



PLGrid PLUS:

Toward Domain-Specific Infrastructure for Supporting International Research Collaboration

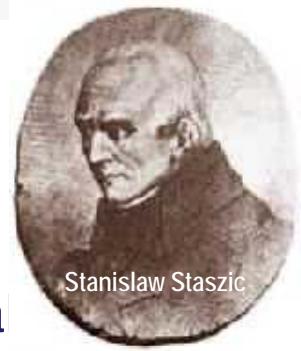
Jacek Kitowski, Tomasz Szepieniec, Mariusz Sterzel
and Lukasz Dutka

ACK Cyfronet AGH, Krakow, Poland
PL-Grid Consortium

EGI Community Forum 2012, Munchen, March 26-30, 2012



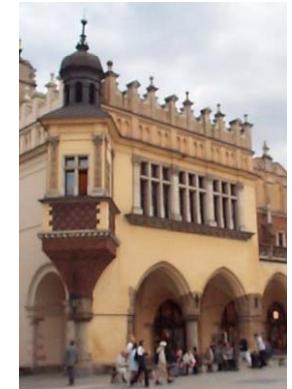
General Information



◆ Krakow - old university centre – former Polish Capital

- ◆ Jagiellonian University (1364)
- ◆ AGH University of Science and Technology (1919)
 - 45,000 students in 200+ courses
 - 16 faculties
 - Many departments
 - Institute (Department) of Computer Science (1980)
- ◆ many others universities and technology centres....
- ◆ Academic Computer Centre CYFRONET-AGH (1973)

◆ Close collaboration between Cyfronet and AGH Institute of Computer Science for many years



Outline



3

- Some basis
 - PL-Grid Consortium
- Road Map of Activities
- Conclusions

PL-Grid Consortium

4

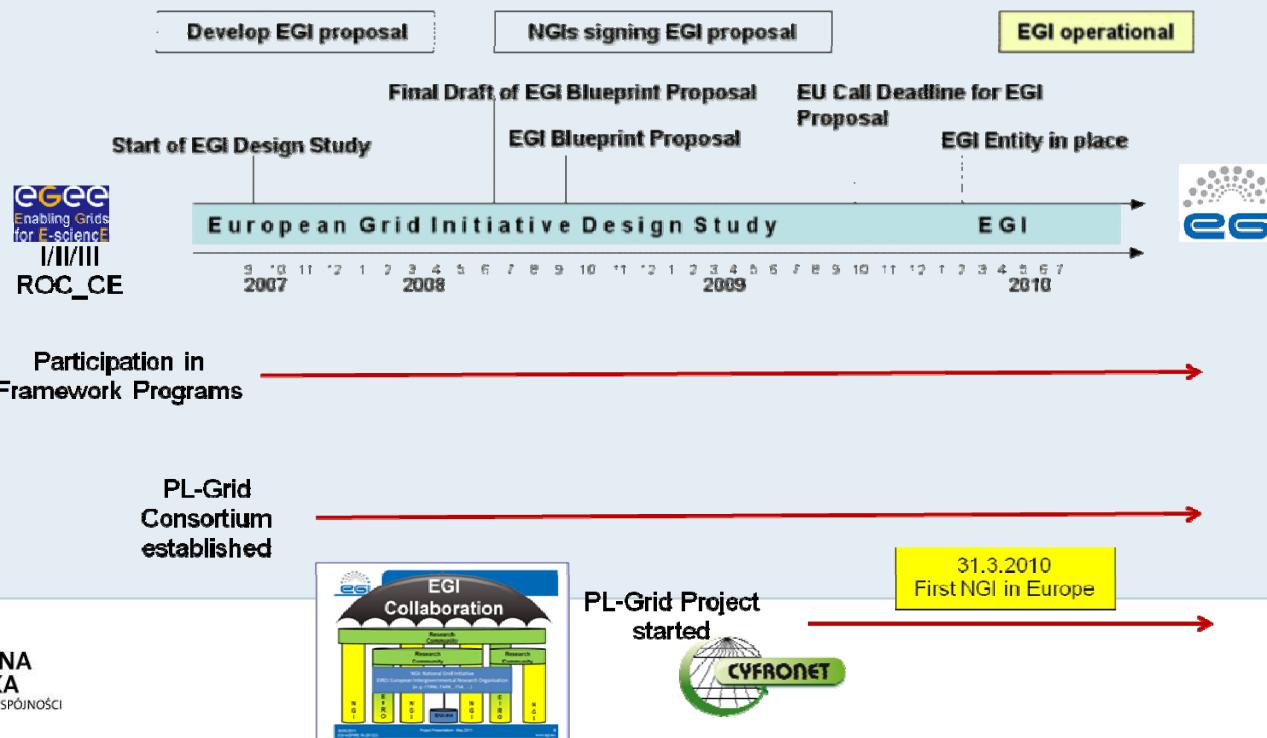


Consortium Creation – January 2007

- Consortium members – make up of 5 Polish centres
- Motivation: World progress in Big Science
(theory, experiments, simulation, 4th paradigm)

PL-Grid Project funded March 2009 by the European Regional Development Fund as part of the Innovative Economy Program

- Duration: 1.1.2009 – 31.3.2012
- Budget: total 21 M€, from EC 17M€

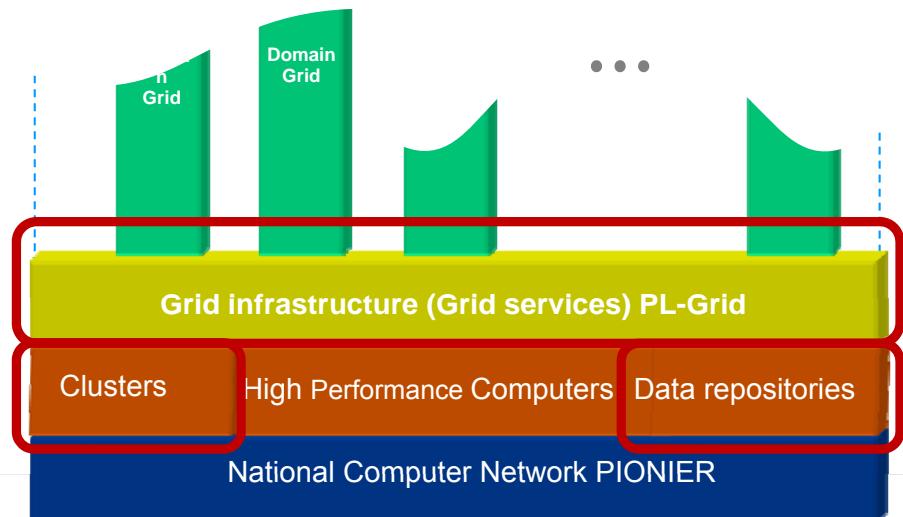


Main Objectives of PL-Grid Project



- Development of a common base infrastructure – National Grid Infrastructure (NGI_PL) – internationally compatible
- Potential capacity to construct specialized, domain Grid systems
- Enabling efficient use of available financial resources
- Plans for HPC and Scalability Computing, including clouds environments

*Polish Infrastructure
for Supporting Computational Science
in the European Research Space*

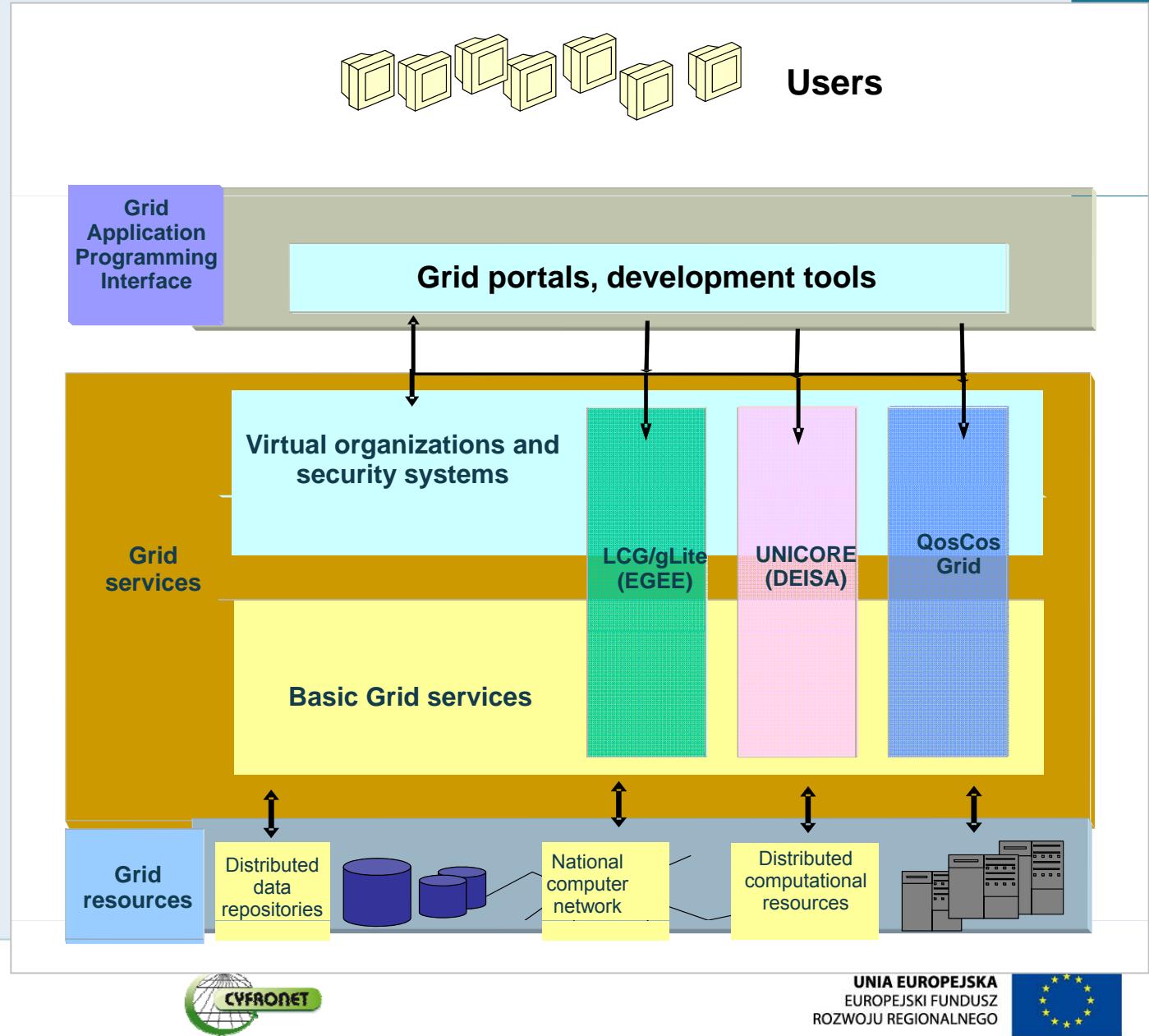


PL-Grid Building Blocks

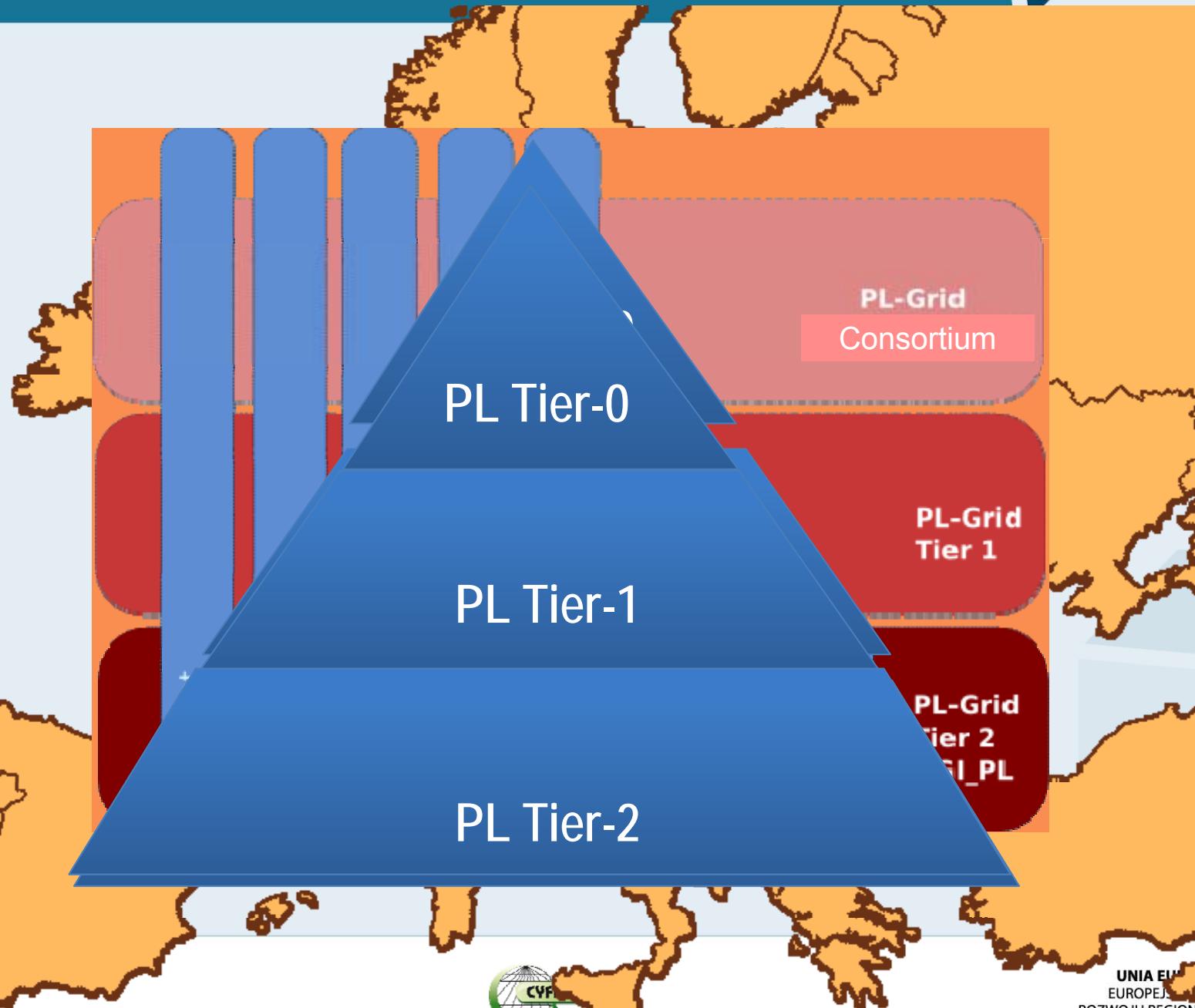


PL-Grid software comprises:

- Unique User Tools
- Unique composition of 3 middleware layers
- software libraries
- Virtual organization systems
- Data/Resource management systems
- Helpdesk system
- System for grants award



European Dimension

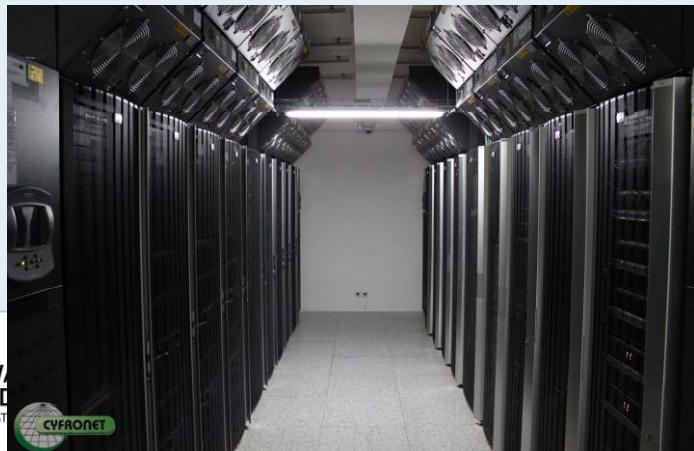


Partners' Computing Resources



TOP500 – November 2011

Rank	Site	System	Cores	Rmax TFlop/s	Rpeak TFlop/s	Power (KW)
88	Cyfronet Kraków	Zeus - Cluster Platform 3000 BL 2x220, Xeon X5650 6C 2.66 GHz, Infiniband, HP	15264	128.8	162.4	
279	TASK Gdańsk	Galera Plus - ACTION Xeon HP BL2x220/BL490 E5345/L5640 Infiniband, ACTION	10384	65.6	97.8	
296	ICM Warsaw	Boreas - Power 775, POWER7 8C 3.84 GHz, Custom, IBM	2560	64.3	78.6	156.7
298	PCSS Poznań	Rackable C1103-G15, Opteron 6234 12C 2.40 GHz, Infiniband QDR, SGI	5640	63.9	136.4	
360	WCSS Wrocław	Supernova - Cluster Platform 3000 BL2x220, X56xx 2.66 Ghz, Infiniband, HP	6348	57.4	67.5	



Aggregated Status



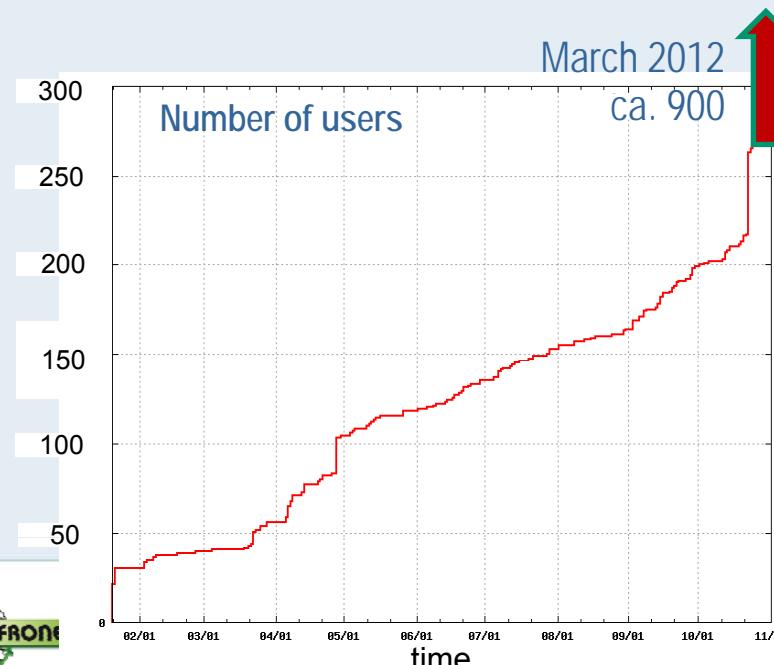
- ❑ Total number of cores (static) 26,000+
- ❑ Number of users (March 2012) 900+
- ❑ Number of jobs per month 750,000-1,500,000
- ❑ High level of availability and reliability
- ❑ Papers published or accepted
 - ❑ Journals: 25+
 - ❑ PL-Grid Book (Springer): 24 chapters

At each Partner site

- Computer Rooms
- Air Conditioning
- Power Lines
- UPS....



Site	Site Size	Availability		
		September	August	July
NGI_PL	24410	89%	80%	97%
CAMK	N/A	46%	98%	100%
CYFRONET-LCG2	10656	85%	98%	97%
IFJ-PAN-BG	12	88%	96%	98%
PSNC	5952	90%	95%	90%
TASK	80	95%	82%	100%
WARSAW-EGEE	2858	64%	95%	100%
WCSS-PPS	4	95%	93%	100%
WCSS64	4848	95%	44%	96%



Selected Tools/Software Achievements

Innovative Infrastructure Environment



- ◆ Efficient Resource Allocation
 - ◆ Grid Resource Bazaar
- ◆ Experimental Workbench
 - ◆ GridSpace2
 - ◆ InSilicoLab
- ◆ Tools and Middleware
 - ◆ Integration of the **Migrating Desktop**, **VineToolkit** and **gEclipse** tools with various PL-Grid middleware services
 - ◆ **QStorMan Toolkit**
 - ◆ Novel Grid Middleware **QosCosGrid**
 - ◆ Integration of the selected tools and web applications with Liferay **Portal** framework and Nagios monitoring system
 - ◆ **HelpDesk** Portal for the users
- ◆ Software Packages implemented
 - ◆ Biology, quantum chemistry, physics, numerical computation, simulation, ,...

Availability of Scientific Software Packages

Porting to PL-Grid Environment



- Access to software packages is provided to users through:
 - gLite
 - UNICORE
 - QCG
- Examples of available packages in various fields:
 - **biology:** AutoDock, BLAST, ClustalW2, CPMD, Gromacs, NAMD
 - **quantum chemistry:** ADF, CFOUR, Dalton, GAMESS, Gaussian, Molcas, Molpro, MOPAC, NWChem, OpenBabel, Siesta, TURBOMOLE
 - **physics:** ANSYS FLUENT, Meep
 - **numerical computations and simulation:** Mathematica, MATLAB, OpenFOAM
 - **other:** Blender, POV-Ray

Selected Use Cases

out of many



Biology

- Main fields

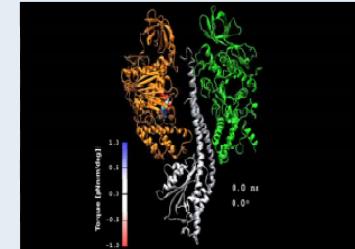
- Biology
- Quantum Chemistry
- Nanotechnology and Material Science
- High Energy Physics
- Astronomy

- Resource utilisation (2010 example)

- Antibiotic simulation – 165 CPU-year
- Molecular simulation – 21 CPU-year
- Modelling of chemical reactions – 17 CPU-year

Wide scope of research

- Protein structures
- Protein folding
- DNA research
- Drug research
- Electrical phenomena in heart
- Biopolymers
- Effects in brain and eye simulation
- Tuberculosis research
-



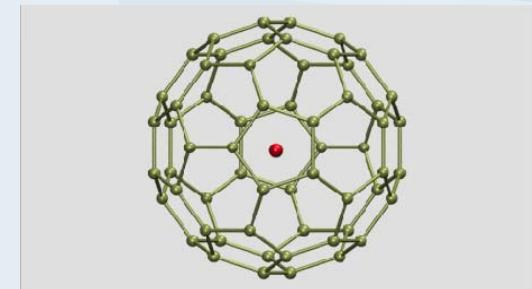
Antimycosis research
Gdansk University of Technology
165 CPU-year

- Accelerated research according to specific workflow and software structure (speedup = 5=)

ACK: Jacek Czub, Anna Neumann, PG

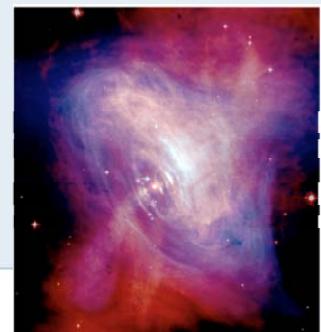
Quantum chemistry

- Simulation of electronic structure of molecules
- Example: calculation of electronic structure and oscillations for fulleren C₆₀ – Cu interaction
- Usage of Turbomole package
 - Available: ADF and Gaussian



Astronomy

- Cherenkov Telescope Array (CTA)
- Network of detectors 10GeV – 100TeV
- ESFRI Project



- Data acquisition by instruments: CANGAROO, H.E.S.S., MAGIC, MILAGRO and VERITAS
- Data kept by infrastructure and analyzed by our users

Physics

- Collaboration with CERN in all LHC experiments
 - Atlas, ALICE, CMS, LHCb
- Collaboration with nEDM project (12 partners)
- Polish-European VO with 20TB storage





Dziedzinowo zorientowane
usługi i zasoby infrastruktury
PL-Grid dla wspomagania
Polskiej Nauki w Europejskiej
Przestrzeni Badawczej

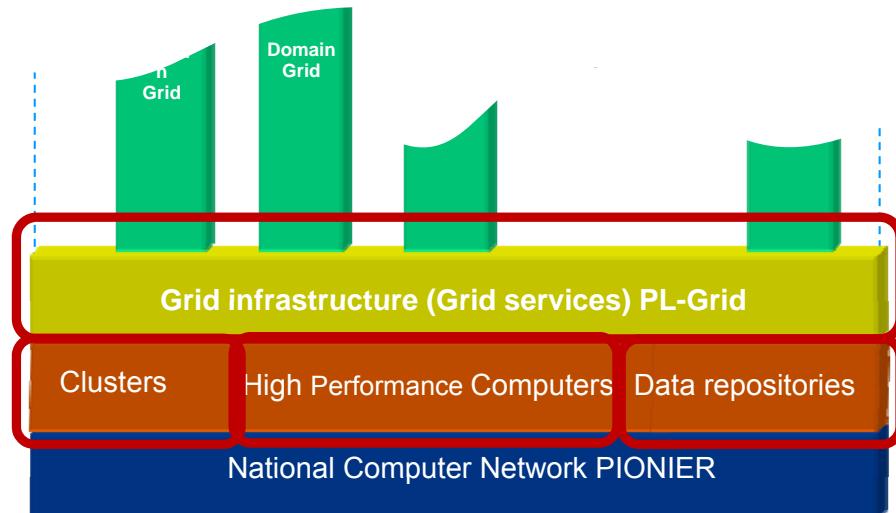
What's next ?

PL-Grid PLUS



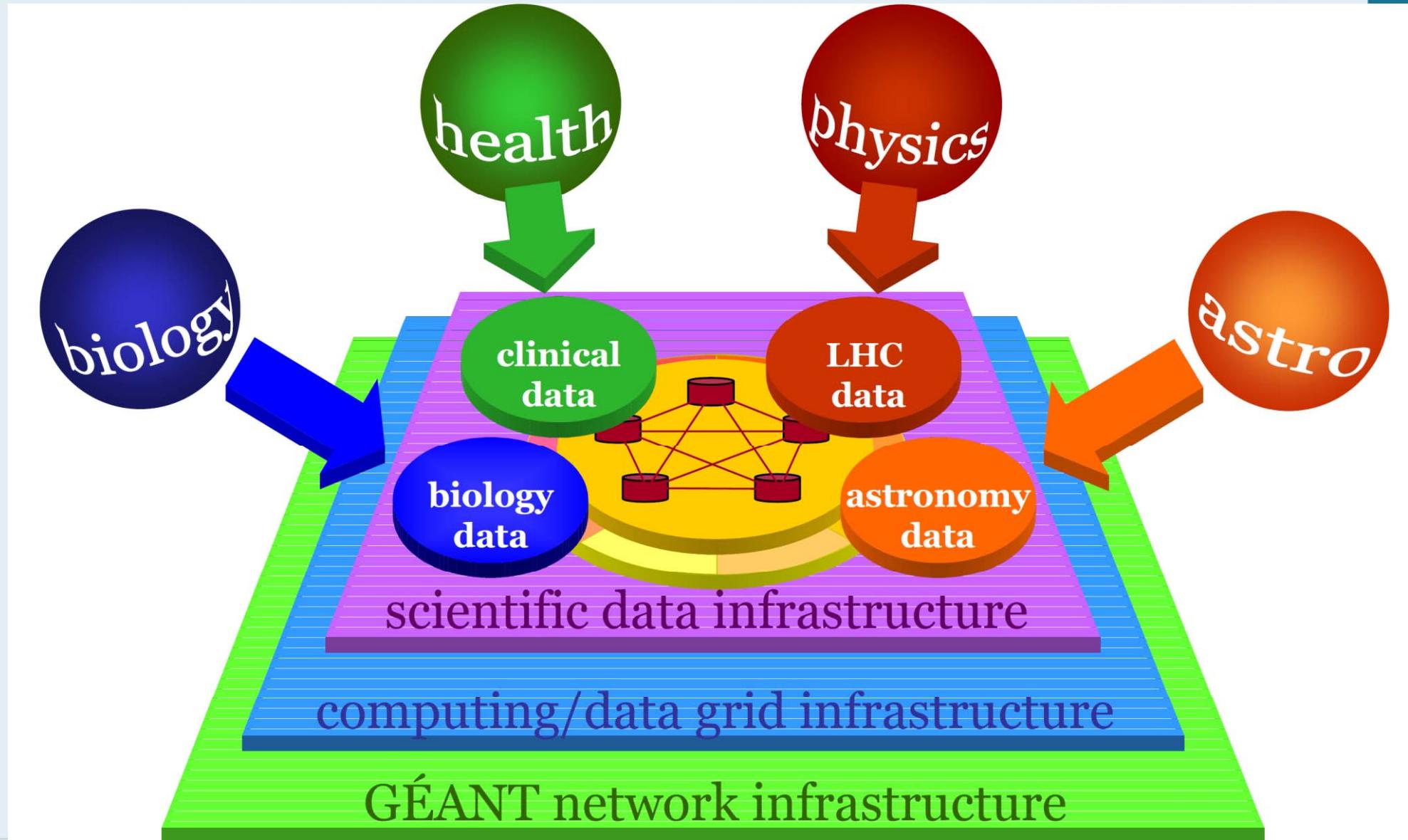
- ◆ Focus on Domain-Specific Federated Infrastructures
- Basic infrastructure
- Basic Services
- Basic Tools and Applications
 - To speed up research in strategic scientific domains
- Capacity to construct domain-specific by several national and international initiatives
 - Unique opportunity to bring together several different science communities
- ◆ Utilisation of existing Infrastructure

*Domain-oriented services and resources
of Polish Infrastructure
for Supporting Computational Science
in the European Research Space*



Fits to European e-Infrastructure Plans

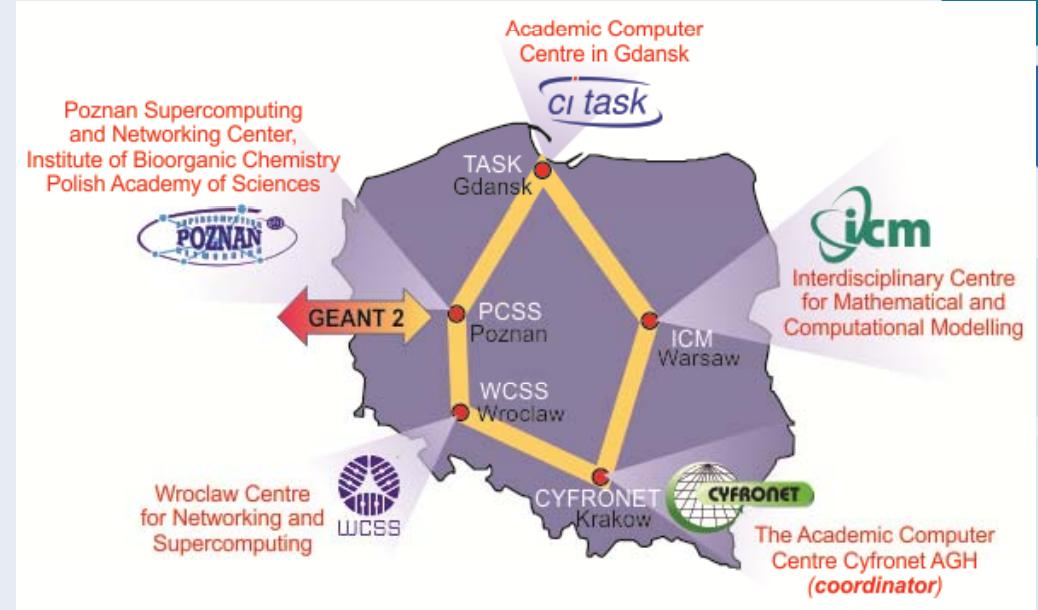
(thanks to Mario Campolargo)



PLGrid PLUS



- PLGrid PLUS Project funded by the European Regional Development Fund as a part of the Innovative Economy Program
- Duration: 1.10.2011 – 30.9.2014
- Budget: total ca. 18 M€
- Five Consortium Partners
- Project Leader:
Academic Computer Centre
CYFRONET AGH



Project Aims



- Design and deployment of “domain grids” – solutions for scientific-domain related services, tools and software packages for 13 identified scientific domains
- Design and start-up of support for possible new domain grids together with trainings
- Deployment of new infrastructure services
- Enabling System-level Research
- Deployment of Quality of Service system for users by introducing SLA agreement
- Implementation of Service Level Management procedures
- Expansion of the existing infrastructure resources and supporting infrastructure
- Deployment of Cloud infrastructure for users

Domain Grids



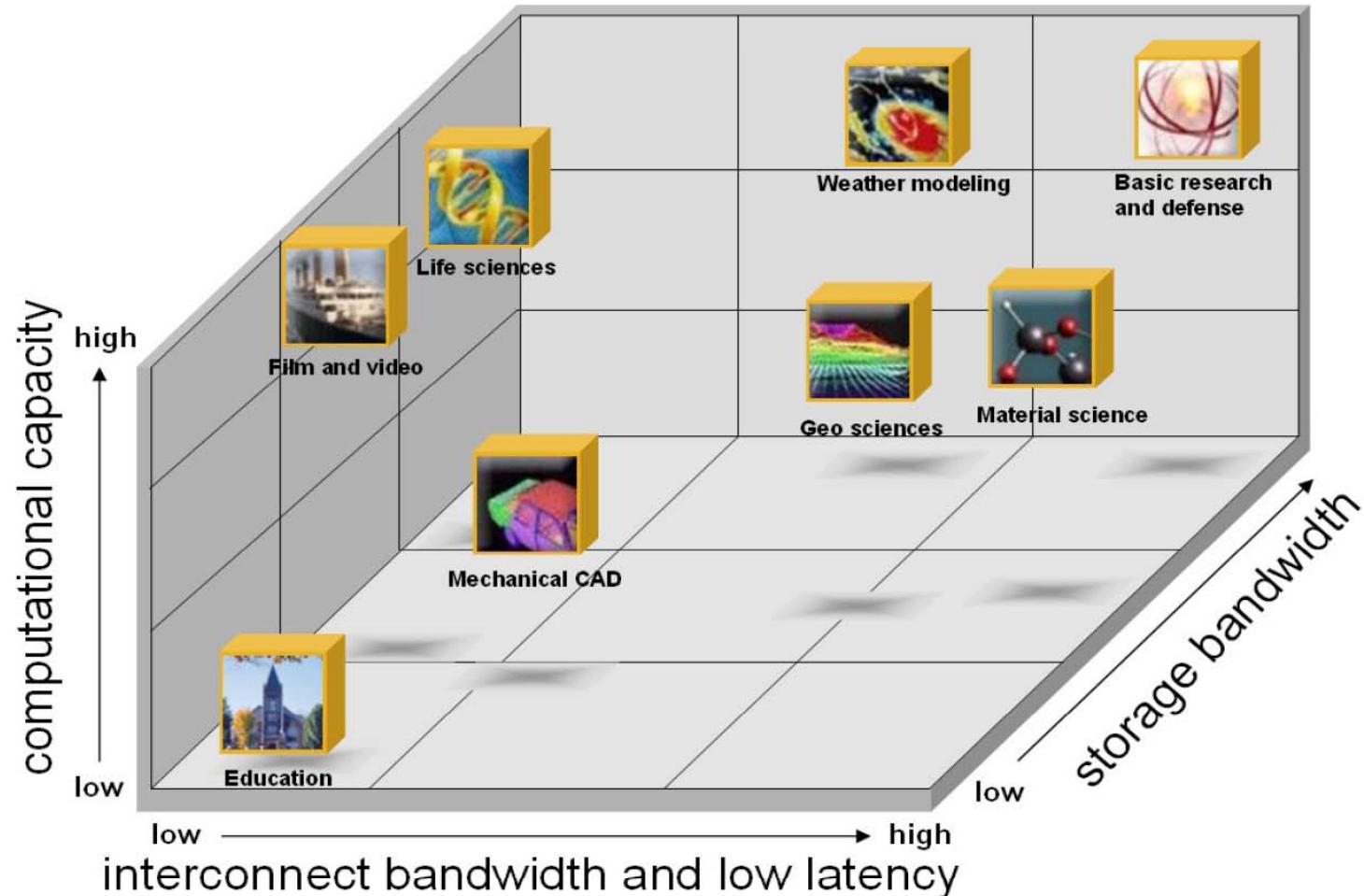
- Pilot program for strategic science domains and important topics of Polish/European Science
- Already identified 13 communities/scientific topics:

- ◆ Astrophysics
- ◆ HEP
- ◆ Life Sciences
- ◆ Quantum Chemistry and Molecular Physics
- ◆ Synchrotron Radiation
- ◆ Power Systems
- ◆ Metallurgy
- ◆ Nanotechnology
- ◆ Acoustics
- ◆ Ecology
- ◆ Bioinformatics
- ◆ Health
- ◆ Material Science

Diversity of Requirements



Market Segmentation



ACK: HP

Activities in Domain Grids in general



- **Integration Services**
 - National and International levels
 - Dedicated Portals and Environments
 - Unification of distributed Databases
 - Virtual Laboratories
 - Remote Visualization
- **Computing Intensive Solutions**
 - Specific Computing Environments (platforms)
 - Adoption of suitable algorithms and solutions
 - Workflows
 - Cloud computing
 - Porting Scientific Packages
- **Data Intensive Computing**
 - Access to distributed Scientific Databases
 - Organization of Scientific Databases
 - Data discovery, process, visualization, validation....
 - 4th Paradigm of scientific research
- **Instruments in Grid**
 - Remote Transparent Access to instruments
 - Sensor networks
- **Organizational**
 - Organizational backbone
 - Professional support for specific disciplines and topics

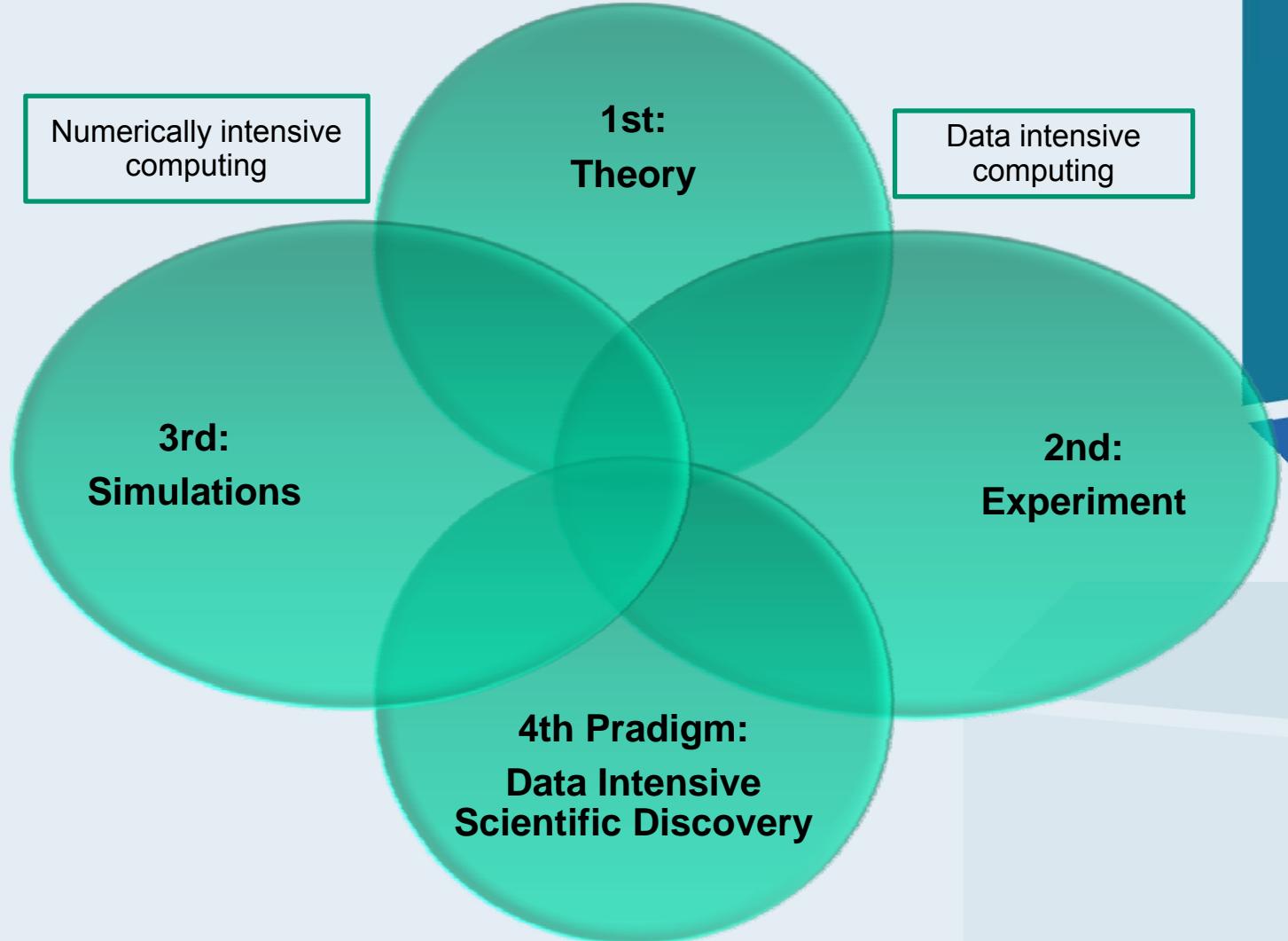
E-Science: Experiments in Silico

Research Paradigms



Required Synergy
between

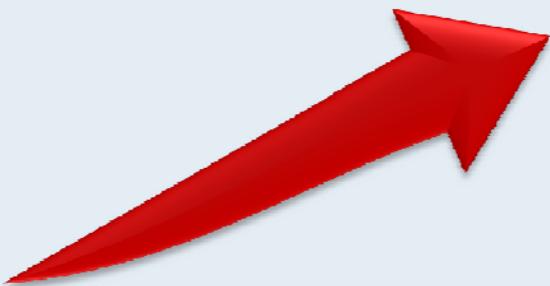
- Theory
- Simulation
- Experiment
- Data analysis



Existing and Planned Resources



- ◆ Current PL-Grid resources:
 - ◆ 260 TFLOPs of CPU
 - ◆ 3.3 PB of storage
- ◆ Planned resource extension for PLGrid PLUS
 - ◆ ca. 500 TFLOPs of CPU
 - ◆ ca. 4.4 PB of storage
- ◆ Accompanying equipment



Extension of Computing Environment



- Keeping diversity
 - Clusters (thin and thick nodes)
 - Clusters with GPGPU
 - SMP machines
 - vSMP

New Services in PLGrid PLUS

(as defined in the Proposal)



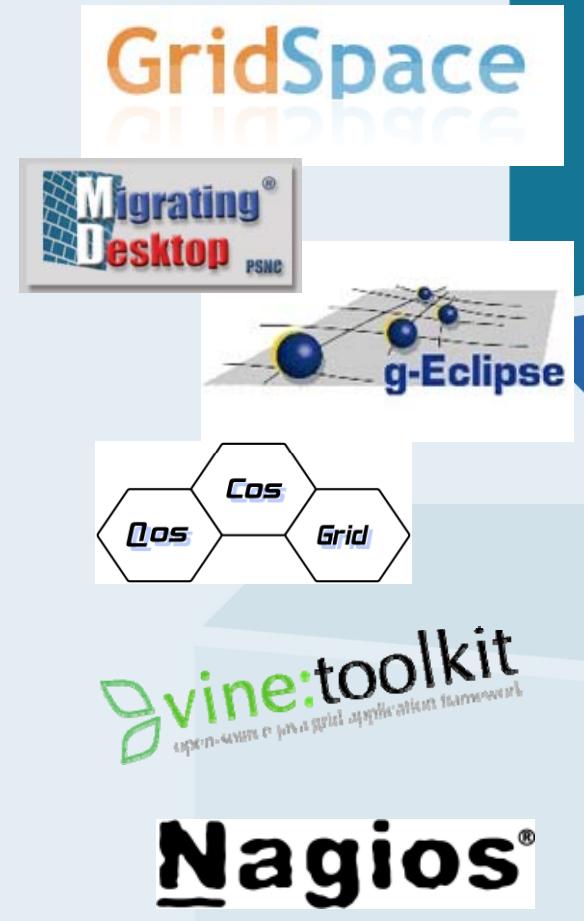
- ◆ Cloud Computing for Polish Science – new computing paradigm foreseen as a natural extension of the current Infrastructure offer
- ◆ Platform for supporting e-Science, resulting from the need for an international cooperation between various disciplines of scientific domains
- ◆ Production infrastructure oriented towards domain specific services, tools, environments and software packages
- ◆ Professional support for specific disciplines and topics important for Polish e-Science
- ◆ Visualisation of the scientific results via shared infrastructure servers equipped with possibility of binding domain specific visualisation tools

Innovative Infrastructure Environment

PL-Grid extensions



- ◆ Efficient Resource Allocation
 - ◆ **Grid Resource Bazaar**, mobile access to the infrastructure, new security modules and other tools for users and systems administrators: -- management of users request
- ◆ Experimental Workbenches
 - ◆ **GridSpace2** platform extension for supporting for new domains and integration with new grid/cloud services
 - ◆ **InSilicoLab** – integrated environment for chemists and biologists
- ◆ Tools and Middleware
 - ◆ **Migrating Desktop**, **VineToolkit** and **gEclipse** tools integration with various PL-Grid domain services
 - ◆ **QStorMan Toolkit** – extension for domain requirements on optimization of data access
 - ◆ **QosCosGrid** continuation of development
 - ◆ **Liferay Portal** framework(s) – adoption to specific needs
 - ◆ **HelpDesk** Portal for the users (specialized versions)



Availability of Scientific Software Packages

Continuation of porting to PL-Grid Environment



- Access to software packages is provided to users through:
 - gLite
 - UNICORE
 - QCG
- Examples of available packages in various fields:
 - **biology:** AutoDock, BLAST, ClustalW2, CPMD, Gromacs, NAMD
 - **quantum chemistry:** ADF, CFOUR, Dalton, GAMESS, Gaussian, Molcas, Molpro, MOPAC, NWChem, OpenBabel, Siesta, TURBOMOLE
 - **physics:** ANSYS FLUENT, Meep
 - **numerical computations and simulation:** Mathematica, MATLAB, OpenFOAM
 - **other:** Blender, POV-Ray



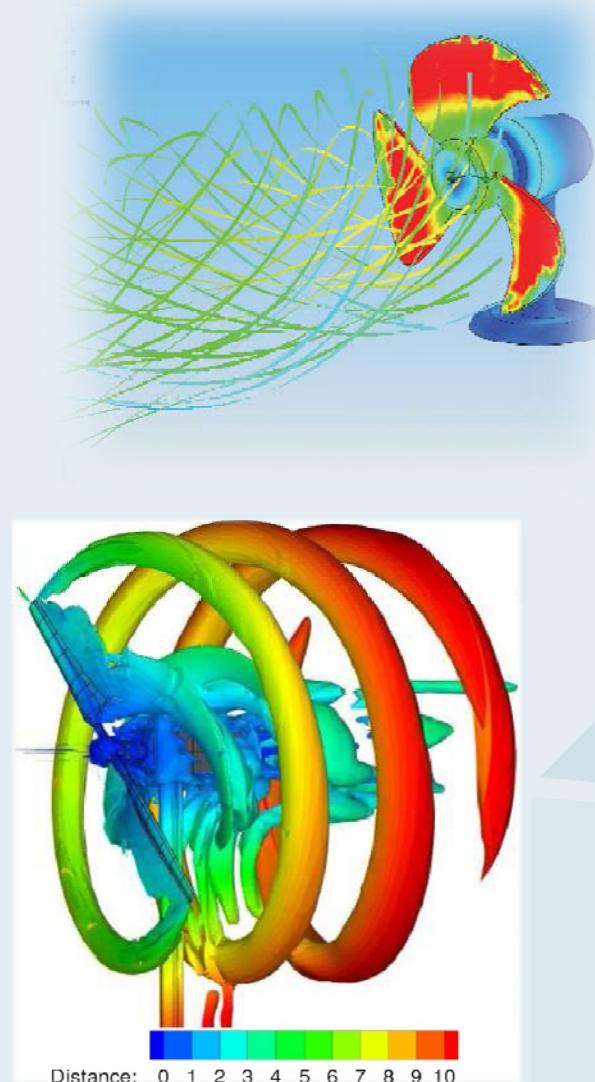
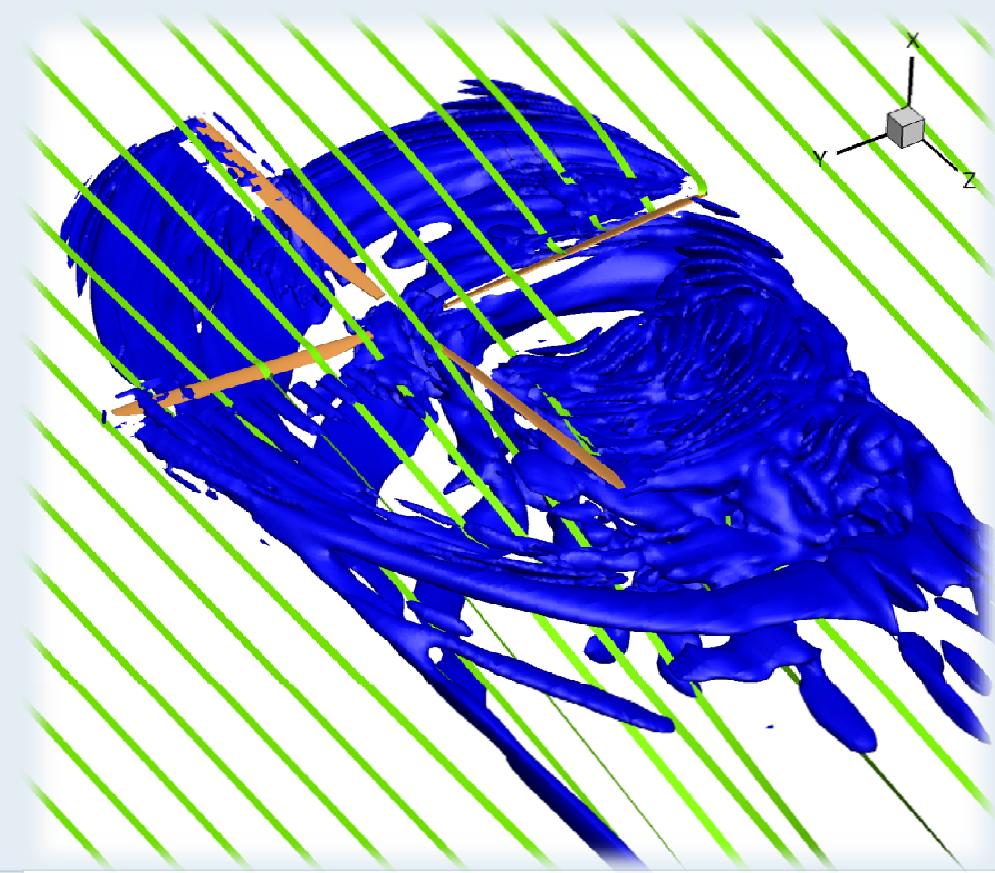
Dziedzinowo zorientowane
usługi i zasoby infrastruktury
PL-Grid dla wspomagania
Polskiej Nauki w Europejskiej
Przestrzeni Badawczej

Some Examples

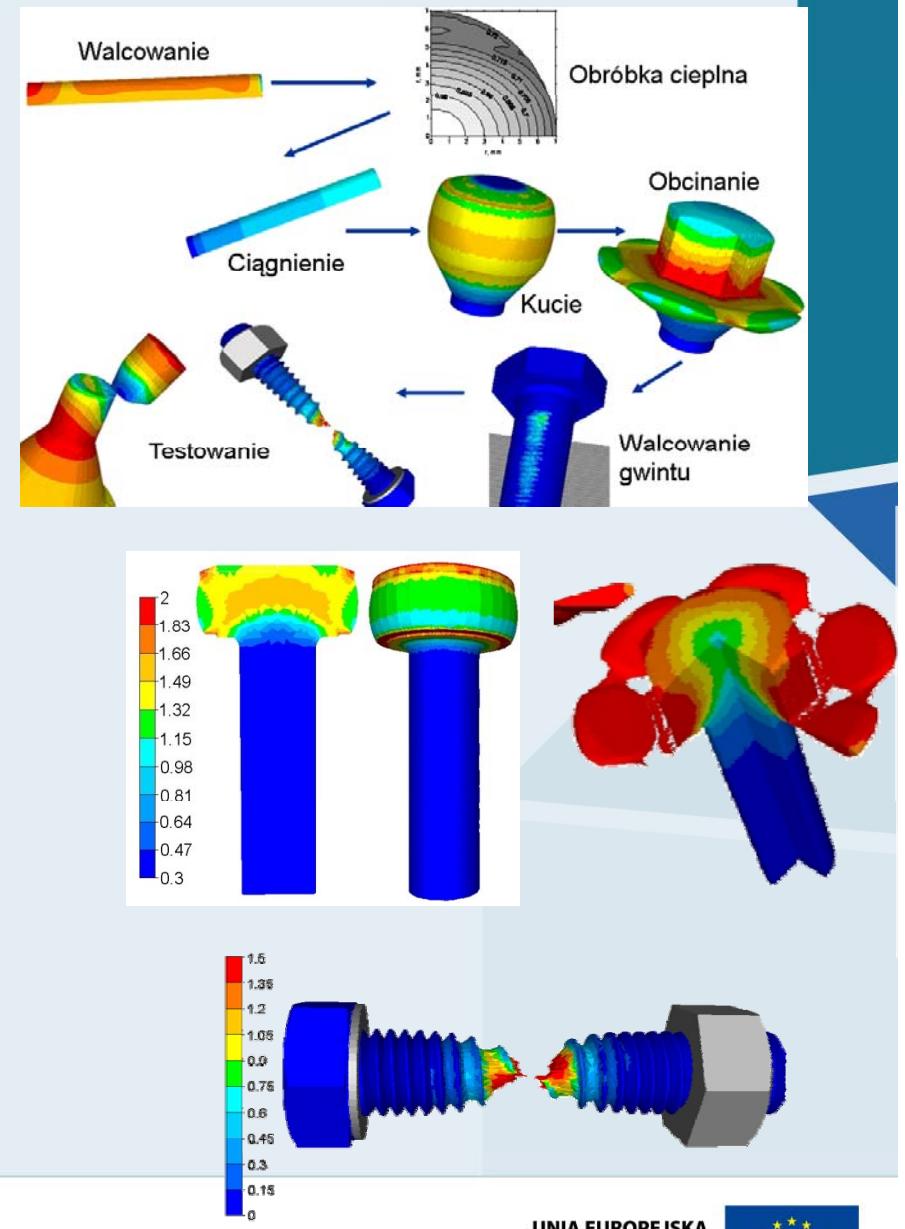
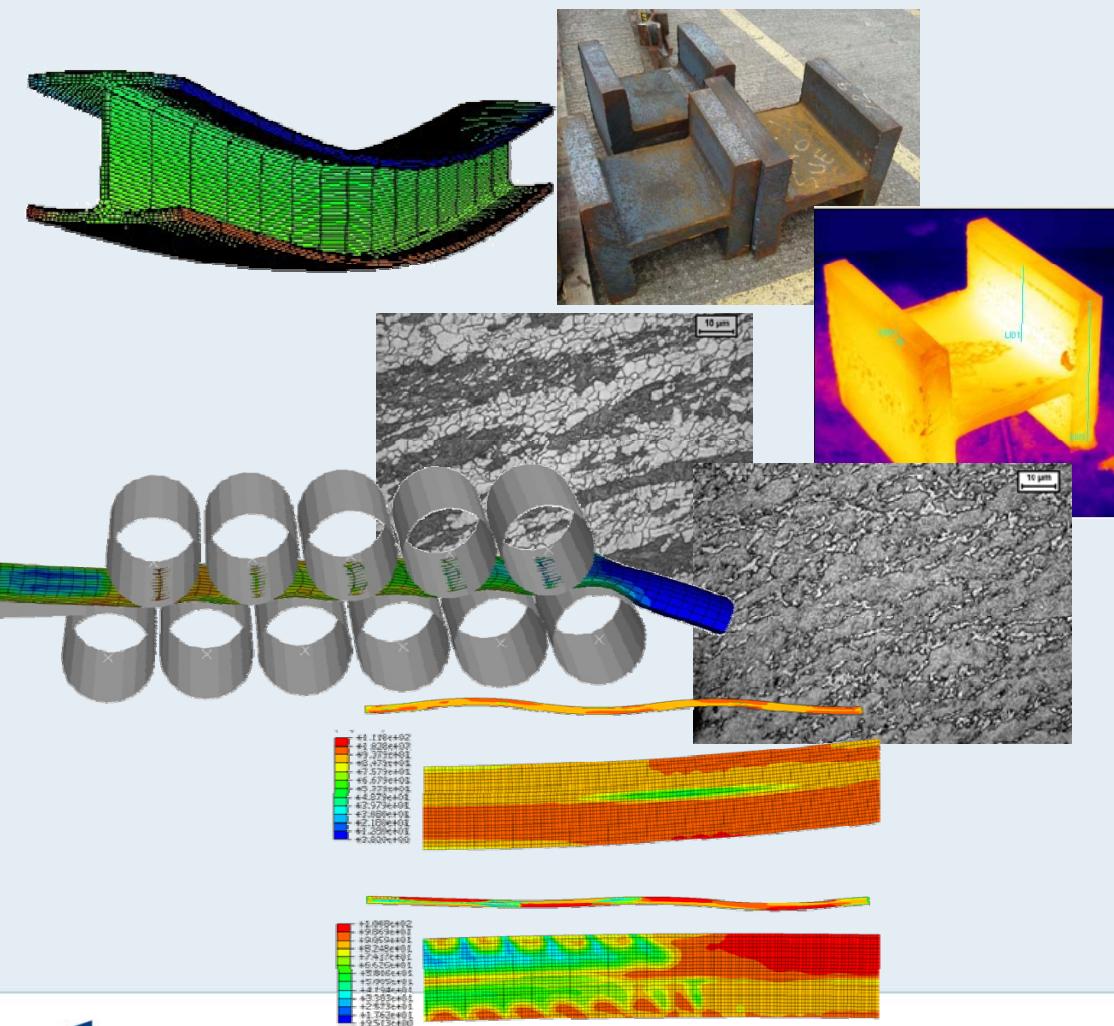
Acoustics



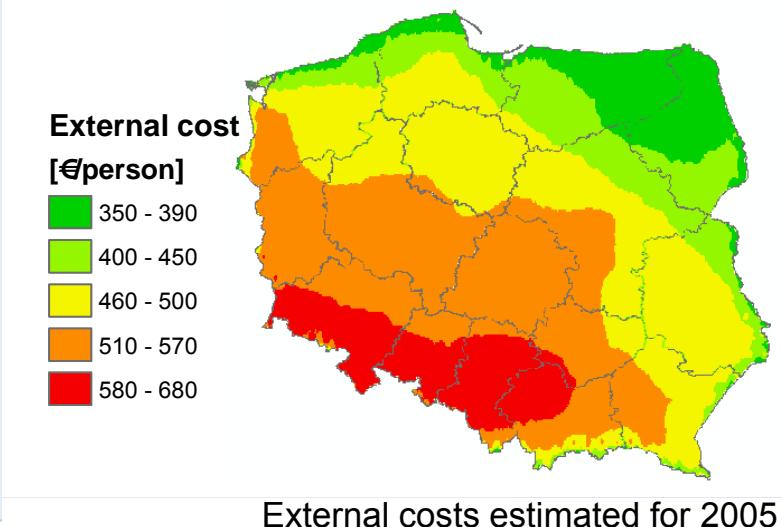
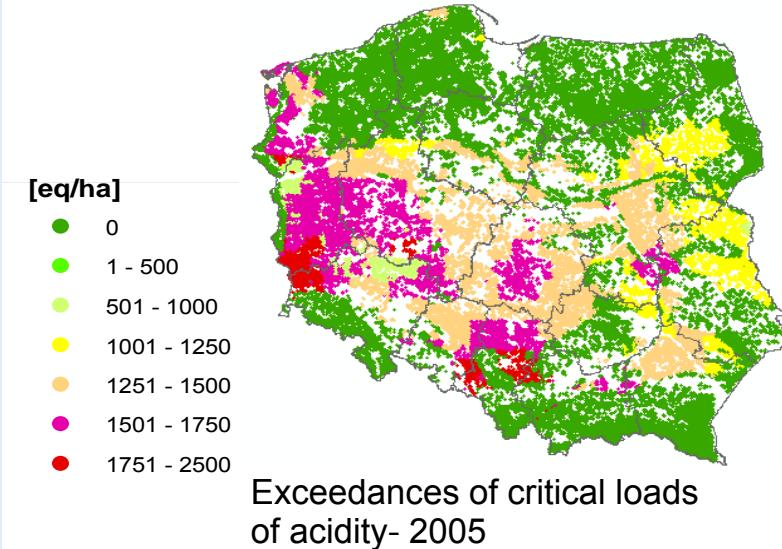
Turbines...choppers..noise



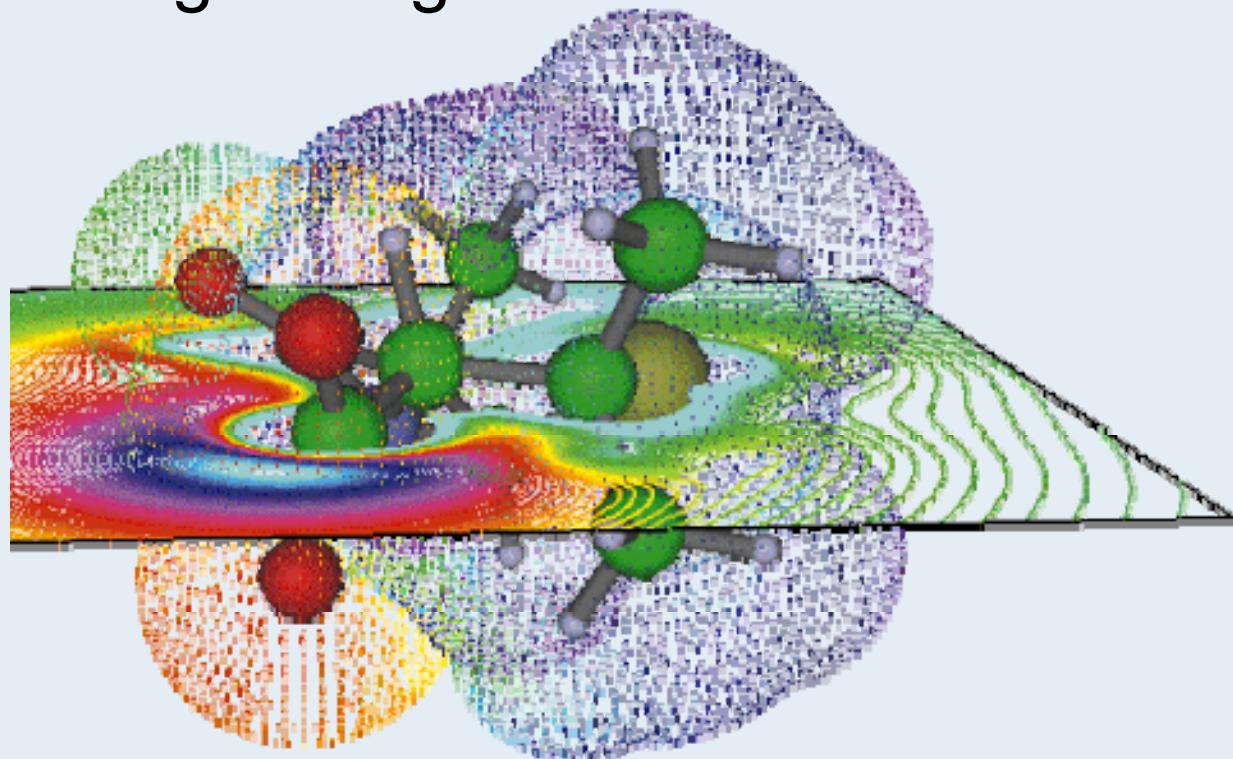
Modelling of different kind of processes



Model for Assessment of Environmental & Health Impacts

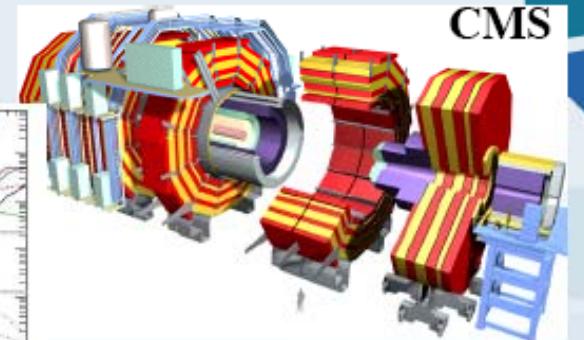
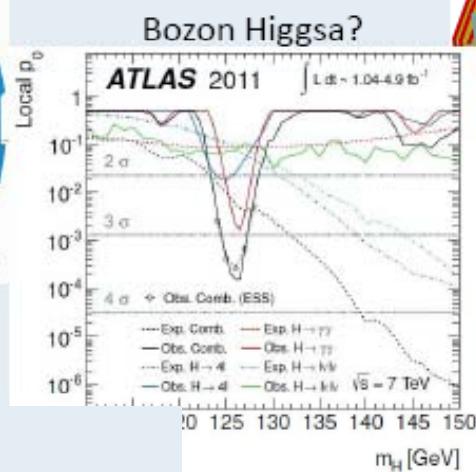
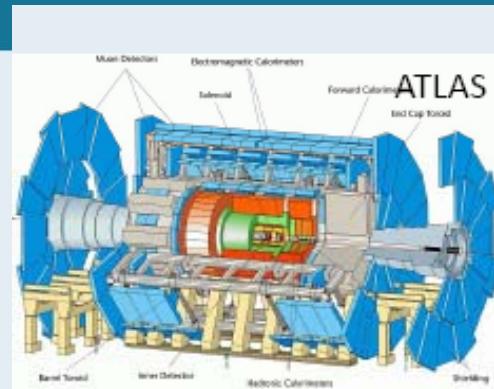


New original algorithms

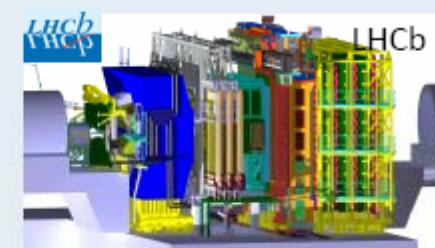
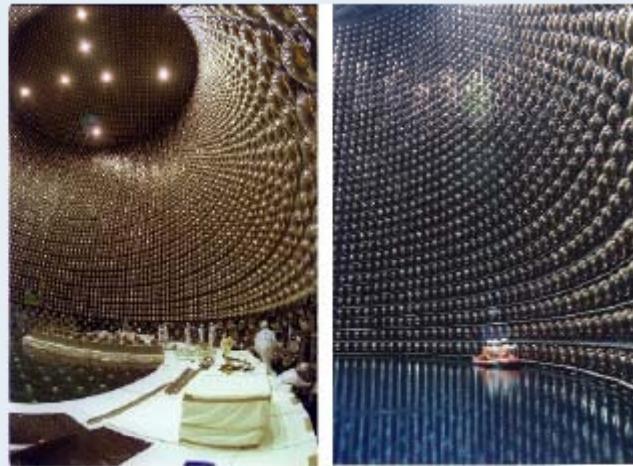
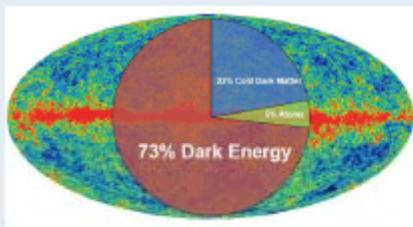


$$H \Psi = E \Psi$$

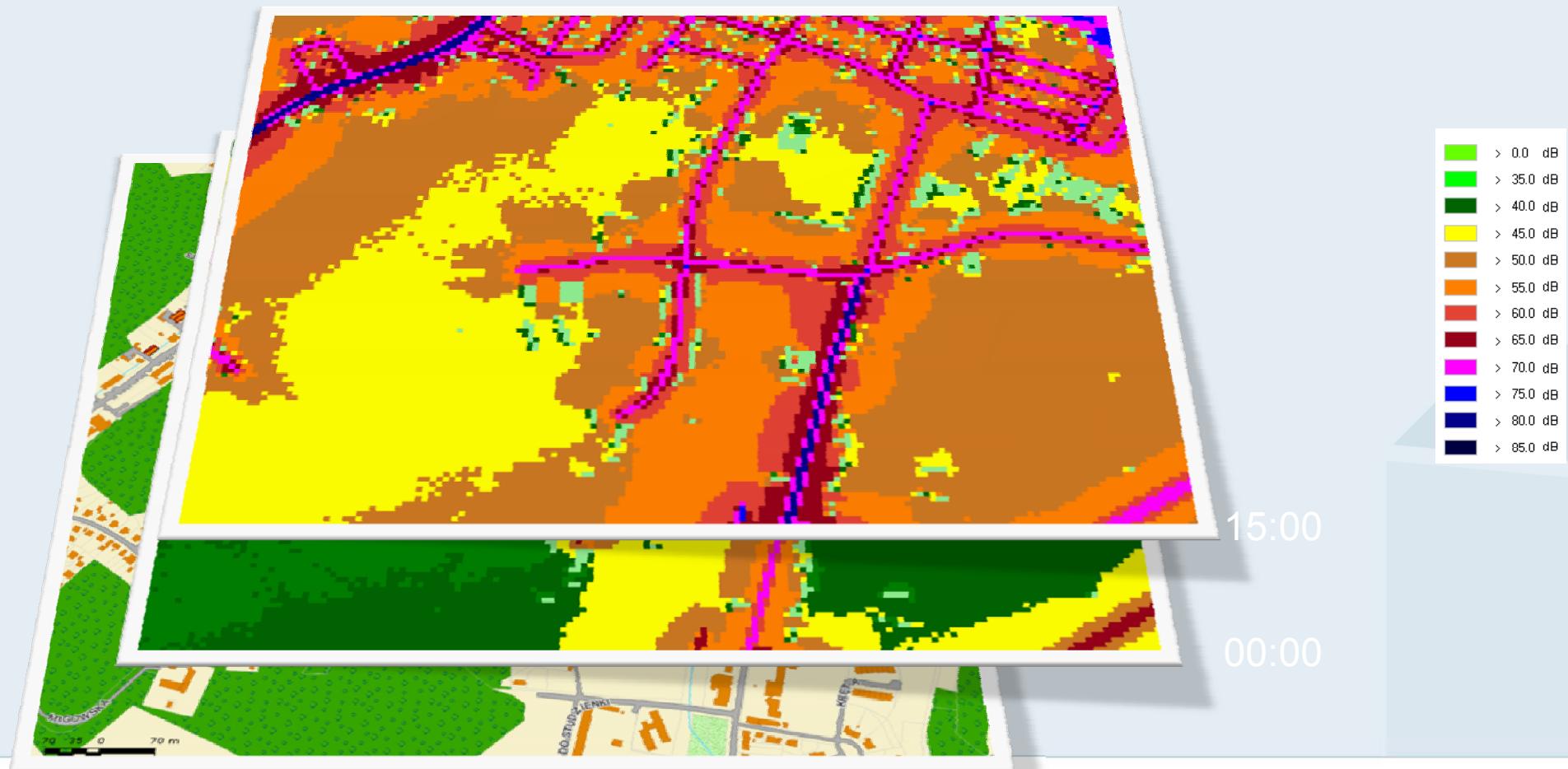
HEP



Eksperyment T2K



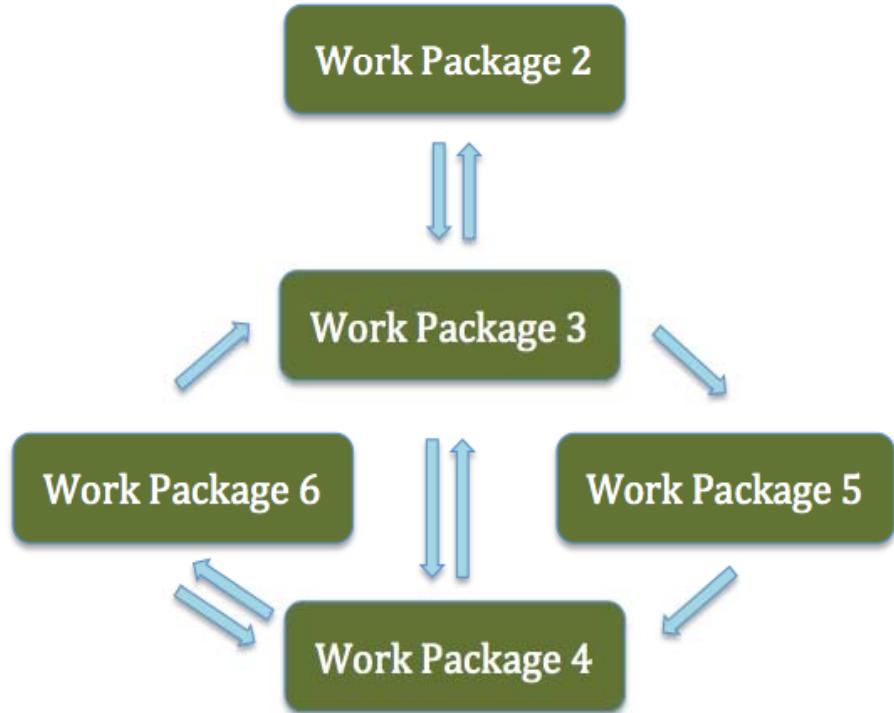
Creation of Country noise maps



Work Packages



- ◆ WP1 – Project Management (Cyfronet)
- ◆ WP2 – Development of Infrastructure (WCSS)
- ◆ WP3 – Operations Centre – Production Management (Cyfronet)
- ◆ WP4 – Pilot program for 13 scientific domain grids (TASK)
- ◆ WP5 – Support for users, training, broadening of existing scientific domain grids (ICM)
- ◆ WP6 – New infrastructure services (visualisation, cloud computing) (PCSS)



Conclusions



- Further development needed, as identified currently, mainly on Domain Specific Grids
- Request from the users' communities
- Capacity for organization of future development according to
 - Expertise and experience
 - Strong scientific potential of the users' communities being represented by PL-Grid Consortium
 - Wide international cooperation concerning the Consortium and individual Partners, good recognition worldwide
 - Good managerial capacity
- Please visit our Web page: <http://www.plgrid.pl/en>
- Credits

Credits



■ ACC Cyfronet AGH

- Kazimierz Wiatr
- Łukasz Dutka
- Michał Turała
- Marian Bubak
- Krzysztof Zieliński
- Karol Krawentek
- Agnieszka Szymańska
- Teresa Ozga
- Andrzej Oziębło
- Maciej Malawski
- Tomasz Szepieniec
- Mariusz Sterzel
- Zofia Mosurska
- Robert Pająk
- Marcin Radecki
- Renata Słota
- Tomasz Gubała
- Darin Nikolow
- Aleksandra Mazur
- Patryk Lasoń
- Marek Magryś
- Łukasz Flis

◆ ICM

- ◆ Marek Niezgódka
- ◆ Piotr Bała
- ◆ Maciej Filocha

◆ PCSS

- ◆ Maciej Stroiński
- ◆ Norbert Meyer
- ◆ Bartek Palak
- ◆ Krzysztof Kurowski
- ◆ Tomasz Piontek
- ◆ Dawid Szejnfeld
- ◆ Paweł Wolniewicz

◆ WCSS

- ◆ Jerzy Janyszek
- ◆ Bartłomiej Balcerek
- ◆ Paweł Tykierko
- ◆ Paweł Dziekoński

◆ TASK

- ◆ Rafał Tylman
- ◆ Młodzież Nakonieczny
- ◆ Jarosław Rybicki

... and many others....

