

# Asia Federation Report on International Symposium on Grid Computing (ISGC) 2010

Francois Grey<sup>1</sup> and Simon C. Lin<sup>2</sup>

1. UNOSAT, CERN, Switzerland

2. Institute of Physics, Academia Sinica, Taiwan

## EXECUTIVE SUMMARY

This report provides an overview of developments in the Asia-Pacific region, based on presentations made at the International Symposium on Grid Computing 2010 (ISGC 2010), held 5-12 March at Academia Sinica, Taipei. The document includes a brief overview of the EUAsiaGrid project as well as progress reports by representatives of 13 Asian countries presented at ISGC 2010. In alphabetical order, these are: Australia, China, India, Indonesia, Japan, Malaysia, Pakistan, Philippines, Singapore, South Korea, Taiwan, Thailand and Vietnam.

As in 2009, the country reports represent a wide range of Grid technology penetration, reflecting to some extent the different countries' socio-economic levels of development. Participation in the Worldwide LHC Computing Grid (WLCG) remains a common incentive for many countries' involvement in Grid technology. For example, Pakistan has established a federated Tier-2 activity, as no single institution was adequately resourced for this.

A second driving force is the European Commission's bold initiatives to spread know-how and capacity for grid-based e-Science worldwide, based on the EGEE projects, and associated projects like EUAsiaGrid (see map below). As a result some countries of the region are even taking the lead in some EGEE-related developments. For example, South Korea's KISTI now leads development of AMGA, an official EGEE gLite middleware component for a metadata catalogue service on the grid.

A third factor that can be gleaned from these reports is the leadership that Taiwan's ASGC has shown in the region, which has enhanced bilateral collaborations between countries, and also led to multilateral initiatives driven by common challenges such as earthquake monitoring and finding cures to neglected diseases such as dengue fever.

While there is much progress to report, there is also a good deal of uncertainty looking to the future. As EUAsiaGrid comes to a close in 2010, a key question, echoed in some of the reports, is how to build on the burgeoning regional collaborations that this project has helped to create. Another element of uncertainty is how to transition from Grids to a grid/cloud hybrid world. Finally, the issue of af-

fordable high bandwidth network connectivity in the region remains a complex one, for political as well as economic reasons.

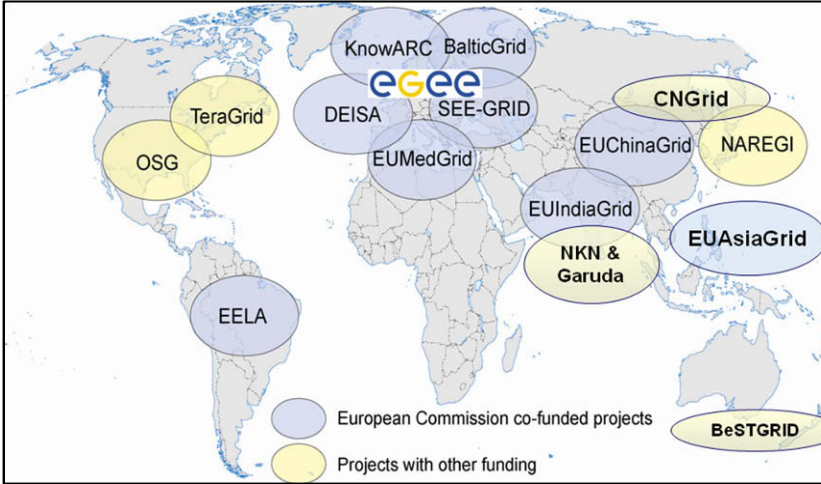


Fig. 1: EGEE-related projects (blue) and other major regional Grid initiatives.

## 1 EUASIAGRID OVERVIEW

EUAsiaGrid is an EU-funded project aiming to bridge a perceived gap between Europe and Asia in the area of Grid computing. The EUAsiaGrid project promotes awareness in the Asian countries of the EGEE e-Infrastructure and related European e-Infrastructure projects and supports capacity building in the region. To this end, it: (a) supports scientific communities through help with application porting and deployment; (b) provides training to further increase grid technology know-how; (c) monitors the results and gives feedback for the definition of policies. The consortium behind the project comprises 17 partners – fourteen in the Asia-Pacific region, distributed over twelve countries.

In order to develop a deeper understanding of the computing, storage, application support and training requirements, EUAsiaGrid has conducted a Requirements Survey in five different languages (English, Mandarin, Thai, Vietnamese, Malaysian). Based on the evidence collected and on the local expertise of the partner institutions in the region, the project has developed coordination and support activities to help foster the establishment of certified resources in the partner countries and to increase human capacity for local support. It has provided support for applications in areas such as high energy physics, computational chemistry, miti-

gation of natural disasters, bioinformatics and biomedical science and social sciences. The establishment of the EUAsia VO as a catch-all VO for the Asia-Pacific region has enabled researchers throughout the region to gain access to grid resources through a simplified process in order to increase uptake of the infrastructure.

The project has produced a number of successful instances of scientific work conducted on the infrastructure. For example, the EUAsiaGrid DC2 (Data Challenge 2) Refined Activity has focused on screening of 20,000 potential ligands for Avian Flu target proteins. EUAsiaGrid has delivered an intuitive and user-friendly productive system that enables biologists to run the simulations and to manage the results on the grid as easily as with a desktop application. In March 2009, virtual screening jobs were run consuming a total of 1,111 CPU-days under the EUAsia VO and more than 160,000 output files with a data volume of 12.8 Gigabytes were created and stored in a relational database. Based on the success of DC2, the EU-AsiaGrid project has also launched the EUAsiaGrid Dengue Fever Activity. Similar activities are being prepared in other application areas such as disaster mitigation.

As many of today's scientific challenges require long-term international collaboration, it is paramount that researchers have access to a persistent, sustainable infrastructure that they can access as needed. Such an infrastructure needs to be supported so that researchers can take its operation for granted and focus on their substantive research work. In order to foster the long-term sustainability of the infrastructure the EUAsiaGrid has supported, the project has developed a roadmap that outlines the collaboration and governance structures for an Asia-Pacific Grid Initiative that will build on and expand on the work done by EUAsiaGrid.

[based on presentation by Marco Paganoni, INFN and University of Milano-Bicocca, Italy and contribution to EC Publication "Towards a Sustainable European e-Infrastructure"]



ARCS is uniquely positioned as the entity which is the cooperative aspect of the key nationally distributed eResearch Service Providers.

ARCS Collaboration Services include:

- Video Collaboration: ARCS provides support for EVO (in collaboration with AARNET) and Access Grid technologies - including High Definition;
- Instant Messaging (IM): Jabber is flexible IM system like MSN, ICQ etc (ARCS runs an ARCS Jabber server);
- Research Oriented Web Tools: Research web-based workspaces and tools - Sakai, Drupal, Plone, Wiki, Twiki, Google Calendar, Google Apps...;
- Customer Service: End-user support is provided via telephone 1800 TOARCS (answered all business hours) & email request tracker system [customer.service@arcs.org.au](mailto:customer.service@arcs.org.au).

The ARCS Data Fabric is available to every Australian researcher and provides:

- Access to Data Fabric uses the AAF;  
A system to store, manage and share data currently based on iRODS (previously SRB);
- Distributed storage and access nodes;
- Some free storage to every researcher (can be extended typically at hardware cost only to arbitrary size);

The data fabric can be accessed with web interface (with Shibboleth) and webDAV and CLI tools (with SLCS); and will soon be unified with Grid/Cloud computing infrastructures.

Keys to a future successful ARCS:

1. deliver well and strongly into these core roles as required by the federal government;
2. development, deployment and operation of highly valued, robust, reliable & easy-to-use tools and services.

The 2008 Strategic Roadmap for Australian Research Infrastructure provided a future perspective and identified significant gaps in infrastructure provision, particularly electronic storage of research data. The result was a plan for a government funded enhanced eResearch infrastructure via a Super Science investment of \$312 million for 2009-2013, with \$97 million to ARCS for Data Storage and Collaboration Tools made up as:

- \$35 million for ARCS Managed Core Services;
- \$12 million for Research Community Projects;
- \$50 million for Data Storage Infrastructure.

Concerning the ARCS Managed Core Services (\$35m), the breakdown is:

- Continuous development, deployment and operation of ARCS managed core services Research Community Projects (\$12m):
- Open competitive process involving national research communities to identify discipline specific services to be developed - typically ongoing operation by the research community themselves, perhaps building on core services.
- Data Storage Infrastructure (\$50m):
- Consultation underway to identify data storage coinvestors and to decide on how many/which sites and how to achieve seamless interoperability.

[based on presentation by Anthony Williams, Executive Director, ARCS]

### 3 CHINA

In the past year, the Beijing Tier-2 Site for WLCG at the Institute of High Energy Physics, Chinese Academy of Sciences has ramped up to 1100 CPU cores, and 400TB disk storage capacities. This site has achieved a high availability in the range 98%-100%.

The analysis efficiency of the LHC detector ATLAS has been increased by a factor of 16 by introducing the FroNTier/Squid system. Frontier is a software subsystem that translates database queries into HTTP, looks up the results in a central database at CERN, and caches the results in an industry-standard HTTP proxy/caching server called Squid.

More than 85,000 jobs of CMS for LHC 2009 first collisions have been completed on the Beijing Tier-2 site in 2009, with a success rate of 95%.

The networking infrastructure used by the Tier-2 site in China is based on a TEIN3 Link to Europe: 1Gbps with timeout <170ms and the GLORIAD Link to America at 622Mbps. Data I/O per day is around 3TB.

To enhance the efficiency of software deployment, the Quattor system, an administration toolkit for optimizing resources, was improved. A customized monitoring system was developed to monitor the status of the local resources. The Tier-2 in Beijing has implemented a multi-grid job management system to facilitate job submission, query and management over different grid systems.

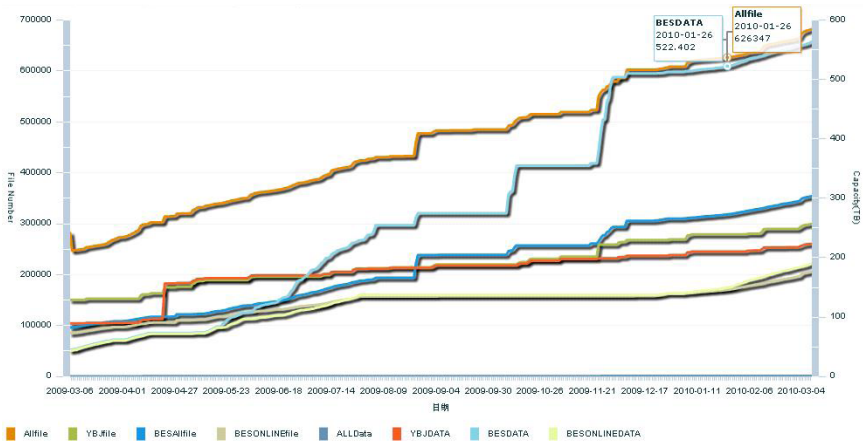


Fig. 3: Real time monitoring of CASTOR, showing rapid increase in file number and capacity handled since March 2009.

[based on presentation by Lu Wang, Computer Centre, IHEP]

## 4 INDIA

EUIndiaGrid has supported the improvement of the Indian network infrastructure both at National and International Level. It has also supported the establishment of the Indian National Certification Authority providing access to grids worldwide for Indian Researchers. The project has had an excellent collaboration with Indian National Grid Initiative, GARUDA, now entering phase three, and with the Department of Atomic Energy Grid Project, which focuses on WLCG.

The GARUDA project is coordinated by the Centre for Development of Advanced Computing (CDAC) and its network infrastructure was designed, set-up and operated by ERNET (Indian Education & Research Network, (<http://www.eis.ernet.in/index.htm>). Both CDAC and ERNET are partners in EU-IndiaGrid2 project.

The National Knowledge Network (NKN) will provide links to all the networks in India by a multi-gigabit, low-latency, OFC-based backbone. The main design consideration for NKN is to create an infrastructure that can scale and adapt to future requirements. The project's ultimate aim is to unite stakeholders in science, technology, higher education, R&D and e-governance using network speeds of tens of gigabits per second coupled with extremely low latencies.

The NKN will eventually cover over 1000 institutions directly. The initial phase of NKN, with 15 Core locations and about 57 institutes covering leading national R&D labs and educational institutes, is operational since December 2008. It is expected to connect more than 100 institutes by the end of March 2009. EU-IndiaGrid2 Indian partners have an active role in NKN with leading responsibilities in the Committee for High Performance and Grid Computing. This national framework is complemented by the transition to phase III of the TransEurasia Information Network (TEIN3) (<http://www.tein3.net>) which sees, for the first time the participation of India (represented by ERNET and the Department of Information Technology) to the feasibility study. NKN and TEIN3 set the premises for important developments and significant strengthening of the cooperation between Europe and India in the domain of international connectivity. This was highlighted in the context of dedicated meetings held in January 2009, with participation of representatives from EU-IndiaGrid project, DANTE, the European Commission and the Indian Government.

In addressing global research challenges, EU-IndiaGrid2 mobilises the provision of global cross-disciplinary research services by supporting specific user communities benefiting from Grid infrastructure in four strategic areas for EU-Indian user communities:



- Climate change – developing studies on South Asian climate change scenarios with a worldwide perspective
- High Energy Physics – driving EU-India collaboration in Large Hadron Collider data research
- Biology – increasing detail and scope of simulation software
- Material Sciences – testing and extension of complex applications

[ Based on presentation by Alberto Masoni, INFN and on the EUIndiaGrid website]

## 5 INDONESIA

Research on Grids in Indonesia started around 2006, together with the development of INHERENT (Indonesia Higher Education Network). INHERENT connects more than 200 universities in Indonesia.

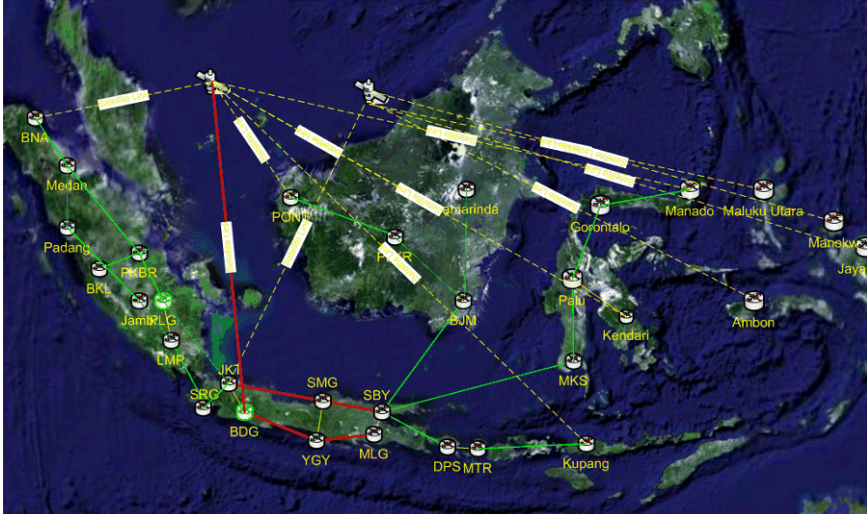


Fig. 4: INHERENT is the Indonesian national Research and Education Network which is currently connecting about 500 universities in Indonesia.

*inGrid* or Indonesian Grid is an infrastructure developed by the University of Indonesia using UCLA Grid Portal consisting Globus Toolkit 4 grid middleware and Gridsphere grid portlet framework. It is based on the INHERENT network. In 2008, there are two clusters connected to *inGrid*: one production cluster from Faculty of Computer Science, University of Indonesia and one research cluster. A range of applications available in the research cluster: Povray (3.1g); gcc; mpiB-LAST; g77; GNU Octave; GROMACS.

Some departments in Institut Teknologi Bandung (ITB) have been using cluster computing to support their Research and Education activity, including for weather prediction <http://weather.geoph.itb.ac.id>, and engineering physics to support data processing, design, and testing. <http://computational.engineering.or.id>.

In 2008 ITB joined the EUAsiaGrid Project, with the objectives of assisting regional integration to the wider Grid Infrastructure, capturing local e-Science requirements and promoting common e-Science applications, both existing and new.

The ITB Grid site was developed in collaboration with EUAsiaGrid Project and uses gLite middleware, consisting of:

- User Interface combined with MON-Box
- DPM/Storage Element
- Computing Element
- Workload Management System

Future Development includes integrating existing Cluster to Grid Infrastructure (current experiment is by Computational Engineering Lab In Engineering Physics Dept.)

In weather forecasting, the experimental application is using MODIS data from satellite images to generate weather forecast information and display it in a web page. Grid utilization to support Numerical Weather Prediction (NWP) research activity is concerned with the development of a common regional platform for NWP application in Southeast Asia. An NWP experiment is performed by implementing WRF4G (WRF for Grid) developed by University of Cantabria under EELA2 to find the most suitable downscaling strategy for NWP in South East Asia.

The result covers Indonesia and its neighbouring countries with horizontal resolution of 30 km, and covers the islands of Java, Bali, and South of Sumatera with a horizontal resolution of 10 km. Plotted parameters are 3-hourly convective rainfall, temperature at 2 m height, and surface equivalent potential temperature. The output is made available via a Google map interface, covering a few Indonesian cities twice a day.

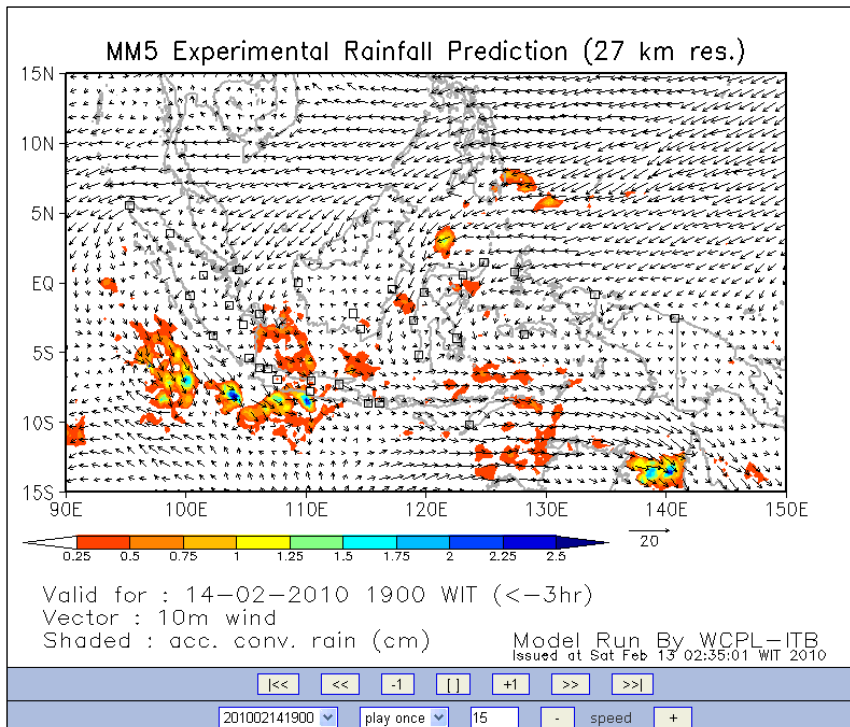


Fig. 5: Predicted rainfall over Indonesia, based on Numerical Weather Prediction.

[Based on a presentation by Basuki Suhardiman, Institut Teknologi Bandung, Indonesia]

## 6 JAPAN

The National Research Grid Initiative (NAREGI) Project originally started as an R&D project funded by the Ministry of Education, Culture, Sports, Science and Technology, MEXT (FY2003-FY2007) with a budget of 2bn Yen (~17M\$) in FY2003. It involved a collaboration of National Labs Universities and Industry in the R&D activities. Nano-science applications were promoted. Starting in 2006, the project was redirected as a part of the Next Generation Supercomputer Development Project.

The goals of NAREGI were to develop a Grid Software System (R&D in Grid Middleware and Upper Layer) as the prototype of future grid infrastructure in scientific research in Japan and to provide a testbed to prove that a high-end grid computing environment (100+Tflop/s expected by 2007) could be practically utilized by the nano-science research community over the academic backbone network, SINET3. A further goal was to participate in international collaboration/interoperability (U.S., Europe, Asian Pacific), and to contribute to standardization activities such as OGF GIN-RG, PGI-WG etc.

NAREGI version 1 middleware was developed in FY2007, to have more flexible scheduling methods, reservation-based scheduling, coexistence with locally scheduled jobs, support of non-reservation-based scheduling, support of “bulk submission” for parameter sweep type jobs. In addition, improvement in maintainability and more systematic logging using Information Service (IS), an easier installation procedure were requirements.

NAREGI Version 1.1.4 was released in September, 2009, with version 1.1.5 to be released end of March, 2010. NAREGI Grid Middleware is being deployed to the national supercomputer centers as an important component of the Japanese Cyber Science Infrastructure Framework. A new project (RENKEI) started in FY 2008 to provide seamless access between NAREGI and the 3rd Tier resources. NAREGI is planned to provide the access and computational infrastructure for the Next Generation Supercomputer System, and is being deployed on the SINET4 upgrade to the research network planned for 2010 (see [Fig. 6](#)).

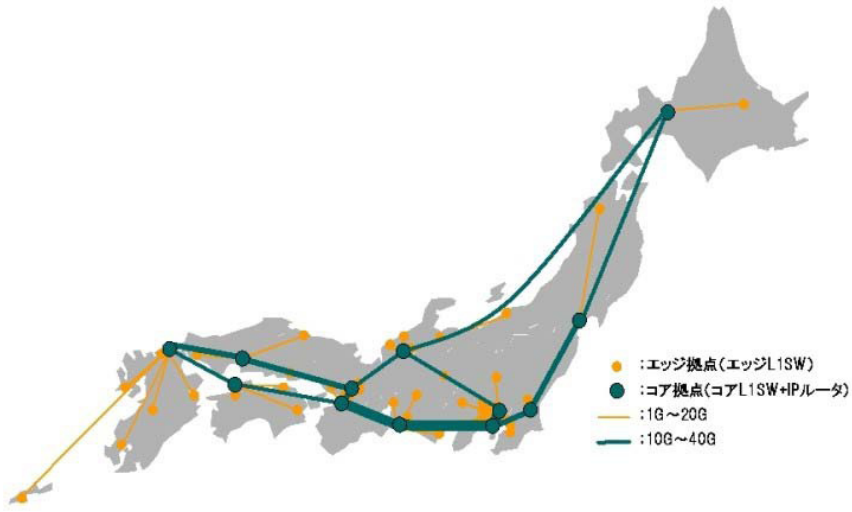


Fig. 6: Map of the SINET4 upgrade to a 40GB backbone (green) planned for FY2010

## 6.1 The RENKEI Project

The RENKEI Project is a new R&D project, which started in September 2008 under the auspices of MEXT. In this project, a new light-weight grid middleware and software tool will be developed in order to provide the connection between the NAREGI Grid environment and wider research communities (3rd Tier resources).

In particular, technology for the flexible and seamless access between the national computing center level and the departmental/laboratory level resources, such as computers, storage and databases is highly emphasized. Also, this newly developed grid environment will be made interoperable with the major international grids.

One highlight of the RENKEI project in 2009 is a demonstration of interoperability conducted at the 5th IEEE eScience Conf.(Oxford UK) where job submission of the application Minem (Plasma Charge Minimization) was realized to multiple grids via HPCBP.

Other activities include development of nation-wide distributed file system technology. Research in this area involves optimal automatic placement of file replicas based on Gfarm 2.0. Development of an interoperable file catalog service between heterogeneous grid environments has also been a research topic, since current file catalog systems (LFC (EGEE gLite), MCAT (SRB), etc.) are not in-

teroperable. Work has focused on development of standardized file catalog based on RNS (Resource Namespace Service) specifications (OGF).

KEK has played a key role in promoting interoperability between gLite and NAREGI, based on GIN and JSAGA. KEK is also investigating the use of Grid and cloud technologies to support the computing infrastructure for the Belle-II accelerator, which is the upgrade of KEKB.

## ***6.2 The Next Generation Supercomputer Project***

The Next Generation Supercomputer Project aims at the development, installation and application of an advanced high performance supercomputer system, as one of Japan's "Key Technologies of National Importance". The total budget is about 115bnYen (\$1.15bn) over the period of FY2006 - FY2012

The goals of the Next Generation Supercomputer Project include:

1. development and installation of the most advanced high performance supercomputer system;
2. development and wide use of application software to utilize the supercomputer to the maximum extent;
3. provision of flexible computing environment by sharing the next generation supercomputer through connection with other supercomputers located at universities and research institutes;
4. establishment of an "Advanced Institute for Computational Science" (tentative name).

It is planned that NAREGI will provide the access and computational infrastructure for the Next Generation Supercomputer Project, and discussions are underway towards this goal.

[Based on reports by Kenichi Miura and Kento Aida, Center for Grid Research and Development, National Institute of Informatics, Takashi Sasaki, KEK]

## 7 MALAYSIA

The original implementation plan for the National Grid Computing Initiative (NGCI) in Malaysia, agreed in 2006, was shared, distributed computing resources to be deployed at several locations throughout the country with a similar facility to be run by a Grid Operations Centre and all sites to be connected via MYREN. In practice, an alternative approach was implemented which is more of an access to HPC services rather than a truly distributed shared resources: it focused on Grid application users with nearly no access for Grid core technology developers.

EUAsiaGrid has contributed greatly towards the next phase of Grid development and in particular gave Malaysian scientists and institutions a better understanding a truly distributed shared compute resources, and emphasize the importance of federated certificates with certificate issuance to users that must be adhered to.

Partly as a result of this, a review of NGCI activities last year, it was agreed to establish a true Grid computing ecosystem that:

1. addresses the needs of both Grid technology developers and users in various domain specific applications;
2. delivers the services of high performance computing and shared computing/storage resources locally, regionally and globally;
3. Facilitates the access of Grid computing resources for all through world-standard technology framework and best practices that are consensually agreed between the Grid service provider and the users.
4. includes cloud computing; both public and private cloud.
5. enables internationally recognized, federated and trusted security framework for Grid resource sharing among government agencies including private companies over the global network;
6. coordinates and facilitate application, review, approval, monitoring, and enforce accountability (including corrective/punitive actions) of grants by Grid R&D Expert Committee;
7. determines distribution of costs and billing on the use of the facility for both public and commercial when sharing resources;
8. introduces new scientific applications and technology components on Grid computing into local R&D;
9. establishes and expands joint research in all aspects of Grid Computing technology and grid-enabled applications locally and globally
10. establishes research collaborations with international Grid computing organizations (such as PRAGMA, SEAGF, CGF, OGF etc);

NGCI is now able to contribute and share computing resources with partners with a federated certificate authority (ASGCCA), and work has begun to establish a Malaysia certificate authority. NGCI will follow closely EGI especially on the middleware. NGCI will promote trust and encourage sharing for example by im-



plementing VOMS or GridShib Digital Certificates and developing policies for harmonization of Raw and Clean Data coming from various disciplines.

Future directions include extending National Grid terms of reference and roadmap to Cloud Computing, exploring Virtual Machine Managers, implementing Fat Nodes for RAM intensive computing for Gene Sequencing Analysis. MYREN Phase 2, the next phase of the Malaysian network, is being deployed and will see the inclusion of polytechnics and community colleges.

KnowledgeGRID is a strategic initiative by the Ministry of Science, Technology and Innovation spear-headed by MIMOS in close collaboration with local universities, research institutions and industry. The initiative aims to maximize the utility of high performance computing resources to accelerate research and industrial development.

KnowledgeGrid combines networked resources – desktops, servers, storage, databases, with scientific instruments, to form a massive repository of computing power to be tapped whenever and wherever it is needed most.

Research areas include AgriGRID for precision farming, BioGRID (DBrain) for dementia studies, VLSI GRID for Integrated circuit design (FRGA) for green motion controller and AutomotiveGRID for fuel efficient concepts in automobiles. The 2nd KnowledgeGRID Malaysia Forum was held 26-29 October 2009.



Fig. 7: Snapshot of homepage of KnowledgeGRID Malaysia.

[based on presentations made by Suhaimi Naipis, CTO and Director, Info-Comm Development Centre and Bukhar Ikhwan Ismail, MIMOS]

## 8 PAKISTAN

### *8.1 Progress of HEP Grid in Pakistan*

The effort to bring Pakistan on the WLCG map as a Grid Node was started in October, 2003. A Grid Technology Workshop was organized by NCP from October 20-22, 2003. The first ever test-bed was deployed during the workshop for tutorial. This test-bed consisted of 9 machines.

Today, Pakistan operates a Tier-2 Federation. No single site had enough resources and bandwidth to become a Tier 2 site so the Idea of a Tier 2 Federation was proposed based on WLCG Tiered Architecture. The National Centre for Physics (NCP) is a Regional Center for this federation. Other participating institutes were The Commission on Science and Technology for Sustainable Development in the South (COMSATS) National University of Science & Technology (NUST) Pakistan Atomic Energy Commission (PAEC1, PAEC2, PAEC3). In 2010, statistics for the Tier-2 federation are as follows: 360 CPU cores, 195TB of storage and 155 Mbps bandwidth.

For example, NCP is connected with 3 ISPs: fiber connection from Nayatel @ 10Mbps, Wireless Radio LinkDotNet@ 3.5Mbps, and Fiber Connection from HEC (PERN2) with 2Mbps internet and 155Mbps R&D link connected with TEIN3, GEANT2, and Internet2. User traffic and Grid traffic is divided among these links.

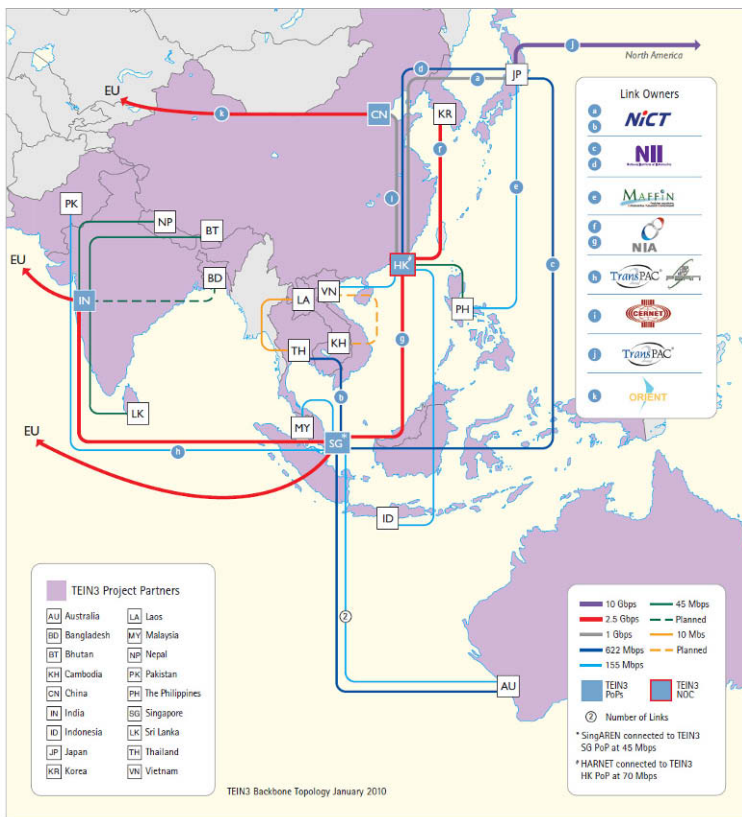


Fig. 8: TEIN3 connectivity in South and South-East Asia, showing the link between Pakistan and Singapore.

High availability and fault tolerance is achieved by installing physically redundant devices, running high availability protocol between redundant devices, Hot Standby Routing Protocol (HSRP), Virtual Router redundancy Protocol (VRRP), Gateway Load balancing Protocol (GLBP). As a result by March 2010 nearly 70,000 KSI2K hours of raw processing has been served by NCP-LCG2, with around 120,000 jobs executed, with similar figures for PAKGRID-LCG2.

PK-GRID-CA was the first certification authority in Pakistan, starting in 2004. It issues X.509 digital certificates, having issued over 150 certificates so far, of which over 50 are active still. There is a web portal for submitting certificate requests.

Despite challenges of limited bandwidth and frequent power outages, the effort to establish a Grid node in Pakistan has been successful, and Pakistan is actively contributing CPU and storage resources to the LHC community.

[based on presentation made by Sajjad Asghar, National Centre for Physics (NCP)]

## 9 PHILIPPINES

The Advanced Science and Technology Institute of the Department of Science and Technology (ASTI) has 16 nodes online: 2 nodes for sandbox; 6 nodes for meteorology; 8 nodes for bioinformatics. Currently it has 39 nodes with 312 cores, 2.5 TFLOPS (6 more nodes soon) and 8 FPGA accelerators. There is also 14TB of storage.

Current local partners include ADM, UP CSRC and PAGASA. The compute clusters are:

- Banyuhay – The Bioinformatics Cluster
- Unos – The Meteorology Cluster
- Dalubhasaan – The Cluster Sandbox
- Buhawi – The General-Purpose Cluster (EGEE certified production cluster)
- Liknayan – The EUAsiaGrid and EGEE Collaboration Cluster (EGEE certified production cluster)

Middleware and Applications include

- Cluster Middleware: ROCKS 5.2.2
- Grid Middleware: gLite
- Bioinformatics (BioRoll and Progeniq BioBoost)
- Seismology (SPECFEM3D Globe)
- Meteorology (WRF , MM5, RegCM)
- Oceanography (SeaDAS)

Current Users are UP Marine Science Institute (Fish Larval Dispersal Model for the Bohol Sea), PAGASA (Forecasting) and UPLB Biotech (Project on Data Warehousing for Drug Discovery). Future Users are Nimbus, Phivolcs, PAGASA (for climatology) and UP National Institute of Physics.

ASTI is a contributing member of EUAsiaGrid and institutional member of the Pacific Rim and Grid Middleware Assembly (PRAGMA). Training activities in 2009 included the Philippines Grid Computing 2009 Forum and Training-Workshop on 17-18 Nov 2009, with participants from UP-MSI, UPLB-BIOTECH, PAGASA, UP-NIP, UP-MathDept., UPM-NTHC, DLSU, IRRI.

Future plans include launching of the Philippine e-Science Grid (April 2010), Launching of ASTI's 3D Visualization Facility (April 2010), acquiring additional 6 computing nodes and 12 TB of storage space (2010), Resource Virtualization (Q1 2010), developing more FPGA applications, installing more grid based applications, continuing to advocate and promote grid technology to local universities and research communities.

Another area of focus for Grid-based research in the Philippines is disaster mitigation. A research team from the School of Science and Engineering of AdMU is collaborating with scientists at the Manila Observatory for Landslides and Flood Disaster Mitigation. The goal is to establish a grid-based database of Philippine geographical data related to landslides and floods that can be analyzed and visual-

ized through the EUAsiaGrid infrastructure. Cellular automata and agent-based models of landslides and floods are being developed and will be used for the simulations.

Biodiversity mapping is another direction being investigated for grid-based data warehousing. In terms of biodiversity, more than 7,100 islands fall within the borders of the Philippine, identified as one of the world's biologically richest countries. The country is one of the few nations that is in its entirety, both a conservation hotspot and a megadiversity country, placing it among the top priority hotspots for global conservation.

Also in 2010, a workshop was held with the Asian Development Bank on evolutionary trade modelling and poverty alleviation. A team from AdMU is collaborating with economists from ADB to model evolutionary trade in some of the poorest regions in South East Asia, such as East India, Bangladesh, Bhutan and Nepal. The goal is to determine the impact of investments in the economic growth of the region using a geographical information system and cellular automata/agent-based modeling techniques, based on a grid infrastructure.

### Risk to Earthquake-Induced Shallow Landslides

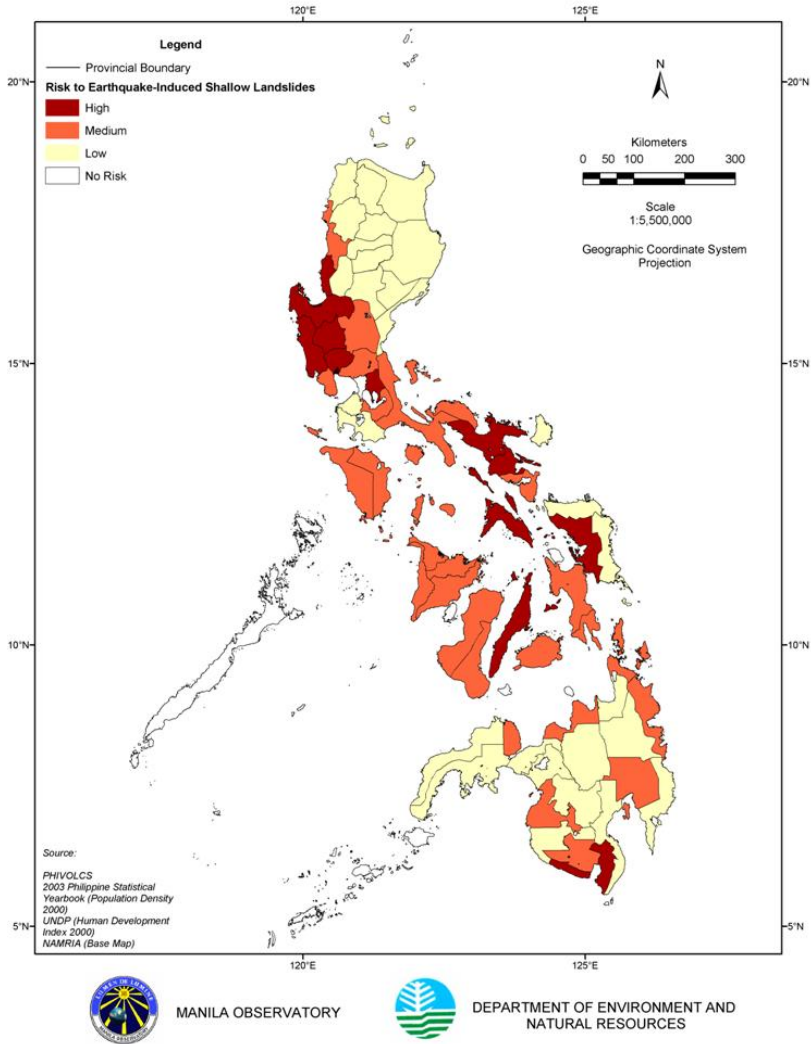


Fig. 9: Landslide and Flood disaster mitigation involves Grid-enabled data warehousing, risk assessment, monitoring and modeling and simulation.

[Based on presentations by Peter Antonio Banzon, Advanced Science and Technology Institute, and Rafael Saldana, Ateneo de Manila University]

## 10 SINGAPORE

Bioinformatics and computational biology are growing fields in the wake of a global digitalisation and quantization of modern biology. The requirement for computing power increases with the volume of biological data, which is exploding due to the great reduction in the cost of procedures such as genome sequencing. This translates into a growing need for computational power at the bench top and on the field, wherever biologists go, and wherever they are being trained.

An example is BioSLAX, a new live CD/DVD suite of bioinformatics tools that has been released by the resource team of the Bioinformatics Center (BIC), National University of Singapore (NUS). Bootable from any PC, this CD/DVD runs the compressed SLACKWARE flavour of the LINUX operating system also known as SLAX. SLAX is becoming the live CD/DVD of choice because of its ability to modularize almost any application and plug it into the system on the fly. The system can also be installed to USB thumbdrives or directly to the PC as a regular Linux using the BioSLAX installer provided.

There are two parts to the SLAX build, the core system and the individual modules. The core system is the OS itself and basic tools, usually referred to as the 'base', while the modules are the individual utilities that a user wants to have on their system. Since these modules can be put in or removed prior to creating the CD/DVD/USB, the system is fully modular and easily customisable.

Grid-enabling of BioSlax has been achieved using United Devices UD MP agents. This builds on a campus wide cycle-harvesting grid called the TeraCampus Grid (TCG@NUS) and enables Rapid recruitment of machines with no need for user installation, as well as push or pull instances. Ongoing research focuses on embedding gLITE inside BioSlax.

In the context of EUAsiaGrid, NUS has collaborated with ASGC, Taiwan and National Yang Ming University to create a next generation BioMirror, which is a compressed set of biodatabases available since 1998 in 12 countries in the Asia-Pacific region. This includes an authorID system, DocID Depository for bioinformatics published datasets, minimum information about bioinformatics investigation MIABi standards. This collaboration has also initiated computationally intensive projects for complex diseases.

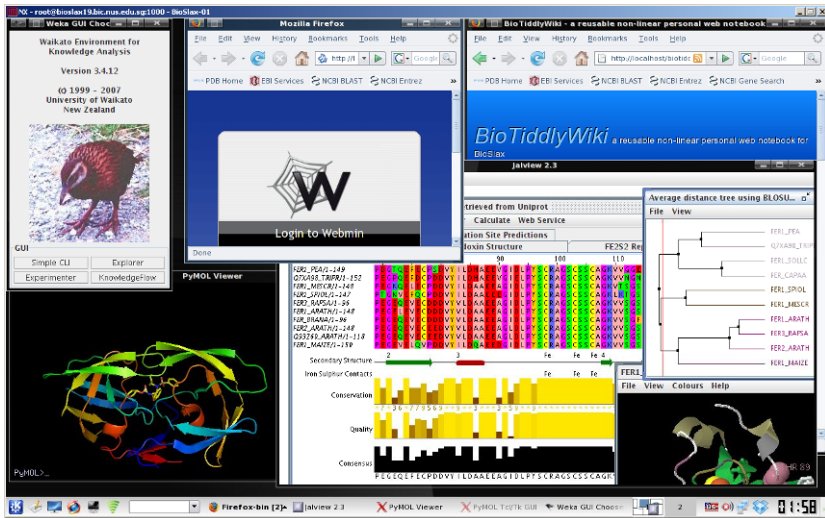


Fig. 10: Screenshot of bioinformatics teaching interface using BioSlax

[based on presentation by Tin-Wee Tan, National University of Singapore]



## 11 SOUTH KOREA

KISTI, the Korea Institute of Science and Technology Information, has participated in the EGEE project since 2006. Since 2007 KISTI has participated in the production grid operation, becoming an ALICE Tier2 in 2007, and contributing on that basis to the joint development of PROOF, a parallel ROOT facility. Since 2009 one full-time equivalent researcher has been devoted to PROOF development.

Currently, KISTI contributes 128 CPU cores and 30 TB storage to the ALICE distributed computing Grid, which is about 1.2% contribution to ALICE computing in the total job execution, and corresponds to processing nearly 8000 jobs per month in average.

KISTI has also been instrumental in establishing a collaborative VO with CC-IN2P3 in France in the area of Grid computing, with the objective to foster the adoption of grid technology and provide researchers in Korea and France with a production Grid Infrastructure. This has been up and running since October 2008, providing about 7,000 CPU cores and 2 TBytes of disk storage. About 50 users have joined the FKPL VO membership.

This VO has been used for deployment of Geant4 applications, run extensively by the National Cancer Center in Korea for compute-intensive simulations relevant to cancer treatment planning as well as deployment of two-color QCD (Quantum Chromodynamics) simulations in theoretical Physics. Several thousand QCD jobs have been run on the Grid, with each job taking about 10 days.



Fig 11: In February, KISTI organized Geant4 and Grid tutorial 2010 for Korean medical physics community. About 30 participants from major hospitals in Korea took part and about 20 Grid Certificates were issued and joined the FKPL VO.

KISTI was also involved in developing AMGA, an official EGEE gLite middleware component for a metadata catalogue service on the grid. KISTI has taken over the leadership of AMGA development since the July of 2009. AMGA 2.0 supporting the OGF WS-DAIR was successfully

released in October in 2009 in collaboration with CERN and INFN. Currently, KISTI is one of the partners of the open glite collaboration, and working with EMI, contributing to the evolution and maintenance of AMGA.

Drug screening has been a focus of Grid-related research in South Korea, through the WISDOM project. Computational methods used include pharmacophore based search and structure based docking. The requirements include 3D structure of target, databases of small molecules, a method to dock and score bound small molecules.

One example of a virtual screening challenge with grid computing concerns human intestinal maltase. The objective is to search for an inhibitor of enzyme action which enables glucose uptake, in order to treat diabetes type 2 related to obesity. Inhibitors already exist, but the search is aimed at more potent ones with less side effects.

Starting with 454,000 chemical compounds from Chembridge, scoring based on docking score results in 3016 compounds selected. Interaction with key resi-

dues reduces the field to 2616 compounds. Focusing on key interactions and using binding models and clustering, the number can be reduced to in vitro testing of just 42 compounds.

During this filtration process, the key numbers concerning the grid-based virtual screening are:

- Total numbers of docking: 308,307
- Total size of output results: 16.3 GBytes
- Estimated duration by 1CPU: 22.4 years
- Duration of experiments: 3.2 days
- Maximum numbers of concurrent CPUs: 4700 CPUs
- Crunching Factor: 2556
- Distribution Efficiency: 54.4 %

Similar grid-enabled searches have focused on:

- Malaria. One of the crucial drug targets in malaria is plasmepsin, the aspartic proteases of the parasite plasmodium. Plasmepsin is involved in the hemoglobin degradation inside the food vacuole during the erythrocytic phase of the malaria parasite life cycle.
- SARS. The global outbreak of SARS (Severe Acute Respiratory Syndrome) in 2002 set in motion a search for an effective vaccine. 3-CL-pro (chymotrypsin-like cystein protease) is an attractive target for development of antiviral drugs directed at SARS, since this protease is essential for the viral life cycle, there are a number of 3D structures available and preparation of the enzyme in large quantities is possible for in-vitro testing.
- H5N1. This flu strain has high fatality and resistance to available drugs such as Tamiflu. Both effective vaccination and antiviral drugs are needed. A suitable target is an inhibitor of Neuraminidase, a glycoprotein on the virion surface which releases the progeny virions from the infected cell. In vitro assay of compounds selected from in-silico screening performed at Academia Sinica.

[Based on presentations made by Hwa Ja Ryu, Chonnam National University and Soonwook Hwang, KISTI]

## 12 TAIWAN

The Mission of Academia Sinica Grid Computing (ASGC) is to build a sustainable research and collaboration infrastructure to support research by e-Science, on data intensive sciences and applications requiring cross disciplinary distributed collaboration.

Development of e-Science infrastructure in Taiwan is not just the result of the global e-Science collaboration (WLCG) but also occurs due to domestic support for new ways of doing science. While WLCG and EGEE usher in the petabyte-scale era, the e-Infrastructure of Taiwan has also had proved to be available for larger scale multiple sciences.

The value of adaption to new computing models and effective resource sharing has been demonstrated by user communities of high energy physics (HEP), drug discovery, long-term digital preservation and high throughput computation.

In 2010, gLite sites in Asia grew to 30 from only 5 sites in 2005, based on the support and coordination of the EGEE Asia Federation and Asia Pacific Regional Operation Centre (APROC) hosted by ASGC.

Other e-Science applications among Asia partners and with Europe and America, such as earthquake simulation, environmental change studies, computational chemistry, social sciences, and life sciences, are taking place endorsed by EUA-siaGrid and other projects.

All these efforts are driving the regional e-Infrastructure towards production quality and sustainable grids. From 2009, data centre energy saving and intelligent operation become one of key focus at ASGC. To continuously improve data management and computation for e-Science, technologies like virtualization, cloud computing and volunteer computing are under test, deployment and integration to Grid infrastructure.

Training for site administrators, e-Science application engineers, trainers and collaboration workshops with user communities were held more often, in most partner sites. In addition, ASGC is assisting more site operators joining APROC support, to take care of the regional collaboration infrastructure and to level up service quality in partnership.

In future, orchestrating users requirements to take advantage of the e-Infrastructure is the best model to grow the technology. ASGC will keep close co-operation with target user groups regionally and evolve the e-Infrastructure.

In terms of technological highlights in 2009, ASGC has promoted Grid-enabled services based on the Grid Application Platform, GAP. This has been used successfully for Grid-enabled virtual screening service for drug discovery called GVSS, which is a JAVA-based user interface developed to facilitate job submission and data management of large-scale molecular docking.

Thanks to the DIANE framework, GVSS allows submitted jobs to be split into multiple independent subtasks and run to completion; this ensures efficient utilization of the GRID resource for the massive molecular dockings empowered by Autodock3.

GVSS hides the complexity of deploying large-scale molecular docking on the GRID while provides users more flexible control over their docking jobs on the GRID. Provide on-line Avian

Flu & Dengue targets and ZINC compounds library (> 300,000). The user prepares the target/compounds gridmap files with on-line tools.

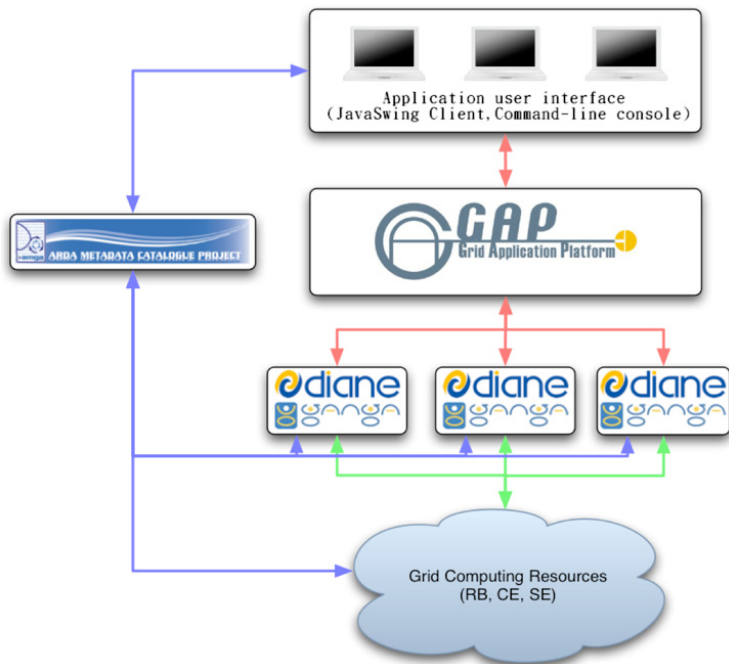


Fig.12 : Illustration of how GAP is used for GVSS

Another significant initiative led by ASGC is the Earthquake Data Center. The objective is to support both data acquisition, data services and risk analysis, as well as Earthquake Research (Earthquake data analysis). The Centre is based on a Federation of Tools and Interfaces (ANTELOP (ADPC, TW), Earthworm (TW), Wilber2, Netdc, and EIDS (Earthquake Information Distribution System). The Data Exchange Protocols used include ArcLink, ORB2ORB and others.

The participants are

- Malaysia: 5 stations
- ADPC: all 4 stations and the coming 10 more ones.
- Philippines: 1 Station (newly established in Jan.'10), more station data will be shared to IRIS and to the region.
- Taiwan: 7 stations (published to IRIS)
- Indonesia: 20 stations (indirectly through GFZ)

[based on presentations by Simon C. Lin, Eric Yen, Hsin Yen Chen & Yin Ta Wu, Academia Sinica]

## 13 THAILAND

Today, in Thailand, there is a need for the “National Health Information System (NHIS)” that provides the transparent and secure access to health information across geographically distributed healthcare centres.

To achieve this, several issues must be resolved altogether: (i) the diversity of health information structures among healthcare centres; (ii) the availability of health information sharing from healthcare centres; (iii) the efficient information access to at least 10,000 healthcare centres; (iv) the privacy and privilege of health information.

To implement this NHIS, we divide our work into 3 main phases starting from the healthcare centre to the information consumer perspectives. The 1st phase focuses on the application of metadata standard to enable the interoperability and usability of health information across healthcare centres.

Basically, two significant desktop tools are developed: Metadata Mapping Tool (MMT) and Metadata Conversion Tool (MCT). MMT and MCT support healthcare centres to efficiently transform their information stored in any relational databases into a standard format with a small effort.

The 2nd phase moves forwards to make information sharing possible and to provide an efficient information access to a large number of healthcare centres. This underlying work is thus based on Web Services, XML, Grid and P2P technologies. Essentially, MCT is wrapped up as a service, namely Metadata Service (MS), to accommodate the availability of information sharing.

The MS connects to a specified relational database and transforms the underlying information into a standard format. Besides MS, a Metadata Broker (MB) as a service is developed in such a way that it not only provides the linkage to a potential set of MS's but also accesses and integrates their information based upon a request. To efficiently handle information access to a large number of healthcare centers, a number of MBs along with their communication are deployed and developed.

Finally, in the 3rd phase, we plan to promote the privacy and privilege of health information with respect to roles of information consumers. To accomplish this, MBs and MS's are extended to handle the dynamic delegation of access rights, single sign on, trust relationship among multiple entities, data privacy and policy related security issues.

This underlying work is thus based on the Public Key Infrastructure (PKI) and Privilege Management Infrastructure (PMI) standards. Particularly, Role Based

Access Control (RBAC) is chosen to implement PKI. The secure health information exchange between entities at the message level relies on Web Services security standards.

Currently, we are at the end of the 1st phase. The field evaluation of MMT and MCT will be conducted soon at about 20 healthcare centres that have at least 3 different health information structures. In conclusion, our work has significantly driven the “sustainable” NHIS due to its intrinsic properties: transparency, data security, availability, extensibility and scalability.

[based on presentations Sornthep Vannarat, NECTEC, Thailand]

## 14 VIETNAM

The advances in biomedical computing, the abundance of biomedical and genomic data, the ubiquity of the internet and general acceptance of Grid Computing is beginning to have a big impact in various aspects of medical, biological and health-care research and practice.

The sharing of knowledge and exchange of diagnosis between physicians can contribute to an improved standard of medical knowledge. Recently some institutions in Vietnam have set up a Grid infrastructure. By using web services technology and grid services provided by the gLite middleware, the HOPE telemedicine Platform that developed at LPC Clermont-Ferrand has been implemented on a grid site at Institute of Applied Mechanics and Informatics (IAMI) in Vietnam.

The physicians access the platform using web portal to access several distributed medical services that manage the traditional alphanumeric information such as patient personal data, diagnosis, results for analysis and investigation, medical images are stores anonymized and encrypted on the grid which their metadata are stored in the local AMGA server.

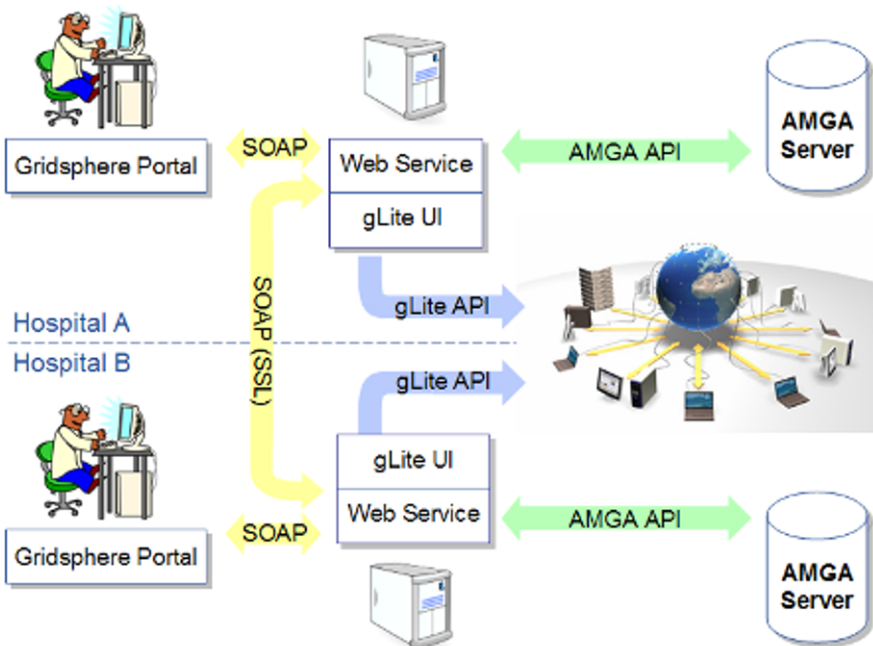


Fig.13: Shows concept of hospital open-software platform for e-Health (HOPE), with information transferred between different locations using SOAP messages over an SSL secured channel.



Currently Vietnamese researchers are also deploying the selected telemedicine applications using servers that connected to VinaREN network. IAMI has also cooperated with Pasteur Institute in HCM City to apply GVSS (GAP Virtual Screening Service) tool based on the power of Grid Computing (EUAsiaGrid) to solve docking problem (first stage in the drug discovery process) for Dengue virus.

The IFI (the Institut de la Francophonie pour l'Informatique – french-speaking Computer Science Institute) created in 2006 the MSI (Modelisation et Simulation des systems complexes) as its research team. MSI has been actively involved in EUAsiaGrid WP3 (Applications) for research, development, deployment, promotion and usage of the grid for biomedical applications. One of the objectives has been building a grid-based flu epidemic surveillance network on top of the EUAsia VO. This has involved building an information system based on AMGA which automatically synchronizes public influenza databases like NCBI with grid resources, as well as deploying a phylogenetic pipeline on the EUASIAGRID resources, and developing/adapting tools for visualization and analysis of obtained results from the pipeline. Developing a web portal for epidemiologists.

IFI is also participating actively in the proposal of CHAIN project as extension of EUAsiaGrid project. The objective would be to develop a simulation platform as an extension of the open-source GAMA platform, as a front-end for managing massive agent-based simulations based on the use of grid computing. Also, the ambition is to develop others grid-based applications in collaboration with national and international partners:

- Biomedical (LPC-IN2P3, HealthGrid, IBT-VAST)
- Earth Science (IGP-VAST, ASGC)
- Nuclear physics (IOP-VAST, IOIT-VAST, CNRS)

Finally, IFI is engaging with other national partners and with the support of the international grid community in order to propose and implemented the Vietnam NGI.

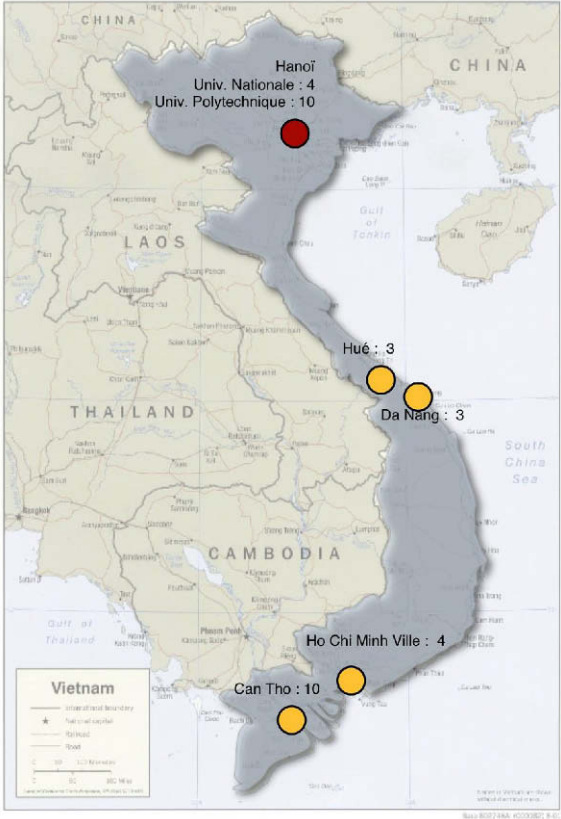


Fig.14: map indicating other centres in Vietnam that exchange researchers, Ph.D. students with IFI.

[based on presentation by Dao Van Tuyet, IAMI, Vietnamese Academy of Sciences, and Nguyen Hong Quang, IFI, Hanoi]