EUAsiaGrid

Marco Paganoni
INFN, Italy

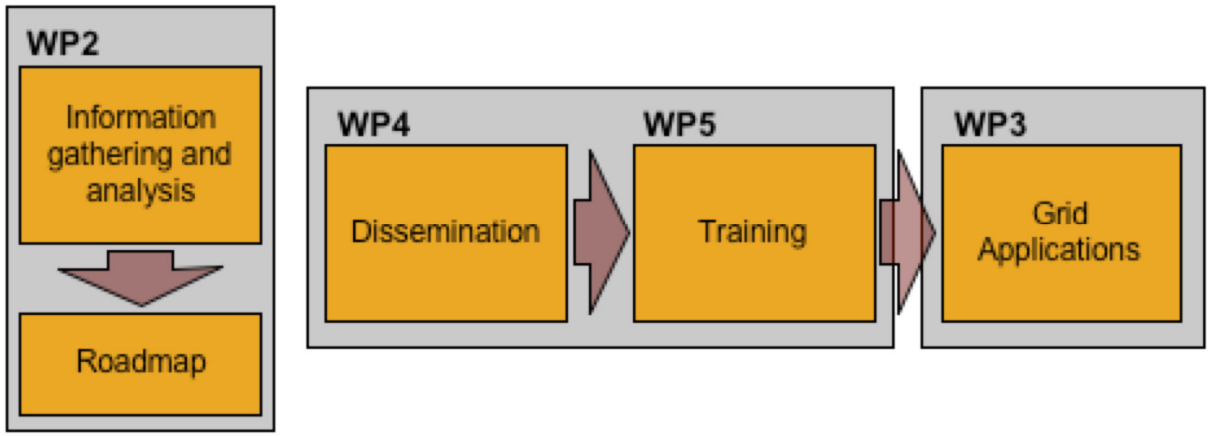
EGEE09 conference
Barcelona, Sep. 23rd, 2009

Partners

1. Istituto Nazionale di Fisica Nucleare (Italy) (coordinator)
2. CESNET (Czech Republic)
3. University of Manchester (United Kingdom)
4. HealthGrid (France)
5. Ateneo de Manila University (Philippines)
6. Australia National University (Australia)
7. Academia Sinica (Taiwan)
8. Advanced Science and Technology Institute (Philippines)
9. Hydro and Agro Informatics Institute (Thailand)
10. Infocomm Development Authority (Singapore)
11. Institute of Information Technology (Vietnam)
12. Institute Teknologi Bandung (Indonesia)
13. National Electronics and Computing Technology Center (Thailand)
14. University Putra Malaysia (Malaysia)
15. MIMOS Berhad (Malaysia)

Breakdown in work packages

Work Package	Title	WP Coordinator
WP1	Project management	INFN (M. Paganoni)
WP2	Requirement capture and coordination policy definition	UNIMAN (A. Voss)
WP3	Support of scientific applications	CESNET (L. Matyska)
WP4	Dissemination	ASGC (V. Huang)
WP5	Training	INFN (M. Fargetta)



Challenges of the Project

First Grid project targeting Asia-Pacific region

- ✓ Geographically large and culturally diverse area
- ✓ Uneven levels of adoption of Grids

Wide range of scientific domains addressed

- ✓ Consolidate on traditional areas and engage new communities
- ✓ Dissemination, Training and Support for applications must create a virtuous cycle

Sustainability of the e-infrastructure

- ✓ Need to define accurately a roadmap for the future (NGI ?)
- ✓ Need to train the trainers to trigger teaching from local groups

Project impact in the 1st year

Results from SPR and Survey

- ✓ First assessment in the AP region of requirements from scientific communities and resources from project partners

"Catch all" VO allows wide area outreach

- ✓ Crossing the borders between nations and disciplines

Letters of support from stakeholders in Taiwan, Thailand and Malaysia

- ✓ Underlining the engagement from partners

All groups of people (basic users, system administrators, application developers) addressed with training

- ✓ Events directly organized by the project and steering of local events

Starting point for 2nd year

Valuable data challenges achieved or launched

- ✓ Avian Flu DC2 refinement (results presented at ISGC09)
- ✓ Earthquake simulation
- ✓ Dengue fever drug discovery
- ✓ Not only single applications but also complex workflows supported

High value of the potential results

- ✓ Addressing rare and neglected deseases
- ✓ Mitigating natural disaster human and economic effects

Interesting new scientific domains joining the effort

- ✓ Digital Culture and Heritage (triggered by Malaysia and Taiwan)
- ✓ Interest in Engineering and Physical Sciences (coming out from the survey)

Roadmap definition and implementation for sustainability

- ✓ First workshop on July 25th in Kuala Lumpur, co-located with APAN28 and EUAsiaGrid CB
- ✓ Outline roadmap structure and discuss sharing of work
- ✓ Follow-up by teleconference over summer and 2nd roadmap meeting in Sep., to reach consensus on a final roadmap by end of project
- ✓ Issues for the roadmap
  - ❑ **Policy Drivers**
  - ❑ **Funding Streams and Mechanisms**
  - ❑ **Stakeholder Analysis**
  - ❑ **Exisiting Initiatives and Infrastructures**
  - ❑ **Coordination Mechanisms**

Future Plans / WP3

Continue support for the identified communities

- ✓ Focus on the benefit of international collaboration (EU – AP)

Identify new areas by analysis of Survey full results

- ✓ Plan special attention to the Physical/Engineering areas
- ✓ Elaborate Natural Disaster Mitigation subtask
- ✓ Dengue Fever Drug Discovery data challenge
- ✓ Earthquake simulation data challenge

Work with WP5, for the planning of training events

- ✓ “Keep an eye” on trained scientists
- ✓ Support trainers

Interoperability with desktop-computing

- ✓ Huge interest, low entry and operation costs
- ✓ Some doubts on usability for mission critical deployment

Supported Applications

- **Applications in the DoW**
  - Computational Chemistry
  - Social Science
  - Bioinformatics and Biomedical research
  - High Energy Physics
  - Mitigation of natural disasters
- **Newly identified areas**
  - Digital culture and heritage
  - Weather forecast and climatology
  - Mathematical modelling
  - Biodiversity

Avian Flu DC2 Refined

- *post-processing of high-score compounds identified during the EGEE DC2 activity*
- *designed and performed purely by EUAsiaGrid user community*
- *test case for utilization of established EUAsia VO and GAP (Grid Application Platform) package developed by ASGC for virtual screening purpose*
- *first key milestone for fostering the e-Science applications in Asia*

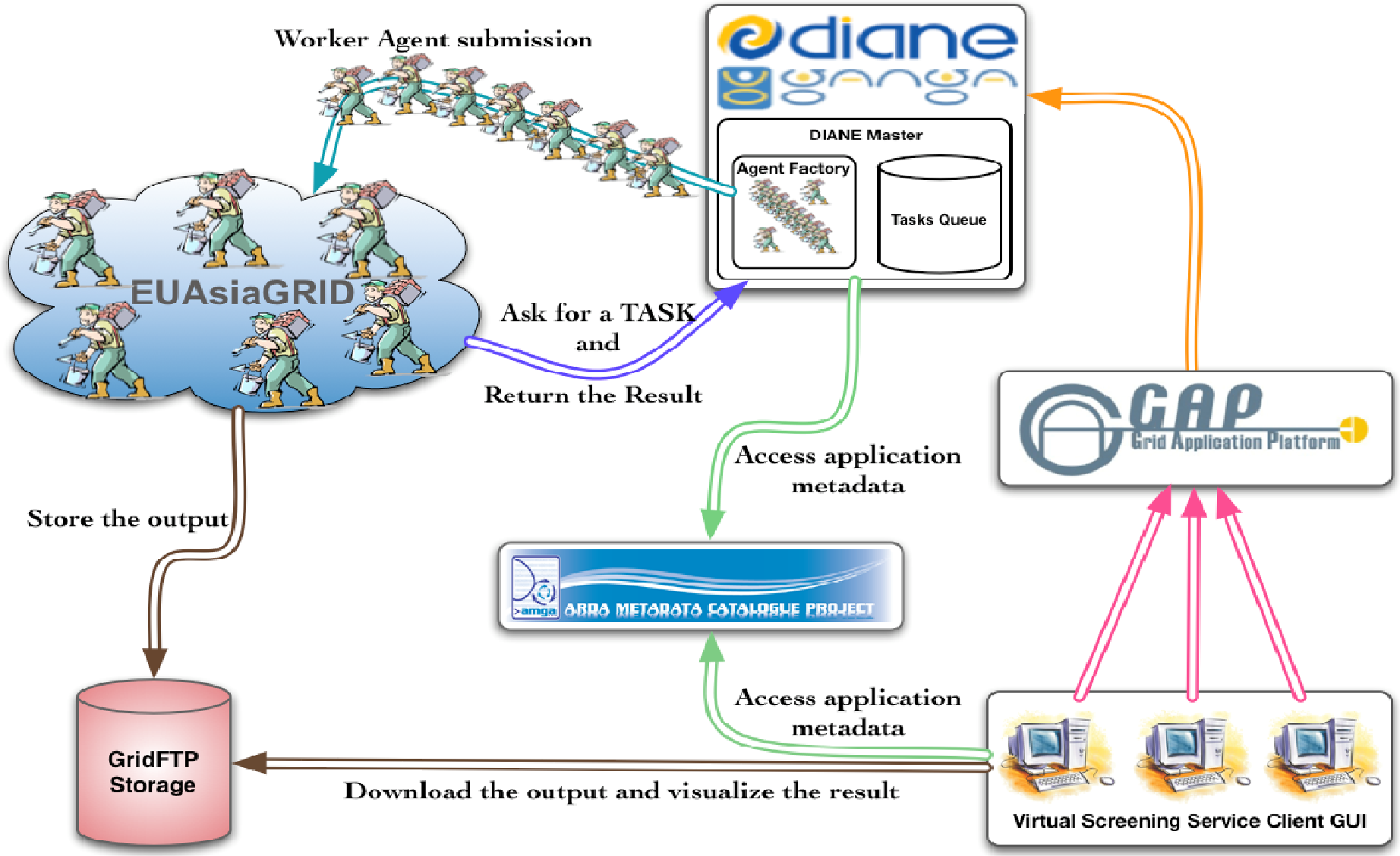
Dengue Fever Data Challenge

- **Objective**
  - ✓ Reduce time and money for drug discovery by Grid in silico simulation
  - ✓ 2.5B people are at risk, 50M cases/year in more than 100 countries
  - ✓ 95% cases are children < 15 yrs in South-East Asia
- **Milestones**
  - *Validate the Pilot run – M15*
  - *Data Challenge on Dengue Fever (NS3 protease structure with ZINC compound library) – M18*
  - *Pilot new designed target structure for in silico simulation – M18*
  - *Post data analysis for the NS3 target – M20*
  - *1st step wet-lab study – M22*
  - *Deploy the VSS on Asia@home – M18*
  - *Enhance the GVSS with APBS – M22*

Dengue Fever Data Challenge /2

- **Grid Computing environment**
  - *EuAsia VO*
  - *GAP Virtual Screening Service (GVSS)*
  - *ASGC provides the core services: GAP-VQS, DIANE2, AMGA, SE*
- **User requirements**
  - *CA*
  - *EuAsia VO*
  - *GAP VQS account*
  - *Download and install GVSS client package*
  - *Registered CDI compounds' data sets*

Dengue Fever Data Challenge / workflow

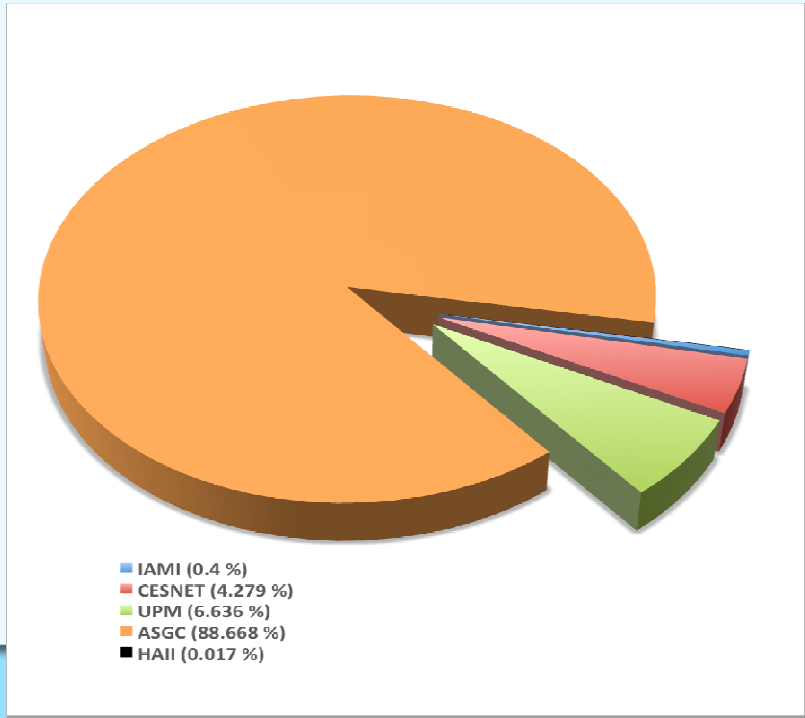


Dengue Fever Data Challenge /resources

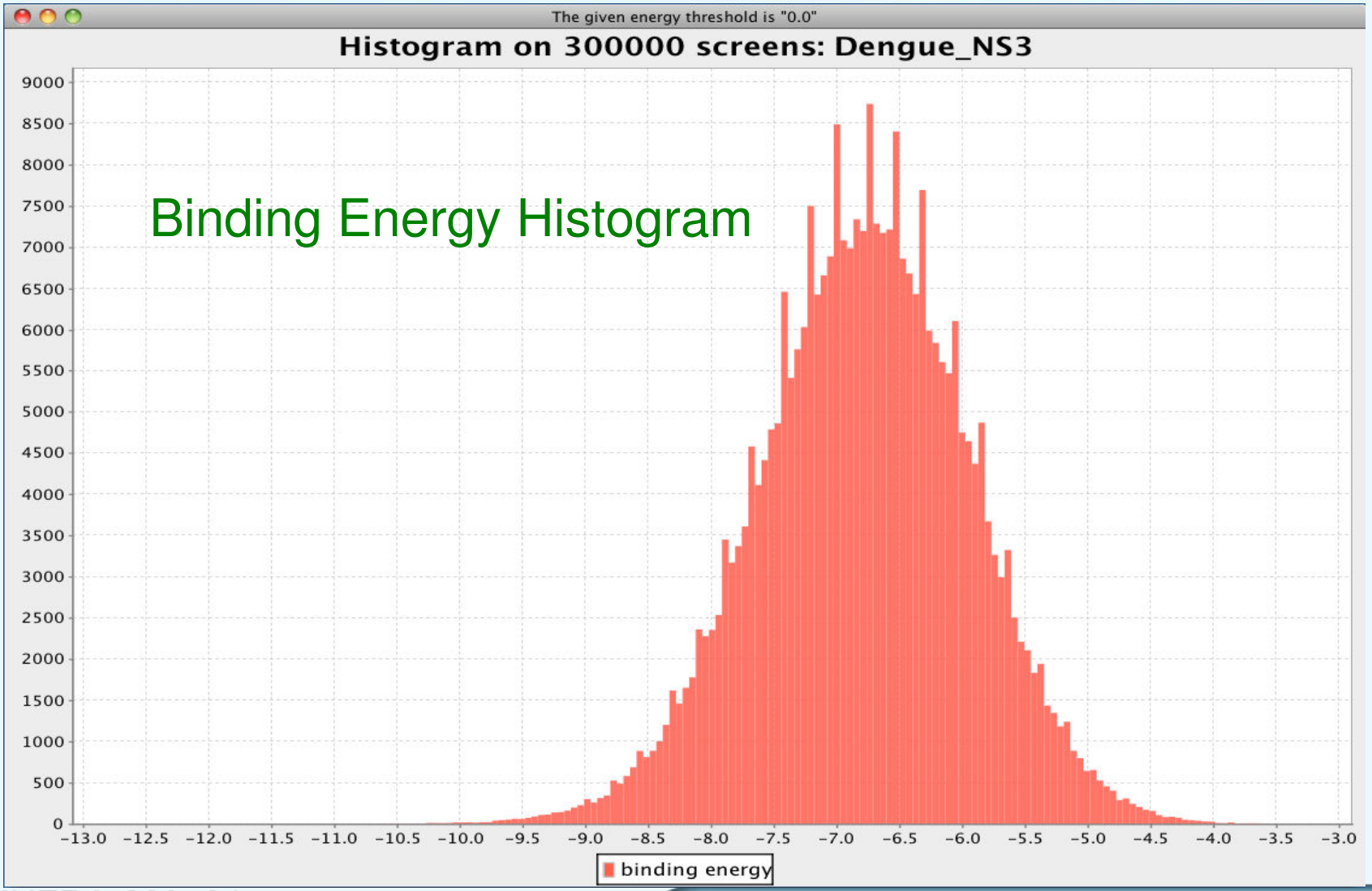
- Accumulating Computing Recourses in EUAsia VO:
  - 268 cpu-cores(100 – ASGC(TW), 2 – TH, 4 - VN, 18 – MIMOS(MY), 80 – UPM(MY), 64 - CESNET(CZ))
  - lcg-infosites --vo euasia ce
- Registered VQS account:
  - 6 users (TW)
  - 17 user (PH, 15 in AdMU, 2 in ASTI)
  - 2 user (TH, 1 in NECTEC, 1 in HAII)
  - 1 user (MY, UPM)
  - 1 user (ID, ITB)
  - 2 user (VN, IAMI)
  - 1 user (FR, HealthGrid)

Dengue Fever Data Challenge / statistics

Total number of completed docking jobs	300,000
Estimated needed computing power	4,167 CPU*days
Duration of the experiment	60 days
Cumulative number of Grid jobs	4,015
Maximum number of concurrent CPUs	150
Number of used Computing Elements	6



Dengue Fever Data Challenge / results



Earthquake simulation

- **Earthquake sensor network**
  - real-time earthquake data collection
  - collaboration with currently available facilities local or international ones
- **Seismic wave propagation analysis**
  - integration of analysis model
  - accurate topographical / geological data
- **Earthquake data center**
  - extension/integration/coordination with SeisGrid in Taiwan

Drug Discovery - DC

1. Avian Flu - DC2 refined
- 8 avian-flu mutant targets from EGEE DC2
- 20,000 highest scored ligands from EGEE DC2 results

2. Dengue fever
- Dengue NS3 target
- 300,000 ligands which are prepared from ZINC.

	Number of dockings	CPU time	CPU-cores used on EUAsia VO	Produced data size	Status
1	20,000	3 years	125	125 GB	Completed
2	300, 000	12 years	125		Completed

Biomed - H1N1 Surveillance

NCBI

AMGA

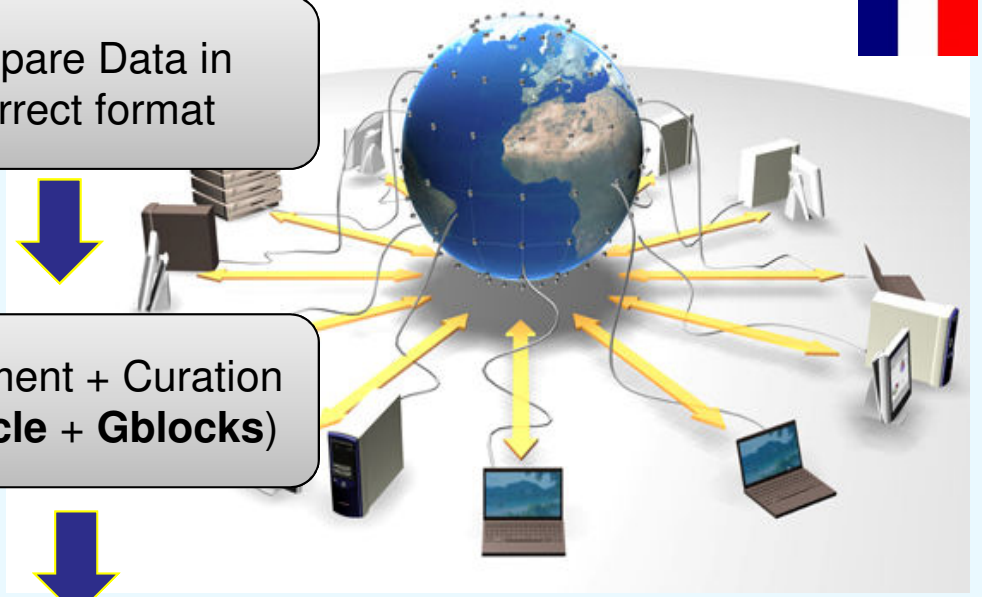
- Metadata
- Sequences
  Protein, Nucleotide, Coding region
- IDs

> Run daily a phylogenetic workflow on the grid

Prepare Data in correct format

Alignment + Curation (**Muscle** + **Gblocks**)

Phylogenetic Analysis (**PhyML**)



grid-based International Network for Flu Observation
g-INFO

DEMO on EUAsiaGrid Booth

Biomed - My STAR

1. Explorative collaboration
   - to develop and deploy production grid and services in Malaysia
   - For e-Health; IPTV and Renewable Energy

2. Funding Request being submitted to Malaysian funding agency (MDeC)

3. Private Company under incorporation

Earthquake Simulation

Objectives

- ✓ Dense sensor networks + historical data + simulation model → accurate simulation
- ✓ Seismic wave propagation analysis to reduce potential impacts
- ✓ Planning for real time hazard mitigation

Methodology

- ✓ Realize regional seismic wave simulation model over gLite mw
- ✓ Validate accurate wave simulation country by country, and implement case studies on disaster mitigation
- ✓ Seismic sensor networking and data federation

Roadmap

- ✓ Porting on Grid of seismic wave propagation analysis – M18
- ✓ Case Studies – M22
- ✓ Feasibility Analysis for tsunami disaster mitigation – M23

Earthquake Simulation / collaboration

	Sensor Network	Seismic Wave Prop. Analysis	Seismic Data Center
Partners	VN, PH, ID, TW		
User Community	Philippine Institute of Volcanology and Seismology (PHIVOLCS), Vietnamese Academy of Science and Technology, The Incorporated Research Institutions for Seismology (IRIS), Global Seismic Network (GSN), Institute of Earth Science & National Central University, Taiwan, Local, Regional, and Global Disaster Mitigation Organization.		
Tech Maturity	TW- Most dense; VN- Sensor Stn ready; PH- expanding	Analysis Model and knowledge available; Cluster and gLite Resources in place; From Global model toward higher resolution regional/country model;	SeisGrid@TW, IRIS, GSN
Exemplar	Integrated Sensor Network by VN, PH and TW	Without local geological data, accurate analysis is not achievable. High resolution historical TW earthquake data sets.	Federation of available Data Centers

Earthquake Simulation / Plans

Facilitate understanding of ground motion mechanism, rupture process, velocity structure and topographic characteristics

GeoGrid framework as the basis

- ✓ *gLite + GAP + DIANE + GANGA*
- ✓ *Portal for data & analysis services*

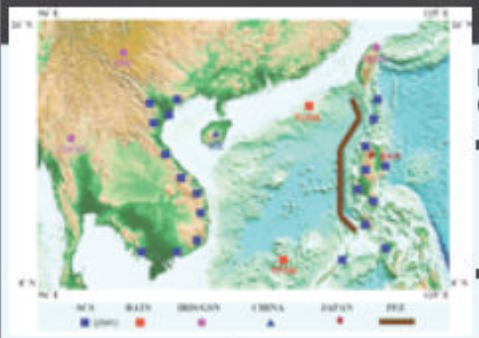
Earthquake Monitoring & Simulation

- ✓ *Earthquake Data Center prototype*
- ✓ *Resource Estimation: 100 CPU*month, 10-30 TB*

Collaboration also with EU experience (Italy, Greece)

e-Science for Earthquake Disaster Mitigation

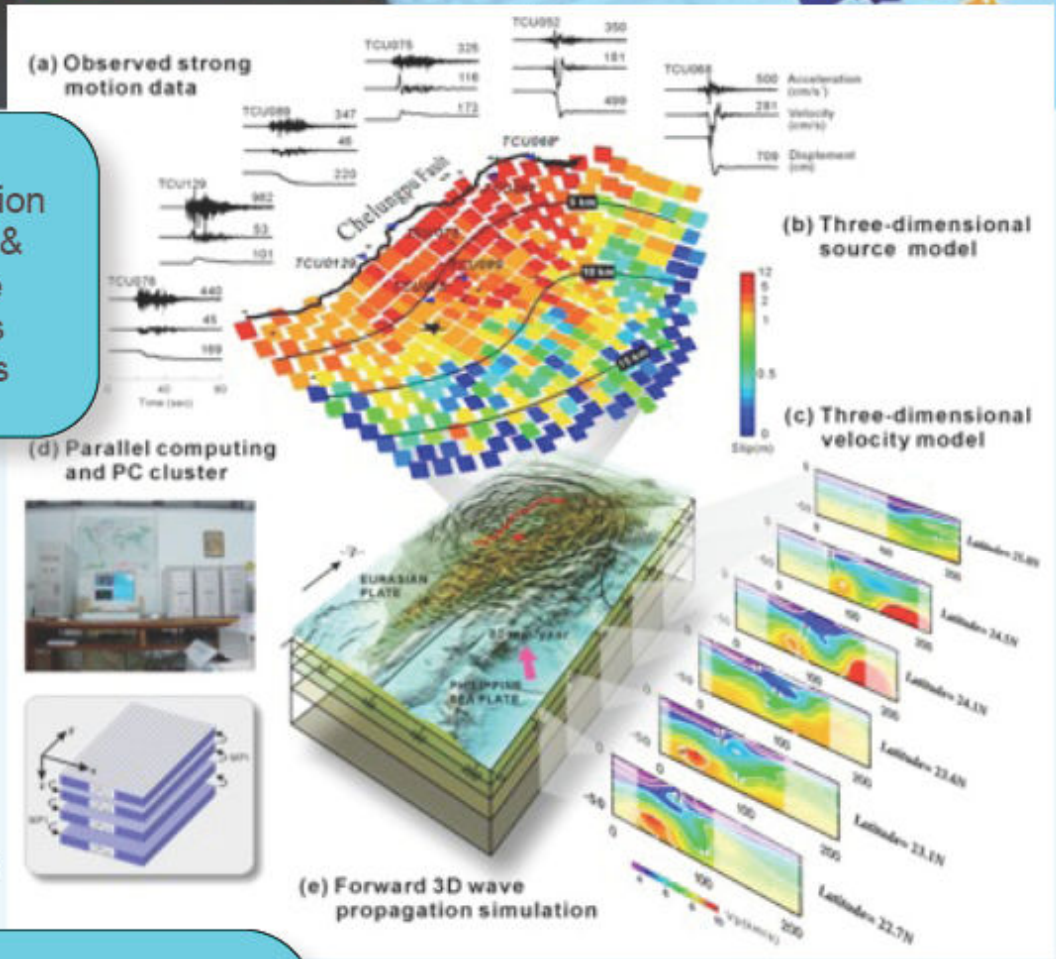
Seismic Sensor Networks



Local Sensor & Observation Data

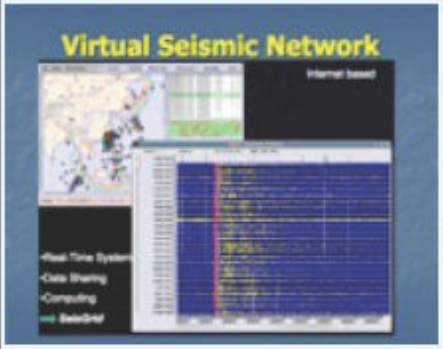
Global/Regional Sensor Data

High Resolution Source & Rupture Process Analysis



Fast Reporting System

Archive



Earthquake Data Center (SeisGrid)

Ref. Historical Events Data

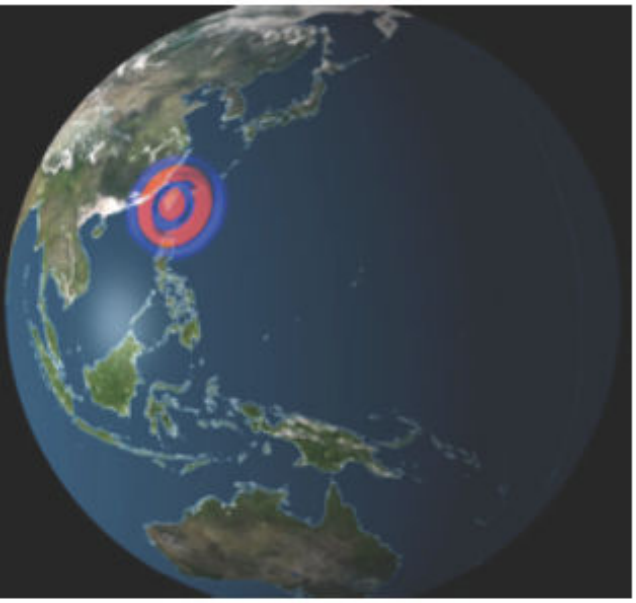
Archive

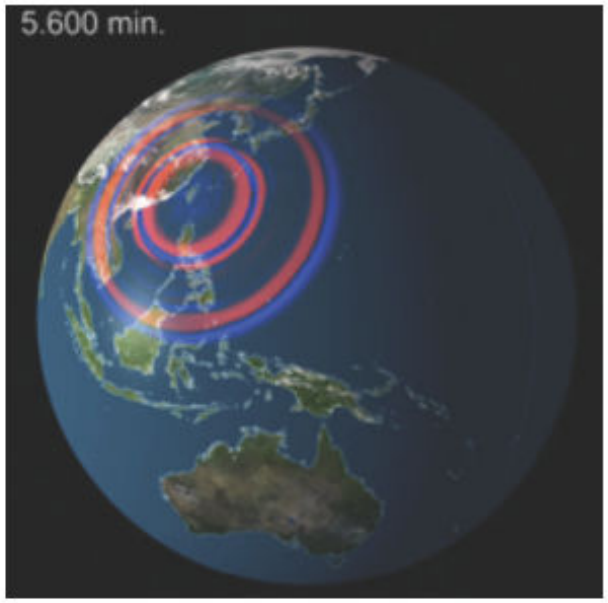
Forward Simulation & Event Construction on Grid

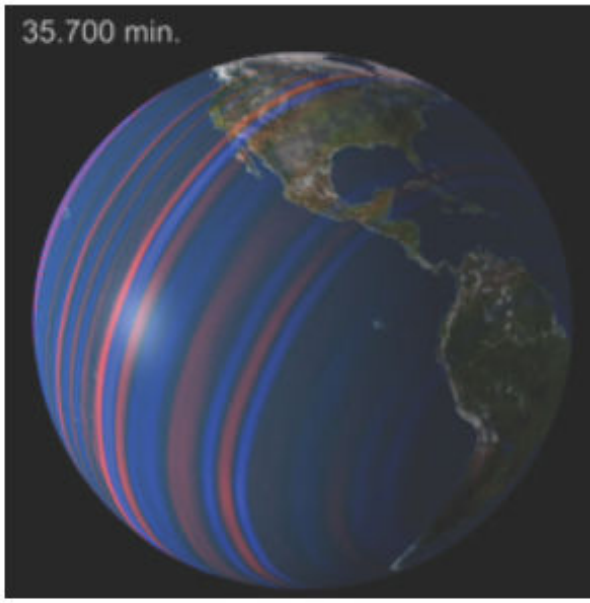
Risk Analysis & Reduction

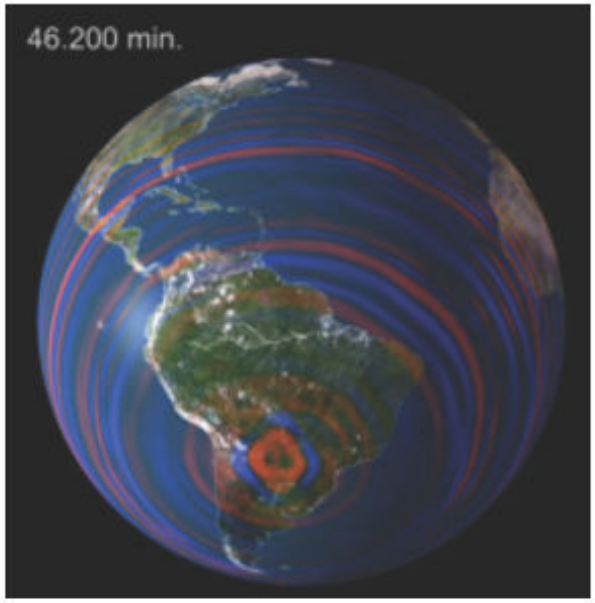
FP7-INFRA-223791

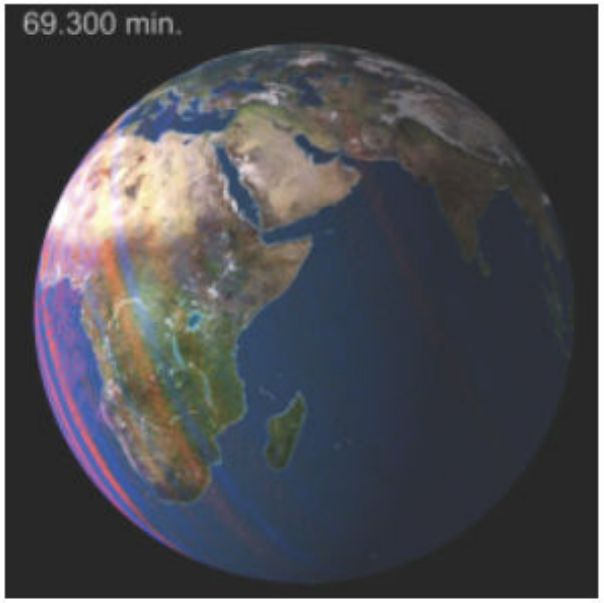
Seismic Wave Propagation Analysis











Planned activities from WP4

Production of new dissemination materials

- ✓ EUAsiaGrid Applications Leaflet

Update and maintain dissemination materials

- ✓ Brochure, Fact Sheet, Poster and Presentation template
- ✓ EUAsiaGrid websites (public/wikis)

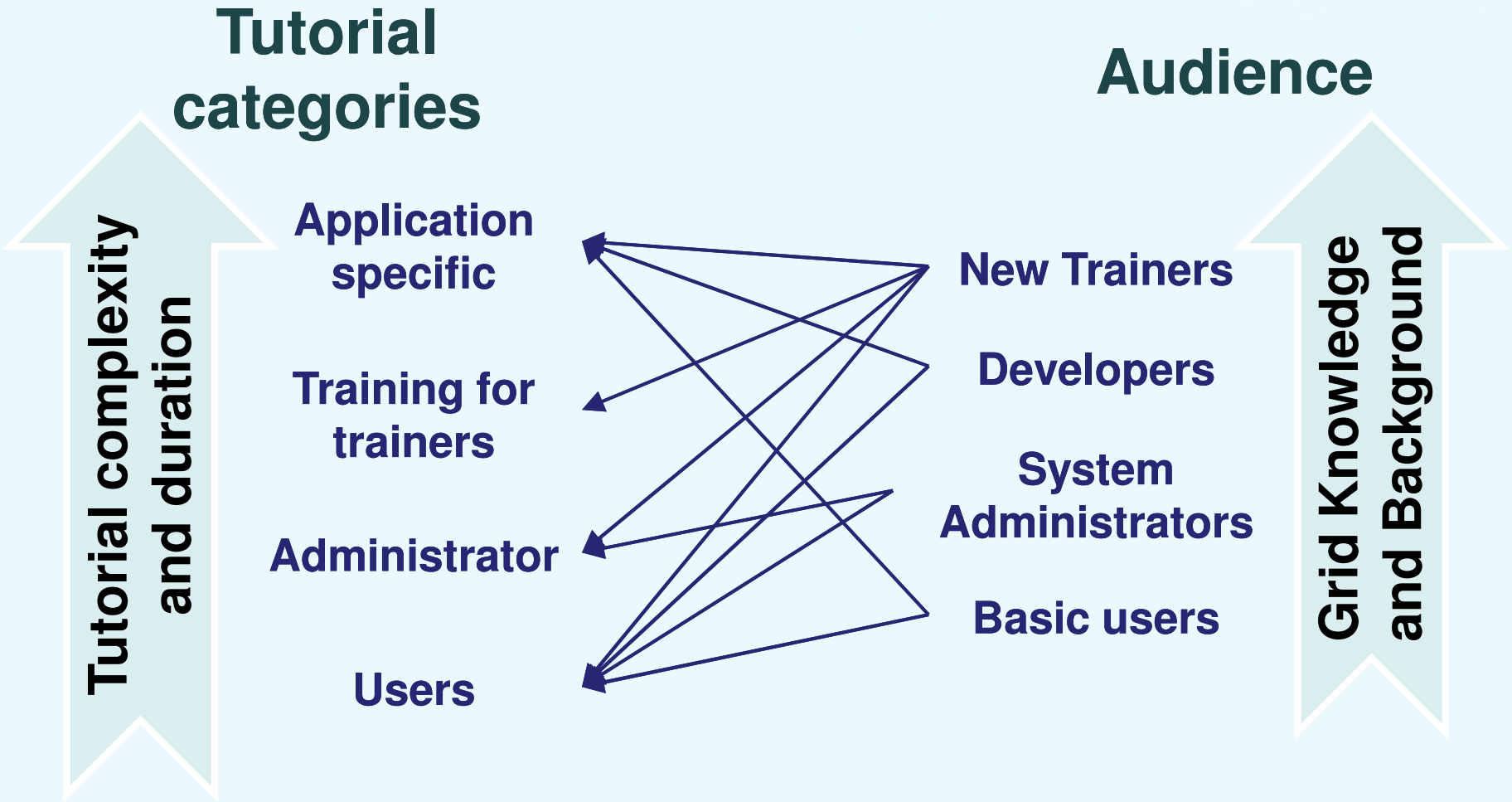
Events organization and participation

- ✓ Expand promotion to scientific events attended by new potential users of the grid
- ✓ Continue to collaborate with other WPs and associated EU projects to increase multi-channel outreach

Publication of results

- ✓ iSGTW as main EUAsiaGrid media partner (F. Grey)
- ✓ Promote the publication of project's achievements on peer-reviewed journals and proceedings of non-ICT conferences (2 papers already !)

Provided Training and Audience



Provided training events

- **Four training events organised by the project:**
  - EUAsiaGrid Training Event at MIMOS
    - Kuala Lumpur, Malaysia, **28-31 July 2008**
  - Grid Camp 2008 at ASGC
    - Taiwan, Taipei, **18-24 October 2008**
  - EUAsiaGrid / EGEE tutorial at ASGC
    - Taiwan, Taipei, **18 April 2009**
  - EAGSS'09 tutorial at UPM
    - Kuala Lumpur, Malaysia, **27 July - 7 August 2009**
- **Trainers from INFN and ASGC**
- **GILDA adopted as training infrastructure**

EUAsiaGrid training event

- **Two days tutorial for System Administrators and two days tutorial for Basic Users**
  - Participants coming only from EUAsiaGrid partners

http://indico.twgrid.org/internalPage.py?pageId=0&confId=476

Grid Camp 2008

- **A two days tutorial aimed at Application developers**
  - co-located with Grid Camp and HEPiX Fall 2008
- **Several modules were dedicated to specific frameworks**
  - e.g. DIANE, GANGA and GAP

http://indico.twgrid.org/internalPage.py?pageId=1&confId=471

EUAsiaGrid / EGEE Tutorial

- **A two days tutorial for Basic Users**
  - co-located with ISGC 2009
- **Include additional modules**
  - Some modules on security and other issues provided by trainers not involved in the project

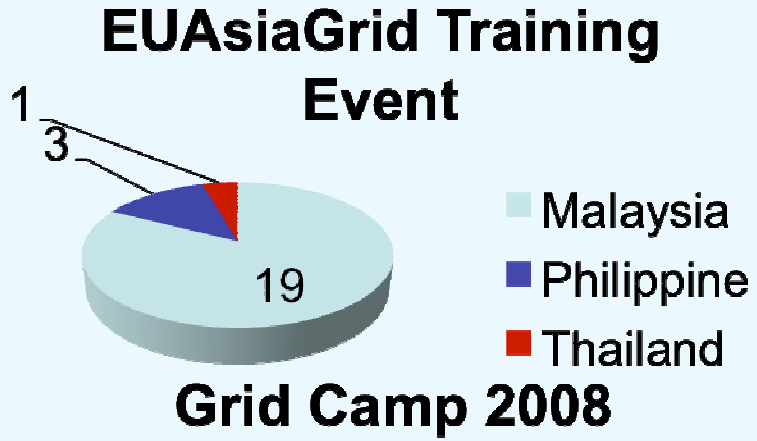
http://event.twgrid.org/isgc2009/home.htm

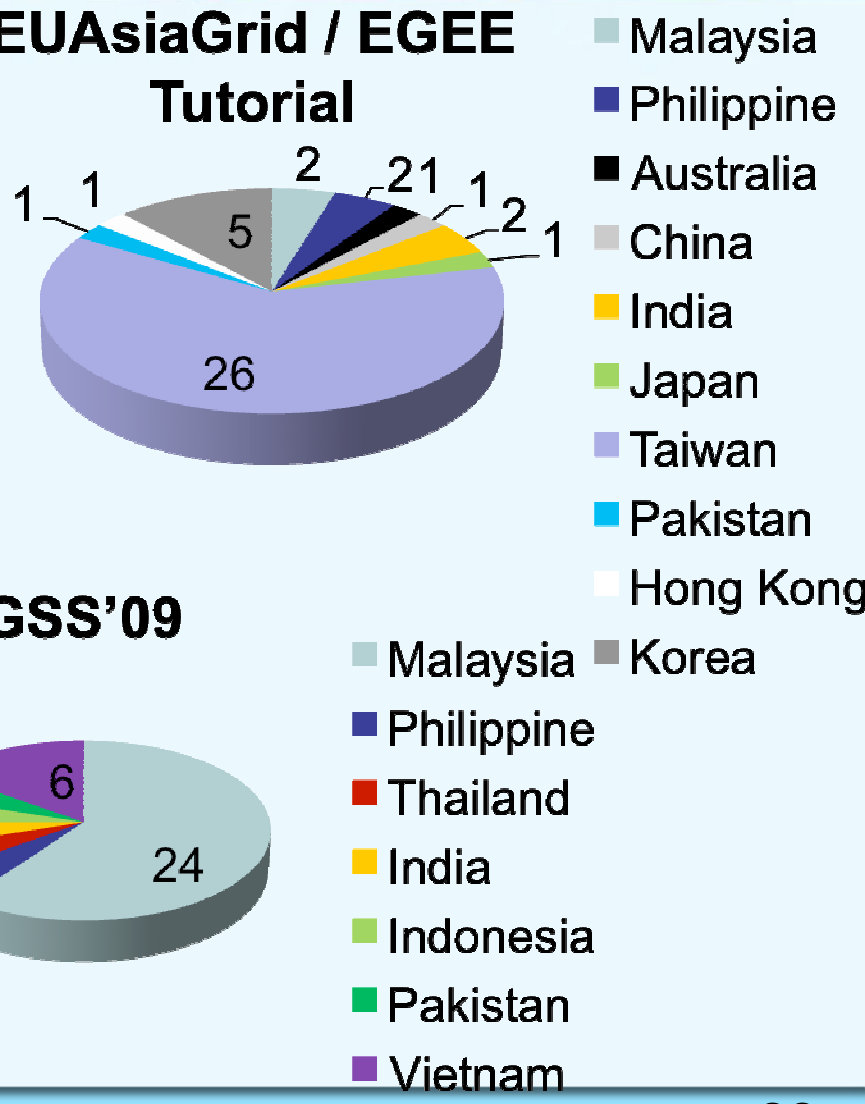
EAGSS'09

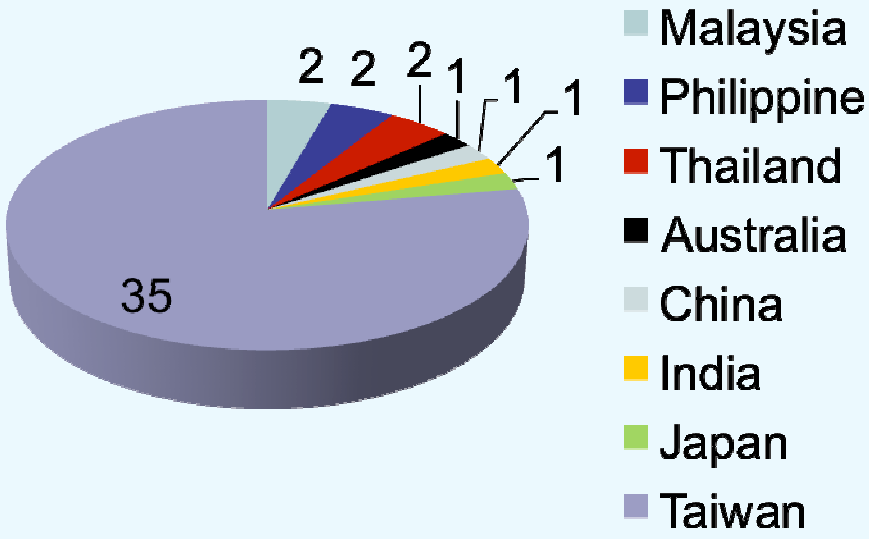
- **A two weeks tutorial**
  - One week for System Administrators and one week for Basic Users
  - co-located with APAN
- **The first event using the official EUAsiaGrid training resources**

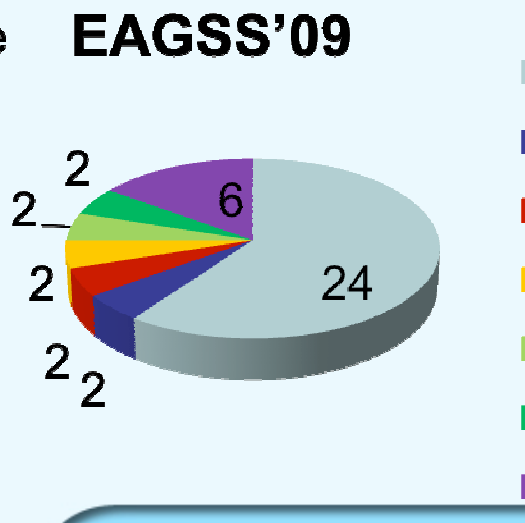
http://indico.twgrid.org/conferenceDisplay.py?confId=703

Provenance by country

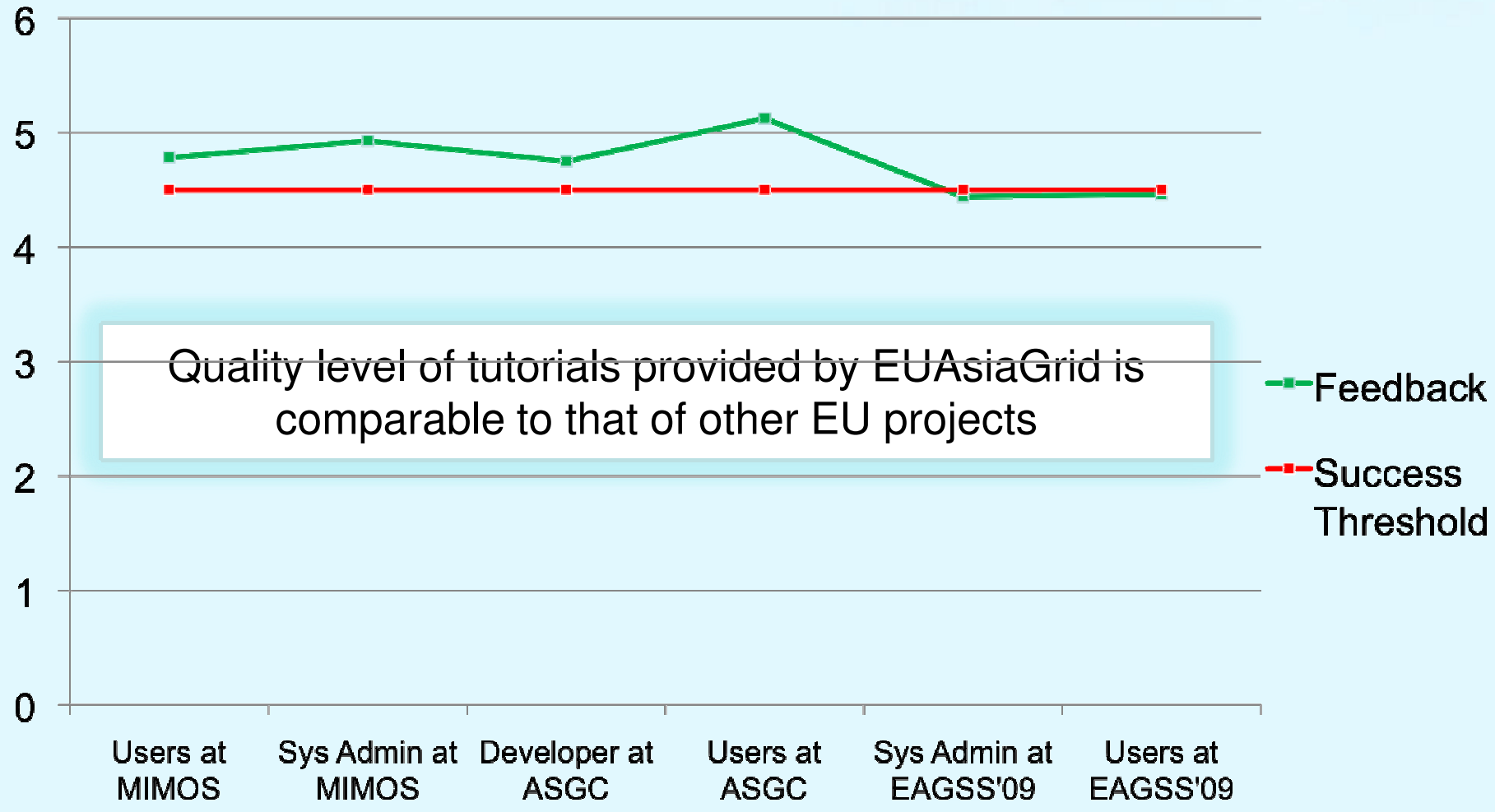




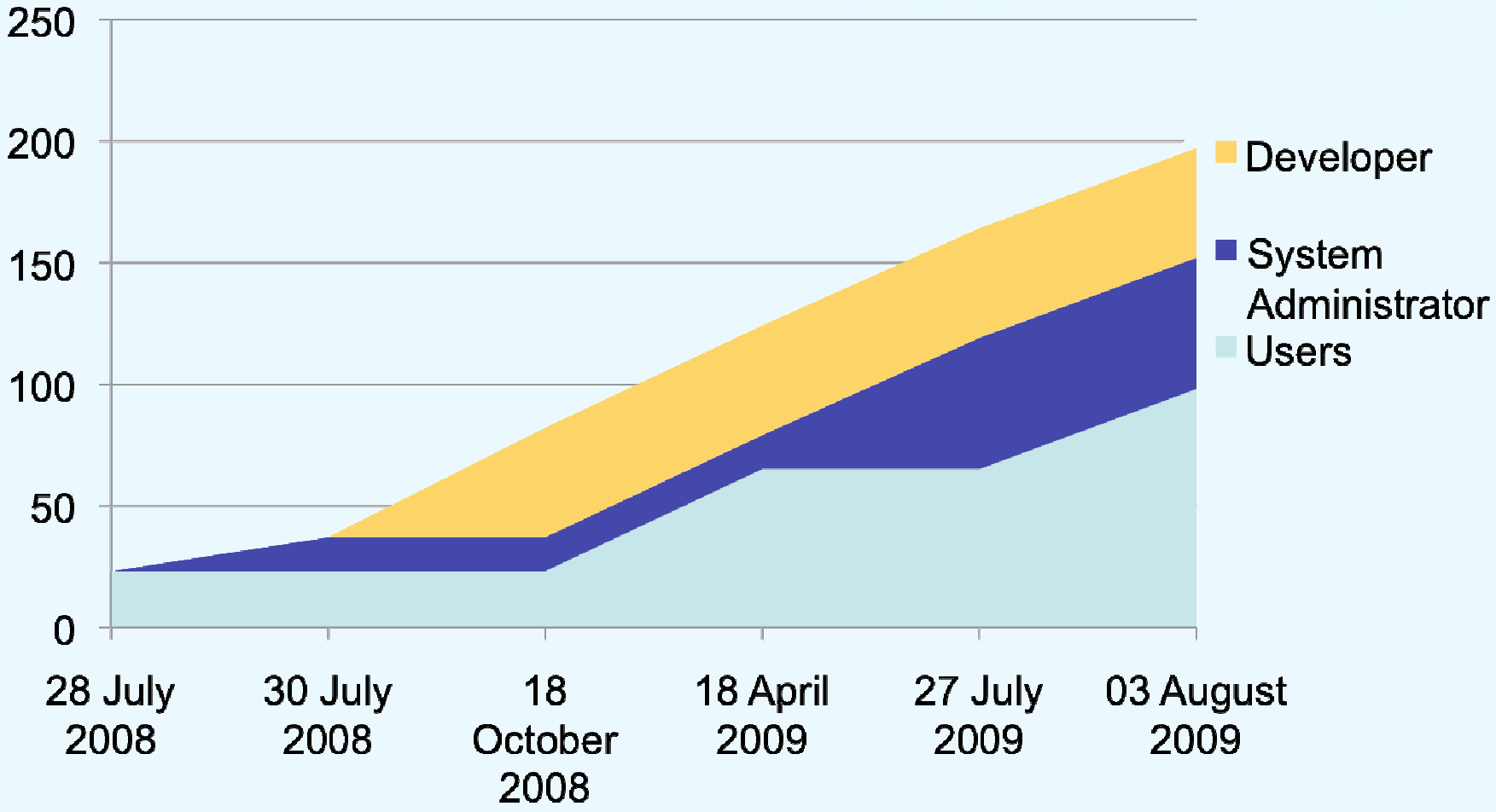




Tutorials Feedback



Cumulative Tutorial Attendance



Training material collection

- **Training material of GILDA accessible from EUAsiaGrid web site**
  - **http://www.euasiagrid.eu**
- **Tutorials presentations available on the agenda web pages**
  - **The list of training events can be found at:** http://www.euasiagrid.org/index.php?option=com_events&task=view_cat&offset=1&catid=40&Itemid=44
- **EUAsiaGrid wiki with additional information**
  - **Wiki page at: http://www.euasiagrid.org/wiki/**

Preparing for online training

- **Following reviewers recommendation for "online training" and "follow up of training":**
  - ***New section for training documents available soon in the EUAsiaGrid web, with material selected and classified to simplify access***
  - *A dedicated wiki for training is available with exercises and step-by-step tutorial both for Users and System Administrators*
- **Survey being addressed to people already trained to follow up their Grid activities**

Next EUAsiaGrid trainers event

- ***Training for Trainers***
  - Period: Oct, 23-27
  - Location: **Vietnam**
- **Participants will get a deeper Grid knowledge and information on how to run a tutorial**
- **The aim is to create a group of persons able to spread the Grid knowledge after the end of the project**
  - Increasing the long term sustainability of the Asian Grid infrastructures

Collaboration in other events

- **Collaborations with other EU institutions and projects are established to improve the training**
- **The project is taking part with CNRS to the organisation of ACGRID school series for 2009**
  - First event
    - Period: 21 September – 2 October
    - Location: **Vietnam**
  - Second event
    - Period: 2 – 14 November
    - Location: **Malaysia**
- **These schools are mainly oriented to local communities but open also to external participants**

Proposed amendments to the DoW

Change of partners in Australia, Singapore and Vietnam

Extension of the project (Apr 2010 – Jun 2010)

- ✓ To fully profit of the discovered potential
- ✓ To reach a solid self-sustainability of Grid infrastructure in AP region

New deliverables added

- ✓ Report on Data Challenges (WP3)
- ✓ Report on sustainability of training provision in the AP region (WP5)

White Paper to provide stakeholders with sustainable model and cost estimate for a stable Grid infrastructure in the Asia Pacific region

- ✓ Build on the EELA-2 experience (MoU being signed)

Conclusions

- ✓ The project, despite its complexity and a slower start has catched up well, by managing risks and overcoming obstacles through clear identification of priorities and leadership in key areas
- ✓ Now we have to focus on application and sustainability, up to the end of the project (June 30th, 2009)
- ✓ Will gather key partners in a JRU to partecipate into CHAIN