Asia Federation Report on International Symposium on Grids and Clouds (ISGC) 2011

Simon C. Lin

Academia Sinica Grid Computing Centre (ASGC)
Taipei, Taiwan
E-mail: Simon.Lin@twgrid.org

Eric Yen

Academia Sinica Grid Computing Centre (ASGC)
Taipei, Taiwan
E-mail: eric.yen@twgrid.org

Overview

This report provides an overview of e-Science developments in the Asia-Pacific region, based on country reports made at the International Symposium on Grids and Clouds 2011 (ISGC 2011) [1], held on 21-25 March at Academia Sinica, Taipei. The document includes a brief overview of 12 Asian countries presented at ISGC 2011. They are: Australia, India, Indonesia, Japan, Malaysia, Pakistan, Philippines, Singapore, South Korea, Taiwan, Thailand and Vietnam.

The country reports represent a wide range of Grid technology penetration and e-Science application, reflecting to some extent the different countries' socio-economic levels of development. In 2010, the participation in the Worldwide LHC Computing Grid (WLCG) continues to be the incentive for many countries to advance the Grid technology. Furthermore, the Grid related EU initiatives under the Seventh Framework Programme (FP7) continue to be the driving force behind the collaboration in the Asia-Pacific region. EUAsiaGrid, a project led by ASGC and INFN, has been the primary driver for the e-science infrastructure and Grid related developments in the region since 2008. Along with Enabling Grid for e-Science (EGEE) project, now the Integrated Sustainable Pan-European Infrastructure for Researchers in Europe (EGI-InSPIRE) project , they facilitated regional collaboration and bridged Asia with the world.
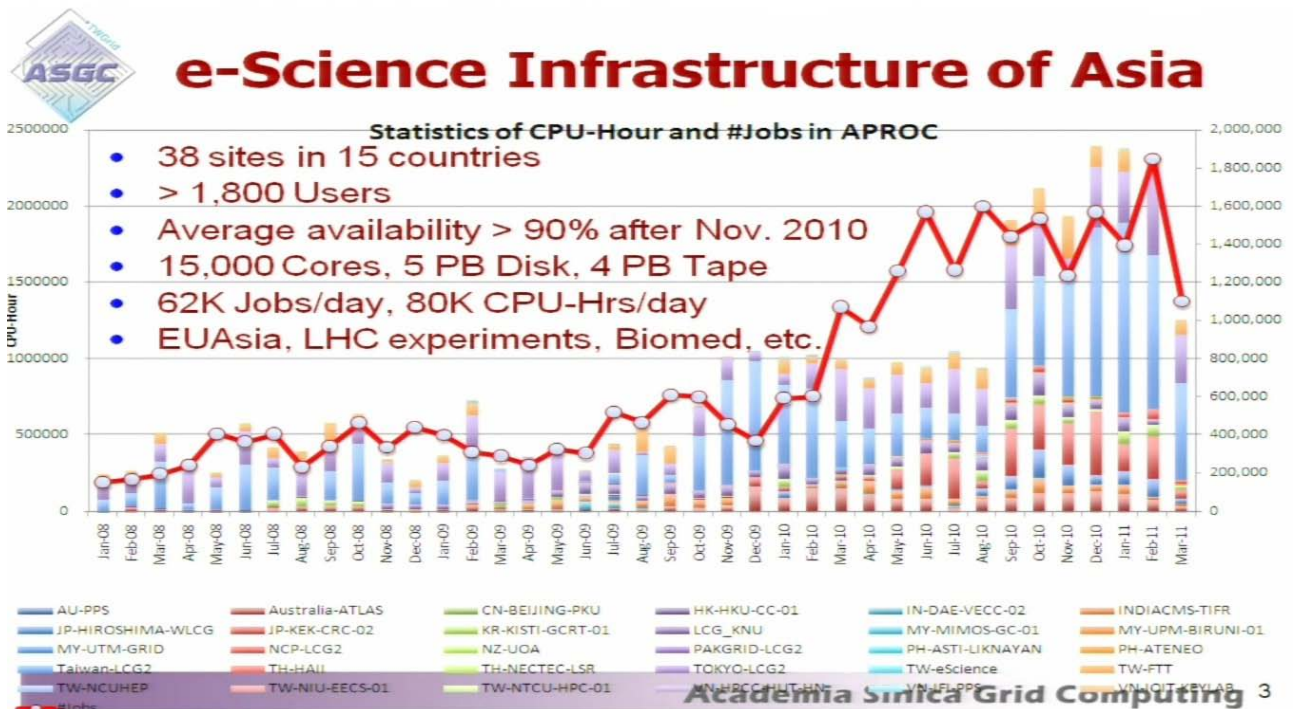
Fig. 1: e-Science Infrastructure in Asia.

The graph above shows the growth of the Asia Pacific e-Science Infrastructure in the past 3 years in terms of CPU-hours and submitted jobs. Starting 2006, the WLCG Asia Pacific regional operation center, run by Academia Sinica Grid Computing Centre (ASGC), has been supporting 38 sites in 15 countries. According to the statistic data of the Asia Pacific Grid Policy Management Authority (APGrid PMA), there are more than 1,800 user certificate issued. Recently, ASGC has gotten each site to reach more than 90% of resource availability. From January 2011, there have been about 62,000 jobs per day, which is 10 times more than 3 years ago. CPU-hours have also increased by 10 times to 80,000 per day compared to January 2008. Most of the resources are used by the LHC experiments, but the EUAsiaGrid project has also drove up the growth since it began in 2008, along with other distributed e-Science collaborations in the region.

The leadership of ASGC, the only WLCG Tier-1 Centre and also the e-Science Asia Centre in the Asia-Pacific region, has influenced the e-Science development in the region. ASGC has been contributing to multilateral collaboration projects, workshops as well as training programs on crucial e-Science research topics in the region such as drug design simulations and disaster mitigation which includes earthquake monitoring, seismic and tsunami simulations.

Even with the progress in each country, there are still challenges facing the Asia-Pacific e-Science community. With the ending of EUAsiaGrid in 2010, sustainability

has been a major concern for regional collaborations. Projects in some countries have been struggling with funding while some in need of high-quality human capital. As for infrastructure, having more stable and affordable high bandwidth international network connections set up continue to be the common need among Asia-Pacific countries. Last but not least, the developments in cloud computing, faster in some countries than others, intertwine with the Grid computing and have become one of the focuses in the region.

Australia

The National eResearch Collaboration Tools and Resources (NeCTAR) project, led by University of Melbourne, is relatively new to the Australian research landscape. It runs on a particular funding called the Super Science Investment for a total of $47M that lasts from 2009 to 2013 from the federal government, which has a main focus in data management. The project objective is to enhance research collaboration through the development of shared eResearch infrastructure.

NeCTAR is mandated to be extremely responsive to the research sector of Australia. Right now, NeCTAR is in the project planning stage. It released a consultation paper in October 2010 and received 69 responses from different institutions all over the world. The project infrastructure will begin to develop and construct in 2011.

Below provides a full glance of the NeCTAR project and its 4 streams for infrastructure development.
1. Virtual Laboratories- Connecting resources and capabilities for research communities.
   - Forming virtual laboratories around engaged Research Communities
   - Integrating existing research and research capabilities. Supporting research workflows.
   - Leveraging other NeCTAR streams: Research Cloud, Research Tools.

2. Research Cloud- National Infrastructure for Research Applications
   - Providing an enduring, open, scalable national platform for deploying Research applications and workflows.
   - Funding Cloud Nodes to form a National Research Cloud
   - Migrating research applications to Cloud infrastructure such as HPC and Data portals, statistics packages, collaboration, visualisation, and data tools.
   - Providing opportunities for Federation, outward with international research clouds and inward with emerging university research clouds

3. Research Tools- Software Development for Research Tools
   - NeCTAR will enhance and develop eResearch tools to improve access to data, research collaboration, and research workflows which are deployed on the NeCTAR Research Cloud.
   - According to researcher-defined needs to open Calls for Proposal, to solve Discipline and Problem-Oriented needs, and to supporting Research Communities.
   - Example Tools include: TARDIS, BioFlows, LORE and GRHANITE.

4. National Server Program- Robust national server infrastructure for core national services.
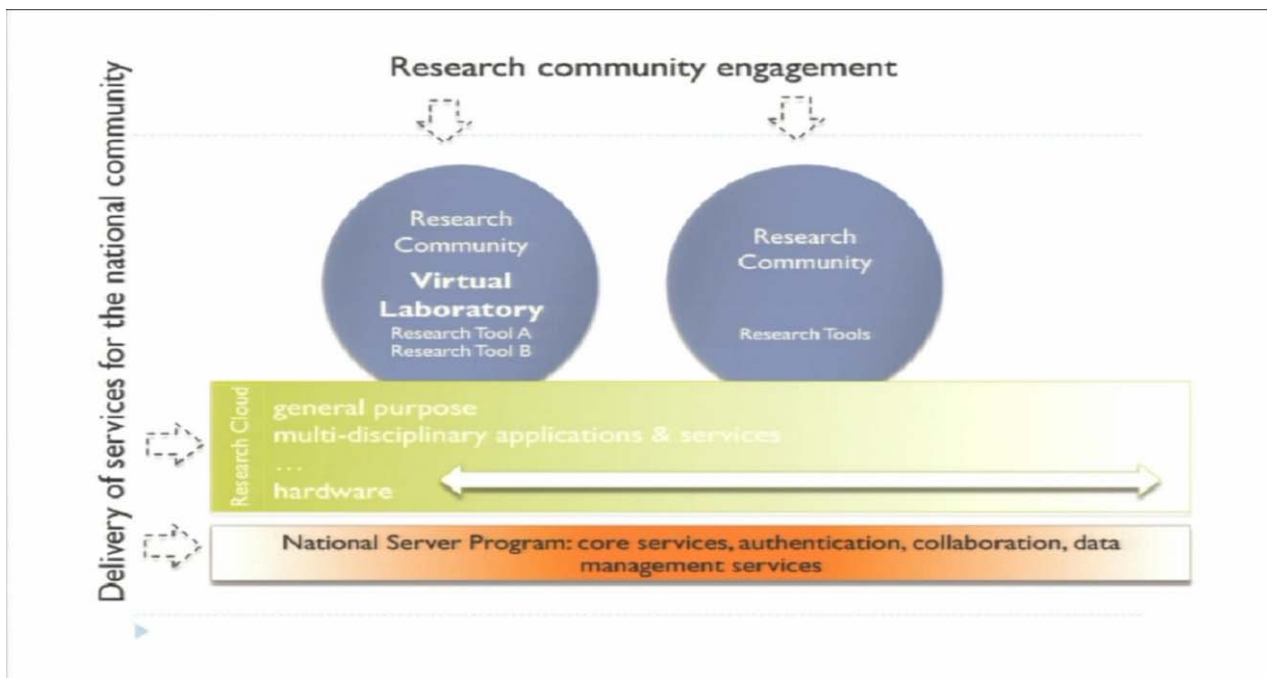

Fig. 2: NeCTAR virtual laboratories and research communities.

When the NeCTAR project ends in 2014, it is expected to have 8-10 Virtual Laboratories, 6 Nodes of the Research Cloud, suite of research tools built, deployed and operated, and 3 National Server Program Nodes.

[Based on the presentation made by Glenn Moloney, NeCTAR] [2]

India

EUIndiaGrid2 (2010-2012) relies on the previous activities done by EUIndiaGrid from 2006-2009. EUIndiaGrid2 project comes with the following objectives:
1. Consolidate and enhance cooperation between European and Indian e-Infrastructures for the benefit of EU-Indian collaboration in e-Science;
2. Support specific user communities in the exploitation of grid infrastructures in areas strategic for EU-Indiana collaboration, for example, biology & material science, climate change, high energy physics;
3. Ensure a sustainable approach to e-Infrastructures across Europe and India through dissemination actions, meetings & workshops;
4. Foster and enhance cooperation with other European Initiatives in the Asian region and worldwide.

Since the launch of the project, EUIndiaGrid and EUIndiaGrid2 tried to be at the core of the European and Indian cooperation. First notable effort was the India-EU Workshop on Research Infrastructures on January 2010 at Delhi. Another effort was to contribute actively to the policy dialogue between e-IRG and the Indian e-Infrastructure management. The project maintained effective cooperation with other projects of GARUDA National Grid Initiative, NKN, EGI, OGF, TEIN3, and WLCG.

Fig. 3: India's cooperation with projects of GARUDA National Grid Initiative.

Three e-Infrastructures in India are managed by the project: GARUDA National Grid Initiative, Indian Component in WLCG, and the National Knowledge Network. There are two Tier-2 centers in India. In addition, the Internet connection speed reached 1 Gbps are also shown in the graph above. The GARUDA project consists of 36 participating institutions.

EUIndiaGrid2 supports several specific user communities. First, EUIndiaGrid2 utilizes NKN-TEIN3 interconnection for LHC data transfer from CERN. Then, it utilizes NKN-TEIN3-GEANT interconnection for crystallopgraphy experiments with the remote access and control of experimental facility (a beamline wthin the ESRF synchrotron) in Grenoble, France directly from Mumbai, India. EUIndiaGrid2 enables EUIndia Virtual Organization on with more than 7300 CPU-cores.   In fact, EUIndiaGrid2 delivers over 26 years of CPU hours in a year to user applications. Applications operated lie in a wide range of areas including biology, climate change, high energy physics, and material science.

EUIndiaGrid2 is committed to foster and enhance cooperation with other European Initiatives. EUIndiaGrid2 has already established effective synergies and cooperation agreements with CHAIN, EUMEDGRIP-Support, EGEE/EGI, EPIKH, GISELA, IGE, SEE-GRID-SCI, TEIN3, and OGF.

"[EUIndiaGrid is] an excellent example of EU-India collaboration underpinning all areas of science" according to Dr. Chidambaram, Principal Scientific Advisor to the Government of India and member of Government Cabinet. The timeline below shows the milestones of EUIndiaGrid.

Fig. 4: The timeline of EUIndiaGrid.
[Based on the presentation made by Alberto Masoni, INFN] [2]

Indonesia

The Grid Computing development in Indonesia started in 2006. InGrid or Indonesian Grid infrastructure is developed by the University of Indonesia using UCLA Grid Portal which consists of Globus Toolkit 4 Grid middleware and Gridsphere Grid portlet framework. In 2008, there were two clusters connected to InGrid: one is the production cluster from the Faculty of Computer Science, University of Indonesia, and the other is the research cluster.
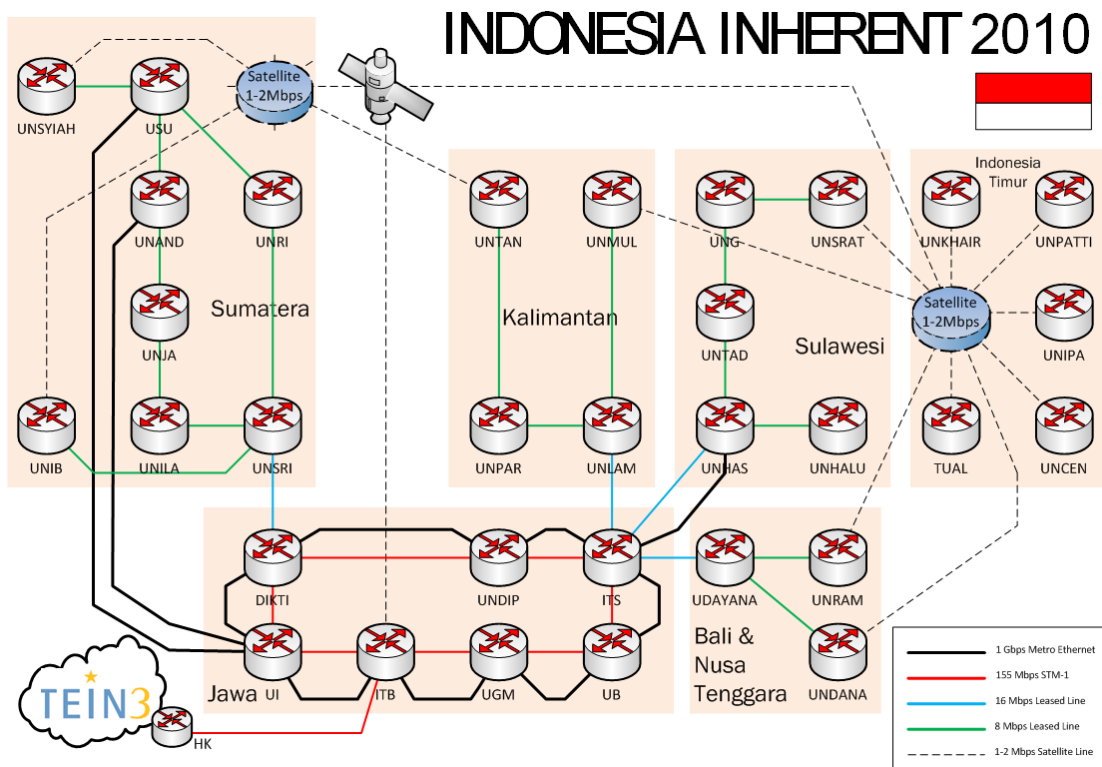

Fig. 5: InGrid's connection through the network.

The development continues when Institut Teknologi Bandung (ITB) joined EUAsiaGrid Project and connected Indonesia Grid to International Grid Infrastructure. ITB Grid site has a user Interface combined with MON-Box, DPM/Storage Element, Computing Element, and the Workload Management System. It would also be able to integrate existing Cluster Computing to ITB Grid Infrastructure. The local applications include: weather forecast, disaster mitigation, chemistry /bio-Informatics, and the future application plans such as volcano and digital library.

For weather forecast application, researchers used MODIS data from satellite image being downloaded from Kochi University in Japan to generate weather forecast information and display in a web page. Grid technology was utilized to support the

Numerical Weather Prediction (NWP) experiment which was performed by implementing WRF4G (WRF for Grid). Current application used for weather forecast is running on 8-node Cluster with Open Source Linux 10.3 as the Operating System. The challenges are the dependency and library issues, especially the conflicting library versions and the compiler. The weather forecast application is now being tested on the GPU cluster.

For disaster mitigation, the focus areas are the hazard map for the mitigation, especially earthquake and volcano, the vulnerabilities such as the impact factors and areas. Grid-enabled high performance computing is needed for the digitization, modeling and data visualization in this domain. Another project is the Firewatch, the collaboration between Government of Indonesia (e.g. Ministry of Forestry, Ministry of Environment, National Institute of Aeronautics and Space (LAPAN), etc.) and Australia. The objective is to achieve automatic real-time satellite processing, high speed access to satellite image products, to install communication related hardware, software, and to obtain the satellite image for digital mapping that covers the forest. It is supported by the Inherent network, and data is delivered from Australia to Indonesia via TEIN3 network with 155 Mbps.

The future plan of Grid application on volcano research aims to add volcano related information and existing databases to a thorough database thus can be used to produce a comprehensive hazard map. As for the Digital Library, a software called Ganesha Digital Library (GDL 3.0) dedicated to ITB's central library was developed at the end of 1999. The plan is to port GDL applications to Grid so that the Grid technology can be utilized in indexing, processing, and storage.

The goals of future dissemination plans are to extend Grid technology to other universities in Indonesia, promote the awareness as well as the interest in Grid infrastructure and application developments, and enable more people to access Grid resources available. There are dissemination activities such as conferences and trainings already planned.

[Based on the presentation made by Basuki Suhardiman, Institut Teknologi Bandung] [2]

Japan

The University of Tokyo

Tokyo Tier-2 Center was developed since 2002 as the ATLAS regional analysis center in Japan. Tokyo is one of the largest sites as a standalone Tier-2 site. It processes

2% of the ATLAS analysis jobs. There are 160 nodes (more than 1,280 cores) in LCG-CE and CREAM-CE(Computing Resource Execution And Management-Computing Element). This is actually a breakout for Tier-3 users because ATLAS Japan collaboration consists of 15 institutes and more than 100 users.
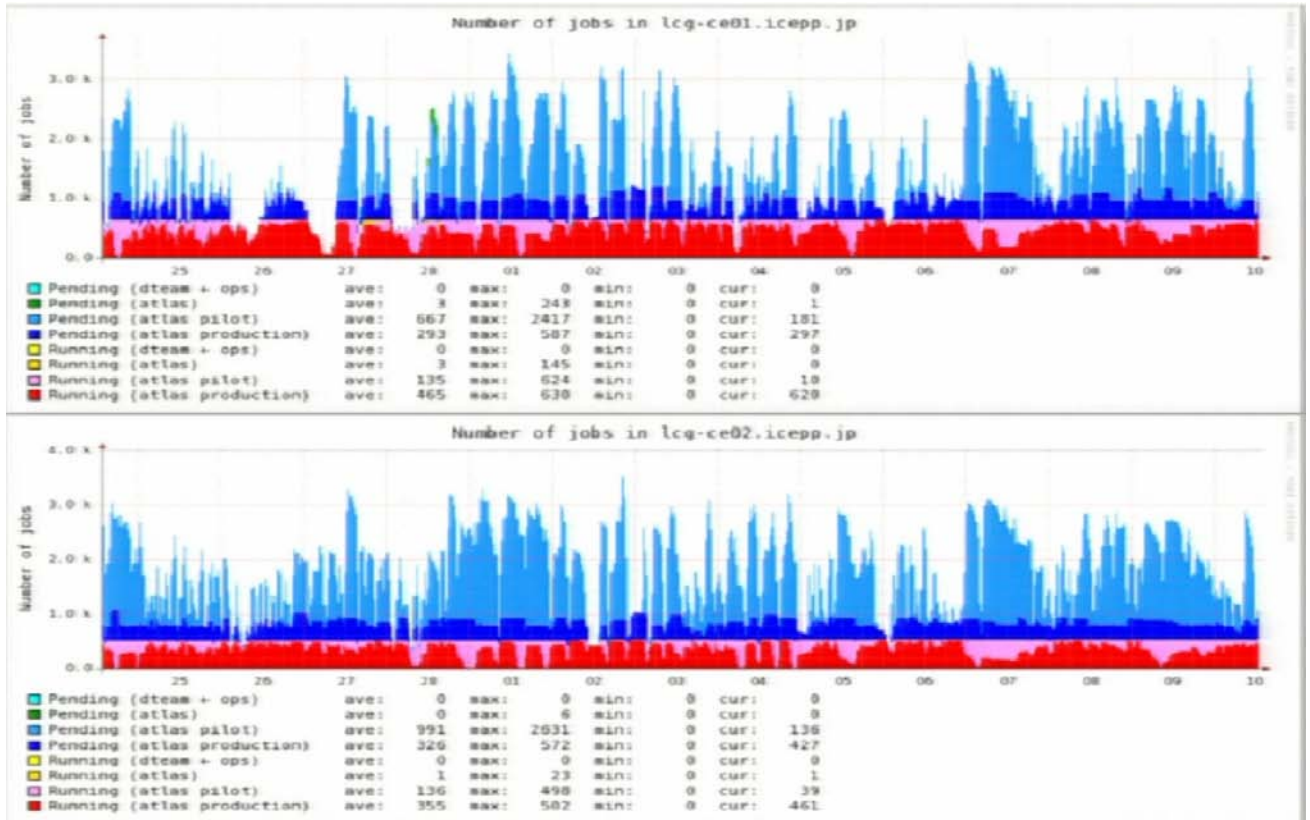

Fig. 6: Jobs processed in the LCG-CE and the CREAM-CE systems.

The graphs show the history of the jobs processed in the LCG-CE and the CREAM-CE systems. 640 job slots (cores) are assigned to the LCG-CE system, while 512 job slots are assigned to the CREAM-CE system. The red region represents jobs being processed, and the blue region represents jobs in queue. About 900 jobs are processed each day with 3000 jobs in the queue. Most of the nodes will be imported to the CREAM-CE system soon.

One of the most important aspects of high energy physics is data distribution, as high energy physics experiments produces large amounts of data. In original ATLAS computing model, data is distributed to Tier-2 priori to be analyzed by user, while Tier-1 to Tier-2 data transfers are restricted in a cloud. In the case for Tokyo Tier-2, data only comes from the IN2P3 computing center.
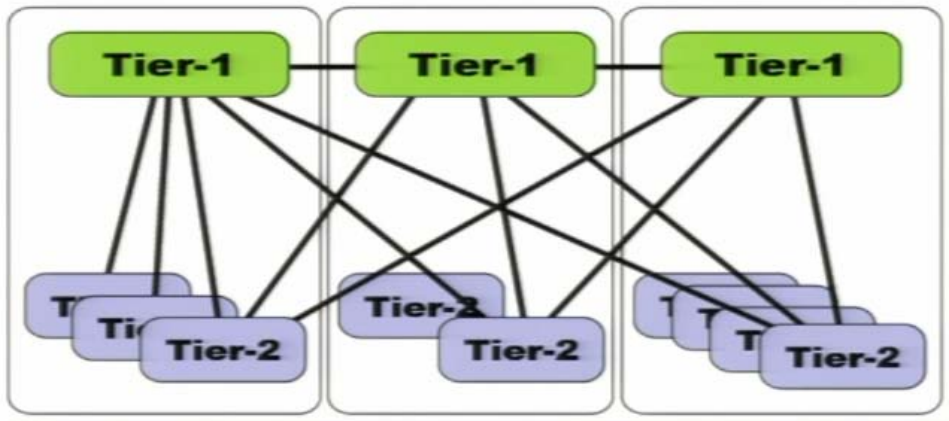
Fig. 7: Recent development in ATLAS data distribution and flow.

ATLAS has been developing dynamic data replication according to requests from user analysis jobs (PanDA Dynamic Data Placement, PD2P). By expanding this scheme, large Tier-2 sites can contribute to production jobs, even in the other cloud (Tier2D). This scheme is still under construction but will be up and running soon. Recently, data distribution in the ATLAS computing model gradually changed. Data will now be transferred over the cloud.
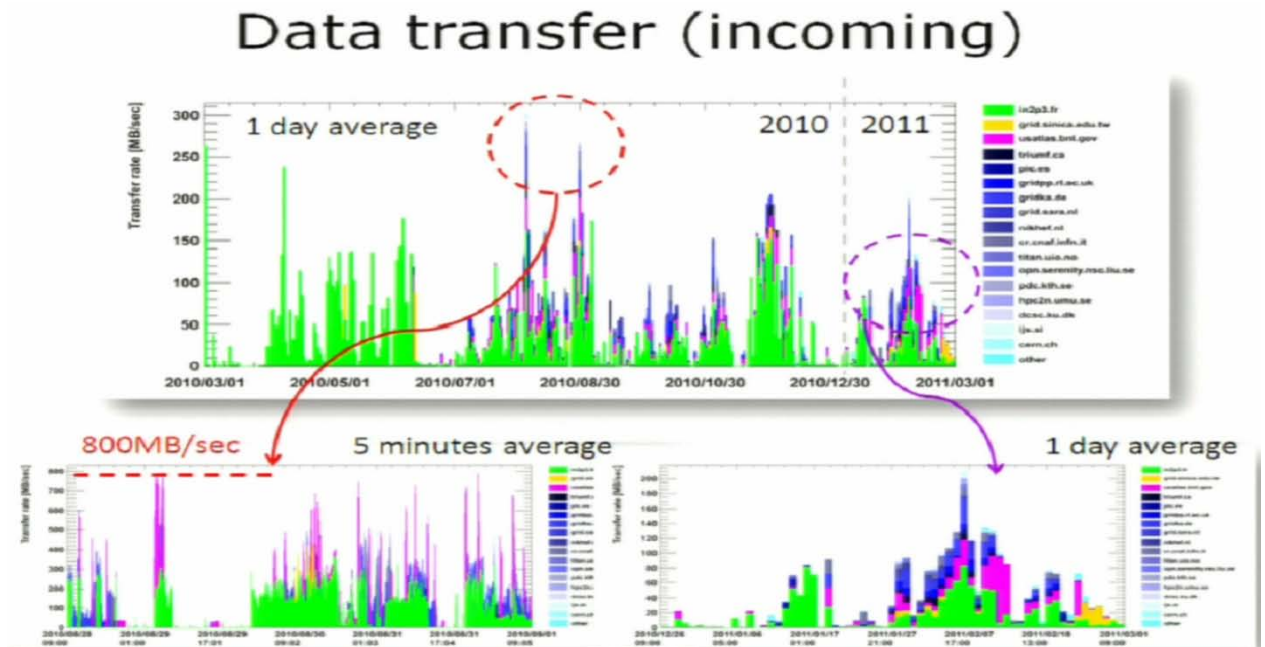

Fig. 8: Data transferred from the Lyon site.

Internet speed of 800MB/sec was reached. And as shown in graph, during 2010, most data were transferred from the Lyon site. However, going into 2011, data transfers from other Tier-1 sites increased visibly. This tendency will be expanded in the near future. A The 98% availability and reliability was maintained in one year operation of

the 2nd (new) system, except for October 2010, due to the annual power outage from October 22 to October 25.

Tokyo Tier-2 has upgrade plans to import all work nodes to CREAM-CE soon, to promote effective use and enhancement of the ATLASLOCALGROUPDISK, and to apply CVMFS for the ATLAS software release.

As for the local system, LTO5 drives and tapes will be installed to the tape robot in March 2011.   There is an ongoing construction of CASTOR based archiving system, and the test for the large scale GPFS operation has started. It will be used for the user's disk storage instead of the NFS.

The role of Tier-2 has gradually changed with ATLAS computing model. Within this model, more steady and reliable bidirectional network (for the incoming data and outgoing data) will be important from now on.

RENKEI

Resources Linkage for e-Science (RENKEI, federation in Japanese) is consisted of five separate sub-projects. Each sub-project has already had their prototypes developed and is in the testing and development phase. The file sharing part of sub-project 2 already reached production level quality, as a function of Gfarm 2. The goal of the project is to develop middleware for the resource federation.

Users preprocess in their local system, then run their main process in large computer centers with grid resources, lastly finish their post process back in their local system. This workflow requires many interactive operations at each phase of the processes, which decreases the efficiency as a whole. The RENKEI project is deemed to develop a solution to this problem. The graph below shows how the RENKEI services will allow seamless access between local and grid resources for jobs to be processed.
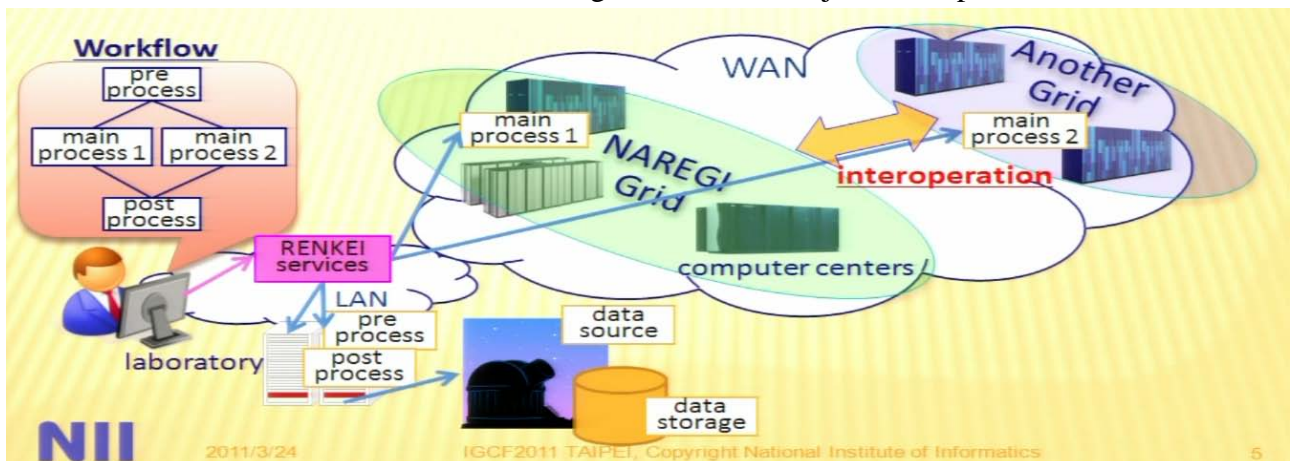
Fig. 9: RENKEI services.

RENKEI is to create a federation of grid and non-grid resources. The workflow engine of the RENKEI portal will assign computing jobs to different resources available in the federation according to the need of jobs. Supporting the interoperation between different grids is also one of the goals of the RENKEI project. Demonstrations are done in several OGF conferences.

Another activity that NII is involved in is the Edubase Cloud project, promoting cloud infrastructure for education. With the RENKEI client VM image prepared, the Edubase project will allow educational programs to utilize supercomputing resources. Plans of Fiscal Year 2011 for NII, including MPI support, integration with other RENKEI and validation test, and the interoperation test with more other grids.

[Based on presentations made by Tomoaki NAKAMURA, University of Tokyo, Kazushige Saga, National Institute of Informatics] [2]

Malaysia

The Malaysia Research and Education Network (MYREN) provides high-capacity broadband to Universities, government ministries and research institutes throughout Malaysia. This includes community colleges and poly-technical institutes. Currently, 64 institutions and organizations within Malaysia are connected and plans are underway to expand this network.

MYREN has been connected to the outside world since January of 2009 through TEIN3 in Singapore. TEIN3 allows researchers in Malaysia to link with research networks in 33 European countries, 10 Asia-Pacific countries, and recently with Internet2 which will allow MYREN-based researchers to collaborate with their colleagues in the US and Canada. This is all part of the MYREN Phase 2 initiative which was launched on 29 April 2010 with the mission of creating a gateway for global research collaboration for over 60 million scientists, academics and students.
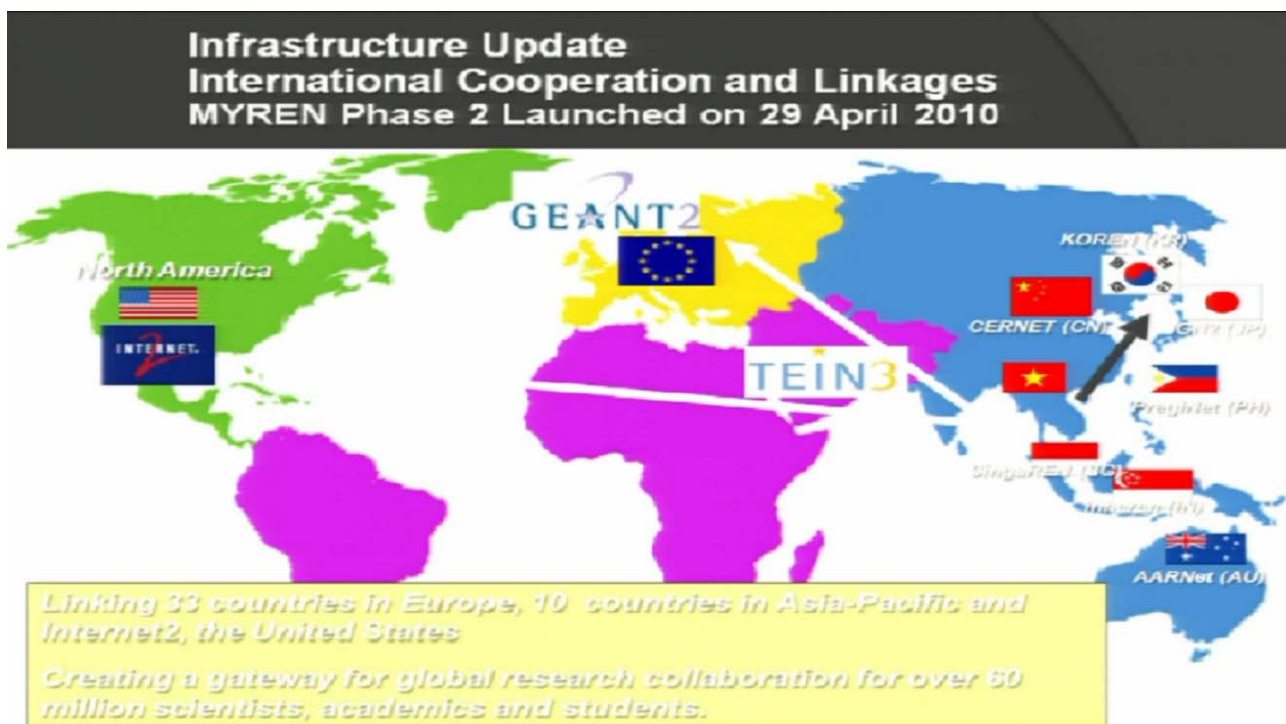
Fig. 10: International cooperation and linkages in Malaysia.

The Malaysian research community uses MYREN to conduct research into such fields as: Bioinformatics, Cheminformatics, Precision Farming, Natural Resources Informatics, High Energy Physics and Nuclear Simulations, and Engineering. MYREN provides an on-demand infrastructure for research and education collaboration technologies (OnDIRECT) which allows Malaysian researchers to work together. OnDIRECT includes video collaboration services and on-demand rich media for learning. These services allow for hassle-free individual or group virtual conferencing, connecting up to 158 concurrent sites (60 in full HD). The Video On-Demand service has been useful for sharing recordings of lectures and conferences.

More importantly, OnDIRECT provides data services for transferring, hosting, and storing of collaborative research data. It also provides a centralized provisioning of development servers so that researchers will have no need to separately buy small servers. As part of this effort, Malaysia seeks to build its own certification authority; the Malaysia Identity Federation and Access Management (MyIFAM). In the future, it is planned to increase the amount of high-performance computing on-demand over private cloud.

BIRUNI Grid: HPC in Private Cloud

UPM has commissioned BIRUNI Grid with the hope of making it the Nucleus HPC Cluster for A-Grid. Hardware racking and Power Up were done by suppliers while the rest of the configuration and deployment were done by the UPM Grid team. This has provided excellent learning opportunities for the staff. UPM and Biruni Grid were aided by their international partners, especially by Academia Sinica. EUSI has also proven instrumental in the development of BIRUNI Grid providing about 400 core. It is the experience gained through collaboration that has proved the most valuable asset.

Developments are currently underway for grid applications for Bioinformatics, Cheminformatics, and Medical Informatics (radiology). Examples include: BLAST, MUSCLE, ProDist (sequence analysis), AUTODOCK (molecular docking), GROMACS (molecular dynamics), Mr Bayes, BEAST (Bayesian Phylogeny), GATE (Grid Analysis Time-series Expression). Ultimately, MYREN will move from a computing intensive analysis to RAM intensive analysis, providing 7.5TB raw data per run and 200GB of clean sequence data. This will allow for next generation sequencing.

As part of MYREN Phase2 a cloud computing service will soon be added to the network known as MYRENCLOUD. This will allow for video collaboration, data disaster recovery, rapid server procurement (up to 1000 virtual servers with 2XQuad, 8GB RAM, 400GB hard-drive), and virtual storage (estimated 100GB for 10,000 users). Ultimately, the hope is to extend MYREN to libraries, hospitals, community centers, and even more schools.

Malaysia seeks to make MYREN a member of the APGridPMA through the creation of a domestic certification authority; MyIFAM. This will allow for greater resource and infrastructure sharing and move MYREN towards SAML-based federation like Shibboleth, Athens.

[Based on the presentation made by Suhaimi Napis, University Putra Malaysia] [2]

Pakistan

NCP-LCG2 is the only operational Tier-2 grid node in Pakistan and has been a partner of WLCG project since 2003. In 2009 another Pakistani grid site, PAKGRID-LCG2, was merged with NCP-LCG2. This will allow for greater cooperation with the ALICE and CMS experiment. Major data analysis will be performed by NCP. To do this NCP will need to secure access to data stored by other WLCG partners and sufficient network bandwidth and services with other Tier-1 sites. Based on collaboration with CMS experiment, PhEDEX (Physics Experiment Data Export) data transfer utility has been deployed to fetch the LHC data for the purpose of physics

analysis for the Experimental High Energy Physicists in Pakistan. Until now there are only three commissioned links with different Tier-1s, whereas testing and commissioning of further links with other Tier-1s is in progress. In the last year, NCP has ramped up to 524 CPU cores and 70 TB of Disk storage capacity connected with 155 Mbps of WAN link.

NCP has a total HEP-SPEC06 of 6365 and a KSI2K equal to 1591. A further breakdown of resources is listed below:
1. 28 servers with SunFire X4150, 2X Quad Intel Xeon CPU X5460 running at 3.16GHz, and 16GB RAM. This includes 56 physical CPUs and 224 Logical Cores.
2. 25 servers with Dell PowerEdge R610, 2X Hex core processor Intel Xeon X5670 CPU running at 2.93GHz and 24GB RAM. This includes 50 Physical CPUs and 300 Logical Cores.
3. 10 Transtec NAS4324M-A, Intel Xeon E5520 -2x2.26GHz, 12GB RAM, 24 SATA Drives providing 23TB of RAW storage.
4. Total online storage of 69.8TB

According to WLCG C-RRB, in 2011 the cumulative requirements of the 33 Tier-2 CMS sites will be as follows: 319500HEP-SPEC06 of computational power, 19900TB of disk storage, a minimum bandwidth between Tier-1 and Tier-2 sites of 1 Gbps. For this purpose NCP has recently purchased 110TB of additional disk space to be ready by April 2011.

A CMS environment was set up, and there has been reliable support for the CMS product and analysis jobs. PhEDEX has been deployed since February of 2009 with an initial network connection of 10Mbps.

VOBOX for ALICE has been deployed and will serve as the entry into the WLCG environment. In order to support ALICE VO, a separate VOBOX and dedicated storage Xrootd of 6TB has been set up. Also, offline Alice Environment (AliEn) is being deployed. This will provide an analysis platform for event generation, event merging, particle transport, generation of digits, reconstruction, particle identification, and generation of event summary data. Necessary tools for ALICE offline Environment like Pythia, GEANT-3, GEANT-4, FLUKA and AliRoot have been deployed.

Unfortunately, many challenges still lie ahead for NCP-LCG2. Network connectivity is still an issue. The network is still plagued by disruptions at the service provider level. The Higher Education Commission of Pakistan has aided NCP by providing 155Mbps at no cost. Electric power supply is yet another pressing issue.

Sever power shortages and extensive load-shedding have harmfully impacted the battery back-up system.

[Based on the presentation made by Fawad Saeed, National Centre for Physics] [2]

Philippines

Established in 1987, the Advenced Science and Technology Institute (ASTI) is the R&D agency of the Philippine Government working on *Information and communication technologies (*ICT) and Microelectronics. It operates and manages the National Research Network and National Grid Infrastructure.

The Philippine e-Science Grid (PSciGrid) Program is fully funded by the Department of Science and Technology which will end in June 2011, but ASTI is trying to continue the program by introducing phase 2. The program objectives are to establish a national e-Science grid infrastructure in the Philippines that will enable collaborative research among local educational and research institutions and to provide seamless access to high-performance computing resources and applications for Life Science and Physical Sciences. Life Science and Physical Sciences are major focus fields because the whole country is declared a mega biodiversity hotspot, and there are high level of typhoons, earthquakes and volcanic activities. There are 2 projects under the PSciGrid Program: boosting Grid computing using reconfigurable hardware technology and boosting social and technological capabilities for bioinformatics research.
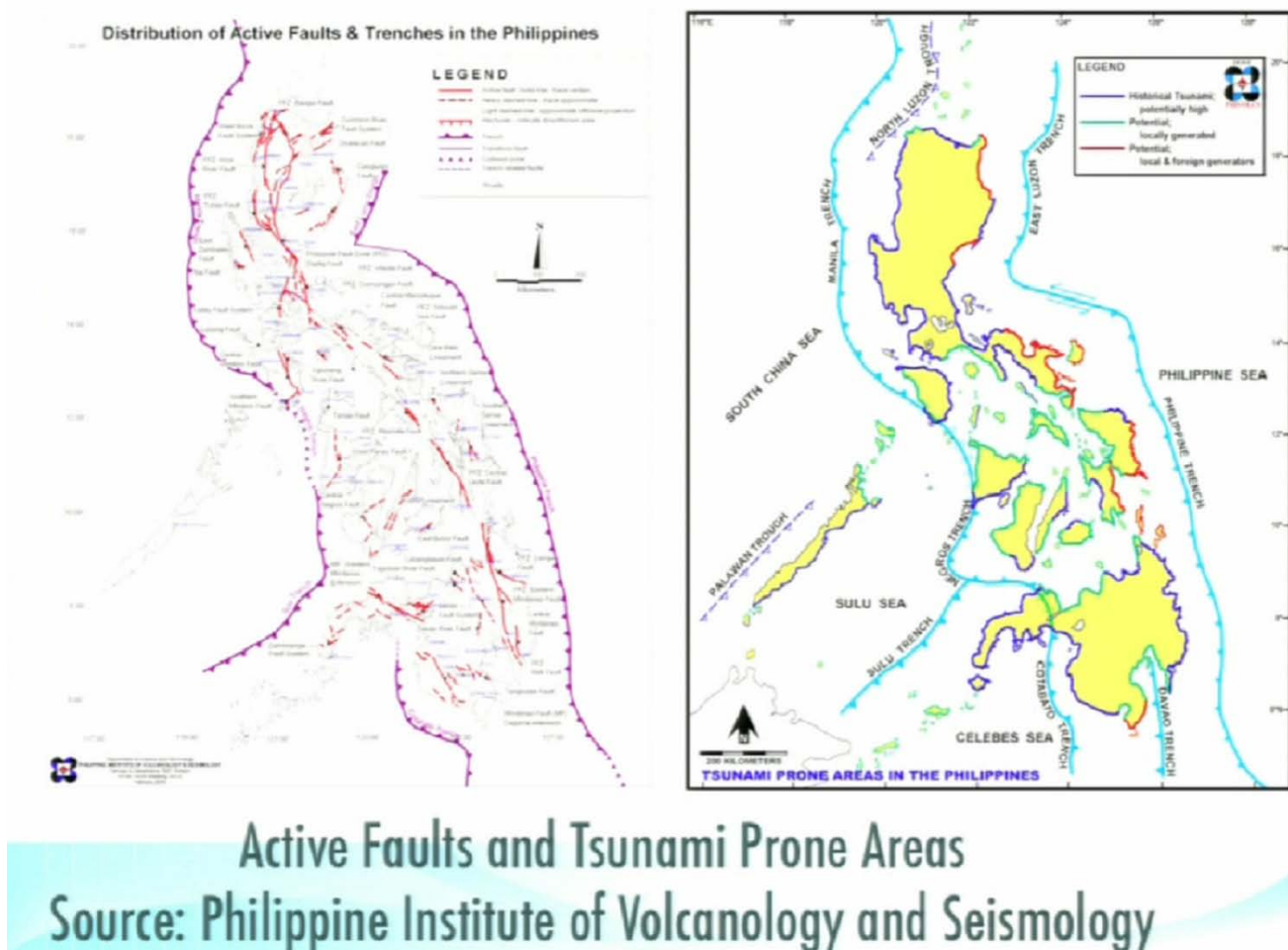
Fig. 11: Active Faults and Tsunami Prone Areas in the Philippines.

ASTI's high performance computing (HPC) facility has 408 cores, 51 computing nodes with 300GB/500 GB of disk space and 16GB/24GB of RAM per node in computing capacity and 30 TB in storage capacity. There are 6 computing clusters using gLite middleware and ROCKS 5.2.2 for the operation system: the bioinformatics cluster (8 nodes), the meteorology cluster (6 nodes), the cluster sandbox (2 nodes), the EUAsiaGrid and EGEE collaboration cluster (EGEE certified production cluster) (7 nodes), virtual cluster 1 (16 nodes), and virtual cluster 2 (13 nodes). The installed applications include Bioinformatics (BioRoll, Progeniq BioBoost, AutoDock), Seismology (SPECFEM3D), Meteorology (WRF , RegCM, MM5), and Oceanography (SeaDAS).

There has been several trainings, including the SPECFEM 3D in August 2010, which is a workshop in collaboration with ASGC for senior programmers on in-depth usage of SPECFEM 3D application for seismic wave propagation simulation. There were also annual Philippine Grid Computing Forum every December since the program

started in 2008, training on UPLB Biotech Data Warehouse Portal in February 2011 and a one-day talk in the Philippine Genome Center.

The current users include: UP Marine Science Institute, UPLB Biotech, International Rice Research Institute, PAGASA (Weather Bureau), UP National Institute of Physics, UP Computer Science Department, PHIVOLCS (Volcanology and Seismology). There are also upcoming users such as Manila Observatory, Philippine Genome Center, and the Energy Development Corporation.

The current and future works of the ASTI are to develop other FPGA applications, to expand virtualization, to procure blade servers, to integrate PsciGrid with Philippine government data center, to establish IPv6-enabled clusters, to propose phase 2 of the Program, and to continue to advocate and promote grid technology to local universities and research communities.

Grid Computing for Biodiversity Conservation

Because of deforestation and other forms of environmental degradation, the Philippines is now considered a biodiversity hotspot. The use of Information Technology to organize, manage and analyze biodiversity data from various collections and experiments, to allow a more coordinated and systematic approach to the sharing of biodiversity information. Data Grid technologies will be applied to biodiversity informatics in Philippines – in particular, in allowing a systematic approach to the process of discovering, accessing, analyzing and integrating huge amount of biodiversity data.

A Grid Computing Application: Bird Biodiversity Conservation

The research presented centers on building an image database (photos and videos) of resident and migratory birds in Ateneo de Manila University and tapping existing Grid infrastructures such as the EUAsiaGrid and the Philippine e-Science Grid. It is founded by Global Biodiversity Information Facility (GBIF) enables free and open access to biodiversity data online. Its mission is to make the world's biodiversity data freely and universally available via the Internet. As a megascience initiative, GBIF aims to provide an essential global informatics infrastructure for biodiversity research and applications worldwide. In the Ateneo de Manila University, there is also a research team called Ateneo Bird Ecology Study Group (ABES). It is an interdisciplinary group of researchers (biology, environmental science, mathematics, engineering, physics, etc.) dedicated to the Biodiversity Informatics.
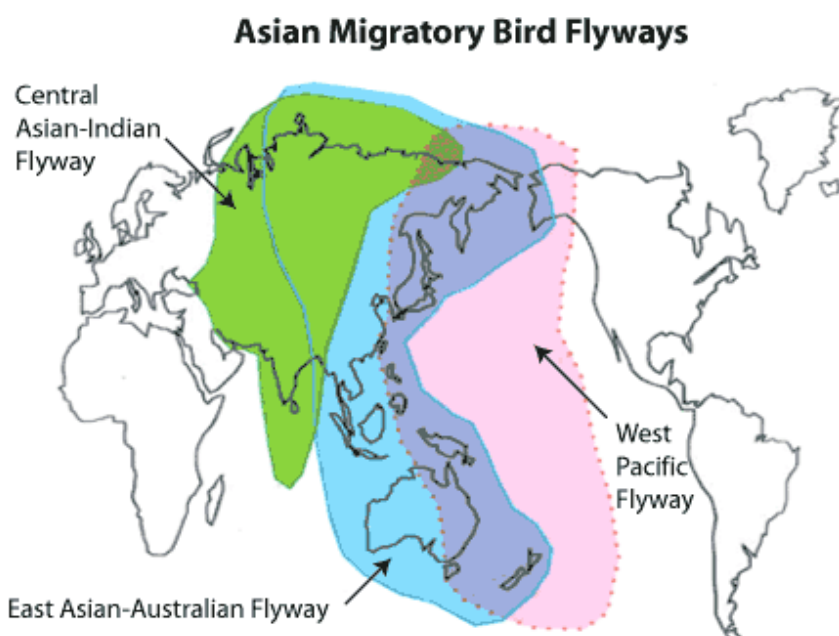
Fig. 12: Asian migratory bird flyways.

The East Asian-Australian Flyway is one of the world's great flyways. In total, the flyway passes through 22 countries with approximately 55 migratory species travelling along it, adds up to about 5 million birds. Launched in November 2006, the Partnership is an informal and voluntary initiative, aimed at protecting migratory water-birds, their habitat and the livelihoods of people dependent upon them.

There are currently 25 partners including 13 countries, 3 intergovernmental agencies and 9 international non government organizations. The Partnership provides a framework for international cooperation, including: development of a Flyway Site Network, collaborative activities to increase knowledge and raise awareness of migratory water-birds along the flyway, building capacity for the sustainable management and conservation of migratory water-bird habitat along the flyway.

[Based on the presentation made by Peter Banzon, Advanced Science & Technology Institute, and Rafael P. Saldaña, Ateneo de Manila University] [2]

Singapore

Under direction of the Singapore government, the Infocomm Development Authority (IDA) plans to establish an Infocomm Resource Marketplace (IRM) that would allow businesses and companies to share, buy and sell infocomm resources such as software and computing and storage resources on an "on demand" basis. Flexibility

would enable new business delivery models and encourage innovation of infocomm services with a business and economic focus.

As part of the Master Plan, the government has instituted a 3-year program and an Open Call for Cloud Computing Proposals to help organizations to undertake trials & POCs in enterprise cloud adoption, before proceeding to production cloud. Much of the program centers on business rather than academic uses of cloud computing. To encourage further adoption of the technology amongst the business community, the government seeks to lower costs and technology barriers to adoption as well as provides business and technology guidance and consultancy services. The government also seeks to create training programs with the goal of training some 600 ICT professionals. Rather, the government seeks to lower the barriers to entry so that companies will see the value-added benefits of cloud computing and put forward their own investments.

Several major corporations are on board with the Singapore government's pro-business cloud computing plans. HP Labs opened at Fusionopolis in Feb 2010 which is an investment of S$50M over 5 years.

iSPRINT

iSPRINT (Increase SME Productivity with Infocomm Adoption & Transformation) is designed to promote the use of IT to improve business operations, resulting in efficiency, productivity, increased revenue or value-added for businesses. However, it does exclude basic IT services (e.g. email, backup, web development). It will support packaged solutions pre-qualified by IDA and customized solutions. SaaS subscription is included as a packaged solution. Capped for 12 months, based on 50% co-payment, SaaS vendors have to be on pre-approved list.

Grid & Cloud Applications in Academics

EUAsiaGrid Partners in Singapore include: IDA, National University of Singapore, Association of Medical and Bioinfomatics in Singapore, and Asia-Pacific Bioinfomatics Network.

NUS runs a grid computing workshop on behalf of EUAsiaGrid as part of their objective of dissemination of knowledge and technology in the Asia-Pacific region. NUS also provides application support for the local life-sciences community. A pilot interface was set up in Singapore with help from Biruni Web and ASGC Grid Portal. The goal is to make a more user-friendly interface which is more easily customized and has more features and software such as: BEAST, BLAST (with formatdb), Gromacs, MrBayes, MUSCLE, ProtDist, and WRF.

The Ministry of Education (MoE) in Singapore is the first in Asia to provide Web 2.0 communication and collaboration tools to 30,000 teachers and staff in over 350 schools. The goal is to adopt innovative technologies within the education profession to drive next-generation learning and communications. Google Apps will replace existing e-mail system and might replace existing word processing software in future.

APBioNet

The Asia Pacific Bioinformatics Network (APBioNet) has been operating since 1998. APBioNet has partnered with the International Society for Computational Biology, Asia (ISCB Asia) to host the tenth annual International Conference on Computational Biology in 2011. What the two partners hope to achieve through this conference is an integrated method to prevent the disappearance of databases on the network. Therefore APBioNet has introduced a new compliance standard; Minimum Information about Bioinfomatics Investigation (MIABi). The plan is to require authors who publish on the network to have unique identifiers (AuthorID) for disambiguation and accountability purposes.
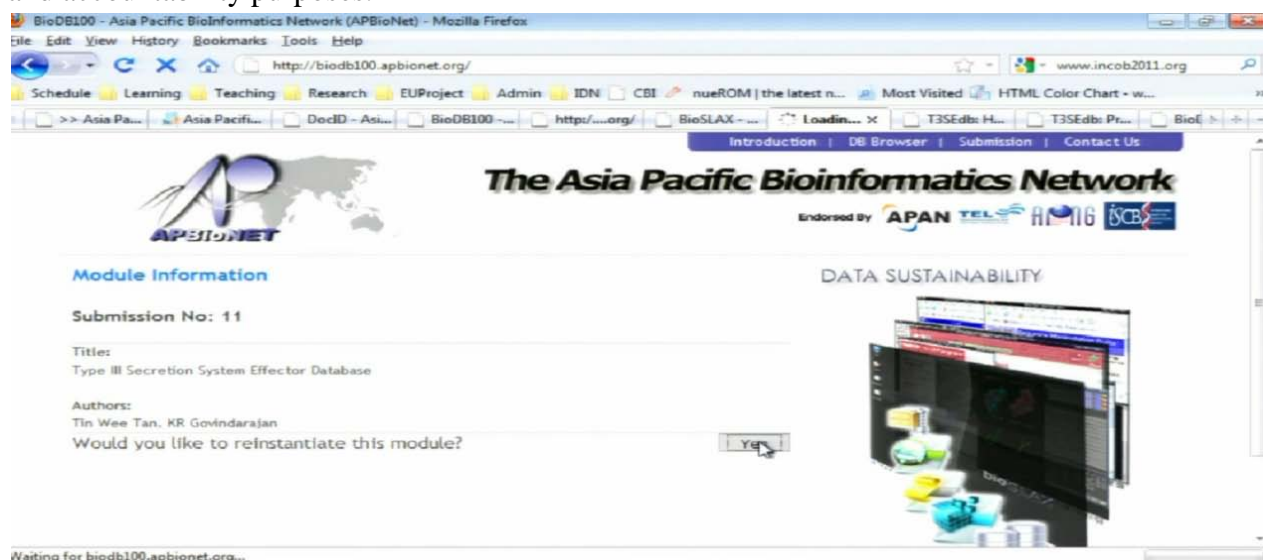

Fig. 13: A demonstration of the APBioNet system.

Other requirements include the deposition of bioinformatics databases in a reinstantiable form. This will promote database longevity and prevent data loss. The MIABi will require every author who publishes papers with regard to a database, to save an image in the archives which can be reinstantiated. To facilitate this process a cloud-based reinstantiation system has been developed. The system is based on the BioSlax Live operating system. The system was designed for full and instant access to unreachable or published but discontinued databases from archives.

[Based on the presentation made by Tan Tin Wee, NUS] [2]

South Korea

KISTI signed WLCG MoU in 2007 and began serving as the ALICE Tier-2 center. KISTI has been part of ALICE distributed computing grid as an official Tier-2, providing a reliable and stable node in the ALICE Grid, funded by Ministry of Education, Science and Technology (MEST). Right now there are 2 people working full-time for the project.

KISTI is contributing 1.2% to ALICE's computing in the total job execution with 120 CPU cores and 50 TB storage dedicated. KISTI processes near 8000 jobs per month in average. KISTI has been delivering 161% of what is expected from KISTI on average over the last six month.
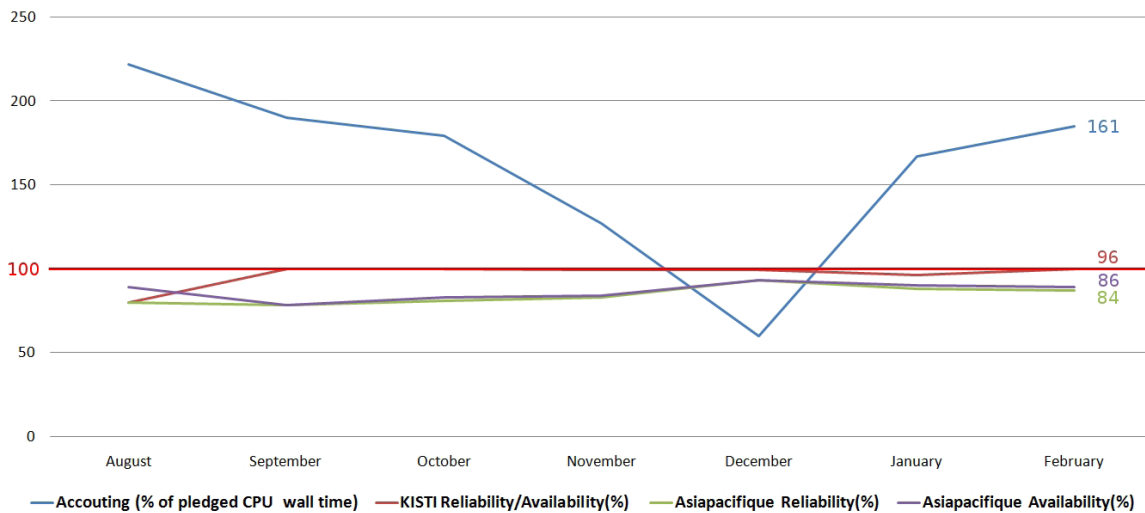

Fig. 14: Accounting and availability of KISTI.

FKPPL VO is a production Grid infrastructure based on gLite. Its objective is to foster the adoption of grid technology and provide researchers in Korea and France with production Grid Infrastructure. It has been up and running since October 2009, providing about 10, 000 CPU cores and 30 TB of Disk Storage. Last December, KEK has joined FKPPL VO, contributing around 1600 CPU cores and 27 TB of disks. FKPPL is also under discussion of moving towards the construction of a France-Asia VO. As of now, around 70 users have joined the FKPPL VO membership.

It is very important to have applications porting support on FKPPL VO. As of right now, KISTI has deployed two applications of Geant4 and two-color Quantum ChromoDynamics (QCD) simulations in theoretical physics. Last year, during 2010, KISTI held two tutorial sessions for users, one in February and one in July.

KISTI Global Science Data Centric (GSDC) Center project launched last year. Right now, KISTI is hosting a Tier-2 center in cooperation with CERN. Partnership and Leadership for Nationwide Supercomputing Infrastructure (PLSI) is a consortium of 14 HPXC Computing Centers in Korea. It distributed HPC computing environments for world-class computational science research. This project was launched in 2007 and is funded internally by KISTI. The target is to reach 400 Tflops by linking 140 HPC centers around Korea. Currently, PLSI have established around 80 Tflops computing capacity by combining 15 computing resources at 10 partner sites over dedicated high-performance networks.
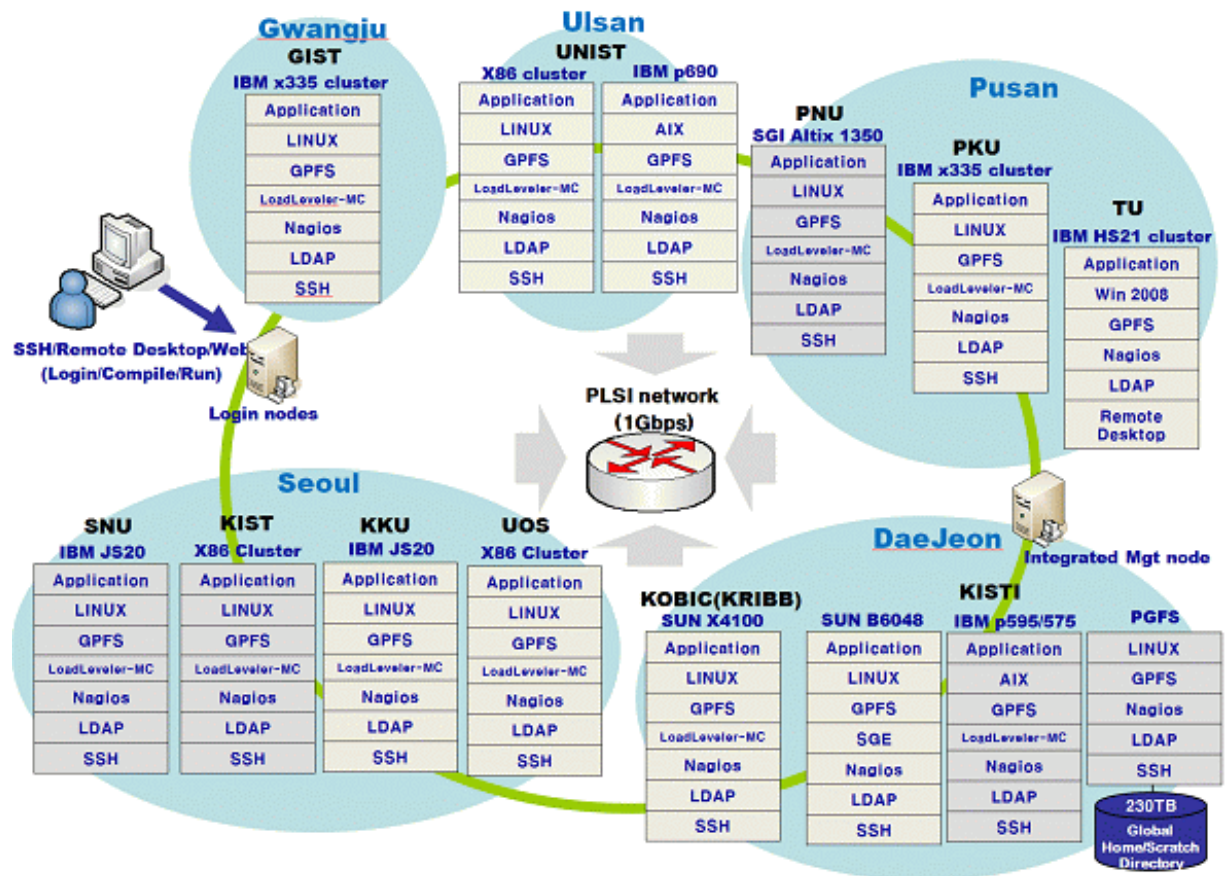

Fig. 15: HPC infrastructure in South Korea.

PLSI user portal is built for convenience for user. A MGrid Portal is also built to support molecular dynamics simulation. AMGA is an official gLite middleware component of metadata catalogue service. AMGA provides access to metadata for files distributed on the grid. KISTI has tken over the leadership of AMGA development since July 2009, as one of the product teams of EMI by contributing to the evolution and maintenance of AMGA.

Likewise in EGEE, KISTI is an official partner of European Grid Projects, EGI and EMI, with its continuing contribution to production-quality Grid operation and the

AMGA development respectively. KISTI is in the middle of moving one step further towards setting up the France-Asia VO starting from FKPPL VO. With KISTI GSDC, KISTI is expected to play a role of a Tier-1 data center as well as its traditional supercomputer center.

[Based on the presentation made by Soonwook Hwang, KISTI] [2]

Taiwan

Academia Sinica Grid Computing (ASGC) is coordinating e-science infrastructure and application collaborations in Taiwan. The e-Science applications mainly focus on Life Sciences, e-Social Sciences, Environmental and Earth Sciences, and High Energy Physics (HEP). In addition to the e-Science application, ASGC also works on technology development including building new generation of Distributed Computing Infrastructure (DCI) and Operation technologies, application-driven technologies as well as service-oriented Computing & Clouds.

International network connectivity is one of the key components of e-Science infrastructure. ASGC provides Taiwan's global network connectivity at speeds of 2.5 Gbps to Japan and US, 5Gbps to Hong Kong, 10Gbps to Netherlands and GEANT. In 2010, there are more than 6 Petabytes of data in and out of ASGC. In addition, ASGC, transferring the data at the peak of 11Gbps/per second, has set a new record in Taiwan. The international bandwidth will continue to increase later this year and next year.

The ASGC currently supports 38 grid computing sites in 15 countries and provides services to over 1800 users. ASGC provides over 15,000 CPU cores with 5PB disk and 4PB tape to the e-Science infrastructure in Asia. In recent years this e-Infrastructure has been handling some 62,000 jobs and 80,000 CPU hours a day which is ten times the number just three years ago. Among the 15,000 CPU cores, one-third of resources is dedicated to the WLCG, one-third to Life Science/Earth Science/EUAsiaGrid and the rest to Clouds and HPC.

To improve the quality of services, ASGC has recently worked on system optimization. Its goals are to reduce cost and energy usage as well as implement further system automation. Take data management and data storage system as an example, ASGC deployed a high density disk array (96 x 2 TB) with large bandwidth, in turn, it helped reduce the total number of storage servers and allowed for a 10Gb Ethernet and 8Gb fiber interface connection.

To introduce elastic resource management, ASGC deployed the Virtual Machine Management Framework to support data storage and virtualization. Via vNode which is developed by CERN, ASGC provided virtual grid services, such as work nodes on demand, customized interface for users, different applications, and storage services. Users are able to selecte their preferred value of each parameters within 10 seconds and receive the virtual environment they need. ASGC is currently working on benchmarking for VM at different scales, VM live migration through global file system, integration between the VM management system and the current grid infrastructure of gLite Work Node on-demand, and applying the VO-based resource policy.

e-Science Applications

Joining the global projects such as WLCG, EGEE, EUAsiaGrid, EGI-InSPIRE, EMI and DEGISCO, ASGC has been providing services, coordination and support to e-Science applications in Asia since 2005. The application data, with unique scientific values, are actually the most essential. Therefore, ASGC not only ported scientific applications to e-Science collaborations, but also established research-oriented production services and long-term scientific collaboration among Asia partners.

EUAsiaGrid has been the major e-Science project in Asia-Pacific region in the past three years. This cooperative relationship includes four European institutions and 11 partnersin Asia. ASGC has served a key role in this partnership; providing training resources, supporting e-science communities, and helping to build the grid infrastructure. The table below describes various applications and projects that have been created and sustained through this partnership.

Discipline	Applications	Partners
HEP	ATLAS, CMS, ALICE, BELLE, CDF, GEANT4	TH, TW, CESNET, INFN
Biomedical	virtual screening for drug discovery-Avian Flu, Dengue Fever. Pandemic disease analysis	MY, TW, VN, CESNET, INFN, FR
Bioinformatics	Grid enabling phylogenetic inference, SVN parameter optimization for prediction of caspases, Genome search to ID T3SS effect, autodock ligand-receptor docking, complex disease studies	SG, TW, VN, CESNET, INFN
Earth Sciences	Disaster mitigation for earthquakes	ID, MY, PH, TH, VN, TW, CESNET, INFN
Computational	Chemical compound property	TH, TW, CESNET

Chemistry	analysis	
Climate Change	Weather simulations, sea level rising	ID, TH, PH, VN, TW
Social Sciences	Social simulations	TW, UK

Fig. 16: e-Science Collaborations in Asia.

Continued the e-Science collaboration in EUAsiaGrid, ASGC coordinated the Asia Pacific Grid Initiative in the EGI-InSPIRE project.   The major e-Science applications are listed as follows:

Life Science: Since 2007, ASGC initiated a series of large scale in-silico screening on the Avian Flu virus by taking into account the possible N1 mutations and a set of 300,000 compounds. The purposes of the drug discovery are to analyze the efficiency of the known drugs to the possible mutations as well as provide information for biochemists in case of large-scale pandemic happens. Over the past few years, ASGC has conducted virtual screening on Avian Flu, Dengue Fever, etc.   Started 2010, ASGC worked on Compound Profiling to extend compound 3-D structure database and compound information characteristics. By using the Autodock 4.2, ASGC generated all possible docking poses and evaluated the interactions. Then the docking poses could be arranged in sequence based on its docking energy. By doing so, we are creating 3-D structure conformers of known compounds for *in-silico* screening and post structure-activity relation analysis.

Earthquake Disaster Mitigation: Earthquake disaster mitigation is another successful e-Science application in Asia. Two major focuses in the earthquake disaster mitigation: one is Seismic Wave Simulation and the other is Shake Movies. For Seismic Wave Simulation, ASGC integrated the seismic sensor networks, earthquake data center and the seismological simulations together to support research and also the high quality forward simulations on earthquake risk analysis.   Via e-Science infrastructure researchers from anywhere with an internet connection access could conduct high-performance seismic wave simulations. For Shake Movies, with pre-computed SGT database (the known velocity structure is also a must), researchers could have quite accurate seismogram at any point to estimate the possible impact to any target location, from hours to seconds. ASGC has established a database with pre-computed 3D SGT for Taiwan (100TB+) so that once a CMT is determined a program will be triggered to retrieve all the Green's functions needed to generate the synthetics to make shake movie and shake maps. The SGT database will have around ~7000 grid point (~64x110 covering all Taiwan area by 25km distance for adjacent two points). Each grid point will have about 15GB SGT data and is independent of any others. All waveforms and shake movies will be available within minutes after earthquake alert. Further improvements includes the I/O performance, higher frequency and resolution with smaller grid size, and improve the accuracy by further understanding of the near-surface site condition and topography etc.

High Energy Physics: CERN's Large Hadron Collider had accelerated its twin beams of protons to an energy of 7 TeV on 30 March 2010. Since then, for ATLAS the inbound transfer data volume to ASGC reached 2,333 TB and outbound transfer volume

from ASGC to other T1s/T2s reached 900 TB. For CMS, the inbound transfer data size to ASGC reached 1,895 TB and outbound transfer rate from ASGC to other T1s and T2s reached 1,704 TB. Taiwan Tier-1 and Tier-2 Centres finished 6.67M jobs in the first EGI years, and provided 90.2M normalised CPU time (HEPSPEC06.Hours) for ATLAS and CMS experiments.

e-Social Science: Population Simulation is the major focus in the e-Siocial Science study now.   SimTaiwan is an initiative project of the collaboration research between NCeSS of UK, CSR and ASGC of Taiwan, Academia Sinica, TW for social simulation. This project aims to create a social model with agent-based modelling techniques and Repast Simphony toolkit. SimTaiwan develops a social simulation model based on the population and household census data of Taiwan. The social model will start from population reconstruction and the first substantive research question will be migration. ASGC has built several evaluations and tests of the social simulation model, including stability analysis, memory usages of different number of seeds for the population initialization, runtime measurements of different number of threads and partitions. The complete Taiwan Migration Model includes three sub-models. The Preferred Departure Model will be run first. The full dataset, 20 million individual data, of Taiwan Census dataset has been run since December 2010.

[Based on the presentation made by Eric Yen, Academia Sinica Grid Computing] [2]

Thailand

The Thai National Grid Project is a four-year project concluded in 2009, producing 192 nodes. Focus area applications include life science, digital media, enterprise computing, computational science, engineering, and education. Projects under areas of digital media and enterprise computing were put under Software Industry Promotion Agency (SIPA), Ministry of ICT, and so was the Office of Thai National Grid Project.

When the project was completed in the early 2009, it was reviewed to have made successful progress. Since then, the project has been operating under an approximate USD $300,000 annual budget for maintenance and services but no expansions. Aside from serving resources to current users, one goal for the project right now is to find a clear direction for Thai National Grid Center.

EUAsiaGrid

EUAsiaGrid Project was a 2 year project (EU-FP7 Coordination and Support Action) spanning from 2008 to 2010. In the end of the project, EUAsiaGrid has four European partners and eleven Asian partners, including National Electronics and Computer Technology Center (NECTEC) and Hydro and Agro Informatics Institute (HAII) from Thailand.

The resources contributed by NECTEC include UI, CE, DPM and MON machines with two 1TB SE, two 2 GHz Intel Xeon, 8 cores, 16GB RAM worker nodes and one P4 2GB RAM worker node. Applications include chemistry and high energy physics, and there is user training for computational chemistry scientists.

HAII also contributes resources to EUAsiaGrid such as UI, CE, MON, and WMS machines, 10 TB SE, and four 2 GHz Intel Xeon, 12 cores, 4 GB RAM worker nodes. There are Disaster Mitigation Workshop Series held in Bankok and Taipei (ISGC 2011). Applications include Grid for water security, image processing, and climate change.

ThaiGrid was involved in other Grid activities such as PRAGMA, EGI-InSPIRE, NECTEC GOC CA, National Health Information System in Thailand and CERN activities.

National e-Science Project

National e-Science Project is a five-year project to provide national infrastructure for e-Science in Thailand. The project runs as a consortium with a total budget of USD $3.7 million. It uses Virtual Organization (VO) to manage application areas and share resource through Grid middleware, GLite. There are five main areas of applications:
1. particle physics
2. climate change
3. water resources & energy and environment
4. computational science and engineering
5. computer science and engineering

ThaiGrid currently has six members: National Electronics and Computer Technology Center, National Nanotechnology Center, Hydro and Agro Informatics Institute Chulalongkorn University, Suranaree University of Technology, and King Mongkut's University of Technology Thonburi. The target of ThaiGrid is to reach 856 cores, 520TB by 2015.

[Based on the presentation made by Piyawut Srichaikul, NECTEC] [2]

Vietnam

4 High Performance Computing (HPC) centers were set up in Hanoi and HCM City between 1995 and 2005: Ho Chi Minh City University of Technology (HCM UT), Hanoi University of Technology (Hanoi UT), Institute of Information Technology-Vietnamese Academy of Science and Technology (IOIT-VAST) and Hanoi University

of Science and Technology (HUST). Almost all activities on grids are from academic research projects due to low computing capacity, lack of interoperability, and the absence of large scale applications and the cooperation between HPCs and customers. There were some Grid research groups linked with the HPC centers in HCM UT, Hanoi UT and Hanoi US offsite between 2003 and 2009.

Vietnam has been participating in Pacific Rim Application and Grid Middleware Assembly (PRAGMA) since 2002. Participating institutes include: IOIT-VAST, HPCC-HUT and HCMC UT. IOIT hosted PRAGMA workshop 17 on Oct. 28-30, 2009 in Hanoi. Vietnam also joined *Asia Pacific Advanced Network* (APAN) in 2007.

VinaGRID project (2008-2010) was funded by the Ministry of Science and Technology. The project goals include building a cyber-infrastructure to connect partners (IOIT-VAST, HPC-VNU, HPCC-HUT, etc.), deploying advanced grid technologies to improve the performance, experimenting with some grid-based applications, and connecting with other grid systems in the region and the world. The project was not very successful because of limited resources and the lack of national coordination; there were mostly research papers and no real applications.

The EGEE Grid has been introduced in Vietnam in 2008 with the support of Centre National de la Recherche Scientifique (CNRS). At the same time, Vietnam has established a dedicated network for research and education (VinaREN) which is funded by the Ministry of Science and Technology and a member of the TEIN 2/3 consortium from 2008 to 2010. The project was not very successful due to the lack of national coordination and limited resources. There were mostly research papers instead of real applications. One productive project of the VinaREN is the RAMS-Grid and MM5-Grid which combined with ensemble forecasting and grid computing to get faster results and improved accuracy. RAMS-Grid is more accurate than RAMS; a test hurricane trajectory was run on the RAMS-Grid. In the graph below, the red line is the real trajectory, the green line is predicted trajectory, and the yellow line is the predicted trajectory by RAMS-Grid.
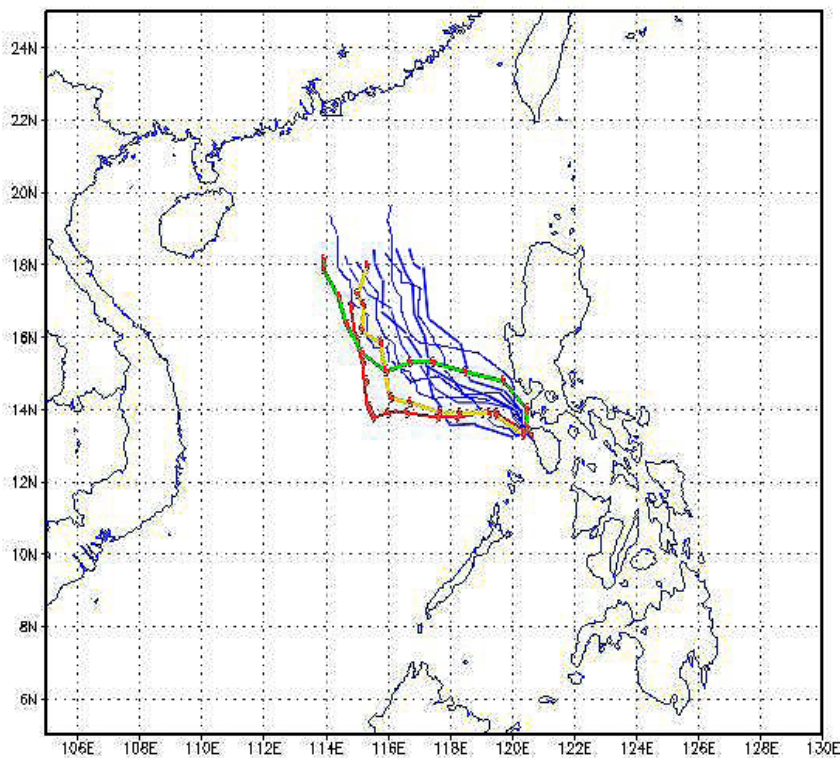
Fig. 17: RAMS-Grid test on Chanchu hurricane, May 13-16, 2006 data.

In 2009, three small EGEE nodes were installed on VinaREN at the IOIT, the Institut de la Francophonie pour l'Informatique (IFI) and the HUST. Last December, the International Workshop on Grid Computing Applications was hosted in Vietnam at IFI. It was concluded that distributed computing and storage infrastructures are opportunities for Vietnamese scientists to play a leading role on world-wide issues such as healthcare, climate change, and natural disasters.

One application called "gINFO – grid-based International Network for Flu Observation" has already been developed and deployed by IFI, a member of the EUAsiaGrid project (2009-2010). CNRS, IRD, and Asia Pacific Grid Initiatives collaborate on these applications and also commit to provide tools and support for modeling, simulation and data management. Though there are large number of young highly educated scientists with strong expertise in programming and two schools were organized in 2007 and 2009, funding and national coordination are still needed. The key issue, however, is that the distributed computing cannot do without a network—it needs to be built on a robust, fast and reliable network linking the grid nodes, clouds and clusters in Vietnam together and connect to other the research and educational institutes in the world.

Four on-going projects for Earth and Life Sciences with the active participation of partners coming from the Vietnam Academy of Science and Technology (VAST) have been identified recently and undertaken at IFI and IOIT.

1. g-INFO (grid-based International Network for Flu Observation), version 2
    - This project is set for the surveillance of emerging diseases.
    - Version 1 of g-INFO is running on biomed and EUAsia VOs.
2. Drug discovery on local biodiversity compounds
    -45 compounds have been tested, and the compound database is ready.
3. Database of pre-calculated earthquake/tsunami scenarios for the Vietnam
    -20 scenarios were calculated using ArcGIS.
4. WN on-demand deployment with Virtualization and Cloud Technologies
    -Tools are being developed for WN on-demand management and WN image configuration based on OpenNebula/Stratuslab.

In conclusion, Grid applications have just begun to be developed in Vietnam, thanks to the support of CNRS and EUAsiaGrid project partners. Real needs have been identified, and related projects have started. However, there are difficulties in the development, including people's lack of experiences, very limited research funding and grids resources (e.g. cores, storage capacity), limited bandwidth and unstable Internet connections, and the absence of a National Grid Initiative (NGI). Despite the obstacles, the goal is for the on-going projects to deploy results to contribute to the community in the coming years.

[Based on the presentation made by Nguyen Hong Quang, Institut de la Francophonie pour l'Informatique] [2]

References

[1] International Symposium on Grids & Clouds 2011, http://www2.twgrid.org/event/isgc2011/index.html

[2] Program of ISGC 2011, http://www2.twgrid.org/event/isgc2011/program.html