Distributed computing and Grid technologies.

Vladimir Korenkov
(JINR, University “Dubna”)

CERN, 05.11.09

Концепция Грид

«Грид - это система, которая:

- **координирует использование ресурсов при отсутствии централизованного управления этими ресурсами**
- **использует стандартные, открытые, универсальные протоколы и интерфейсы.**
- **обеспечивает высококачественное обслуживание»**

(Ian Foster: "What is the grid? ", 2002 г.)

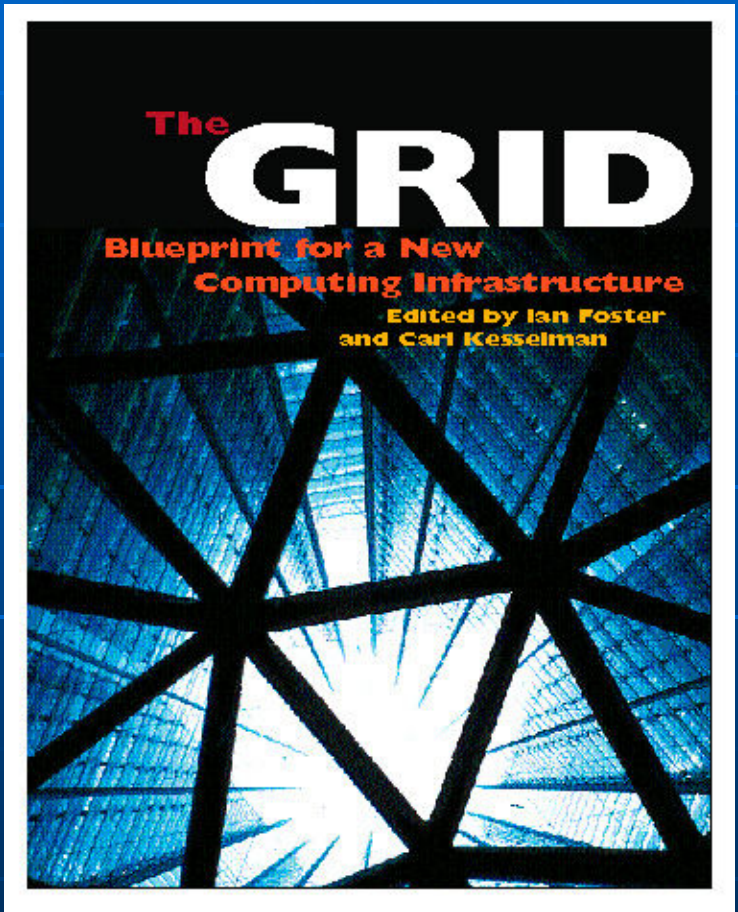
Создание компьютерной инфраструктуры нового типа, обеспечивающей глобальную интеграцию информационных и вычислительных ресурсов на основе управляющего и оптимизирующего программного обеспечения (middleware) нового поколения.

Междисциплинарный характер грид: развиваемые технологии применяются в физике высоких энергий, космофизике, микробиологии, экологии, метеорологии, различных инженерных и бизнес приложениях.

Виртуальные организации (VO)

Five Emerging Models of Networked Computing From *The Grid*

- **Distributed Computing**
  - **|| synchronous processing**
- **High-Throughput Computing**
  - **|| asynchronous processing**
- **On-Demand Computing**
  - **|| dynamic resources**
- **Data-Intensive Computing**
  - **|| databases**
- **Collaborative Computing**
  - **|| scientists**



Ian Foster and Carl Kesselman, editors, "The Grid: Blueprint for a New Computing Infrastructure," Morgan Kaufmann, 1999, http://www.mkp.com/grids

Развитие научных исследований в физике высоких энергий, астрофизике, биологии, науках о Земле и других требует совместной работы многих организаций по обработке большого объема данных в относительно короткие сроки. Для этого необходимы географически распределенные вычислительные системы, способные передавать и принимать данные порядка десятков терабайт в сутки, одновременно обрабатывать десятки тысяч задач и долговременно хранить петабайтные объемы данных.

Современные Грид-инфраструктуры обеспечивают интеграцию аппаратурных и программных ресурсов, находящихся в разных организациях в масштабах стран, регионов, континентов в единую вычислительную среду, позволяющую решать задачи по обработке сверхбольших объемов данных, чего в настоящее время невозможно достичь в локальных вычислительных центрах.

Five big ideas

Resource sharing: **Global sharing** is the very essence of grid computing.

Secure access: **Trust** between resource providers and users is essential, especially when they don't know each other. Sharing resources conflicts with security policies in many individual computer centers, and on individual PCs, so getting grid security right is crucial.

Resource use: **Efficient, balanced use of computing resources** is essential.

The death of distance: **Distance should make no difference:** you should be able to access to computer resources from whereever you are.

Open standards: Interoperability between different grids is a big goal, and is driven forward by the adoption of open standards for grid development, making it possible for everyone can contribute constructively to grid development. Standardization also encourages industry to invest in developing **commercial grid services and infrastructure.**

Grid is a result of IT progress

Network vs. computer performance:

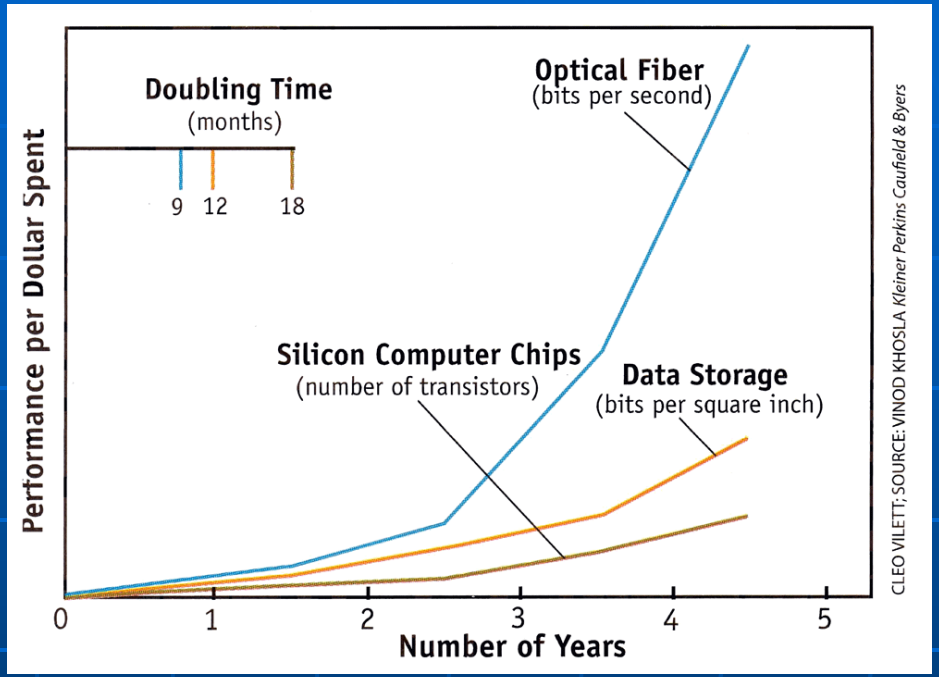
- Computer speed doubles every 18 months
- Network speed doubles every **9** months

1986 to 2000:

- Computers: 500 times faster
- Networks: 340000 times faster

2001 to 2010 (projected):

- Computers: 60 times faster
- Networks: 4000 times faster



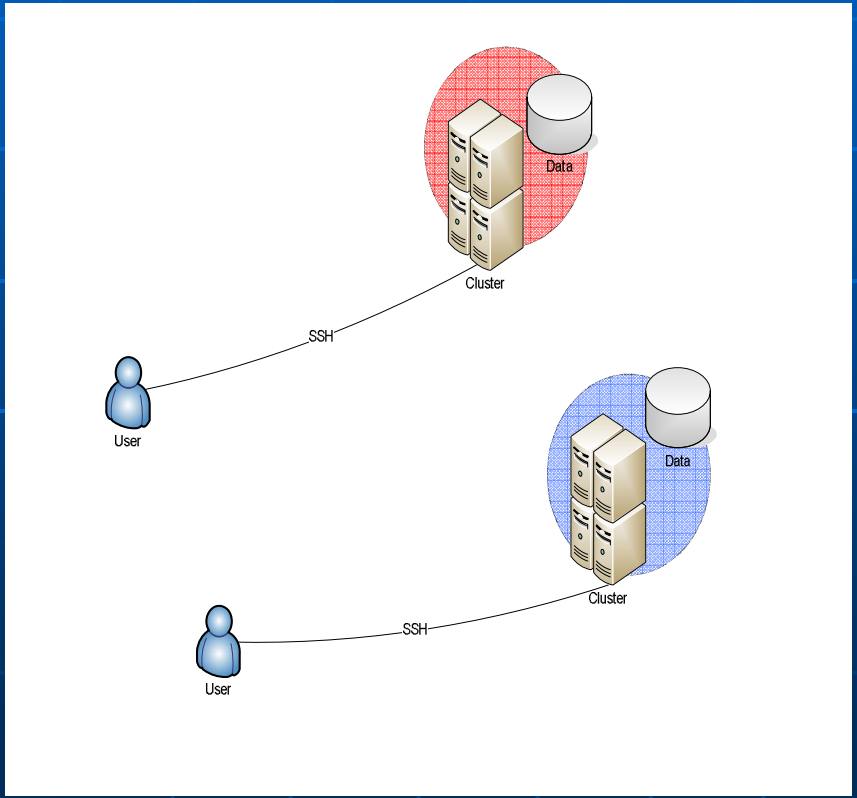
Excellent wide area networks provide for a distributed supercomputer – ***the Grid***

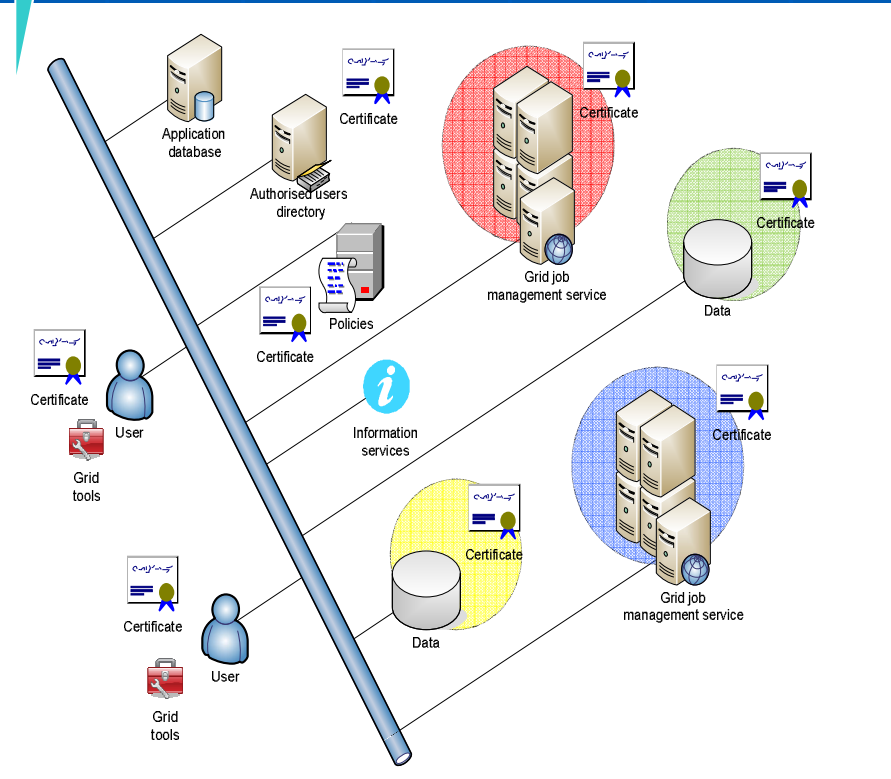
"Operating system" of such a computer is ***Grid middleware***

Graph from "The Triumph of the Light", G. Stix, Sci. Am. January 2001

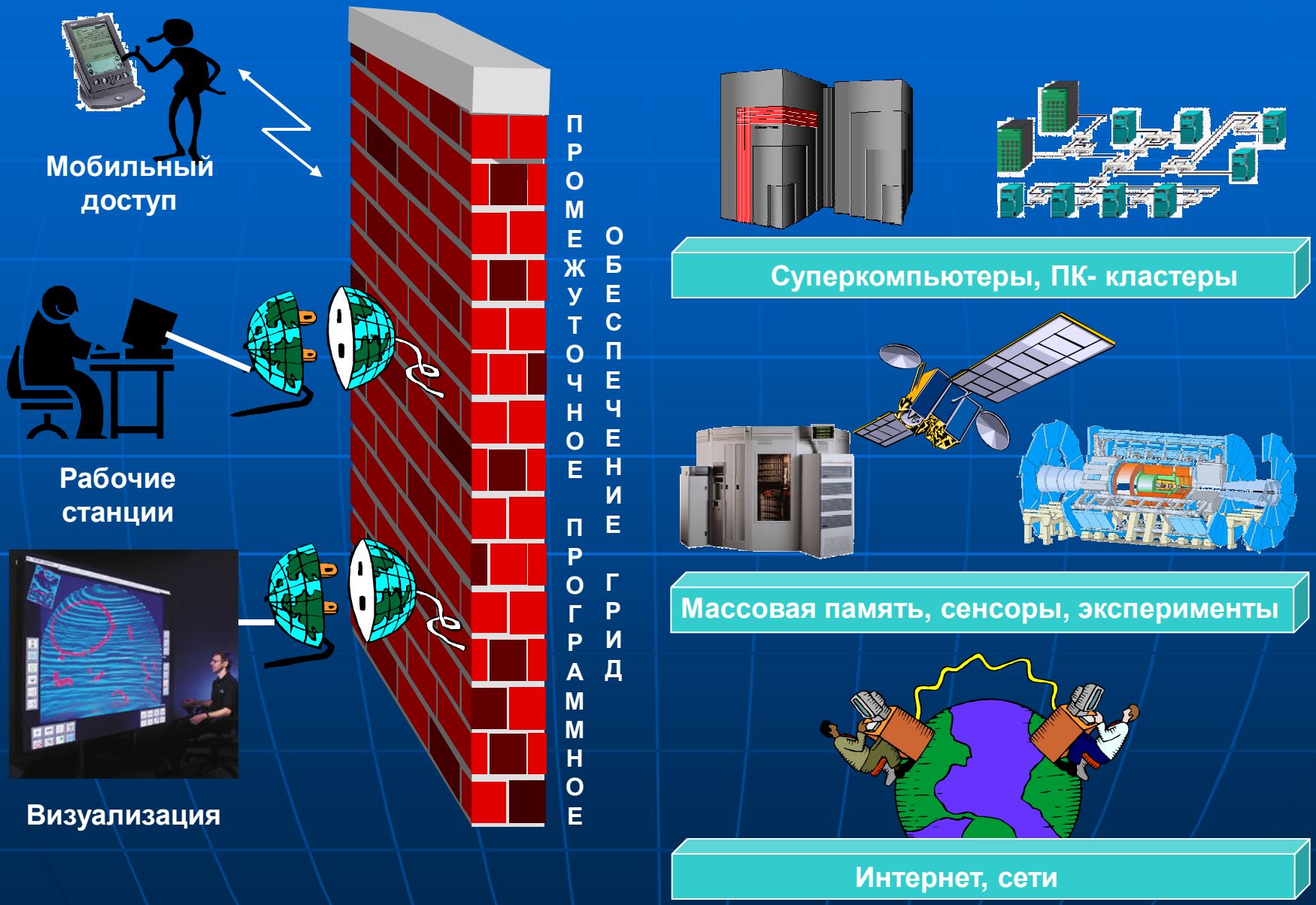
From the conventional HPC…

To the Grid





Грид - это средство для совместного использования вычислительных мощностей и хранилищ данных посредством интернета



Grid-архитектура с точки зрения программного обеспечения

Приложения для конкретных научных сфер:

- Химия
- Биология
- Космология
- Физика высоких энергий
- Экология

Наборы инструментальных средств:

- Распределенные вычисления
- Удаленная визуализация
- Средства программирования
- Интенсивное использование данных
- Удаленные измерения
- Коллективные приложения

Grid-сервисы (middleware):

Независимые от ресурсов и приложений: аутентификация, авторизация, размещение и распределение ресурсов, получение результатов, статистика и служебная информация, удаленный доступ к данным, стратегия, способы обнаружения неисправностей.

Grid-ресурсы (Grid Fabric):

Транспортные протоколы, сервера имен, планировщики использования процессоров, инфраструктура открытого ключа, статистика по отдельным сайтам, сервис каталогов и т.п.

Global Community

TERAGRID

INFN GRID

EUROGRID

K*GRID

PPDG

iVDgL

Open Science Grid

DataGRID

National Grid NG SINGAPORE

Grid3 2003

ACCESS GRID

grangenet

GRIDPP UK Particle Physics

DataTAG

NEESgrid Building the National Virtual Collaboratory for Earthquake Engineering.

GriPhyN Data Intensive Science

CERN openlab for DataGrid applications Developing Solutions for the Data-Intensive Science of the Large Hadron Collider

LCG

WestGrid

NORDUGRID Grid Solution for Wide Area Computing and Data Handling

GridCat

eGee Enabling Grids for E-science in Europe

grid.it project

APGrid Asia-Pacific Grid

Grid Consortium Japan

Ninf

NAREGI 超高速コンピュータ網形成プロジェクト National Research Grid Initiative

Grid Applications Grid Middleware Networking

国立情報学研究所グリッド研究開発推進拠点 NII -The National Institute of Informatics

GridLab

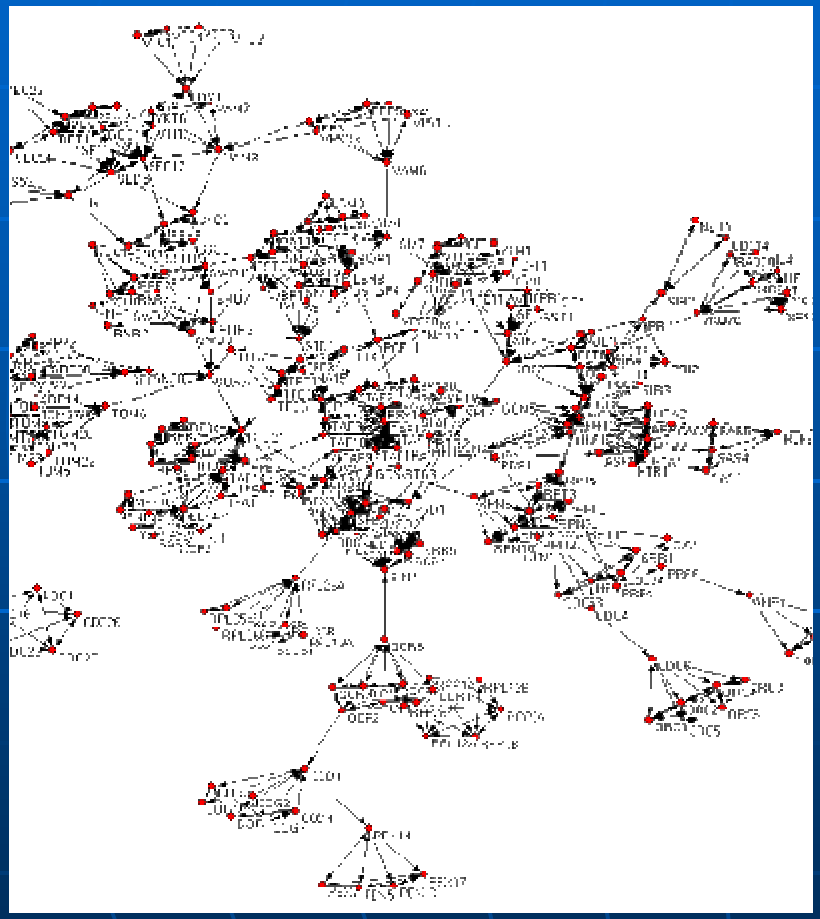
Перспективные области применения грид

- **Физика высоких энергий (LHC, CERN)**
- **TeraGrid**
- **HealthGrid (GEMSS, MammoGrid, Pharma Grid, BioGrid, Infogenmed)**
- **Молекулярная биологии** для моделирования структуры белка, анализа последовательностей ДНК
- **Бизнес** — нефтяная отрасль, геологоразведка
- **В банковском деле**
- **В автомобильной промышленности** Grid позволяет ускорить и удешевить расчет виртуальных крэштестов
- **Airbus и Boeing** используют технологию для постоянного контроля состояния двигателей и других агрегатов и узлов самолета показания датчиков, непрерывно собирающих информацию, через спутник передаются на землю, где анализируются в сетях Grid
- **IBM** активно использует технологию Grid для внутренних нужд (моделирование при создании микропроцессоров следующих поколений, таких, как Power5 и Power6”

Grids in LHC experiments

- Almost all Monte Carlo and data processing today is done via Grid
- There are 20+ Grid flavors out there
  - Almost all are tailored for a specific application and/or specific hardware
- LHC experiments make use of only 3 Grid flavors:
  - gLite
  - ARC
  - OSG
- All experiments develop own higher-level Grid middleware layers
  - ALICE – AliEn
  - ATLAS – PANDA, GANGA, DDM
  - LHCb – DIRAC, GANGA
  - CMS – ProdAgent, CRAB, PhEDEx

Bioinformatics and Grid



- Many large clusters utilized for
  - Services
    - Sequence similarity (BLAST queues)
  - Research
    - Molecular modeling (folding, docking)
    - Training of novel predictors
- Jobs are typically short (3 minutes)
- But plenty (all against all → 10^{12})
- Considerable preparation for single job (couple of gigabytes of data to transfer)

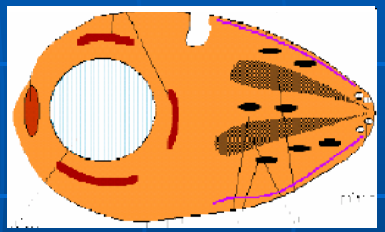
Biomedical applications

- Biomedicine is also a pilot application area
- More than 20 applications deployed and being ported
- Three sub domains
  - Medical image processing
  - Biomedicine
  - Drug discovery

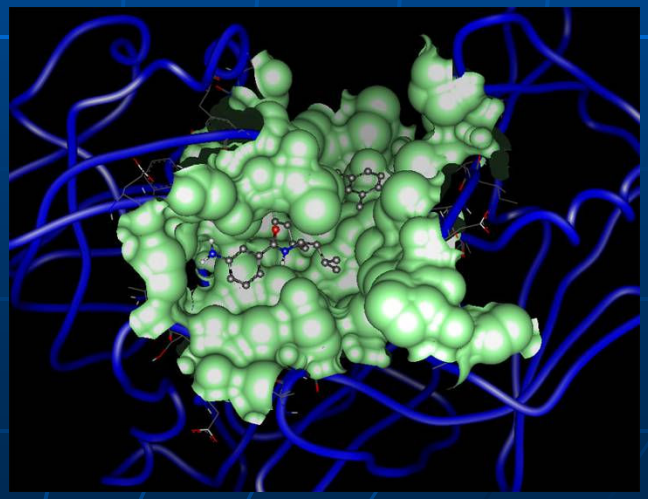
- Use Grid as platform for collaboration
  (don't need same massive processing power or storage as HEP)

Applications Example: WISDOM

- Grid-enabled drug discovery process for neglected diseases
  - *In silico* docking
    - compute probability that potential drugs dock with target protein
  - To speed up and reduce cost to develop new drugs



- WISDOM (World-wide In Silico Docking On Malaria):
  - Three large-scale deployments with more than 6 centuries of computations achieved in 190 days
  - 3,5TB of data produced
  - Up to 5000 computers in 50 countries
  - Some promising in-vitro tests, with relevant biological results.



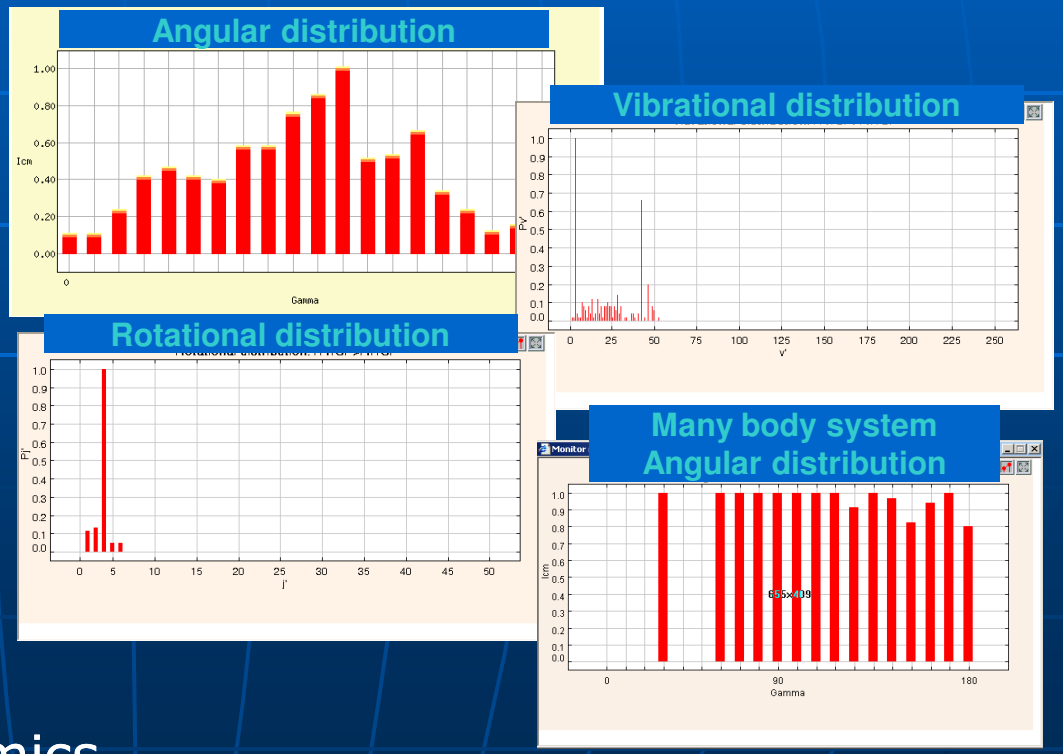
Radio astronomy needs Grid, too

- Enormous datasets, massive computing, innovative instrumentation
  - Dozens of new surveys launched recently
  - Many (10 – 100) terabytes per survey
  - High data rates
  - 10 – 100 researchers per survey
  - International collaborations (almost always)
  - Data is non-proprietary (usually)

Computational Chemistry

- GEMS (Grid Enabled Molecular Simulator) application
  - Calculation and fitting of electronic energies of atomic and molecular aggregates (using high level *ab initio* methods)
  - The use of statistical kinetics and dynamics to study chemical processes

- Virtual Monitors
  - Angular distributions
  - Vibrational distributions
  - Rotational distributions
  - Many body systems

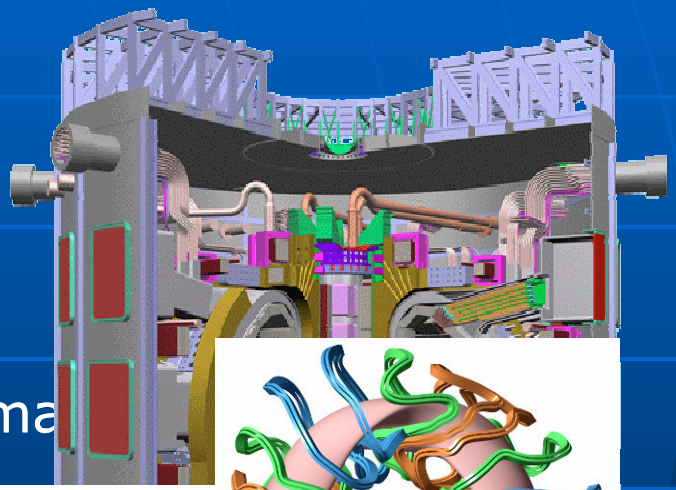


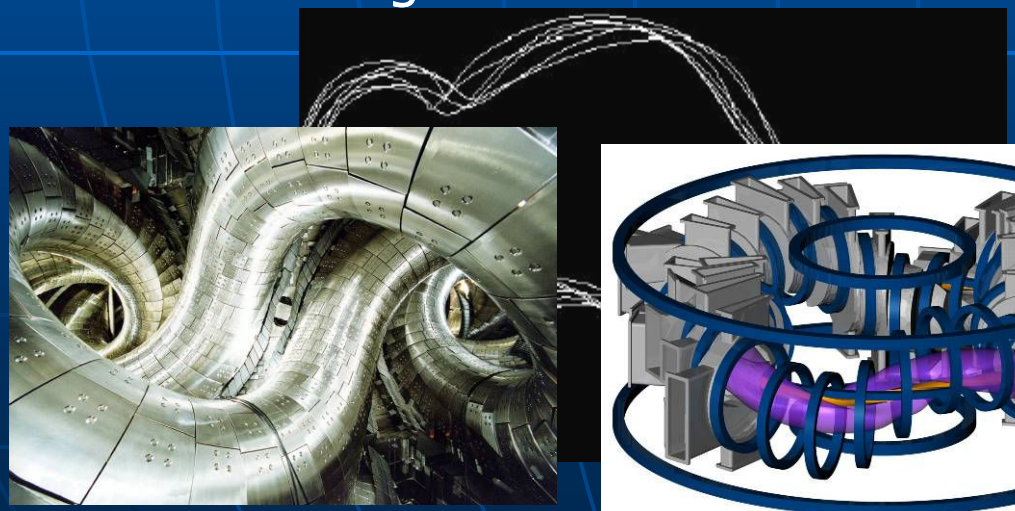
- End-User applications
  - Nanotubes
  - Life sciences
  - Statistical Thermodynamics
  - Molecular Virtual Reality

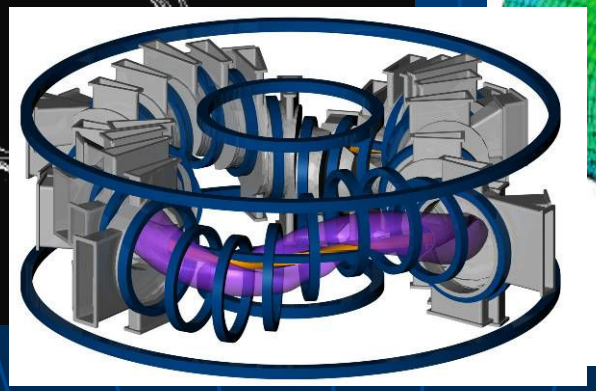
T.Strizh (LIT, JINR)

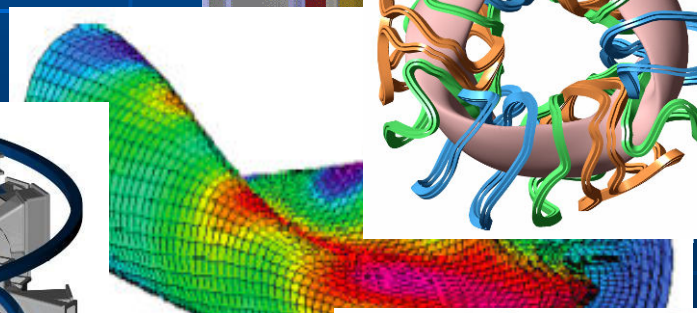
Fusion

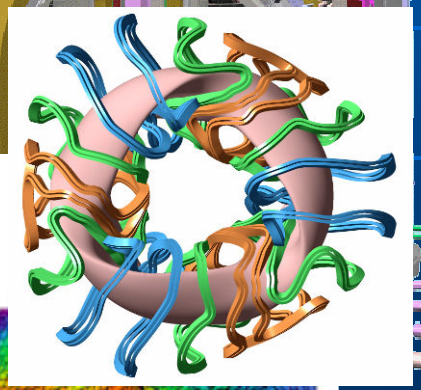
- Large Nuclear Fusion installations
  - E.g. International Thermonuclear Experimental Reactor (ITER)
  - Distributed data storage and handling needed
  - Computing power needed for
    - Making decisions in real time
    - Solving kinetic transport
      → particle orbits
    - Stellarator optimization
      → magnetic field to contain the plasma

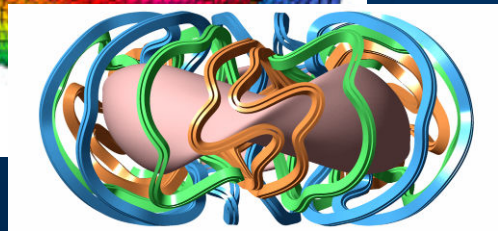






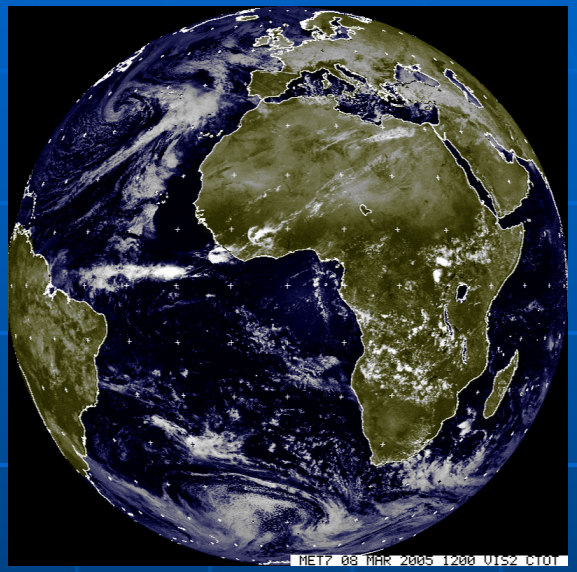






Earth Science Applications

- Community
  - Many small groups that aggregate for projects (and separate afterwards)

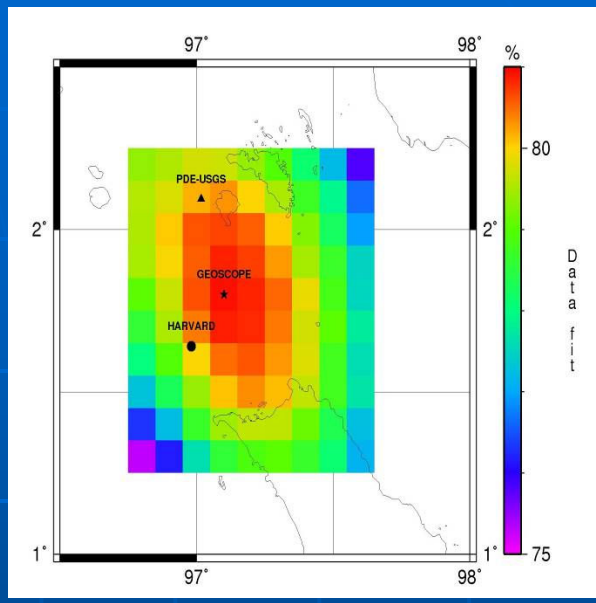


- The Earth
  - Complex system
  - Independent domains with interfaces
    - Solid Earth – Ocean – Atmosphere
  - Physics, chemistry and/or biology

- Applications
  - Earth observation by satellite
  - Seismology
  - Hydrology
  - Climate
  - Geosciences
  - Pollution
  - Meteorology, Space Weather
  - Mars Atmosphere
  - Database Collection

Earth Sciences: Earthquake analysis

- Seismic software application determines:
  Epicentre, magnitude, mechanism
  → May make it possible to predict future earthquakes
  → Assess potential impact on specific regions



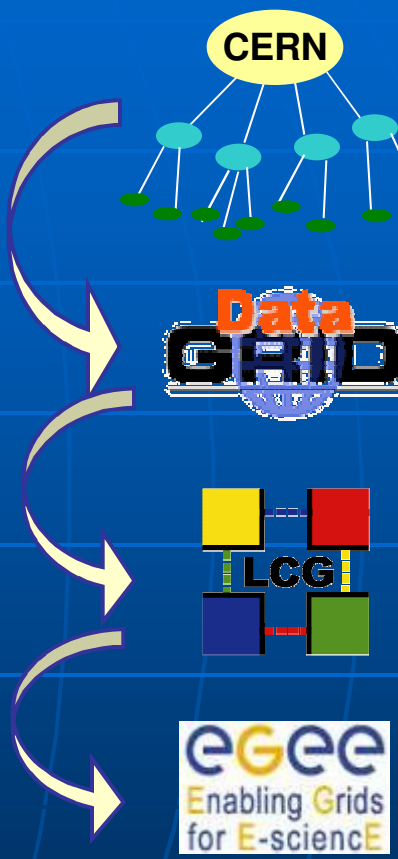
- Analysis of Indonesian earthquake (28 March 2005)
  - Data from French seismic sensor network GEOSCOPE transmitted to IPGP within 12 hours after the earthquake
  - Solution found within 30 hours after earthquake occurred
    - 10 times faster on the Grid than on local computers
  - Results
    - Not an aftershock of December 2004 earthquake
    - Different location (different part of fault line further south)
    - Different mechanism
- Rapid analysis of earthquakes is important for relief efforts

How to Grid-enable your science

- Application-specific software is formalized as *Run-Time Environment (RTE)*
  - Installed and configured at each site, typically by an authorized user/manager
  - Advertised via information system for match-making
  - Jobs only need to specify the RTE name and version
  - Special authorization plugins are available in case of a commercial application software
- Users with similar applications are grouped into *Virtual Organisations (VOs)*
  - A VO typically corresponds to a project
  - A VO contributes the hardware
- Some VOs at JINR:
  - CERN VOs: ALICE, ATLAS, CMS, LHCb
  - HONE
  - Fusion
  - BioMed
  - eEarth
  - Panda

Some history

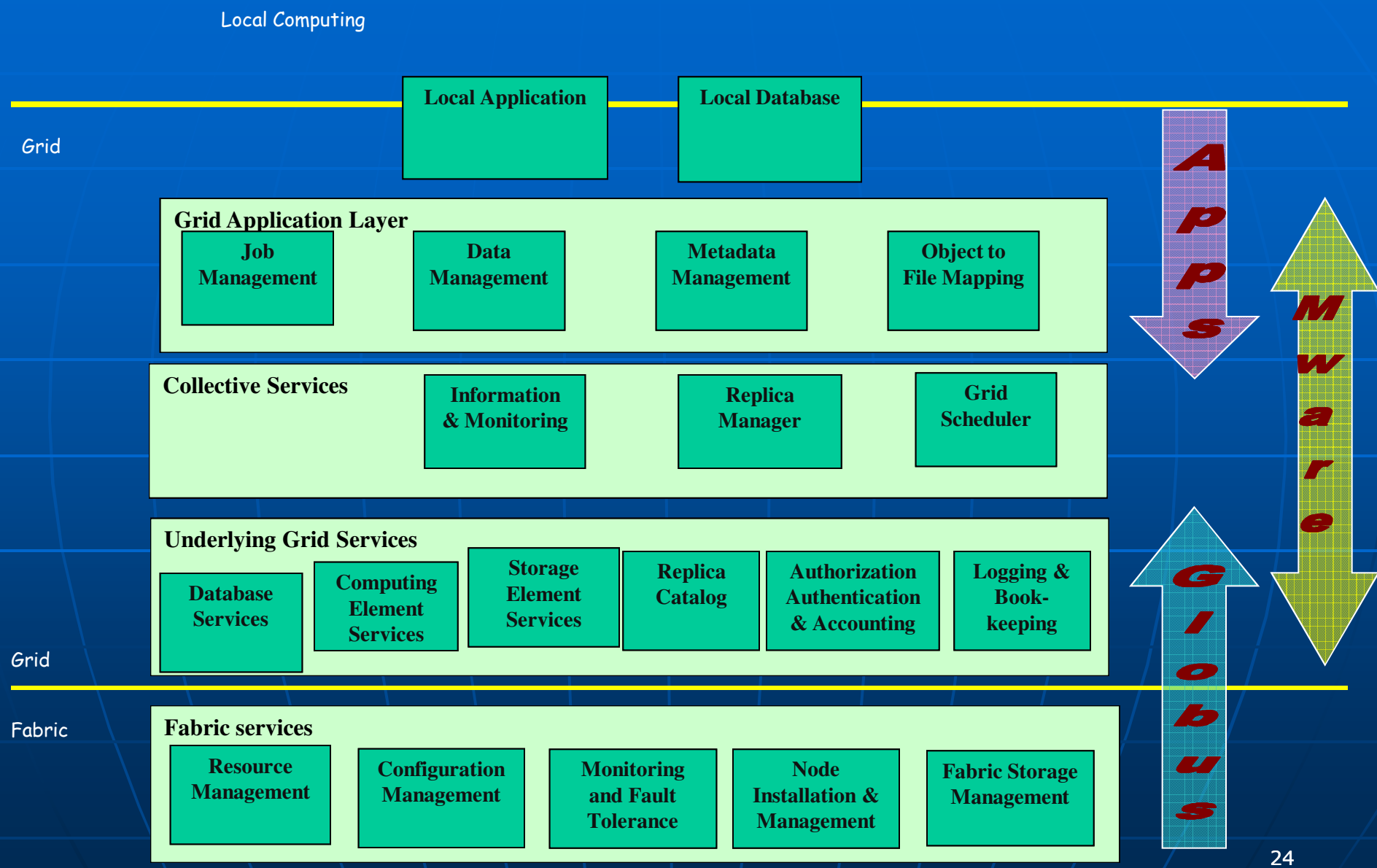
- 1999 – Monarc Project
  - Early discussions on how to organise distributed computing for LHC
- 2001-2003 - EU DataGrid project
  - middleware & testbed for an operational grid
- 2002-2005 – LHC Computing Grid – LCG
  - deploying the results of DataGrid to provide a production facility for LHC experiments
- 2004-2006 – EU EGEE project phase 1
  - starts from the LCG grid
  - shared production infrastructure
  - expanding to other communities and sciences
- 2006-2008 – EU EGEE-II
  - Building on phase 1
  - Expanding applications and communities …
- 2008-2010 – EU EGEE-III



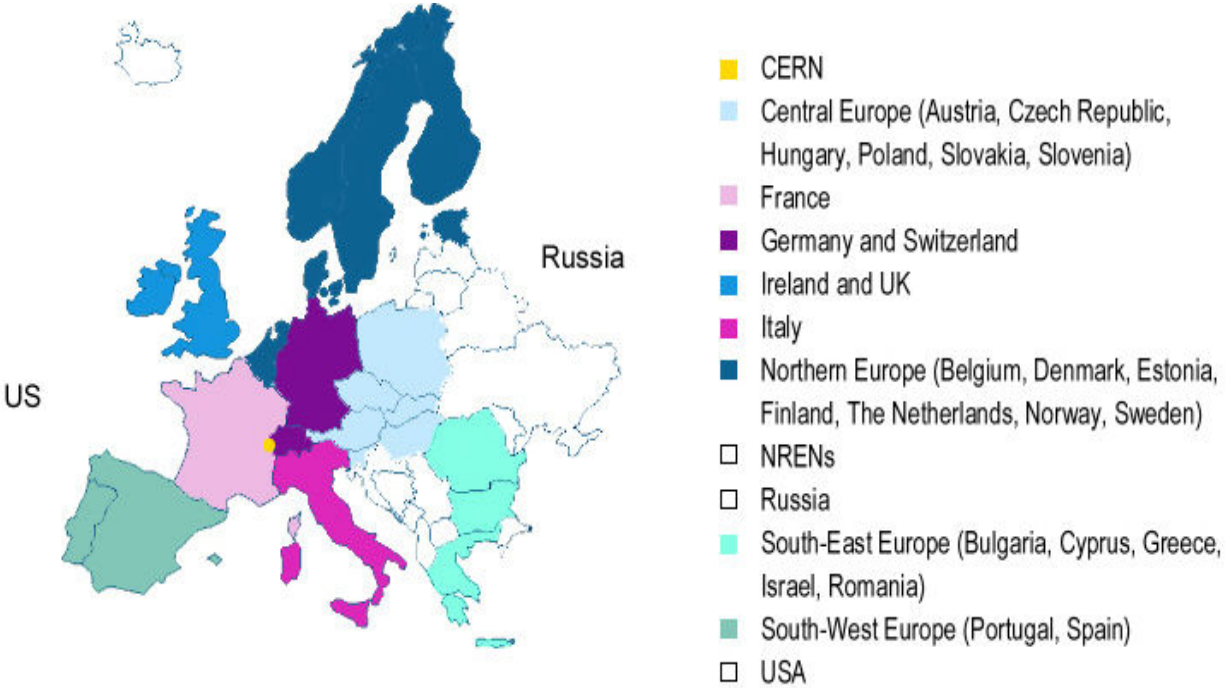
MONARC project

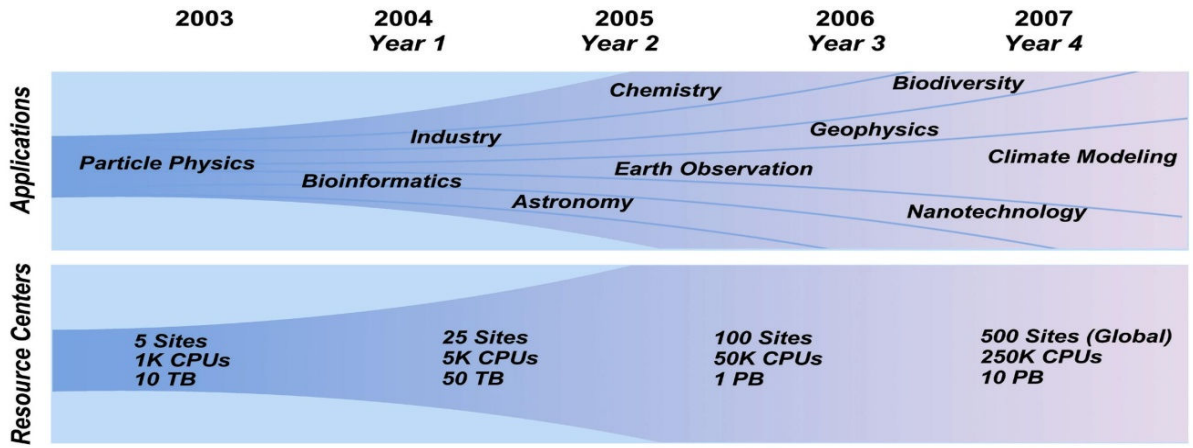
- 1998 – MONARC project
  - a distributed model
    - Integrate existing centres, department clusters, recognising that funding is easier if the equipment is installed at home
    - Devolution of control– local physics groups have more influence over how local resources are used, how the service evolves
  - a multi-Tier model
    - Enormous data volumes → looked after by a few (expensive) computing centres
    - Network costs favour regional data access
    - Simple model that HEP can develop and get into production ready for data in **2005**

DataGrid Architecture



EGEE (Enabling Grids for E-sciencE)

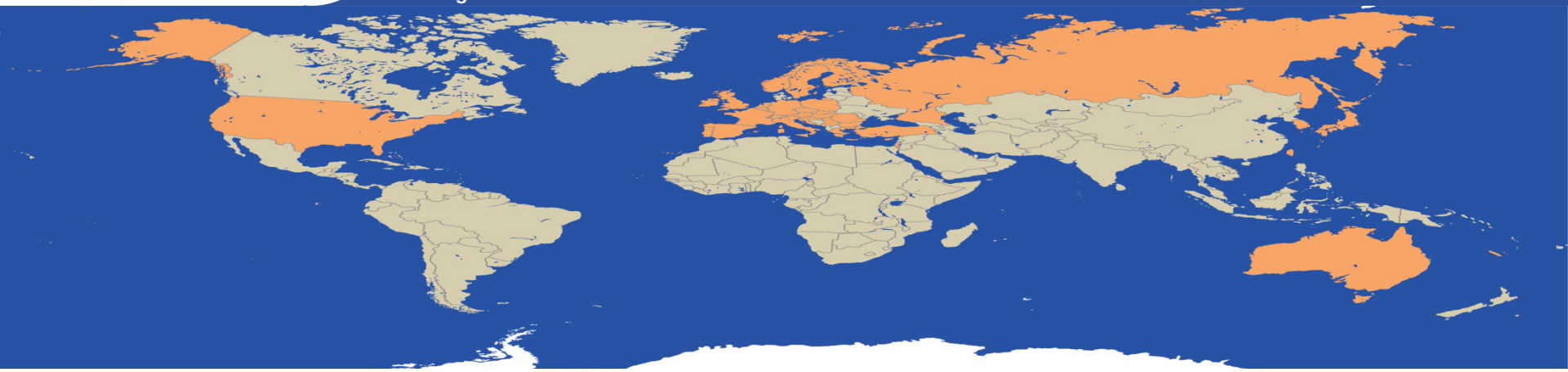




Проект EGEE - Enabling Grids for E-sciencE направлен на создание глобальной инфраструктуры, основанной на технологиях грид.

На первой фазе проекта участвовали **70 организаций из 27 стран**.
В настоящее время в проекте участвуют около 120 организаций из 45 стран.

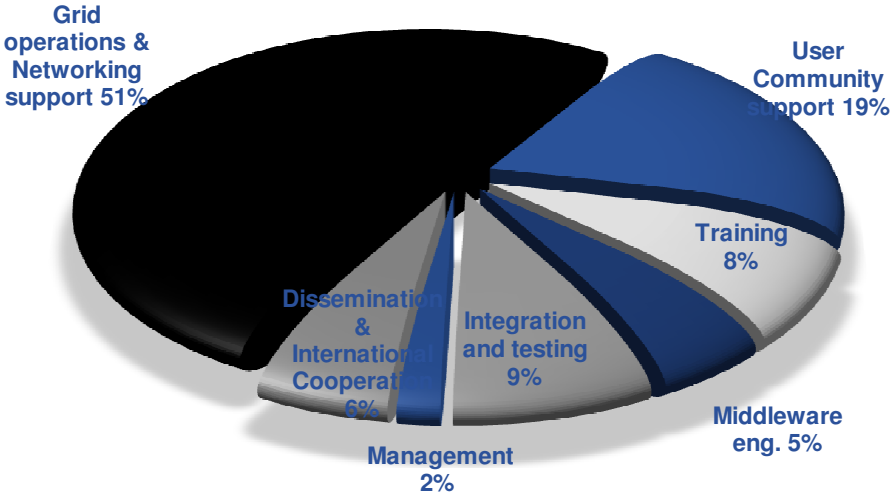
EGEE-III



Flagship Grid infrastructure project co-funded by the European Commission

Main Objectives

- Expand/optimise existing EGEE infrastructure, include more resources and user communities
- Prepare migration from a project-based model to a sustainable federated infrastructure based on National Grid Initiatives



Duration: 2 years
Consortium: ~140 organisations across 33 countries
EC co-funding: 32Million €

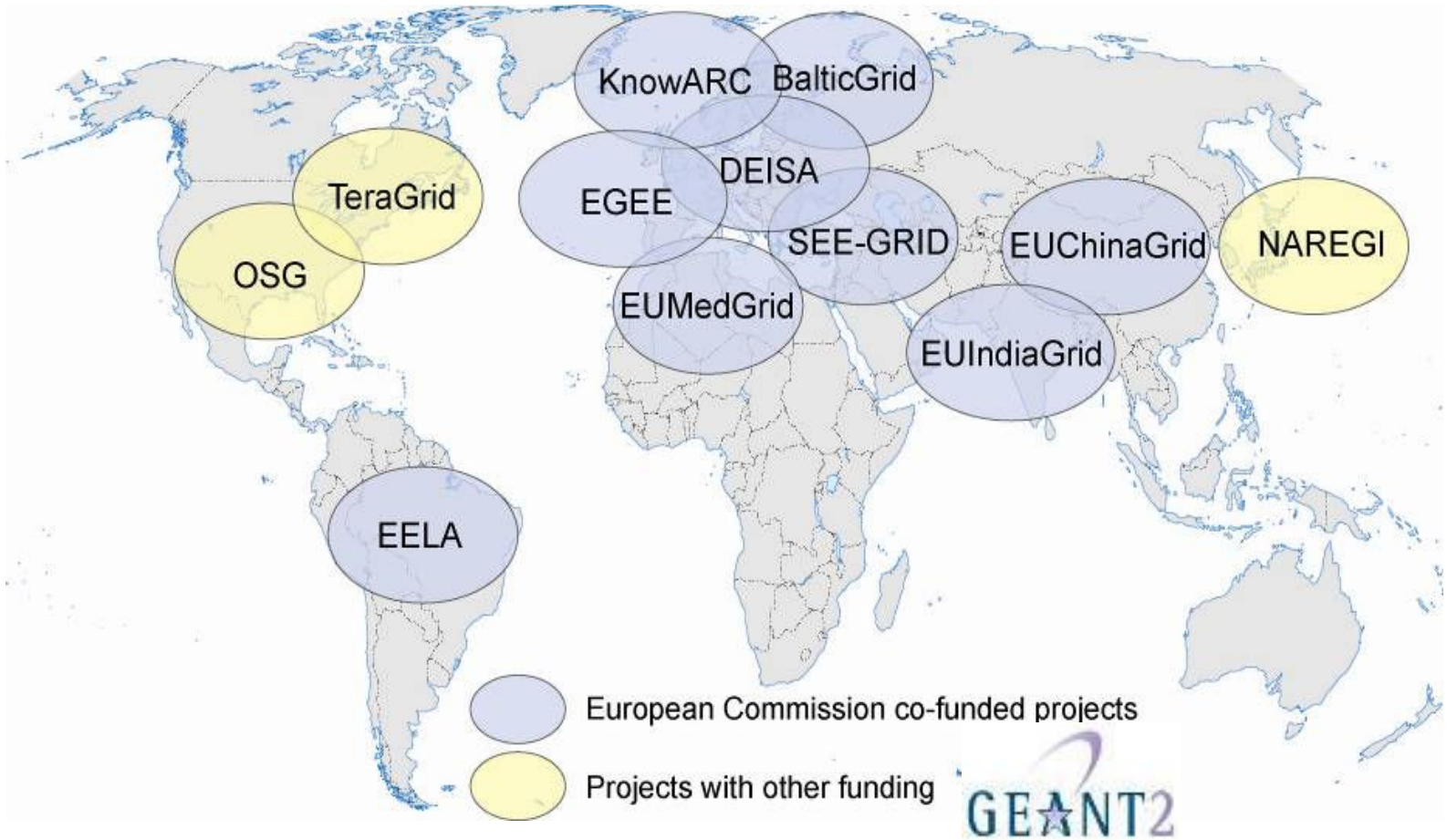
EGEE – What can we deliver?

- **Infrastructure operation**
  - Currently includes >270 sites across 50 countries
  - Continuous monitoring of grid services & automated site configuration/management
  - Support ~300 Virtual Organisations from diverse research disciplines

- **Middleware**
  - Production quality middleware distributed under business friendly open source licence

- **User Support** - *Managed process from first contact through to production usage*
  - Training
  - Expertise in grid-enabling applications
  - Online helpdesk
  - Networking events (User Forum, Conferences etc.)

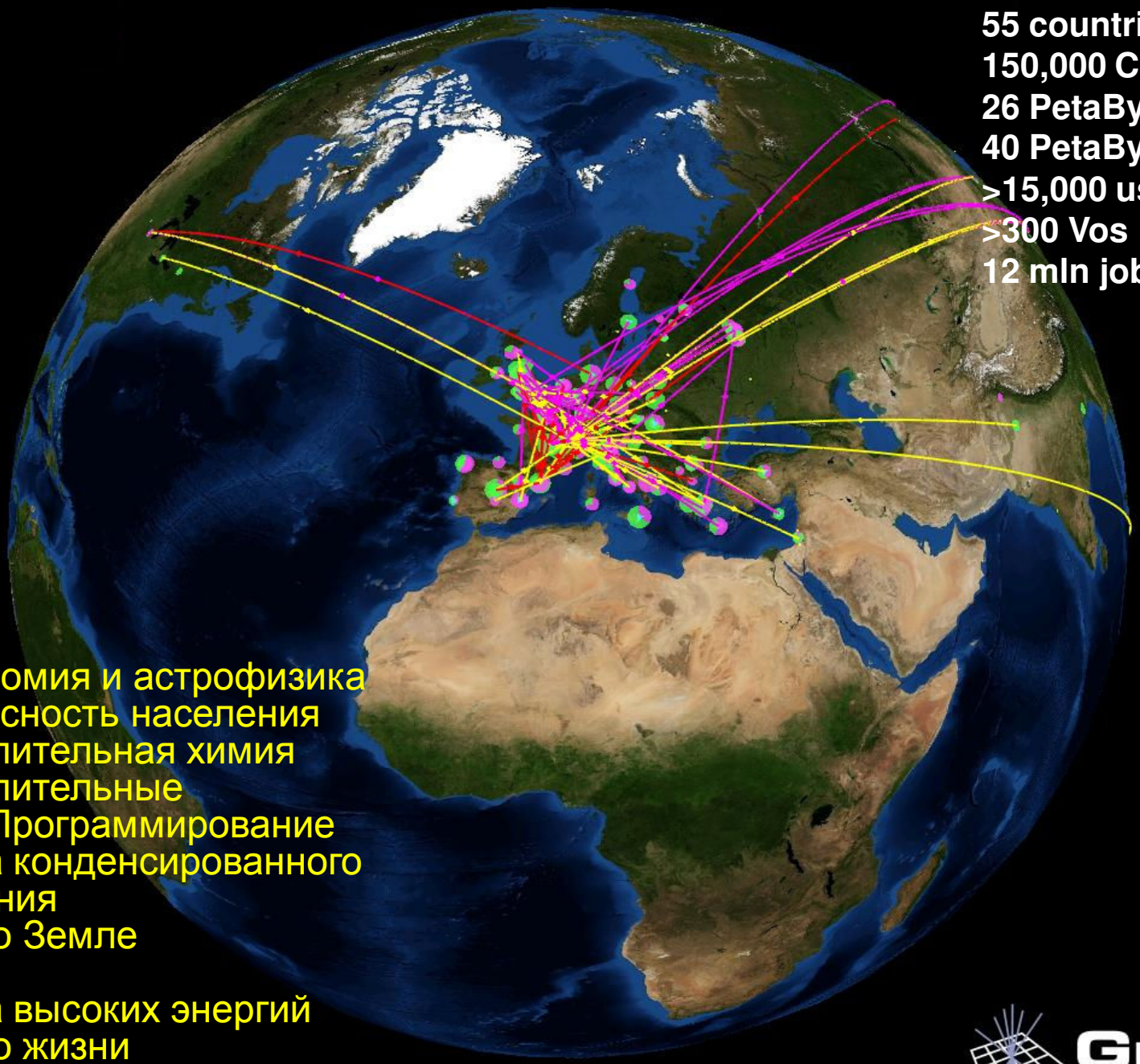
Collaborating e-Infrastructures



Potential for linking ~80 countries by 2008

egee
Enabling Grids for E-sciencE

350 sites
55 countries
150,000 CPUs
26 PetaBytes (Disk)
40 PetaBytes (Tape)
>15,000 users
>300 Vos
12 mln jobs/month

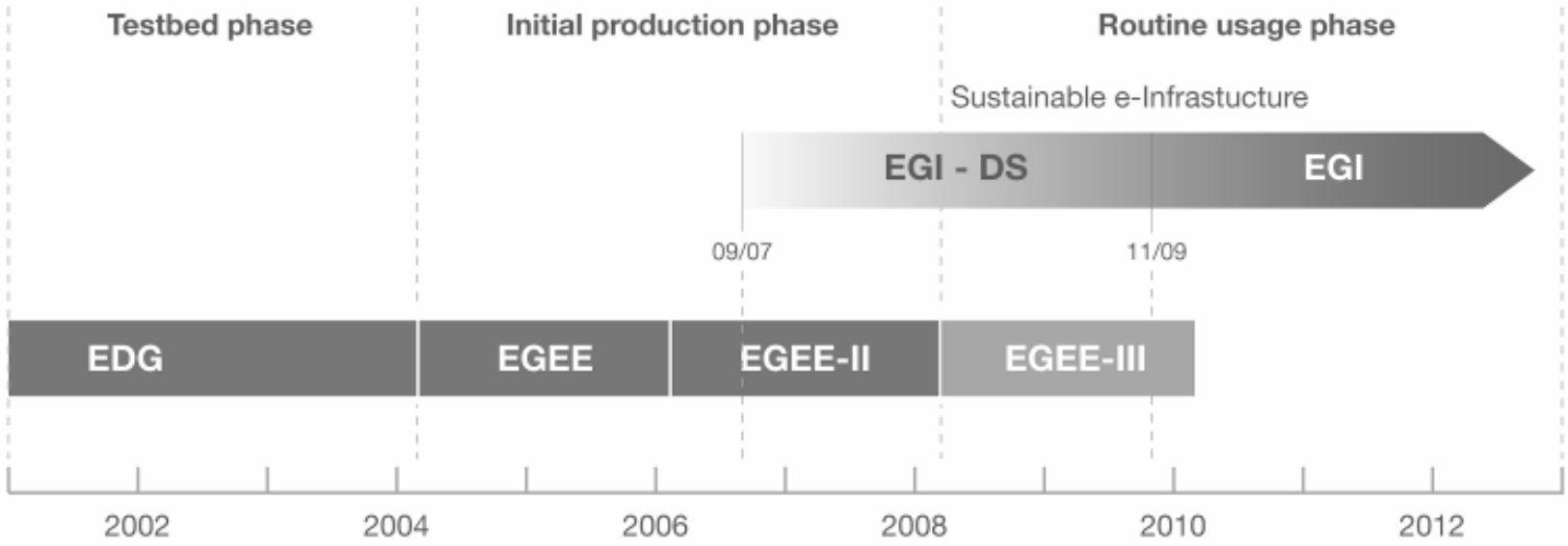


Астрономия и астрофизика
Безопасность населения
Вычислительная химия
Вычислительные науки/Программирование
Физика конденсированного состояния
Науки о Земле
Синтез
Физика высоких энергий
Науки о жизни

GridPP
UK Computing for Particle Physics

European e-Infrastructure

Need to prepare permanent, common Grid infrastructure
Ensure the long-term sustainability of the European e-infrastructure independent of short project funding cycles
Coordinate the integration and interaction between National Grid Infrastructures (NGIs)
Operate the European level of the production Grid infrastructure for a wide range of scientific disciplines to link NGIs

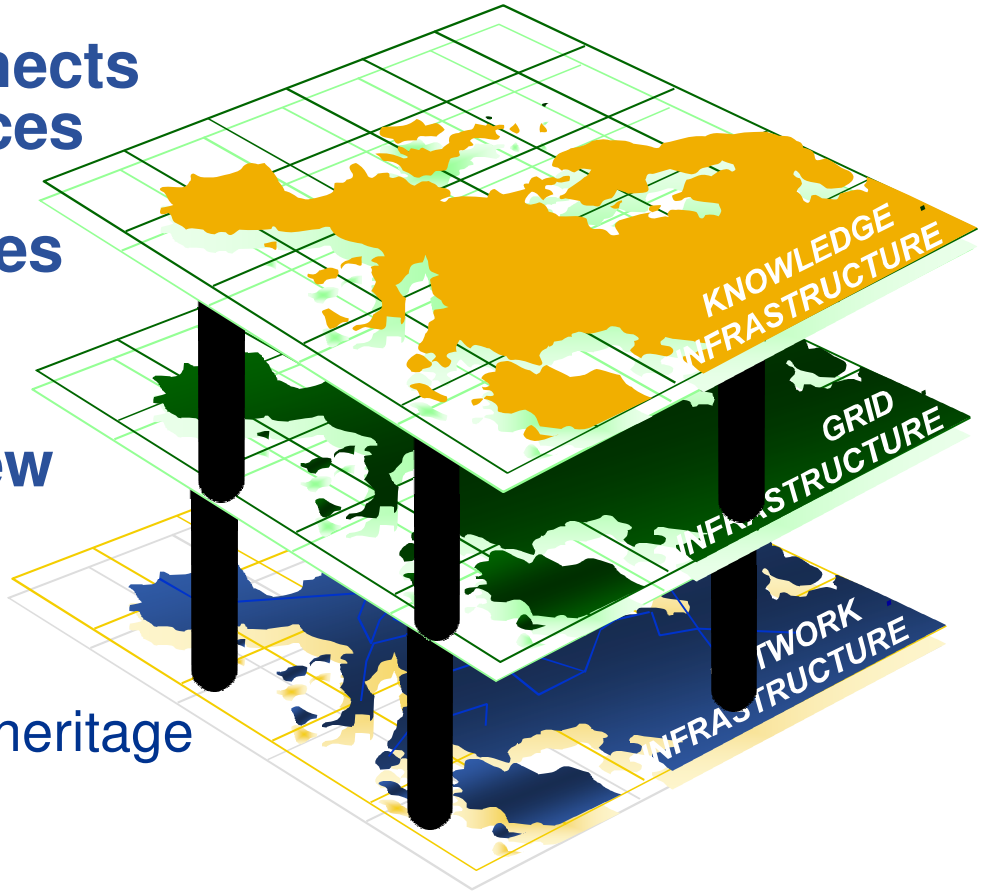


The Future of Grids

- **From e-Infrastructures to Knowledge Infrastructures**

- **Network infrastructure connects computing and data resources and allows their seamless usage via Grid infrastructures**

- **Federated resources and new technologies enable new application fields:**
  - Distributed digital libraries
  - Distributed data mining
  - Digital preservation of cultural heritage
  - Data curation



→ Knowledge Infrastructure

Major Opportunity for Academic and Businesses alike

EGEE & OSG

WLCG depends on two major science grid infrastructures

EGEE - Enabling Grids for E-Science

OSG - US Open Science Grid

LHC Computing Grid Project (LCG)

Основной задачей проекта LCG является создание глобальной инфраструктуры региональных центров для обработки, хранения и анализа данных физических экспериментов LHC. Новейшие технологии GRID являются основой построения этой инфраструктуры.

Проект LCG осуществляется в две фазы.

1 фаза (2001-2005 гг.) - *создание прототипа и разработка проекта системы (LCG TDR).*

2 фаза (2005-2008 гг.) - *создание инфраструктуры LCG, готовой к обработке, хранению и анализу данных на момент начала работы ускорителя в 2008 году.*

WLCG activities

Applications Area
Common projects
Libraries and tools,
data management

Distributed Analysis
Joint project on distributed analysis with the LHC experiments

Middleware Area
Provision of grid middleware – acquisition, development, integration, testing, support

Grid Deployment Area
Establishing and managing the Grid Service - Middleware certification, security, operations. Service Challenges

CERN Fabric Area
Cluster management
Data handling
Cluster technology
Networking (WAN+local)
Computing service at CERN

LHC Computing Grid Project
http://lcg.web.cern.ch/LCG/

Applications | Fabric | Grid Deployment | Grid Technology | ARDA

CERN Home > The LHC Computing Grid Project (LCG)

Find

- All CERN
- IT Department
- LCG

LCG Home
Operations Centre
Project Structure
Calendar
Job Opportunities
Contact Us...
LCG Logos

Operating Committees

Implementation (PEB)
GRID Deployment Board
Architects Forum

High Level Committees

Overview (POB)
Software & Computing Committee (SC2)
Computing Resources Review Board

LHC Computing Grid Project

The world's largest and most powerful particle accelerator, the Large Hadron Collider (LHC), is being constructed at CERN, the European Organization for Nuclear Research, near Geneva on the border between France and Switzerland.

The accelerator will start operation in 2007 and will be used to answer the most fundamental questions of science by some 6,000 people from universities and laboratories all around the world. The computational requirements of the experiments that will use the LHC are enormous: 12-14 PetaBytes of data will be generated each year, the equivalent of more than 20 million CDs. Analysing this will require the equivalent of 70,000 of today's fastest PC processors.

The goal of the LCG project is to meet these unprecedented computing needs by deploying a worldwide computational grid service, integrating the capacity of scientific computing centres spread across Europe, America and Asia into a virtual computing organisation.

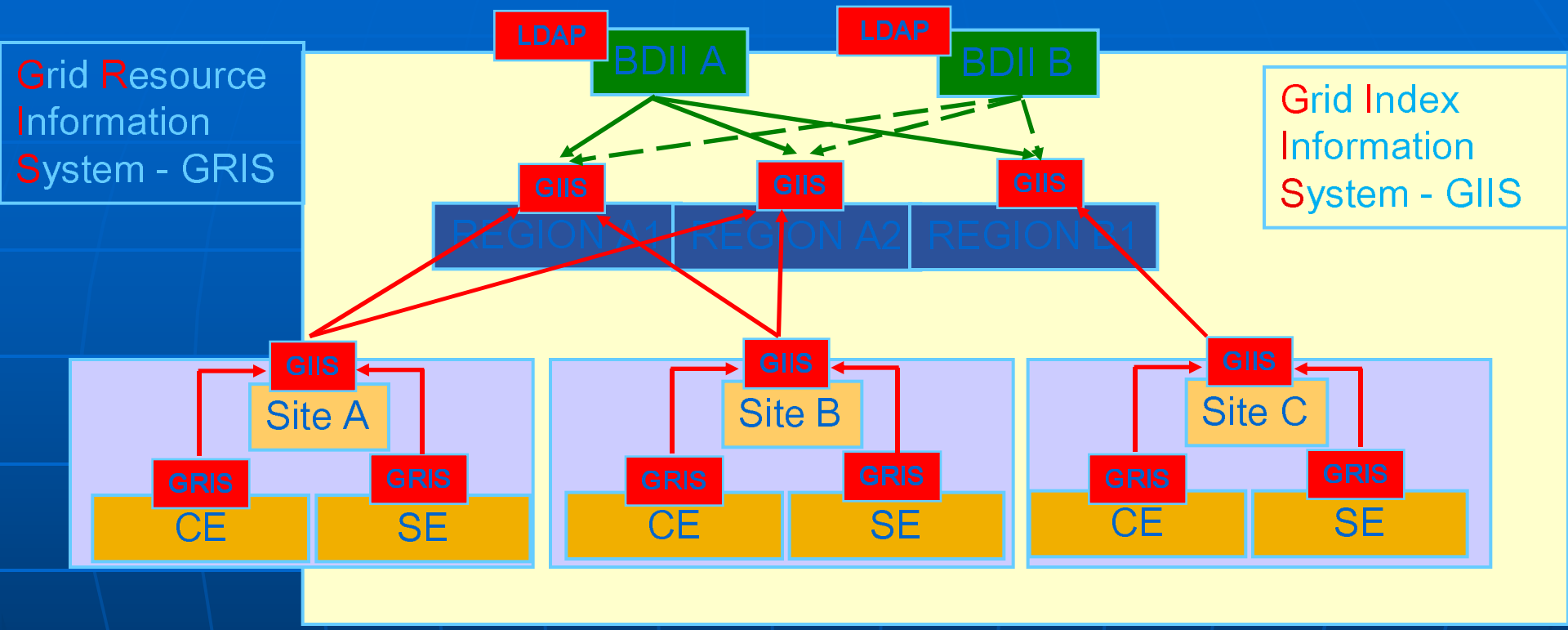
LCG Links

- Project Overview
- Project Planning
- Documents/Presentations
- Meetings
- LCG User's Overview
- Grid Application Group (GAG)
- Technology Tracking
- Regional Centre Resources
- Requirements (RTAGs)

External Links

- LHC Experiments
- Industrial Collaboration
- European Grid Projects
- Other Grid Projects
- In the Press

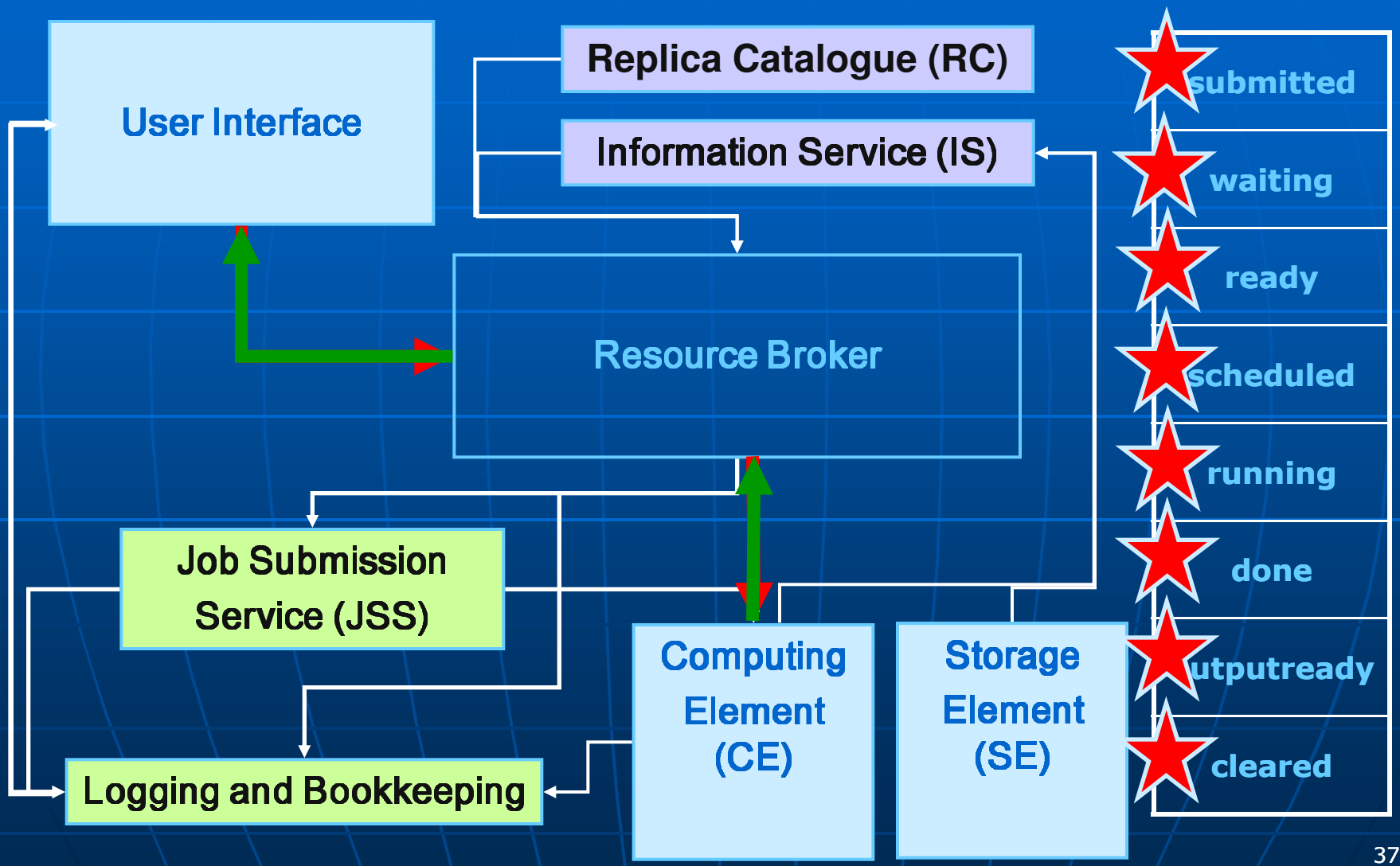
Information service LCG



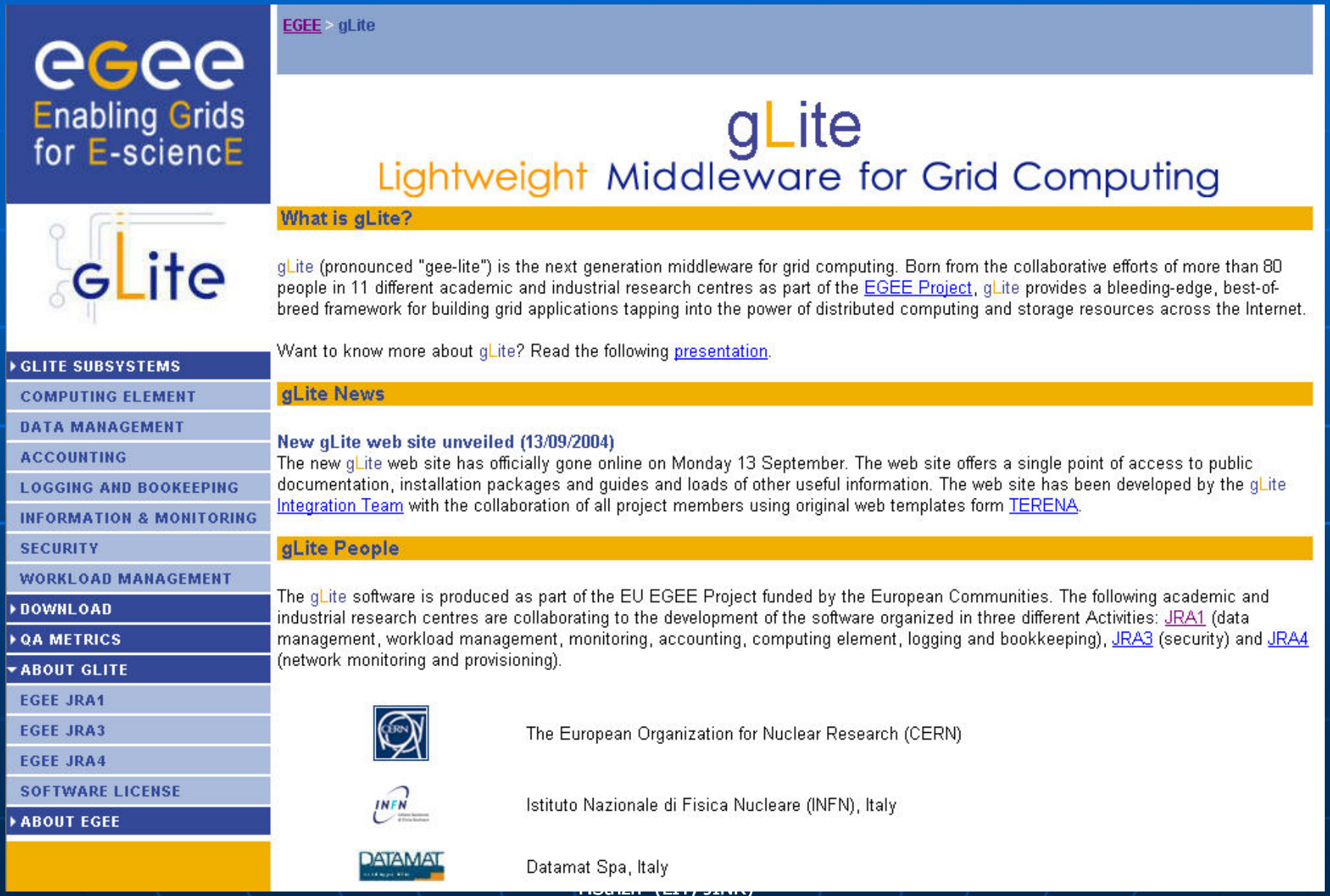
CE,SE → GRIS → GIIS → BDII

GLUE Schema

Job submission to the WLCG



gLite http://www.glite.org



EGEE > gLite

gLite
Lightweight Middleware for Grid Computing

- GLITE SUBSYSTEMS
  - COMPUTING ELEMENT
  - DATA MANAGEMENT
  - ACCOUNTING
  - LOGGING AND BOOKEEPING
  - INFORMATION & MONITORING
  - SECURITY
  - WORKLOAD MANAGEMENT
- DOWNLOAD
- QA METRICS
- ABOUT GLITE
  - EGEE JRA1
  - EGEE JRA3
  - EGEE JRA4
  - SOFTWARE LICENSE
- ABOUT EGEE

What is gLite?

gLite (pronounced "gee-lite") is the next generation middleware for grid computing. Born from the collaborative efforts of more than 80 people in 11 different academic and industrial research centres as part of the EGEE Project, gLite provides a bleeding-edge, best-of-breed framework for building grid applications tapping into the power of distributed computing and storage resources across the Internet.

Want to know more about gLite? Read the following presentation.

gLite News

New gLite web site unveiled (13/09/2004)
The new gLite web site has officially gone online on Monday 13 September. The web site offers a single point of access to public documentation, installation packages and guides and loads of other useful information. The web site has been developed by the gLite Integration Team with the collaboration of all project members using original web templates form TERENA.

gLite People

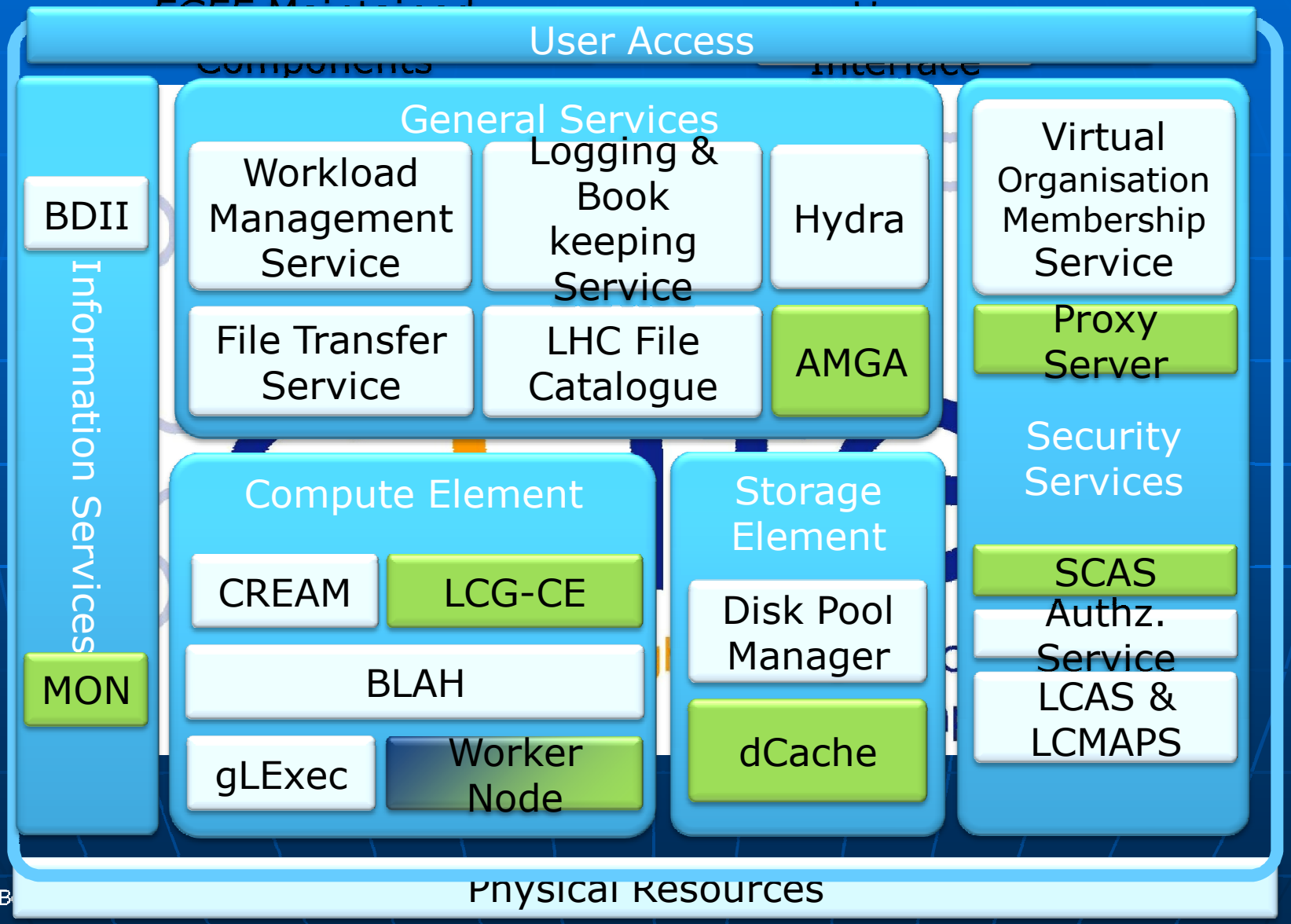
The gLite software is produced as part of the EU EGEE Project funded by the European Communities. The following academic and industrial research centres are collaborating to the development of the software organized in three different Activities: JRA1 (data management, workload management, monitoring, accounting, computing element, logging and bookkeeping), JRA3 (security) and JRA4 (network monitoring and provisioning).

The European Organization for Nuclear Research (CERN)

Istituto Nazionale di Fisica Nucleare (INFN), Italy

Datamat Spa, Italy

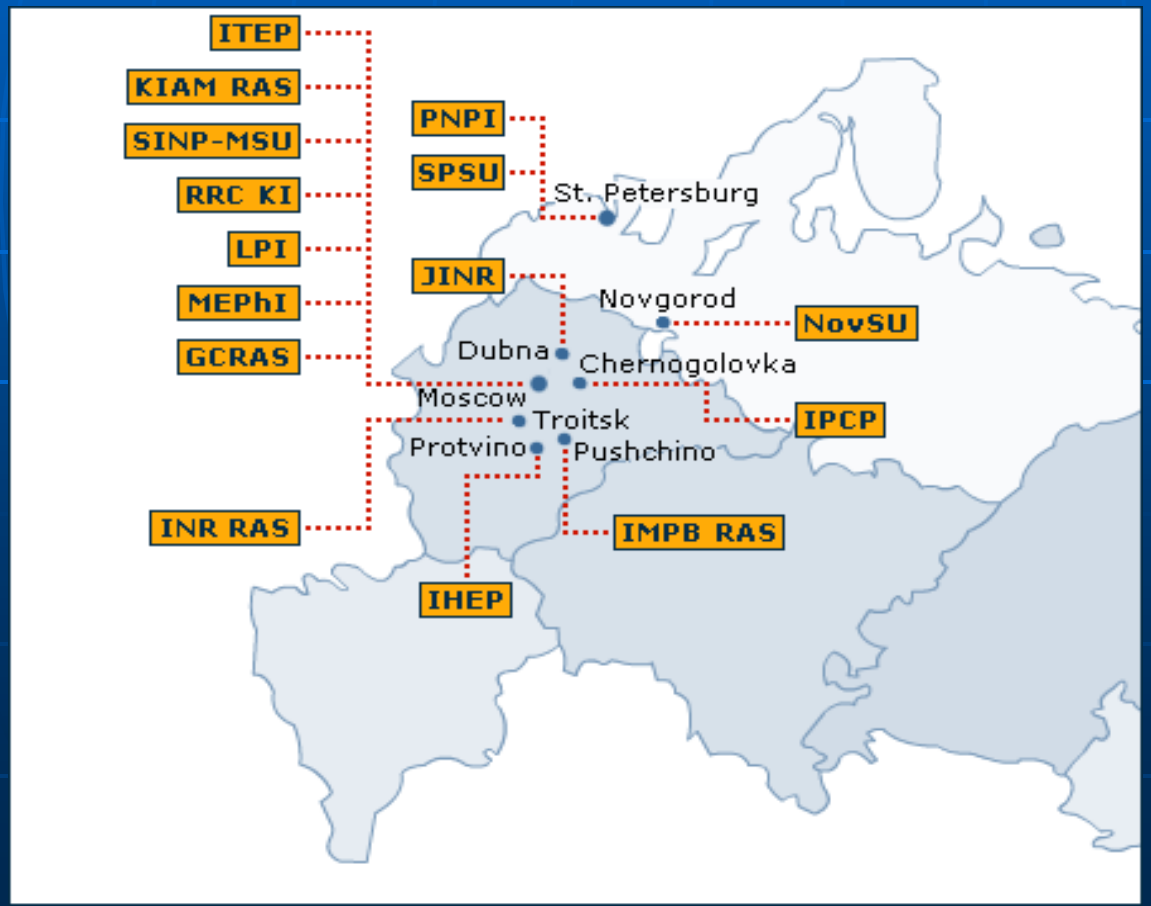
gLite Middleware



RDIG инфраструктура

Российский консорциум RDIG (Russian Data Intensive Grid) был создан в 2003 году как национальная грид-федерация в прокте EGEE.

В настоящее время RDIG –инфраструктура состоит из 15 ресурсных центров, в которых доступно около 5000 процессоров и 2000 ТВ дискового пространства.



RDIG Resource Centres:

- ITEP
- JINR-LCG2
- Kharkov-KIPT
- RRC-KI
- RU-Moscow-KIAM
- RU-Phys-SPbSU
- RU-Protvino-IHEP
- RU-SPbSU
- Ru-Troitsk-INR
- ru-IMPB-LCG2
- ru-Moscow-FIAN
- ru-Moscow-GCRAS
- ru-Moscow-MEPHI
- ru-PNPI-LCG2
- ru-Moscow-SINP

Структурные элементы грид-инфраструктуры в России

ROC – региональный операционный центр в России

- Обеспечение функционирования базовых грид сервисов 24x7
- Мониторинг и реагирование на черезвычайные ситуации
- Учет ресурсов и выполненных работ (accounting)
- Управление и поддержка виртуальных организаций
- Поддержка безопасности инфраструктуры, CA
- Сертификация (валидация) грид ПО
- Поддержка пользователей
- Операционная поддержка ресурсных центров

RC – ресурсные центры *(CPU, Disk, Tape)*

VO – виртуальные организации:

- физика высоких энергий: LHC – ATLAS, ALICE, CMS, LHCb; *PHOTON, HONE …*
- биомед
- Fusion
- *российские VOs: e-Earth (геофизика)*,
- *химфизика (*Черноголовка, МГУ ++*)*,
  нанотехнологии и наноиндустрия

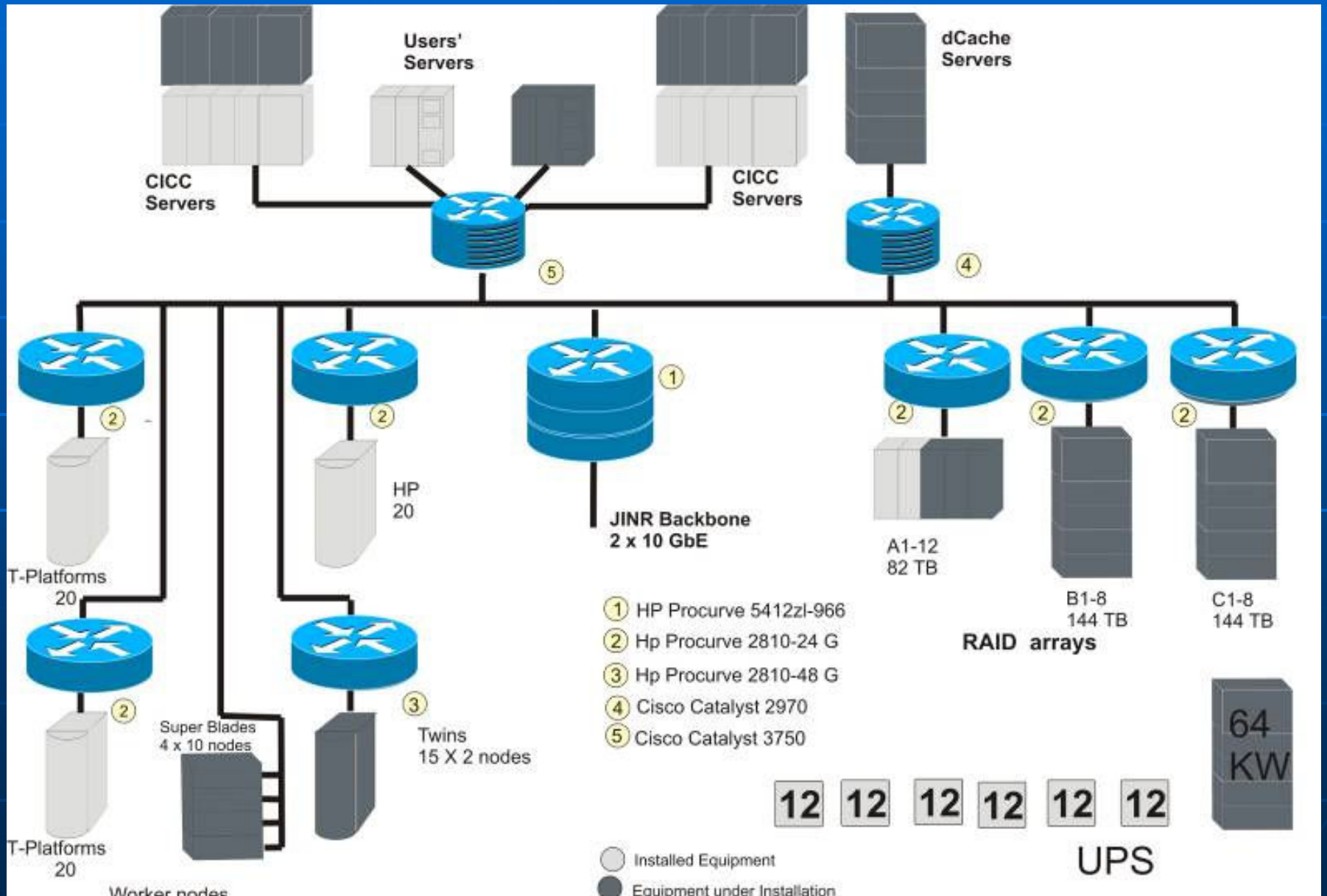
ОИЯИ в проектах WLCG/EGEE

Протокол между ЦЕРН, Россией и ОИЯИ об участии в проекте LCG был подписан в 2003 году. MoU об участии в проеке WorldwideLHC Computing Grid (WLCG) был подписан в 2007 году.

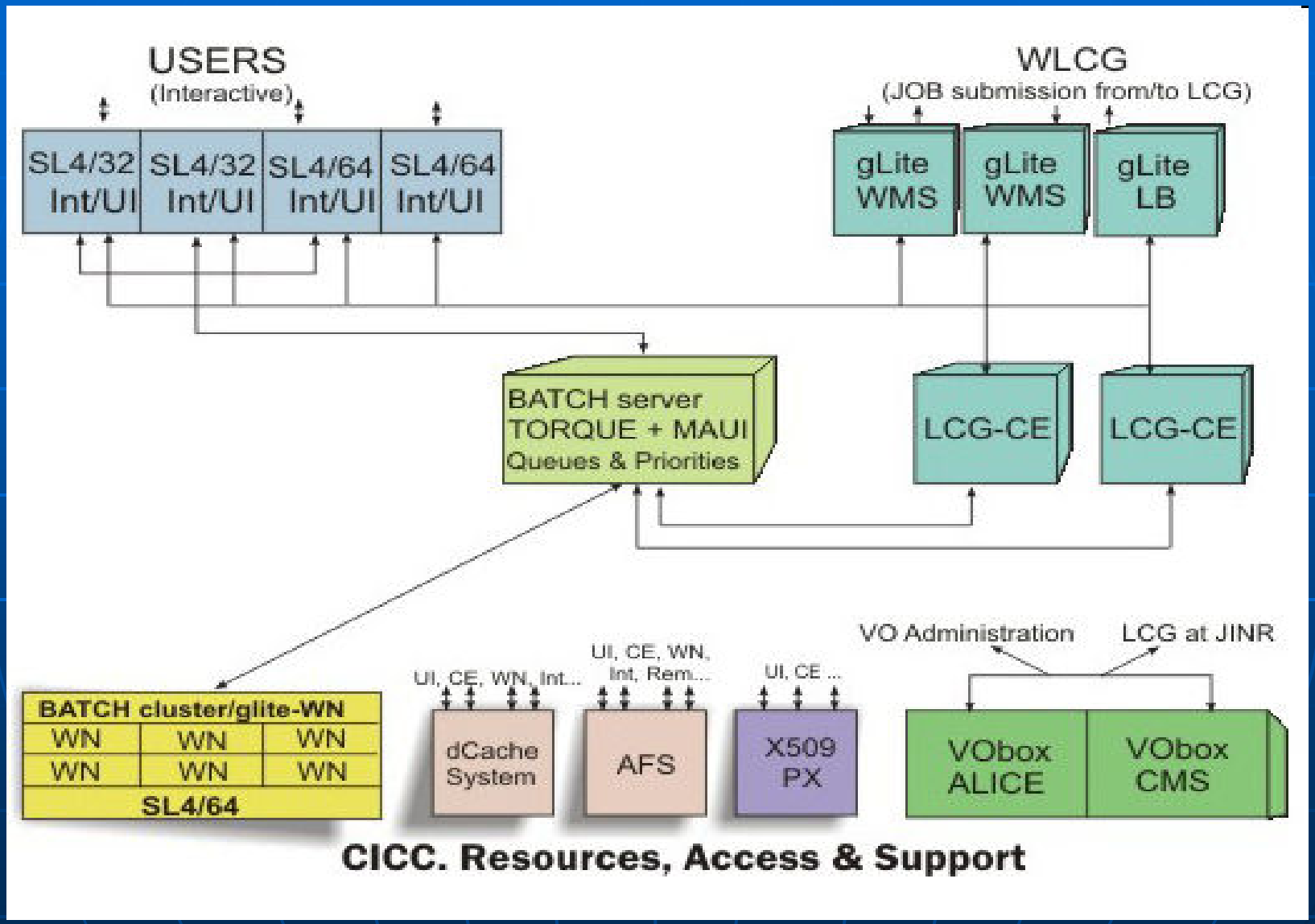
Задачи ОИЯИ в проектах WLCG/EGEE

Поддержка и развитие WLCG/EGEE инфраструктуры;
Участие в тестировании и сертификации нового промежуточного ПО (middleware testing/evaluation),
Участие в Data and Service Challenges,
Развитие систем grid мониторинга и учета ресурсов;
Развитие системы FTS-monitoring;
Развитие информационной системы MCDB;
Участие в проекте ARDA;
Развитие приложений в области физики высоких энергий;
Организация обучения пользователей и системных администраторов
Поддержка стран-участниц ОИЯИ в развитии грид-технологий

Сетевая инфраструктура ЦИВК ОИЯИ



JINR CICC structural scheme



JINR WLCG infrastructure

CICC comprises:

65 servers

4 interactive nodes

960 computing nodes, Xeon 5150, 8GB RAM (GEthernet)

(160 computing nodes, Xeon X5450, 16GB RAM, InfiniBand).

Site name: JINR-LCG2

Internal CICC network – 1Gbit/sec

Operating system -Scientific Linux CERN 4.6 (transfer to Scientific Linux SL release 5.2.) ;

Middleware version GLITE-3.1

File Systems – AFS (the Andrew File System) for user Software and home directories is a world-wide distributed file system. AFS permits to share easily files in an heterogeneous distributed environment (UNIXes, NT) with a unique authentication scheme (Kerberos).

dCache- for data.

User registration system – Kerberos 5 (AFS use Kerberos 5 for authentication)

JINR WLCG infrastructure

JINR provides the following services in the WLCG environment:

Basic services:

- ***Berkley DB Information Index (top level BDII);***
- ***site BDII;***
- ***2 x Computing Element (CE);***
- ***Proxy Server (PX);***
- ***2 x Workload Management System (WMS);***
- ***Logging&Bookkeeping Service (LB);***
- ***RGMA-based monitoring system collector server (MON-box);***
- ***LCG File Catalog (LFC);***
- ***Storage Element (SE), dCache 400 TB, 4 x gridftp door, 14 x pool;***
- ***4 x User Interface (UI), installed in AFS.***

Special Services **- VO boxes for ALICE and for CMS; ROCMON;**
PPS and testing infrastructure **- Pre-production gLite version;**

Software for VOs: dCache xrootd door, AliROOT, ROOT, GEANT packages for ALICE; ATLAS packages; CMSSW packages for CMS and DaVinchi, Gauss packages for LHCb.

RDIG monitoring&accounting http://rocmon.jinr.ru:8080

- **Monitoring – allows to keep an eye on parameters of Grid sites' operation in real time**
- **Accounting - resources utilization on Grid sites by virtual organizations and single users**

Monitored values

CPUs - total /working / down/ free / busy

Jobs - running / waiting

Storage space - used / available

Network - Available bandwidth

Accounting values

Number of submitted jobs

Used CPU time

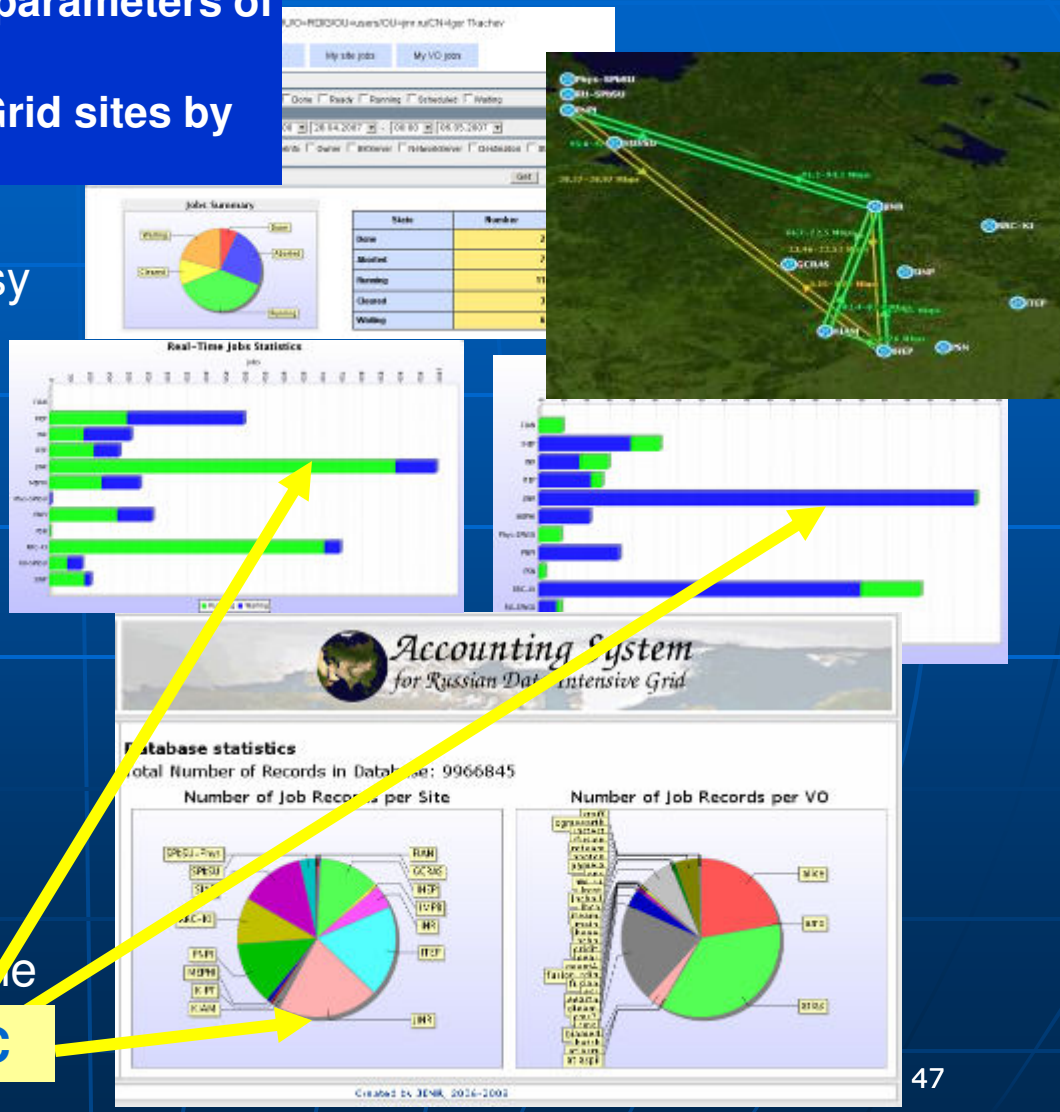
- Totally sum in seconds
- Normalized (with WNs productivity)
- Average time per job

Waiting time

- Totally sum in seconds
- Average ratio waiting/used CPU time per job

Physical memory

- Average per job



JINR CICC

Portal www.egee-rdig.ru

EGEE-RDIG - Microsoft Internet Explorer

Адрес: http://www.egee-rdig.ru

LogIn

Home

rdig
- News
- Meetings/Seminars
- Agenda
- Partners
- Boards
- Documents

services
- Certification Authority
- CIC
- ROC
- Resource Centers
- Mail Lists

activities
- SA1. European GRID Support, Operation & Management
- SA2. Network Resource Provision
- NA2. Dissemination & Outreach
- NA3. User Training & Induction
- NA4. Application Identification & Support

general info

EGEE & RDIG

The RDIG MEMORANDUM

EGEE (En… integrate c… Grids to c… support of … generated … and nation… Trillium, etc… countries, … Germany/Switzerland, Ireland/UK, Italy, North… US), the largest international Grid infrastructur…

The EGEE vision is to provide distributed … computing, offering round-the-clock access … location. The resulting infrastructure will provid… e-Science. EGEE will work to provide interope… Cyberinfrastructure, establishing a worldwide G…

EGEE is a two-year project in a four-year prog… Particle Physics applications (LHC Comput… exclusively on a Grid infrastructure to store an… **Biomedical Grids**, such as the proposed Hea…

Eight Russian Institutes made up the cons… federation in the EGEE project: IHEP, IMPB RAS…

EGEE-RDIG - Microsoft Internet Explorer

Адрес: http://rus.egee-rdig.ru/

Research Centers | Physics | Science | Math | Astronomy | Relativity

Английская Версия | Главная | Сегодня : 14/1/2005

РДИГ
- Новости
- Встречи/Семинары
- Календарь
- Участники
- Правление

services
- Certification Authority
- CIC
- ROC. Региональный Операционный Центр
- Resource Centers

Направление Работ
- SA1. Поддержка, эксплуатация и управление Грид-системами
- SA2. Обеспечение сетевых ресурсов
- NA1. Руководство проектом I3
- NA2. Распространение информации
- NA3. Обучение и включение в число пользователей
- NA4. Идентификация и поддержка приложений

EGEE & RDIG

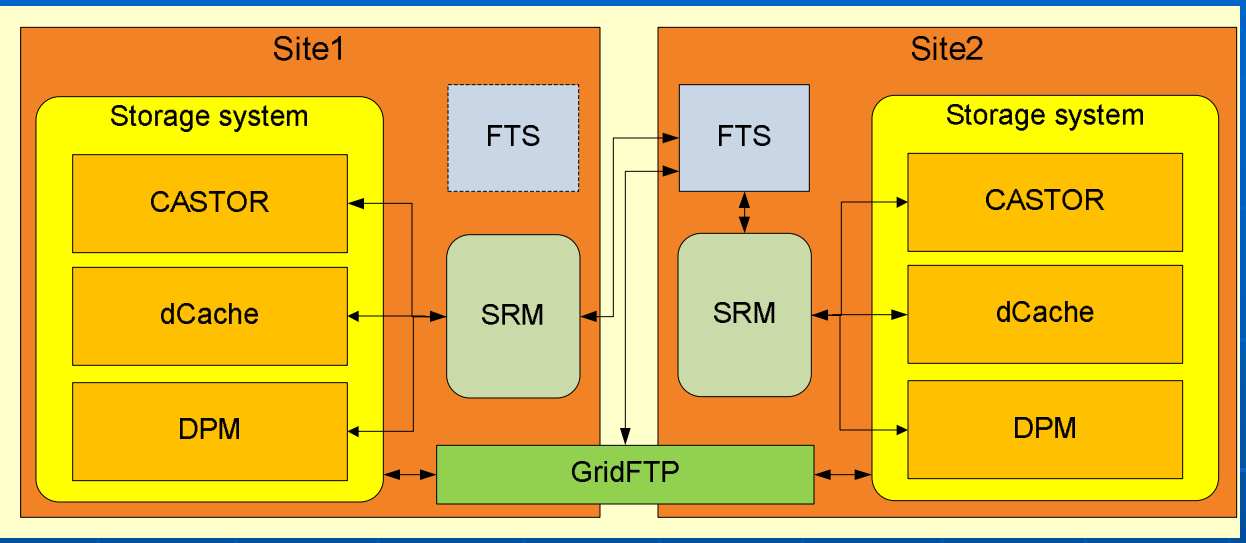
МЕМОРАНДУМ РДИГ

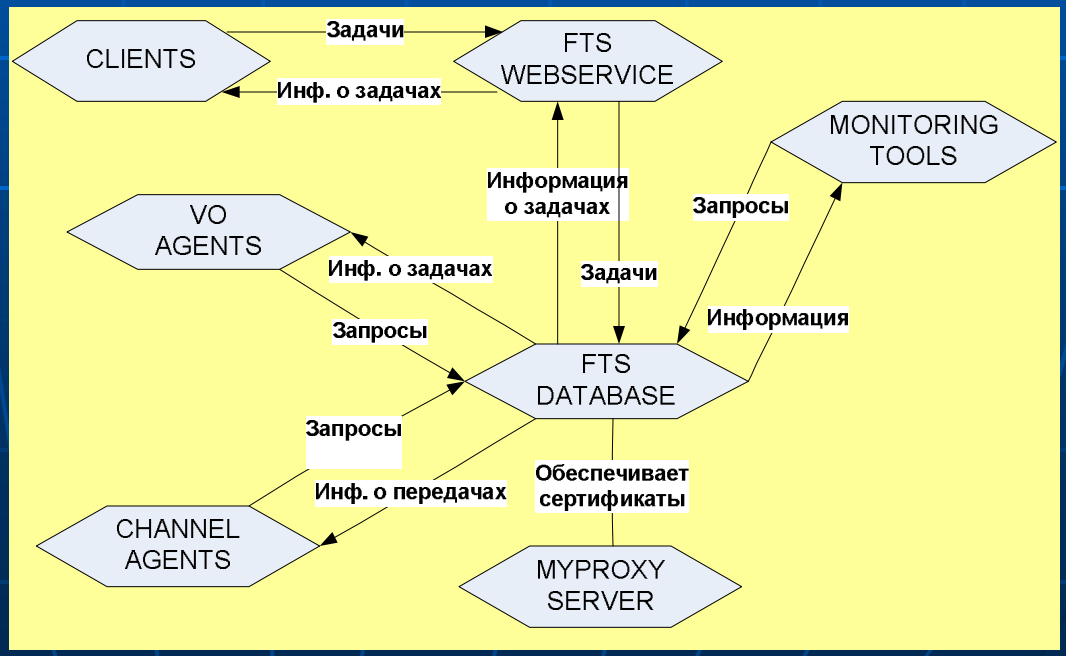
В недалёком будущем Вам как научному работнику станут доступны не имеющие себе равных вычислительные мощности и объёмы информации. Это станет возможным благодаря Грид-инфраструктуре, которая развивается в рамках финансируемого ЕС проекта Enabling Grids for E-science in Europe ("Развёртывание Грид-систем для развития е-науки в Европе"), более известного как EGEE. Настоящая публикация представляет краткий обзор важнейших сторон проекта EGEE, знание которых понадобится Вам как потенциальному пользователю Грид-инфраструктуры.

Что такое EGEE?

Цель проекта EGEE - объединить уже ведущиеся национальные, региональные и тематические Грид-разработки в единую цельную Грид-инфраструктуру для поддержки научных исследований. EGEE предоставит исследователям как в академических кругах, так и в разных областях экономики круглосуточный доступ к самым высокопроизводительным вычислительным ресурсам независимо от их географического положения. Пользоваться инфраструктурой смогут географически распределённые сообщества исследователей, которые нуждаются в общих для них вычислительных возможностях Грид-систем, готовы объединить свои собственные вычислительные инфраструктуры и согласны с принципами общего доступа. Проект поддерживают, в основном, финансирующие учреждения ЕС, но предназначен он для работы во всём мире. Значительные средства поступают от США, России и других участников проекта, не входящих в ЕС.

File Transfere System Monitoring and Testing

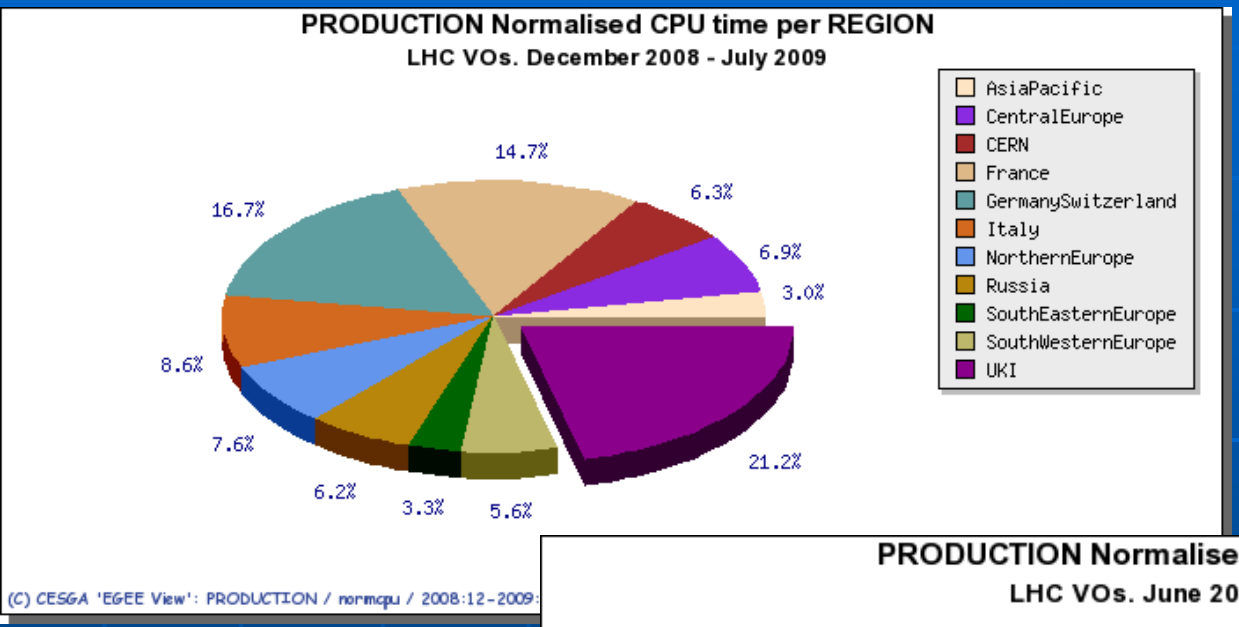


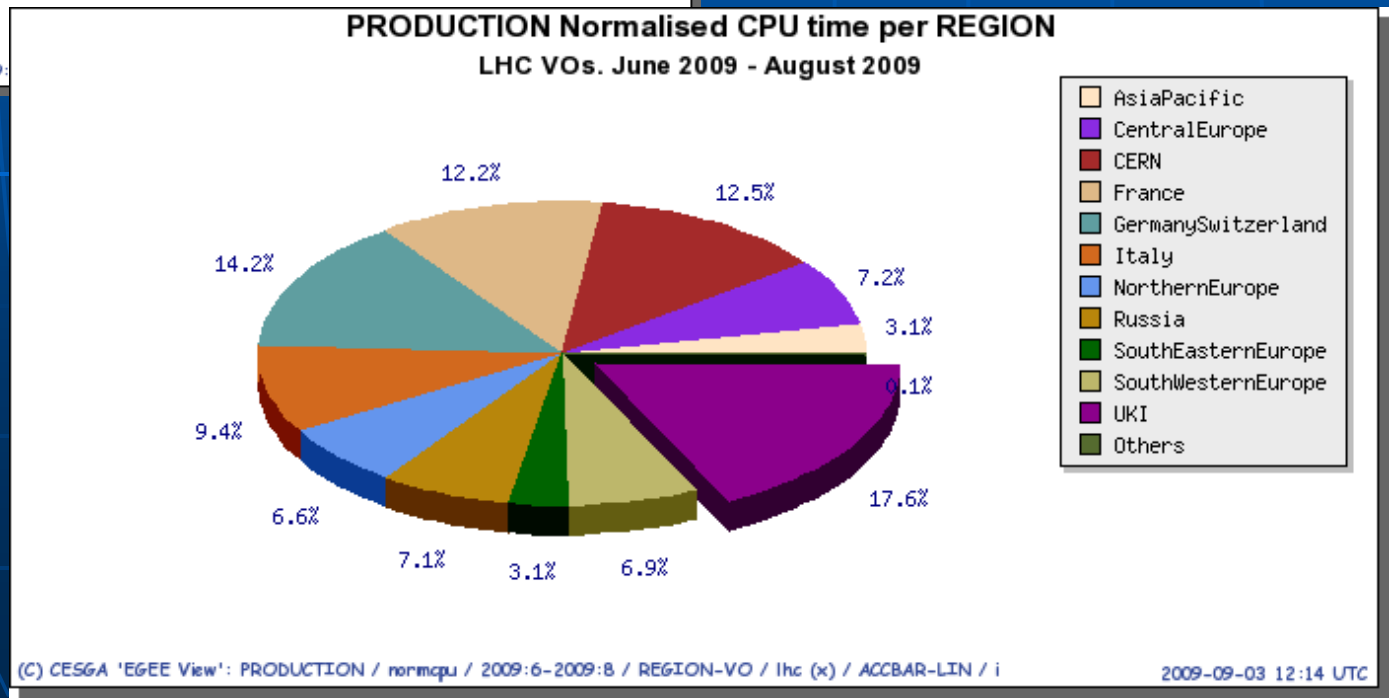


Integration with Google Earth

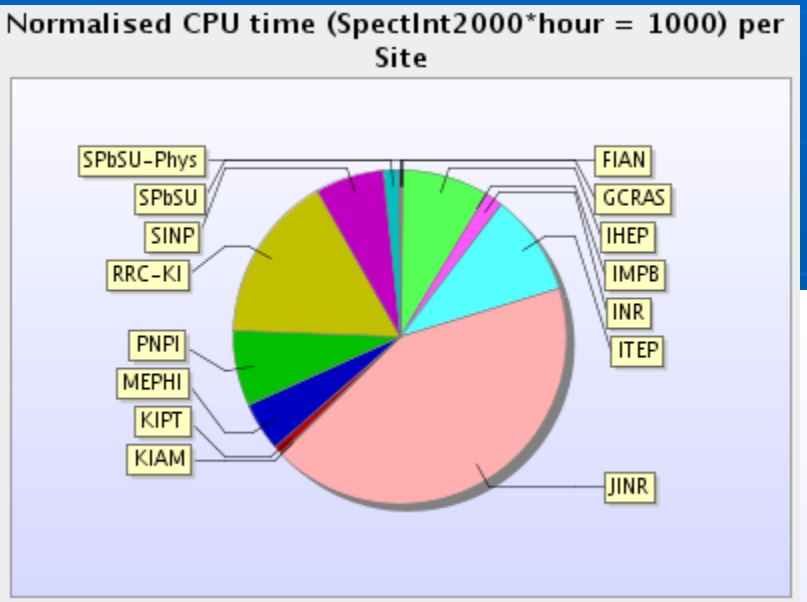
Experiment specific monitoring systems provide input data
Dashboard agents publish this information in the KML format
Strong contribution to the development of Sergey Mitsyn (JINR)
Application will be shown during the LHC demo at the EGEE conference in Barcelona

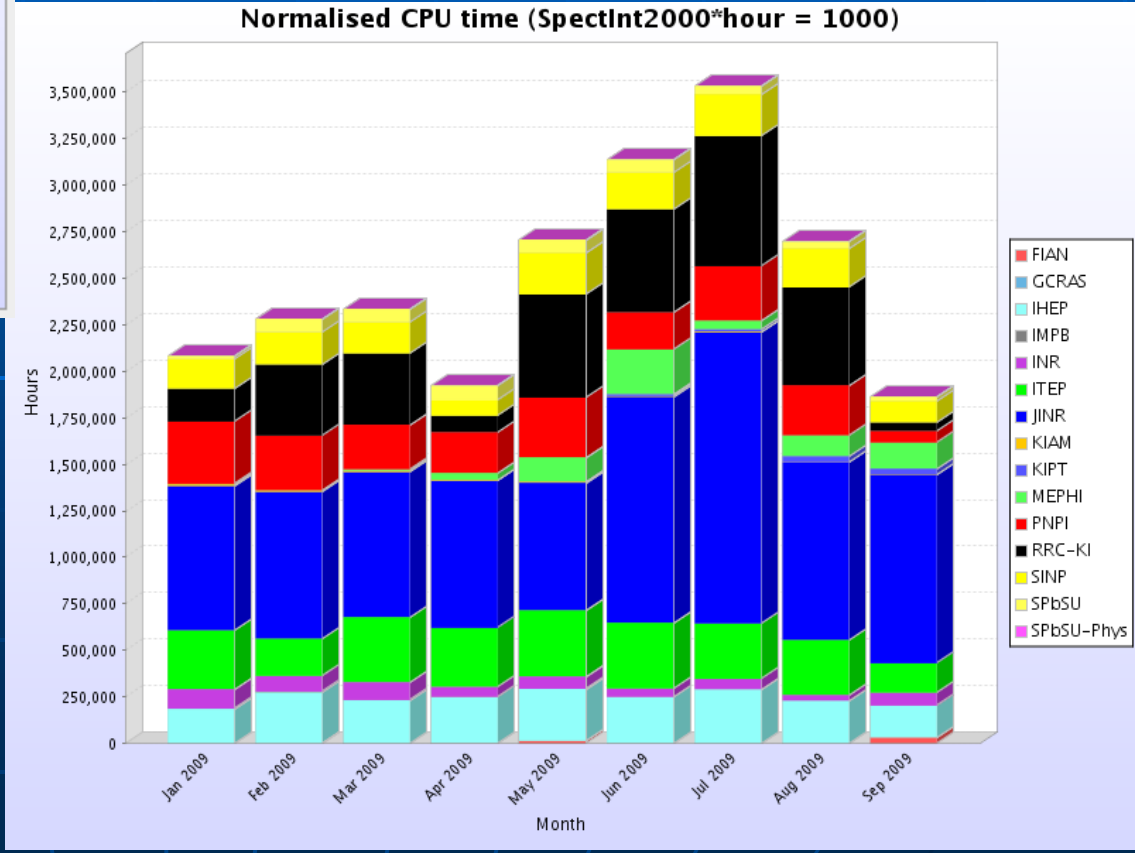
Production Normalised CPU time per EGEE Region (June-August 2009)





Russia and JINR Normalized CPU time per SITE (January 2009 - September 2009)





Production Normalised CPU time per EGEE site for VO LHC (June – September 2009)

GRID-site	CPU time	Num CPU
FZK-LCG	8,095,787	8620
CERN-PROD	4,552,891	6812
INFN-T1	4,334,940	2862
GRIF	4,089,269	3454
JINR	3,957,790	960
CYFRONET-LCG	3,948,857	2384
PIC	3,921,569	1337
UKI-GLASGOW	3,860,298	1912
RAL-LCG2	3,793,504	2532
UKI-LT2-IC-HEP	3,752,747	960
IN2P3-CC	3,630,425	4544

Проекты в области развития grid-технологий

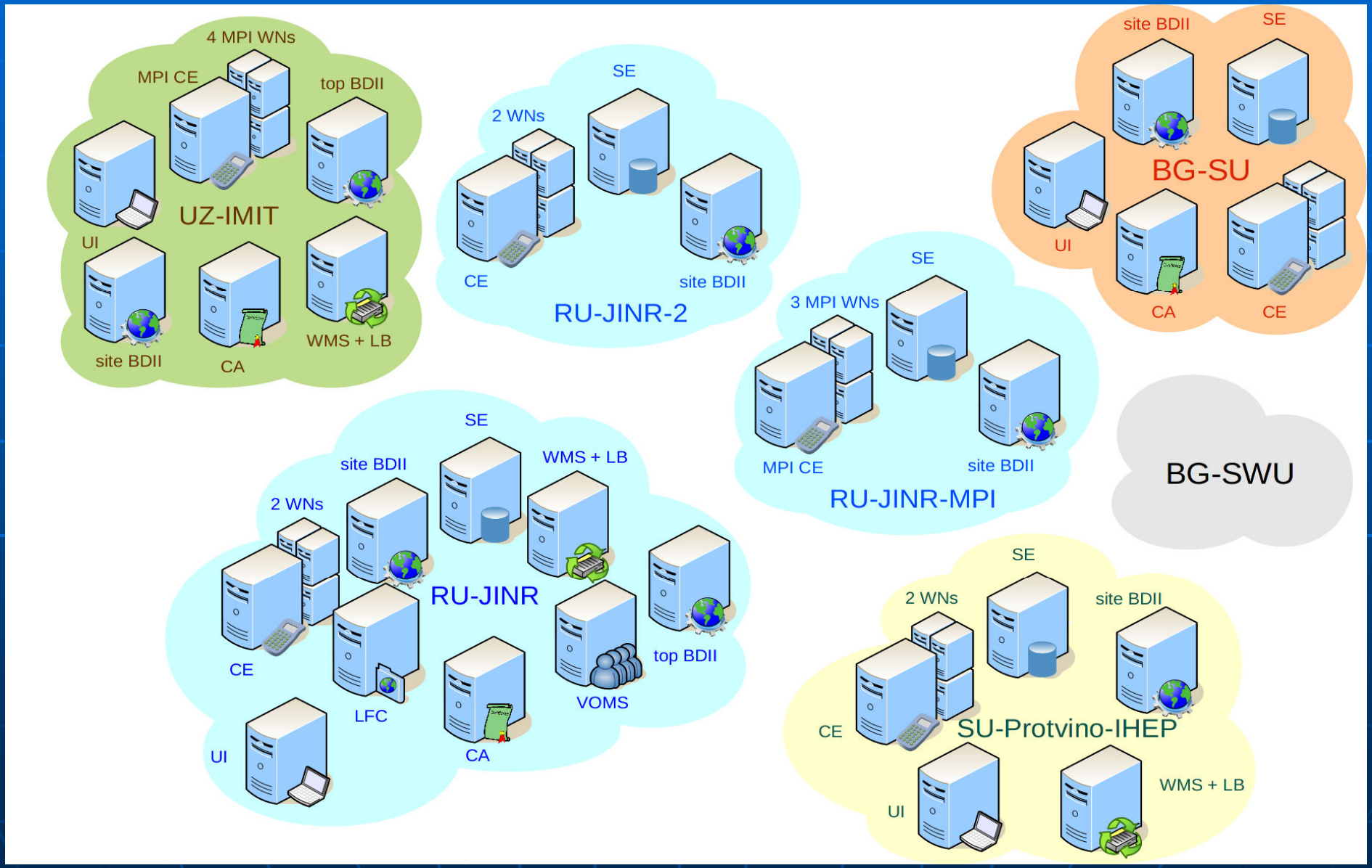
- WLCG: проект грид для Большого Адронного Коллайдера (Worldwide LHC Computing GRID)
- EGEE: развертывание грид-систем для е-науки (The Enabling Grids for E-sciencE)
- RDIG: Российский грид для интенсивных операций с данными
- Проект СКИФ-ГРИД: программа развития высокопроизводительных вычислений союзного государства Россия - Белорусь
- Подготовка вычислительных комплексов ЛИТ ОИЯИ (г. Дубна) и ННЦ ХФТИ (г. Харьков) к распределенному анализу данных эксперимента CMS (CERN) на основе грид-технологий
- Грид мониторинг для эффективного функционирования виртуальных организаций (совместный с ЦЕРН)
- Развитие Грид –ифраструктуры и ПО для обеспечения совместных исследований с участием ОИЯИ и центров Германии (финансирование BMBF)
- Развитие грид-сегмента для LHC экспериментов в кооперации с ЮАР (финансирование ЮАР);
- NATO проект "DREAMS-ASIA“ (Развитие грид-технологий для науки и медицины в Центральной Азии)
- ОИЯИ-Румыния кооперация в программе Хулубей-Мещеряков
- ОИЯИ-Институт физики АН Чехии проект «Грид-инфраструктура для физических экспериментов»
- «Развитие распределенной грид-инфраструктуры ОИЯИ-Армения для научных исследований»
- ГридННС- инфраструктура региональных центров программы развития нанотехнологий и наноиндустрии
- Дубна-Грид: создание городской грид-инфраструктуры
- Совместные проекты по развитию грид-технологий с научными центрами Армении, Белоруссии, Болгарии, Грузии, Польши, Румынии,Словакии, Узбекистана, Украины, ФРГ, Чехии, ЮАР

User Training and Induction

Russian and JINR physicists participants of ATLAS experiment train and practise with Grid and the GANGA

COURSES **LECTURES** **PRACTICAL TRAINING**

Distributed training infrastructure



RDMS CMS Data Bases

RDMS CMS Data Management system

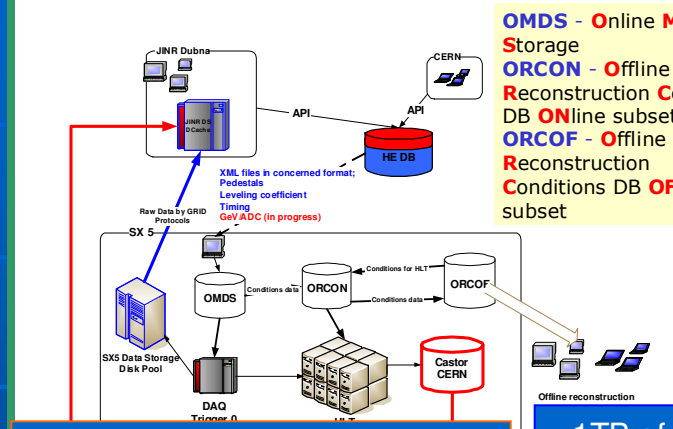
Data Bases

- **Equipment Data Bases**
  ME 1/1
  HE
- **Condition Data Base**
  Calibration data base HE

Services

- API to databases
- WEB Interfaces
- **Data storage and data management**
- Export data and integration with other databases

HE Calibration and Raw Data Flow



OMDS - **O**nline **M**aster **D**ata **S**torage
ORCON - **O**ffline **R**econstruction **C**onditions DB **ON**line subset
ORCOF - **O**ffline **R**econstruction **C**onditions DB **OF**fline subset

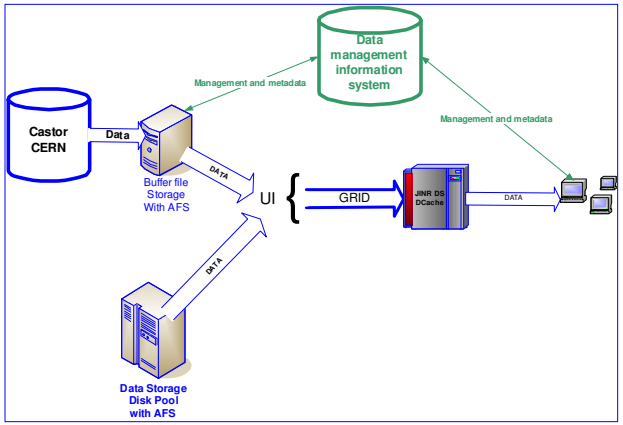
Raw data to JINR by GRID protocols

~1TB of raw data already transferred

HE Calibration DB Status

- System is online
- Full calibration cycle support
- Integrated into CMS computing environment
- ~30000 records
- ~500Mb
- ~600Gb raw data transferred to JINR

CERN - JINR Data management for MTCC: common schema of realization



Удалённый мониторинг

- Коллаборация АТЛАС:
  - 164 институтов из 35 стран
- Для эффективного участия в коллаборации пользователи должны иметь возможность:
  - Отслеживать общий статус эксперимента
  - Отслеживать состояние подсистем, разработанных в их институтах
- Эти задачи решаются с помощью системы удалённого мониторинга

Remote ATLAS Control Room in Dubna

MOTIVATION

- Monitoring of the detector at any time
- Participation of the subsystem experts from Dubna in the shifts and data quality checks remotely
- Training the shifters before they come to CERN

ACR at CERN

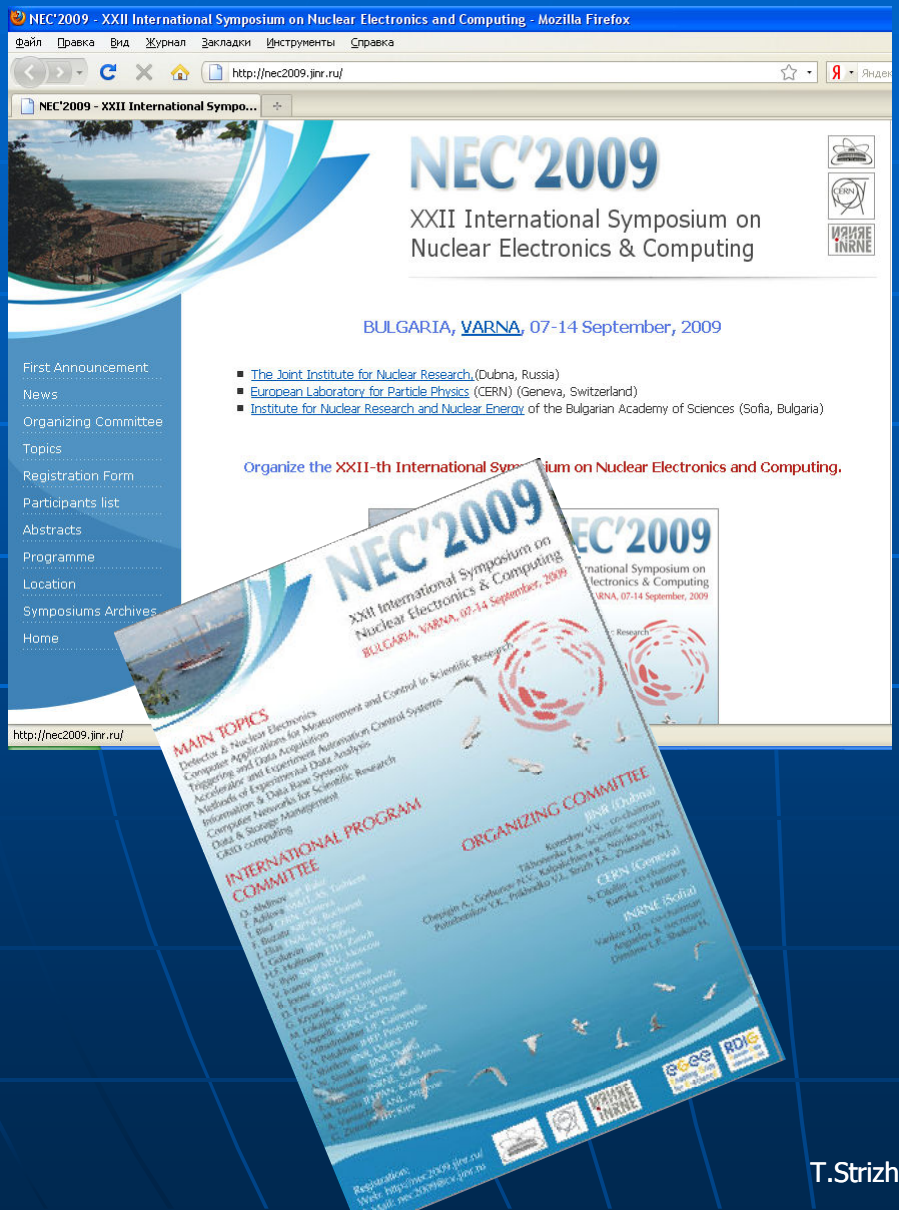
The goal - to have reduced copy at JINR

3-rd International Conference "Distributed Computing and Grid-technologies in Science and Education“ 30 June – 4 July, 2008

126 reports, 211 participants

http://grid2008.jinr.ru

The XXII International Symposium on Nuclear Electronics and Computing (NEC'2009) Bulgaria, Varna, 7-14 September, 2009.



The main topics of the symposium are:

- Detector & Nuclear Electronics
- Computer Applications for Measurement and Control in Scientific Research
- Triggering and Data Acquisition
- Accelerator and Experiment Automation Control Systems
- Methods of Experimental Data Analysis
- Information & Data Base Systems
- Computer Networks for Scientific Research
- Data & Storage Management
- Grid computing

http://nec2009.jinr.ru

Development of the JINR Grid-environment – 2010-2016

Network level:

links between Moscow and Dubna on the basis of state-of-the-art technologies DWDM and 10Gb Ethernet.

JINR Local area network :

JINR High-speed backbone construction – 10Gbps

Resource level:

requirements of the LHC experiments stimulate the development of a global Grid-infrastructure, together with the resource centers of all the cooperating organizations. First of all, this is of primary concern for such large research centers as the JINR. To reach effective processing and analysis of the experimental data, further increase in the JINR CICC performance and disk space is needed.

	2010-2011	2012-2013	2014-2015	2016
CPU (kSI2k)	3500	5000	8000	12000
Disk systems (TB)	1500	2500	4000	8000
Mass storage (TB)	1000	2000	5000	10000

Useful References:

- Grid Café: http://www.gridcafe.org/
- OPEN GRID FORUM: http://www.ogf.org
- GLOBUS: http://www.globus.org
- TERAGRID: http://www.teragrid.org
- Open Science Grid: http://opensciencegrid.org/
- LCG: http://lcg.web.cern.ch/LCG/
- EGEE: http://www.eu-egee.org
- EGEE-RDIG: http://www.egee-rdig.ru
- EGI: http://web.eu-egi.eu/
- International Science Grid this Week: http://www.isgtw.org/

The blind men and the elephant in the room