



Enabling Grids for E-scienceE

Gridové počítanie a projekt EGEE

Miroslav Dobrucký
Ústav informatiky
Slovenská akadémia vied
Bratislava

Použitý materiál z iných prezentácií konzorcia EGEE.
Copyright (c) Members of the EGEE-II Collaboration 2006. See www.eu-egee.org for details.

www.eu-egee.org



- **Čo je to gridové počítanie?**
- **Charakteristiky gridu**
- **Aplikácie**
- **Gridové projekty a EGEE projekt**

Vedci a výskumníci vykonávajú svoje aktivity v kooperácii v rámci celého sveta, komunikujú medzi sebou a zdieľajú spolu údaje.

Grid: zosieťované centrá na spracovanie údajov a "middleware" ako "lepidlo" na spojenie zdrojov.



Virtuálne organizácie

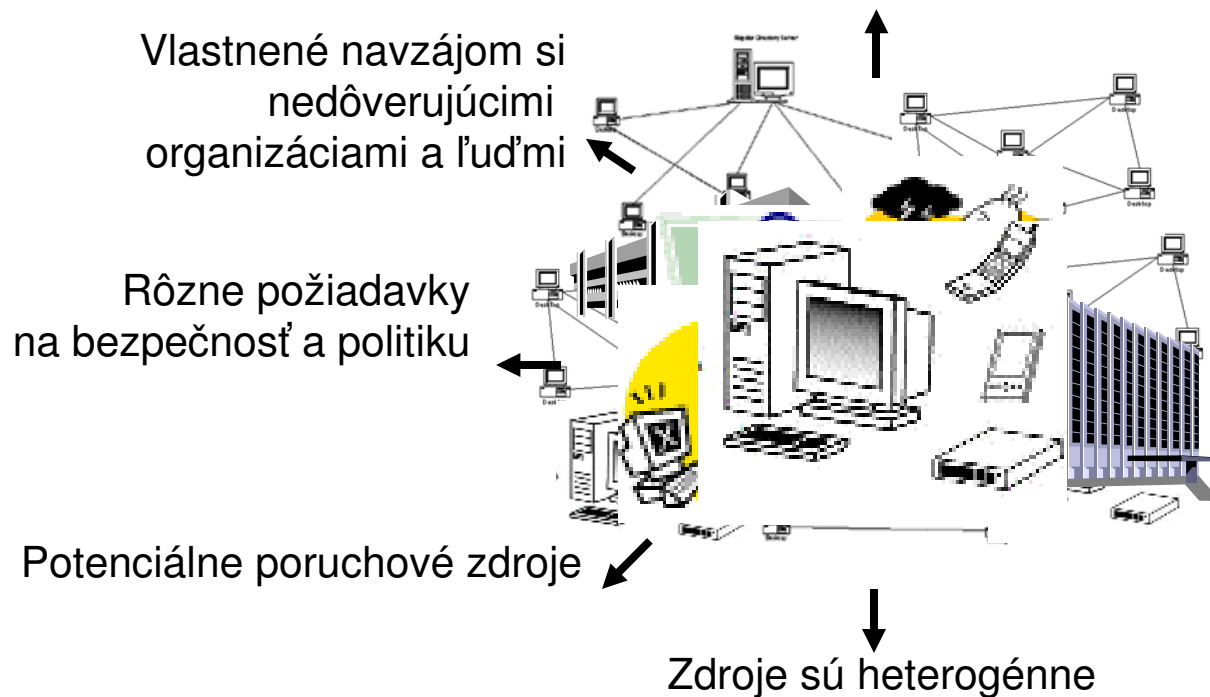
Vedecké nástroje a experimenty poskytujú obrovské množstvo údajov.

- **Súčasné distribuované aplikácie:**
 - Smerujú k *špecializovaným* systémom
 - Sú zamerané na *jeden* účel alebo skupinu používateľov

- **Gridy idú ďalej a berú do úvahy:**
 - Rôzne druhy zdrojov
 - Nie je vždy rovnaký hardvér, údaje a aplikácie
 - Rôzne druhy *interakcií*
 - Skupiny používateľov alebo aplikácie chcú interagovať s gridom rôznymi spôsobmi
 - *Dynamická* povaha
 - Zdroje a používatelia sú pridávaní/odoberaní často

- Čo je to gridové počítanie?
- **Charakteristiky gridu**
- **Aplikácie**
- **Gridové projekty a EGEE projekt**

Veľké množstvo zdrojov



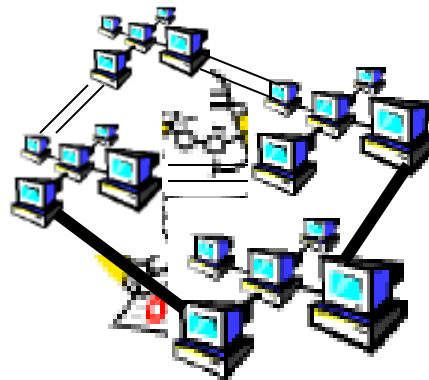
potrebujeme štandardy

Veľké množstvo zdrojov

Vlastnené navzájom si
nedôverujúcimi
organizáciami a ľuďmi

Rôzne požiadavky
na bezpečnosť a politiku

Potenciálne poruchové zdroje



Zdroje sú heterogénne

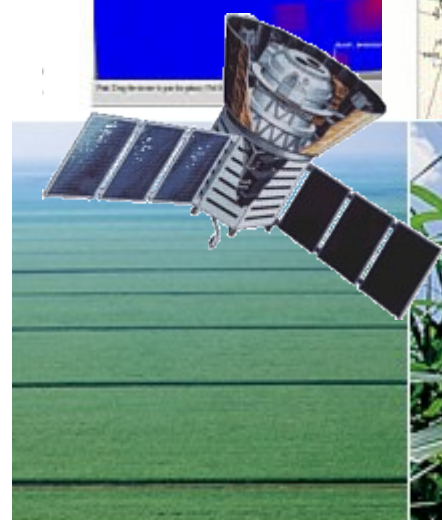
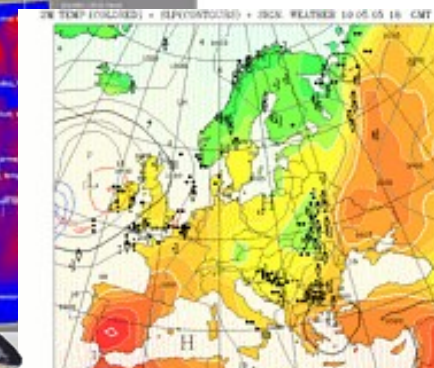
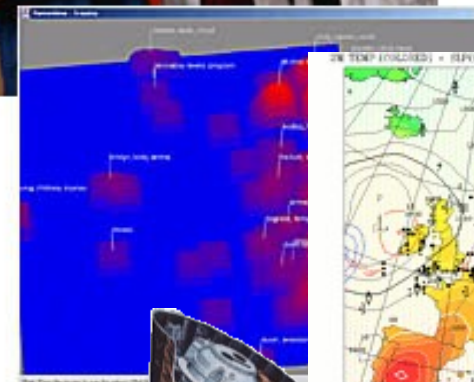
Prepojené heterogénnymi,
viac-úrovňovými sieťami

Rôzne politiky
manažmentu zdrojov

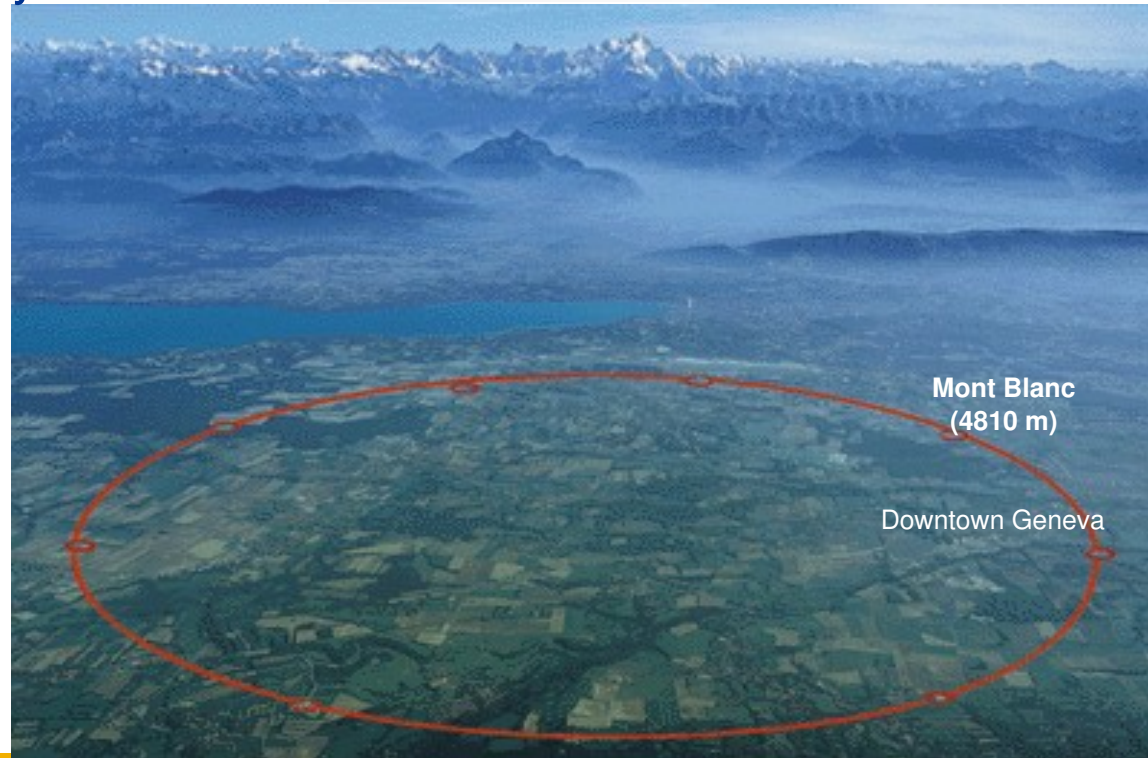
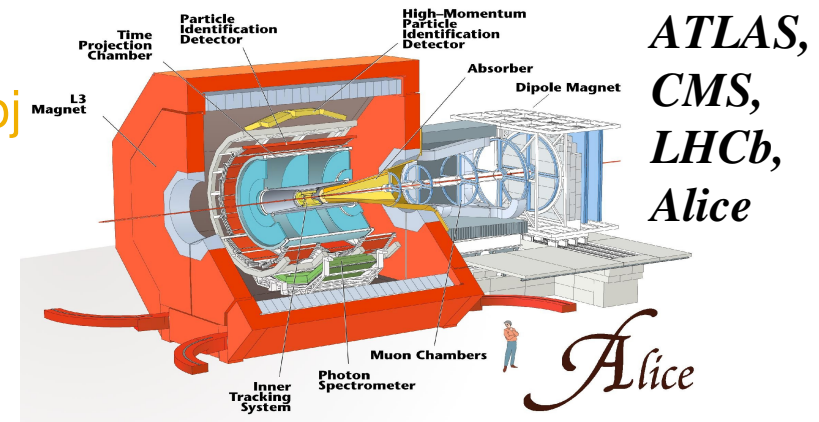
Geograficky
oddelené

- Čo je to gridové počítanie?
- Charakteristiky gridu
- **Aplikácie**
- **Gridové projekty a EGEE projekt**

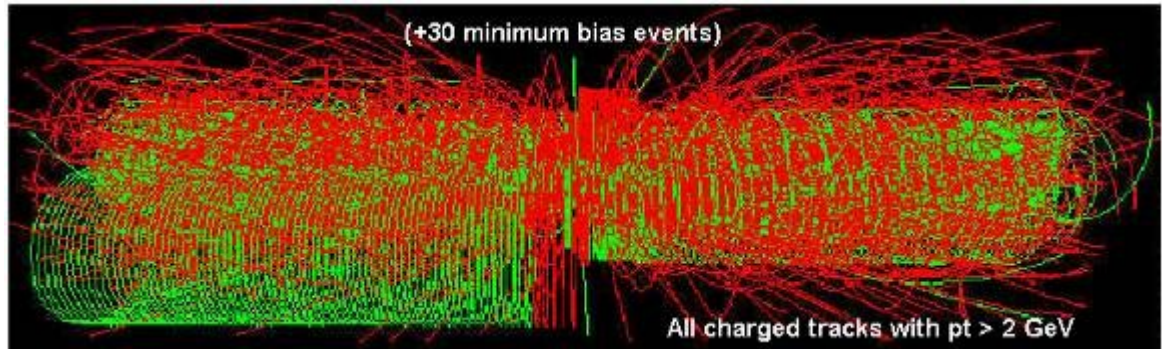
- **Medické/zdravotnícke** *(snímkovanie, diagnostika a liečenie)*
- **Bioinformatické** *(skúmanie ľudského genómu aby sme porozumeli genetickým chorobám)*
- **Nanotechnológie** *(návrh nových materiálov na molekulárnej úrovni)*
- **Strojárstvo** *(optimalizácia návrhu, simulácie, analýza porúch a vzdialená manipulácia a riadenie)*
- **Prírodné zdroje a životné prostredie** *(predpoveď počasia, pozorovanie zeme, modelovanie a predikcia zložitých systémov)*



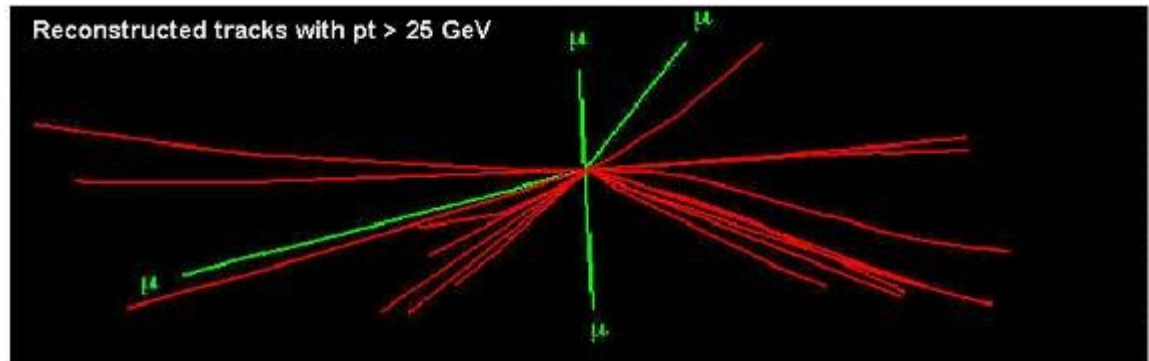
- **The Large Hadron Collider (LHC)**
 - doteraz postavený **najvýkonnejší nástroj** na vyšetovanie elementárnych častíc
- **Údajová výzva:**
 - **10 Petabajtov/rok** údajov !!!
 - **20 miliónov CD-čiek** každým rokom!
- Simulácia, rekonštrukcia, analýza:
 - Údaje z LHC vyžadujú výpočtový výkon porovnateľný so **~100,000** najrýchlejšími **súčasnými PC** procesormi!



Štartujúc z tejto udalosti



Hľadáme túto “signatúru”



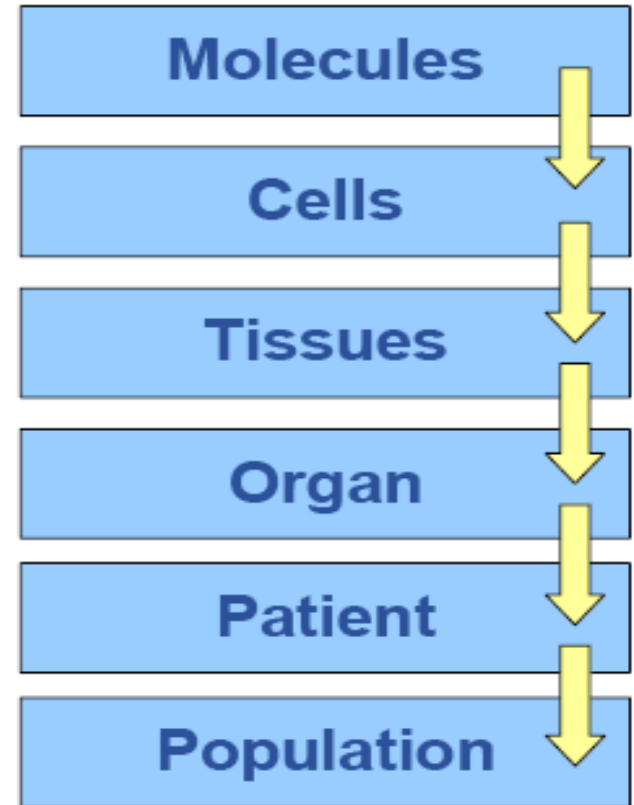
→ **Selektivita: 1 in 10^{13}**

(Asi ako hľadanie ihly v 20-tich miliónoch kôp sena)

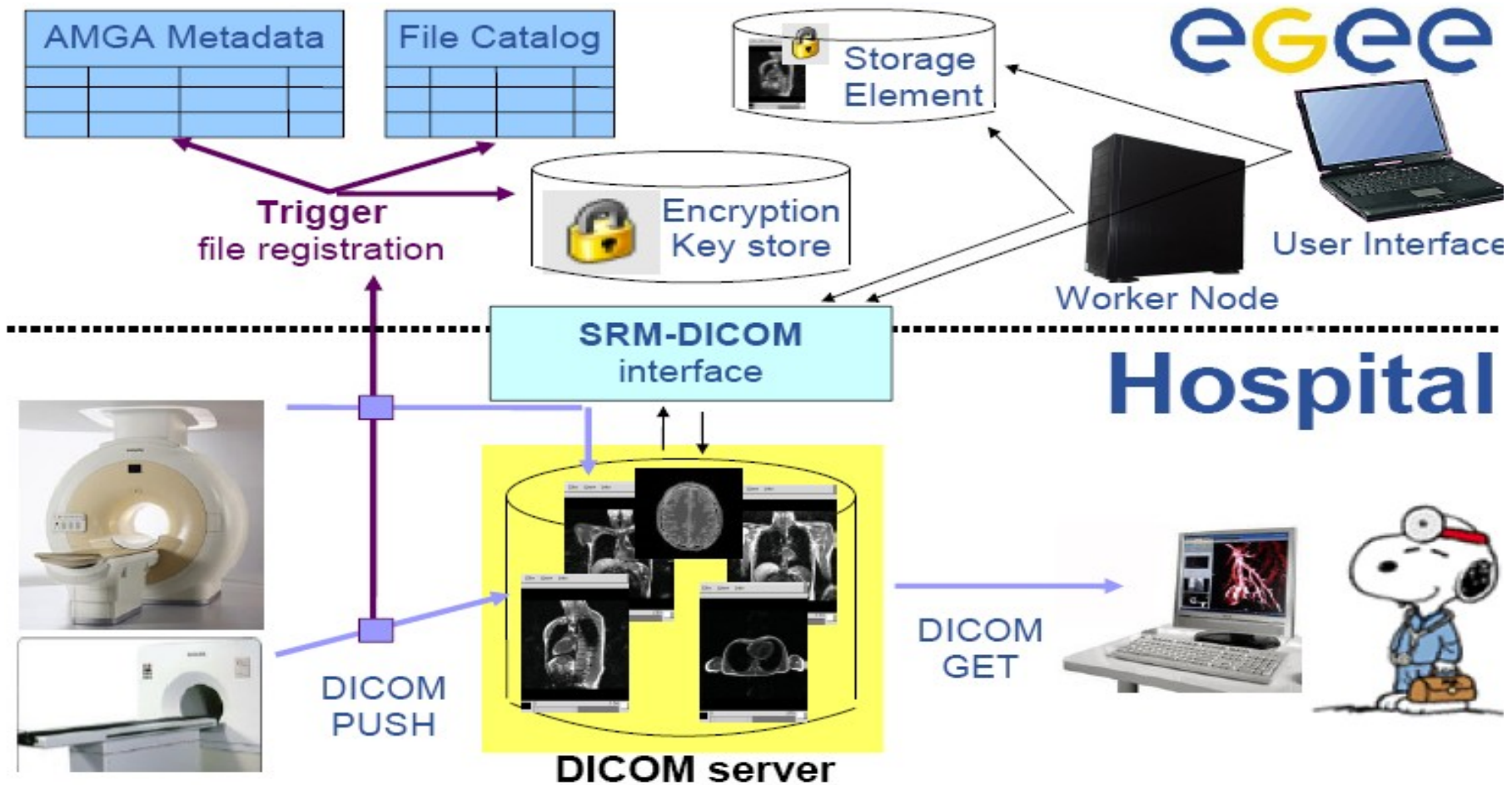
- **Bioinformatics**
 - Genomics
 - Proteomics
 - Phylogeny...

- **Medical imaging**
 - Medical imaging
 - Computer Aided Diagnosis
 - Therapy planning
 - Simulation...

- **Life sciences**
 - Drug discovery
 - Epidemiology
 - ...

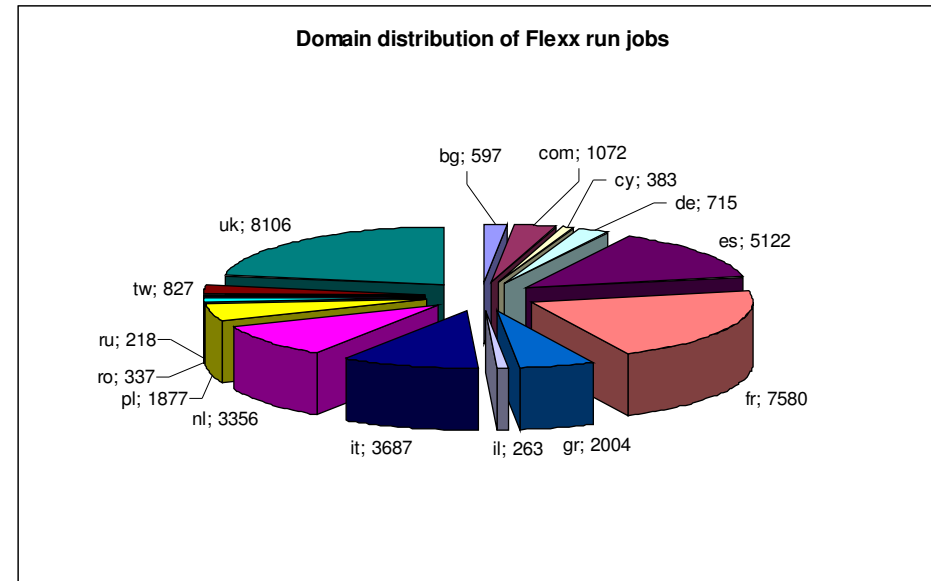


Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin



Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin

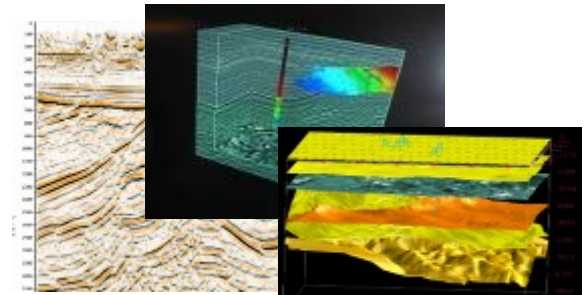
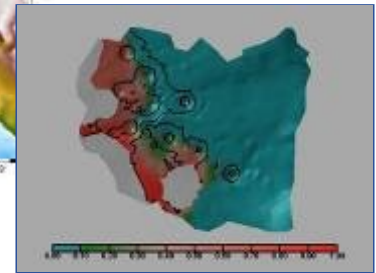
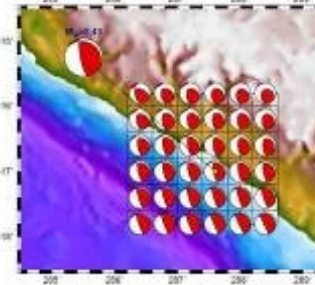
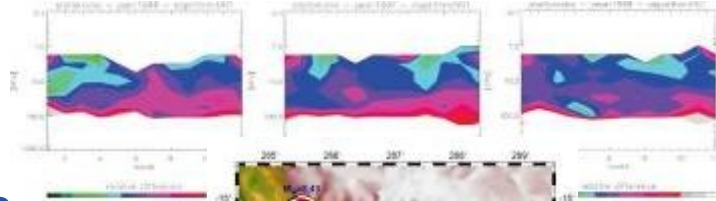
- Significant biological parameters
 - two different molecular docking applications (Autodock and FlexX)
 - about one million virtual ligands selected
 - target proteins from the parasite responsible for malaria
- Significant numbers
 - Total of about 46 million ligands docked in 6 weeks
 - 1TB of data produced
 - Up to 1000 computers in 15 countries used simultaneously for a total of about 80 CPU years
- Significant results
 - Best hits to be re-ranked using Molecular Dynamics



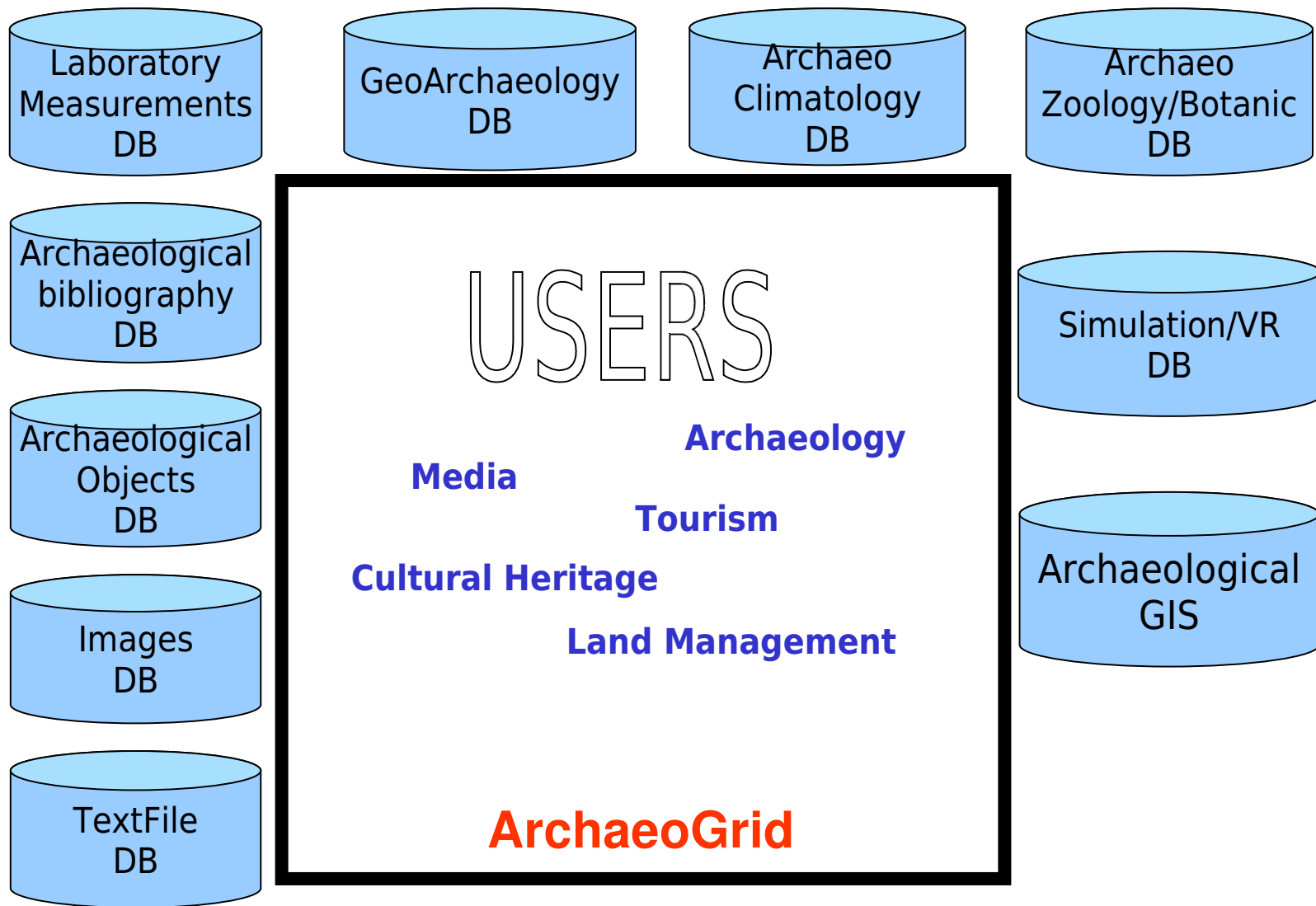
New data challenge in the fall of 2006
 New malaria targets
 Focus on other neglected diseases
 Enlarged collaboration
 (possibly including related projects)

Roberto Barbera, 1st EGEE User Forum, CERN, 1st March 2006

- **Pozorovanie Zeme zo satelitov**
 - Ozónové profily
- **Fyzika pevného zemského jadra**
 - Rýchle určenie mechanizmu dôležitých zemetrasení
- **Hydrológia**
 - Manažment vodných zdrojov v oblasti okolo stredozemného mora (SWIMED)
- **Geológia**
 - Geo-klaster: V&V iniciatíva Compagnie Générale de Géophysique



➤ **Veľa rôznorodých aplikácií prenesených do EGEE**



- Čo je to gridové počítanie?
- Charakteristiky gridu
- Aplikácie
- **Gridové projekty a EGEE projekt**

Mnohé snahy vyvíjať Grid — na celom svete

- NASA Information Power Grid
- DOE Science Grid
- NSF National Virtual Observatory
- NSF GriPhyN
- DOE Particle Physics Data Grid
- NSF TeraGrid
- DOE ASCI Grid
- DOE Earth Systems Grid
- DARPA CoABS Grid
- NEESGrid
- DOH BIRN
- NSF iVDGL
- ...

- DataGrid (CERN, ...)
- EuroGrid (Unicore)
- DataTag (CERN,...)
- Astrophysical Virtual Observatory
- GRIP (Globus/Unicore)
- GRIA (Industrial applications)
- GridLab (Cactus Toolkit)
- CrossGrid (Infrastructure Components)
- EGSO (Solar Physics)
- ...

- UK – OGSA-DAI, RealityGrid, GeoDise, Comb-e-Chem, DiscoveryNet, DAME, AstroGrid, GridPP, MyGrid, GOLD, eDiamond, Integrative Biology, ...
- Netherlands – VLAM, PolderGrid
- Germany – UNICORE, Grid proposal
- France – Grid funding approved
- Italy – INFN Grid
- Eire – Grid proposals
- Switzerland - Network/Grid proposal
- Hungary – DemoGrid, Grid proposal
- Norway, Sweden – NorduGrid
- ...

European DataGrid (EDG)

www.edg.org



LHC Computing GRID (LCG)

cern.ch/lcg



CrossGRID

www.crossgrid.org



DataTAG

www.datatag.org



GridLab

www.gridlab.org



EUROGRID

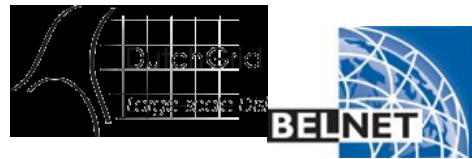
www.eurogrid.org



Európske národné projekty:

- INFN
- UK e-Science Programme
- NorduGrid



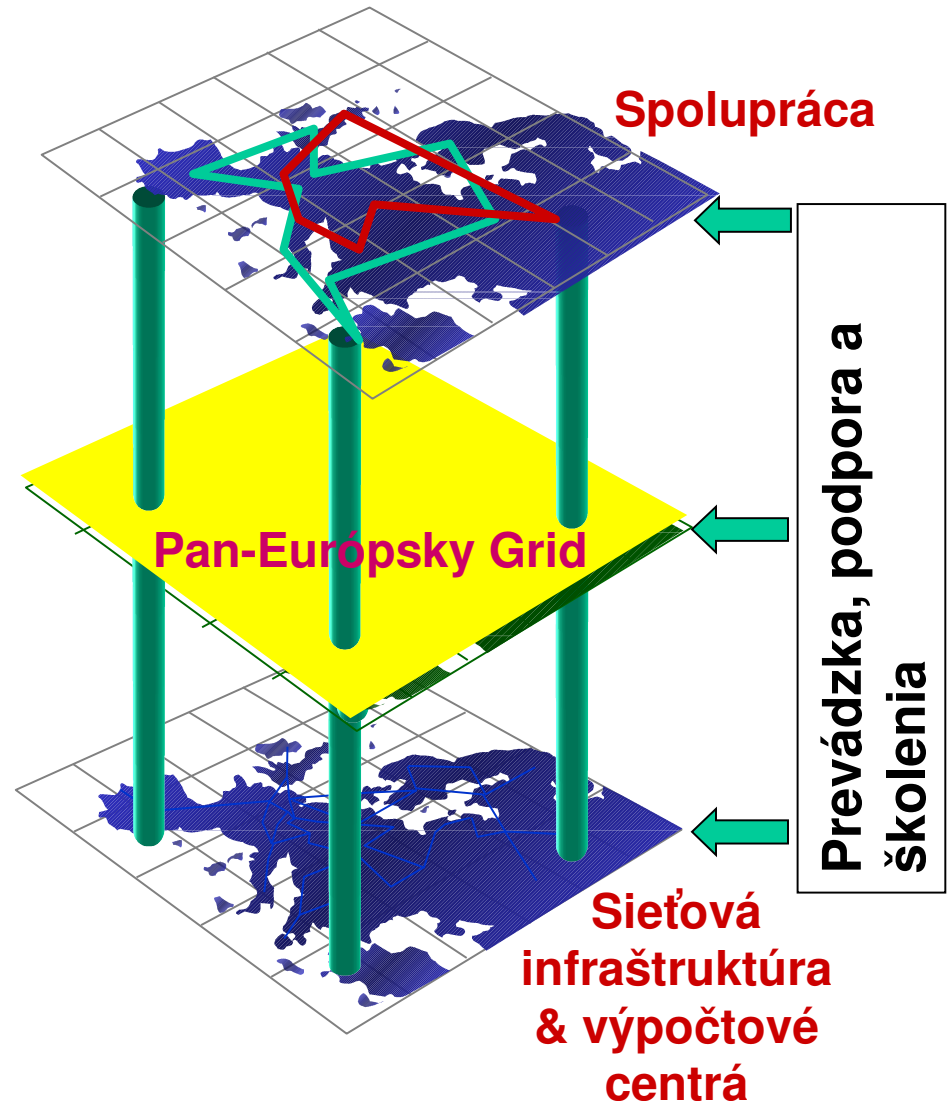


CroGrid



4+2 ročný program:

- Vybudovať, nasadiť a prevádzkovať konzistentné, robustné a rozsiahle produkčné gridové služby
 - Ktoré prepájajú a sú založené na národných, regionálnych a medzinárodných iniciatívach
- Zlepšiť a udržiavať middleware za účelom poskytovania spoľahlivej služby pre používateľov
- Pritiahnuť nových používateľov z výskumu aj priemyslu a zaistiť školenia a pomoc/podporu pre nich

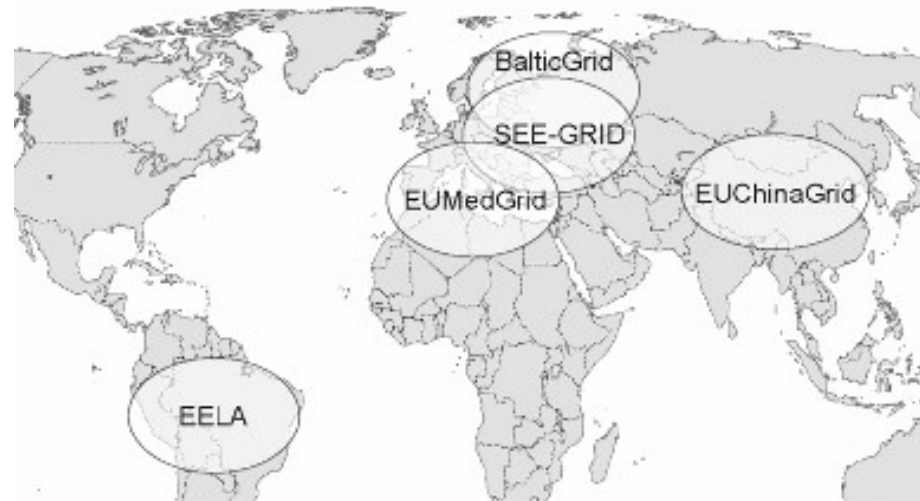


- vyše 90 partnerov
 - 32 krajín
 - 12 združení (federácií)
- Hlavné a národné gridové projekty v Európe, USA, Ázii



+ 27 ďalších krajín cez spriatelené projekty:

- BalticGrid
- SEE-GRID
- EUMedGrid
- EUChinaGrid
- EELA



Test-beds & Services

Certification testbeds (SA3)

Pre-production service

Production service

Infrastructure:

- Physical test-beds & services
- Support organisations & procedures
- Policy groups

Support Structures

Operations Coordination Centre

Regional Operations Centres

Global Grid User Support

EGEE Network Operations Centre (SA2)

Operational Security Coordination Team

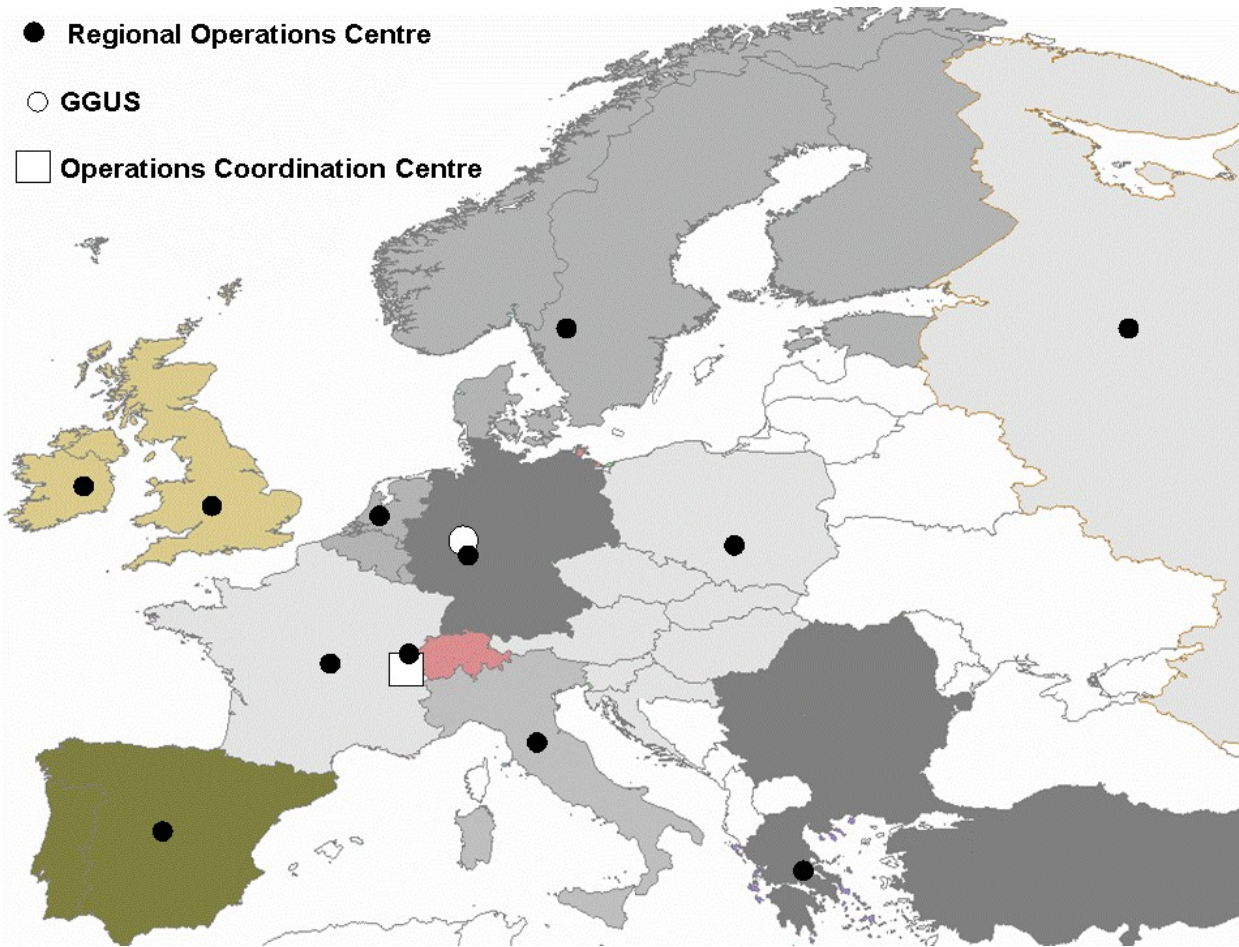
Security & Policy Groups

Joint Security Policy Group

EuGridPMA (& IGTF)

Grid Security Vulnerability Group

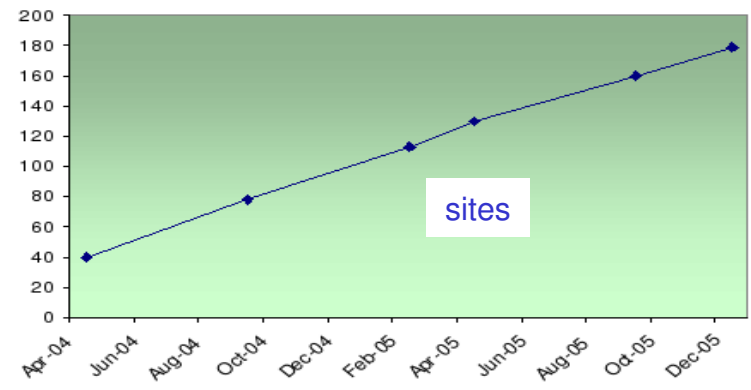
Operations Advisory Group (+NA4)



- Operations Coordination Centre (OCC)
 - management, oversight of all operational and support activities
- Regional Operations Centres (ROC)
 - providing the core of the support infrastructure, each supporting a number of resource centres within its region
 - **Grid Operator on Duty**
- Resource centres
 - providing resources (computing, storage, network, etc.);
- Grid User Support (GGUS)
 - At FZK, coordination and management of user support, single point of contact for users

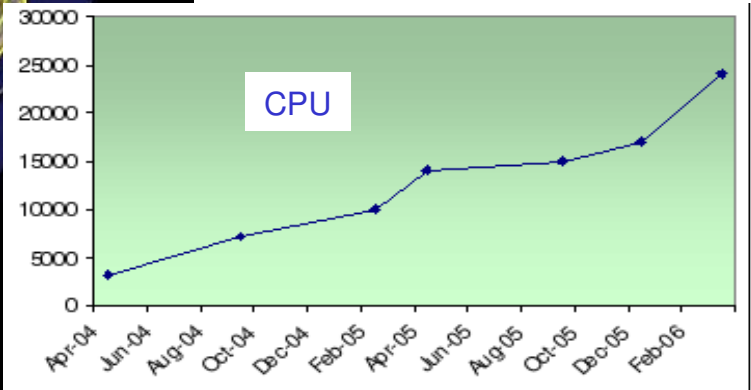


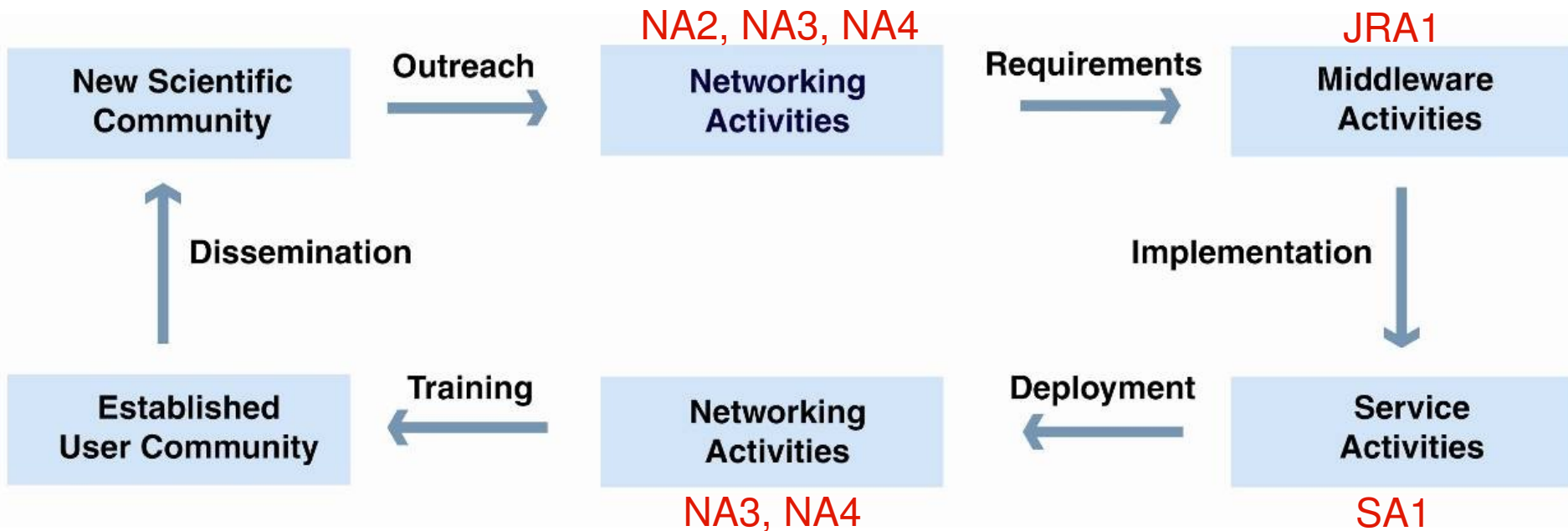
© 2006 Europa Technologies
 Image © 2006 NASA
 Image © 2006 TerraMetrics



Parametre infraštruktúry dnes:

- 249 uzlov v 50 krajinách
- ~56 000 CPU
- ~ 5 PB disk, + páskové MSS





Building effective user communities



User Interface (UI):

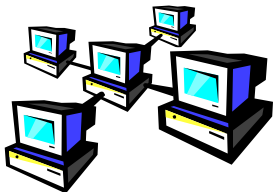
The place where users logon to the Grid



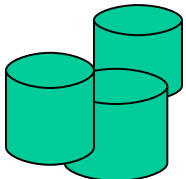
Resource Broker (RB): Matches the user requirements with the available resources on the Grid



Information System: Characteristics and status of CE and SE
(Uses “GLUE schema”)

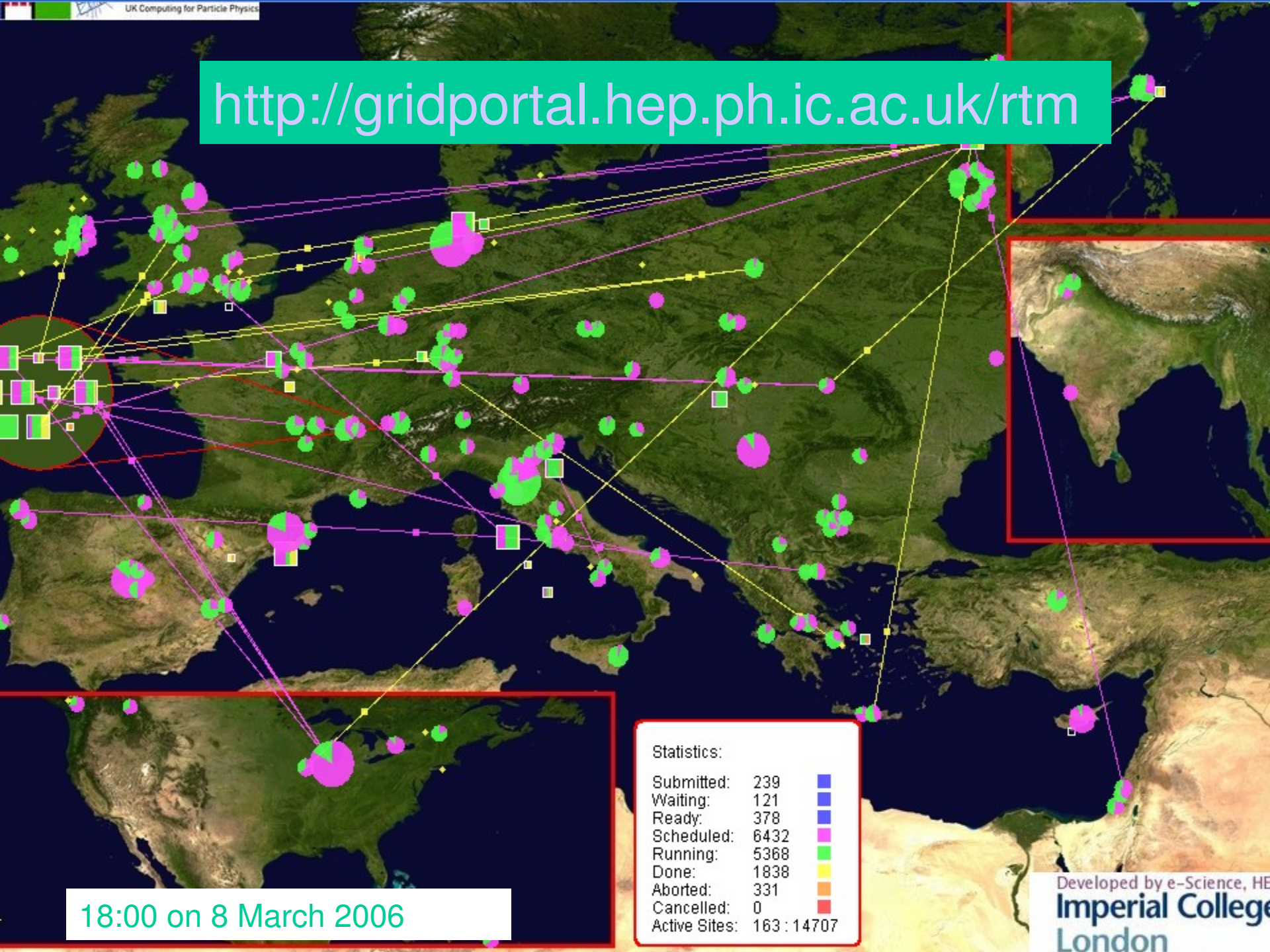


Computing Element (CE): A batch queue on a site’s computers where the user’s job is executed



Storage Element (SE): provides (large-scale) storage for files

<http://gridportal.hep.ph.ic.ac.uk/rtm>

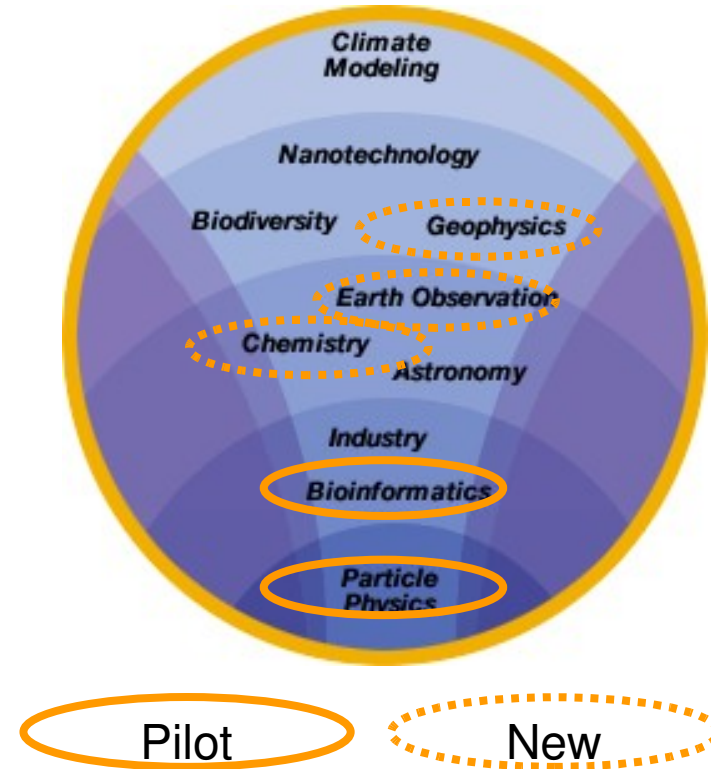


Statistics:

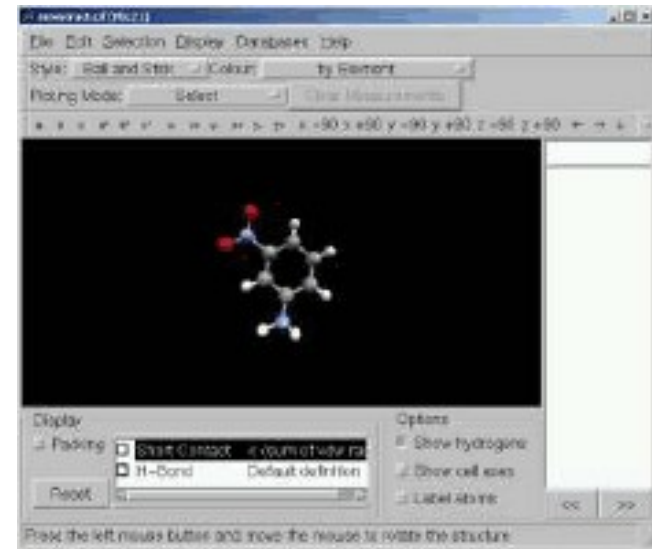
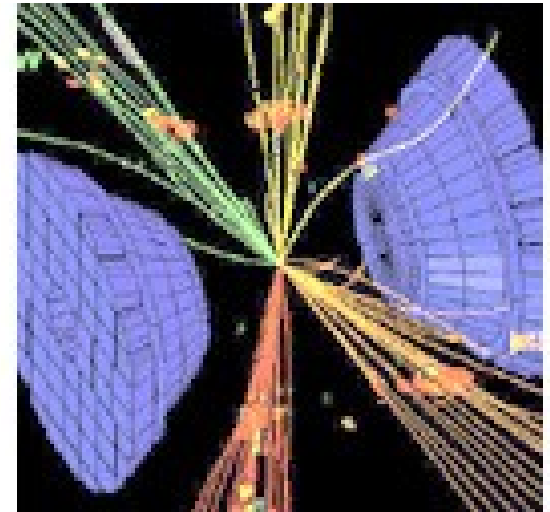
Submitted:	239	■
Waiting:	121	■
Ready:	378	■
Scheduled:	6432	■
Running:	5368	■
Done:	1838	■
Aborted:	331	■
Cancelled:	0	■
Active Sites:	163 : 14707	

18:00 on 8 March 2006

- Rozsah projektu EGEE: plná podpora pre akademické aplikácie (je taktiež otvorený pre priemyselný a socio-ekonomický svet)
- Hlavné kritérium úspechu EGEE projektu: koľko spokojných používateľov z koľkých rôznych oblastí ?
- Cieľ: 5000 používateľov (3000 po 2 rokoch) aspoň z 5-tich disciplín
- Dve vybrané pilotné aplikácie pomohli zhodnotiť výkonnosť a funkčnosť vyvíjanej infraštruktúry: Fyzika & Bioinformatika
- Ciele na prvé 2 roky boli splnené, aj prekročené (50 -> 180 uzlov)



- **Fyzika vysokých energií (High Energy Physics)** s výpočtovým gridom LHC (www.cern.ch/lcg) využíva gridovú infraštruktúru na ukladanie a analýzu petabajtov (10^{15} bajtov) reálnych a simulovaných údajov. LCG projekt je hlavným zdrojom požiadaviek a tvrdých termínov, ktoré sa nedajú konvenčne riešiť.
- **V biomedicíne** mnohé komunity čelia rovnako skľučujúcim výzvam zvládnuť záplavu bioinformatických a zdravotníckych údajov. Potrebujú mať prístup k veľkým a distribuovaným nehomogénnym údajom a majú dôležitú požiadavku na okamžité výkonné počítanie.



- **Prirodzené pokračovanie EGEE**

- Konzorcium bolo zväčšené
- Dôraz na poskytovanie infraštruktúry
 - zvýšená podpora pre aplikácie
 - prepojenie s inými infraštruktúrami
 - priemysel viac zapojený

SA: služby (service activities)

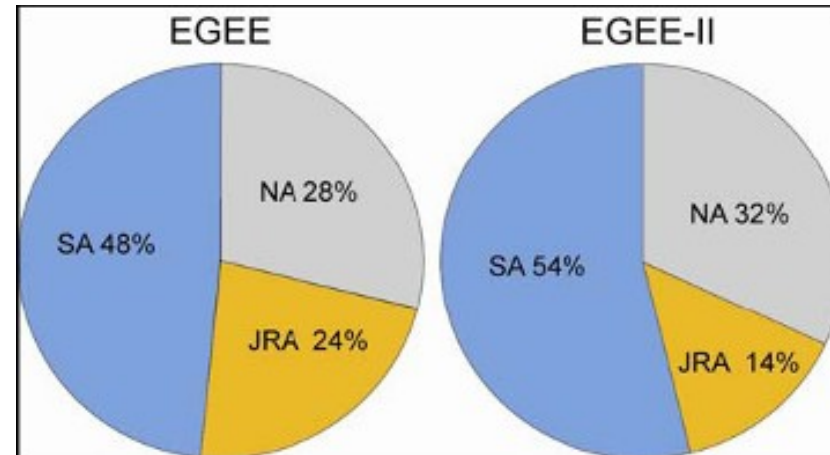
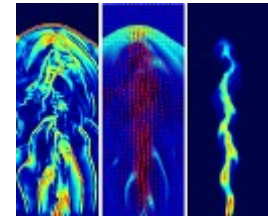
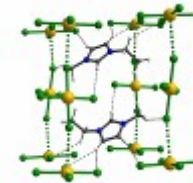
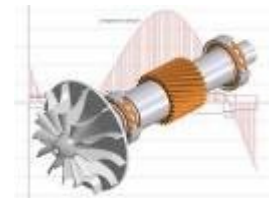
- zriadenie prevádzky

NA: „spoločenské“ aktivity (network)

- podporovanie virt.organizácií (VOs)

JRA: výskum (“joint research activities”)

- napr. spevnenie middleware



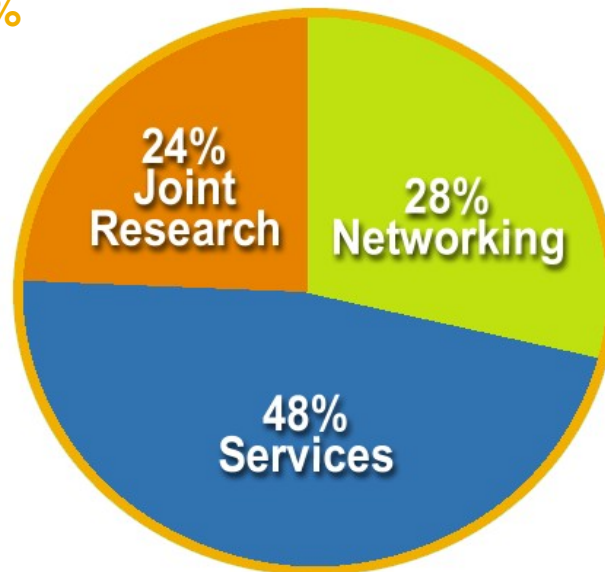
24% Združený výskum

JRA1: Príprava middleware a integrácia – **17%**

JRA2: Zabezpečenie kvality – **1.5%**

JRA3: Bezpečnosť – **3%**

JRA4: Vývoj sieťových služieb - **2.5%**



48% Služby

SA1: Prevádzka gridu, podpora a správa

SA2: Zabezpečovanie sieťovej infraštruktúry

28% Sieťové aktivity

NA1: Manažment

NA2: Disseminácia a propagácia

NA3: Trénovanie používateľov a vzdelávanie

NA4: Identifikácia aplikácií a ich podpora

NA5: Politika a medzinárodná spolupráca

Dôraz v EGEE je na prevádzke produkčného gridu a na podpore koncových používateľov

- **EGEE prevádzkuje najväčší multi-VO grid na svete**
- **vytvára “gridovú vrstvu” v e-infraštruktúre pre výskum, verejné služby a priemysel**
- **Kľúčové koncepty pre EGEE**
 - udržateľnosť – dlhodobé plánovanie
 - Produkčná kvalita
 - ...
- Gridy sú v podstate o ľuďoch
- ... ako ľudí v rôznych organizáciách prinútiť spolupracovať
- ... a ako táto spolupráca môže byť umožnená prevádzkou gridu, školeniami, podporou a (najviac meniacim sa?) middleware-om

Ďakujem za pozornosť.

egee.ui{AT}sav.sk

Miroslav Dobrucký

Ústav informatiky

Slovenská akadémia vied

Bratislava