

PRAGMA: Enabling cyberinfrastructure for long-tail of international science communities

Shava Smallen
Co-chair, PRAGMA Steering Committee
University of California, San Diego

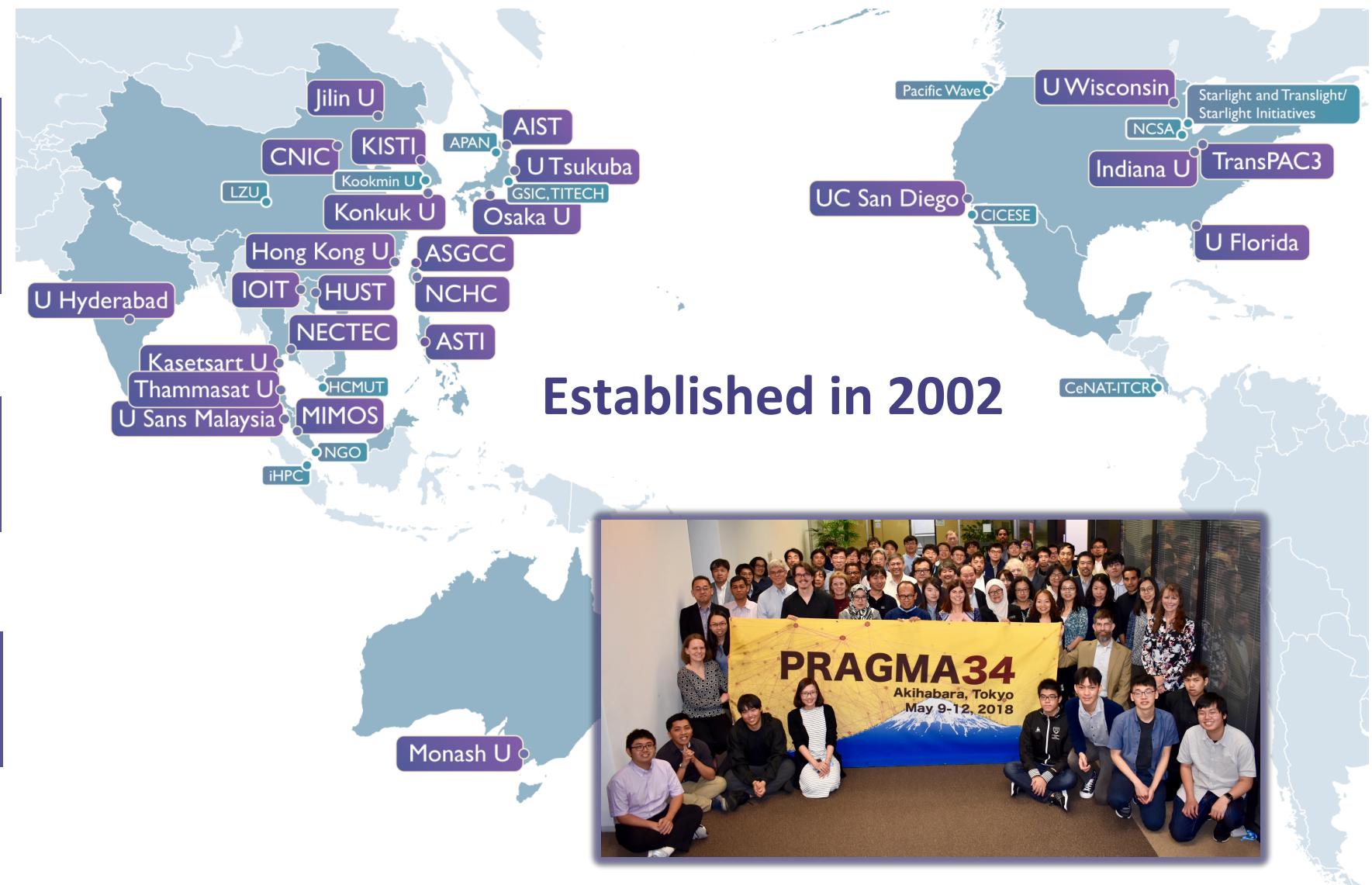


PRAGMA is an Open Community of Practice

Focused on researchers
and institutions on the
Pacific Rim

Engages “Long Tail”
science communities

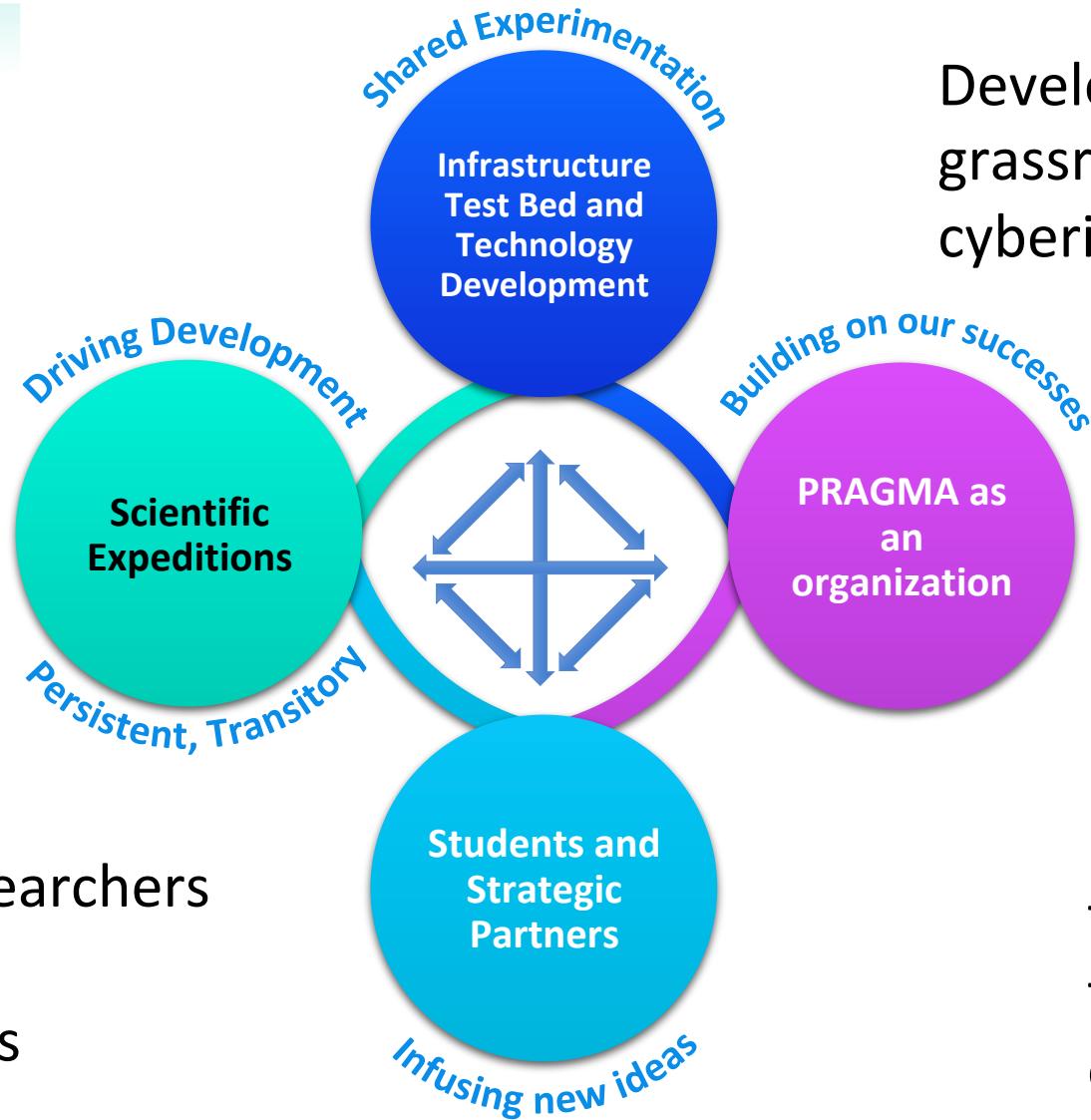
Twice yearly meetings
for measured progress



PRAGMA – Building Trusted Community of Practice through four strategies of collaborating

Fostering international scientific expeditions by forging teams of domain scientists and CI researchers

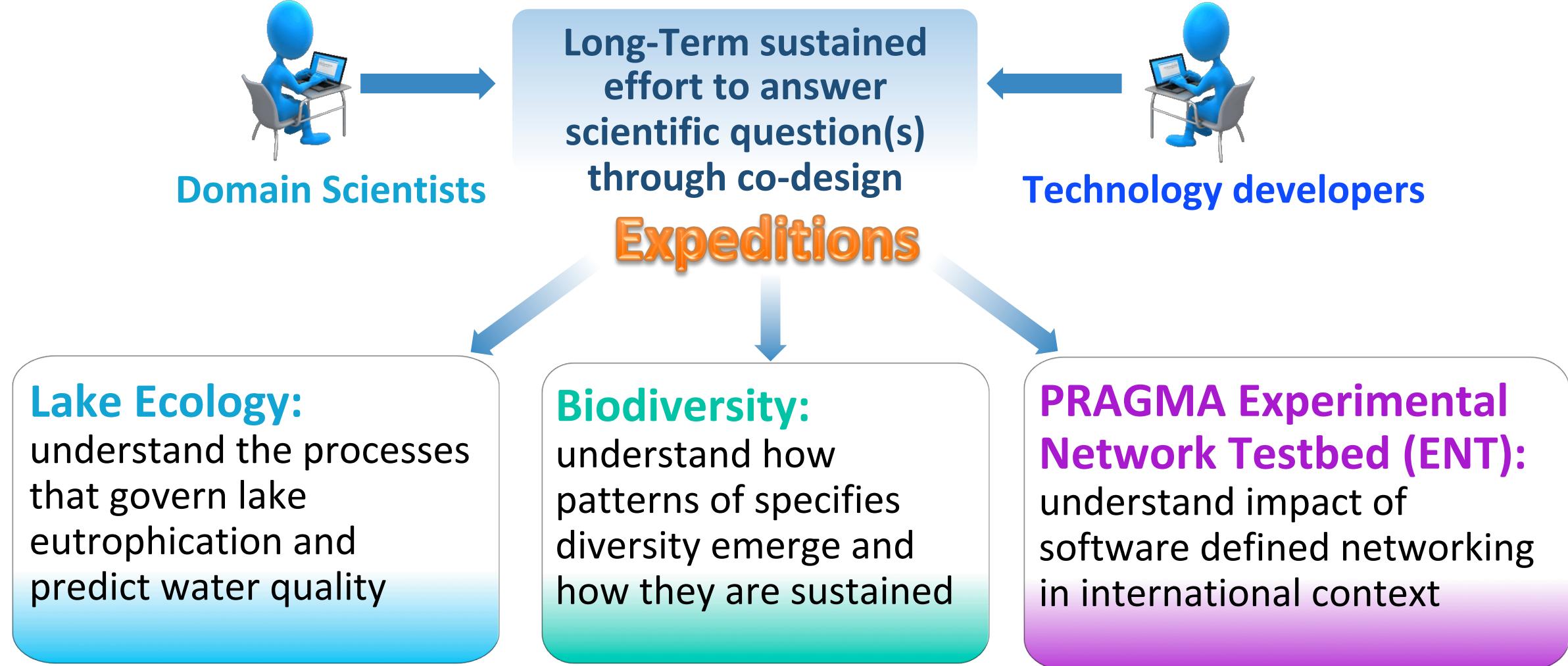
Infusing new ideas by developing young researchers and by engaging with strategic partners



Developing and improving a grassroots, international cyberinfrastructure

Building and enhancing the essential people-to-people trust and organization

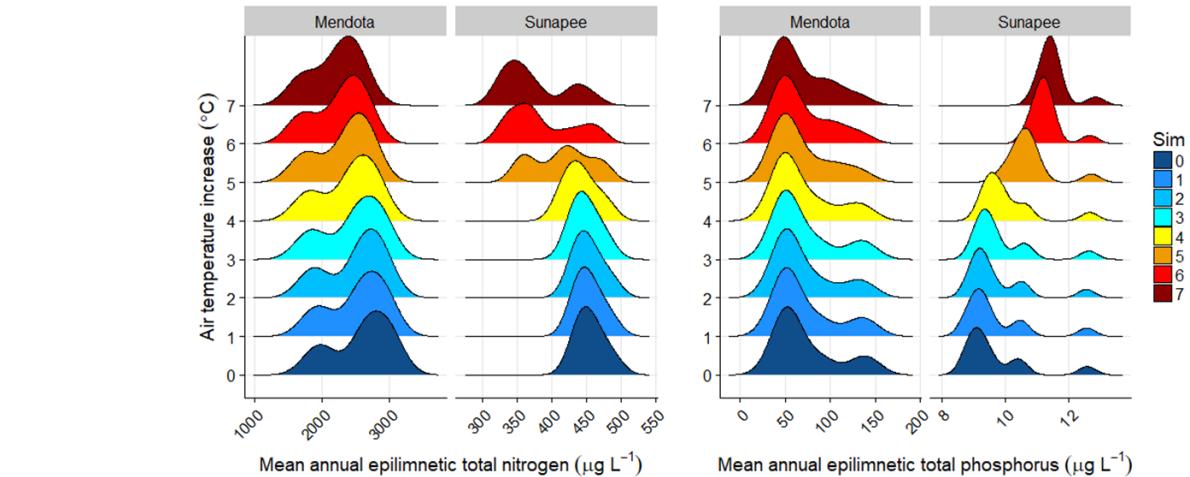
PRAGMA's Expeditions: A Model of Collaboration



Lake Expedition: Predicting Water Quality in Lakes

Paul Hanson (U. Wisconsin), Cayelan Carey (Virginia Tech), Renato Figueiredo (U. Florida)

- **Eutrophication:** excessive richness of nutrients in a lake or other body of water, frequently due to run-off from the land, which causes a dense growth of plant life.
- **Lake eutrophication** is global issue, results in **degraded water quality**
- Goal: Integrate sensor data as inputs to **computational lake models** → science goal is to *predict* water quality



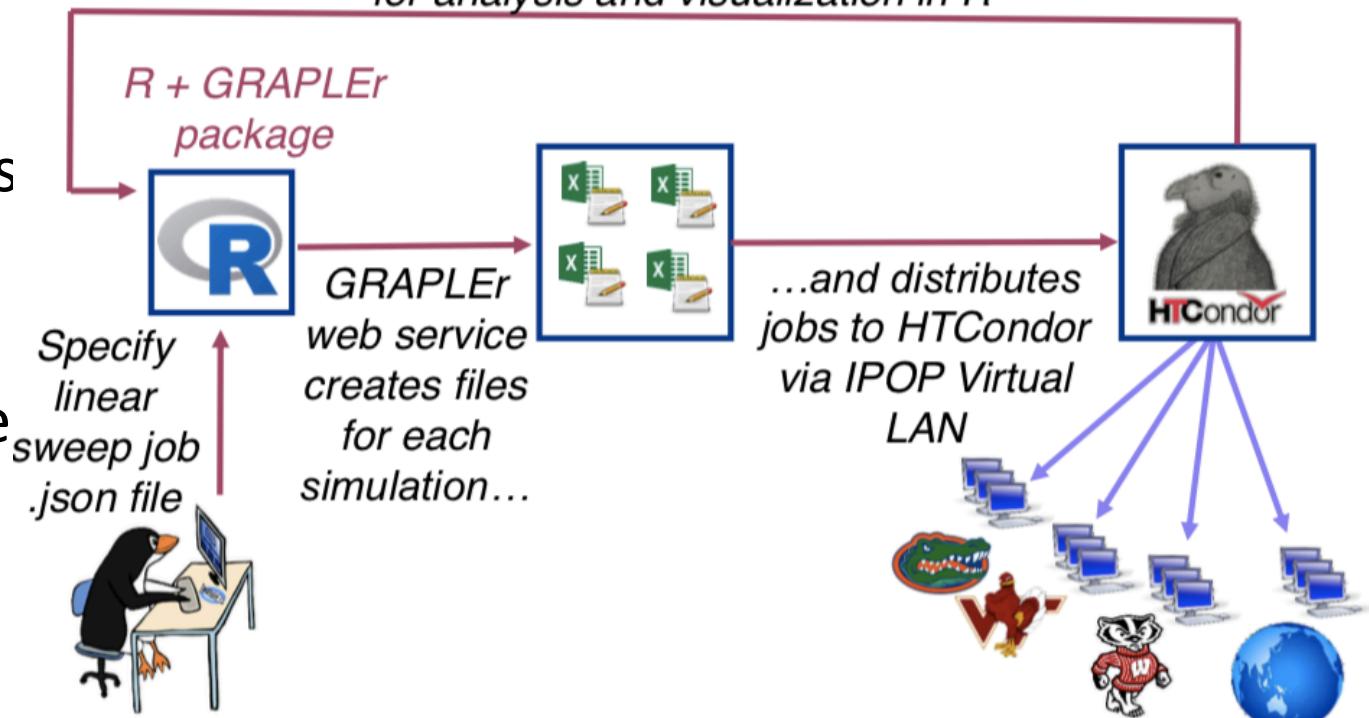
Lake Expedition: Developing Predictive Models using High-Throughput Computing and Overlay Networking

- **Sensor gateways** link sensors to internet (connecting to storage and models)
- **IPOP overlay network** create virtual private network for sensor data and computation
- **GRAPLER distributed computing** uses an R interface to allow ecologists access to distributed computing
- **Educational Modules** educate future generations in computational skills (currently at 12 universities)

Prof Carey teaching students to use GRAPLER at VaTech



Outputs from model runs are aggregated and returned to user for analysis and visualization in R



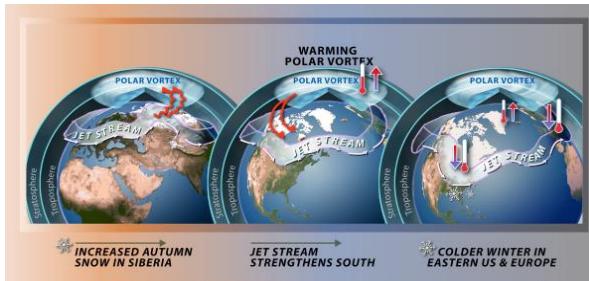
Virtual Biodiversity Expedition: Research and Predict Biodiversity of Terrestrial Species



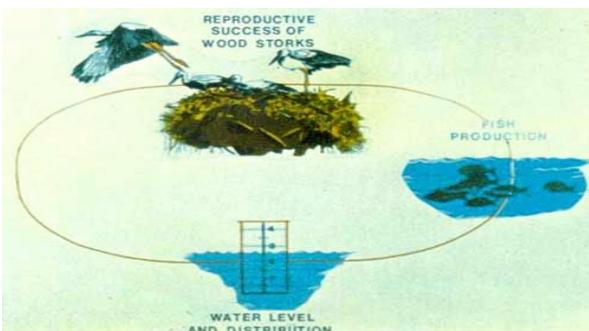
Biodiversity Inventory



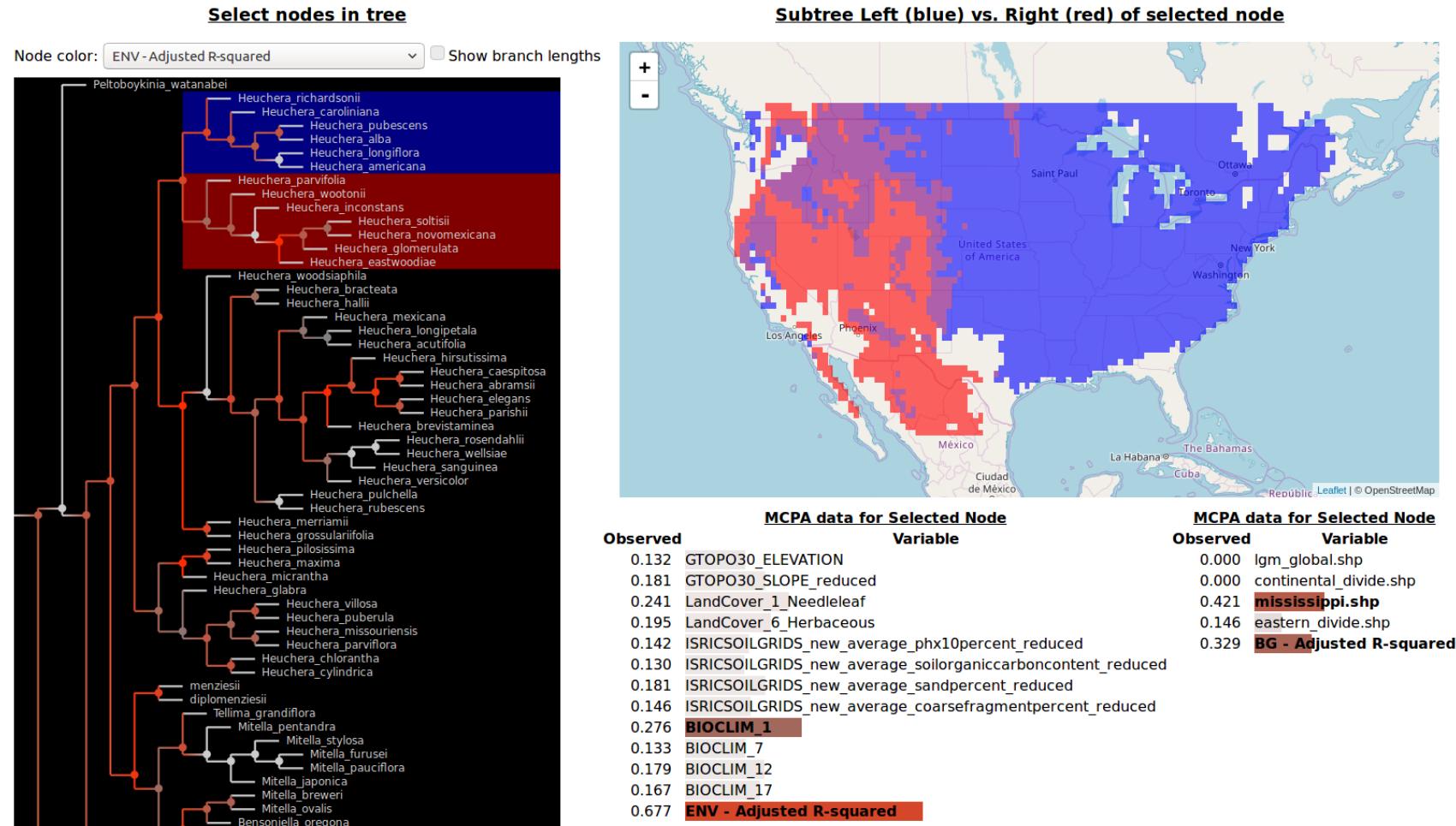
Global Climate Change



Macro-Ecological Modeling

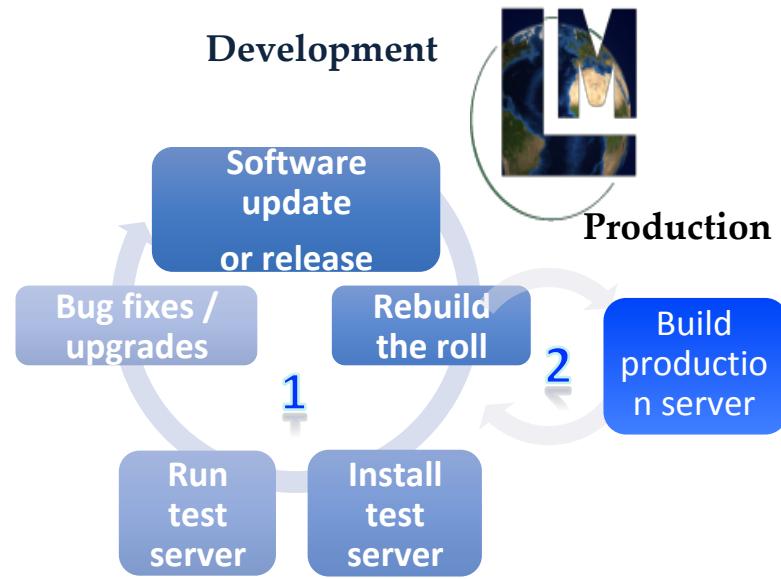
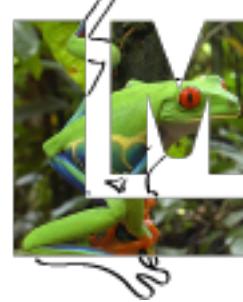


Aimee Stewart (U. of Kansas), Nadya Williams (UCSD)



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

Virtual Biodiversity Expedition: Expanding and Enhancing Lifemapper Deployments



- **Software engineering and Rocks increased availability and flexibility of Lifemapper**
- **Lifemapper deployed in a variety of environments** such as researcher laptops, project servers, HPC resources (US XSEDE Comet)
- **Working to facilitate data ingestion** to enable regional Lifemappers that use locally-available high resolution data
- **Training**

PRAGMA Experimental Network Testbed (ENT) Expedition

Motivation:

an International Software-defined Networking (SDN) Testbed for use by PRAGMA researchers and collaborators

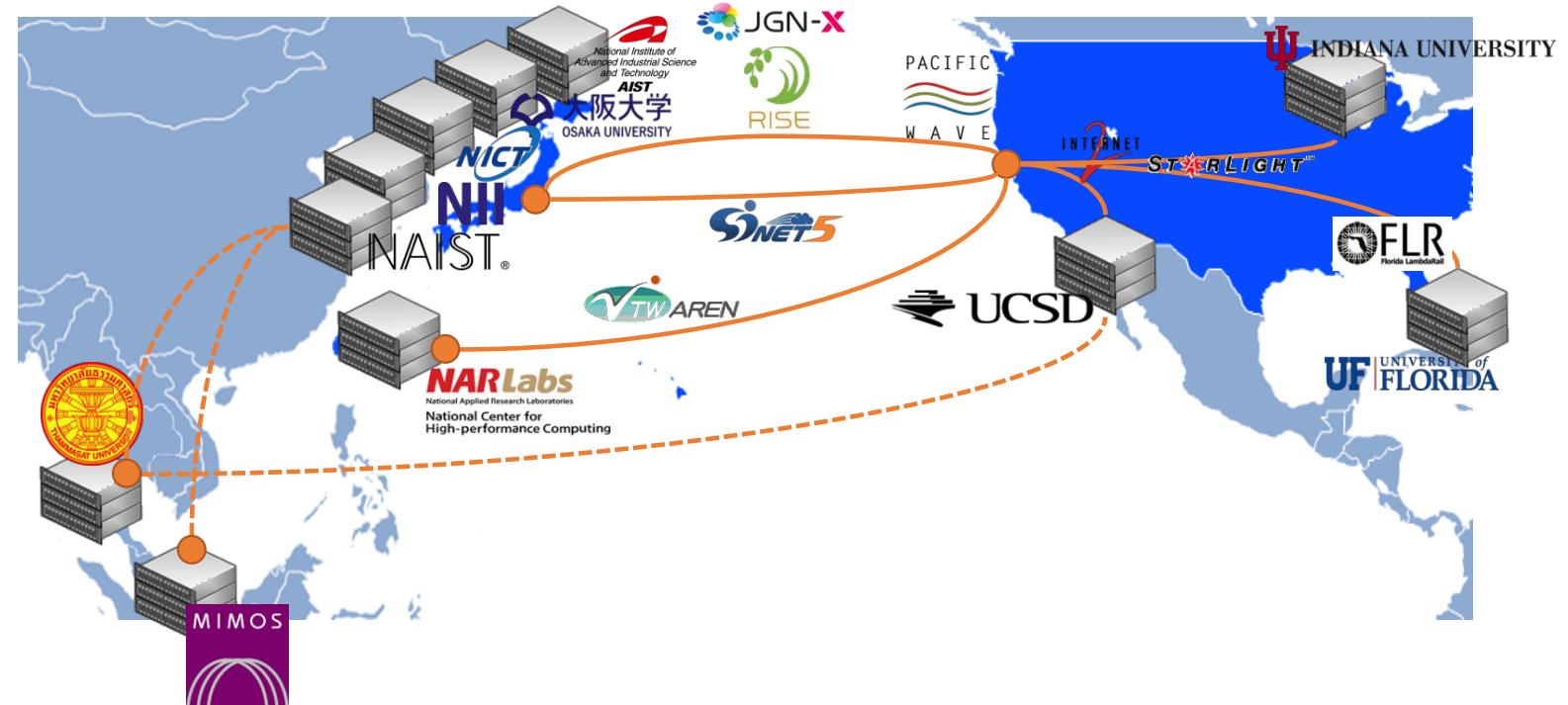
Accomplishments:

4 countries and 11 institutions have been involved in building the large scale SDN testbed

Using NAIST's AutoVFlow as a openflow meta controller

Numerous Publications

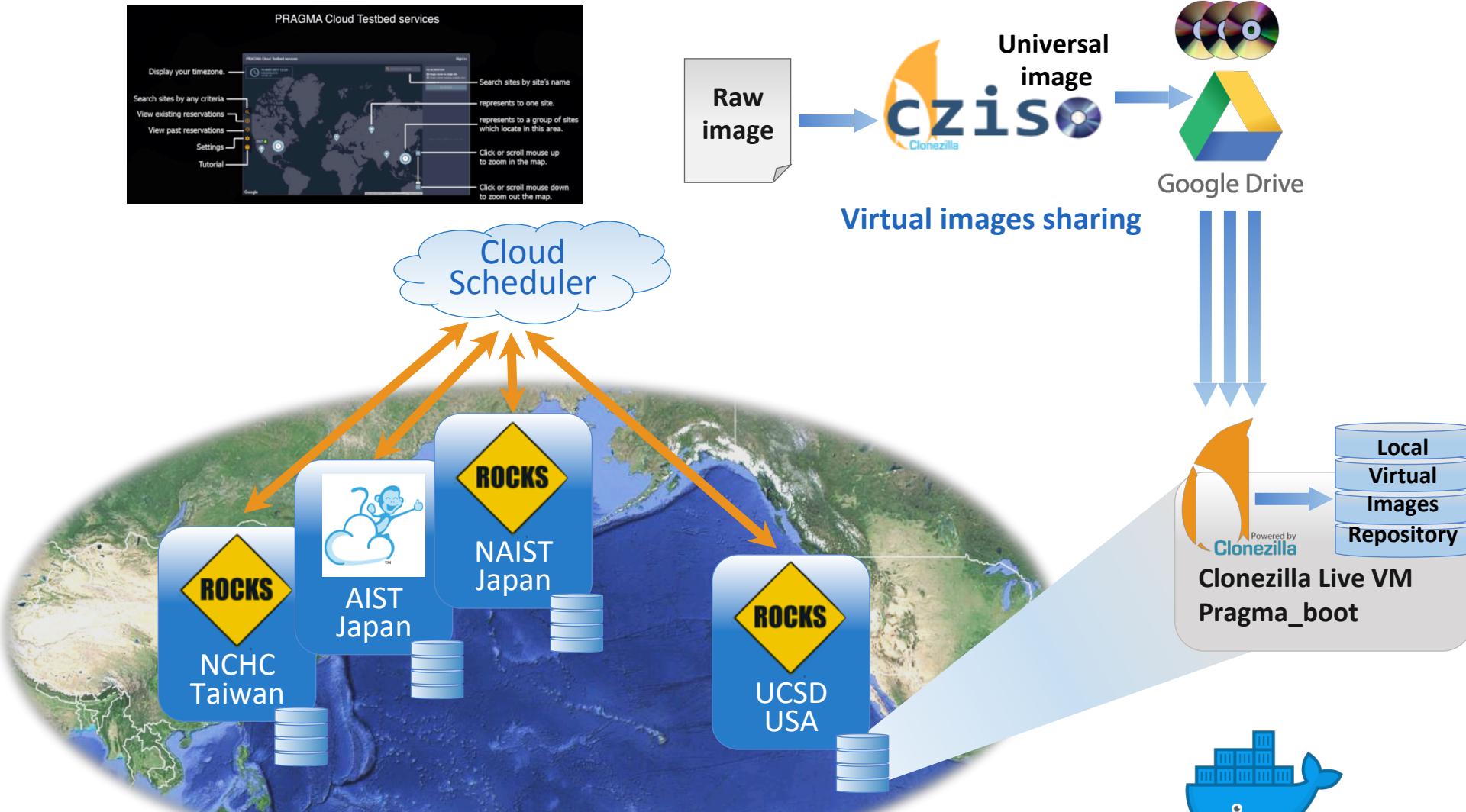
- Efficient Packet Header Rewriting (PARES) – (U. Florida, NAIST)
- Multipath TCP – NAIST with student visitation to UCSD.
- SMOC – Simple Multipath Openflow Controller
- Overseer: Latency and Bandwidth aware routing.



IPOP Integration

- OpenFlow hardware switches no longer required to participate in ENT
- IPOP overlay network allows tunneling of traffic from/to SDN-programmed switch ports across the Internet
- Software-based SDN switches (Open vSwitch)

Building on the International Development and Collaboration to Create PRAGMA's Multi-Cloud Testbed for Advancing Science



Using deep learning to detect museum specimens that had been contaminated with mercury salts

Contaminated



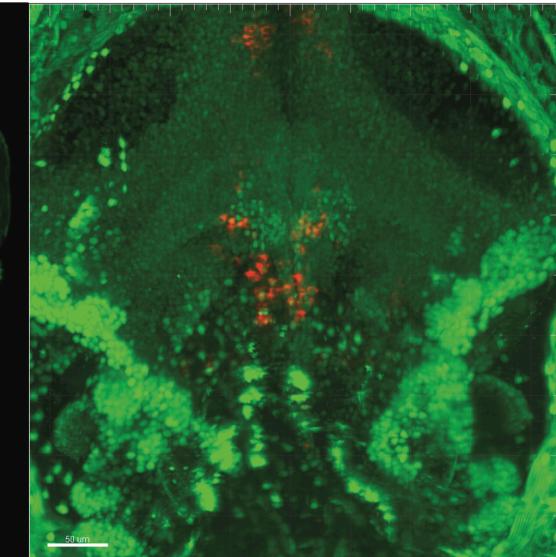
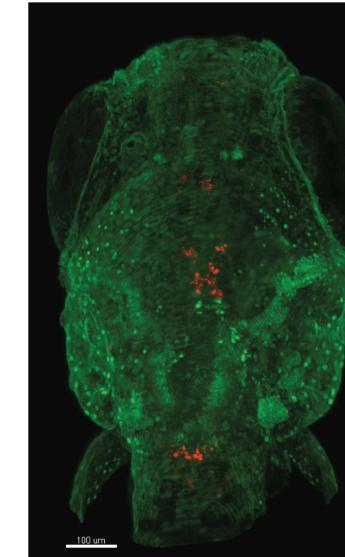
Unsure



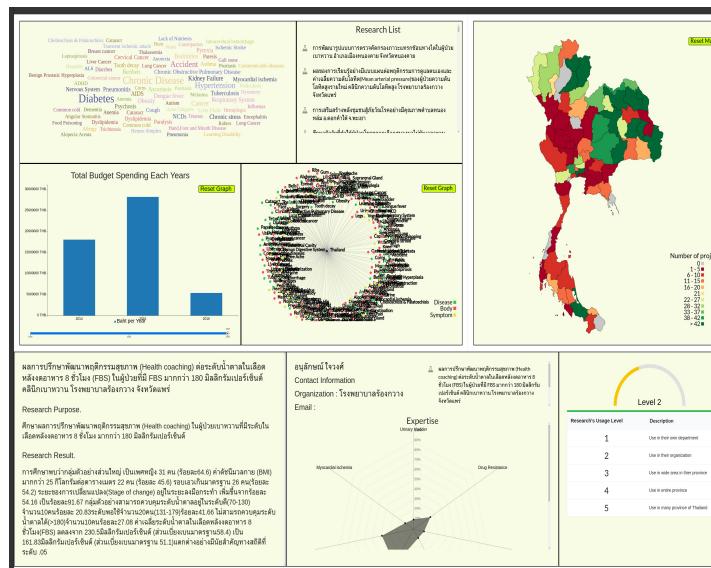
Clean



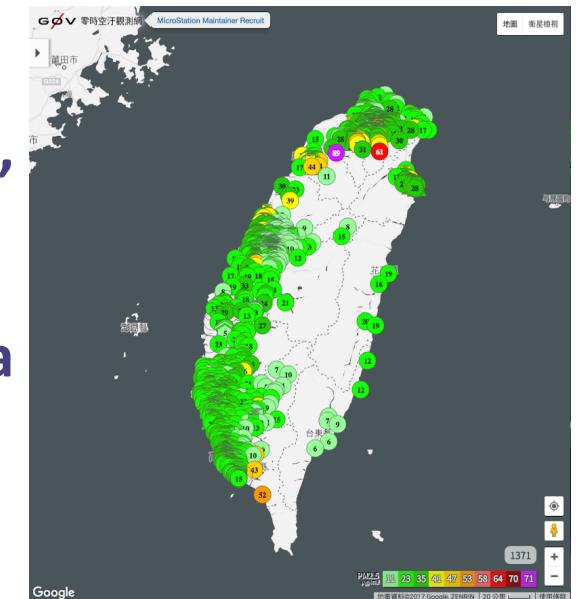
Using neural deep learning to learn architectures of the zebra fish brain



Visualizing data using SAGE2 display walls



**Enabling FAIR
(findable , accessible,
interoperable,
reusable) access to
Airbox pollution data
using persistant
identifiers**



PRAGMA: Co-designing useful Cyberinfrastructure (CI), transforming long-tail science communities

- Science is **inherently international** and requires collaboration.
- Collaboration is **enabled by sharing and exchanging data, algorithms and codes**
- Cyberinfrastructure is only one dimension; **people and trust are also essential**
- There are fundamental challenges in matching existing CI to hundreds of communities. **Deep interactions with long-tail communities has the potential to transform both science domains and CI development.**

PRAGMA seeks to address these challenges to improve *CI and scientific impact via practical implementation*. We focus on international collaborations that uniquely team technology specialists and domain scientists



Thank you!

Thank you to Nurul and Big Data Summit 2 Committee!

We look forward to more collaboration so please talk to us about your ideas.

Future Meetings

- PRAGMA 35 Penang, Malaysia, October 3-5, 2018 <http://www.pragma-grid.net/pragma35>
- PRAGMA 36 Jeju, South Korea, April 26-27, 2019

Web: www.pragma-grid.net

Software: <https://github.com/pragmagrid>

Info: pragma-grid-team@googlegroups.com

