NA4: Application Identification and Support

C. Loomis (CNRS), V. Floros (GRNET)

EGEE-II 2nd EU Review (CERN)
8-9 July 2008

www.eu-egee.org

Information Society and Media

Contents

- NA4: Application Identification and Support
- Adoption of Grid Technology
- EGEE User Community
- Common APIs and Tools
- Exploitation Plans
- Summary

- Presentation highlights important pointed raised in DNA4.2.2 and DNA1.2.2.
  - *Statistics, where possible, cover EGEE-II project.*
  - *Status statistics are for June 2008.*

Goals & Organization

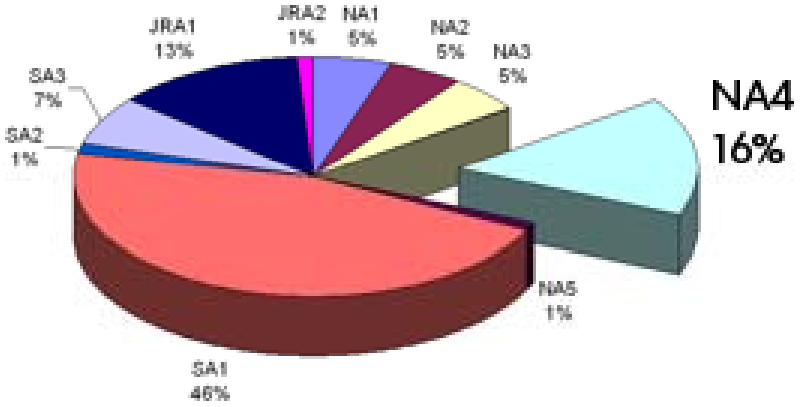
- **Expand use of EGEE infrastructure:**
  - **User**: Person exploiting EGEE services.
  - **Virtual Organization**: Groups of users federating resources.
  - **Applications**: User codes, programs, and algorithms.
- **Ensure current users are satisfied.**

Steering Committee	
Coordinator	C. Loomis
Deputy Coordinator	V. Floros
VO Mgr. Group	F. Schaer
NA4/NA1 Liaison	F. Harris
Astron. & Astrophysics	C. Vuerli
Comp. Chemistry.	M. Sterzel
Earth Science	M. Petitdidier
Fusion	F. Castejon
High-Energy Physics	M. Lamanna
Life Sciences	C. Blanchet V. Breton J. Montagnat
GILDA	R. Barbera
GASuC	G. Sipos

NA4 in Numbers

40 (42) Partners, 25 (27) Countries

EGEE-II Budget



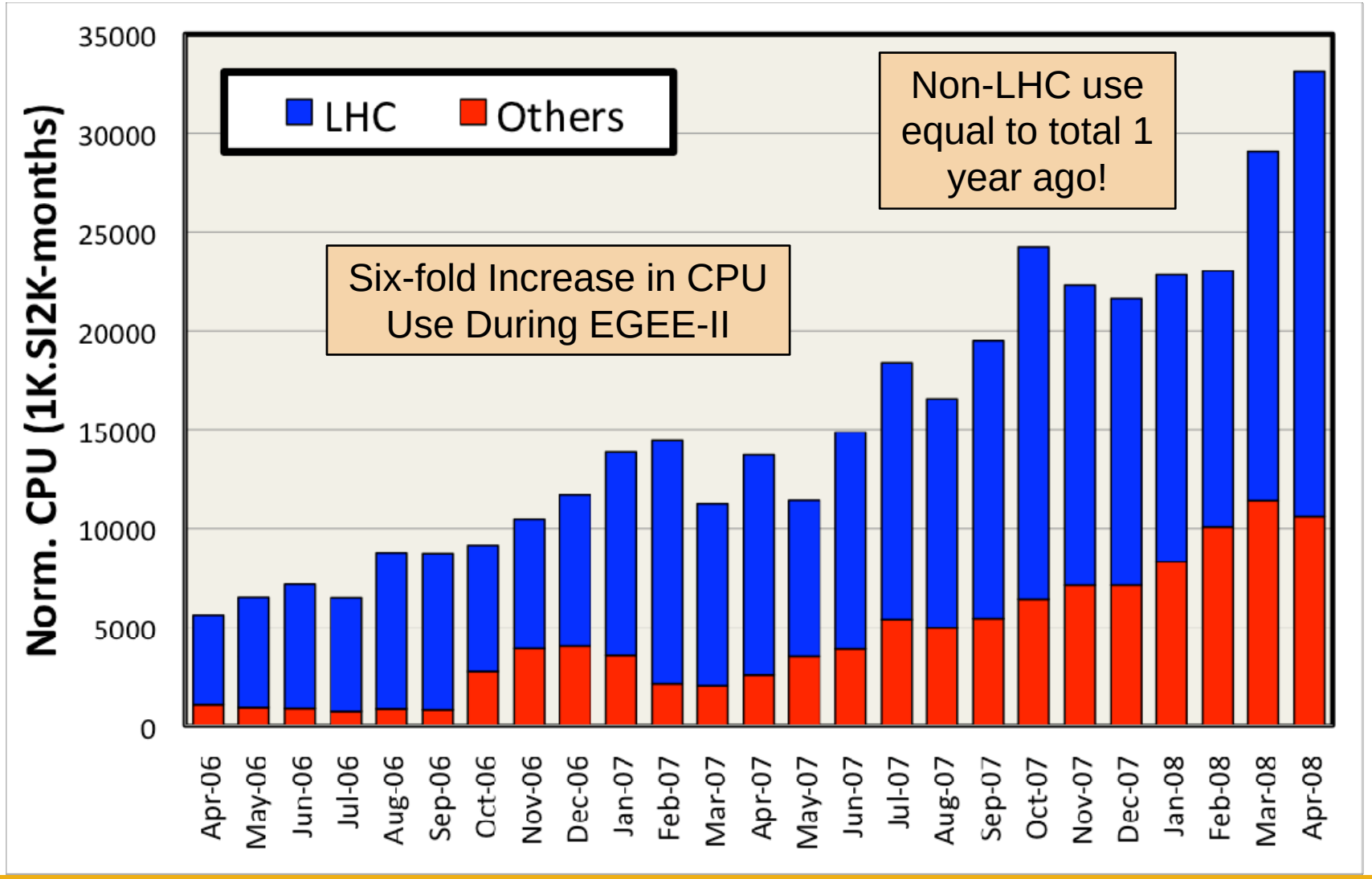
Federation	FTE	People
Cent. Europe	6	56
CERN	12	17
FR	16	72
DE/CH	3	12
IT	18	44
N Europe	3	26
Russia	2	6
SE Europe	5	51
SW Europe	11	25
UK/IRE	1	4
Asia	0	3
US	0	0
Total	77	316

25%

Adoption of Grid Technology

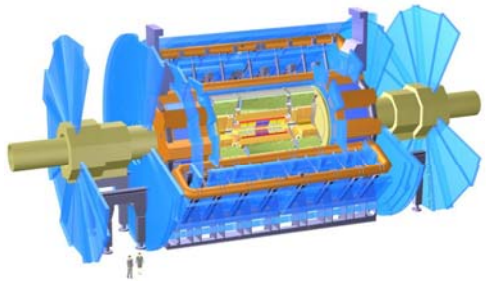
CPU Utilization

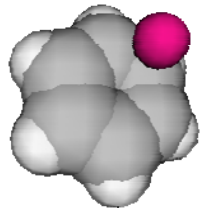
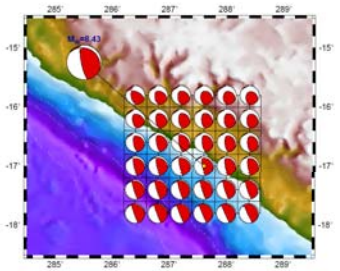
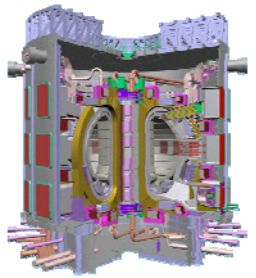
- **Recent level equal to ~32000 CPUs in continuous use.**



Reported Applications

- **1st year**
  - Growth in reported apps.
- **2nd year**
  - Transition: prototype to production

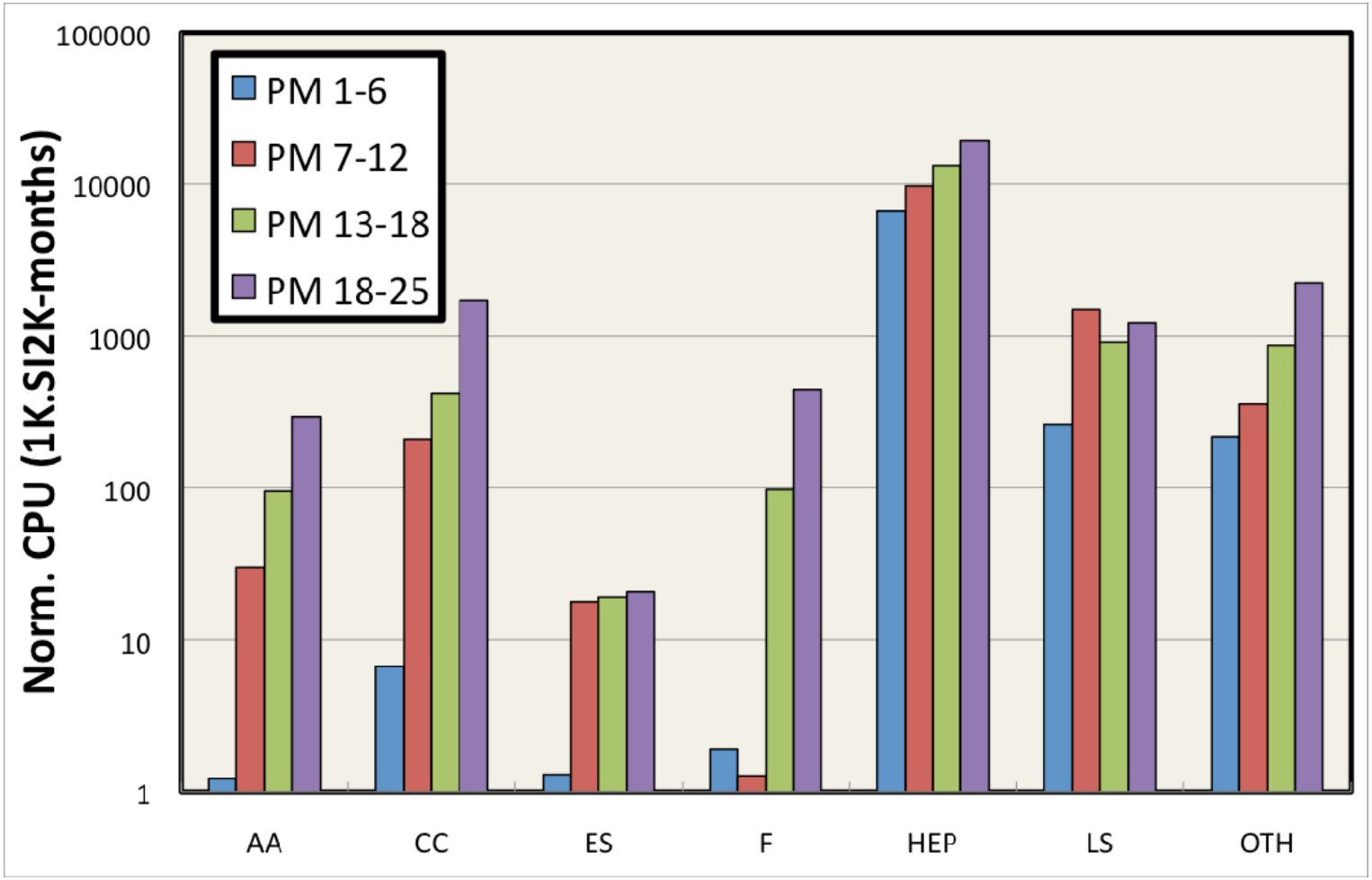




	6/2006	2/2007	1/2008
Astron. & Astrophysics	2	8	9
Comp. Chemistry	6	27	21
Earth Science	16	16	18
Fusion	2	3	4
High-Energy Physics	9	11	7
Life Sciences	23	39	37
Others	4	14	21
Total	**62**	**118**	**117**

Condensed Matter Physics
Comp. Fluid Dynamics
Computer Science/Tools
Civil Protection

Maturing User Communities

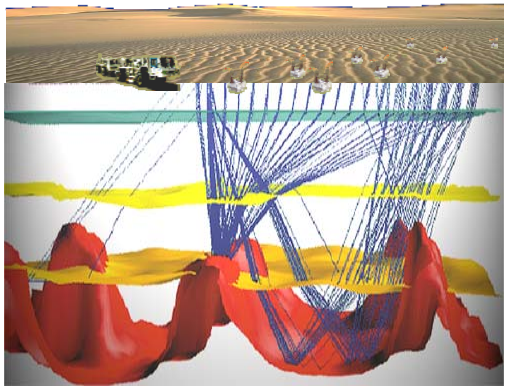
- **Continued strong use in developed disciplines.**
- **Prototyping to production for younger disciplines.**

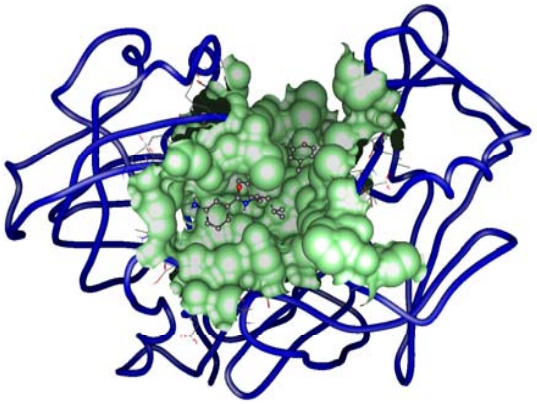


Industrial Participation

- **Collaboration with industry in NA4 limited:**
  - Mostly academics involved with limited industrial contacts.
  - Commercial network restrictions discourage direct collaboration.

- **Geocluster (CGGVeritas)**
  - Geoscience software package, used for example in petroleum search.
  - Made available to researchers on EGEE grid infrastructure.
- **WISDOM**
  - Collaboration with BioSolveIT.
  - Provided free licenses for docking calculation on EGEE.





License Models

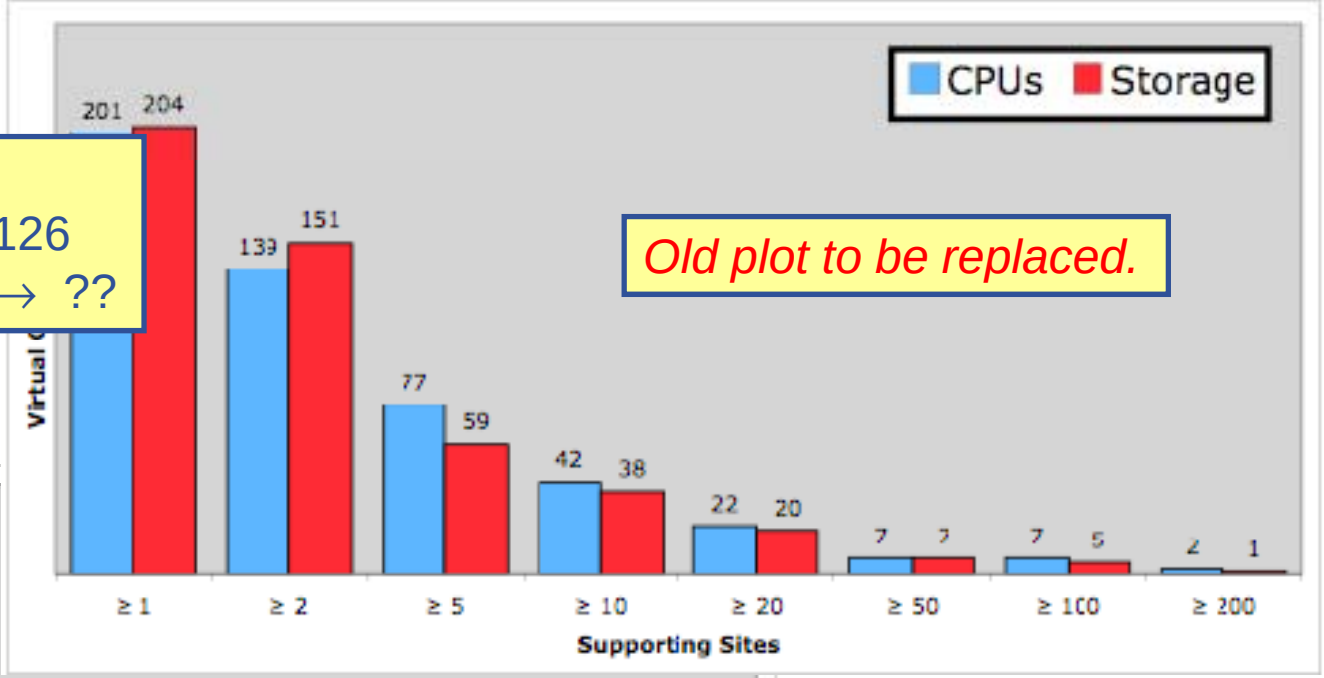
- **Expanding community needs commercial software on the grid. Working to find good license models:**
  - VO License Model
    - Used for comp. chemistry packages Gaussian and Turbomole.
    - Puts burden of enforcement on VO manager.
    - Inflexible and poorly adapted to workflow.
  - Client/Server License Model
    - Used by MATLAB Parallel Computing Toolkit.
    - Allows separate licenses for client and server.
    - More flexible and allows sites to provide a “resource”.
    - Currently in process of running a trial with EGEE users.

- **Collaboration with MathWorks shows that EGEE starting to become a target platform for software vendors.**

EGEE User Community

Virtual Organizations

Total VOs: 204 → 233
Registered VOs: 116 → 126
Median sites per VO: 3 → ??



Old plot to be replaced.

Old plot to be replaced.

Total Users: 5034 → ??
Affected People: 10200 → ??
Median members per VO: 18 → ??

EGEE is an open infrastructure and not all VOs register with the project.

Support Services

- **EGEE have comprehensive and efficient support system to ensure that users are satisfied.**

- **Support from other activities:**
  - GGUS (SA1)
  - Training (NA3)
  - Middleware (JRA1)
- **Support activities within NA4:**
  - Administrative support: OAG, VO Mgrs. Group
  - User support: UIG, NA4 Portal
    - § http://egee-uig.web.cern.ch/egee-uig/production_pages/UIGindex.html
  - Application porting support: GILDA, GASuC

Previous Support Issues

- **All of these issues have been resolved in EGEE-III via structural changes to the NA4 activity and tasks.**

- **Resource Allocation:**
  - *Issue: No EGEE computing and storage resources to allocate to new virtual organizations as bridge to production use.*
  - Solution: Create seed resources to for new communities.
- **Application Porting Support:**
  - *Issue: Porting support is most efficient "in person". How to finance travel for unfunded people?*
  - Solution: Fund to partially finance travel to GASuC.
- **Direct User Support:**
  - *Issue: Providing user support to "outside" users.*
  - Solution: Team within NA4 to provide this type of support.

Application Porting Support

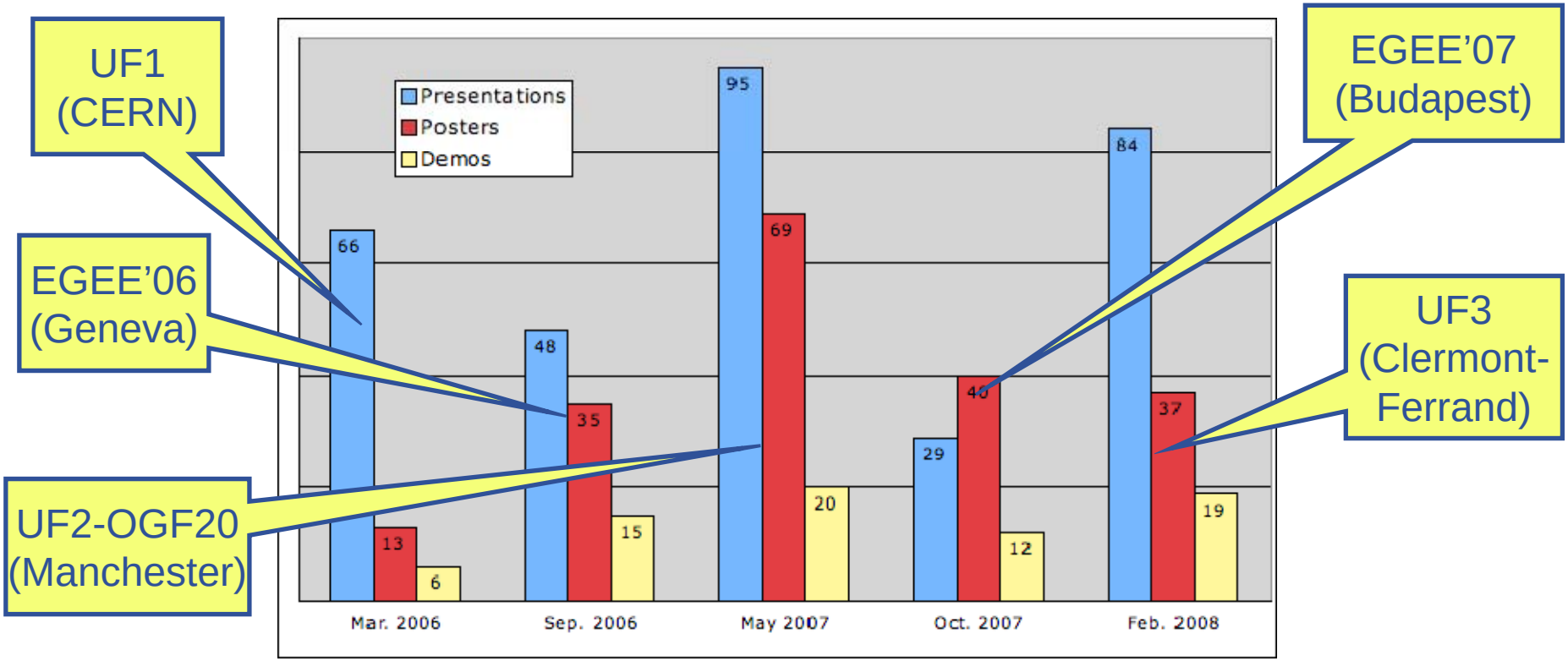
- **Porting an application to the grid usually requires expertise that new virtual organizations do not have.**

- **Training infrastructure (with NA3):**
  - https://gilda.ct.infn.it/
  - GILDA team advises new users on EGEE grid technology.
  - t-infrastructure provides resources for testing new applications.
- **Porting to production service:**
  - http://www.lpds.sztaki.hu/gasuc/
  - Some prefer porting directly to production service.
  - GASuC (SZTAKI) now offers hands-on consulting to do this.

- **Direct support from NA4 partners:**
  - Motivated to port “local” applications.

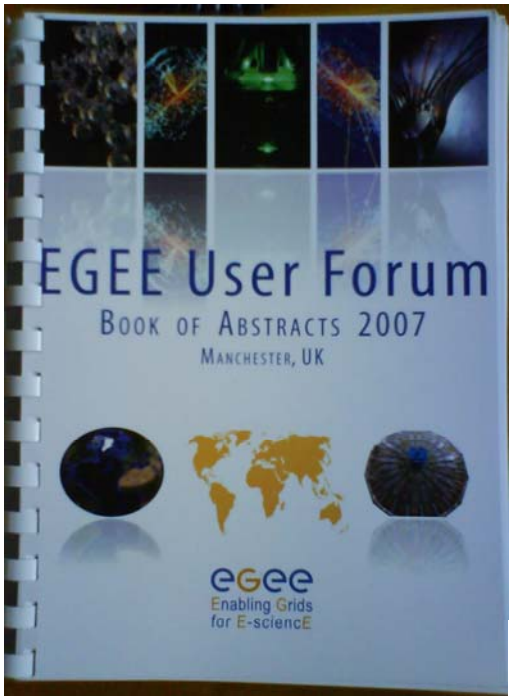
Community Building

- **Meetings for specific scientific disciplines.**
- **Strong participation in technical working groups.**
- **User Forums & EGEE Conferences.**

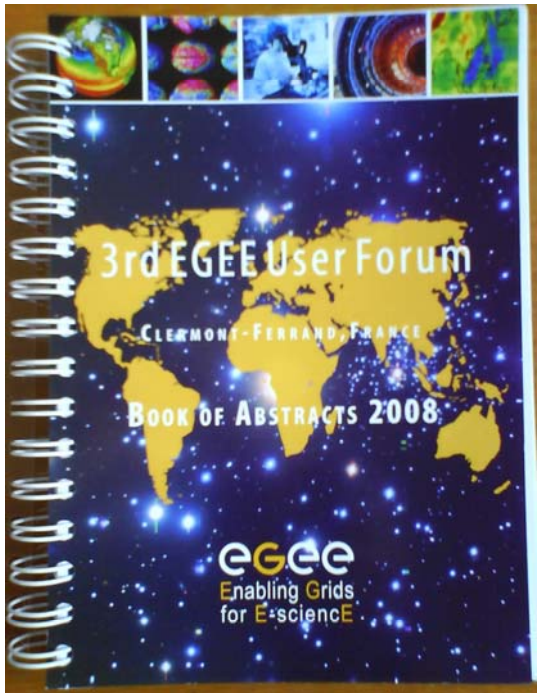


User Forum Results

- **Rich scientific program allowing users to share their grid expertise and demonstrate benefits of grid technology for science.**



http://indico.cern.ch/conferenceDisplay.py?confId=7247



http://indico.cern.ch/conferenceDisplay.py?confId=22351

Common APIs and Tools

Middleware

- **Middleware critical for success of NA4:**
  - gLite provides important core services.
  - Application-level code and services supplements those services.

- **NA4 contributions:**
  - Improvements to gLite and gLite deployment.
  - Development of high-level services.
  - Identification of external services and packages.

Middleware

- **Extensive testing of services**
  - HEP and life science communities leaders in this area
  - Recent work with gLite WMS indicative of positive results.
  - Advanced testing of prototypes, like Hydra for data encryption.

- **Collaborate through targeted working groups:**
  - MPI: improve parallel job support on grid
  - SDJ: reduce scheduling latencies for quasi-interactive apps.
  - MDM: mgt. of medical data on the grid
  - Priority: provide mechanisms to define VO-level job priorities
  - Portal: define best practices for grid portals
  - VO Config.: improve sharing via simpler VO configuration
  - DB Access: improve database access from grid
  - *Issue: Ensure recommendations are acted upon.*

Development

- Direct development has usually resulted in generic service used by several scientific communities.
- AMGA: Metadata catalog.
- Ganga: Job submission framework.
- DIANE: Master/slave task manager.
- Dashboard: VO and user-level monitoring.
- MOTEUR: Workflow engine.
- MDM: Medical Data Management

RESPECT

- **Rec. External Software Pkgs. for the EGEE Community**
  - Identify useful, 3rd-party software that works with gLite.
  - Make people aware of that software to avoid duplicated efforts.
  - http://egeena4.lal.in2p3.fr/index.php?option=com_content&task=view&id=71&Itemid=63

- **Current packages:**
  - GridWay: Grid metascheduler.
  - Ganga: Job submission framework.
  - DIANE: Master/slave task manager.
  - i2glogin: Interactive login to grid nodes.
  - GReIC: Database access and management.

- **Discussing with int.eu.grid to add more of their products to the RESPECT program.**

Exploitation Plans

- **Detailed exploitation plans for each sector given in final periodic report. Generally, plans are to expand number, size, and complexity of ported applications.**

- **EGEE-III**
  - Provides grid infrastructure for next two years.
  - All current disciplines will continue into EGEE-III.
  - Add Grid Observatory activity.

- **Long-term exploitation depends on having a stable, production platform available.**

Summary

- **Adoption of grid tech. and growth of user community:**
  - 6x increase in CPU utilization over life of project
  - Use by diverse set of VOs and scientific disciplines.
- **Work on common APIs and tools:**
  - Improvement of gLite itself through testing and enhancements.
  - Direct development of tools.
  - Identification of third-party tools via RESPECT.
- **Scientific disciplines continue with EGEE-III:**
  - Expand number, size, complexity of ported applications.
  - Structural and task changes should address previous issues.
  - Challenge: Effectively support large and growing community.