Reports from Resources Working Group

Philip Papadopoulos (UCSD)
Yoshio Tanaka (AIST)

PRAGMA 31

Bangkok, Thailand

Agenda of breakouts

1st day (14:00 - 15:30, 8, Sep.): Discussions for short term goals.

Lake Expedition
- Paul, Cayelan, Renato

PRAGMA-RDA
- Beth

PRAGMA-ENT
- Kohei

2nd day (14:30 - 16:00, 9, Sep.): Discussions for long term vision.

Presentation: A new data set gathering idea for the Lower Mekong region (Tho Nguyen)

PRAGMA Cloud followed by the discussions on the future vision
- Shava, Nadya

Outline of the talk on Lake Expedition – Grand Challenges

1. Setup: GLEON has never been able to identify THE science questions that the network should address; rather, GLEON provides the framework for bubble-up science.
2. With #1 in mind, I think it makes to frame this in terms of how science is done
3. Two slides on what GLEON is
4. Three slides on the current Lake Expedition, with science focus (Renato will give separate technology talk)
5. Nine slides on three GLEON Grand Challenges

How can PRAGMA & GLEON work together to address these challenges?

- Expand upon our existing GRAPLEr framework for lake modeling
- Develop new compute strategies/infrastructure to aid modeling in service of the science
- Collaborate with GLEON scientists for managing and visualizing diverse datasets (e.g., citizen science app data)
- And in multiple other ways…

PRAGMA-ENT Accomplishments & Activities (PRAGMA 31)

- ENT backbone
  - Connection to Thammasat University with GRE
  - RISE OpenFlow SW at LA joined as a node.
- IPOP access points on ENT
- Monitoring & visualization tool [Demo]
- Integration with Rocks and PRAGMA boot [Student Hackathon]
- Authentication & Secure SDN [Poster]

Future Plans

- Network expansion (more sites)
  - Interconnecting RISE (JGN-X) & FSFW (Internet2)
- End user support
  - Visualization
  - Automation and API for AutoVFlow
  - Wiki: https://github.com/pragmagrid/pragma_ent/wiki
- More applications
  - Disaster management applications
  - Tiled Display Wall
  - Security, Authentication
  - Trust data sharing
    (HathiTrust Digitalized Books Corpus)

Roadmap of ENT

- Pragma32
  - Reservation or scheduling service
    - VM & Network
  - User friendly UI
    - Visualization
  - Making reservation from Web or API & automation
    - AutoVFlow needs to expose APIs
- Pragma33
  - Expand ENT to at least 10 institutions

RAGMA-RDA: Demo Phases

Phase 3 (Sep 2016, PRAGMA 31):

- Demo: Rice bioinformatics researchers can execute Tassel pipeline using PRAGMA Cloud resources. The new PRAGMA data services are used to identify, download, and faithfully rep the computational experiment. Data objects are also stored in a repository.
- PRAGMA data services: PID Information Types API service (with PRAGMA extension), Data Type Registry service, experimental Handle service, client
- Core functionality achieved, packaging of services for other domains

Phase 3.1 (Sep 2016, RDA P8 Denver):

- Demo same workflow as in PRAGMA 31 as part of adoption report plenary session
- Feedback to RDA PID/DTR working groups

Phase 4 (Apr 2016, PRAGMA 32):

- User testing of service
- Hardening of code and usability improvements in UI and landing pages
- Exploration of contextual provenance inclusion, other PID types, other pipelines

PRAGMA-RDA: Goals for PRAGMA 32

Integrate more virtual cluster images and get feedback from early users (Bioscience WG)

- Biolinux VM up and running
- Modularization to accommodate additional use cases
  - AirBox in collaboration with Tele Science WG.
- AIST resources for PRAGMA Cloud ready
  - DELL PowerEdge M610, 2 Blades ready, CentOS 6.7
    - Intel Xeon E5620 2.4 GHz 4core * 2 = 16 nodes
    - memory: 24 GB, storage: 600 GB HDD * 2
    - Network: 10GbE: for public/guest networks, 1GbE: for management/storage networks
- Cloudstack drivers for pragmaboot
  - Core functionality implemented
- Tassel pipeline installed
  - Other pipelines are in progress
- Have 3 IRRI users ready to test in upcoming months
  - Feedback for both cluster usage and data service

tablishing a Central Research Database
r the Lower Mekong Region

Report on Project Planning Meeting (Sept. 6, 2016)

September 9th, 2016

Tho H. Nguyen
University of Virginia
Gordon Holtgrieve
University of Washington

overview

e need for data sharing:

Research increasingly focusing at the ‘nexus’ of domains and systems

Researchers wants to share data, and ultimately return value to the Mekong

Need for data discovery, awareness & coordination

Difficulties due to data owners’ lack of control over data

capable and accessible database:

Distributed architecture bringing the database to data owner

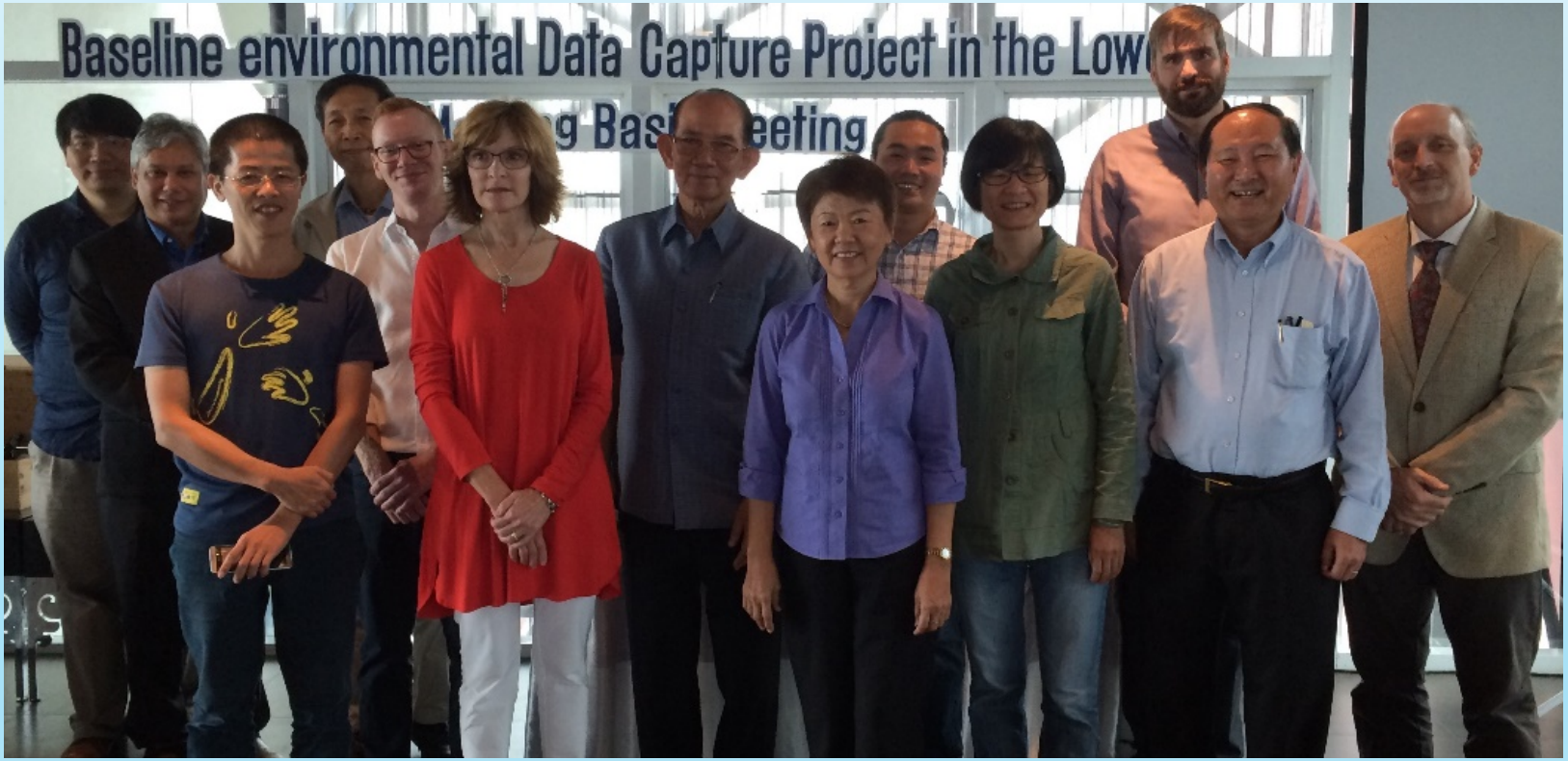
Workflow to support custom protection control over datasets

Bottom-up architecture to “crowd-source” data

Tools and proceses to support data curation

strategy – step 1

- Announce, solicit input, build support



strategy – step 2 …

- Identify partners to develop/implement database
  - PRAGMA, CENTRA, regional partners, …
- Design input from data owners, data users
  - SEAIP 2016, PRAGMA 32, …
- Secure funding for US, Asia Pacific, and Mekong regional activities
- Develop, deploy proof-of-concept system
- Build regional support, scale up

PRAGMA Cloud: Goals for PRAGMA 32

- Integrate 3 additional sites: NAIST, IU, UF
- Integrate PRAGMA-ENT, enhancements to openvswitch
- Finish image management with Google drive and Clonezilla (in progress)
- Develop Openstack driver boto (NCHC)
- Integrate more virtual cluster images? Wa-dock – newer version?
- Cleanup documentation on how to become a pragma site
- Add user documentation
- Longer term goals
  - Comet integration ?
  - Access to common datasets? Handled within VM