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

PLGrid PLUS Project
Status and Current Achievements

Jacek Kitowski, Łukasz Dutka, Tomasz Szepieniec, Mariusz Sterzel and Robert Pająk

ACK Cyfronet AGH, Krakow, Poland
PL-Grid Consortium

CGW'12, Krakow, Poland, October 22-24, 2012

Outline

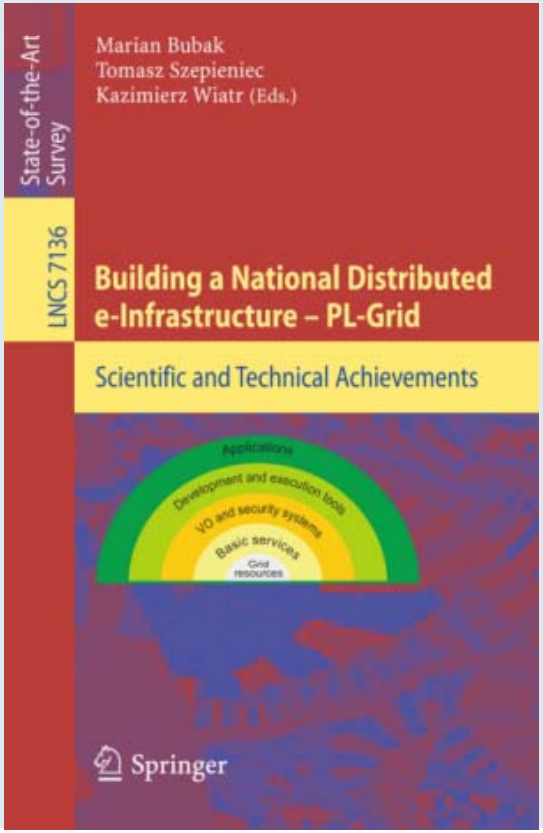
- Consortium and PL-Grid Project
- PLGrid Plus – basic facts
- Domain-specific solutions and services
- Conclusions

Consortium and the PL-Grid Project

Polish Infrastructure for Supporting Computational Science in the European Research Space – PL-Grid

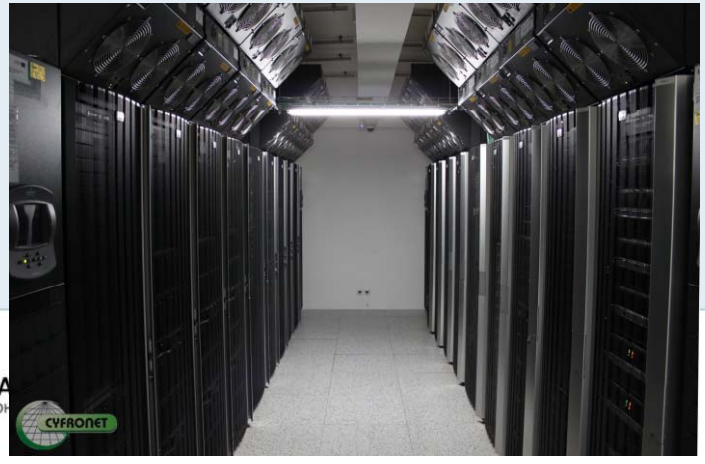
- Consortium created (Jan. 2007) with goal:
  - Significant extension of amount of computing resources provided to the Polish scientific community
- PL-Grid Project (1.1.2009-31.3.2012) by the European Regional Development Fund as part of the Innovative Economy Program with outcome:
  - Common base infrastructure
    – National Grid Infrastructure (NGI_PL)
    – internationally compatible
  - Potential capacity to construct specialized, domain Grid systems
  - Resources: 230 Tflops, 3600 TB
    users: 1000+
    jobs/month: 500,000-1,500,000
    - innovative grid services and end-user tools like Efficient Resource Allocation, Experimental Workbench and Grid Middleware
    - Scientific Software Packages
    - User support: helpdesk system, broad training offer

- Publication of the book (Springer 2012) with PL-Grid achievements
- Content: 26 papers



TOP500 Nov.2011 Polish Sites

Rank	Site	System	Cores	Rmax TFlop/s	Rpeak TFlop/s
88	**Cyfronet**	**Zeus** - Cluster Platform 3000 BL 2x220, Xeon X5650 6C 2.66 GHz, Infiniband, HP	15264	128.8	162.4
279	**Gdansk Cl Task**	**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
298	**PCSS**	**Rackable** C1103-G15, Opteron 6234 12C 2.40 GHz, Infiniband QDR, SGI	5640	63.9	136.4
348	Grupa Allegro	Cluster Platform 3000 BL 2x220, Xeon L5420 4C 2.50 GHz, Gigabit Ethernet, HP	10748	59.1	107.5
360	**WCSS**	**Supernova** - Cluster Platform 3000 BL2x220, X56xx 2.66 Ghz, Infiniband, HP	6348	57.4	67.5



Cyfronet at TOP500 lists

List	Rank	System	Cores	Rmax (GFlop/s)	Rpeak (GFlop/s)
06/2012	89	**Zeus** - Cluster Platform SL390/BL2x220, Xeon X5650 6C 2.660GHz, Infiniband QDR, NVIDIA 2050/2090, HP	13944	185316	271113
11/2011	88	**Zeus** - Cluster Platform 3000 BL 2x220, Xeon X5650 6C 2.66 GHz, Infiniband, HP	15264	128790	162409.0
06/2011	81	**Zeus** - Cluster Platform 3000 BL2x220, L56xx 2.26 Ghz, Infiniband, HP	11694	104765.1	124424.2
11/2010	85	**Zeus** - Cluster Platform 3000 BL2x220, L56xx 2.26 Ghz, Infiniband, HP	9840	88050.7	104697.6
06/2010	161	Cluster Platform 3000 BL2x220, L56xx 2.26 Ghz, Infiniband, HP	6144	39934.5	55541.8
11/2008	311	**Zeus** - Cluster Platform 3000 BL2x220, L54xx 2.5 Ghz, Infiniband, HP	2048	16179	20480
11/1996	408	SPP1600/XA-32, HP (Convex)	32	5.5	7.7
06/1996	408	SPP1200/XA-32, HP (Convex)	32	4.0	7.7

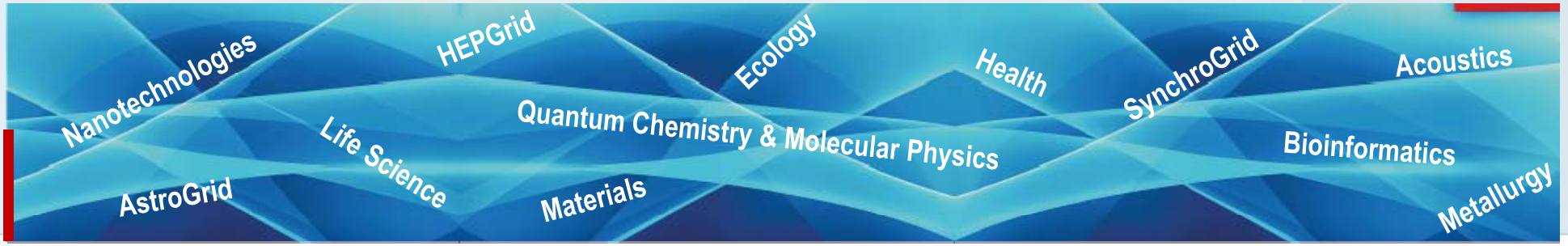
PLGrid Plus:

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

- Budget: total ca. 18 M€ including funding from the EC: ca.15 M€
- Duration: 1.10.2011 – 30.9.2014
- Five PL-Grid Consortium Partners; Coordinator: Academic Computer Centre CYFRONET AGH

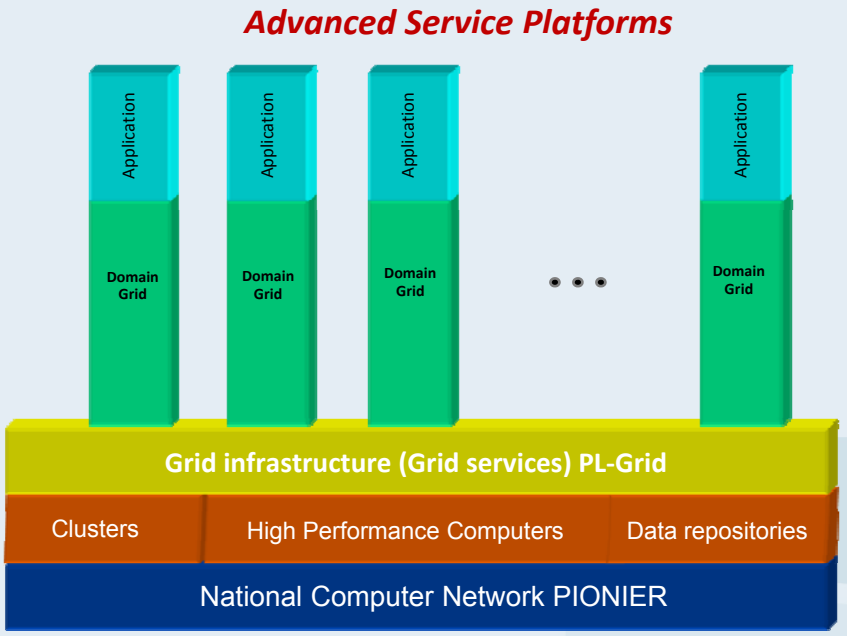
The main aim of the PLGrid Plus project is to increase potential of the Polish Science by providing the necessary IT services for research teams in Poland, in line with European solutions.

- Preparation of specific computing environments, i.e. solutions, services and extended infrastructure (including software), tailored to the needs of different groups of scientists.
- These domain-specific solutions will be created for already identified 13 groups of users, representing the strategic areas and important topics for the Polish and international science:



Project tasks:

- Design and start-up of support for new domain-specific federated grids
- Development of new infrastructure services tools, environments and resources
- Extension of the resources available in the PL-Grid infrastructure by
  - 500 Tflops
  - 4.4 PB
- Keeping diversity
  - Clusters (thin and thick nodes)
  - Clusters with GPGPU
  - SMP
  - vSMP
- Deployment of Cloud infrastructure for users
- Deployment of Quality of Service system for usersby introducing SLA agreement
- Broad consultancy, training and dissemination offer

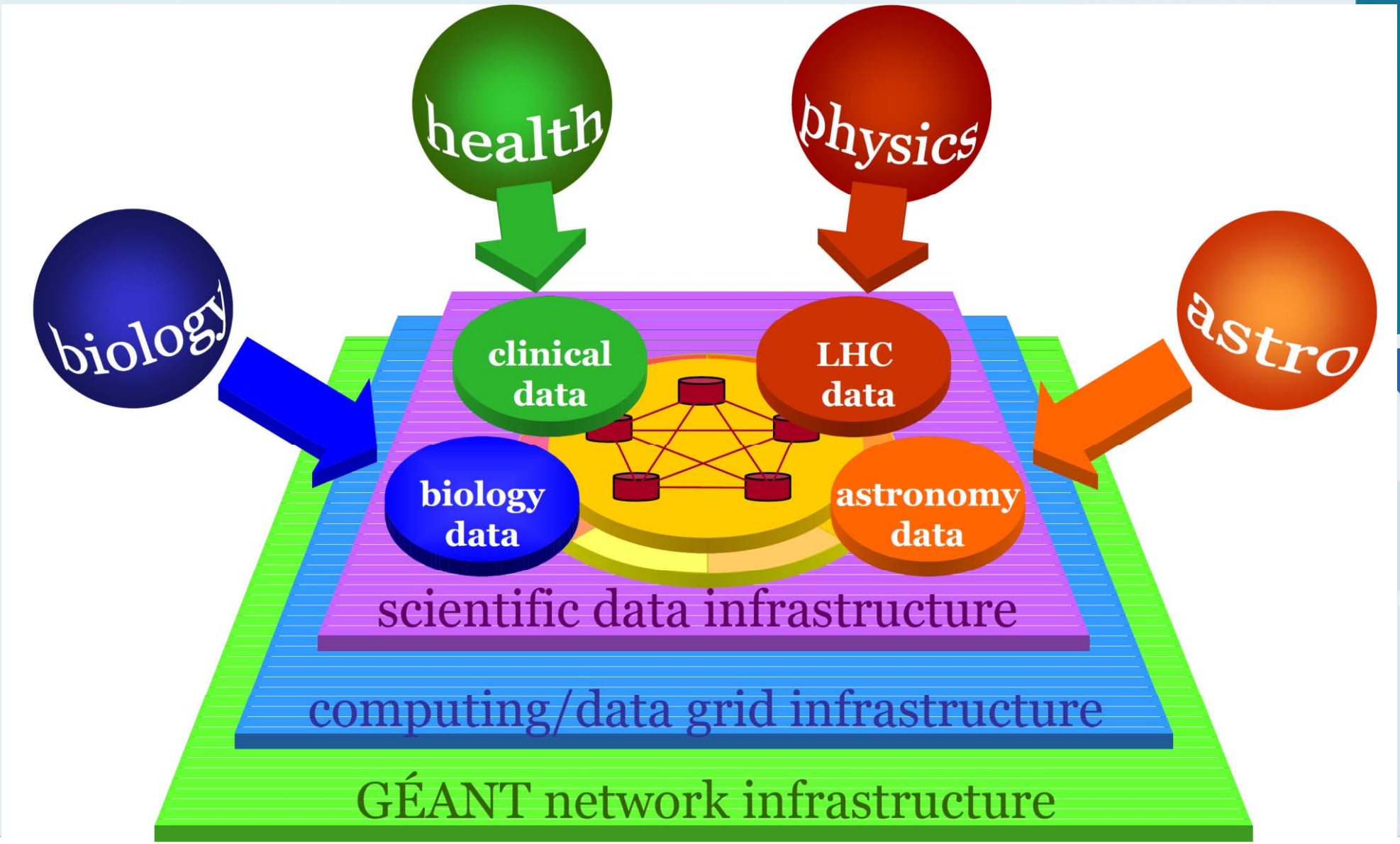


The scope is not limited to the selected domains. By using the developed general services and experience in building the domain ones, the integration of new groups will proceed smoothly and at lower cost.

PLGrid Plus:
Activities in Domain Grids in general

- Integration Services
  - National and International levels
  - Dedicated Portals and Environments
  - Unification of distributed Databases
  - Virtual Laboratiories
  - Remote Visualization
  - Service value = utility + warranty
  - SLA management
- Computing Intensive Solutions
  - Specific Computing Environments
  - 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

Fits to European e-Infrastructure Plans
(thanks to Mario Campolargo)



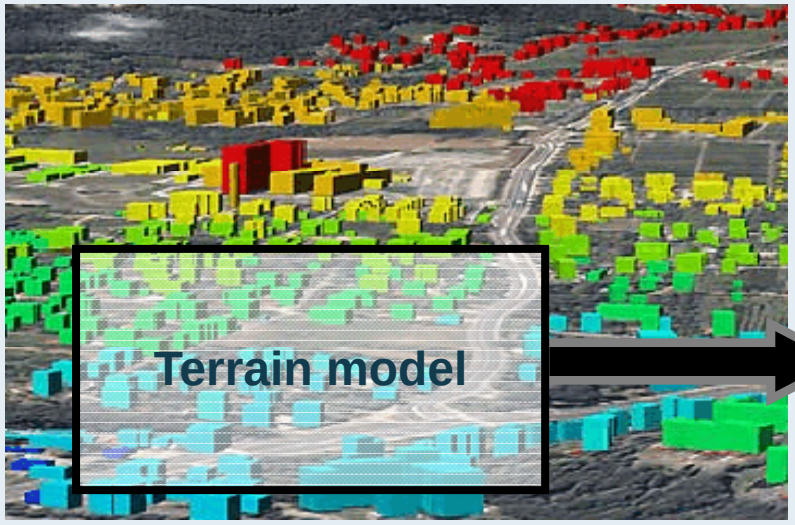
Domain Grids

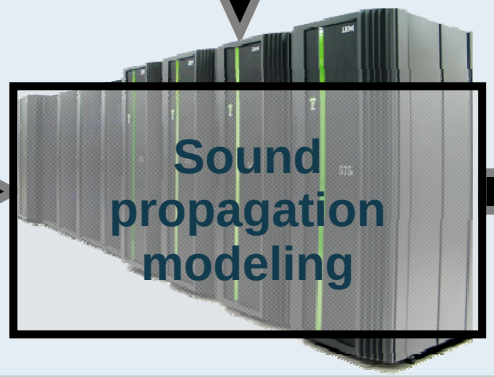
- Pilot program for strategic science domains and important topics of Polish/European Science
- Access to the software packages is provided by:
  - Glite
  - Unicore
  - QCG
- Already identified 13 communities/scientific topics:
  - Astrophysics
  - HEP
  - Life Sciences
  - Quantum Chemistry and Molecular Physics
  - Synchrotron Radiation
  - Power Systems
  - Metallurgy
  - Nanotechnology
  - Acoustics
  - Ecology
  - Bioinformatics
  - Health
  - Material Science

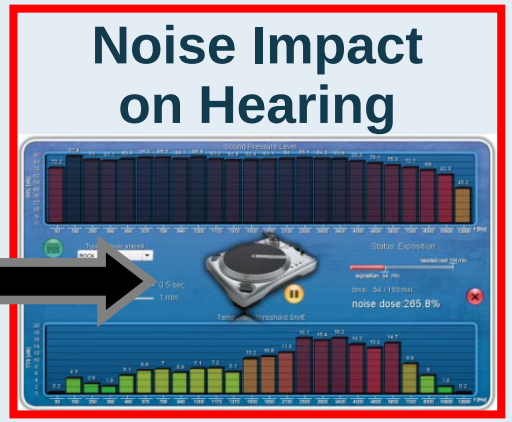
Examples of domain specific solutions and services

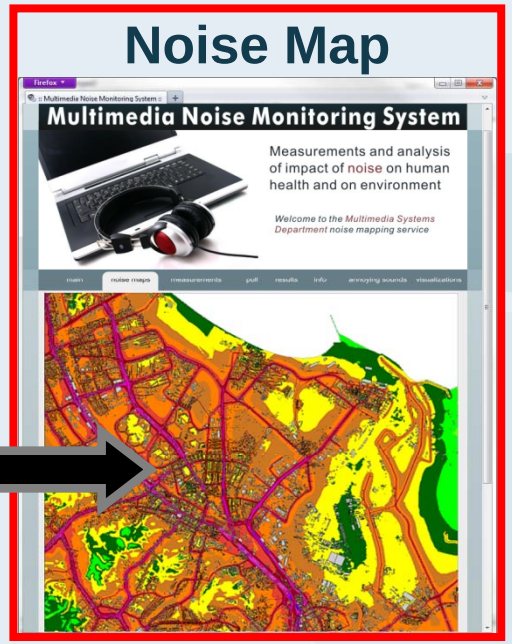
Acoustics

Multimedia Noise Monitoring System and Simulation

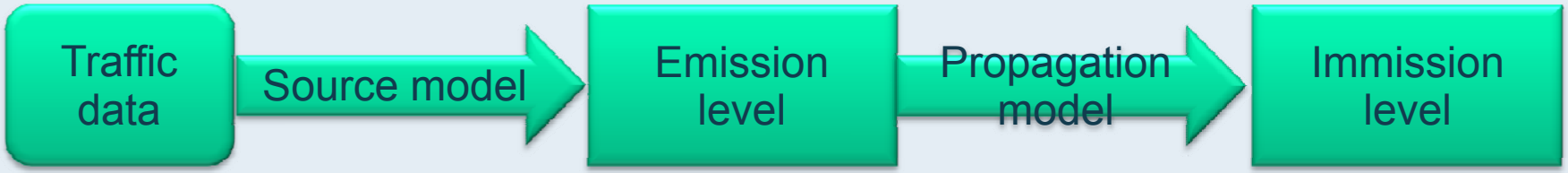








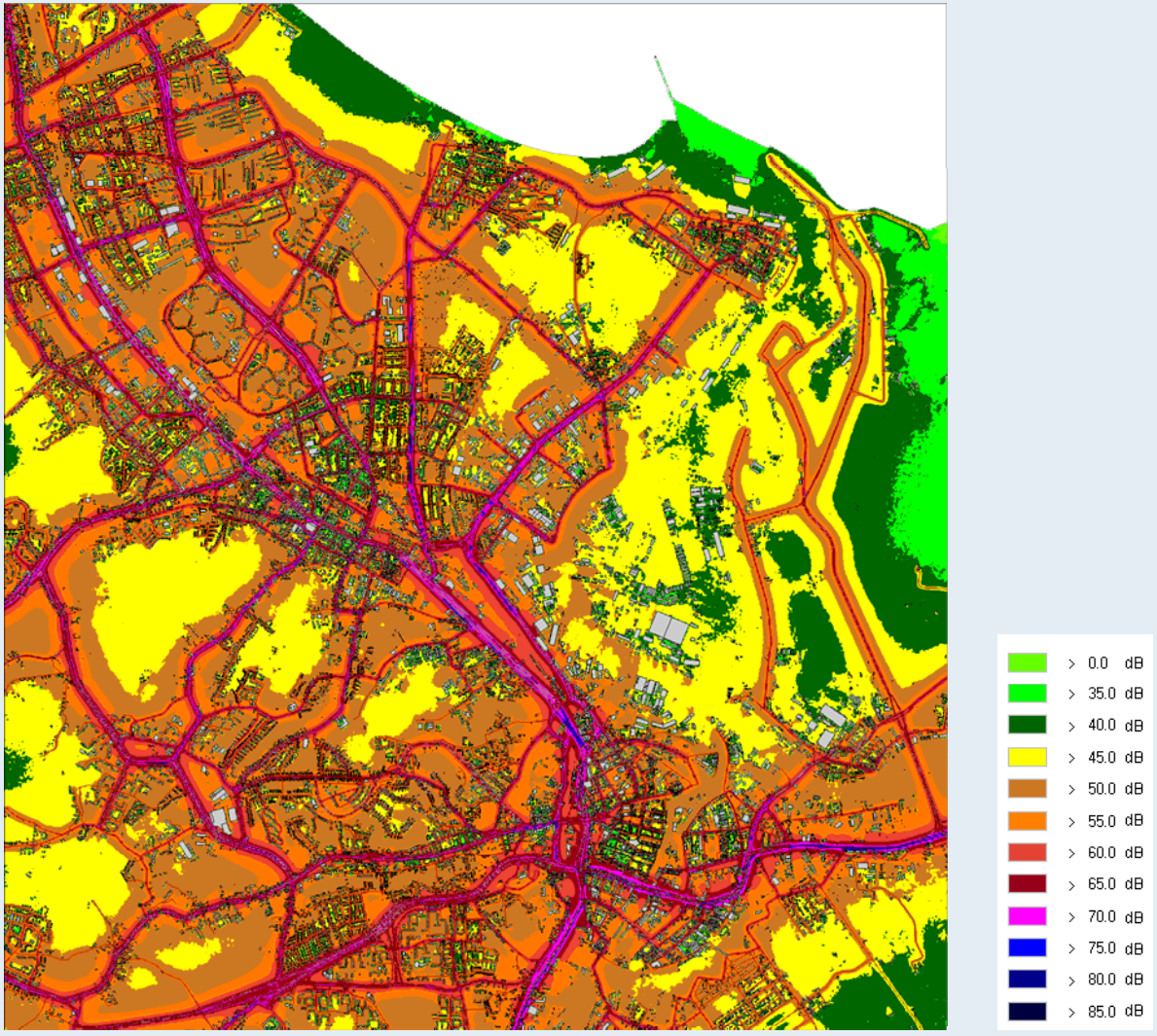
Acoustics
Methodology



- Source part, the Harmonoise model
  - Traffic volume
  - Vehicle speed
- Propagation part
  - The acoustic ray tracing method
  - Additional libraries: Harmonoise, CGAL, Tardem
  - Geometrical description of sources and buildings

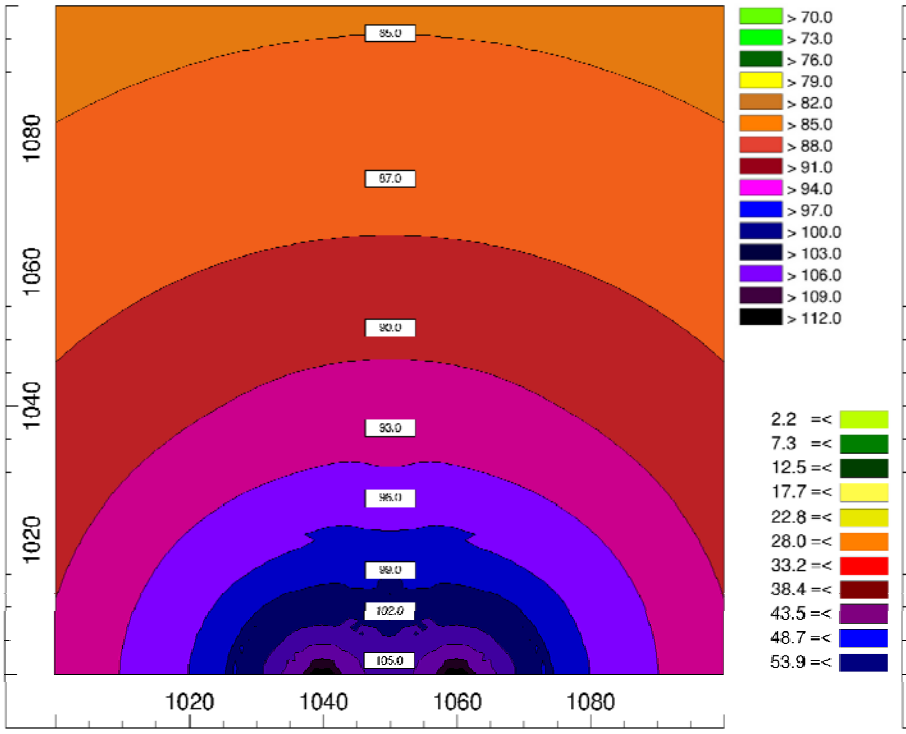
Acoustics

Computed noise map of large urban area (road noise)

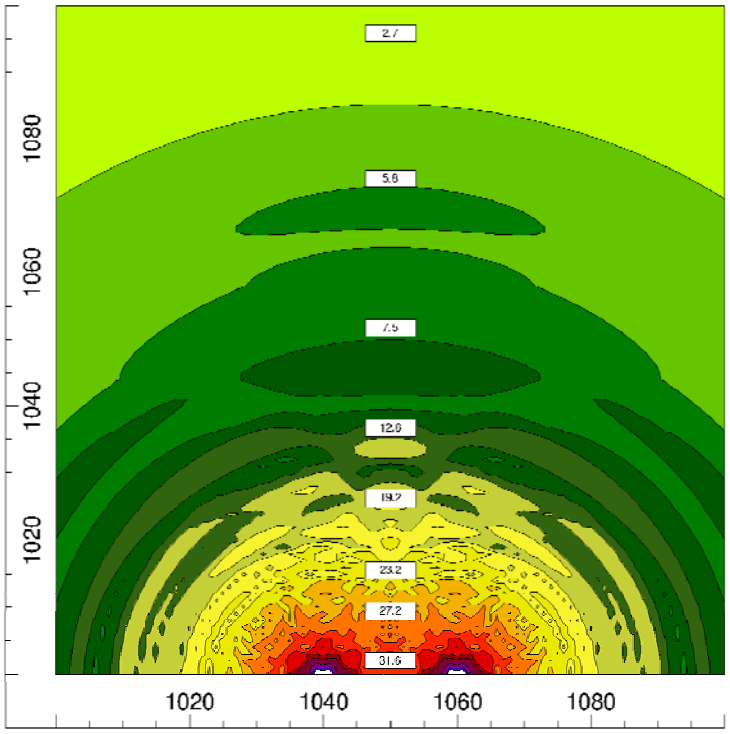


Acoustics

Simulation of noise exposure during outdoors concert



Noise map for the outdoor loud acoustic event (open field musical concert)



The map of maximum TTS values

Nanotechnology

The AuxEx Service: goals and features

Goals:

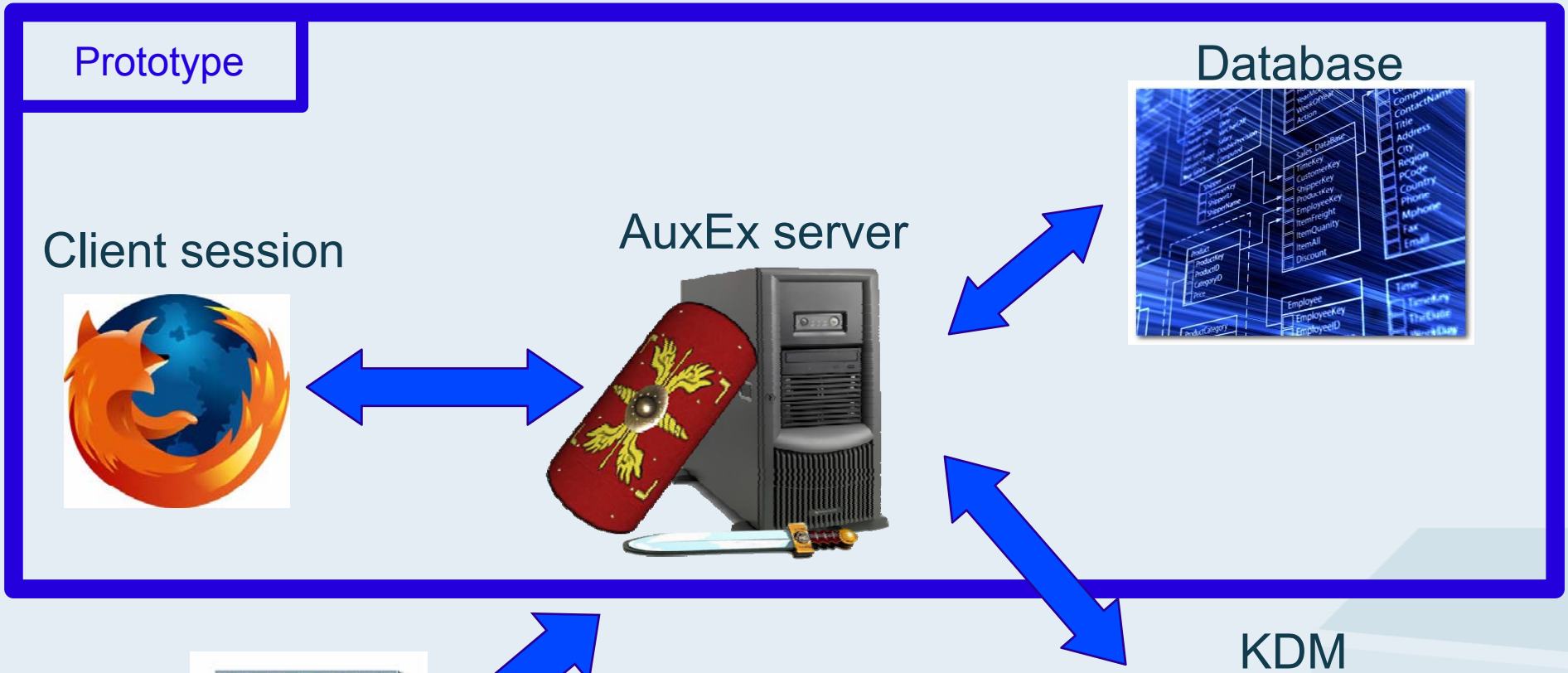
- creation of central point of collection and distribution of experimental data
- reduction of the likelihood of duplication of work on getting the same results
- improvement of data security (access rules, backup)
- automated data processing
- providing a consistent interface to the advanced computing application

Features:

- dedicated software (does not require the user to spend time to learn)
- requires no client installation Web computing (web interface)
- application tailored to the needs of a particular research project (for now NewLoks and Organometallics)
- the ability to install the server in the PLGrid Plus resource and locally on the user's machine

Downloaded from http://broeder10.wordpress.com 15.10.2012

Nanotechnology
Application structure



Synchrogrid
Overview

- Synchrogrid builds services for the synchrotron radiation scientific community
- It is expected that the establishment of the Research Centre of Polish Synchrotron (2014) will influence the rapid rise in the users of synchrotron radiation to 1,000 people
- Involved institutions:
  - Jerzy Haber Institute of Catalysis and Surface Chemistry Polish Academy of Sciences
  - AGH University of Science and Technology
  - Jagiellonian University
  - University of Silesia in Katowice
  - Adam Mickiewicz University
- Launching a synchrotron makes its radiation easily accessible
- The world witnessed the rapid increase of the number of scientists using synchrotron radiation when the first synchrotron was opened in their country

Synchrogrid
Services in development

- **Elegant Service**
  - Elegant („ELEctron Generation ANd Tracking”) is a fully 6D accelerator simulation program that now does much more than generation of particle distributions and tracking them
  - Matlab configured to use the Self Describing Data Sets (SDDS) file protocol
  - Additional scripts that ease submission of jobs locally on the cluster and with use of grid middleware
  - **Status:** prototype delivered to specific users for evaluation
- **Virtual Accelerator Service**
  - Requires Elegant service
  - TANGO (The TAco Next Generation Objects) control system – open source object-oriented control system for controlling accelerators, experiments and any kind of hardware or software being actively developed by a consortium of (mainly) synchrotron radiation institutes
  - Virtual Machine with User Interface for submitting grid jobs
  - **Status:** prototype development (installation on cluster and preparation of modules)

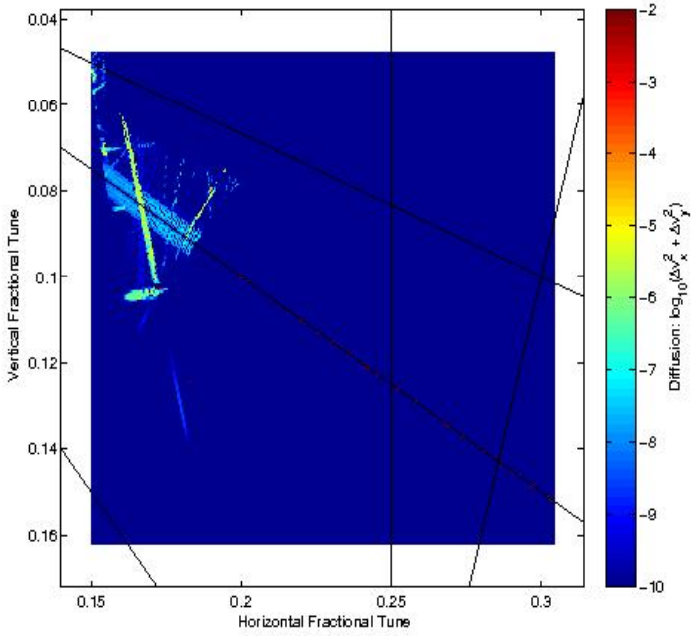
Synchrogrid
Usage of Elegant service prototype

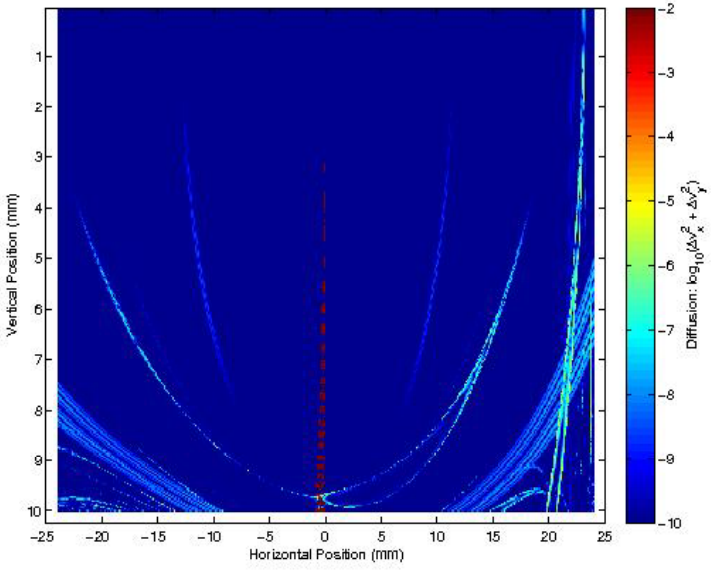
Preparation of input files

- Submission of jobs (local or grid) of parallel version of Elegant software
- Intermediate output file is automatically analysed by pre-prepaired Matlab scripts
- Final output (as figures on the right) is analysed by a user. Then, corrections to the input files are entered manually and the iteration is repeated.
- If the results are satisfactory for the user, the figures (prepared in eps and jpeg format) could be inserted into publication

Calculations in basic version (calculations in two dimensions only) required ca. 100h of walltime on iteration.
The introduction of scan in additional dimension is foreseen in the next release.

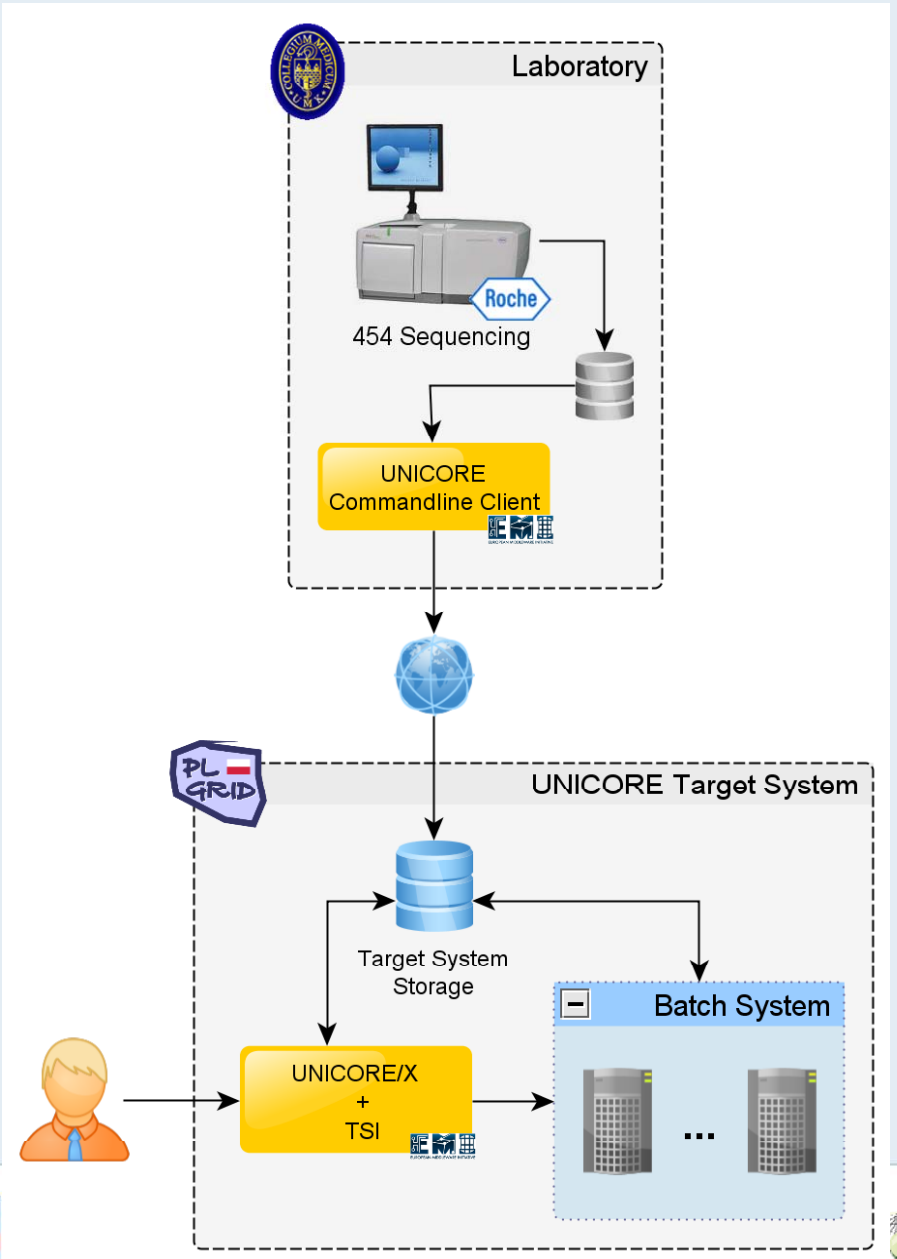
Storage ring requires a sufficiently large dynamic aperture in order to achieve high injection efficiency and long Touschek lifetime. In order to predict the performance of the storage ring, different simulations are done, e.g. to calculate the dynamic aperture in the 6D space, the frequency map analysis and diffusion maps analysis are made.**The Figure shows the diffusion map and frequency map for the electron bunch circulating in the storage ring. The diffusion is low for a blue color and for light blue where you see such circles it is a bit higher.**





Bioinformatics

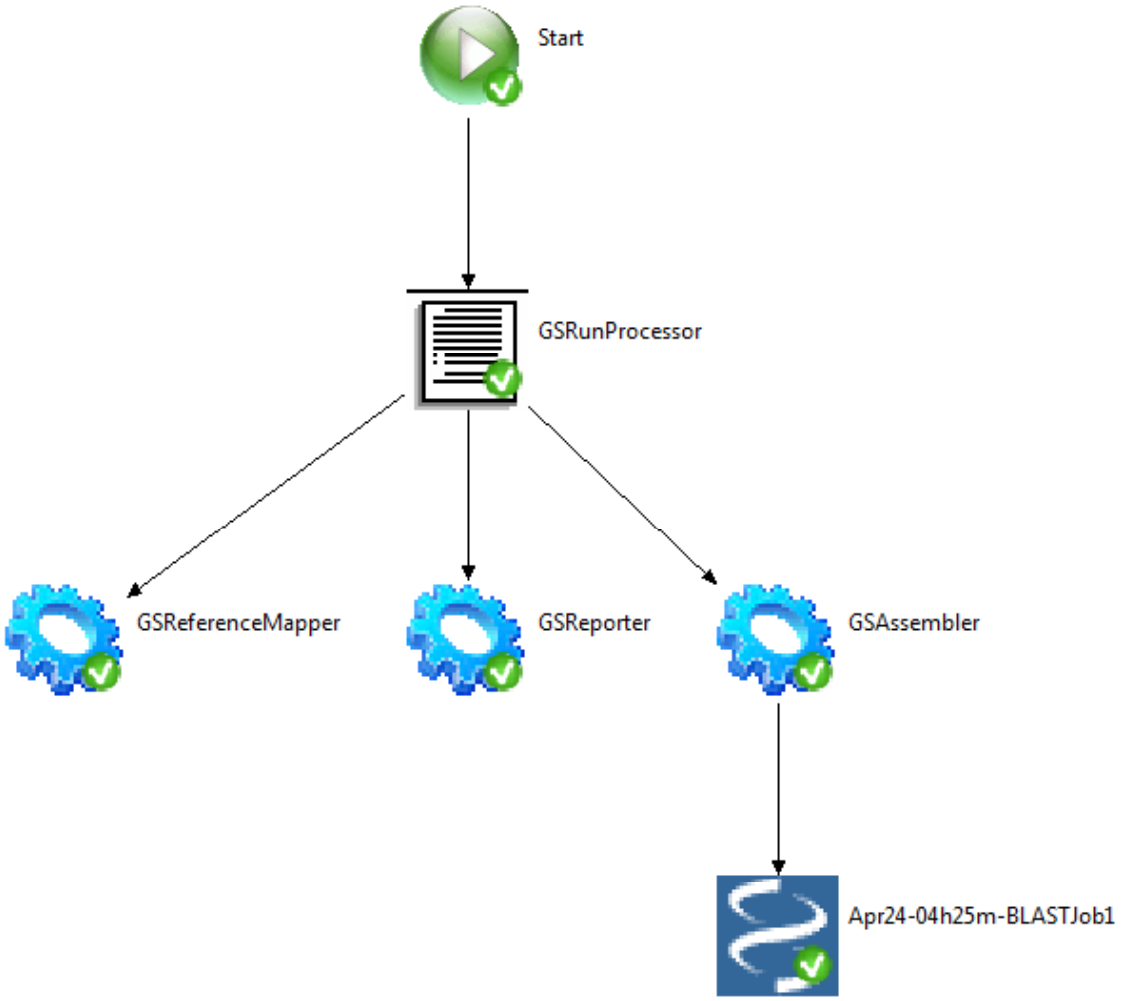
Processing of DNA sequencing data



- High-throughput GS FLX Instrument (Roche Diagnostics)
- UNICORE Commandline Client (UFTP)
- Target System Storage (PL-Grid)
  - data storage
- Target System (PL-Grid)
  - workflow execution

Bioinformatics

Workflow for processing DNA sequencing data



PL-Grid allows to process data in 4h instead of 80h

- GS Run Processor
- GS Reporter
- GS Reference Mapper
- GS Assembler (run in parallel)
- BLAST

Health
Processing of medical images

An example result of the matching procedure visualized via volume rendering. Red image represents reference dataset that is transformed in the registration process, while the green image is the current object under examination. M Chlebiej *et al.*

- UNICORE Rich Client
- UNICORE Command line Client (UFTP)
- Web access

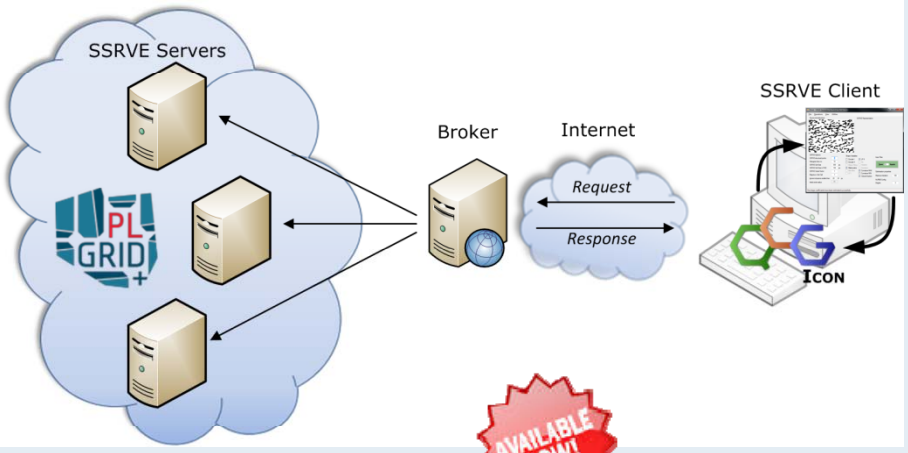
- UNICORE Storage (PL-Grid)
  - secure storage for medical images

- Target System (PL-Grid):
  - dedicated software to process medical images

Metallurgy
SSRVE – Existing Prototype

SSRVE objective:

Creation of Grid Service, which supports creation of Statistically Similar Representative Volume Element by parallelization of optimization procedure allowing massive parallel calculations on grid infrastructure.



SSRVE main workflow:

- SSRVE Client (MS Windows OS) – import of micrographs of dual phase steel, image processing, analysis of shapes coefficients, detection of histograms characteristics, export of *ssrve* input file, launch of QCG Icon,
- QCG Icon – automatic import of *ssrve* input file, configuration of a job (automatic configuration for basic and intermediate users is available, however all options can be set up by advanced users),
- SSRVE Server (Linux OS) – launch path includes *ssrve* file, which contains startup parameters for parallel optimization methods; the results are sent back to SSRVE Client.

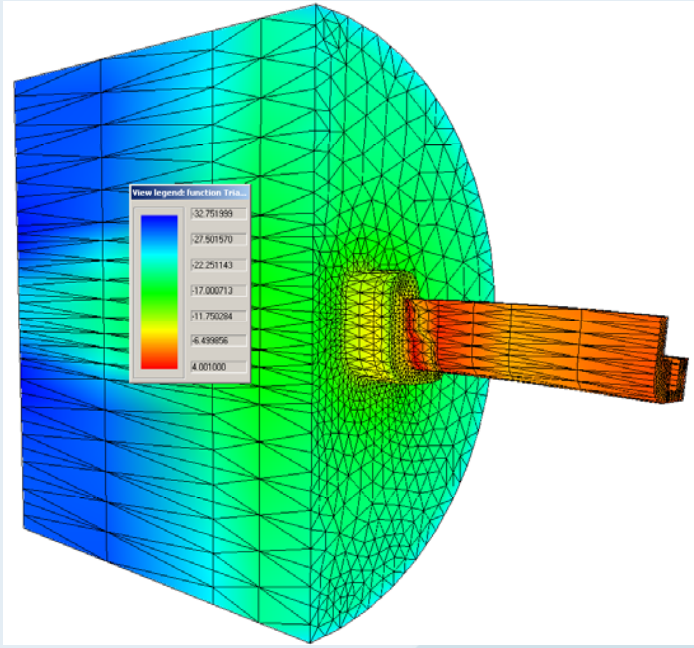
Metallurgy

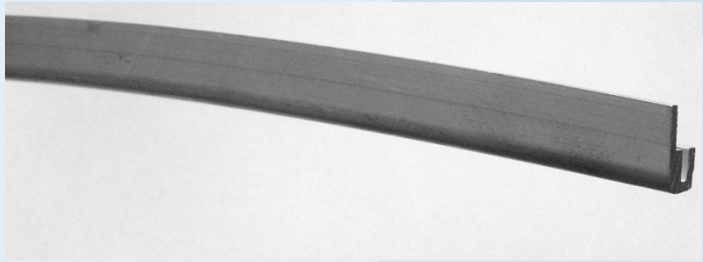
Extrusion – Prototype planned in February 2013

Main Objectives:

Support of engineers in demanding calculations for optimization of the metallurgical process of profiles extrusion. Optimization includes:

- shape of foramera,
- channel position on a die,
- calibration stripes,
- extrusion velocity, ingot temperatures, tools.





Extrusion Service Realization:

- Implementation of parallel parts of Finite Element Method, used as the part on a Server side (implemented in Fortran),
- GUI will be created as Windows standalone application, integrated with QCG Icon (implemented in C++).

Metallurgy

MCMicro – Prototype planned in May 2013

Main Objectives:

Material engineers support in numerical simulations of static recrystallization by using Monte Carlo approach.

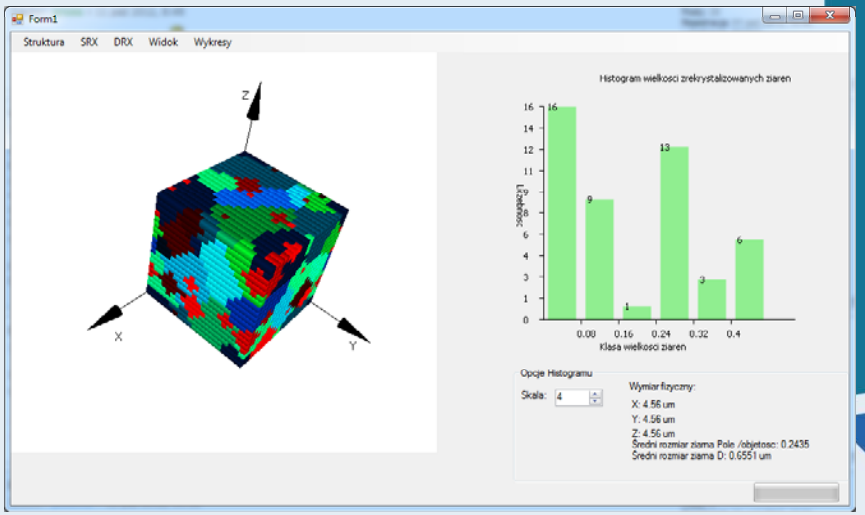
Tasks:

- Implementation of graphical user interface for configuration of numerical calculations of the material model and visualization of obtained results,
- Implementation of module for generation of initial microstructure and export of input file for parallel calculations,
- Design and implementation of parallel Monte Carlo model for static recrystallization.

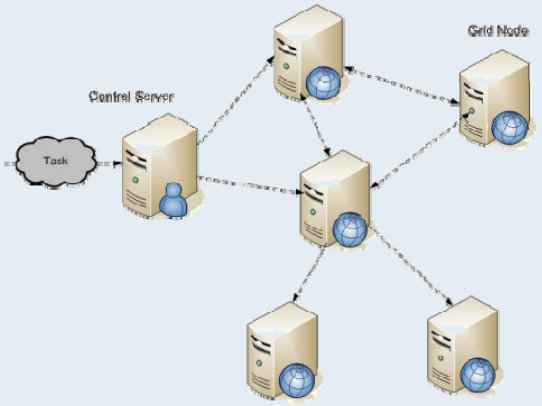
Implementatoin issues:

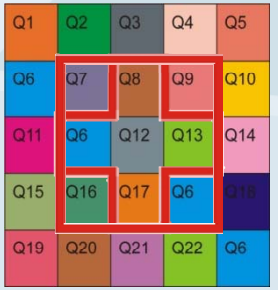
- Client's GUI – C++/CLI, ASP.NET,
- Server side – unmanaged C++, MPI.

Graphical User Interface on Client's Side



Parallel Model Calculation on Server's Side





$\Omega = \{Q_0, \ldots, Q_{n-1}\}$

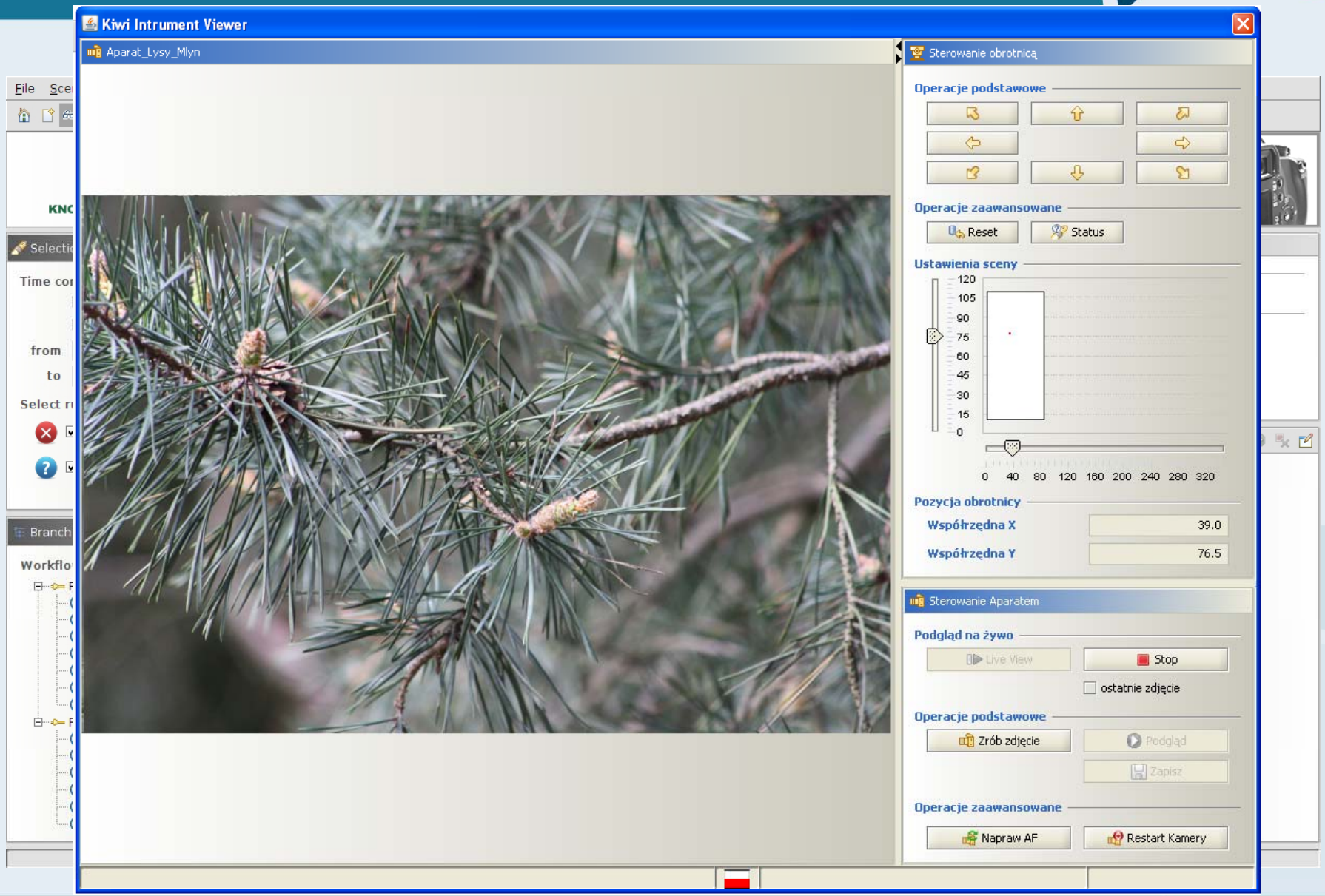
Ecology
KIWI Eye (1.0) observation set

A set for automatic photo observations

Functionality:

- Remote control of the camera
- Remote viewing of the video from the camera
- High resolution photos (DSLR type camera with interchangeable lenses)
- Scene and measuring points definition
- Workflow-type management software and remote access

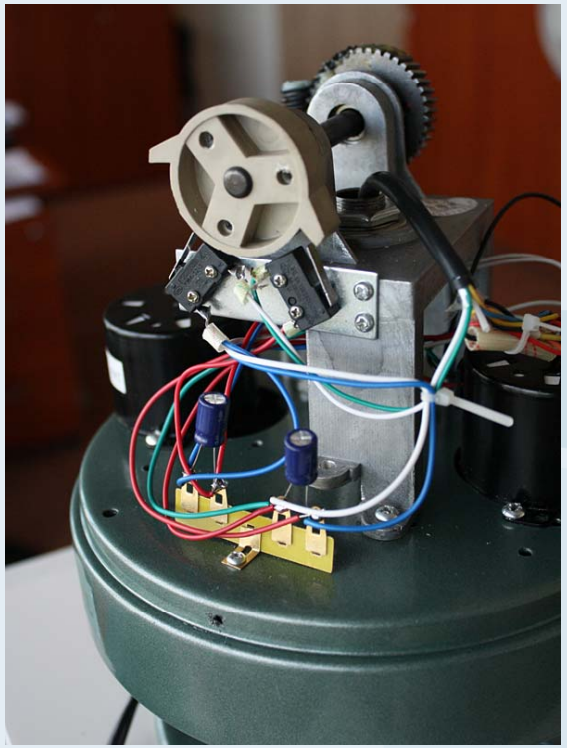
Ecology
KIWI – workflows and remote access



Ecology
KIWI – universal monitoring platform

Achievements:

- **KNOW-HOW** associated with the implementation of the observational sets for the **WLIN project**
- **Observation equipment**, hardware, software, materials, technologies, **problems and their solutions**



AstroGrid-PL

- Whole-community astronomical grid
- Main Polish astronomical institutes involved CAMK PAN (coordinator), CA UMK, OA UJ and a few others
- Integrated platform for various areas of research
- Universal core services
- Data management very important in astronomy!
- Use of the Grid brings huge benefits (professional infrastructure, ease of data sharing, direct access from computational resources)
- Simple WWW & CLI interfaces for advanced data/metadata management (e.g. data receiving and archiving, cataloguing, sharing raw data or catalogues)
- Dedicated support for specific instrumental projects
- iRods solution is being considered & tested
- This service is basis for other services

AstroGrid-PL core sevices

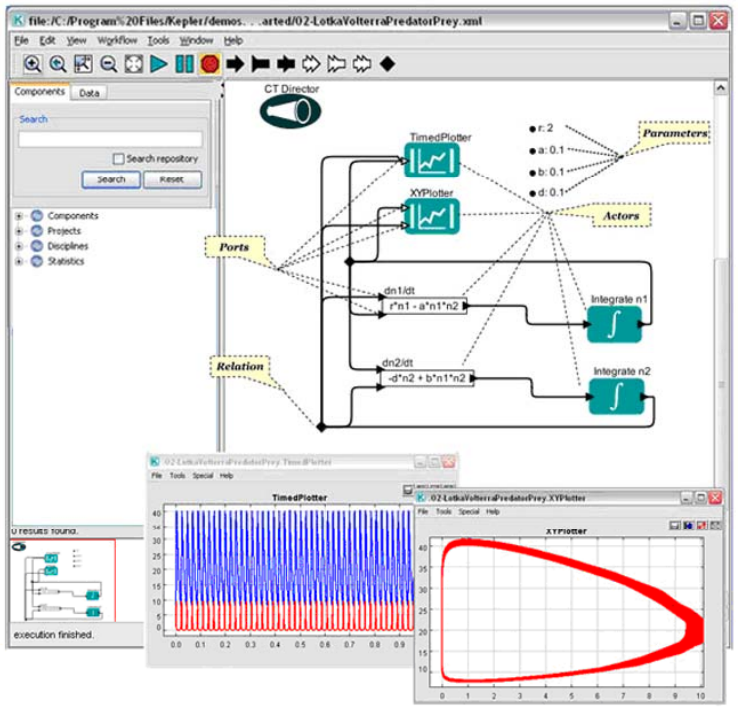
Polish Virtual Observatory and Workflow Environment for Astronomy

- The Virtual Observatory (VObs) is an international astronomical community-based initiative aiming to provide transparent and distributed access to data available worldwide. The VObs consists of a collection of data centers each with unique collections of astronomical data, software systems and processing capabilities.

- Service goals:
  - Setup National VObs Data Center, integration of Polish data and join international efforst

- Workflow Environment
  - Provide astronomers with workflow environment
  - Enable easy resource switching: desktop – cluster – grid
  - Library of template scenarios for popular activities
  - Dedicated support for selected projects
  - Kepler environment is being tested...

- Universal Fluid Dynamics code Piernik



Conclusions

- Further developement 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
  - Michał Turała
  - Marian Bubak
  - Krzysztof Zieliński
  - Karol Krawentek
  - Agnieszka Szymańska
  - Maciej Twardy
  - Teresa Ozga
  - Angelika Zaleska-Walterbach
  - Andrzej Oziębło
  - Zofia Mosurska
  - Marcin Radecki
  - Renata Słota
  - Tomasz Gubała
  - Darin Nikolow
  - Aleksandra Pałuk
  - Patryk Lasoń
  - Marek Magryś
  - Łukasz Flis
- ICM
  - Marek Niezgódka
  - Piotr Bała
  - Maciej Filocha
- PCSS
  - Maciej Stroiński
  - Norbert Meyer
  - Krzysztof Kurowski
  - Bartek Palak
  - Tomasz Piontek
  - Dawid Szejnfeld
  - Paweł Wolniewicz
- WCSS
  - Jerzy Janyszek
  - Paweł Tykierko
  - Paweł Dziekoński
  - Bartłomiej Balcerek
- TASK
  - Rafał Tylman
  - Mścislaw Nakonieczny
  - Jarosław Rybicki

… and many others….