Training in "Using Digitized Herbarium Data in Research: Applications for Ecology, Phylogenetics, and Biogeography" [Workshop]. Held at Botany 2018, Rochester, MN, 20-26 July 2018. www.2018.botanyconference.org/engine/search/index.php?func=detail&aid=55

Stewart, A.M, Grady, C.J. (4 May 2018) Lifemapper Workshop. Full day workshop hosted by NCHC, Hsinchu, Taiwan.

Williams N., Stewart A., Papadopoulos P. Virtualizing Lifemapper software infrastructure for biodiversity expedition. *Concurrency Computat: Pract Exper.* 2017;e4137. https://doi.org/10.1002/cpe.4137

SOFTWARE

Lifemapper website lifemapper.org.

The core Lifemapper code is available at 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 github.com/pragmagrid/lifemapper-server and github.com/pragmagrid/lifemapper-compute contain code for installation of the two primary components (server and compute) of Lifemapper; these components may be installed on the same or separate Rocks clusters.

PRAGMA EXPERIMENTAL NETWORKING TESTBED (PRAGMA-ENT)

Extending PRAGMA ENT's capabilities and using ENT to evaluate new tools and approaches

The PRAGMA Experimental Network Testbed (PRAGMA-ENT) expedition has the goal of constructing an international software-defined networking (SDN)/OpenFlow testbed for use by PRAGMA researchers and collaborators. PRAGMA-ENT is breakable in the sense that it offers complete freedom for researchers to access network resources to develop, experiment with, and evaluate new ideas without concerns of interfering with a production network. PRAGMA-ENT also provides networking support to the PRAGMA multi-cloud and user-defined trust envelopes. This exposes SDN to the broader PRAGMA community and facilitates the long tail of eScience by creating new collaborations and infrastructure among institutions in the Pacific Rim area.

The PRAGMA-ENT team has been connecting resources in the United States (University of Florida, UC San Diego and Indiana University); Japan (Nara Institute of Science and Technology (NAIST), National Institute of Information and Communications Technology (NICT), Osaka University, National Institute of Advanced Industrial Science and Technology (AIST), and National Institute of Informatics (NII)); Taiwan (National Applied Research Laboratories (NARLabs)); Thailand (Thammasat University), and Malaysia (MIMOS) (Figure 4). 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. This year, researchers used PRAGMA-ENT to evaluate their SDN monitoring tool in the wide-area environment (Watanakeesuntorn et al. 2017). Another researcher developed a hybrid overlay/SDN environment for IoT/edge computing (See Hybrid Overlay/SDN virtual network section).

We are currently working to develop a Network Function Virtualization (NFV) testbed that uses PRAGMA-ENT resources together with PRAGMA Cloud resources. We are trying to develop the environment so that researchers can conduct various experiments on NFV using our globally distributed computing and networking resources. Some researchers currently are evaluating SDN-IP using the environment (Aguirre et al 2018).

We have been also working to develop an on-demand VLAN service to deploy an international path between the U.S. and Japan in an on-demand manner. One of our international paths, Science Information NETwork (SINET), provides network services interface (NSI) to control international VLAN paths. Using the on-demand VLAN service will allow us to develop more dynamic and flexible environments for the experiments.

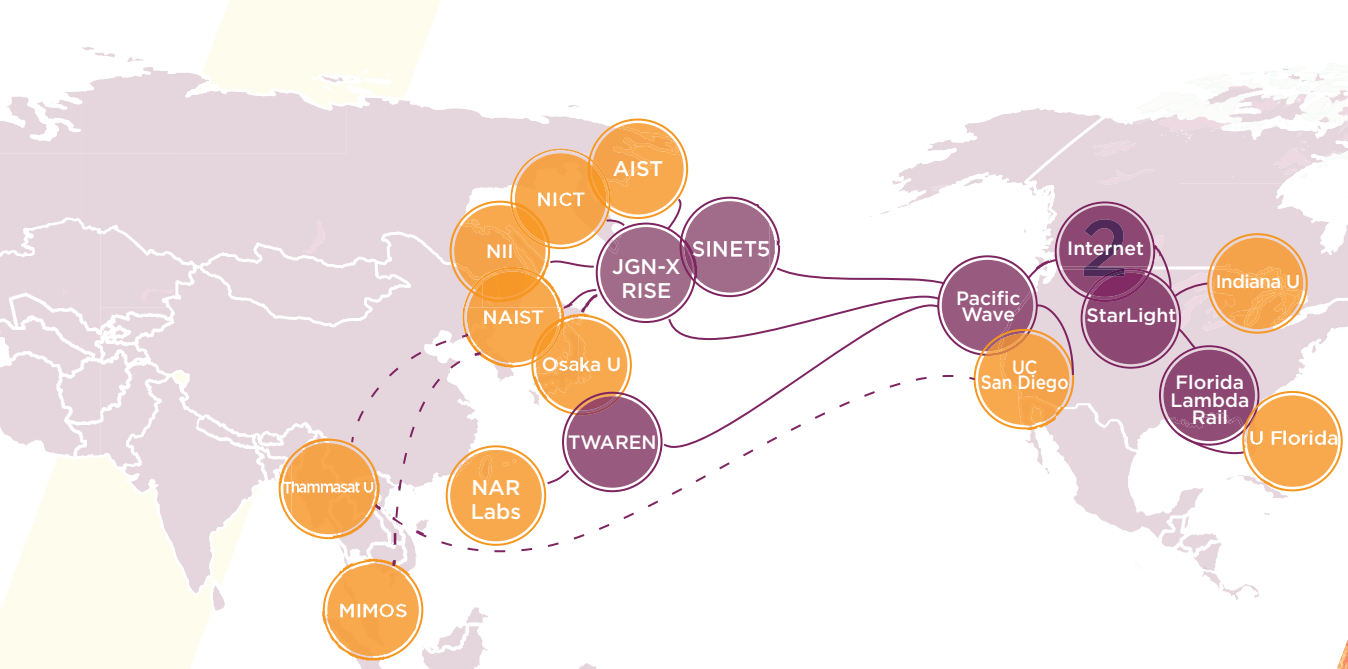


Figure 4: PRAGMA-ENT L2 Backbone. High-speed research networks (Florida Lambda Rail, Internet2, JGN-X, SINET, TWAREN, PacificWave and StarLight) interconnect OpenFlow-enabled hardware switches in the United States (University of Florida, UC San Diego, and Indiana University), Japan (Nara Institute of Science and Technology, National Institute of Information and Communications Technology, Osaka University, National Institute of Advanced Industrial Science and Technology, and National Institute of Informatics), Taiwan (National Applied Research Laboratories), Thailand (Thammasat University), and Malaysia (MIMOS). The solid and dashed lines indicate physical connections and virtual connections, respectively.

PARTICIPANTS

Nara Institute of Science and Technology (NAIST): Kohei Ichikawa, Wassapon Watanakeesuntorn, Pongsakorn Uchupala (now at Sony), Chawanat Nakasan (now at Kanazawa Univ.), Che Huang (now at NTT); *University of Florida:* Matthew Collins, Kyuho Jeong, Renato Figueiredo, Jose Fortes; *National Institute of Information and Communication Technology (NICT):* Hiroaki Yamanaka, Jin Tanaka; *UC San Diego:* Nadya Williams, Shava Smallen, Philip Papadopoulos; *Osaka University:* Juan Sebastian Aguirre, Yoshiyuki Kido, Susumu Date, Yasuhiro Watashiba, 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, FangPang 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

Aguirre, J. S., Ichikawa, K., Kido, Y., Date, S., Shimojo, S. (2018) Application aware traffic engineering functionality for an SDN transit network. PRAGMA 34 workshop, May 2018.

Watanakeesuntorn, W., Uthayopas, P., Chantrapornchai, C., Ichikawa, K. (2017) Real-time monitoring and visualization software for OpenFlow network. 15th International Conference on ICT and Knowledge Engineering, November 2017.

HYBRID OVERLAY/SDN VIRTUAL NETWORK

Extending SDN technologies to IoT devices via hybrid software defined virtual networks

Software defined networking (SDN) is a mature technology widely used by network administrators in the datacenter. Its use within the datacenter and on backbone connections across datacenters leverages the fact that a single or few administrative entities own the data plane (switches, routers, links) and can manage it through a centralized control plane. Leveraging SDN techniques to connect the multitude of IoT (Internet of Things) devices across the Internet, such as those found in smart cities, health, or environmental applications, presents a very different set of challenges; nonetheless, SDN exposes key primitives to packet handling, which provides a basis for software-defined edge networking.

We are investigating the use of SDN in IoT/edge computing with hybrid solutions that utilize both SDN switches and overlay networks to create software-defined virtual networks. In our approach, a software-managed overlay network takes the responsibility of creating peer-to-peer virtual links, by establishing tunnels across the public Internet—even when devices are in different network providers and constrained by middleboxes, such as network address translators (NATs) and firewalls. The overlay network virtual links terminate in ports of SDN-controlled software switches, allowing packets sent/received by IoT, edge, and cloud resources that reach a switch to be forwarded across the overlay to other nodes. The resulting virtualized overlay network can create a software abstraction of the layer 2 (physical addressing) network and, in doing so, simplify the complexities associated with the deployment of middleware and applications across edge and cloud resources on the wider Internet. In addition, the overlay network allows dynamic membership and grouping of resources and enforces authentication and privacy in communication, addressing key management and security concerns.

This year, we have enhanced the IPOP overlay network to allow tunneling of traffic from/to SDN-programmed switch ports across the Internet (Subratie et al 2018). In our approach, switches are software modules, and virtual links are overlay tunnels also implemented as software running on endpoints to encapsulate one protocol within another (Ethernet/IP over IP in our approach with IPOP tunnels). Together, the hybrid overlay/SDN approach is thus fully software-defined and can be instantiated in a variety of resources: e.g., physical machines, virtual machines, or containers. As a next step, we are investigating the use of PRAGMA-ENT as a platform for evaluation of the hybrid approach, with software-based SDN switches (Open vSwitch) at the edge and ENT hardware switches at the core.

PARTICIPANTS

Renato Figueiredo, Ken Subratie *(University of Florida)*

REFERENCES

Subratie, K., Figueiredo, R., “Towards Island Networks: SDN-enabled Virtual Private Networks with Peer-to-Peer Overlay Links for Edge Computing”, to appear, *Proceedings of IDCS,* Tokyo, Japan, 2018

PRAGMA CLOUD AND VIRTUALIZATION

Enabling experiments on the PRAGMA Cloud testbed

The PRAGMA Cloud testbed provides a persistent set of resources for researchers to run their applications. To reduce administrative overhead for participation in the PRAGMA Cloud Testbed, we leverage existing virtualization technologies (e.g., Rocks, CloudStack, OpenNebula) at our partner sites. We have developed lightweight tools, such as pragma_boot, which enables a virtual cluster (VC) to be seamlessly booted at different PRAGMA partner sites and automatically "fixes" the instantiated images to fit local site requirements (e.g., local IP addresses, DNS names, number of compute nodes, and PRAGMA-ENT interfaces); cziso, which leverages Clonezilla to store virtual cluster images centrally in Google Drive in a universal format and then converts to the locally supported image format (e.g., KVM+ZFS volume or KVM+QCOW2) at a partner site; and a Cloud Scheduler GUI that enables users to schedule when and where to execute their virtual clusters (Figure 5). The PRAGMA Cloud Testbed currently has a total of more than 500 CPUs and ~1.5 TB of memory amongst five resources in Japan, Taiwan, and the United States. The PRAGMA Cloud GUI is accessible from cloud.pragma-grid.net. Interested users should contact pragma-cloud-admin@googlegroups.com for an account.

Last year, a new version of Cloud Scheduler GUI was developed by undergraduate students Nannapas Banluesombatkul and Prapansak Kaewlamul, advised by Prapaporn Rattanatamrong, from Thammasat University. Since then, we deployed it on our production server, configured it with current PRAGMA cloud resource details, and demonstrated its capabilities during PRAGMA 33 in Brisbane, Australia. The students authored a paper that was presented and published in the 21st International Computer Science and Engineering Conference in Bangkok, Thailand in November 2017.

During January–February 2018, an undergraduate student from Chungnam National University in South Korea, Eunjeong Jang, worked on the development of a prototype administrative interface for the Cloud Scheduler GUI.

Other updates to the PRAGMA Cloud testbed are described in the following subsections.

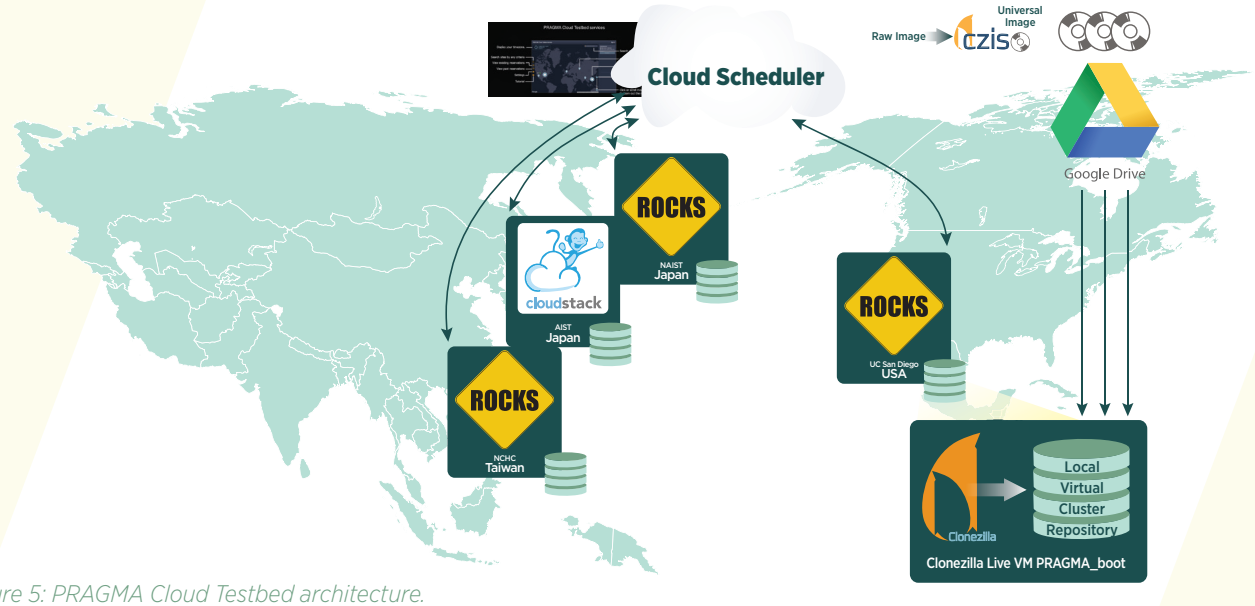


Figure 5: PRAGMA Cloud Testbed architecture.

WORK WITH PRAGMA–ENT AND MULTI–SITE VIRTUAL CLUSTERS

The pragma_boot tool enables a virtual cluster (VC) to be seamlessly booted at PRAGMA partner sites and automatically "fixes" the existing VC images to fit local site requirements. This past year, pragma_boot was enhanced to have the ability to configure a PRAGMA-ENT interface on a user's virtual cluster.

Enabling a PRAGMA-ENT interface on a user's VC can be used to enable a VC's access to a protected dataset or to create a multi-site VC. During PRAGMA 33 in Brisbane, Australia, we demonstrated the capability of creating a multi-site VC with one site at Indiana University (USA), another site at NAIST (Japan), and did a live boot at a third site at UC San Diego (USA) as shown in Figure 6. This multi-site capability is also being leveraged for the Network Function Virtualization (NFV) testbed described in the PRAGMA Experimental Network Testbed highlight.

To integrate this new multi-site capability to the Cloud Scheduler GUI, we engaged two additional undergraduate students from Thammasat University, Visaruth Punnium and Pasit Pongpojkasem who were advised by Prapaporn Rattanatamrong. The students added the ability to select two or more sites for a single multi-site virtual cluster reservation and also enabled advanced searching capabilities for a multi-site reservation (see Figure 7). The students presented the results of their efforts in a poster during the PRAGMA 34 workshop in Tokyo, Japan.

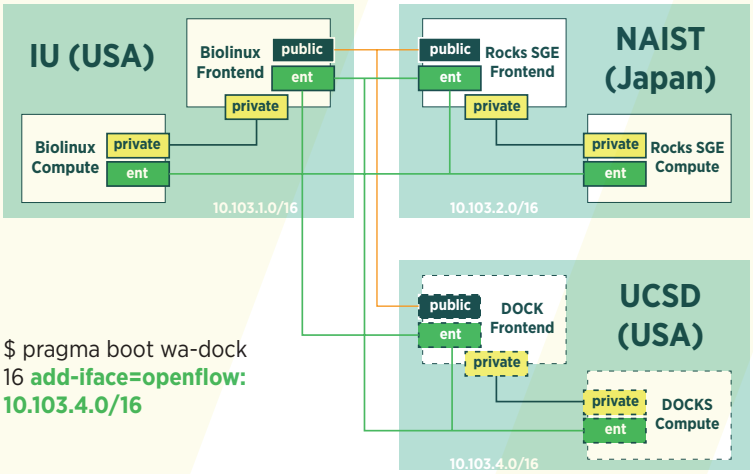


Figure 6: Multi-site virtual cluster at three geographically distributed sites, each configured with PRAGMA-ENT interfaces that enable a private network overlay amongst the resources.

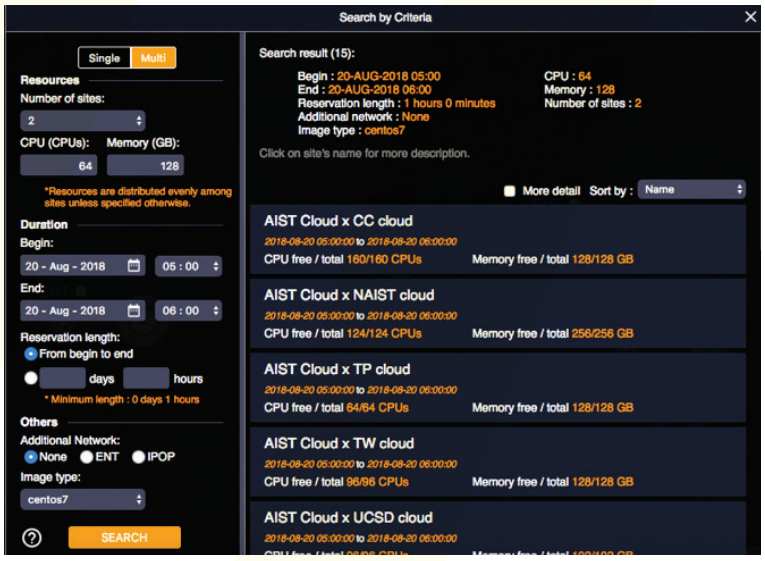


Figure 7: Screenshot of the new multi-site advanced searching functionality in the PRAGMA Cloud Scheduler GUI. Multiple choices are shown for a request of a total of 64 CPUs and 128 GB of memory—the users can select the best one that suits their requirements.

EARLY EXPERIMENTATION WITH GPUS AND MACHINE LEARNING

GPUs can offer improvements for machine learning applications involved with video and audio processing. To facilitate a more rapid exploration of these technologies, PRAGMA developed a new Rocks gpupt roll[1] that enables GPUs to be allocated or passed through to virtual machines. We installed eight Tesla and GeForce GTX type GPUs on UC San Diego PRAGMA Cloud resources and allocated GPUs to two virtual clusters for two separate experiments that were demonstrated during the PRAGMA 34 workshop in Tokyo, Japan. One virtual cluster was allocated to University of Florida for a deep learning experiment to find museum specimens that had been contaminated with mercury salts as part of the Integrated Digitized Biocollections (iDigBio) project[2]. A second virtual cluster was allocated to a team from National Chiao Tung University (Taiwan) for a deep learning experiment to process traffic monitoring images.

[1] github.com/pragmagrid/gpupt
[2] www.idigbio.org

DEPLOYING CONTAINERS ON PRAGMA CLOUD

Over the past few years, container-based virtualization has matured and become pervasive as a lightweight method to package complex application environments and microservices. A container is a lightweight, stand-alone, executable package that includes binaries, libraries, runtime environment, and configuration to run applications (see Figure 8). Containers have much shorter startup times of 1–2 minutes compared to virtual machines' 15–30 minutes. While some PRAGMA applications (e.g., lake ecology) may benefit from being packaged in containers, it may not be appropriate for an application like Lifemapper, which has complex software dependencies. We have provided an environment in the PRAGMA Cloud where we can explore the use of containers and quantitatively assess usage of both technologies, containers, and full virtualization.

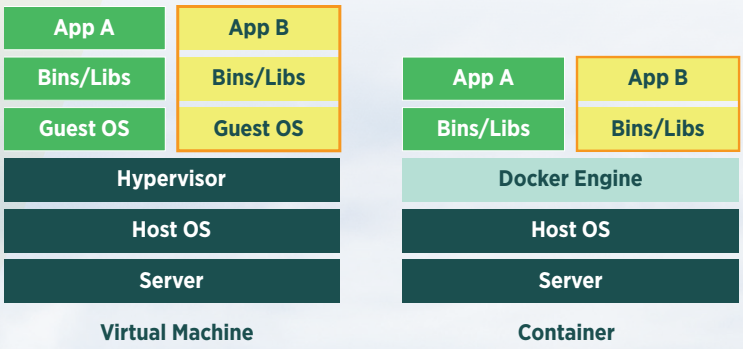


Figure 8: The architecture of a container versus a virtual machine.

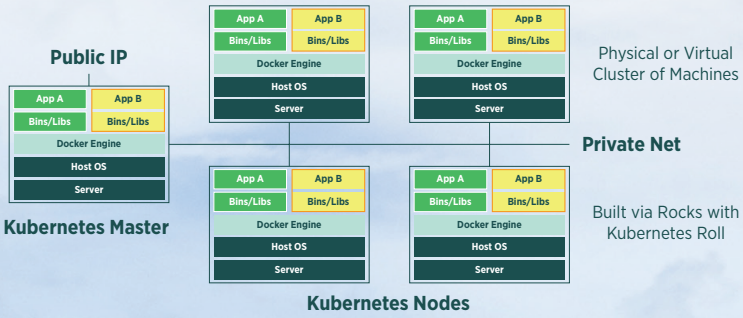


Figure 9: Kubernetes deployment on a standard Rocks virtual cluster.

Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. During the PRAGMA 34 meeting we demonstrated a new Kubernetes virtual cluster instance for PRAGMA Cloud. To create the Kubernetes virtual cluster, a new Rocks roll[1] was developed that enables easy deployment and configuration of Kubernetes on a standard Rocks cluster (see Figure 9). Kubernetes requires a specific OS kernel version, so packaging it as a virtual cluster enables it to be run on any PRAGMA Cloud testbed site. The Kubernetes virtual cluster image is based on Rocks 7, which was released in December 2017.

Over the next year, we will leverage the Kubernetes virtual cluster and use it to experiment with application containers. We believe Kubernetes will enable PRAGMA

[1] github.com/rocksclusters/kubernetes

to deploy new applications for users more quickly, and we will investigate which services could be re-packaged as Kubernetes applications. Containers will also facilitate network performance monitoring on the PRAGMA Cloud testbed as each deployed application container can inherit a common instrumented container that will be built to include monitoring tools like Tstat. Kubernetes cluster federation is emerging now, and we plan to utilize this capability with eduGAIN/OAuth2-based authentication and to investigate federation with other Kubernetes deployments, such as UC San Diego's Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI) and Pacific Research Platform (PRP) projects.

PARTICIPANTS

UC San Diego: Shava Smallen, Nadya Williams, Philip Papadopoulos; *Thammasat University:* Nannapas Banluesombatkul, Prapansak Kaewlamul, Visaruth Punnium, Pasit Pongpojkasem, Prapaporn Rattanatamrong; *National Institute of Science and Technology:* Kohei Ichikawa

REFERENCES

Banluesombatkul, N., Kaewlamul, P., Rattanatamrong, P. Williams, N., and Smallen, S. (2017) PRAGMA Cloud Scheduler: Improving usability of the PRAGMA Cloud testbed. 2017 21st International Computer Science and Engineering Conference (ICSEC). IEEE, 2017.

Punnium, V., Pongpojkasem, P., Smallen, S., Williams, N. and Rattanatamrong, P. (2018) Multi-site Cluster Reservation for PRAGMA cloud (Poster). PRAGMA 34, May 9–12, 2018, Tokyo, Japan.

SOFTWARE

PRAGMA's software products discussed in this highlight are available at www.pragma-grid.net/products.

AI BRIDGING CLOUD INFRASTRUCTURE (ABCI)

Deploying world-class infrastructure to push the envelope of artificial intelligence research

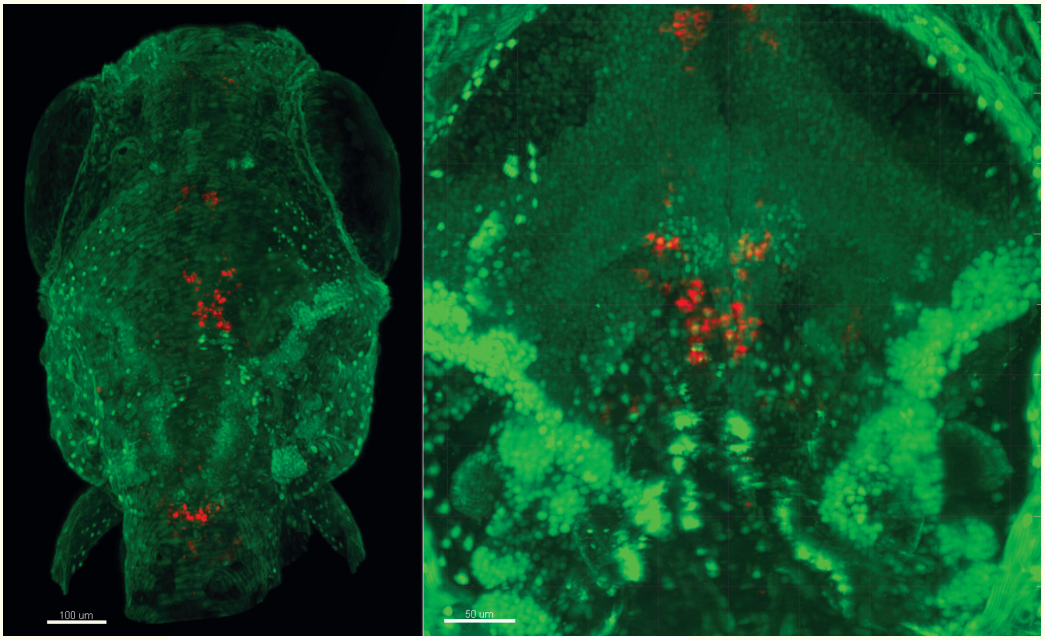


Figure 10. Fluorescent image of nuclear targeted neural activity sensors in the zebrafish brain. Activity of neurons that responded to low oxygen were identified. A particularly interesting subpopulation responding to the low oxygen environment are highlighted in red. All neurons appear in green.

With the recent global interest in artificial intelligence (AI), The National Institute of Advanced Industrial Science and Technology (AIST) in Japan is positioning itself to be one of the leaders in the development and deployment of AI applications for both industry and society. As a community resource, ABCI[1] will foster collaboration between academia, industry, and vendors of IT solutions. Researchers will be able to leave the AI "fabrication" to ABCI and focus on their research. AI technologies, especially machine/deep learning, are highly dependent on calculations at scale using big datasets to train AI algorithms and improve their accuracy. In order to address this computational challenge, AIST built ABCI, the world's first large-scale, open innovation platform for AI research and development. ABCI consists of 1,088 computation nodes with 4,352 NVIDIA Tesla V100 GPUs, 2,176 Intel Xeon Gold 6148 CPUs, Infiniband EDR, 476 TiB memory, 1.74 PB NVMe SSD, 22 PB large-capacity GPFS storage system, and high-speed network connections, all housed in an ultra-green datacenter building. By commoditizing and introducing supercomputer cooling technologies, the expected average power usage effectiveness (PUE) over a year is less than 1.1. This infrastructure achieves more than 550 petaflops (half-precision floating point) and 37 petaflops (double-precision floating point) theoretical peak performance. ABCI is currently the fastest computer in Japan and placed fifth on the global Top500 list as of June 2018, with a high performance Linpack (HPL) benchmark of 19.9 petaflops.

[1] abci.ai

This platform integrates high performance computing (HPC), big data, ultra-high bandwidth, and low latency hardware in a modern design to enable better cooperation between HPC applications and machine learning, which will become a very important factor in the future of AI. To facilitate academic AI research, we are connecting ABCI to a global, high-speed network research platform (the Pacific Research Platform[2]) to extend research opportunities to a broad number of international institutions. One of the first grand challenge projects we are supporting is mapping whole-brain-scale signal integration events in zebrafish (Figure 10) with researchers from the Salk Institute for Biological Studies and the Scripps Institute of Oceanography in San Diego. This will open the possibility of creating novel neural deep-learning architectures of the fish brain to be used in a wide variety of application areas, including robotics, sensor design, and behavioral sciences.

ABCI started full-fledged operation in August 2018. AIST is creating an ABCI utilization service to provide learned models, open data, and training data sets to its users for the acceleration of joint academic-industry AI R&D. For the PRAGMA community, this platform will be a valuable resource to further AI research between our members. Additionally, the community will gain informative "lessons learned" from operating such a platform and a reference architecture for PRAGMA members interested in establishing data centers combining HPC and AI to drive future innovations. Members from the National Center for High Performance Computing (NCHC) in Taiwan have already exchanged information several times with AIST as they move to build an AI-focused infrastructure. It is expected that the PRAGMA community will build a large-scale distributed AI research platform by connecting ABCI with AI supercomputers operated by other PRAGMA member institutions.

PARTICIPANTS

National Institute of Advanced Industrial Science and Technology (AIST): Hirotaka Ogawa, Ryousei Takano, Satoshi Sekiguchi, Yoshio Tanaka, Jason Haga; *Salk Institute for Biological Studies:* Gerald Pao, *Scripps Institution for Oceanography:* George Sugihara

[2] prp.ucsd.edu

TOWARDS EXTENDING THE PRAGMA CLOUD TO INCLUDE AN OPEN DATA PLATFORM

Integrating data servers from international partners to support a notion of FAIR data

Driven by the needs of key applications as well as the growing opportunities for artificial intelligence (AI) research infrastructure, PRAGMA has taken steps to extend its PRAGMA Cloud infrastructure to an international shared data platform. Additionally, guiding our initial steps in this extension are the Force 11 FAIR (Findable, Accessible, Interoperable, and Reusable) Principles, starting with persistent IDs, especially for data not necessarily associated with publications. By expanding the PRAGMA Cloud testbed to include an open data platform, we intend to boost the efficiency of data-driven research, facilitate novel AI research that utilizes computational resources within the PRAGMA network, and enable piloting and testing of strategies for making early-life data FAIR.

MOTIVATION AND ACCOMPLISHMENTS

Two examples of our work illustrate and underpin both the benefit and the need for the international data platform. First, PRAGMA's virtual biodiversity expedition's Lifemapper system creates distribution maps and predictions using a wealth of existing species occurrence data, combined with global climate, terrain, and land cover information. The NSF-funded, University of Florida-based project iDigBio's[1] servers are a major source of species occurrence data but are not designed to host Lifemapper cluster-based simulations. To address this need, by the end of 2018, we will have deployed an iDigBio mirror at UC San Diego, co-located to utilize the NSF-funded XSEDE resource Comet, and we will have created a data platform between U Florida and UC San Diego (via standard network paths to connect to UC San Diego's DMZ). Regular performance monitoring and synchronization of data will give us insights on how performance varies over time and make Lifemapper more efficient by having data co-located with virtual cluster capabilities on Comet.

Second, through a partnership developed through the PRAGMA community, Japan's Artificial Intelligence Bridging Cloud Infrastructure (ABCI) is initially available for some U.S. researchers' use. During summer 2018, we explored a practical method of efficient transfer of data across the Pacific from San Diego to Japan and are testing to gain insight on acceptable performance. The motivating research involves the use of novel neuromorphic deep learning architectures running on GPU-enabled nodes at ABCI, using light sheet microscopy imaging data of zebrafish brain activity that scientists from the Salk Institute in the U.S. have collected. The data and results need to be transferred efficiently and without error and need to be stored at the destination (ABCI) in a manner that makes the data immediately usable by the GPUs and deep learning architectures. (See AI Bridging Cloud Infrastructure highlight for details of the science.)

To date, our study of data movement across international boundaries has been technical in nature and has focused on the first step of making data FAIR, namely evaluating how and when to assign globally unique, persistent IDs (PIDs) to data that is early in its lifecycle. Furthermore, our efforts focus on data that needs sharing long before it is associated with research publications, of which many choices exist.

PRAGMA's strategy for FAIR data evaluation in the PRAGMA Cloud testbed draws on the Research Data Alliance (RDA) discussions around constructing an international fabric composed of data objects, where all data objects are find-

[1] Integrated Digital Biocollections, https://www.idigbio.org/

able. This vision for FAIR data assigns PIDs (handles) to data objects, with information about objects stored to a globally available type registry. PRAGMA has pioneered the notion of PID Kernel Information (PID KI), which embeds a small amount of metadata about a data object into the object's handle record. We have evaluated how and when to assign PIDs to data through two project efforts that have significant societal benefit: Rice Galaxy and AirBox.

RICE GALAXY: Rice genomic researchers seek to improve rice yields, thereby reducing world hunger, through research and access to tools and infrastructure. Rice molecular genetics, breeding, and diversity researchers have adopted sequencing-based technologies and high-density genotyping chips for genome variation discovery (e.g., single nucleotide polymorphism (SNPs), indels, large structural variants) (Juanillas et al. 2018). Rice Galaxy adopted the Galaxy bioinformatics framework[2] and built federated Galaxy resources, with shared datasets, software tools, and analysis workflows tailored to the needs of rice genetic/genome/diversity researchers and breeders. Pilot extension of Rice Galaxy to open science has the goal of advancing the platform for open access, hence making Rice Galaxy consistent with the data access policy (CGIAR[3]) that guides the Rice Galaxy community and data sharing. The design employs a hands-off technique (data provenance capture) to gather information about a researcher's rice genomics analysis as the analysis is running. Through this technique, acquired information is compiled with pre-analysis information that is available at the beginning of the analysis workflow (Luo et al. 2018). Key metadata is stored to the PID as PID Kernel Information (Luo Oct 2017, Luo Dec 2017). The data management software services were ported in 2018 to virtual machines (VMs) running at the Advanced Science and Technology Institute (ASTI) in the Philippines while using the NSF-funded RPID testbed (NSF 1659310) for obtaining test IDs. The specially instrumented and configured Rice Galaxy VMs, developed by PRAGMA, are now deployed at ASTI for beta testing.

AIRBOX: Air pollution is a major concern in modern life. Fine grained particulate matter measuring less than 2.5 micrometers, i.e., PM2.5, is particularly harmful to health. The Airbox project aims to put equipment and data sharing analysis tools in the hands of the public to monitor air quality.

At the March 2017 PRAGMA 32 meeting in Gainsville Florida, Indiana University, the University of Florida, and the National Center for High-performance Computing (NCHC) defined a strategy for "chunking" sensor data so that immutable chunks could be findable (Plale et al. 2018). The strategy was then implemented as a set

[2] galaxyproject.org
[3] a global research partnership for a food secure future dedicated to reducing poverty, enhancing food and nutrition security, and improving natural resources, see www.cgiar.org

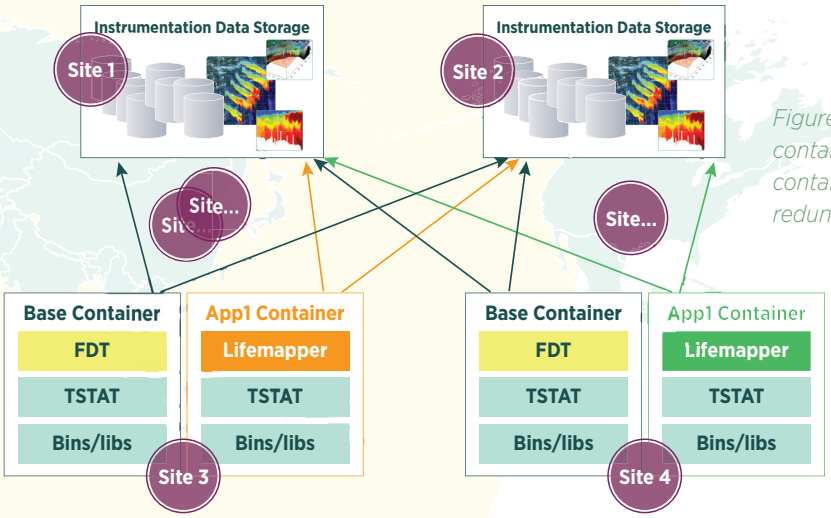


Figure 11. Instrumented PRAGMA Cloud testbed using containers that inherit a common TSTAT logging container. Application and Tstat metrics get sent to redundant centralized instrumentation data storage.

of type definitions that are stored to an instance of the Data Type Registry that is managed by the NSF-funded Robust Persistent Identification of Data (RPID) testbed (see Data subsection below). The PRAGMA AirBox project utilizes the services of the RPID testbed, including requesting test PIDs from the handle service, storing type definitions to the Data Type Registry, and experimenting with PID Kernel Information representations.

The AirBox pilot has been used in a classroom training exercise for student projects that are carried out using cloud resources. Students analyze the AirBox sensor data that is regularly published to the cloud. The data are accessed through Azure VMs that are configured with tools for analyzing the sensor data. The purpose is to demonstrate that online courses can utilize cloud resources in the classroom (Plale and Kouper 2017). The pilot utilizes sensor data from the Airbox Project with permission from NCHC in Hsinchu City, Taiwan.

FUTURE PLANS

Currently, in the PRAGMA Cloud, we are actively testing the use of a Kubernetes virtual cluster to support application containers (see PRAGMA Cloud and Virtualization highlight). We will use performance monitoring containers that will facilitate instrumentation of the PRAGMA testbed (see Figure 11). The instrumented container will contain monitoring tools like Tstat and data transfer tools like FDT. The plan for the data platform consists of infrastructure monitoring, PID assignment, and data storage:

INFRASTRUCTURE MONITORING: Infrastructure monitoring will begin with existing tools, with selection approached from an application's point of view of the infrastructure, an uncommon and novel perspective. Initial work performed during Spring 2018 by a student from Chungnam National University (South Korea) displayed Tstat, TCP STatistical and Analysis Tool[4], and data using the TICK stack (Telegraf, InfluxDB, Chronograf, and Kapacitor)[5]. We also intend to leverage work from the NetSage project[6], whose service collects networking data from active and passive measurements (e.g., perfSONAR[7], Tstat, SNMP), stores the data centrally, and visualizes it with Grafana[8].

PID ASSIGNMENT: PID assignment leverages the NSF-funded RPID testbed (see Data subsection below). The RPID testbed services support assignment and resolution of temporary PIDs (handles). The ability to assign temporary PIDs is critical for testing. RPID supports PID Kernel Information profiles, which are exploratory extensions to handles that we continue to utilize to study connections between data objects generated during both Rice Galaxy and Airbox projects.

DATA STORAGE: There are now multiple open source data storage systems that are suitable for managing time series data. PRAGMA partners at AIST, NCHC, and MIMOS are beginning to develop expertise around the Ceph object store[9], which holds promise as a common distributed data storage platform for the PRAGMA Cloud testbed data platform. Other storage servers are under discussion within PRAGMA. Evaluation criteria for data storage include the

[4] tstat.polito.it
[5] Influx Data Company, www.influxdata.com
[6] www.netsage.global
[7] www.perfsonar.net
[8] Grafana Labs
[9] ceph.com

following: How well do PIDs map to identity/naming strategy used in the data storage system? Does the storage system support multiple unique IDs? How are data objects discoverable through unique IDs? How well do the systems federate across administrative boundaries? How efficient is the system in moving large amounts of data in/out for purposes of analysis? How well do the systems support access control where data are restricted?

PARTICIPANTS

UC San Diego: Nadya Williams, Philip Papadopoulos, Shava Smallen; *National Center for High-performance Computing (NCHC):* Weicheng Huang, Steven Shiau, Ceasar Sun, Hsiu-Mei Chou; *Indiana University, Bloomington:* Beth Plale, Yu Luo; *National Institute of Advanced Industrial Science and Technology (AIST):* Jason Haga, Ryosei Takano, Yoshio Tanaka; *Salk Institute for Biological Studies:* Gerald Pao, Chen-min Yeh, Eric Saberski; *Rice Galaxy Project: International Rice Research Institute (IRRI):* Ramil Mauleon, Venice Juanillas; *Advanced Science and Technology Institute (ASTI):* Jelina Tanya H. Tetangco; *AIST:* Jason Haga; *Indiana University: Airbox Project: Academia Sinica:* Ling-Jyh Chen; *NCHC:* Hsiu-Mei Chou; *Indiana University:* Beth Plale; *TSTAT (Infrastructure Monitoring) Project: Chungnam; National University: Chungnam National University: Min*sung Joh, Minsun Lee

SOFTWARE

Tstat roll: github.com/pragmagrid/tstat

PRAGMA Data Service: github.com/Data-to-Insight-Center/RDA-PRAGMA-Data-Service

REFERENCESS

Juanillas, V. M., Dereeper, A., Beaume, N., Droc, G., Dizon, J., Mendoza, J. R., Perdon, J. P., Mansueto, L., Triplett, L., Lang, J., Zhou, G., Ratharanjan, K., Plale, B., Haga, J., Leach, J. E., Ruiz, M., Thomson, M., Alexandrov, N., Larmande, P., Kretzschmar, T., and Mauleon, R. P. (2018) Rice Galaxy: an open resource for plant science, *bioRxiv* 358754. 2018. doi: doi.org/10.1101/358754.

Luo, Y., Zhou, Q., Ratharanjan, K., and Plale, B. (2018) Persistent IDs: Application to Workflow and Sensor Applications (poster). PRAGMA 34, Tokyo, Japan, May 9, 2018. www.pragma-grid.net/images/pragma34/Luo.pdf

Luo, Y. (2017) Towards FAIR Open Science with PID Kernel Information. Research Data Alliance Plenary 10, Montreal, Canada, Oct 2017. Retrieved from www.rd-alliance.org/sites/default/files/RDA%20Plenary%20Yu%20.pdf

Luo, Y. (2017) Toward FAIR Open Science with PID Kernel Information: the RPID testbed. IEEE Workshop on Big Data Metadata and Management BDMM'17 (invited talk). Boston, MA, USA, Dec 2017.

Plale, B., Kouper, I., et al. (2017) SEADTrain data analysis: Hands-on tutorial. Presentation materials. ESIP Summer 2017 meeting in Bloomington, IN, figshare.com/articles/SEADTrain_Data_Analysis/6873800.

Plale, B. et al. (2018). Persistent IDs for AirBox Air Quality Monitors (poster). Figshare. Presentation PRAGMA 32/CENTRA Mar 2017, doi.org/10.6084/m9.figshare.6875567.v2.

DATA

PRAGMA PID AIRBOX DATA: All devices operational as of May. The data is from 2 channels: i) MAPS channel – data from May 2017–Jan 2018 (9 months); size: ~500MB and ii) Airbox-Copy channel—data from Sept 2017–Jan 2018 (5 months); size: ~850MB. Total ~1.3GB Airbox data. github.com/Data-to-Insight-Center/SEADTrain/tree/master/data and github.com/Data-to-Insight-Center/SEADTrain/wiki/SEADTrain-Data-Description.

The data used in this training exercise is made available through a collaboration of the Pacific Rim Applications and Grid Middleware Assembly (National Science Foundation award #1234983), the Center of Excellence for Cyber Enablement of Applications (MOST 104-2923-E-492-002-MY3), and Academia Sinica (MOST 105-2221-E-001-016-MY3). Special thanks to Edimax Inc. and the LASS community for their support, technical advice, and administrative assistance.

ROBUST PERSISTENT IDENTIFICATION OF DATA (RPID): The NSF-funded (#1659310) robust PID testbed, called the RPID testbed (pronounced "rapid"), stimulates and enables evaluation of powerful new complementary outputs of the Research Data Alliance (RDA) in PID oriented data management. See rpidproject.github.io/rpid. See also Luo 2017.

VISUALIZATION INFRASTRUCTURE AND ACTIVITIES

Improving understanding through visualizing multiple integrated data sources

The use of large format tiled display walls for immersive visualization and analytics continues to be an active area in PRAGMA and builds on strong ties with the Collaborations to Enable Transnational Cyberinfrastructure Applications (CENTRA) partnership and the University of Hawaii, Manoa. The application areas have expanded beyond the disaster management platform developed using SAGE2 (Scalable Amplified Group Environment 2), LAVA (Laboratory for Advanced Visualization Applications) at the University of Hawaii, Manoa, and EVL (Electronic Visualization Laboratory) at the University of Illinois, Chicago. One growing application area is in e-health, in which a visualization workflow to provide decision support for medical research projects in Thailand (MedThaiSAGE) is being developed through a collaboration between the National Institute of Advanced Industrial Science and Technology (AIST) and Mahidol University. The idea is to enable practitioners and policy makers to explore a variety of different data in order to make informed decisions when implementing future policies. With this goal in mind, a prototype platform has been developed (see Figure 12) that links together different types of data in the backend to provide various views of the data on the display wall.

A second application area is developing a visualization tool for traffic monitoring in Taiwan using machine learning (Figure 13). This smart city application, in a collaboration between AIST and Taiwan's National Chiao Tung University (NCTU), would enable urban planners to analyze traffic patterns in drone and webcam videos when making decisions. Thus, these applications serve as working prototypes of systems designed to support decision-making capabilities for different domains in a shared collaborative environment. The future vision is to create an immersive platform that provides a visual "workflow" that can facilitate the understanding of data by decision-makers during both non-critical and critical situations. These applications can leverage the stable PRAGMA com-

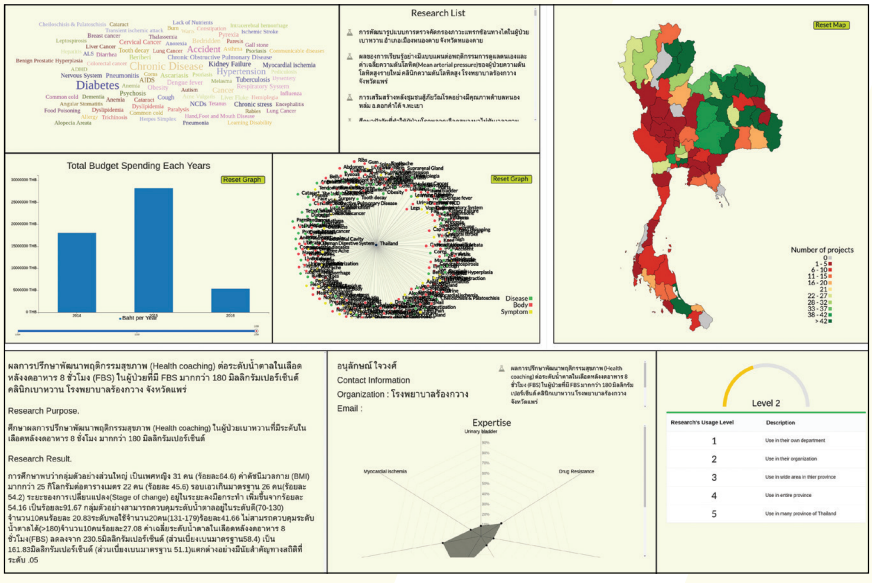


Figure 12. Image of the unified, single visualization interface of MedThaiSAGE. Multiple types of data are visible at once and are linked together to propagate any changes in the data as the user explores it. The application is designed to run natively in SAGE2 workspaces. This work is a collaboration between AIST and Mahidol University, with technical collaboration from the University of Hawaii.

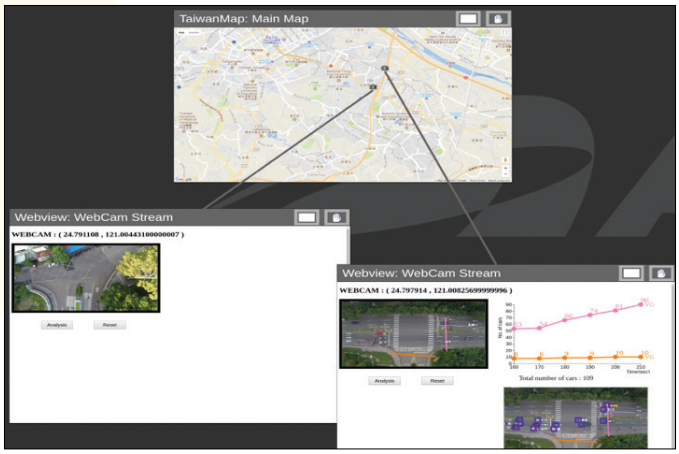


Figure 13. Screen shot of SAGE2 workspace of machine learning application for traffic monitoring. Multiple camera locations can be visualized and analyzed for the number of cars that pass a virtual gate set by the user.

pute infrastructure and are drivers for the Pan-Pacific Visualization Alliance established within CENTRA.

At the PRAGMA 35 conference, a keynote presentation was given in conjunction with the Big Data Summit 2 that focused on visualization applications for a variety of domains. One highlight was illustrating how complex workflows could be facilitated across multiple applications by linking the data passed between applications and incorporating voice commands. This keynote was an important outreach opportunity for extending PRAGMA's goal of developing strong international collaborations and engaging new partners.

PARTICIPANTS

National Institute for Advanced Industrial Science and Technology (AIST): Jason Haga; *National Chiao Tung University (NCTU):* Jimmy Ching-Yu Liu (student), Chih-Wei Yi; *Mahidol University:* Jarernsri Mitrpanont, Jirayu Roungsuriyaviboon (student), Thada Sathapornwatanakul (student), Wudhichart Sawangphol; *University of Hawaii, Manoa:* Dylan Kobayashi (student)

WEB–BASED SCIENCE AND ENGINEERING PLATFORM FOR ADVANCED CYBER–LEARNING

Implementing new workbench and workflow environments in the EDISON platform for computational science and engineering

EDISON, which stands for EDucation-research Integration through Simulation On the Net, is the well-known e-Science platform developed to meet the scientific computing needs of domain scientists. The number of disciplines and users of EDISON have increased annually since its launch in 2011 and currently there are eight academic disciplines and more than ten thousand users of the platform each year. The vigorous activity has been accelerated by the deployment and use of several new e-Science technologies, such as a scientific workbench and workflows. EDISON scientific workbench is a user-friendly browser-based simulation running panel that contains the customized work areas of input, output, visualization, monitoring, job control, and et cetera for developers of apps (which are simulation software programs). This tool enables users to conduct their simulations efficiently and lowers the barriers to users with fewer computational technology skill, as in the case of computational medicine users. Figure 14 shows an example situation in which the workbench was used to provide an advanced lecture environment for a Physiology course where the users (normally medical interns) were able to read documents of theory while simultaneously running simulation apps to visualize and test the theories being learned. The tutor and interns participating in the course reported that using the cyber-learning environments provided them much greater understanding of the subject.

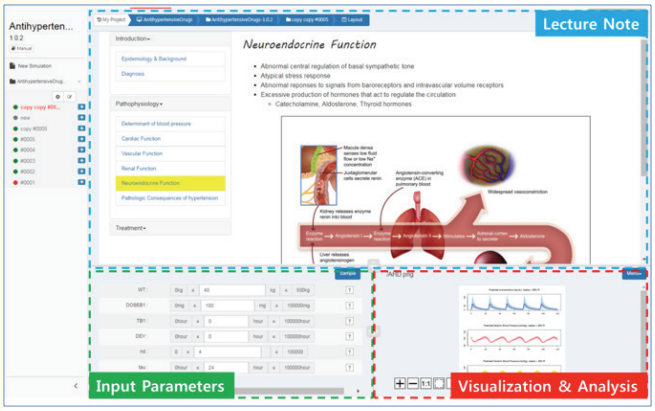


Figure 14. EDISON's Web-based scientific workbench.

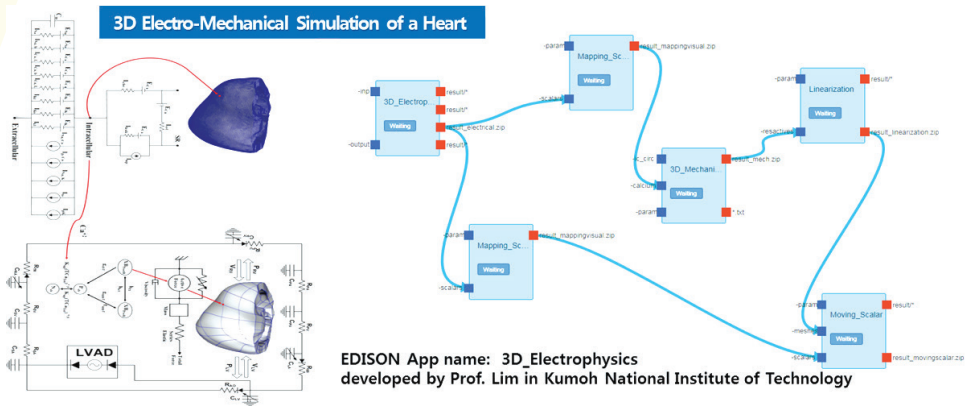


Figure 15. An example of Web-based scientific workflow: Multiphysics simulation for a human heart.

Another important technical advancement in 2017-2018 was the development and application of the scientific workflow tool. The EDISON portal provides a workflow execution environment that can incorporate scientific apps developed by various simulation developers from different disciplines. This means that the workflow tool conducts consecutive executions of multiple apps by applying various functions: serial, parallel, conditional, loop, format converting, and et cetera. This workflow tool can be applied in college courses to integrate learning as well as individual cyber-learning. Figure 15 shows the connection of six EDISON apps for a multi-physical analysis of a human heart as a workflow. After an actual application in a college lecture, it was demonstrated that by reflecting the clarity of the structure of the computation flow, the level of learners' understanding of the given problem increased. In addition to this example, there have been examples of other possible application areas, such as automation of presolving and post processes, optimal engineering design, percolation analysis, and more. Therefore, it is highly anticipated that beginners in engineering can be encouraged to understand their specialties in a bigger picture of the analysis processes.

Figure 16. EDISON was a proud recipient of a 2017 National R&D Top 100 Outstanding Achievement Award.

By the virtue of the above efforts, the EDISON program won a 2017 National Top 100 Outstanding Achievement Award for R&D from Ministry of Science and ICT, Korea (Figure 16). Nowadays, STEAM (Science, Technology, Engineering, Art, and Mathematics) education is increasingly emphasized throughout the world. To be fully successful in STEAM in K-12 and higher education, various useful cyber-learning tools are now being developed and provided. In response to this requirement, the EDISON workbench and workflow tools will be applied to a variety of new science and engineering problems and upgraded with proper technologies as part of future efforts.

PARTICIPANTS

Korea Institute of Science and Technology Information (KISTI): Dr. Ruth Lee (PI), Dr. Junghun Shin (Team Leader), Dr. Jeonghyun Jerry Seo (Team Leader), Sunil Ahn (Team Leader), Namgyu Kim (Team Leader)

SOFTWARE

EDISON portal: www.edison.re.kr/home. A Korean (full version) and English (partial version) are available.

Scientific Workbench: www.edison.re.kr/web/portal/scienceappstore. Log-in with user ID and password are required for running simulation apps. Please contact Ruth Lee (jsruthlee@kisti.re.kr) if interested in using the scientific workbench or scientific workflow applications.

Scientific Workflow: www.edison.re.kr/web/portal/workflow-workbench. Log-in with user ID and password are required for running simulation apps.

REFERENCES

Kim, N., On N. R., Lee, J. S., Nam, D., Cho, K., Choe, K. T., and Yu, J. (2015). An empirical analysis of the user acceptance of computer simulations utilized for higher education in science and engineering, *ICIC Express Letters, Part B: Applications,* 6 (4), pp. 1197–1203.

Ma, J., Lee J. S., 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) pp. 1089–1104.

Park, S. R., Lee, J. S., and Cho, K. (2016). Effectiveness of online simulation-based e-learning for computational science and engineering in higher education. *Advanced Science Letters,* 22(9) pp. 2257–2261.

Ryu, H., Yu, J. L., Jin, D., Lee, J. H., Nam, D., Lee, J. S., Cho, K., Byun, H. J., and Byeon, O. (2014). EDISON science gateway: A cyber-environment for domain-neutral scientific computing, *IEICE Transactions on Information and Systems,* E97-D(8), pp.1953–1964.

EDUCATION

STUDENTS IN PRAGMA

Creating globally minded problem solvers for 21st century challenges

PRAGMA is committed to engaging new communities of researchers and building a broader scientific 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 our experiments and encouraging development between workshops.

PRAGMA's efforts to engage students are aimed at creating globally minded problem solvers for 21st century challenges. In particular, we want to inspire students to become research leaders who can function well in a collaborative, multidisciplinary, and multi-cultural world. Our approach focuses on two strategies:

- **Enhancing professional experiences for and development of students at PRAGMA workshops** and
- **Providing students with research experiences throughout the PRAGMA network.**

To these ends, we established the PRAGMA Student organization in 2012 and work with PRAGMA members to

encourage student research mobility experiences through programs like the Queensland Research Projects Abroad (QURPA) and Monash Undergraduate Research Projects Abroad (MUPRA) programs, Internships at AIST, and project-based research exchanges at member sites (for more examples, see PRAGMA Students and PRAGMA Data Infrastructure highlights). This section describes these activities.

PRAGMA STUDENTS

Building a future generation of international researchers

PRAGMA 35 student workshop participants and mentors, Penang, Malaysia October 3, 2018

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.

To date, activities of PRAGMA Students have included organizing PRAGMA-affiliated student workshops and poster sessions as part of the biannual PRAGMA Workshops and developing a unique model to provide multiple opportunities for students to participate in PRAGMA's collaborative scientific research. PRAGMA provides a trusted network of people and opportunities in leadership that help students gain valuable professional experience.

There are several advantages to joining PRAGMA Students and the PRAGMA community. These opportunities include, but are not limited to, the following activities:

1. Participating in workshops and conferences in an international setting, which stimulates inspiration, information sharing, and collaboration;
2. Working with PRAGMA mentors and advisors on challenging, state-of-the-art research projects, which strengthens research skills and ability for scientific exploration;
3. Visiting other PRAGMA sites as part of short-term residential research opportunities, which provide both scientific and cultural experiences.

For the May 2018 PRAGMA 34 workshop in Tokyo, Japan, PRAGMA Students organized the poster session. Members of the PRAGMA Students committee reviewed and selected posters from all workshop participants. Around 21 posters were carefully evaluated, and detailed review feedback was provided to the submitters. Workshop participants were asked to put stickers on their favorite posters, and an award was presented to the top four posters that received the most stickers from participants during the poster session. The poster winners came from Taiwan, the United States, Thailand, and Japan institutes. This activity also took place at PRAGMA 35. (See Posters subsections for lists of Best Posters).

In addition, PRAGMA Students organized a lightning talk session for all PRAGMA participants that gave poster authors an opportunity to briefly talk about their research in one-minute presentations. The session allowed participants to get feedback on their work while gaining valuable presentation experience. Eighteen authors gave brief talks about their respective research projects.

During the PRAGMA 35 workshop in Penang, Malaysia, PRAGMA Students held a new session, Student Presentation Session, which gave an opportunity for students to do a 15-minute presentation and receive feedback from PRAGMA mentors. Nine students participated in this session. Additionally, prior to PRAGMA 35, on September 26, 2018, we held a webinar, led by Jason Haga (AIST), to help students prepare for academic presentations.

For the PRAGMA 35 poster and lightning talk sessions, 25 posters were accepted and 23 poster authors gave 1-min lightning talks to introduce their posters. Best student poster recipients were from Malaysia and Indonesia institutes.

Outside of the workshop venue, some students in PRAGMA partook in short-term residential research opportunities. Wassapon Watanakeesuntorn, a master student from the Nara Institute of Science and Technology (NAIST), was a visiting scholar from August to September 2018 at The National Institute of Advanced Industrial Science and Technology (AIST) in Tsukuba, Japan where he worked on a project called "Novel Neuromorphic Deep Learning Architectures from Empirical Dynamical Modeling of the Zebrafish Brain." As part of this project, PRAGMA aims to optimize the code and use the AI Bridging Cloud Infrastructure (ABCI) data center to run the deep learning algorithm on this project. This internship project is a collaboration project with UC San Diego. (See also AI Bridging Cloud Infrastructure highlight.)

We would like to note that five members of the PRAGMA Student Steering Committee, Pongsakorn U-chupala (NAIST), Quan Zhou (Indiana University), Chawanat Nakasan (NAIST), Che Huang (NAIST), and Giljae Lee (U Florida) all received their PhDs during this year. Congratulations!

PRAGMA STUDENT STEERING COMMITTEE

Wassapon Watanakeesuntorn *(NAIST),* Can Wu *(Chinese Academy of Sciences),* Chiao-Ning Chuang *(National Chiao Tung University)*

STUDENT MENTORS (to October 2018)

Beth Plale *(Indiana University),* Karpjoo Jiang *(Konkuk University, Korea)* (through May 2018), Putchong Uthayopas *(Kasetsart University). Note: Jason Haga has assumed responsibility for the Mentoring Functional Area Role on the Steering Committee.*

PRAGMA 34 BEST POSTERS

1st: Chiao-Ning Chuang, Chien-Heng Wu, Wen-Yi Chang, and Whey-Fone Tsai. *National Chiao Tung University, Taiwan.* **Implementation of Deep Learning Algorithm on Personal Big Data Platform for Engineering Applications**

2nd: Yu Luo, Quan Zhou, Kunalan Ratharanjan, Beth Plale, Ramil Mauleon, Jason Haga, and Hsiu-Mei Chou. *Indiana University Bloomington, USA.* **Persistent IDs: Application to Workflow and Sensor Applications**

3rd: Suchanat Mangkhangjaroen. *Thammasat University, Thailand.* **ONE: Online Note Extraction to Music Cheat**

Honorable Mention: Juan Sebastian Aguirre Zarraonandia. *Osaka University, Japan.* **Application Aware Traffic Engineering Functionality for an SDN Transit Network**

PRAGMA 35 BEST STUDENT PRESENTATIONS

1st: Chiao-Ning Chuang. *National Chiao Tung University, Taiwan.* **Image Restoration for CCTV Images for Natural Disaster Management**

2nd: Muhammad Jaziem bin Mohamed Javeed. *Universiti Sains Malaysia, Malaysia.* **LINGO Profiles Fingerprint (LPFP) and Association Rule Mining (ARM) for Predicting Drug Target Interactions (DTIs)**

3rd: Nuraisah. *Universitas YARSI, Indonesia.* **Tuberculosis (TB) Disease Interactive Map in Jakarta Capital Special Region**

PRAGMA 35 BEST POSTERS

1st: Saravanan Sagadevan, Nurul Hashimah Ahamed Hassain Malim, Nurul Izzati Ridzuwan, and Muhammad Baqir Hakim Mohammad Bashir. *Universiti Sains Malaysia, Malaysia.* **Criminality Linguistics Detection on Social Networks Through Personality Traits**

2nd: Muhammad Reza Aditya, Ressky-tha Permata Sari, Adri Nursimarsiyan, and Nova Eka Diana. *Universitas YARSI, Indonesia.* **Mobile-based Augmented Reality for Sundanese Alphabets Education**

3rd: Andi Batari Ahmad and Nova Eka Diana. *Universitas YARSI, Indonesia.* **Deep Learning Classification for Liver Disease**

Honorable Mention: Muhammad Jaziem Bin Mohamed Javeed, Aini Atirah Binti Rozali, Siti Zuraidah Binti Mohamad Zobir, Habibah Binti Abdul Wahab, and Nurul Hashimah Ahamed Hassain Malim. *Universiti Sains Malaysia, Malaysia.* **Curating Target-Activity Information for NADI Compounds Based on CHEMBL Using Similarity Searching and Pattern Matching**

MURPA/QURPA

Providing 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, 53 students have participated in the programs and have travelled to UC San Diego, the National Center for Supercomputing Applications (NCSA) in Illinois, The Technion in Israel, the Institute for Infocomm Research (I2R) in Singapore, the University of Warwick in Coventry, England, 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, four Monash students travelled to UC San Diego, the University of Auckland in New Zealand, the Max Plank Institute in Germany, and AIST in Japan.

Monash University Computer Science student Michael Franklin travelled to the Institute for Advanced Industrial Science and Technology, in Tsukuba, Japan. He worked with Dr. Jason Haga on immersive visualization for big data for river disaster management. Michael's project aimed to expand an existing river disaster management virtual reality (VR) application, providing the ability to read live data and perform data analysis to identify regions of interest.

Computer Science Honors student Jason Xu travelled to the Max Planck Institute for Evolutionary Biology, in Germany. He worked with Professor Arne Traulsen and Dr.

Laura Hindersin studying evolutionary processes on networks. He showed, using large-scale simulations, that calculations in the limit of weak selection break down when populations are embedded in a network structure. Jason's work was instrumental in formulating some of the research questions he's now tackling with Dr. Julian Garcia, as part of his honors thesis.

Alexander Zenin from Monash travelled to UC San Diego where he worked with Winston Armstrong at the San Diego Supercomputer Center. Alex is interested in cybersecurity, and he worked on a project identifying security vulnerabilities in processes for storing, accessing, and processing sensitive medical data. He devised a workflow for penetration tests assembling and using different tools to assess vulnerability.

Engineering and computer science student Laura Cummack travelled to the University of Auckland in New Zealand. She worked on the ICON bionic joint project, under the supervision of Dr. Desney Greybe. The team Laura worked on was tasked with developing joint sensing exoskeletons to support the motion of patients who need movement and strength assistance. Laura worked in the part of the project that models specific features of individual patients. She produced a tool for model generation using Python, based on legacy code. The process calibrates the model to data specific to the user.

Both the MURPA and QURPA programs include an advanced seminar scheme in which students can attend

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

virtual seminars given by world renowned experts. These seminars allow students to "meet" potential mentors and get information about potential projects. In 2018, the seminar series focused on presenters from the National Center for Supercomputing Applications (NCSA) at the University of Illinois in Urbana-Champaign, Illinois.

As done in the past, seminars were broadcast simultaneously to Monash University (in Melbourne) and the University of Queensland (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.

AIST INTERNSHIP PROGRAM

Providing information technology research opportunities and experiences with cultural context

With the increase in globalization of science and the ability of information technologies to cross international boundaries, it is critical to provide information technology research opportunities and experiences with cultural context. To address this challenge, the National Institute of Advanced Industrial Science and Technology (AIST) is leveraging its past extensive experience with the Pacific Rim Experiences for Undergraduates (PRIME) program (see References) to create an international hub for computer science research through visiting student internships designed to broaden, strengthen, and maintain AIST's presence on the global research stage. Started in 2014, AIST has expanded its internship activity each year. In 2017, we hosted 14 internship students from France, Taiwan, USA, Thailand, and Australia. There was a total of seven publications from the internships, including two Best Paper awards at the 2017 International Conference on Information Technology (InCIT 2017) in Bangkok and the 2018 Computer Games Multimedia and Allied Technologies (CGAT) conference in Singapore, respectively. For 2018, we have continued our internships with 11 students. Our student research has covered a wide-range of topics, including immersive visualization, networking for the Internet of Things (IoT) era, artificial intelligence, gamification, and cyberinfrastructure.

Importantly, we are strategically aligning these student activities to both facilitate international collaborations and support interactions under different formal agreements, such as memorandums of understanding (MoUs). For example, a recent MoU between AIST and Mahidol University, combined with more than 24 student internships, helped establish a new satellite laboratory, Mahidol AIST Research Unit (MARU), on the Mahidol University campus.

Earlier in 2018, AIST completed a new five-year MoU with UC San Diego to develop collaborations using AIST's AI Bridging Cloud Infrastructure (ABCI) as the key infrastructure for researcher and student exchanges. One student from the Nara Institute of Science and Technology (NAIST) worked on building containers and deploying them on ABCI for our zebrafish grand challenge project (see PRAG-

AIST Internship students on final presentation day, 2017.

MA Data Infrastructure and AI Bridging Cloud Infrastructure highlights). We are tapping into the PRAGMA and CENTRA communities to find partners that are interested in helping to advance research in this area through our research opportunities. For example, we recently hosted an intern from the National Chiao Tong University, supported by the Center of Excellence Cyber-Enablement of Applications (CECEA), who studied ways to improve real-world webcam images using a machine learning approach. The student, Chiao-Ning Chuang, won a best poster award at PRAGMA 34 and a best student presentation award at PRAGMA 35. We believe these activities are a critical step toward the goal of creating strong international collaborations. Moreover, these activities demonstrate AIST's long-term commitment to enhancing global IT research and supporting PRAGMA's strategy of providing students with research experiences throughout the PRAGMA network in order to create globally minded problem solvers for 21st century challenges.

PARTICIPANTS

AIST: Jason Haga, Ryousei Takano

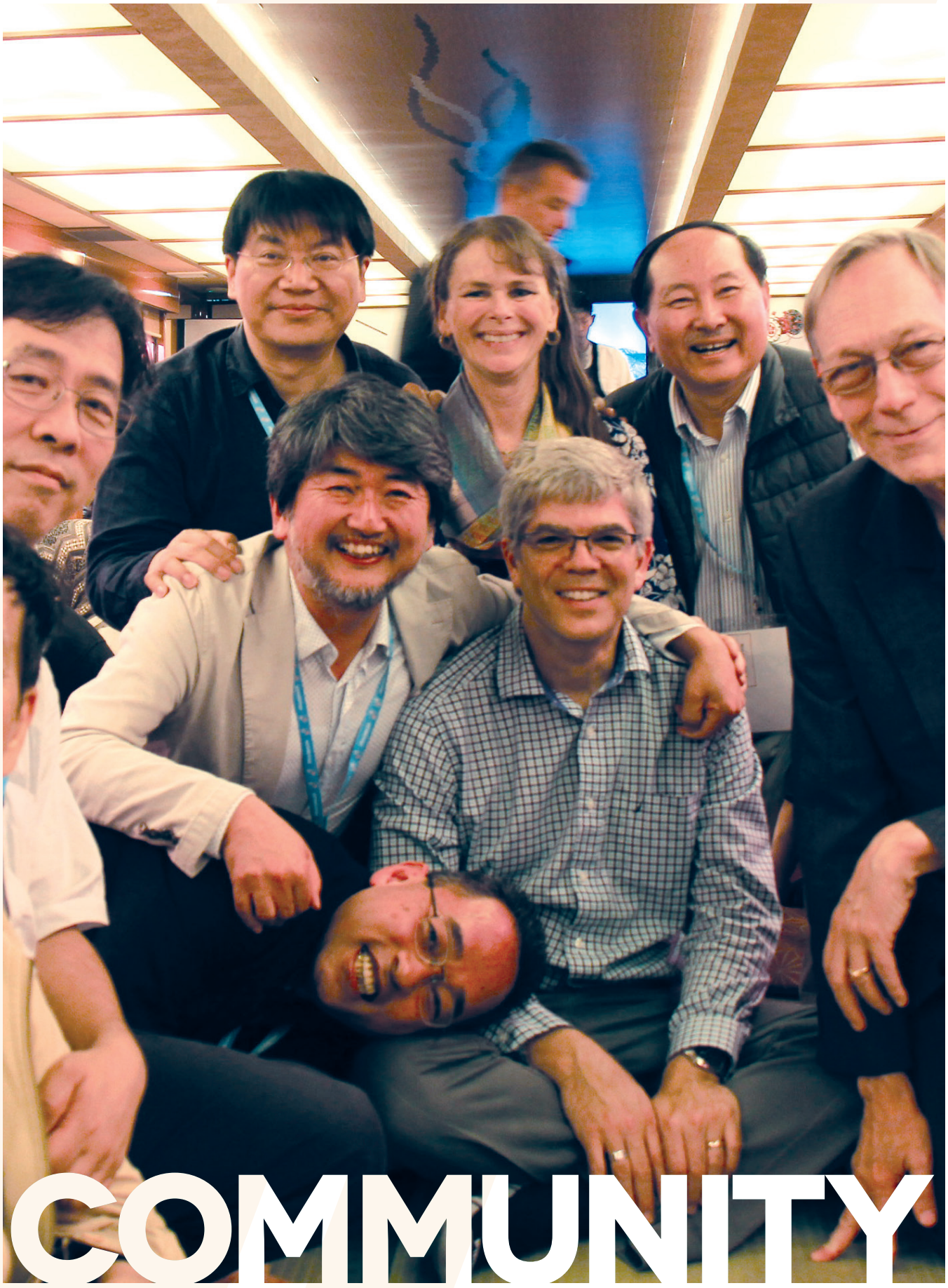
REFERENCES

Haga J., Galvin J., Arzberger P., Wienhausen G. PRIME: A Model of Experiential Learning within a Global Research Community. In Education in Action: Experiential Learning in Higher Education, editors Forbes D., Lin J., and Losh E. UCSD Sixth College. pp 297-306. June 2012.

Arzberger, P., Wienhausen, G., Abramson, D., Galvin, J., Date, S., Lin, F-P., Nan, K., Shimojo, S. PRIME: An Integrated and Sustainable Undergraduate International Research Program. *Advances in Engineering Education.* 2010, 2(2).

PRIME Website: prime.ucsd.edu

A GULL DRIFTS
LOW-ALTITUDE
ROCKET SHAG
GATE FOR LUNCH
TURTLES ACROSS
THIS IS THE LIFE THAT GO
ODLESSLY
POEM WITHOUT WO
BEET WITHOUT CONDIT
O.K. S



COMMUNITY

BUILDING COMMUNITY: INFUSING IDEAS AND BROADENING IMPACT

PRAGMA actively looks for opportunities to build and broaden 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 hosting face-to-face meetings to engage new participants, we disseminate our practices, experiences, and results via publications, presentations, and posters. Over the last six years, PRAGMA collectively published approximately 100 articles, more than 40 of which involved students. In 2017, 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 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 Software). Over the last five years, PRAGMA members have created, improved, or added components to 26 published software products and services: seven to enable the PRAGMA Multicloud (PRAGMA Cloud Scheduler, Cziso, DynIP, PRAGMA Boot, PRAGMA Boot Google Drive Virtual Cluster, Vc-out-parser, Clonzilla, Tensorflow Rocks Roll, Tstat Rocks Roll, Kubernetes Rocks Roll, Gpupt, and Virtual Cluster Image repository), 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.

SEAIP 2017 Meeting, National Chi Nan University, Nantou Taiwan

REFERENCES

Plale B. and Chen M., Eds. Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) special issue on international clouds for data science, Concurrency and Computation: Practice and Experience, John Wiley and Sons, Vol. 29, Issue 13, 2017. DOI: 10.1002/cpe.v29.13/issuetoc (onlinelibrary.wiley.com/doi/10.1002/cpe.v29.13/issuetoc).

SOFTWARE

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

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

ENGAGING NEW SOUTHEAST ASIA MEMBERS

Southeast Asia International Research and Training Program (SEAIP) and PRAGMA Institute: Taichung and Nantou, December 4–8, 2017

The SEAIP's ongoing series of workshops (see seaip.narlabs.org.tw) are organized by the National Center for High-performance Computing (NCHC) and are supported by the National Applied Research Laboratories (NARlabs) and the Ministry of Science and Technology in Taiwan. 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.

The theme of the 13th SEAIP workshop, held December 4–8, 2017, was "Cloud Computing, Internet of Things, and AI for Discovery and Innovation." Organized by NCHC with funding from the Ministry of Science and Technology, as well as from NARlabs, the atmosphere of the workshop was informal and conducive to learning and developing collaborations. The first sessions were held at the National Center for High-performance Computing's Taichung Office in Taichung, Taiwan. A second session was held on the campus of National Chi Nan University in Nantou, Taiwan. Part of this year's SEAIP meeting focused on community, technical, analytic, and application challenges in a time of rapid changes in information and communications technology research. Each year, the SEAIP workshop brings together researchers from Europe, the United States, Northeast Asia, and Southeast Asia. This year, the meeting was held in conjunction with the ICT Virtual Organization of ASEAN Institutes and NICT (ASEAN IVO) meeting. The SEAIP venue provides a forum for PRAGMA researchers to share their experience and to meet researchers from Southeast Asia that might not be able to attend PRAGMA workshops, thus broadening PRAGMA's impact and involvement with others.

In 2018, this series of meetings was renamed the International Joint-Research and Training Program to more accurately reflect its global perspective. The next meeting will be held in Taiwan, November 26–30, 2018.

CHAIR OF SEAIP AND THE INTERNATIONAL JOINT-RESEARCH AND TRAINING PROGRAM: Fang-Pang Lin, *National Center for High-performance Computing.*

EXPLORING NEW RESEARCH TOPICS AND BROADENING ENGAGEMENT

PRAGMA explores new areas of research and is always looking for new communities of users who will benefit from advancing cyberinfrastructure technologies. As examples, we highlight several efforts either associated with PRAGMA workshops or from separate workshops that PRAGMA members attended in order to start a dialog with new communities.

Understanding the New AI Bridging Cloud Infrastructure

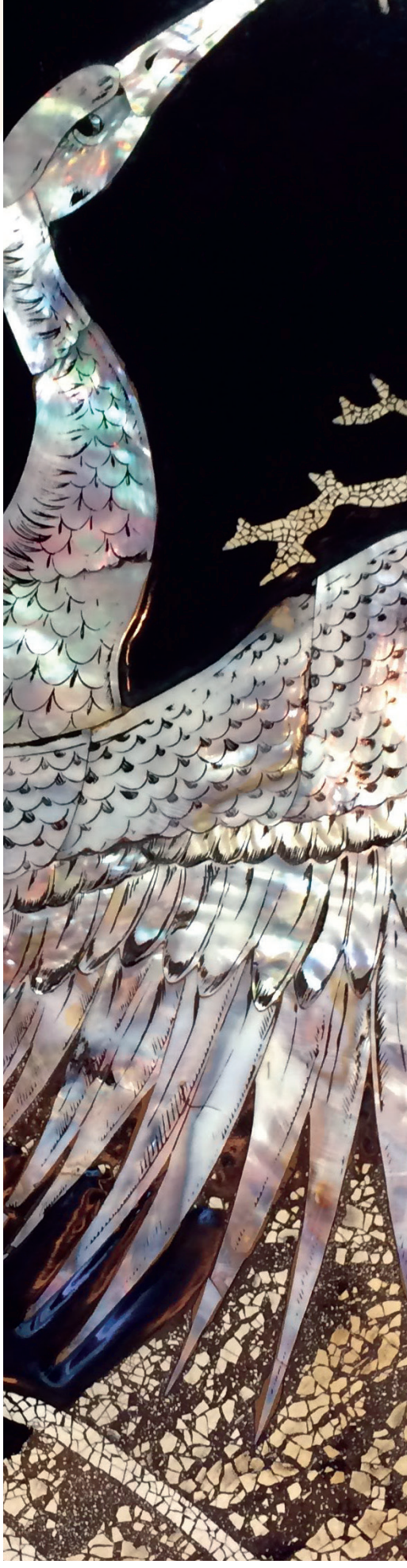
The Workshop on High Performance Infrastructure for AI, Tokyo, May 9, was held in conjunction with PRAGMA 34. This meeting provided attendees with an overview and update of the National Institute of Advanced Industrial Science and Technology's (AIST) investment in and development of the state-of-the-art AI Bridging Cloud Infrastructure (ABCI). The ABCI platform, housed on the University of Tokyo's campus, is aimed at pushing the research envelope of artificial intelligence, machine learning, and deep learning applications. In addition to the overview of the architecture and engineering of the deployment, attendees also heard about pilot applications and a related activity by the National Center for High-performance Computing in Taiwan. At the end of the presentation, participants were given a tour of the new ultra-green datacenter and computer facility. See the AI Bridging Cloud Infrastructure highlight for more information.

Empowering Malaysian and Pacific Rim Researchers with Current Trends in Data Analytics

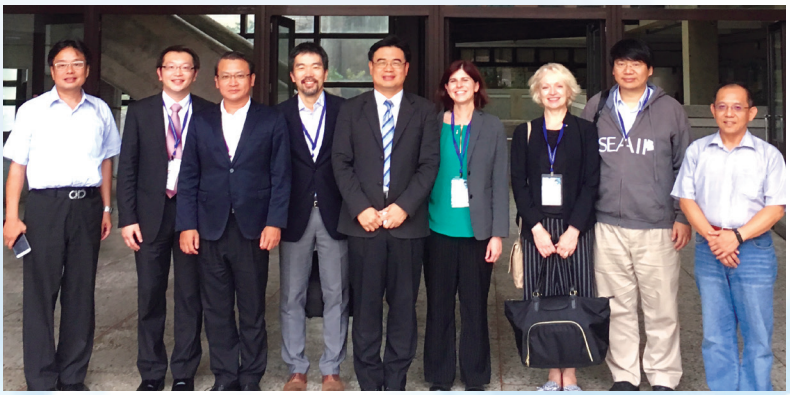
As a part of Big Data Week Asia, the 2018 Big Data Summit 2 was held in conjunction with the PRAGMA 35 conference in Kuala Lumpur and Penang, October 2–6. The objective of the workshop was to empower researchers with current trends in data analytics, focusing on the application of high-performance computing and artificial intelligence for data analytics, to enable sharing advances both among Malaysian researchers and also between Malaysian and international researchers. PRAGMA elected to hold its 35th workshop in conjunction with Big Data Week Asia to foster greater collaborations in this emerging area, and it was a great success. At the PRAGMA 35 meeting, PRAGMA members discussed advances in the newly created AI Bridging Cloud Infrastructure, advances in visualization, and developments in the Pacific Research Platform. In addition, Malaysian attendees learned more about PRAGMA and its activities and were invited to participate as collaborators. Several participants stayed for the PRAGMA 35 workshop and engaged in PRAGMA's working groups. Several projects were identified to pursue, including use of visualization for understanding neurological signals in the brain and the application of AI to smart cities data. These participants want to become part of the PRAGMA.

Exploring Opportunities with an International Network of Bays

The 2018 World Congress of the Most Beautiful Bays in the World Club was held in Penghu Bay, Taiwan during September and October 2018 (see world-bays.com). One of the efforts of PRAGMA's Telescience Working Group has been to explore the application of distributed technologies in smart cities and in the environment. Interactions between Telescience Working Group Chair Fang-Pang Lin and Con-



gress organizers led to several members of the Most Beautiful Bays in the World Club attending PRAGMA 34 in Tokyo in May. At the PRAGMA conference, representatives of the Most Beautiful Bays Club explained the goals of their organization and invited PRAGMA members to participate in the XIV World Bays Congress held on Penghu Bay in Taiwan, which had an emphasis on environmental health and sustainability, climate change, and renewable energy. Several PRAGMA members accepted the invitation and attended as guests of the Penghu County Government, giving a presentation that provided an overview on PRAGMA and continued the dialog to understand the needs of the members of the Most Beautiful Bays Club. PRAGMA members along with the Penghu County Government representative also met with several members from National Penghu University of Science and Technology, including the University President, Dean of Student Affairs, and Professors, to discuss project ideas and student involvement related to "blue economy" environment and tourism projects in Penghu as starting points for collaboration.



Most Beautiful Bays meeting.

Assisting Mekong River Basin Researchers Address Technical Barriers to Modeling the Environment and Sharing Data

The 2018 Mekong Research Symposium was held in Ho Chi Minh City, Vietnam on September 6–7, 2018 (see www.lowermekong.org/events/lower-mekong-research-symposium). As a follow-up to a co-located event at PRAGMA 31 (Bangkok, September 7, 2016), researchers at the University of Virginia helped organize the symposium, which engaged both PRAGMA and CENTRA members to attend.

As noted in the invitation to the event (see website previously mentioned), "the Mekong, one of the world's most significant rivers, is central to its inhabitants' livelihood, imperative to an emerging world economy, and a critically important paragon of the challenges in achieving global environmental sustainability. The Lower Mekong Basin (LMB) is placed in the Indo-Burma hotspot by biologists; an area of great interest to scholars due to an extraordinary diversity of plant and animal life. Today, rapid infrastructure development and land use conversions coupled with climate change threats such as sea level rise and extreme weather conditions are resulting in accelerated changes to the basin's physical and ecological processes."

At the end of the workshop, participants had agreed that better data access will allow more researchers to participate and contribute to sustainable development in the region, starting with creating an inventory of existing Mekong data and assessing infrastructure resources available in the region. The infrastructure for data sharing is a starting point for future collaboration among regional scientists and their international colleagues in PRAGMA and CENTRA.

CENTRA–PRAGMA COLLABORATIONS

Diversifying opportunities for the CENTRA and PRAGMA communities

CENTRA: YEAR 3

The Collaborations to Enable Transnational Applications (CENTRA) aims to advance and catalyze international collaborations and enable them to be long lived through support of travel, exchange visits, and collaboration mechanisms. CENTRA grew out of interactions among PRAGMA members, with CENTRA's founding members being the Advanced Computing and Information Systems (ACIS) laboratory of the University of Florida with support from the National Science Foundation, 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 of Taiwan, and the ASEAN International Virtual Organization (IVO) funded by the National Institute of Information and Communications Technology (NICT) of Japan.

This year, CENTRA turned three years old. It held its third annual meeting (CENTRA 3) in Tokyo, Japan on May 14–16, 2018, hosted by NICT. It was timed to follow the PRAGMA 34 meeting and leverage the continuing interactions between CENTRA researchers and PRAGMA investigators. A total of 86 participants from 35 institutions in 12 countries attended CENTRA 3.

At the CENTRA 3 meeting, several ongoing collaborative projects reviewed their progress since CENTRA 2, including "Dynamically Aggregating Smart Community Sensors, Edge, and Cloud Resources with Overlay VPNs," led by Renato Figueiredo of the University of Florida; "SDN-IP Peering for IoT Data Transmission," led by Te-Lung Liu of NCHC; "IT for Natural Disaster Management & Visualization Alliance," led by Jason Haga of AIST; "ASEAN IVO: Software Defined System on Disaster Mitigation and Smart Cities," led by Hong-Hoe Ong, with Luke Jing Yuan of MIMOS; "AirBox: a Participatory Ecosystem for PM2.5 Monitoring," led by Ling-Jyh Chen of Academia Sinica in Taipei ; and "Distributed Lifemapper," led by Aimee Stewart of the University of Kansas. In addition, new projects

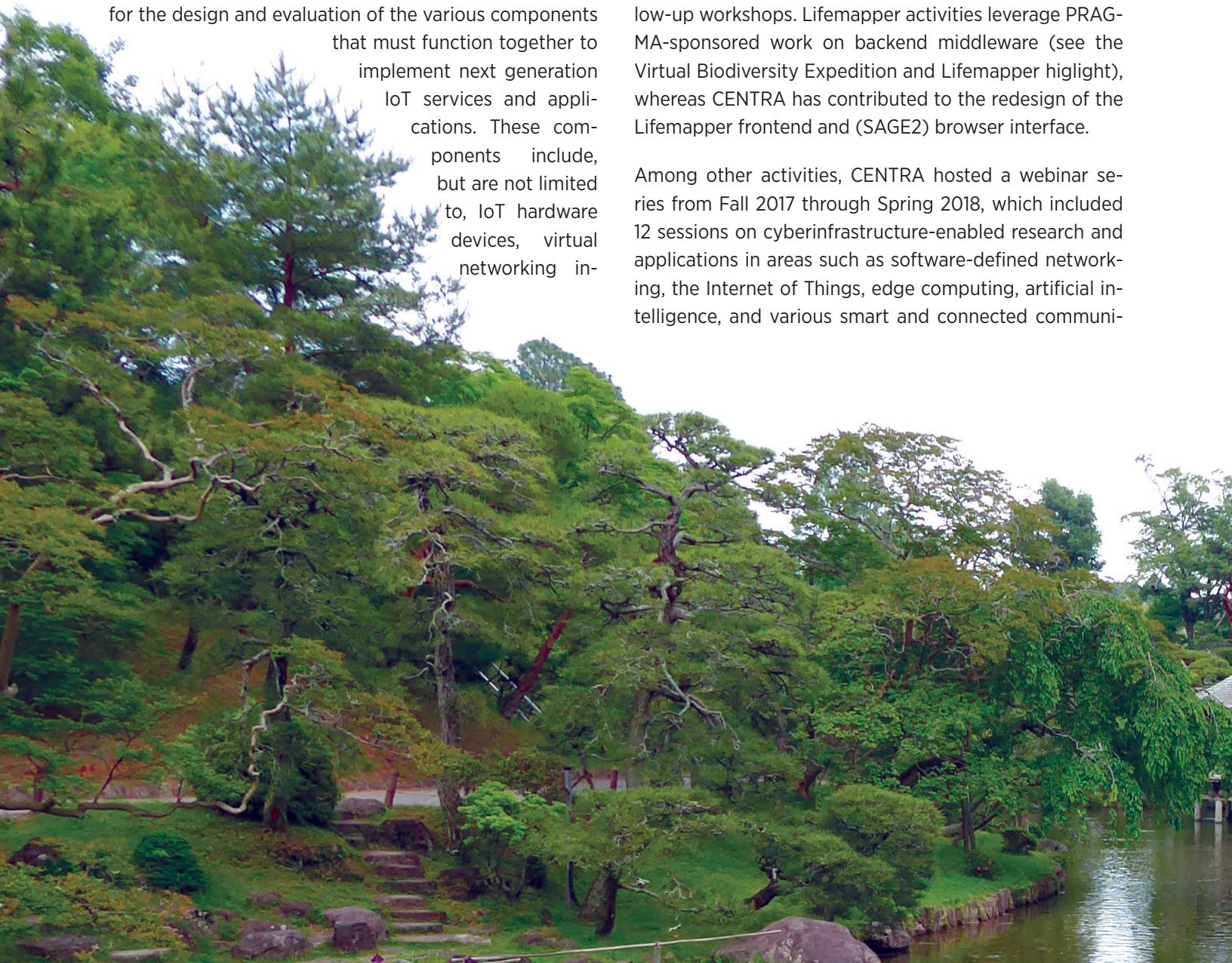
CENTRA 3 All-hands group photo, Tokyo, May 2018

were proposed on "Privacy-preserving Cloud Computing for IoT," led by Joao Paulo of INESC TEC; "Network Congestion Issues in Disaster Scenarios," led by Krishna Kant of Temple University in Philadelphia; "PADE: Precision Agriculture with Drones and Edge," led by Christopher Stewart of Ohio State University; "Using UAV Images to Monitor Rice Paddies with Artificial Intelligence," led by Hui-Ping Tsai of the National Chung Hsing University in Taichung, and "Secure Data Collaboration Using GFFS," led by Andrew Grimshaw of the University of Virginia.

Examples of exchange activities enabled by CENTRA include a visit by U.S. PhD student Kensworth Subratie (University of Florida) to CENTRA member site NICT in Japan on March 4–16, 2018. The purpose of the short stay was to set up the necessary networking infrastructure for an international Internet of Things (IoT) testbed between NICT and ACIS. The testbed is to be used as a collaborative environment among multiple CENTRA research groups for the design and evaluation of the various components that must function together to implement next generation IoT services and applications. These components include, but are not limited to, IoT hardware devices, virtual networking infrastructure and application packages, as well as their orchestration and management. Parts of this work benefit from earlier research led by Renato Figueiredo (Ken's adviser) under PRAGMA's auspices on trusted networking.

Another example is the visit by U.S. researchers Aimee Stewart and C.J. Grady, both from the University of Kansas, to Taiwan, May 2–7, 2018 to lead CENTRA's SAGE-Lifemapper Workshops and Marine Science Collaboration Meetings. One workshop was hosted at NCHC's Hsinchu headquarters and included using Lifemapper on a SAGE2-enabled visualization wall to test its capabilities and future implementation at the CENTRA member site at the University of Kansas (KU). The other was presented at the National Museum of Marine Science and Technology (NMMST). Both were very successful, well attended, hands-on workshops. Results include plans for Lifemapper deployments at NMMST and participating universities and plans by NCHC and KU for additional follow-up workshops. Lifemapper activities leverage PRAGMA-sponsored work on backend middleware (see the Virtual Biodiversity Expedition and Lifemapper higlight), whereas CENTRA has contributed to the redesign of the Lifemapper frontend and (SAGE2) browser interface.

Among other activities, CENTRA hosted a webinar series from Fall 2017 through Spring 2018, which included 12 sessions on cyberinfrastructure-enabled research and applications in areas such as software-defined networking, the Internet of Things, edge computing, artificial intelligence, and various smart and connected communi-



ties applications. The series was attended by a total of 117 unique participants from 14 countries, with 55% from Asia, 34% from North America and 11% from Europe and Oceania. The webinar series engaged a high percentage of early-career researchers (38%), including students (34%) and post-docs (4%). All webinar video recordings are documented on the CENTRA website for on-demand viewing. The series speakers included two members of PRAGMA, and many of the webinar attendees are involved in PRAGMA activities.

PRAGMA 34, held in conjunction with CENTRA 3, Tokyo, May 2018

CENTRA continues to follow its intentional growth path. Last year, the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC) became a member, and after the CENTRA 2 meeting, the Korea Institute for Science and Technology Information (KISTI) became CENTRA's fifth member.

PARTICIPANTS

Leaders of Global CENTRA: US-CENTRA at *U Florida:* Jose Fortes; *ASEAN-IVO at NICT:* Shinji Shimojo; *CECEA at NCHC:* Fang-Pang Lin; *INESC TEC:* Rui Oliveira; *KISTI:* Kum Won Cho

WEBSITES

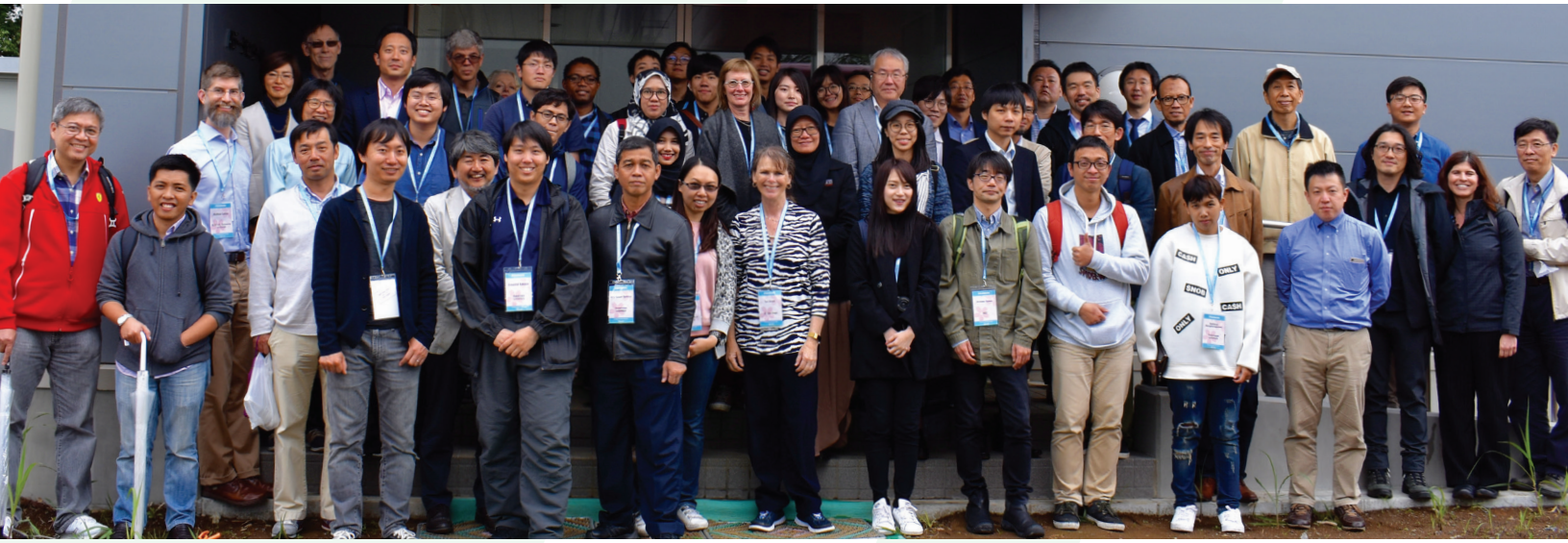
CENTRA: www.globalcentra.org. CENTRA Webinars are documented at vimeo.com/globalcentra.

The Collaborations to Enable Transnational Cyberinfrastructure Applications (CENTRA) is supported by NSF ACI award #1550126 and aims to advance the scientific understanding and applications of distributed software-defined cyberinfrastructure.

WORKSHOPS
WORKING GROUPS

WORKSHOPS AND WORKING GROUPS

Reviewing, planning, and organizing activities and introducing new members

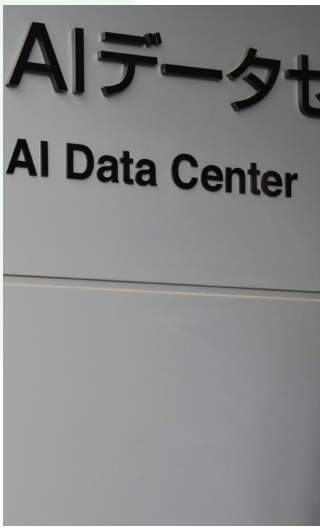


PRAGMA workshops are meetings of all members of the PRAGMA community. They are the major vehicle for information exchange between and among working groups, expeditions, researchers, and institutions; they also provide excellent opportunities to engage new researchers and students at the host sites. New 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 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.

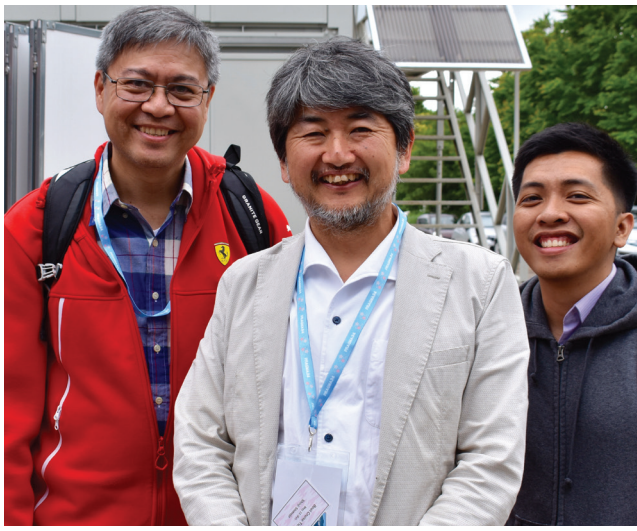
The workshops are organized according to the activities of the following four working groups in PRAGMA:

- **RESOURCES WORKING GROUP:** Working to make the distributed resources of PRAGMA useful to diverse applications via development and use of tools and technologies like PRAGMA Cloud Testbed, PRAGMA ENT, Open Data Platform and others. Co-leaders: Nadya Williams (SDSC/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 WORKING GROUP:** Focusing on the use of technologies to improve understanding in several areas of computational science through the use and improvement of EDISON. Co-leaders: Ruth Lee (KISTI) and Hsi-ching Lin (NCHC).



HIGHLIGHTED WORKING GROUP 2018: TELESCIENCE

The Telescience working group focuses on technologies that are related to smart and connected communities, which include collaborative and advanced visualization, collaboration on SDN clouds, and the Internet of Things (IoT). The application areas focus on both natural to urban environments as well as on newly developed robotics. Traffic and crowd monitoring and behavior reasoning is just one example of an application area that was recently investigated in the group. In that study, traffic information was translated to define the contribution of car exhaust to air particulate PM2.5 and was added to the contribution of power plants, which was then modeled by down-scale meso weather forecasting. Collaborative visualization based on the SAGE2 network on disaster management and marine biology study is another case that exemplified the group's efforts. The group also helps build communities, including the newly formed Smart Bay Cities collaboration, which will expand on the experiences and technologies developed this year to project future opportunities for societal impact.



PRAGMA WORKSHOPS

In 2018, two PRAGMA Workshops were held:

- **PRAGMA 34:** May 9–12, 2018, Akihabara, Tokyo, Japan. Co-hosted by AIST and Osaka University. PRAGMA 34 was held in conjunction with a Workshop on High Performance Infrastructure for AI, in conjunction with the 3rd annual CENTRA meeting.
- **PRAGMA 35:** October 3–6, 2018, Penang, Malaysia. Hosted by Universiti Sains Malaysia. Co-sponsored by Big Data Week Asia. Held in conjunction with Big Data Week Asia and the Big Data Summit 2. In addition, a MYREN (Malaysia Research and Education Network) and Network Startup Research Center co-located a workshop, and Kamal Hisham Kamaruddin, head of MYREN, gave a keynote at PRAGMA 35, and discussed connectivity between Malaysia and other PRAGMA sites.

More information about the PRAGMA 34 and 35 workshops can be found at www.pragma-grid.net/pragma34 and www.pragma-grid.net/pragma35, respectively.

In addition, training programs such as the Southeast Asia International Joint Research and Training Program (SEAIP, seaip.narlabs.org.tw), newly renamed the International Joint-Research and Training Program, provided PRAGMA with new members (See Building Community section).

Looking to the future, we will continue to use 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 36:** April 25–27, 2019, to be hosted by the Korea Institute of Science and Technology Information (KISTI) in Jeju, Korea in conjunction with the 4th CENTRA Workshop (April 22–24, 2019).
- **PRAGMA 37:** September 11–13, 2019, to be hosted by UC San Diego, La Jolla, California, USA.

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.

PRAGMA 35 Workshop, USM, Penang Malaysia, October 2018

MEMBERS & PARTNERS

*PRAGMA is an institution- and people-based organization governed by a steering committee (SC) that invites new members, determines locations of workshops, and sets overall direction for PRAGMA initiatives. More information about SC members (denoted with an asterisk *) can be found at www.pragma-grid.net/people.php.*

ACTIVE MEMBERS

A key component of PRAGMA is active involvement by participating in workshops, contributing resources, hosting workshops, and/or promoting and supporting student and researcher exchanges. The following institutions and members 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*: Yoshiyuki Kido, *kido@cmc.osaka-u.ac.jp*

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

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

KONKUK UNIVERSITY (Konkuk): Karpjoo Jeong* (ex-officio), *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* (ex-officio), *wftsai@nchc.narl.org.tw;* Fang-Pang Lin*, *fplin@nchc.narl.org.tw;* Weicheng Huang*, *whuang@nchc.narl.org.tw,* Hsiu-Mei Chou, *hmchou@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*; Ryousei Takano, *takano-ryousei@aist.go.jp*

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

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

UC SAN DIEGO: including the California Institute for Telecommunications & Information Technology (Calit2), San Diego Supercomputer Center (SDSC), Center for Research in Biological Systems (CRBS): Shava Smallen*, *ssmallen@sdsc.edu;* Philip Papadopoulos* (ex-officio, as of July 1, 2018 at UC Irvine), *philip.papadopoulos@gmail.com;* Peter Arzberger*, *parzberg@ucsd.edu;* Nadya Williams, *nwilliams@ucsd.edu;* Teri Simas, *simast@sdsc.edu* (until June 1, 2018)

UNIVERSITY OF FLORIDA (UF), 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;* Matthew Collins, *mcollins@acis.ufl.edu,* Grace Hong, *gshong@acis.ufl.edu*

UNIVERSITY OF HONG KONG (HKU): Wing Keung Kwan* (ex-officio), *hcxckwk@hku.hk;* P.T. Ho, *hcxchpt@hku.hk;* Lilian Y.L. Chan, *lilianyl@hku.hk*

UNIVERSITY OF WISCONSIN-MADISON (UW-Madison), the Center for Limnology: Paul Hanson, *pchanson@wisc.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*

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@nectec.or.th*

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

NETWORKING MEMBERS

Networking partners provide access to expertise to improve the efficiency of the resource's 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 & SDX, AND GLOBAL RESEARCH PLATFORM INITIATIVES: Maxine Brown, *maxine@uic.edu;* Joe Mambretti, *j-mambretti@northwestern.edu;* Thomas DeFanti, *tdefanti@ucsd.edu*

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

For more information about PRAGMA 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 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 were piloted at eight universities in the 2015–2016 academic year and were assessed to determine how participation in the module activities alter student reasoning about climate change and computing.

PARTNERS

GLEON (*www.gleon.org*) 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.

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 recently worked with PRAGMA in Southeast Asia to support researchers from Myanmar to attend PRAGMA 24. In addition, NSRC has been able to encourage participation in PRAGMA workshops, such as collocating a NSRC tutorial for MYREN (Malaysian Research and Education Network) concurrently with PRAGMA 35, and encouraging MYREN to provided a plenary talk at PRAGMA 35, bringing together these two communities.

SPONSORS

PRAGMA is supported by its member institutions and the U.S. National Science Foundation (NSF OCI 1234983, PI: Smallen, Co-PI: Williams), involving and supporting researchers at UC San Diego, University of Florida, Indiana University, University of Wisconsin, University of Kansas, and Virginia Tech. 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. In addition, previous support came from NSF OCI-0627026, PI: Papadopoulos and involved support from NSF's Office of Shared Cyberinfrastructure, Office of International Science and Engineering, Division of Information and Intelligent Systems, and Division of Biological Infrastructure.

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

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 Grant Numbers JPMJCR1303 and JPMJCR1414, and JSPS KAKENHI Grant Number JP17H01748.

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.

Indiana University receives additional support from NSF 1659310 funded robust PID testbed, called the RPID testbed (pronounced "rapid").

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 MSIP through the EDISON Project and KISTI through its Supercomputing R&D Program.

Monash University's 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 receives its funding through Thailand's National Science and Technology Development Agency.

Osaka University and National Institute of Information and Communications Technology (NICT) research was supported in part by institutional funding for their collaborative research project titled "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 and SINET5 of the National Institute of Informatics (NII), Japan.

StarLight receives major funding from NSF (ACI-1450871) for the project titled "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 through 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.

Virginia Tech and University of Florida receive additional funding from NSF 1737424 for the Smart and Connected Water Systems project as part of its Smart and Connected Communities program.

PRAGMA would like to acknowledge and thank those who helped prepare this report: Jennifer Matthews for graphic design; Molly Wofford for editing; Grace Hong coordination at PRAGMA 34 and 35; and all of the individuals who contributed photos, including Jason Haga, Paul Hanson, Grace Hong, Shava Smallen, Nadya Williams, Wassapon Watanakeesuntorn, Gerald Pao, and students in the MURPA/QURPA program.

PRAGMA
www.pragma-grid.net