Enabling GRID for Computer Algebra Applications

Victor Edneral
Skobeltsyn Institute of Nuclear Physics, Moscow State University
Moscow, Russia

Partially supported by grant CERN-INTAS 03-52-4297

Acknowledgements

- This talk includes slides from previous tutorials and talks delivered by:
  - the EDG training team
  - Dr Fabrizio Gagliardi, ***EGEE 1 Project Director***
  - Dr Bob Jones, ***EGEE 1Technical Director***
  - Oxana Smirnova
    JINR / Lund / CERN
- Prepared by Victor Edneral, SINP MSU

Overview

- Motivation for Grid Computing
- The basic ideas of Grid technology
- What is EGEE?
- Some Examples
- Grid and Computer Algebra

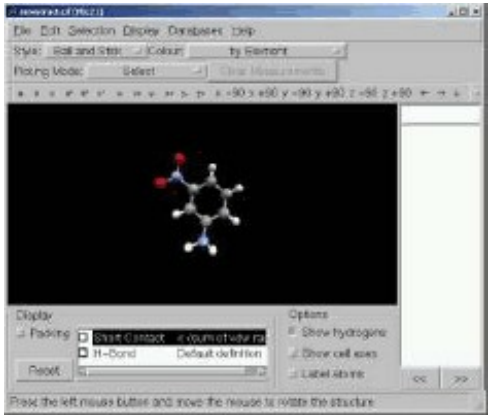
The terms of the problem

- Technological progress produces more sophisticated digital sensors (particle physics detectors, satellites, radio-telescopes, synchrotrons…)
- Much of science is therefore becoming increasingly “data-intensive”
- Huge amounts of data need to be analyzed by large and geographically distributed scientific communities
- Consequently, single computers, clusters or supercomputers are not powerful enough for the necessary calculations and the data processing

Result:

access to large facilities is difficult and expensive for the scientific community, particularly in less favored countries

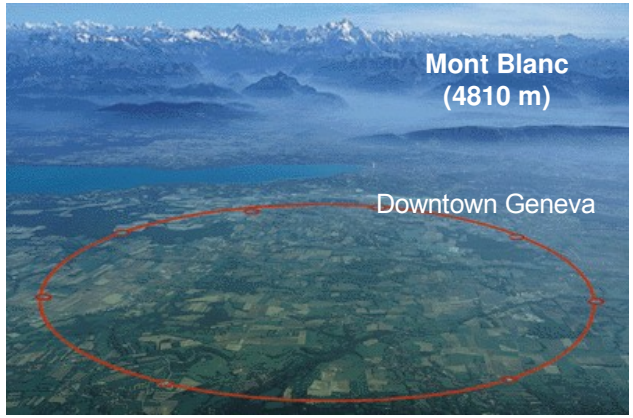
=> increase of the “electronic divide”



CERN: an example of data intensive science and a large international facility

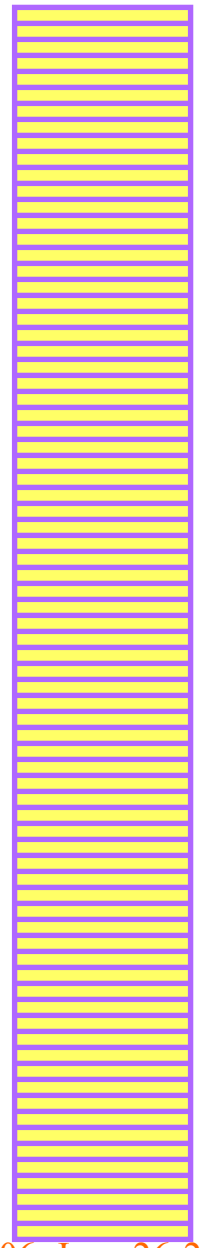
- CERN is building the Large Hadron Collider (LHC) the most powerful instrument ever built to investigate elementary particles physics
- Data Challenge:
  - One Megabyte of data digitised for each collision
  - 10^{10} collisions recorded each year = 10 Petabytes/year of data !!!
  - LHC data correspond to about **20 million CDs each year!**
- Simulation, reconstruction, analysis: LHC data handling requires a computing power equivalent to ~100,000 of today's fastest PC processors!

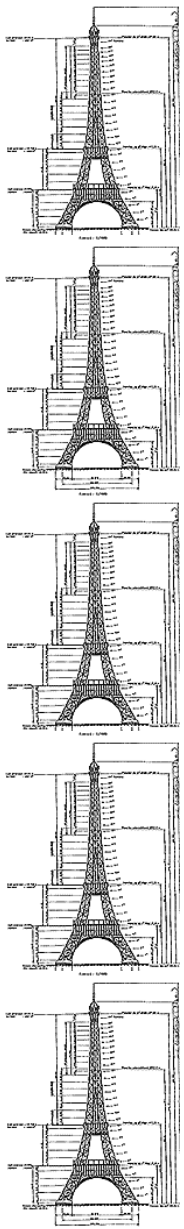
(10^6 mega; 10^9 giga; 10^{12} tera; 10^{15} peta)



Just a comparison…

6-8 Petabytes
~10.000.000 CD-ROM

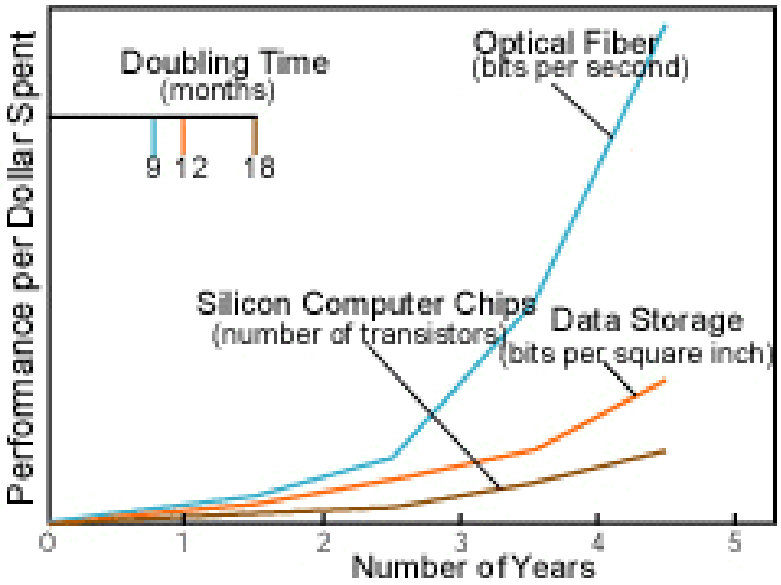




5 times the
Eiffel Tower
~1500 m

IT progress: some facts

- **Network vs. computer performance:**
  - Computer speed doubles every **18** months
  - Network speed doubles every **9** months
- **1986 to 2000:**
  - Computers: 500 times faster
  - Networks: 340000 times faster
- **2001 to 2010 (projected):**
  - Computers: 60 times faster
  - Networks: 4000 times faster



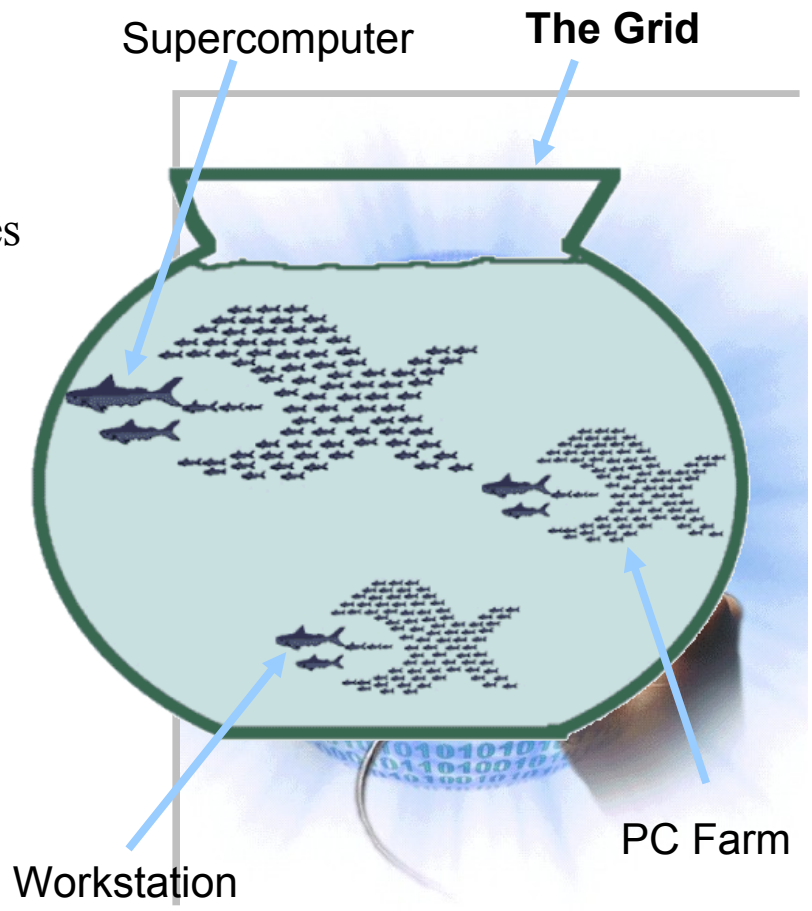
Bottom line: CPUs are fast enough; networks are very fast – gotta make use of it!

What is Grid computing?

- **"Coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations" (I.Foster)**
  - Resources are controlled by their owners
  - The Grid infrastructure provides access to collaborators
- **A *Virtual Organization* is:**
  - People from different institutions working to solve a common goal
  - Sharing distributed processing and data resources
- **Enabling *People* to Work Together on Challenging Projects**
  - Science, Engineering, Medicine, …
  - Public service, commerce too!

The Grid Paradigm

- Distributed supercomputer, based on commodity PCs and fast Networks
- Access to the great variety of resources by a single pass – certificate
- A possibility to manage distributed data in a synchronous manner
- A new commodity



Information Society Technologies

What is EGEE?

Enabling Grid to E-sciencE

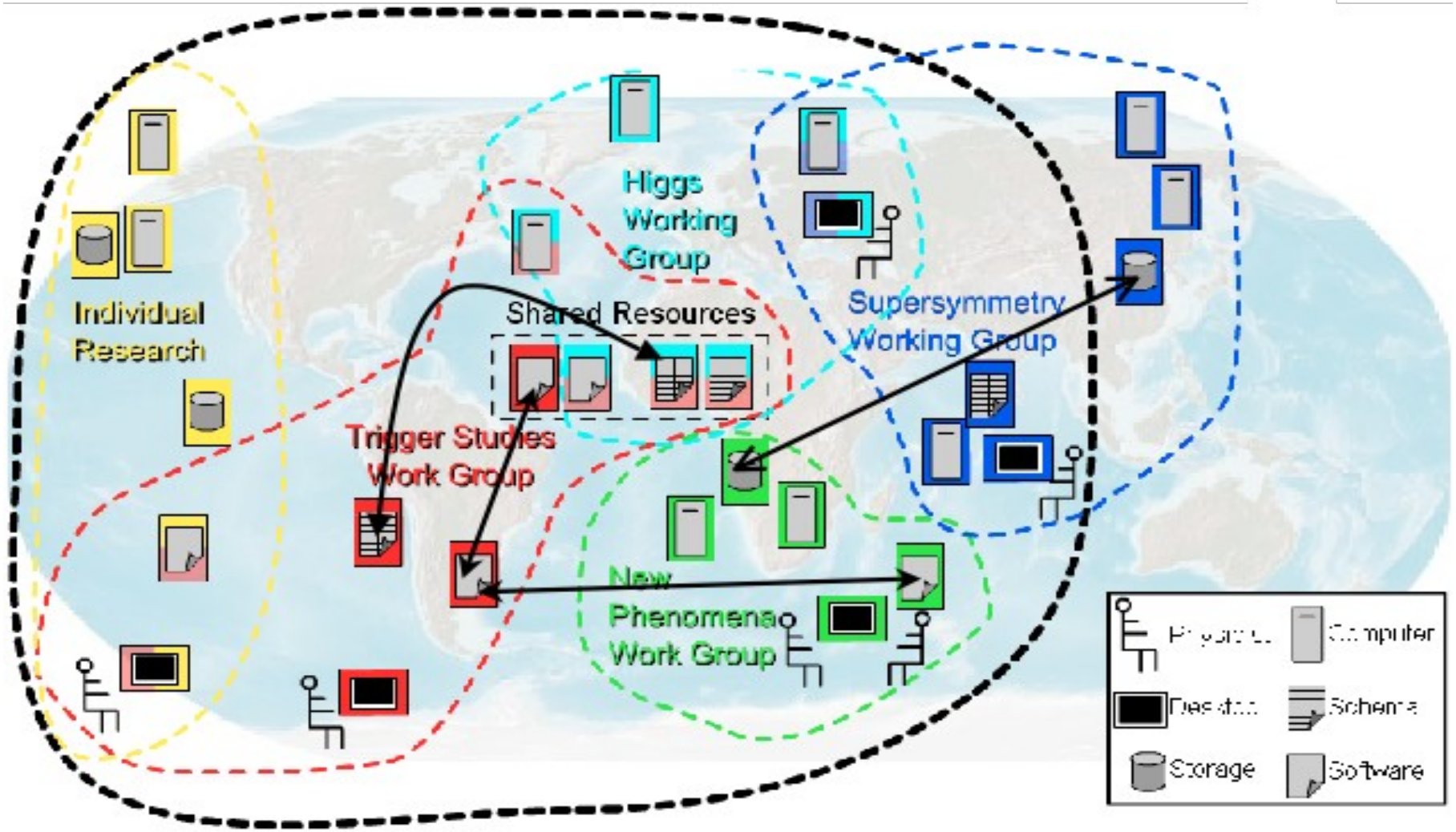
EGEE is funded by the European Union under contract IST-2003-508833

Complex Infrastructure

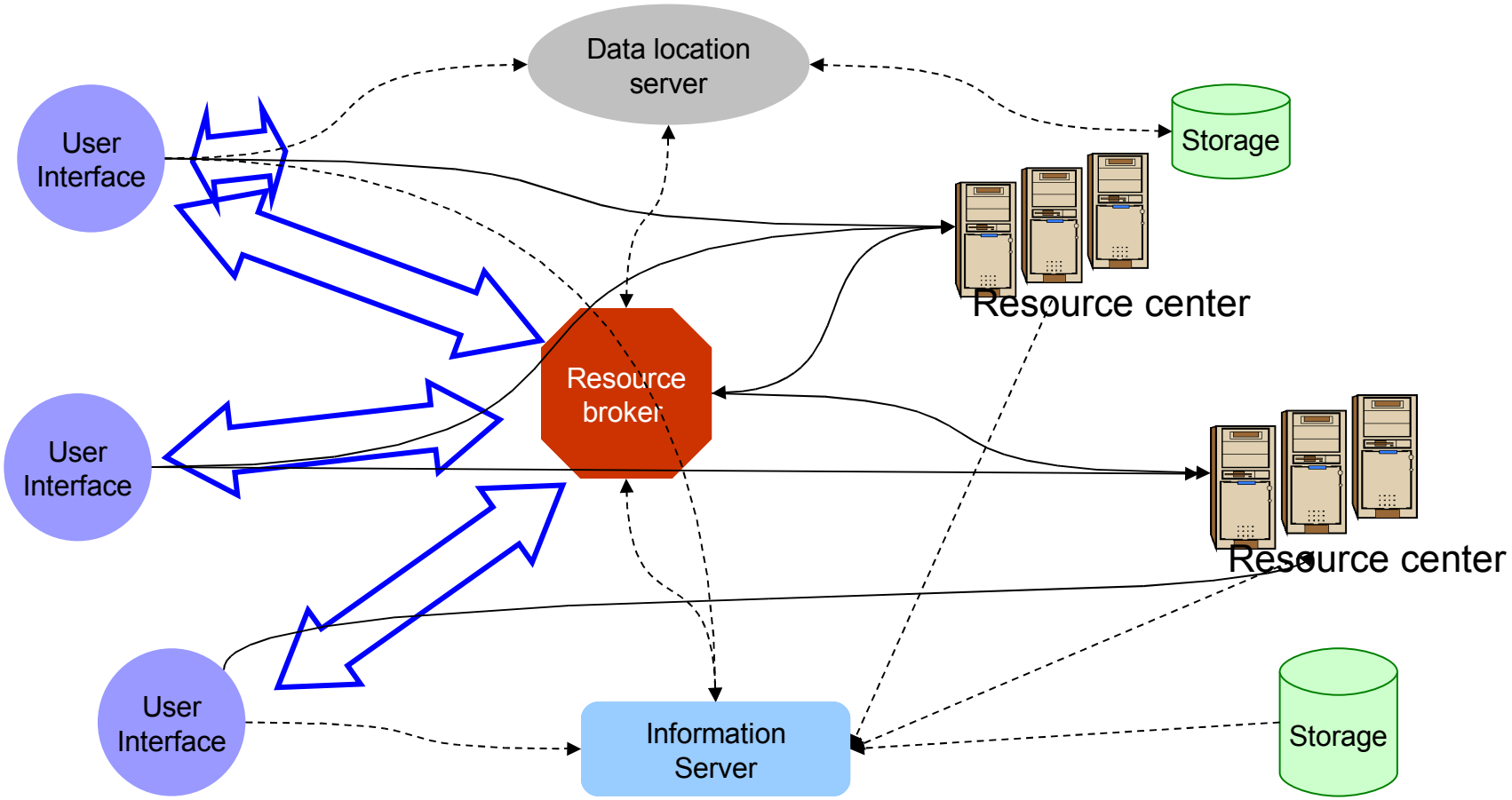
- **Users want access to compute power and data**
  - With security, reliability, trust, …
- **This requires a complex infrastructure**
  - Registries
  - Brokers
  - Administration
  - Policy
  - Negotiation
  - Etc.
- **Users shouldn't need to know the details**

Slide from Ian Foster's ssdbm 03 keynote

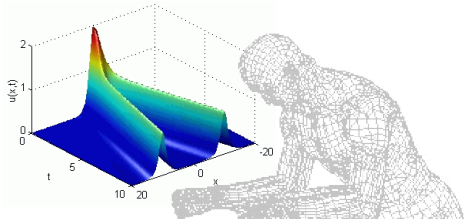
Virtual Organizations



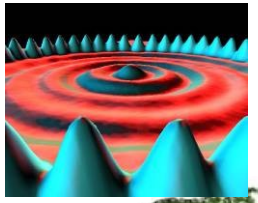
Some architectural thoughts



From Phase I to II



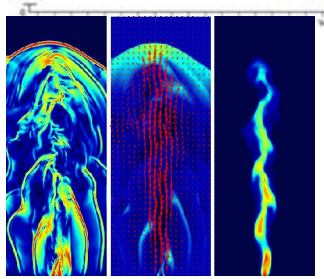
- **EGEE I.** ***April 2004***
  - Large scale deployment of EGEE infrastructure to deliver production level Grid services with selected number of applications

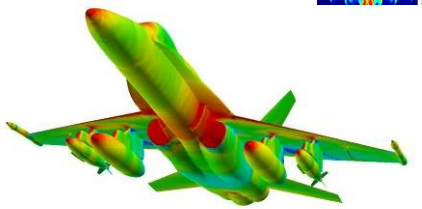


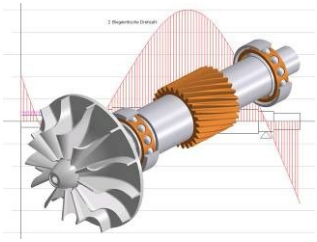
Enabling Grids for E-science in Europe

- **EGEE II.** ***April 2006***
  - Natural continuation of the project's first phase
  - Opening up to a larger user community
    - → increased multidisciplinary Grid infrastructure
    - → more involvement from Industry
  - Extending the Grid infrastructure world-w
    - → increased international collaboration

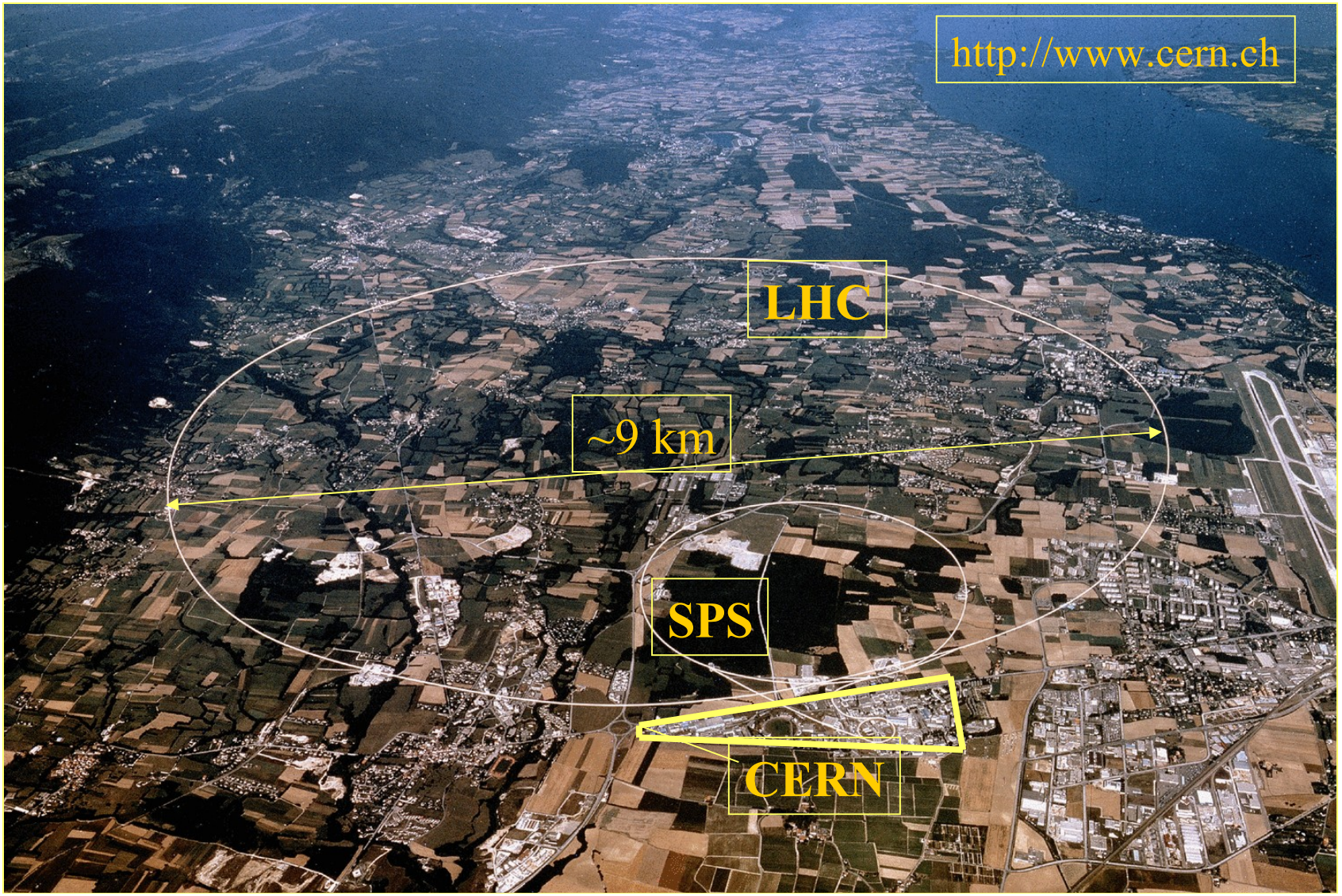
Enabling Grid to E-sciencE







The Large Hadron Collider



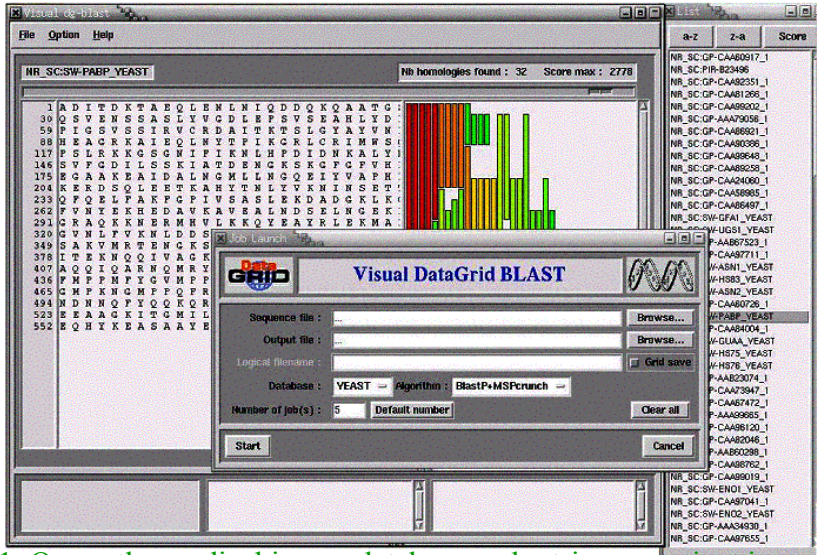
Applications

- **Medical and biomedical:**
  - Image processing (digital X-ray image analysis)
  - Simulation for radiation therapy
  - Protein folding
- **Chemistry**
  - Quantum
  - Organic
  - Polymer modelling
- **Climate studies**
- **Space sciences**
- **Physics:**
  - High Energy and other accelerator physics
  - Theoretical physics, lattice calculations of all sorts
  - Neutrino physics
  - Combustion
  - Thermonuclear synthesis
- **Genomics**
- **Material sciences**
- **Computer algebra?**

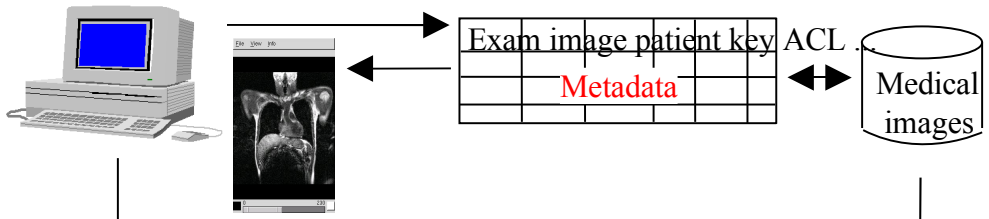
And many others

Bio-medicine applications

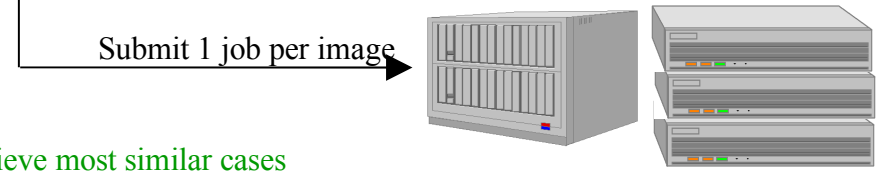
- **Bio-informatics**
  - Phylogenetics
  - Search for primers
  - Statistical genetics
  - Bio-informatics web portal
  - Parasitology
  - Data-mining on DNA chips
  - Geometrical protein comparison
- **Medical imaging**
  - MR image simulation
  - Medical data and metadata management
  - Mammographies analysis
  - Simulation platform for PET/SPECT



1. Query the medical image database and retrieve a patient image



2. Compute similarity measures over the database images



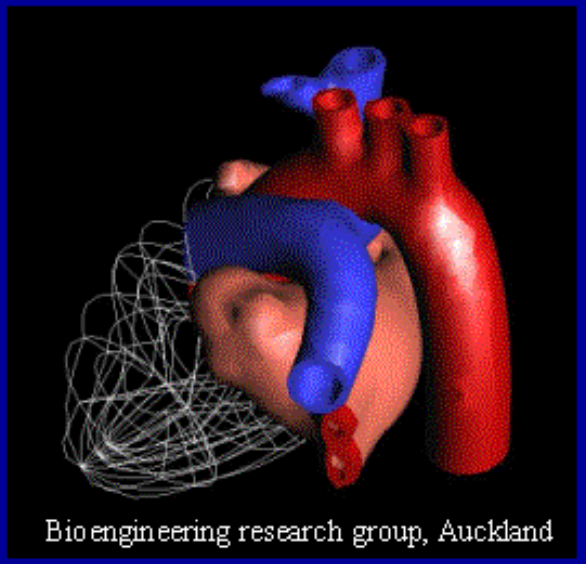
3. Retrieve most similar cases

Applications deployed
Applications tested
Applications under preparation

Bio-medicine applications

3.3 Heart Modeling

- Objectives: modeling heart anatomy, dynamics and physiology for heart image processing
  - bio-mecanical model
  - electrical model
  - very complex structure
  - biological scale out of range



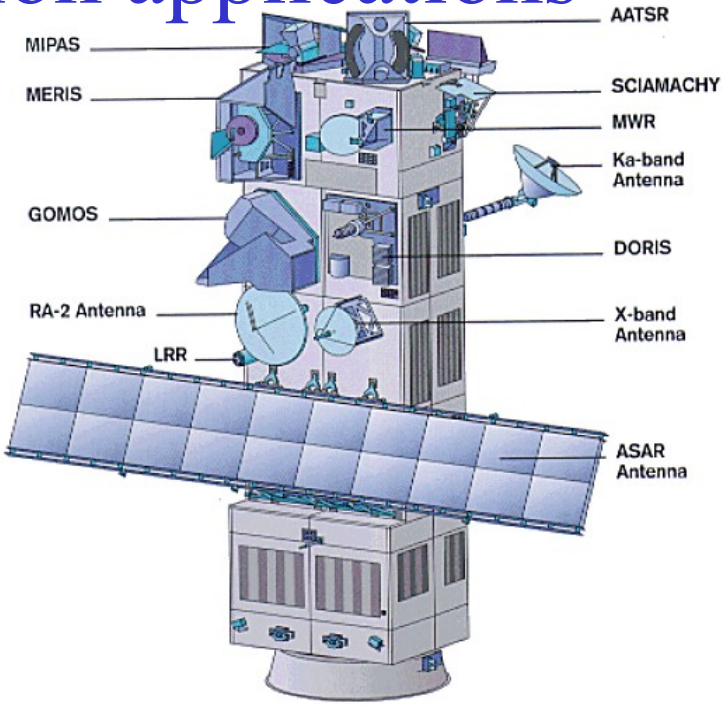
Bioengineering research group, Auckland

- Finite Element modeling
  - elements oriented in heart fibers direction: fine resolution
  - electrical propagation model based on bidomain theory
  - 4D model (3D+T)

Creatis

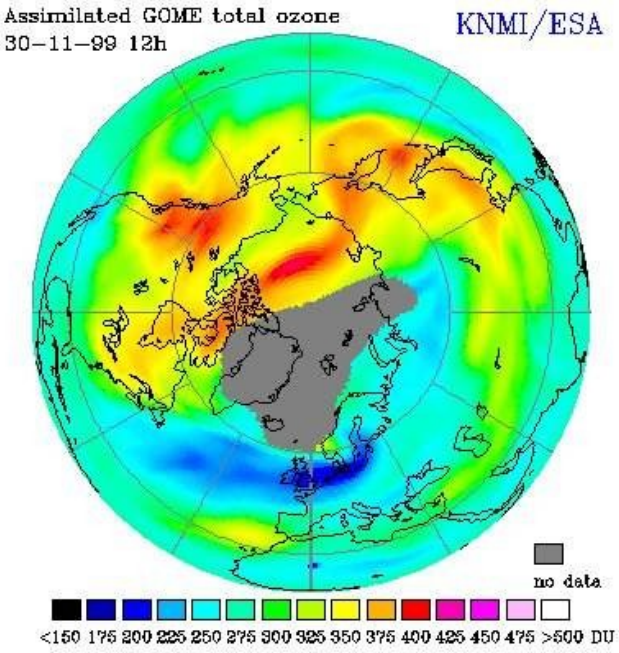
GGF1 - DataGrid WP 10 - March 2001

Earth observation applications



ESA missions:

- **about 100 Gbytes of data per day (ERS 1/2)**
- **500 Gbytes, for the ENVISAT mission (2002).**

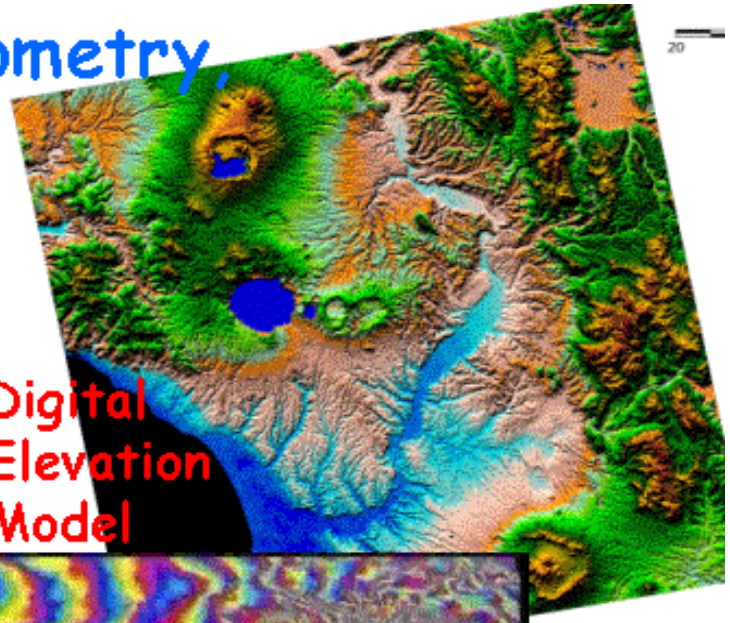


Grid contribute to EO:

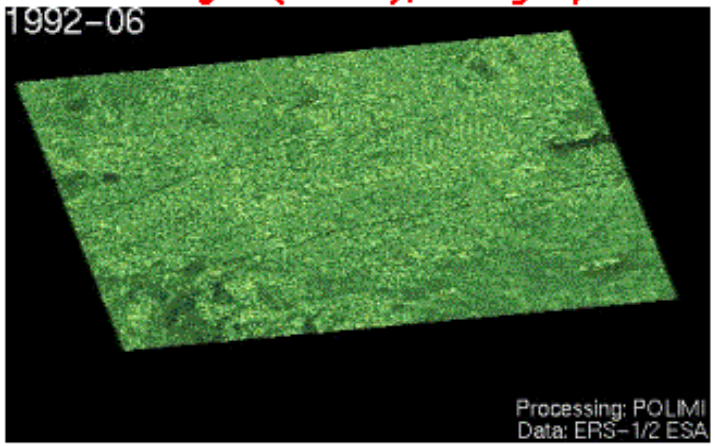
- **enhance the ability to access high level products**
- **allow reprocessing of large historical archives**
- **improve Earth science complex applications (data fusion, data mining, modelling ...)**

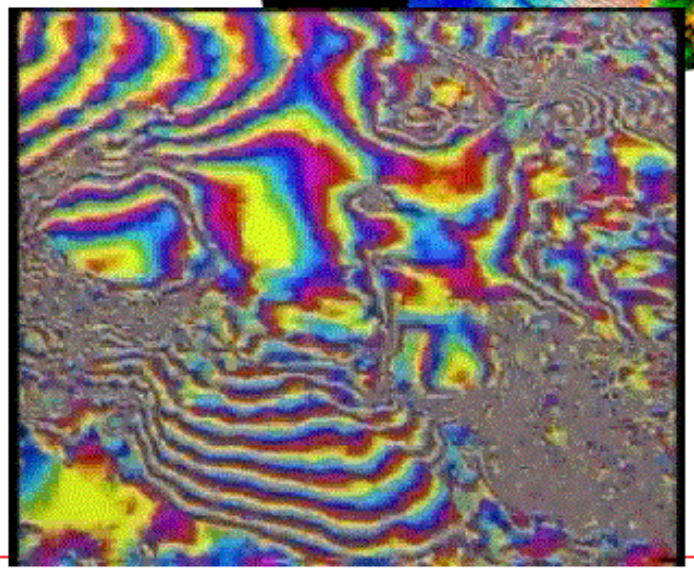
Earth observation applications

Number crunching: interferometry, subsidence, DEM generation



Pomona (Cal): subsidence velocity fields
40 ERS1/2 images (92-99), Ambiguity: 28 mm





GRID requirements:

- large data files (10+ GB)
- stages with intensive processing
- science driven value adding

Grid Geography

The Grid on This Monday

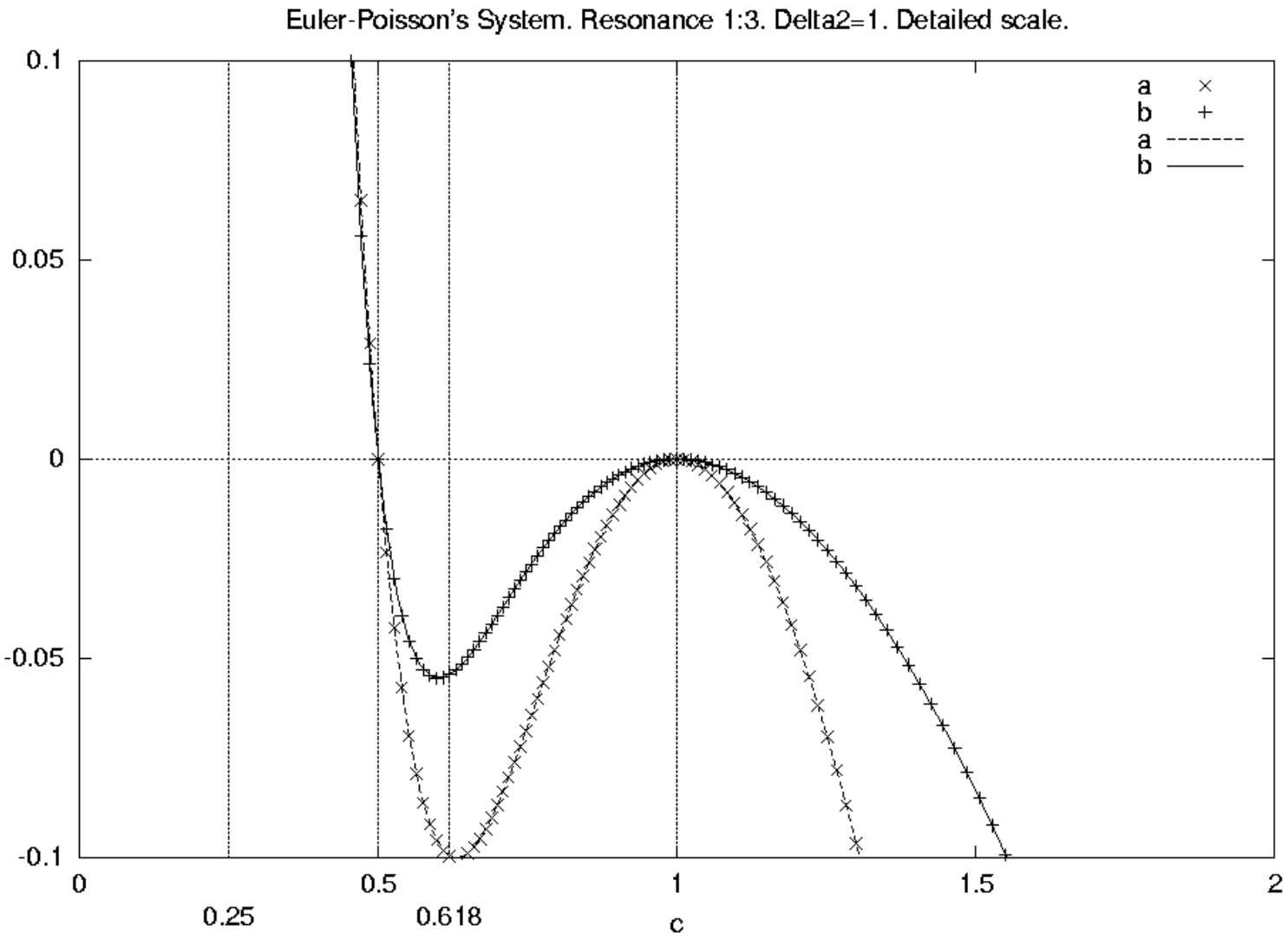
http://goc.grid.sinica.edu.tw/gstat/

182	UPV-GRyCAP	ramses.dsic.upv.es	.	.	ok	ok	ok	GLITE-3_0_0	2	2	0	0	0.12	0.00	26	20	OK ok
183	USC-LCG2	lcg-ce.usc.cesga.es	.	.	ok	ok	ok	LCG-2_7_0	126	3	133	0	0.05	0.00	126	123	OK ok
184	USCMS-FNAL-WC1	cmslcgce.fnal.gov	.	.	ok	warn	ok	GLITE-3_0_0	956	445	1030	580	0	0	984	745	OK ok
185	VU-MIF-LCG2	grid2.mif.vu.lt	ok	ok	ok	note	ok	GLITE-3_0_0	30	23	0	0	0.14	0	32	27	OK info
186	WARSAW-EGEE	ce.polgrid.pl	.	.	ok	ok	ok	LCG-2_7_0	206	126	16	35	3.20	0	206	147	OK .
187	WCSS	ce.egee.wcss.wroc.pl	.	.	ok	ok	ok	LCG-2_7_0	12	0	11	29	0.03	0.00	12	11	OK ok
188	WEIZMANN-LCG2	wipp-ce.weizmann.ac.il	ok	ok	ok	ok	ok	LCG-2_7_0	49	2	17	219	0.12	0.02	49	48	CT ok
189	wuppertalprod	grid-ce.physik.uni-wuppertal.de	.	.	ok	ok	ok	GLITE-3_0_1	299	56	243	259	0.84	0.17	410	189	OK ok
							sites	countries	totalCPU	freeCPU	runJob	waitJob	seAvail TB	seUsed TB	maxCPU	avgCPU	
						Total	189	43	25239	11183	11535	258457	103343.01	4615.92	43458	27664	

Grid for Computer Algebra

- **Which problems**
  - Many similar calculations
    - Experimental Mathematics. Scanning some domain in parameter space. For example searching domains of integrability of an ODE system.
    - Massive calculation of Feynman diagrams
- **Which software we can use**
  - Batch calculations
  - Statically compiled programs for testing
  - GNU license

Case of Resonance (1:3)



Russian Data Intensive Grid

95	ITEP	ceitep.itep.ru	ok	ok	ok	ok	ok	GLITE-3 0 0	64	2	62	185	4.41	0.79	64	62	OK	ok
98	JINR-LCG2	lcgce01.jinr.ru	ok	ok	ok	ok	ok	GLITE-3 0 0	30	2	29	251	14.89	0.42	30	29	OK	ok
121	RRC-KI	gate.grid.kiae.ru	.	.	ok	ok	ok	LCG-2 7 0	80	1	79	10	0.69	0.09	80	78	OK	ok
122	ru-IMPB-LCG2	lcgce.psn.ru	.	.	ok	ok	ok	LCG-2 7 0	26	3	25	9	0.66	0.01	30	27	OK	ok
123	ru-Moscow-FIAN-LCG2	ce1.grid.lebedev.ru	.	.	ok	ok	ok	LCG-2 7 0	12	2	10	0	1.69	0.00	12	11	CT	ok
124	ru-Moscow-GCRAS-LCG2	grid8.wdcb.ru	.	.	ok	ok	ok	GLITE-3 0 0	4	4	0	0	0.09	0.00	8	6	JS	info
125	RU-Moscow-KIAM-LCG2	ce.keldysh.ru	.	.	ok	ok	ok	LCG-2 7 0	16	1	15	5	1.26	0.00	16	15	OK	ok
126	ru-Moscow-SINP-LCG2	lcg06.sinp.msu.ru	ok	ok	ok	ok	ok	LCG-2 7 0	94	57	37	0	4.24	0.78	106	96	OK	ok
127	RU-Phys-SPbSU	ce.phys.spbu.ru	.	.	ok	ok	ok	LCG-2 7 0	4	4	0	0	0.08	0.00	4	3	OK	.
128	ru-PNPI-LCG2	cluster.pnpi.nw.ru	.	.	ok	ok	ok	GLITE-3 0 0	58	0	60	132	1.66	0.07	60	59	OK	.
129	RU-Protvino-IHEP	ce0001.m45.ihep.su	.	.	ok	ok	ok	LCG-2 7 0	85	0	81	6	2.78	0.17	90	70	OK	info
130	RU-SPbSU	alice01.spbu.ru	.	.	ok	ok	ok	LCG-2 7 0	8	8	0	2	0.73	0.15	8	7	OK	.
131	Ru-Troitsk-INR-LCG2	grce001.inr.troitsk.ru	.	.	ok	ok	ok	LCG-2 7 0	10	5	5	2	1.24	0.06	10	9	OK	.

491 Processors 34.42 TByte

Virtual Organization

- **RDIG rather than EGEE**
- **The calg VO at Skobeltsyn Institute of Nuclear Physics (MSU) & Joint Institute of Nuclear Research (Dubna)**
- **We have hardware now**
  - **124 processors x (1-4) GByte RAM**
  - **19.13 TByte mass storage**
- **Appropriate software**
  - **CompHEP (SINP MSU)**
  - **GINV (JINR)**
  - **Singular**
  - **LISP**

To know more about EGEE:

http://www.eu-egee.org

Questions?