e-Science Introduction

Eric Yen

e-Science Workshop, March 2011
Outline

• Workshop Overview
• E-Science Basics
• Landscape of e-Science
• Application Development Concept
• Security Infrastructure
• Exemplar Applications
e-Science Workshop Overview

• Objectives
  • Help user communities to take advantage of the global DCI – World Wide Grid
  • Engage close collaboration among regional user communities and with the Grid community

• Target Audience
  • Both users and Grid/e-Science engineer
  • Of course, this is also good for novice to understand the e-Science, application development, related technology and the collaboration.

• Two workshop on Natural Disaster Mitigation and Life Science are arranged.
<table>
<thead>
<tr>
<th>Time</th>
<th>e-Science Application Workshop I: Natural Disaster Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30 ~ 09:45</td>
<td>e-Science in Asia</td>
</tr>
<tr>
<td></td>
<td>Simon C. LIN &amp; Eric YEN</td>
</tr>
<tr>
<td></td>
<td>Academia Sinica, TW</td>
</tr>
<tr>
<td>09:45 ~ 10:30</td>
<td>General Introduction to e-Science Applications</td>
</tr>
<tr>
<td></td>
<td>Simon C. LIN &amp; Eric YEN</td>
</tr>
<tr>
<td></td>
<td>Academia Sinica, TW</td>
</tr>
<tr>
<td>10:30 ~ 11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:00 ~ 12:30</td>
<td>e-Science Application I – Typhoon</td>
</tr>
<tr>
<td></td>
<td>Introduction of Computational Meteorology</td>
</tr>
<tr>
<td></td>
<td>Chuan Yao LIN</td>
</tr>
<tr>
<td></td>
<td>Research Center for Environmental Changes, Academia Sinica,</td>
</tr>
<tr>
<td></td>
<td>TW</td>
</tr>
<tr>
<td>12:30 ~ 13:00</td>
<td>gLite-based WRF</td>
</tr>
<tr>
<td></td>
<td>Yun-Pin SUN</td>
</tr>
<tr>
<td></td>
<td>ASGC, TW</td>
</tr>
<tr>
<td>13:00 ~ 14:00</td>
<td>Lunch (Recreation Hall, 4F)</td>
</tr>
<tr>
<td>14:00 ~ 15:30</td>
<td>e-Science Application II – Earthquake</td>
</tr>
<tr>
<td></td>
<td>Introduction of Computational Seismology</td>
</tr>
<tr>
<td></td>
<td>Li ZHAO</td>
</tr>
<tr>
<td></td>
<td>Institute of Earth Sciences, Academia Sinica, TW</td>
</tr>
<tr>
<td>15:30 ~ 16:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>16:00 ~ 16:30</td>
<td>gLite-based SPECFEM3D and Finite Difference</td>
</tr>
<tr>
<td></td>
<td>Yi-Tao HE</td>
</tr>
<tr>
<td></td>
<td>ASGC, TW</td>
</tr>
<tr>
<td>16:30</td>
<td>Discussion</td>
</tr>
</tbody>
</table>
# e-Science Application Workshop II: Life Science

**Date:** 20 March 2011

<table>
<thead>
<tr>
<th>Time</th>
<th>Conference Room 1, BHSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00 ~ 09:15</td>
<td>e-Science in Asia&lt;br&gt;Simon C. LIN &amp; Eric YEN&lt;br&gt;Academia Sinica, TW</td>
</tr>
<tr>
<td>09:15 ~ 10:00</td>
<td>General Introduction to e-Science Applications&lt;br&gt;Simon C. LIN &amp; Eric YEN&lt;br&gt;Academia Sinica, TW</td>
</tr>
<tr>
<td>10:00 ~ 10:30</td>
<td>Coffee break</td>
</tr>
<tr>
<td>10:30 ~ 12:00</td>
<td>e-Science Application I – Drug Discovery&lt;br&gt;Introduction of Modern Drug Discovery (Remote)&lt;br&gt;Yin-Ta WU&lt;br&gt;Genomics Research Centre, Academia Sinica, TW</td>
</tr>
<tr>
<td>12:00 ~ 12:30</td>
<td>Drug Discovery Tools and GVSS Production System&lt;br&gt;Hsi-Kai WANG&lt;br&gt;ASGC, TW</td>
</tr>
<tr>
<td>12:30 ~ 14:00</td>
<td>Lunch (Recreation Hall, 4F)</td>
</tr>
<tr>
<td>14:00 ~ 15:30</td>
<td>e-Science Application II – Next Generation Sequencing (NGS)&lt;br&gt;Introduction of Computational Biology&lt;br&gt;Yeng-Cheng YANG&lt;br&gt;Institute of BioMedical Informatics, National Yang Ming University, TW</td>
</tr>
<tr>
<td>15:30 ~ 16:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>16:00 ~ 16:40</td>
<td>Implementing iRODS for Next Generation Sequencing Data Management&lt;br&gt;Gen-Tao Mike CHIANG&lt;br&gt;Wellcome Trust Sanger Institute</td>
</tr>
<tr>
<td>16:40</td>
<td>Discussion</td>
</tr>
</tbody>
</table>
e-Science Basics

• “e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it. ... e-Science will change the dynamic of the way science is undertaken.”

• By John Taylor, former Director General of Research Councils UK

• Vision: a globally connected scholarly community promoting the highest quality scientific research

• e-Science refers to either computationally intensive science or data intensive science that is carried out in highly distributed computing environment.

• WLCG, EGEE, EGI, TeraGrid, OSG, EUAsiaGrid, …
e-Infrastructure/Cyberinfrastructure

- Driven by Data Deluge
  - Turning data into insight and knowledge base efficiently
  - Open, consistent and well-designed data format, interface, protocol and quality code
  - Searchability, accessibility and sustainability

- Resources and Tools are shared cross-disciplinarily

- Enable Service-Oriented Science
  - “scientific research enabled by distributed networks of interoperating services”
  - New e-Infrastructure is required to host both the data and services
Data Centric Sciences

- **Data Deluge:** going to Exa-scale Era
  - Data is inherently **distributed**
  - Data is produced in **large quantities**
  - Data is produced at a **very high rate**
  - Data is **needed** by many people

- More complicated data management required
  - Data has **complex** interrelations
  - Data has many free **parameters**
  - Data **Integration**
  - **Co-Scheduling, Streaming, Caching, & Replication**

- Mass Collaboration
- Large Scale Computing
The Changing Nature Of Research

**e-Science**

**Experimental**

- Thousand years ago
  - Description of natural phenomena

**Theoretical**

- Last few hundred years
  - Newton’s laws, Maxwell’s equations…

**Computational**

- Last few decades
  - Simulation of complex phenomena

**The Fourth Paradigm**

- Today and the Future
  - Unify theory, experiment and simulation with large multidisciplinary data
    - Using data exploration and data mining (from instruments, sensors, humans…)
    - Distributed communities

Simulation of complex phenomena

\[
\left(\frac{a}{a_0}\right)^2 = \frac{4\pi G \rho}{3} - K \frac{c^2}{a^2}
\]
## Terabyte → Petabyte (2008)

<table>
<thead>
<tr>
<th></th>
<th>Terabyte</th>
<th>Petabyte</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RAM time to move</strong></td>
<td>2.5 minutes</td>
<td>~2 days</td>
</tr>
<tr>
<td><strong>1GB WAN move time</strong></td>
<td>10 minutes</td>
<td>6 days</td>
</tr>
<tr>
<td><strong>Disk cost</strong></td>
<td>2 disks = $200 (SATA)</td>
<td>2000 Disks + 42 units + 5 racks = $500000</td>
</tr>
<tr>
<td><strong>Disk power</strong></td>
<td>20 Watts</td>
<td>50 Kilowatts</td>
</tr>
<tr>
<td><strong>Disk weight</strong></td>
<td>2 Kg</td>
<td>5.5 Tonnes</td>
</tr>
<tr>
<td><strong>Disk footprint</strong></td>
<td>Inside machine</td>
<td>4 m²</td>
</tr>
</tbody>
</table>

Source: P. Kunst et. Al, ADSSS, 2009
Distributed Computing Infrastructure for e-Science

- Enabling collaboration to realize that the whole is greater than the sum of parts
- WWG realized the global e-Infrastructure to share resources over Internet

- Cloud offers versatile granularity and new usage patterns to the DCI services
  - Granularity: service-oriented layers in infrastructure, platform, software, data, network, etc.
  - Usage pattern: on-demand elasticity
  - More user customized and user controlled environment on remote resources
The Grid

• “a software infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources”.

• Foster, Kesselman and Tuecke

• Features
  • No central control
  • Production quality
  • Open standards and open architecture
Data-Driven Multiscale Collaborations for Complexity

*Great Challenges of 21st Century*

- Multiscale Collaborations
  - General Relativity, Particles, Geosciences, Bio, Social...
  - And all combinations...
- Science and Society being transformed by CI and Data
  - Completely new methodologies
  - “The End of Science” (as we know it)
- CI plays central role
  - No community can attack challenges
  - Technical, CS, social issues to solve
- *Places requirements on computing, software, networks, tools, etc*
  
  Source: Ed Seidel

*Small groups still important!*
Embracing a new vision for Science

- Virtual Research Communities
- Improved scientific process... role of simulation
- Cross-disciplinarity
- Data deluge - Wet-labs versus ICT infrastructures
Connecting the finest minds

Sharing the best scientific resources

Building virtual global research communities

- Linking ideas at the speed of light
- Harnessing the unlimited power of computers, instruments and data
- Innovating the scientific process

e-infrastructure

géant | grids | scientific data | supercomputing
“We humans have built a creativity machine. It’s the sum of three things: a few hundred million of computers, a communication system connecting those computers, and some millions of human beings using those computers and communications.”

Future perspectives for e-Infrastructures

- e-Infrastructures in transition
  - Towards infrastructure-as-a-service
  - From connectivity and grids to an integrated offer involving networks, data, all computing and software
  - Progressive and disparate involvement of users
  - Governance and financial models in evolution
  - What role for innovation?
- More emphasis on Scientific Data Infrastructures
- International dimension continues to be important
- Enabling open Science, research and innovation
• Integrated Sustainable Pan-European Infrastructure for Researchers in Europe
• A 4-year project with €25M EC contribution
  – Project cost €70M
  – Activity cost ~€330M
• EGI – European Grid Initiative
  – Deploying Technology Innovation
    • Distributed Computing continues to evolve
      – Grids, Desktops, Virtualisation, Clouds
  – Enabling Software Innovation
    • Provide reliable persistent technology platforms
      – Today: Tools built on gLite/UNICORE/ARC
  – Supporting Research Innovation
    • Infrastructure for data driven research
      – Support for international research (e.g. ESFRI)

AS leads 10 Countries in Asia to join
European Grid Infrastructure

Status April 2010 (yearly increase)
- 10000 users: +5%
- 243020 LCPUs (cores): +75%
- 40PB disk: +60%
- 61PB tape: +56%
- 15 million jobs/month: +10%
- 317 sites: +18%
- 52 countries: +8%
- 175 VOs: +8%
- 29 active VOs: +32%

Taiwan is a partner of EGI, coordinating Asia
Pacific Grid Initiative (APGI) JRU

10-14-26 UTC (3 minutes ago)

11/01/2011
EGI-InSPIRE RI-261323
e-Science in Asia

- Diversity
  - Geographically large and culturally diverse in nature
  - Level of scientific collaboration often reflected by network connectivity
  - The region as a whole traditionally inexperienced in regional collaboration
- Grids and Clouds in Asia
  - Inhomogeneous Grids and Clouds with limited operations experience, making collaboration difficult.
- Why e-Science in Asia?
  - Global infrastructure is establishing quickly
  - Take advantage of sharing and collaboration to bridge the gap between Asia and the world
  - To address the challenge of regional cooperation

EGEE/EUAsiaGrid have helped building the unseen Regional Collaboration. One hopes many others will happen soon!
EUAsiaGrid

- Identify and engage scientific communities which can benefit from the use of state-of-art Grid technologies;
- Disseminate EGEE middleware in Asian countries by means of public events and written and multimedia material; 12 Countries, 15 partners, led by INFN and ASGC
- Provide training resources and organise training events for potential and actual Grid users;
- Support the scientific applications and create a human network of scientific communities by building on and leveraging the e-Science Grid infrastructure.

Excellent Progress Exceeded Expectations!

AS role was key for the project success.

Computational Chemistry
Social Science
Bioinformatics and Biomedical
High Energy Physics
Mitigation of natural disasters
Features of Distributed Applications

- Interoperability: Work across multiple distributed resources
- Distributed Scale-Out: Utilize multiple distributed resources concurrently
- Extensibility: Support new patterns/abstractions, programming models, functionality & infrastructure
- Adaptivity: Respond to fluctuations of dynamic resource and availability of dynamic data
- Simplicity: Accommodate distributed concerns at different levels easily

Source: SAGA
<table>
<thead>
<tr>
<th><strong>Middleware Services for Grid App</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Application</strong></td>
</tr>
<tr>
<td>GUI, CLI or Portal, application packages, together with client services</td>
</tr>
<tr>
<td><strong>Collective</strong></td>
</tr>
<tr>
<td>(application-specific)</td>
</tr>
<tr>
<td>Application specific services, such as checkpointing, job management, failover, staging, distributed data discovery and backup, and workflow engine, customized services, etc.</td>
</tr>
<tr>
<td><strong>Collective</strong></td>
</tr>
<tr>
<td>(Generic)</td>
</tr>
<tr>
<td>Resource discovery, resource brokering, system monitoring, community authorization, certificate revocation</td>
</tr>
<tr>
<td><strong>Resource</strong></td>
</tr>
<tr>
<td>Access to computation, data; access to information about resource matchmaking, system structure, status, and performance.</td>
</tr>
<tr>
<td><strong>Connectivity</strong></td>
</tr>
<tr>
<td>Communication (IP), service discovery, authentication, authorization, delegation</td>
</tr>
<tr>
<td><strong>Fabric</strong></td>
</tr>
<tr>
<td>Storage system, computers, networks, code repositories, catalogs</td>
</tr>
</tbody>
</table>

Need to explore in more detail the requirements and scientific workflows.
Grid and Cloud Logical Architecture

Dynamic Computing Model (Application Environment)

- Life Science
- Earth Sciences
- Environ. Changes
- Social Science
- Security, Information, Accounting & Monitor

EMI Stacks

- VM & Dynamic Resource Management
- Hardware Fabric
- Distributed Resource Management & Services
- Job Management Service
- Data Service

API
From the Basic Grid Use Case

- On behalf of an authorized user, AAI and single sign on services for versatile IaaS, PaaS and SaaS.
- Tools interrogate the information system, resource discovery or dynamic provisioning
- Locates an optimal execution resource, submits the job to the execution resource, which in turn interprets the submitted job description and locates and fetches the necessary input data from a remote storage - also on behalf of the user.
  - Resource on-demand provisioning model with customized application environment
  - Elastic and efficient resource matchmaking model according to user-defined metrics and requirements
  - Storage space and file system on-demand
  - Job overflow and scalable automatically
- Upon the completion, the newly created data is uploaded to a storage resource where this user is authorized (as a member of a Virtual Organization), registered in the necessary data indexing catalogs, and the job record is updated in the accounting and monitoring system.
  - Support streaming and minimization congestion, avoid duplicate transmission, by P2P technology
  - Enhanced accounting and monitoring system
  - Network virtualization
Common Application Requirement I

- Many areas of science could benefit from a common IT infrastructure to support multi-disciplinary and distributed collaborations Full usage of available resources
- Inter-Infrastructure migration: support transfer of data and cross-execution of jobs, including transportation of data, accounting, service availability information between infrastructures (Grids and Clouds, eg., from local infrastructure to national/global infrastructure)
- Effective resource match-making: collect information from sites and provide community based matchmaking services, based on information services such as GLUE, workload management and workflow engine.
- All requirements should cover computation, data, and networking services.
Besides being able to access information from different sites, to integrate, federate and analyze information from many disparate and distributed data sources and to access and control computing resources and experimental equipment at remote sites are all required.

Searching for new scientific tools

- Search, access, move, manipulate and mine distributed data repository
- Tools to create and maintain the distributed data repository (data structure, metadata, etc.)
- Driven mainly by the imminent deluge of data from new generation of scientific experiments and surveys (from petabyte towards exabyte)
- Also expedite the evolution of research infrastructure
To Manage Long-term Preservation

• Define desired preservation properties
  – Authenticity / Integrity / Chain of Custody / Original arrangement
  – Life Cycle Data Requirements Guide
• Implement preservation processes
  – Appraisal / accession / arrangement / description / preservation / access
• Manage preservation environment
  – Minimize costs
  – Validate assessment criteria to verify preservation properties
Security Infrastructure
Separating responsibilities

- **Single Authentication token** ("passport")
  - key issue: provide a persistent, trusted identifier
  - issued by a party trusted by all,
  - recognised by many resource providers, users, and VOs
  - satisfy traceability and persistency requirement
  - in itself does not grant any access, but provides a unique binding between an identifier and the subject

- **Per-VO Authorisations** ("visa")
  - granted to a person/service via a virtual organisation
  - based on the identifier
  - acknowledged by the resource owners
  - today largely *role-based access control*
    - but providers can also obtain lists of authorised users per VO,
  - can still ban individual users
  - most of the real *liability* and *responsibility* goes here
A Federation Model for Grid Authentication

- A Federation of many independent CAs
  - Policy *coordination* based on common minimum requirements (not ‘*policy harmonisation*’)
  - Acceptable for major relying parties in Grid Infrastructures

- No strict hierarchy with a single top
  - Leverage of national efforts and subsidiarity
  - Allow incorporation of many pre-existing CAs
Building the CA federation

- Providers and Relying Parties together shaped the common minimum requirements
  - Authorities compliant with minimum requirements (profile)
  - Peer-review process within the federation to (re) evaluate members on entry & periodically
  - Reduce effort on the relying parties
    - single document to review and assess for all Authorities
    - collective acceptance of all accredited authorities
  - Reduce cost on the authorities
    - but participation in the federation comes with a price

- ... the ultimate decision always remains with the RP
New CAs: the Accreditation Process

Accreditation Guidelines for EUGridPMA

Basic elements:

- Codification of procedures in a CP(S) for each CA
  - *de facto* lots of copy/paste, except for vetting sections
- Peer-review process for evaluation
  - comments welcomed from all PMA members
  - two assigned referees
- In-person appearance during a review meeting
- Accreditation after remaining issues are addressed (by e-mail)

*Discussions* remain important, as not all details are codified!

- Accreditation model for other PMAs typically embedded in their charter ...
- Periodic re-appearance and re-discussion are needed
The Grid security model

- Started to build an X.509 PKI in 2000 → IGTF (2005)
  - EU DataGrid, CrossGrid, LCG, EGEE, USA, Asia ...
- Single electronic ID to be used everywhere
  - All Grids, All VOs (but needs Trust)
- Single registration at Virtual Organisation (VO)
- Single Login (per session)
  - Requires (identity) delegation
- AuthZ attributes come from a VO authority
- Common security policies (JSPG)
- IGTF AuthN policies also essential for building trust
  - TAGPMA + EUGridPMA + APGridPMA
Policy Interoperability

• All about building TRUST – **mutual development** helps
• Wherever possible, JSPG aimed to
  • prepare **simple** and **general** policies
  • For primary stakeholders
  • But also of use to other Grid infrastructures (NGI's etc)
• Policies **augment** local security policies
• Common policy eases the problems of interoperability
• Users, VOs and Sites all accept the same policies during their (single) registration (with Grid or VO)
• Other participants know that their actions are already bound by the policies
  • No need for additional negotiation, registration or agreement
IGTF – International Grid Trust Federation

- common, global best practices for trust establishment
- better manageability and coordination of the PMAs

![Map of IGTF Partnerships]

<table>
<thead>
<tr>
<th>Partners</th>
<th>23/9</th>
<th>48/25</th>
<th>15/10</th>
<th>86/43</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Cert</td>
<td>~1,800</td>
<td>~4,850</td>
<td>1,607</td>
<td>~8,300</td>
</tr>
<tr>
<td>Host Cert</td>
<td>~4,000</td>
<td>~8,150</td>
<td>2,433</td>
<td>~14,500</td>
</tr>
</tbody>
</table>
Guidelines: common elements in the IGTF

- **Coordinated namespace**
  - Subject names refer to a unique entity (person, host)
  - Usable as a basis for authorization decisions

- **Common Naming**
  - *One-stop shopping* for all trust anchors in the federation
  - Trusted, redundant, sources for download

- **Concerns and ‘incident’ handling**
  - Guaranteed point of contact -> IGTF RAT
  - Forum to raise issues and concerns

- **Requirement for documentation of processes**
  - Detailed policy and practice statement
  - Open to auditing by federation peers
Guidelines

• Classic Secured X.509 CAs
  • Aimed at long-lived identity assertions
  • Identity vetting procedures
    • Based on (national) photo ID’s
    • Face-to-face verification of applicants via a network of Registration Authorities
    • Periodic renewal (once every year)
  • off-line signing key or HSM-backed on-line secured systems

• Short-lived Credentials Services (SLCS)
  • Leveraging either organisational IdMs or federations
  • Life time < 11 days

• Member Integrated CS (MICS)
  • Higher IdM quality required, for 13-months certs
  • Revocation requirements off-sets risks
Certificate Statistics

Graph showing certificate statistics from Oct. 2009 to Dec. 2010. The x-axis represents the months of the year, and the y-axis represents the number of certificates. The graph includes data for various categories such as PRAGMA-UCSD CA, IHEP, NGO-Nettrust CA, NECTEC GOC CA, NCHC Grid CA, NAREGI CA, KISTI Grid CA, IGCA, HKU Grid CA, SDG CA, CNIC Grid CA, KEK, ASGC CA, APAC Grid CA, and AIST GRID CA. The Host Cert category is also shown with a red line.
The Worldwide Trust Framework is in Production since 2005, serving as the foundation of DCI

No Need to re-Invent the Wheel
EUAsiaGrid/ EGEE/ EGI was Facilitating Regional Collaboration and Bridging Asia with the World
# e-Science Collaborations in Asia

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Applications</th>
<th>Partners</th>
<th>Going DG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEP</td>
<td>ATLAS, CMS, ALICE, BELLE, CDF, GEANT4</td>
<td>TH, TW, CESNET, INFN</td>
<td>X</td>
</tr>
<tr>
<td>BioMedical</td>
<td>Virtual Screening for Drug Discovery – Avian Flu, Dengue Fever</td>
<td>MY, TW, VN, CESNET, INFN</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Pandemic disease analysis</td>
<td>VN, FR</td>
<td></td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>Grid enabling phylogenetic inference</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SVM Parameter optimization for prediction of Caspases</td>
<td>SG, TW, VN, CESNET, INFN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genome search to identify T3SS effect</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Autodock ligand-receptor docking</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Complex diseases studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earth Science</td>
<td>Disaster Mitigation on Earthquake</td>
<td>ID, MY, PH, TH, VN, TW, CESNET, INFN</td>
<td>X</td>
</tr>
<tr>
<td>Comp Chemistry</td>
<td>Chemical compound property analysis</td>
<td>TH, TW, CESNET</td>
<td>X</td>
</tr>
<tr>
<td>Climate Change</td>
<td>Weather simulation, sea level rising</td>
<td>ID, PH, TH, VN, TW</td>
<td></td>
</tr>
<tr>
<td>Social Sci.</td>
<td>Social Simulation</td>
<td>TW, UK</td>
<td>X</td>
</tr>
</tbody>
</table>
# Application Repository

## Applications

Click on the application name for details.

<table>
<thead>
<tr>
<th>APP. NAME</th>
<th>DOMAIN</th>
<th>COUNTRY / INSTITUTION(s)</th>
<th>STATUS</th>
<th>VIDEO?</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMBER</td>
<td>Chemistry</td>
<td>(NECTEC)</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>APBS</td>
<td>Chemistry</td>
<td>(ASGC)</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>Autodock</td>
<td>Biomedicine</td>
<td>(ASGC) (NECTEC)</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>BEAST</td>
<td>Bioinformatics</td>
<td>(ASGC)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>BLAST</td>
<td>Bioinformatics</td>
<td>(ASGC)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>DiffSST</td>
<td>Climate</td>
<td>(HAI)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>g-INFO</td>
<td>Biomedicine</td>
<td>(IFI) (CNRS) (HealthGrid)</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>GAP</td>
<td>Information Technology</td>
<td>(ASGC)</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>Gaussian</td>
<td>Chemistry</td>
<td>(NECTEC)</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>GROMACS</td>
<td>Biomedicine</td>
<td>(ASGC) (NECTEC)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>GVSS</td>
<td>Biomedicine</td>
<td>(ASGC)</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>HOPE</td>
<td>Biomedicine</td>
<td>(HealthGrid) (IAMl)</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>LIBSVM</td>
<td>AI / Optimization</td>
<td>(ASGC)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>MM5</td>
<td>Climate</td>
<td>(ASTI)</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>R</td>
<td>Environmental Sciences</td>
<td>(NDHU) (ASGC)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>RePast</td>
<td>Other</td>
<td>(UniStAnd) (ASGC)</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>TURBOMOLE</td>
<td>Chemistry</td>
<td>(NECTEC)</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>VASP</td>
<td>Physics</td>
<td>(ASGC)</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>WRF4G</td>
<td>Climate</td>
<td>(ITB) (HAI)</td>
<td>3</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Total of Applications: 19**

Application Status: S1 (in consideration), S2 (running but not ported to gLite yet), S3 (ported to gLite, unavailable in EUAsia VO), S4 (available in EUAsia VO), S5 (ready for production).
Convenient access to grid infrastructures for individual users

Provides, through the portal interface, support to:

- Submission of jobs
  - Specific forms for individual applications
  - Helping to prepare the job description and input data
- Data management
  - Allow sharing with other users
- Job Monitoring
- Life Sciences
  - Autodock 4, Beast, Blast, Gromacs, MrBayes, Muscle, Prodist
  - GVSS*
- Earth Science: Earthquake*
- Weather Simulation: WRF*
- Statistics: R
- Other User Defined Applications
Roadmap for AP region

- Detailed and improved analysis of the survey results to define the requirements from scientific communities and the resource provision from project partners
- APGI-Union based on individual resource-providing institutions proposed as an interim structure to overcome the absence of mature NGI in all the countries
- Operation towards a sustainable, scalable, persistent e-Infrastructure, easy to use and embedded in the research environment is coordinated by APROC incorporated with EGI.
- Asia regional infrastructure is compliant with EGI structure and similar federated initiatives in the rest of the world (e.g. Latin America)
Exemplar Applications
Bio-Portal and Virtual Screening Services

Best Demo Award of EGEE’07 Conference

Avian Flu Drug Discovery

Interactive scoring Visualization

DIANE Master Process

Docking task pulling Docking complex returning

Virtual Cluster (DIANE workers)

Resource Broker

Grid Application Portal

EGEE Grid Resources

The 8 Avian Flu Mutations were proposed by GRC, AS
Virtual Screening Service by AutoDock

- One-click job submission
- View the best conformation of a simulation
- Submit the docking job to the Grid with just one click
- Generate the histogram with a given energy
- Visualize your job status

SG + DG
# Dengue Fever Data Challenge in 2009

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of completed docking jobs</td>
<td>300,000</td>
</tr>
<tr>
<td>Estimated needed computing power</td>
<td>4,167 CPU*days</td>
</tr>
<tr>
<td>Duration of the experiment</td>
<td>60 days</td>
</tr>
<tr>
<td>Cumulative computing results</td>
<td>42.5 GB</td>
</tr>
<tr>
<td>Total Computing Resources in EUAsia VO</td>
<td>268 Cores</td>
</tr>
<tr>
<td>Number of used Computing Elements</td>
<td>6</td>
</tr>
</tbody>
</table>

Collaborators: UPM, MIMOS, MY, IAMI, VN; HAI, TH, Cesnet, CZ; GRC, TW.
e-Science for Earthquake Disaster Mitigation

Collaborators: PH, VN, TW, ID, MY, TH

Seismic Sensor Networks

Global/Regional Sensor Data

Ref. Historical Events Data

Local Sensor & Observation Data

High Resolution Source & Rupture Process Analysis

Archive

Fast Reporting System

Forward Simulation & Event Construction on Grid

Risk Analysis & Reduction

(e) Forward 3D wave propagation simulation

(c) Three-dimensional velocity model

(b) Three-dimensional source model

(d) Parallel computing and PC cluster

Local Sensor & Observation Data

Global/Regional Sensor Data

Ref. Historical Events Data

Archive

Virtual Seismic Network

Earthquake Data Center (SeisGrid)

Forward Simulation & Event Construction on Grid

Collaborators: PH, VN, TW, ID, MY, TH

e-Science for Earthquake Disaster Mitigation
Seismic Wave Propagation Analysis

• SWPA Application Porting and Data Challenge
  • Porting of both SPECFEM3D and FDM packages
  • Verification of application environment at each partner sites
  • Finish the simulation (from 1 to 144 cores) by end March’10

• Seismogram Simulation Application Development
  • Develop Simulation Wizard for the whole process
    • Location and tomography model: global, South East Asia and Taiwan
    • Earthquake source
      • Upload Event Source
      • Search from Global CMT Historical Data (and Map Visual.)
  • Seismogram Visualization at Stations
    • From established station or any identified virtual sites
1. Location and Tomography Model Selection

Welcome!
This wizard will help you set up simulation parameters:
- Location and tomography model
- Earthquake Source (CMTSOLUTION format)
- Stations

Available models:
- Global (ID_isotropic prm)
- Global (ID_isotropic prm)
- Southeast Asia (ID_isotropic prm)
- Taiwan (ID_isotropic prm)

2. Epicenter Data Preparation

3. Choose Position for Seismogram

4. Seismogram Access & Visualization
Future Works – Hazard Maps

• Achieving full process of quantitative seismic hazard assessment
  • Collecting and analyzing event data
  • Understanding fault characteristics in details
  • Facilitating accurate simulation on seismic waves
  • Assessing anticipated earthquake and potential damages by the correct seismic and engineering models
• Maps of disaster coverage, risk and also evacuation are pragmatic to better preparedness
Environmental Changes

• Support researches on understanding of the changing world
  – Probing the warming world
  – Investigate impacts of extreme weather

• WRF over gLite is available since March 2010

• Focus on Meteorological Researches in Taiwan and South East Asia
  – Heavy rainfall system during Mei-Yu season
  – Typhoon route and precipitation
  – East Asian Climate
  – Climate change in Mekong River Basin
  – Landslide modeling

Collaborators:
HAII, TH
EUAsiaGrid
RCEC, AS, Taiwan
Typhoon Morakot (2-11 August, 2009)

Highest winds: 140 km/h (10-min sustained)
Fatalities : 789 total
Rainfall : 2777 mm (total)
Damage : $6.2 billion (2009 USD)

Jhiben (知本), Taitung, Taiwan

Xiaolin village (小林村), Taiwan
Isobaric Contour (Morakot)
Simulation during (2009-08-05-12 ~2009-08-07-12)

Central pressure deviation: \( \frac{(980-956)}{980} = 2.5\% \)
Precipitation (Morakot) Simulation (2009-08-06-00:00~2009-08-07-12:00)

WRF Simulation

CWB history data
Social Simulation

- Project on population migration simulation from 2010
  - TW-UK Collaboration
  - Porting the UK-based Migration Model to gLite/EUAsiaGrid
  - Customization of the model for/of Taiwan
    - Taking into account the birthrate, fertility, and mortality
    - Deploy the local model based on regional researches
  - Feedback cycle for model verification:
    - Based on the real Census data of Taiwan
  - Deployment of agent-based modeling/simulation methods
- Further extension
  - Financial model
  - Social changes
- Collaborators: U. Manchester; U. St. Andrews; Survey Research Center, AS; EUAsiaGrid

Social Resilience in the future!
Conclusions
Natural Disasters are Regional Issues

- Earthquake, Tsunami, Typhoon, Flood, Pandemic are regional issues and cannot be dealt with by individual countries alone.
- It takes experts from different scientific disciplines, simulation, networking, computing resources, grids and clouds to mitigate the disasters.
- Detailed, quantitative scientific understandings are becoming possible.
- We are building a bottom-up SE Asia regional collaboration with the help of EU e-Infrastructure projects.
Support from Global e-Infrastructure is Critical

- Most of the existing regional collaborations on the above areas are in lack of the bottom-up approach taken by the EU-Asia e-Infrastructure projects.
- Bottom-up approach enables unprecedented collaboration which may raise the general standard of the academia communities in Asia.
- Interdisciplinary nature will lead to new scientific findings of disaster mitigation.
- However, continuing support from advanced countries such as EU and leading countries in the region is still required in order to reach the critical point.
e-Science for the Masses

• Not only porting scientific applications to e-Science collaboration, but also establishing research oriented production services and long term scientific collaboration among partners
• Unique scientific values of e-Science Application Data, e.g.
  – LHC data, unprecedented energy frontier, new fundamental understanding of the Universe
  – Earthquake data, first-principle simulation, archival and re-use
  – Drug Discovery data, neglected diseases information, open access and generating more knowledge
  – Regional collaborative data often related to Disaster Mitigation
• Common concerns such as Disaster Mitigation address the challenge of regional cooperation
• Take advantage of sharing and collaboration to bridge the gap between Asia and the world, an opportunity to leapfrog
• However, one must reduce the entry barriers for e-Science in Asia
• In Asia, e-Science for the masses is more strategic than the big science!