

JOINT INSTITUTE FOR NUCLEAR RESEARCH
ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ



Проект WLCG, RDIG, МИВК ОИЯИ, RDIG-M

Кореньков Владимир Васильевич

**Научный руководитель
Лаборатории информационных технологий
имени М.Г. Мещерякова ОИЯИ**

**Совещание консорциума RDIG-M
ОИЯИ, ЛИТ
12 апреля 2024 года**

Грид технологии – путь к успеху



На торжестве по поводу получения Нобелевской премии за открытие бозона Хиггса директор ЦЕРНа Рольф Хойер прямо назвал **грид-технологии одним из трех столпов успеха** (наряду с ускорителем LHC и физическими установками).

Без организации грид-инфраструктуры на LHC было бы невозможно обрабатывать и хранить колоссальный объем данных, поступающих с коллайдера, а значит, совершать научные открытия.

Сегодня уже ни один крупный проект не осуществим без использования распределенной инфраструктуры для обработки данных.



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»

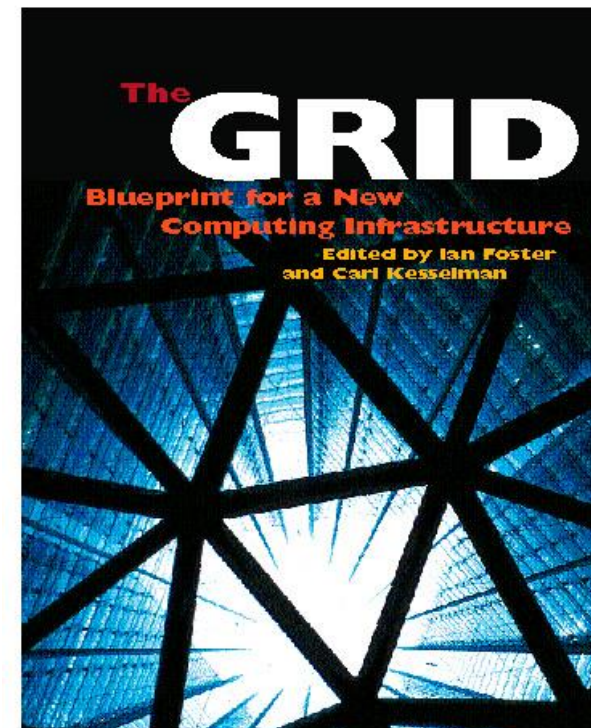
(Ian Foster: "What is the grid? ", 2002 г.)

Модели грид:

- ❖ Distributed Computing
- ❖ High-Throughput Computing
- ❖ On-Demand Computing
- ❖ Data-Intensive Computing
- ❖ Collaborative Computing

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

Виртуальные организации (VO)

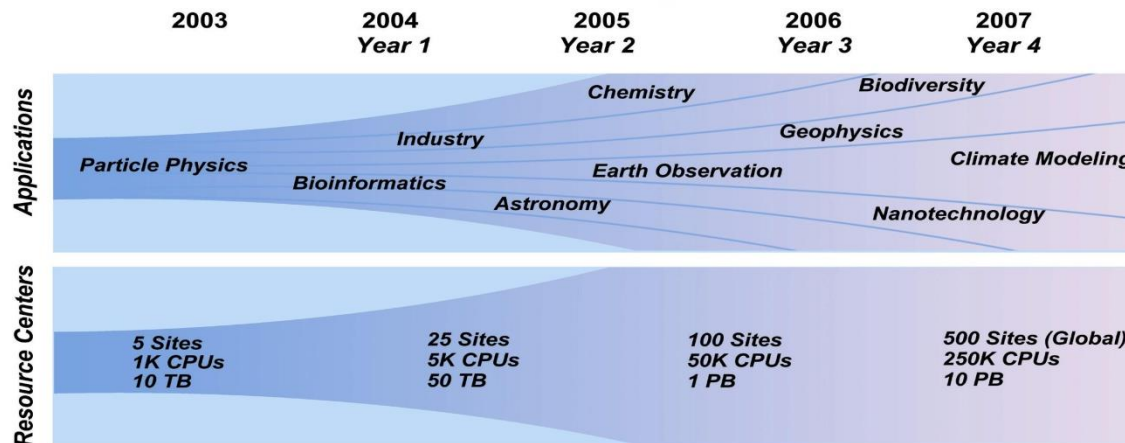
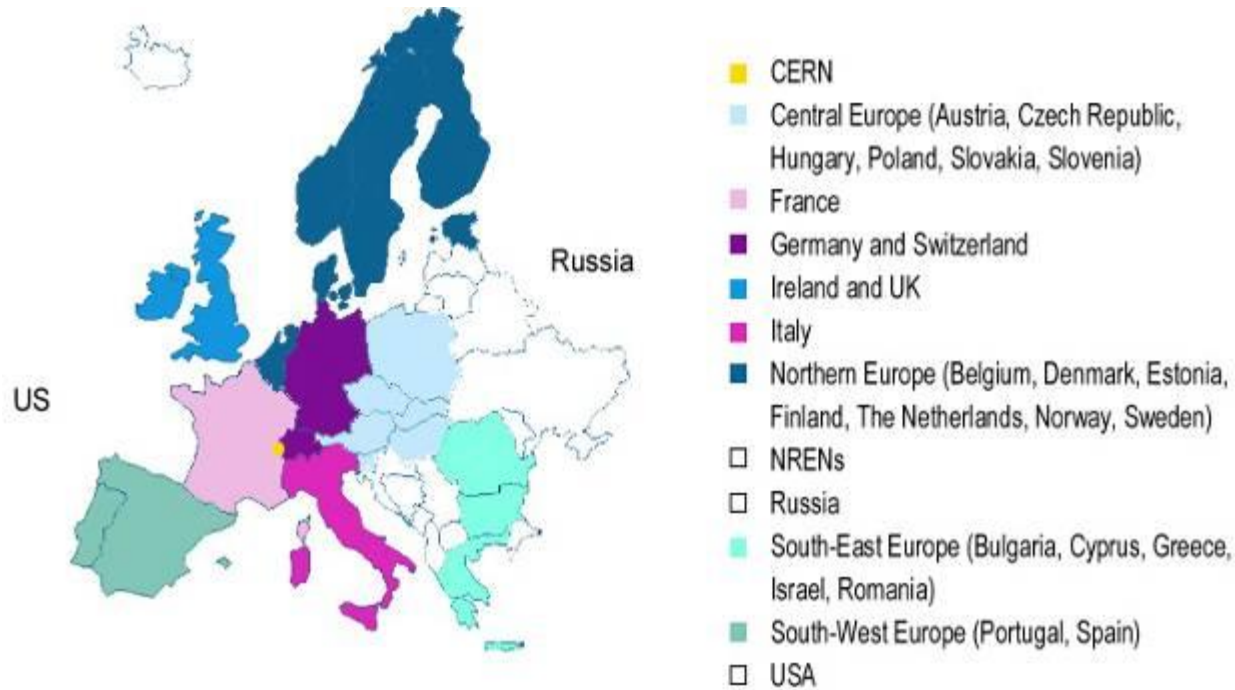


Some history

- **1999 – Monarc Project**
 - Early discussions on how to organise distributed computing for LHC
- **2001-2003 - EU DataGrid project**
 - middleware & testbed for an operational grid
- **2003 - RDIG**
- **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 - Worldwide LHC Computing Grid (WLCG)**
- **2006-2008 – EU EGEE-II**
 - Building on phase 1
 - Expanding applications and communities ...
- **2008-2010 – EU EGEE-III**
- **2010- EGI**
- **Tier1 in Russia (NRC KI, JINR)**



EGEE (Enabling Grids for E-science)

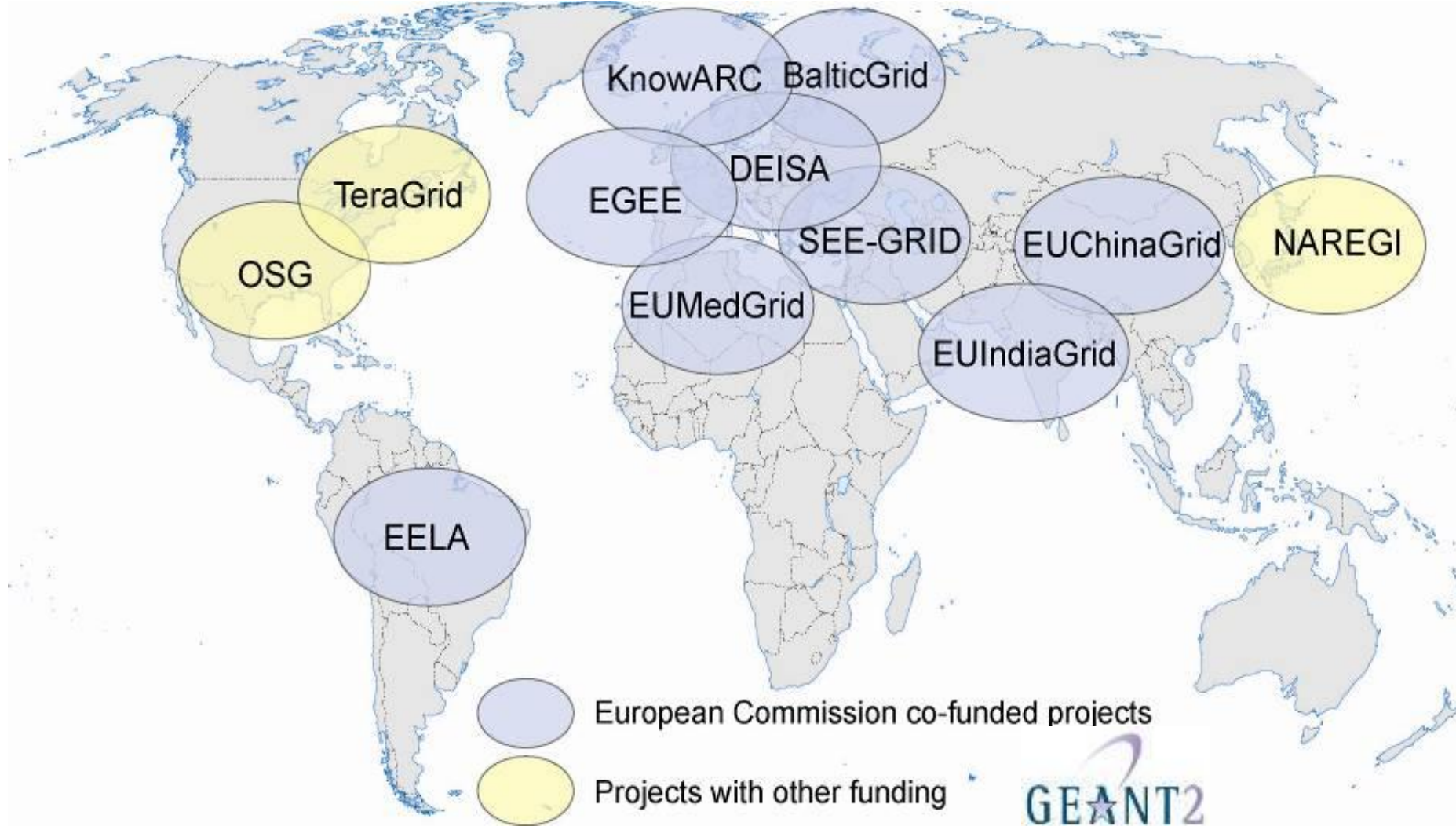


The aim of the project is to create a global Pan-European computing infrastructure of a Grid type.

- Integrate regional Grid efforts
- Represent leading grid activities in Europe

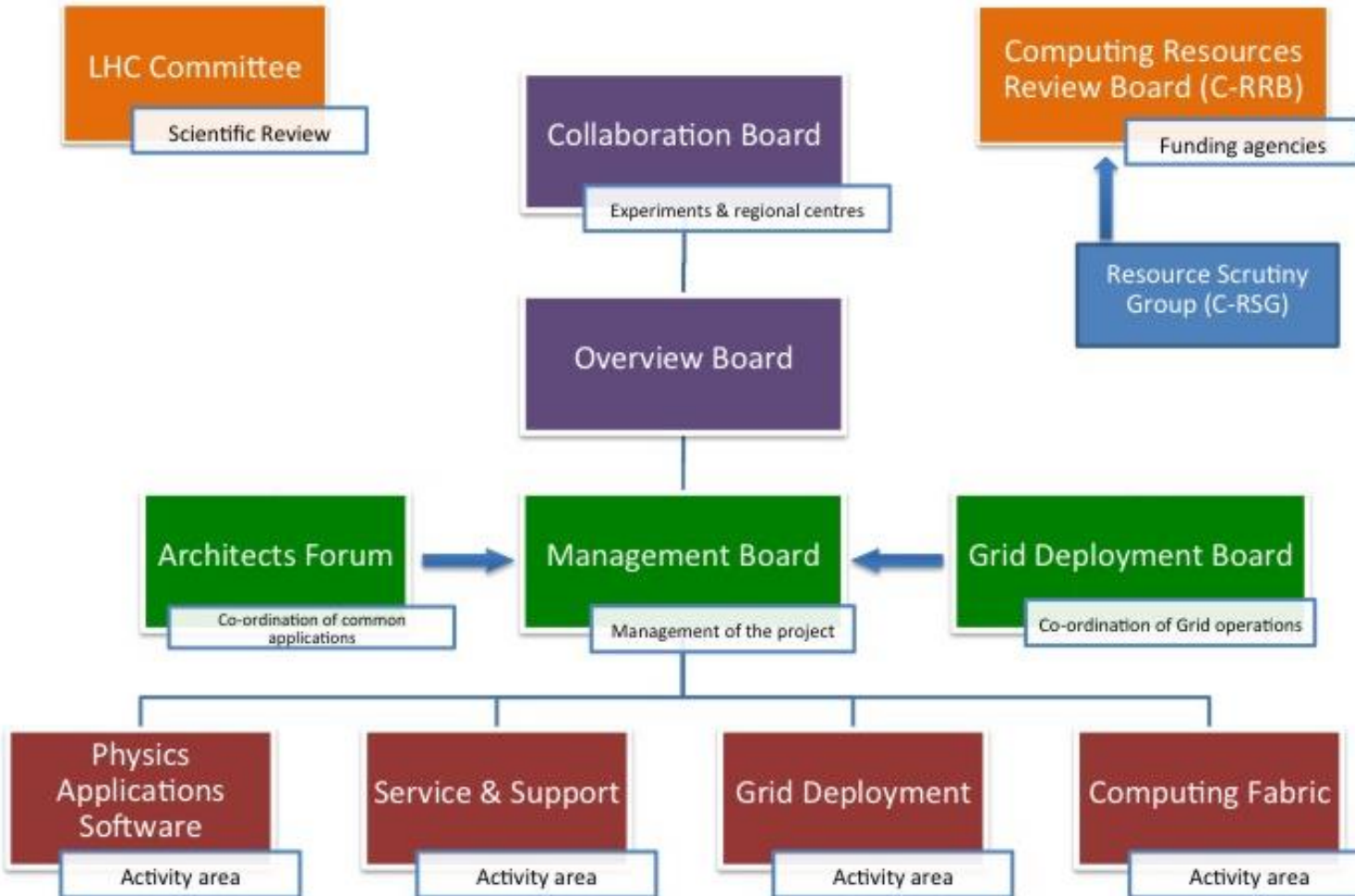
**10 Federations,
27 Countries,
70 Organizations**

Collaborating e-Infrastructures



Potential for linking ~90 countries by 2010

Worldwide LHC Computing Grid - Organisation



Collaboration Board (CB)

The Collaboration Board (CB) provides the main technical direction for LCG.

Overview Board (OB)

A standing committee of the CB, the Overview Board (OB), has the role of overseeing the functioning of the Collaboration. It also acts as a clearing-house for conflicts that may arise within the Collaboration.

Management Board (MB)

The Management Board (MB) supervises the work of the Project, maintaining the overall program

Grid Deployment Board (GDB)

The Grid Deployment Board (GDB) is the forum within the Project where the computing managements of the experiments and the regional computing centres discuss and take, or prepare, the decisions necessary for planning, deploying, and operating the LHC Computing Grid.

Architects Forum (AF)

An Architects Forum (AF) consisting of the Applications Area Manager (chair), the software architects of the four LHC experiments, the leaders of the various AA software projects and other invited members provides for the formal participation of the experiments in the planning, decision-making and architectural and technical direction of applications area activities.

Computing Resources Review Board (CRRB)

The Resources Review Board (RRB) comprises the representatives of each Experiment's Funding Agencies and the managements of CERN and of each Experiment's Collaboration. It is chaired by the CERN Director for Research and Computing.

Computing Resources Scrutiny Group (C-RSG)

The purpose of the C-RSG is to inform the decisions of the Computing Resources Review Board (C-RRB) for the LHC experiments.

Worldwide LHC Computing Grid Project (WLCG)

The primary goal of the WLCG project is to create a global infrastructure of regional centers for processing, storage and analysis of data of the LHC physical experiments. The grid-technologies are a basis for constructing this infrastructure.

MoU with 42 countries about participation in the WLCG project.

A protocol between CERN, Russia and JINR on participation in the LCG project was signed in 2003. MoU about participation in the WLCG project was signed in **2007**.

Russia	FASI	Tue 03 Jul 2007	Sergei Mazurenko
Russia	JINR	Tue 04 Sep 2007	A. N. Sissakian

Boards: GDB, MB OB C-RRB Russia-JINR V. Velikhov V.Korenkov

Services WLCG and EGI: Certification Authorities (CA), GGUS, VOMS, VO, M&A, NOC ...

Worldwide LHC Computing Grid (WLCG)

- The mission of the [Worldwide LHC Computing Grid \(WLCG\)](#) is to provide global computing resources for the storage, distribution and analysis of the data generated by the LHC.
- WLCG combines about 1.5 million computer cores and 1.8 exabytes of storage from over 170 sites in 42 countries. This massive distributed computing infrastructure provides more than 12 000 physicists around the world with near real-time access to LHC data, and the power to process it.
- It runs over 2 million tasks per day and, at the end of the LHC's LS2, global transfer rates exceeded 260 GB/s.
- These numbers will increase as time goes on and as computing resources and new technologies become ever more available across the world.
- CERN provides about 20% of the resources of WLCG.
- Simone Campana Project leader / chairperson WLCG

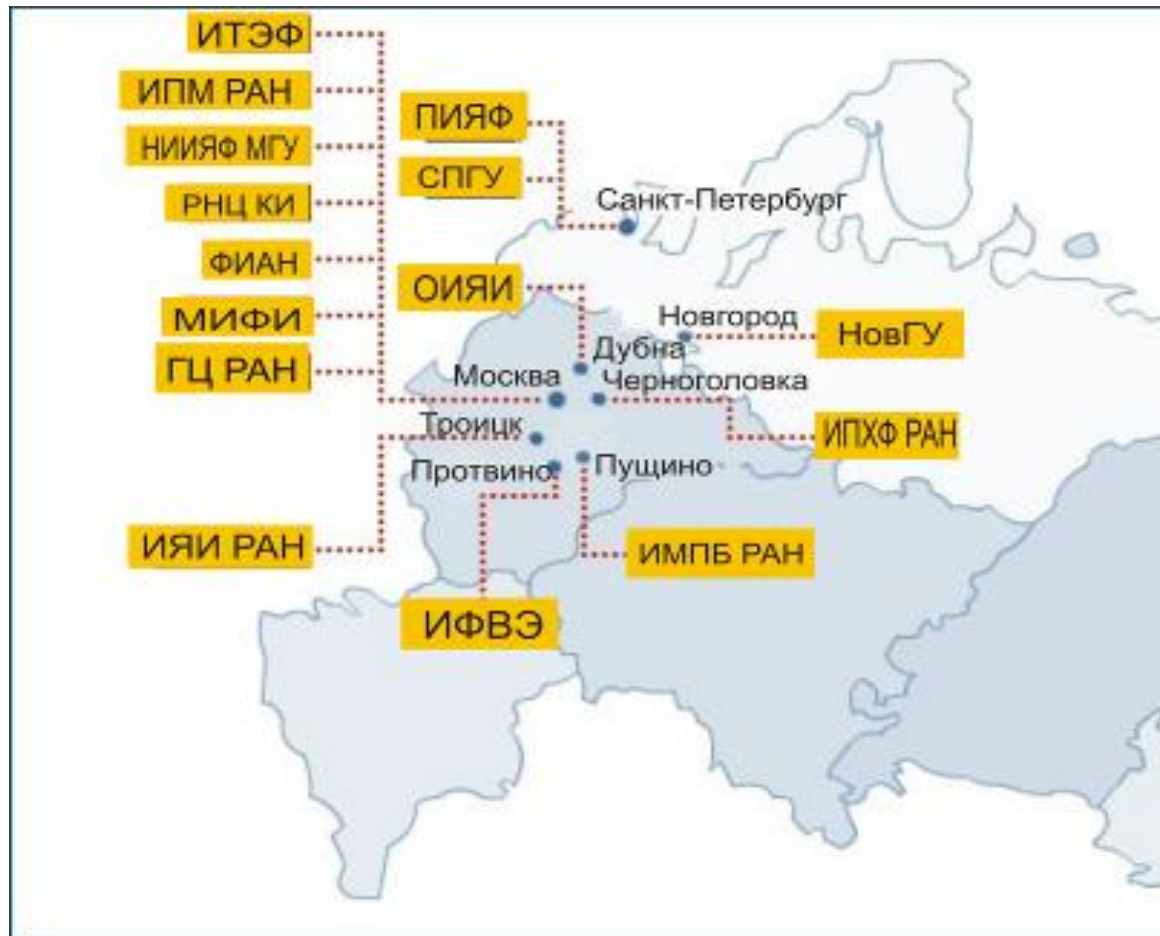
- WLCG project supports LHC experiments (ATLAS, CMS, ALICE, LHCb) and others:
 - Heidi Schellman DUNE
 - Giuseppe Andronico JUNO
 - Ikuo Ueda (deputy: Cedric Serfon) Belle-II
 - Stefano Bagnasco VIRGO



Russian Data Intensive Grid infrastructure (RDIG)

The Russian consortium RDIG (Russian Data Intensive Grid), was set up in September 2003 as a national federation in the EGEE project.

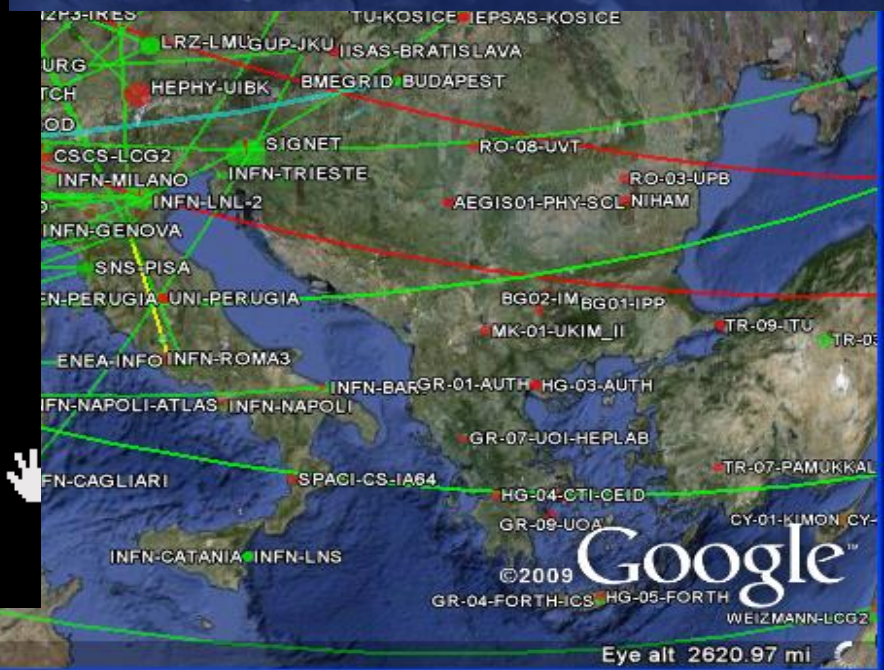
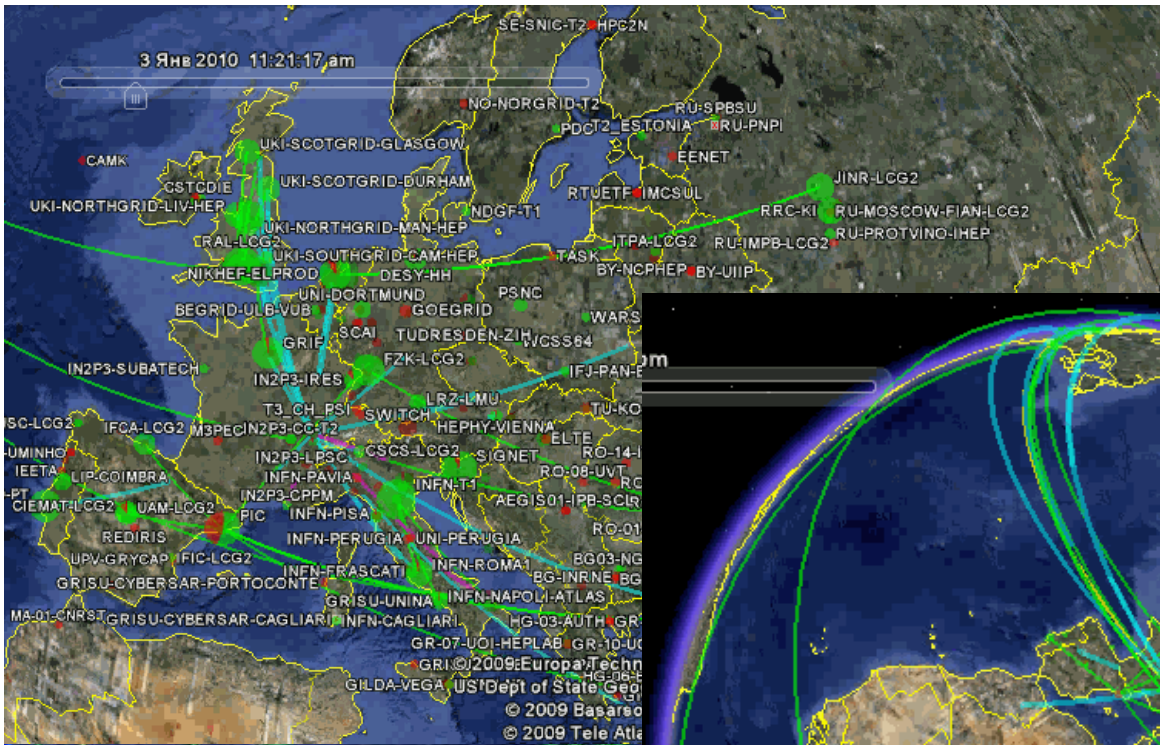
In 2010 the RDIG infrastructure comprises **10 Resource Centers with > 3000 CPU and > 5000 TB of disc storage.**



RDIG Resource Centres:

- ITEP
- JINR-LCG2 (Dubna)
- RRC-KI
- RU-Moscow-KIAM
- RU-Phys-SPbSU
- RU-Protvino-IHEP
- RU-SPbSU
- Ru-Troitsk-INR
- ru-IMPB-LCG2
- ru-Moscow-FIAN
- ru-Moscow-MEPHI
- ru-PNPI-LCG2 (Gatchina)
- ru-Moscow-SINP
- **Kharkov-KIPT (UA)**
- **BY-NCPHEP (Minsk)**
- **UA-KNU**
- **UA-BITP**

The Worldwide LHC Computing Grid (WLCG)



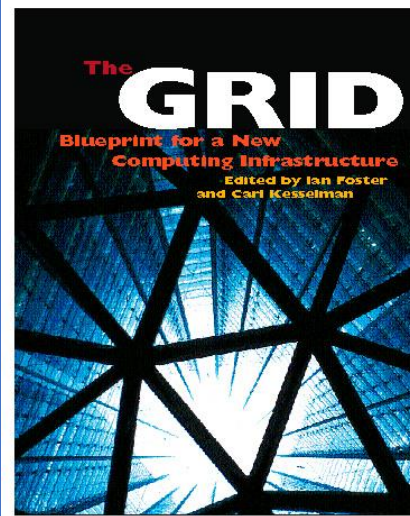
© 2009 Cnes/Spot Image
47°27'19.45" N 6°33'30.91" E

- Расширение компьютерных ресурсов за счет использования внешних невыделенных ресурсов (HLT, Clouds, HPC...)
- Изменения модели компьютеринга в каждом эксперименте, с целью оптимизации использования ресурсов
- Значительные усилия вкладываются в развитие программного обеспечения, чтобы улучшить общую производительность при использовании современных архитектур (многоядерность, GPU...)
- Оптимизации процессов обработки, количество хранящихся реплик данных и др.

Grids, clouds, fog, edge, supercomputers...

Grids

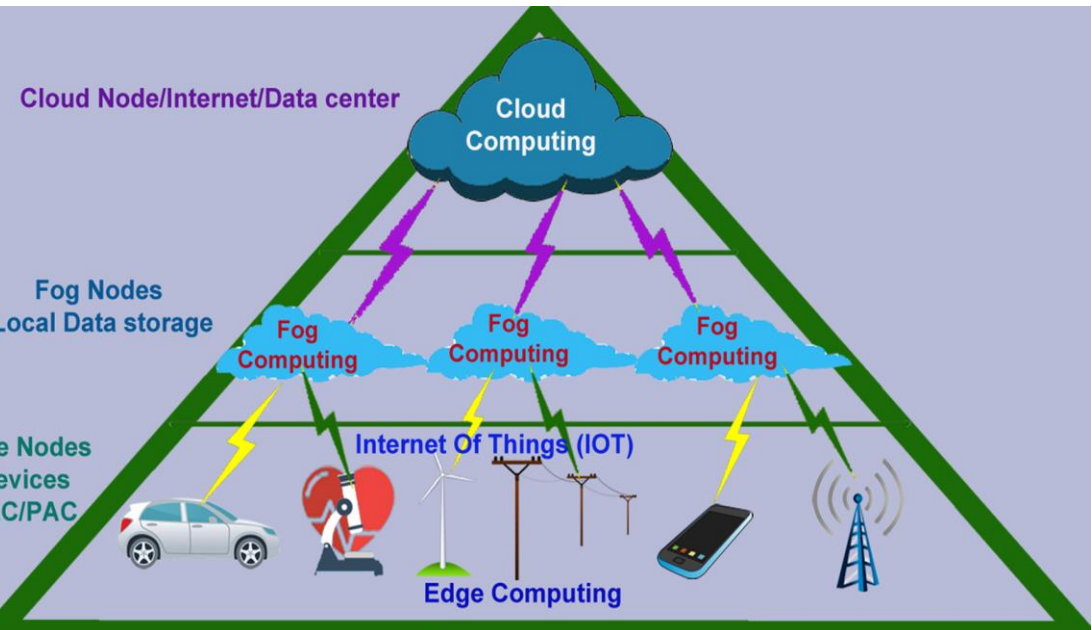
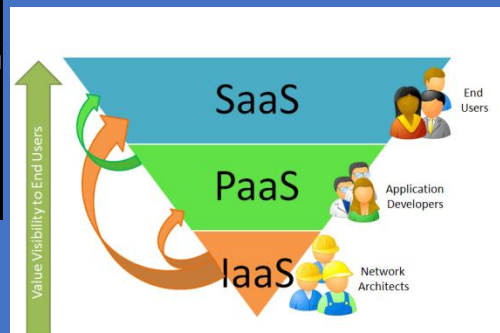
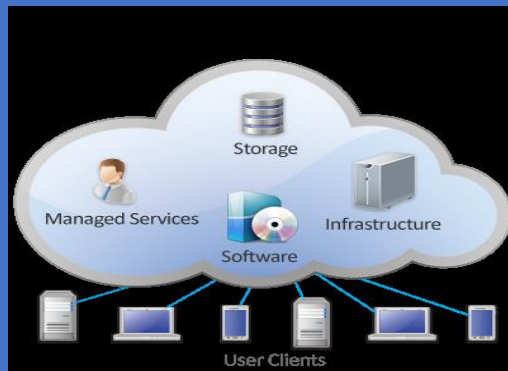
- Collaborative environment
- Distributed resources



Supercomputers

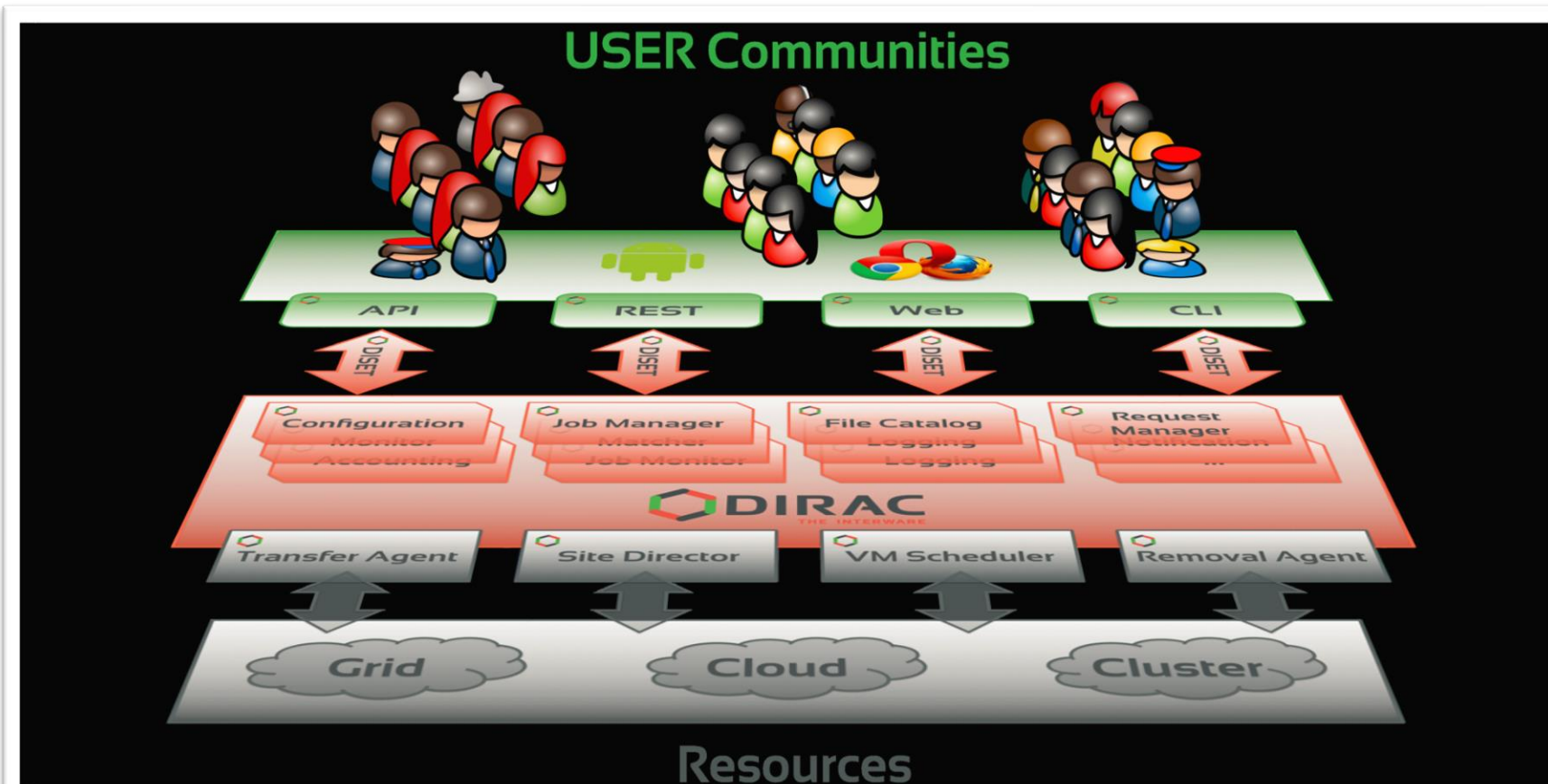


Clouds

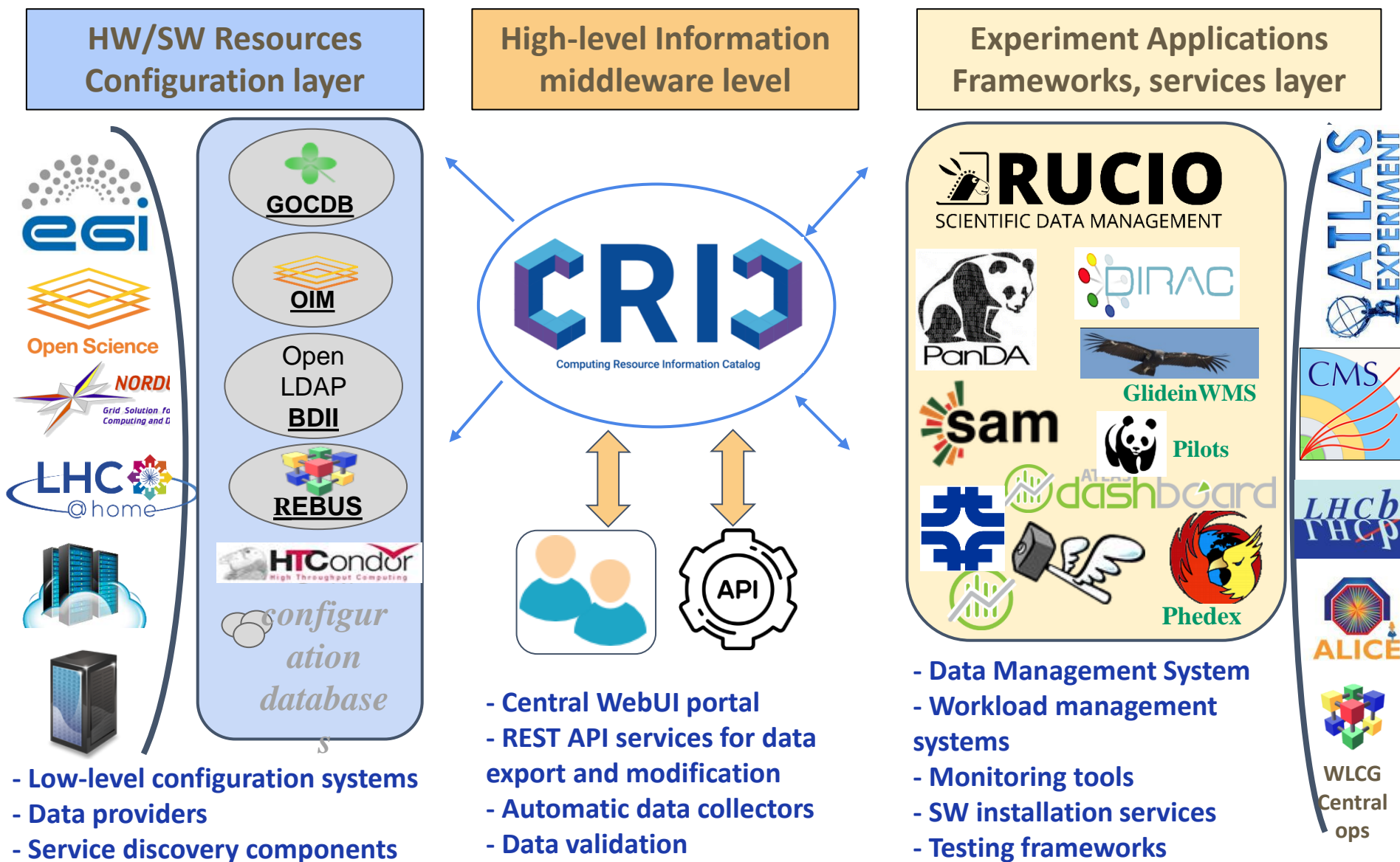


Платформа DIRAC

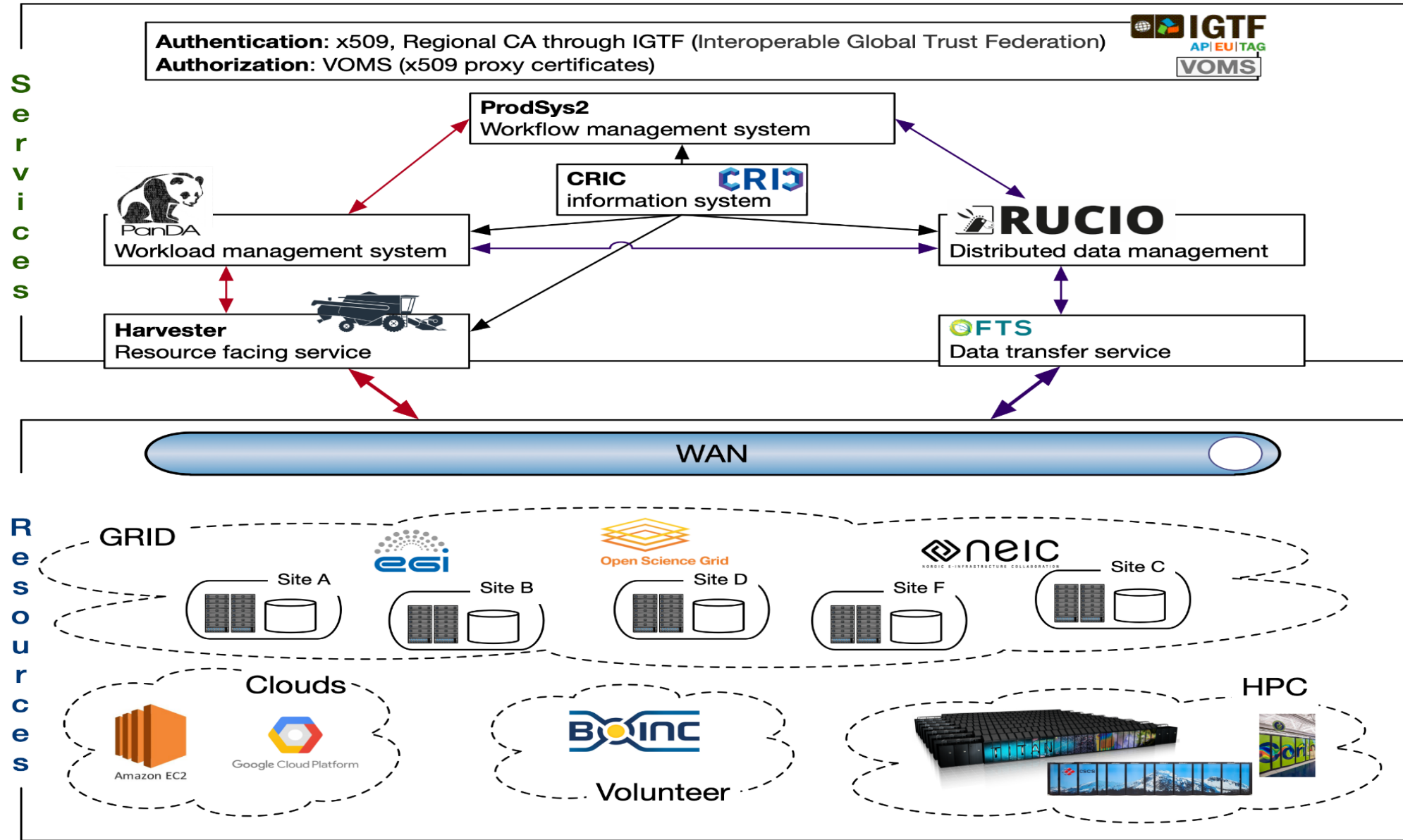
- DIRAC has all the necessary components to build ad-hoc grid infrastructures **interconnecting** computing resources of different types, allowing **interoperability** and simplifying **interfaces**.
- This allows to speak about the DIRAC *interware*.



CRIC: a unified topology system for a large scale, heterogeneous and dynamic computing infrastructure



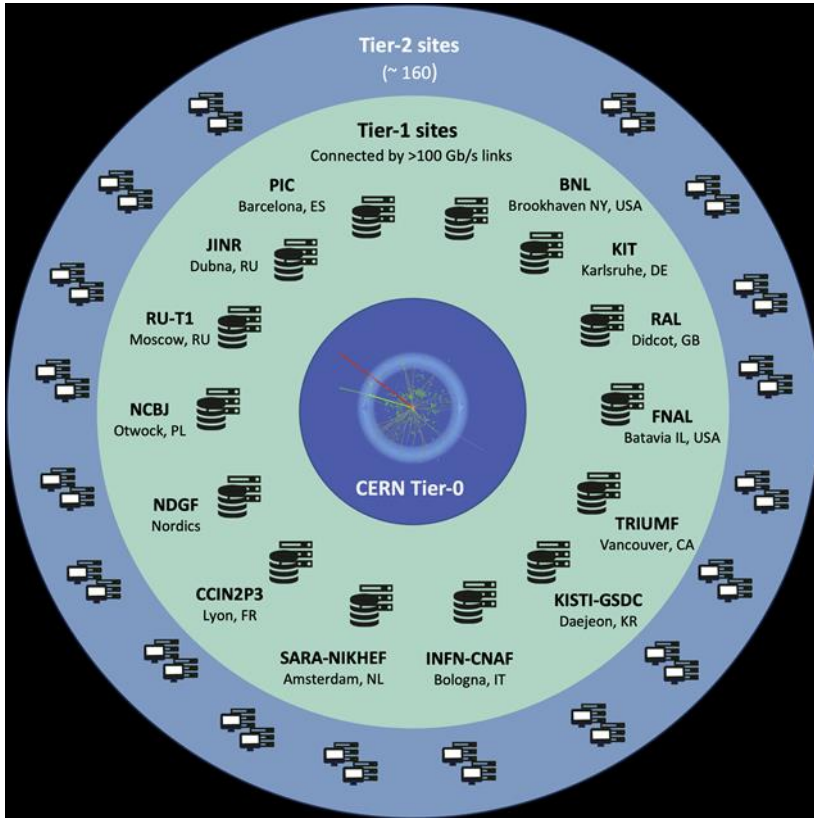
ATLAS computing



The Worldwide LHC Computing Grid



WLCG: an International collaboration to distribute and analyse LHC data. Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists



The mission of the WLCG project is to provide global computing resources to store, distribute and analyze the **~250-300 Petabytes** of data expected every year of operations from the Large Hadron Collider.

WLCG computing enabled physicists to announce the discovery of the Higgs Boson.

170 sites

42 countries

> 12k physicists

~1.6 M CPU cores

~2 EB of storage (1 EB - CERN)

> 2.5 million jobs/day

100-400 Gb/s links



Tier0 (CERN):
data recording,
reconstruction
and distribution

Tier1:
permanent
storage,
re-processing,
analysis

Tier2:
Simulation,
end-user
analysis

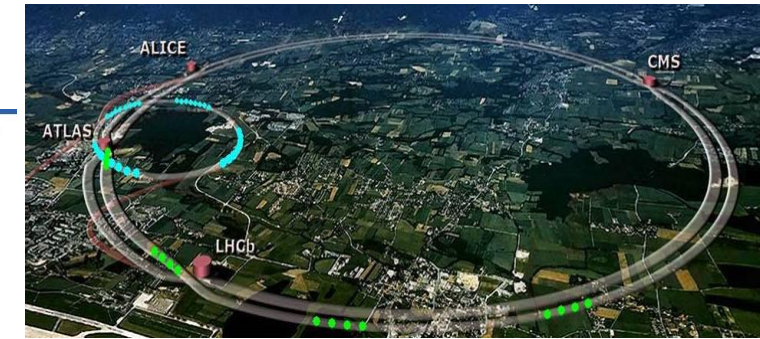
Countries — Sum CPU Work (HS23 hours) by Country and VO (LHC VOs)

Country	alice	atlas	cms	lhcb	Total	Percent
USA	236,929,906	3,104,319,728	2,937,679,000	0	6,278,928,634	25.95%
Switzerland	1,572,811,270	1,928,489,548	1,470,994,067	881,870,100	5,854,164,985	24.19%
Germany	353,571,160	1,430,078,402	781,539,502	411,641,240	2,976,830,304	12.3%
United Kingdom	47,102,022	1,566,571,819	329,396,977	990,134,417	2,933,205,234	12.12%
Russia	68,049,212	166,288,406	648,423,550	85,445,220	968,206,387	4%
Italy	121,408,726	207,166,953	506,188,230	63,976,557	898,740,466	3.71%
Canada	0	793,395,846	0	0	793,395,846	3.28%
France	44,994,230	438,917,399	108,632,855	175,730,168	768,274,653	3.17%
Spain	0	249,945,715	257,009,608	76,045,702	583,001,025	2.41%
Sweden	46,686,214	265,958,243	0	0	312,644,456	1.29%
Total	2,710,117,792	11,048,840,149	7,425,042,222	3,015,019,683	24,199,019,847	
Percent	11.20%	45.66%	30.68%	12.46%		

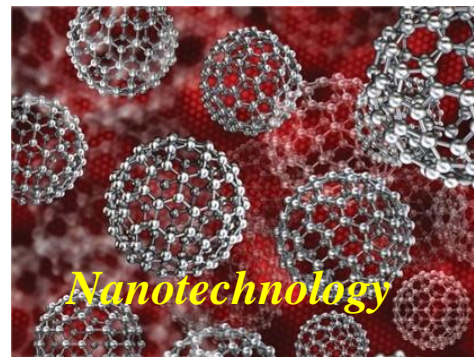
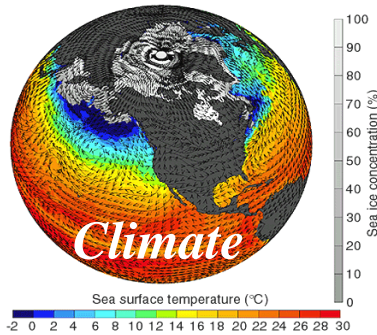
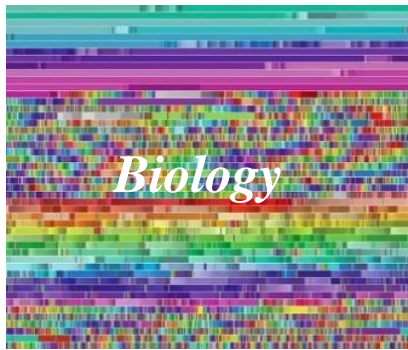
HPC+Big Data+Artificial intelligence



High Energy Physics



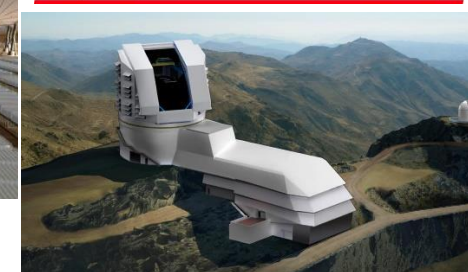
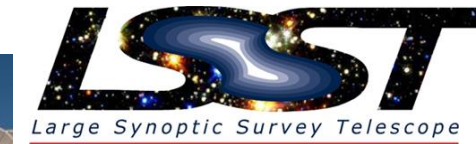
CERN Large Hadron Collider > 600 Pb/Year



Square Kilometer Array radio telescope
> 1 Eb/Year raw data (estimation)



International radiotelescope for the 21st century



Large Synoptic Survey Telescope > 10 Pb/Year (estimation)



... et cetera

International Large-scale projects



International large-scale projects are being prepared in Russia:

- **NICA**, JINR, Dubna (proton and heavy ion collider)
- **PIK**, PNPI, Gatchina (high-flow reactor complex)
- **SKIF**, INP SB RAS Novosibirsk (Siberian ring photon source)
- **Super S-Tau Fabric**, Sarov (electron-positron collider)
- **Нейтринная программа (Байкал, JUNO, NOVA, DUNE ...)**
- **синхротронно-нейтронная программа, науки о жизни**

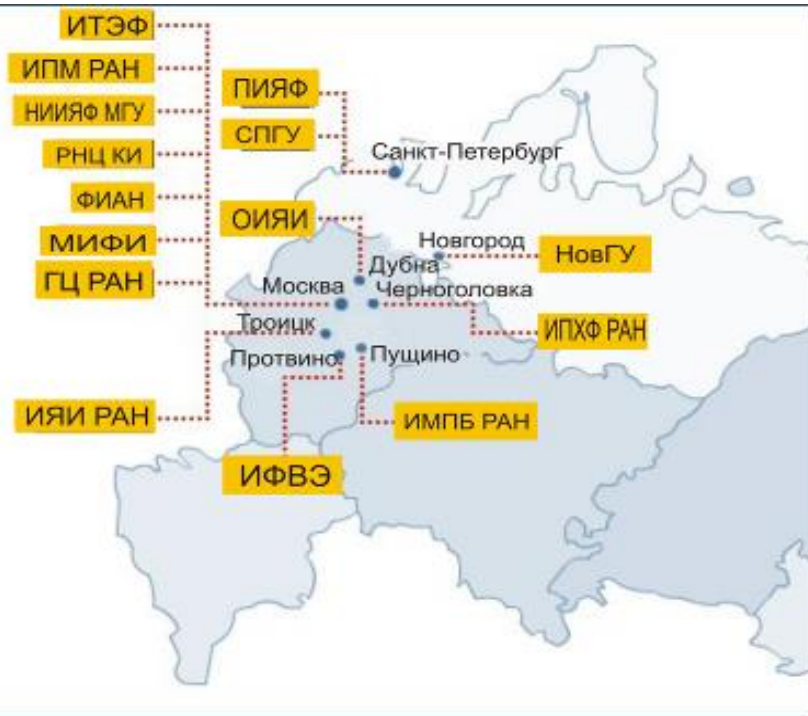


Joint Institute for Nuclear
Research

SCIENCE BRINGING NATIONS
TOGETHER



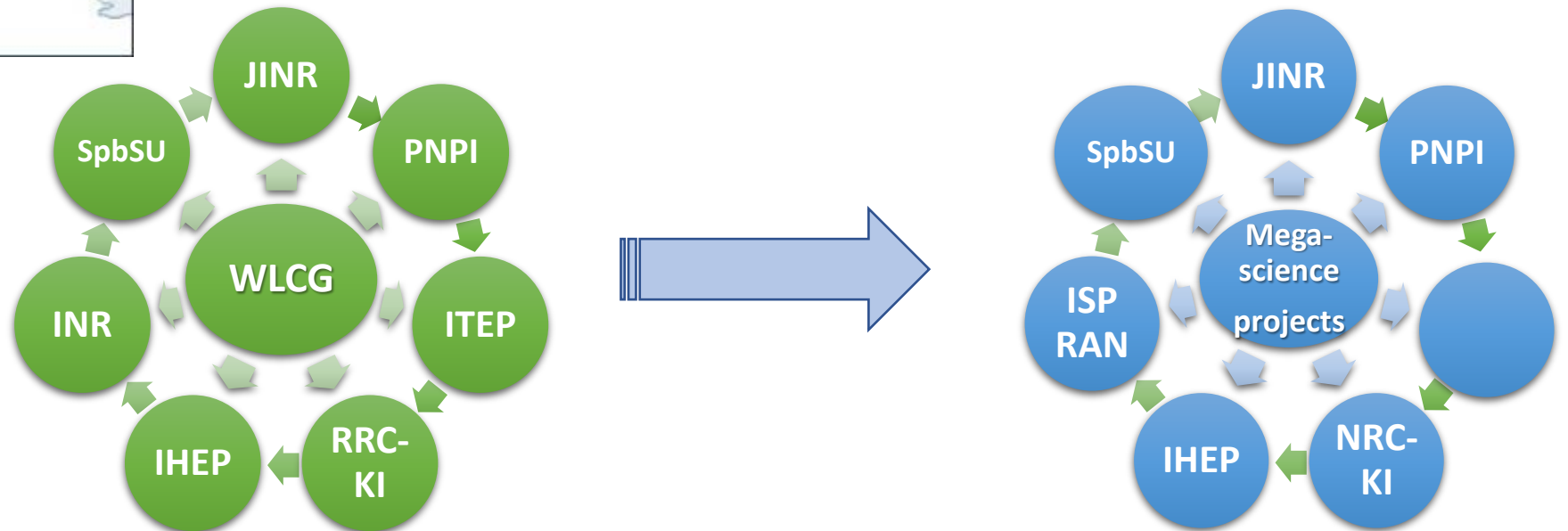
Институт ядерной физики
имени Г. И. Будкера СО РАН



The Russian consortium RDIG (Russian Data Intensive GRID) was set up in September 2003 as a national federation in the EGEE project.

A protocol between CERN, Russia and JINR on participation in the LCG project was signed in 2003. MoU on participation in the WLCG project was signed in 2007.

Consortium RDIG-M – Russian Data Intensive GRID for Megascience projects



Создание консорциума для IT-обеспечения исследовательской инфраструктуры класса «мегасайенс»

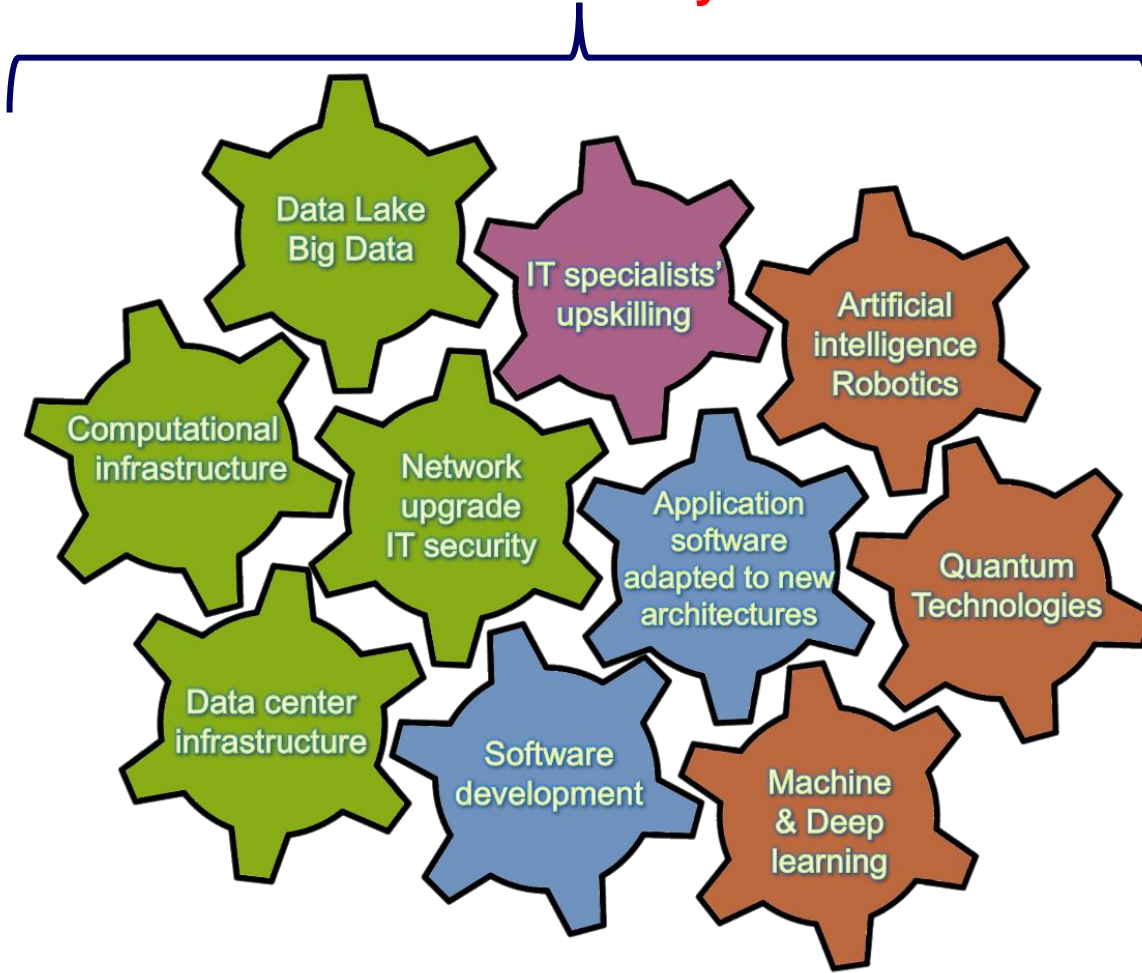


- Консорциум Российский GRID для интенсивных операций с данными (РДИГ) был создан в 2003 году для активного участия в распределенной обработке данных экспериментов на Большом адронной коллайдере LHC в рамках научной коллаборации LHC WLCG (Worldwide LHC Computing Grid). Созданная инфраструктура RDIG имеет огромное значение для эффективного участия ученых России в научной программе экспериментов на LHC.
- В России реализуется программа масштабных научных проектов, важнейшей частью которых является развитие распределенных гетерогенных компьютерных систем (включая системы с экстремально параллелизмом) для обработки, хранения, анализа экспериментальных данных, разработка и внедрение эффективных методов, алгоритмов и программного обеспечения для моделирования физических систем, математической обработки и анализа экспериментальных данных, развитие методов машинного обучения, искусственного интеллекта, квантовых вычислений.
- Для решения этой масштабной задачи необходимо развивать распределенную компьютерную инфраструктуру, объединяющую ключевые научные и образовательные институты, участвующие в проектах мегасайенс - РДИГ-М. Созданный консорциум на базе ОИЯИ, НИЦ Курчатовский институт, ИСП РАН должен стать ядром для IT-обеспечения исследовательской инфраструктуры класса «мегасайенс».

Strategy for Information Technology and Scientific Computing at JINR



Scientific IT ecosystem:



Coordinated development of interconnected IT technologies and computational methods

It will be a steady implementation/upgrades of

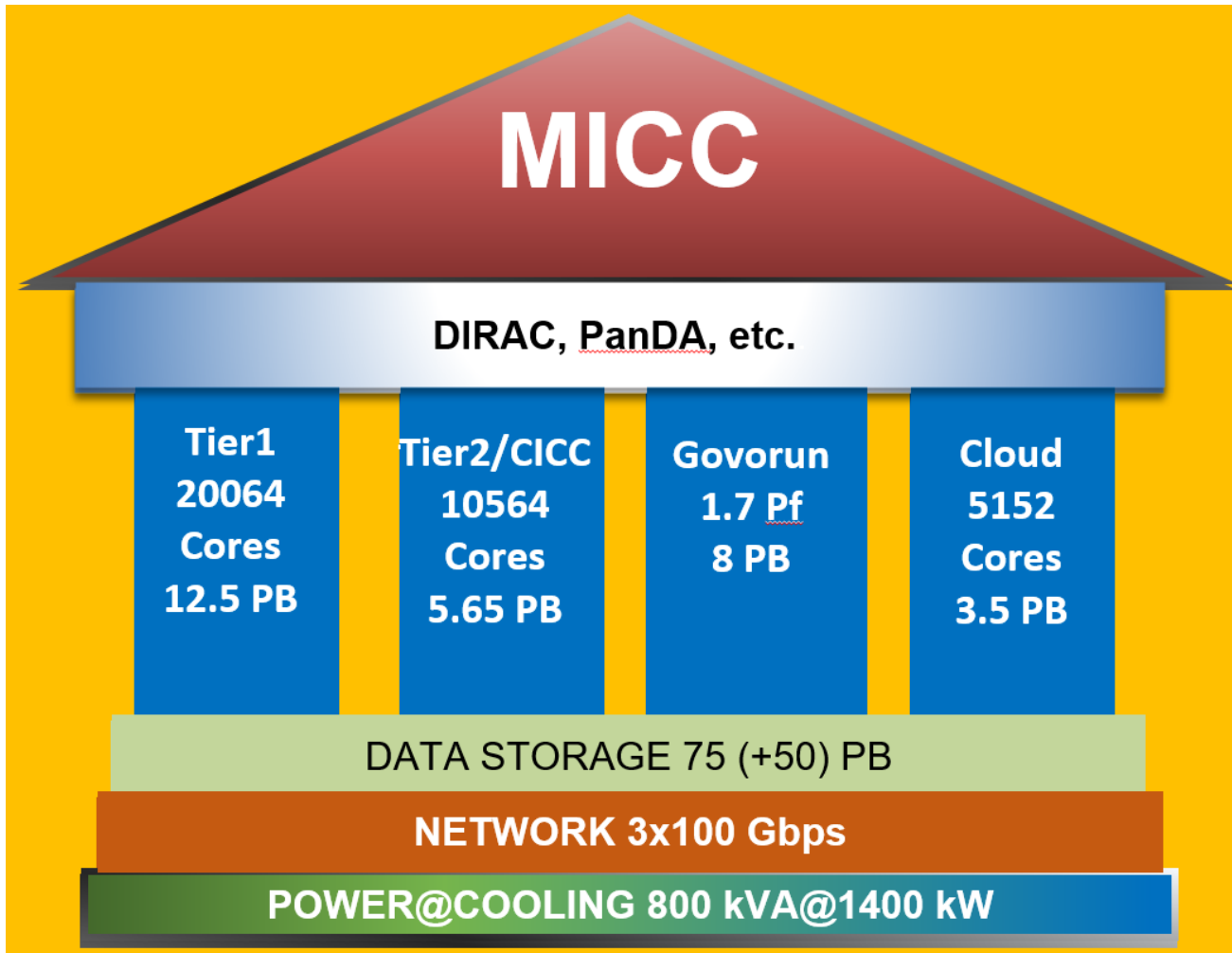
- Networking (**Tb/s** range),
- Computing infrastructure within the **Multifunctional Information & Computing Complex (MICC)** and
- “Govorun” Supercomputer,
- Data center infrastructure,
- Data Lake & long-term storage for all experiments.

The **development of new data processing and analysis algorithms** based on

- ML/DL,
- Artificial intelligence,
- Big Data
- Quantum technologies.

A variety of means will be used for IT specialists' upskilling.

Multifunctional Information and Computing Complex (MICC)



4 advanced software and hardware components

- Tier1 grid site
- Tier2 grid site
- hyperconverged “Govorun” supercomputer
- cloud infrastructure

Distributed multi-layer data storage system

- Disks
- Robotized tape library

Engineering infrastructure

- Power
- Cooling

Network

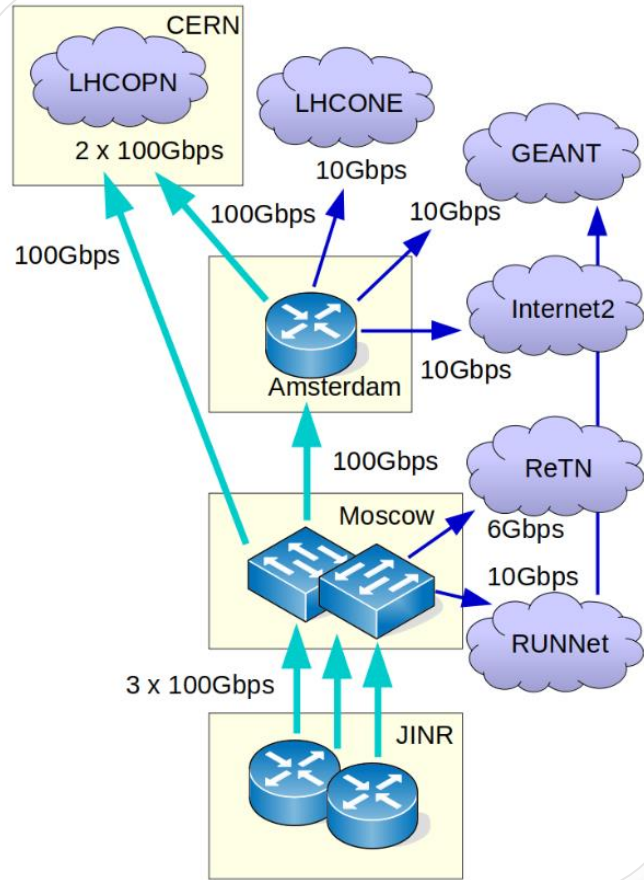
- Wide Area Network
- Local Area Network

The main objective of the project is to ensure multifunctionality, scalability, high performance, reliability and availability in 24x7x365 mode for different user groups that carry out scientific studies within the JINR Topical Plan

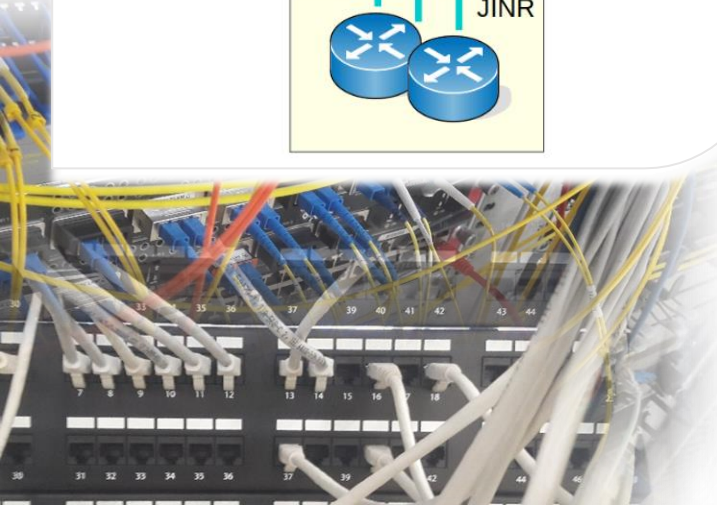
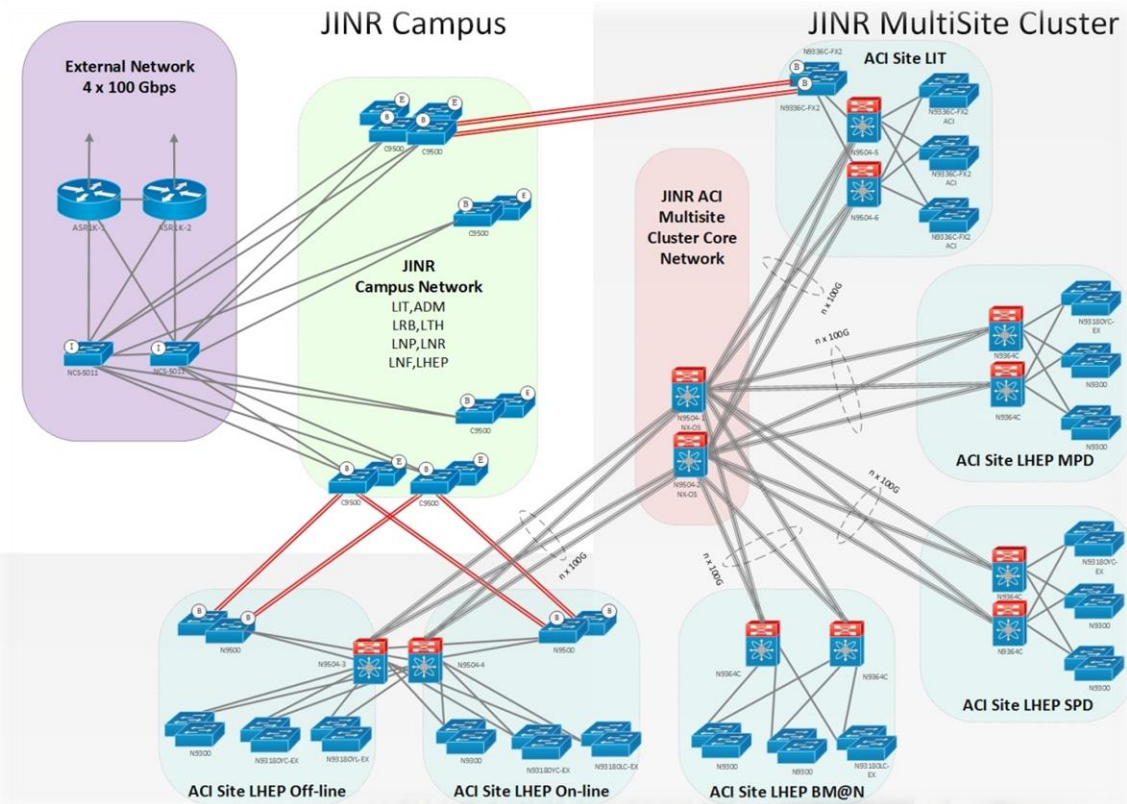
Networking



- JINR-Moscow **3x100 Gbit/s**
- JINR-CERN - **100 Gbit/s** and JINR-Amsterdam **100 Gbit/s** for LHCOPN, LHCONE, GEANT networks
- Direct channels up to 100 Gbit/s for communication with NIKS networks
- The multi-site cluster network with a bandwidth **4x100 Gbit/s** between VBLHEP and MLIT



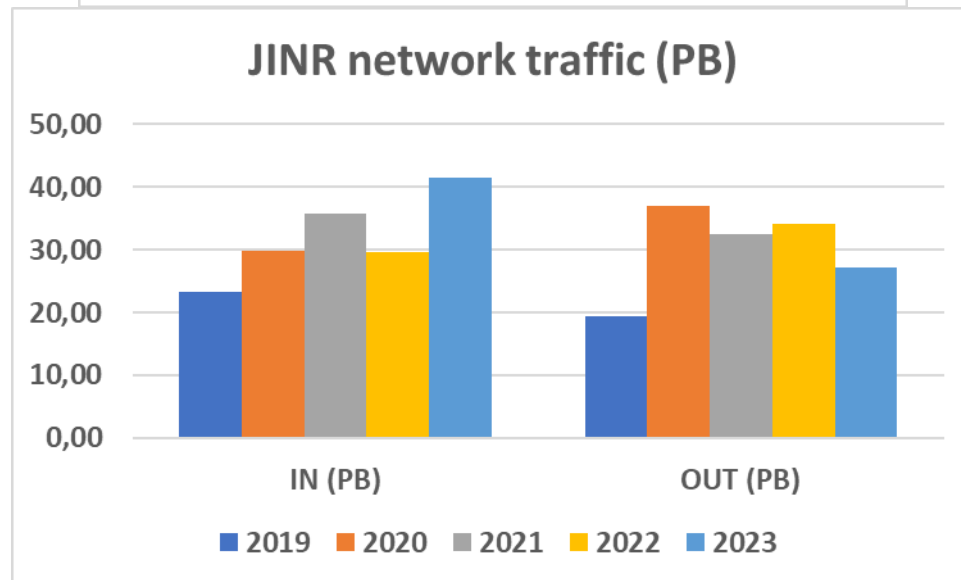
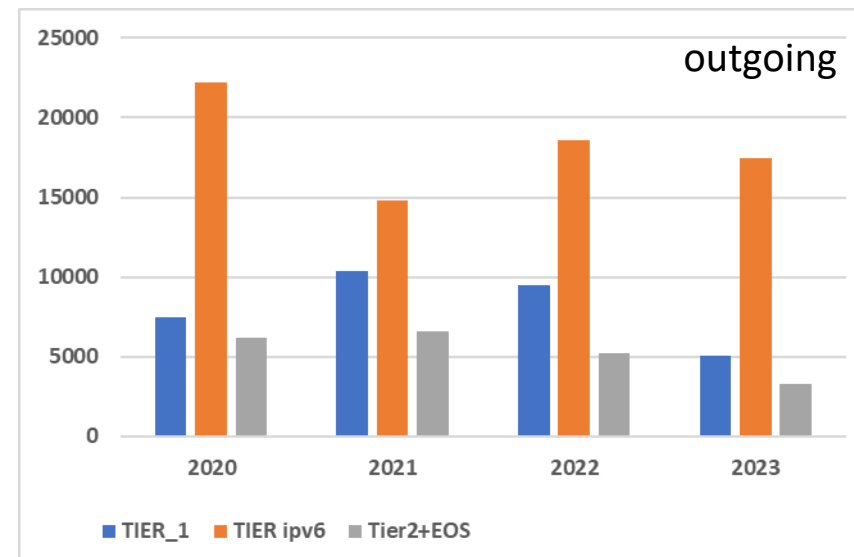
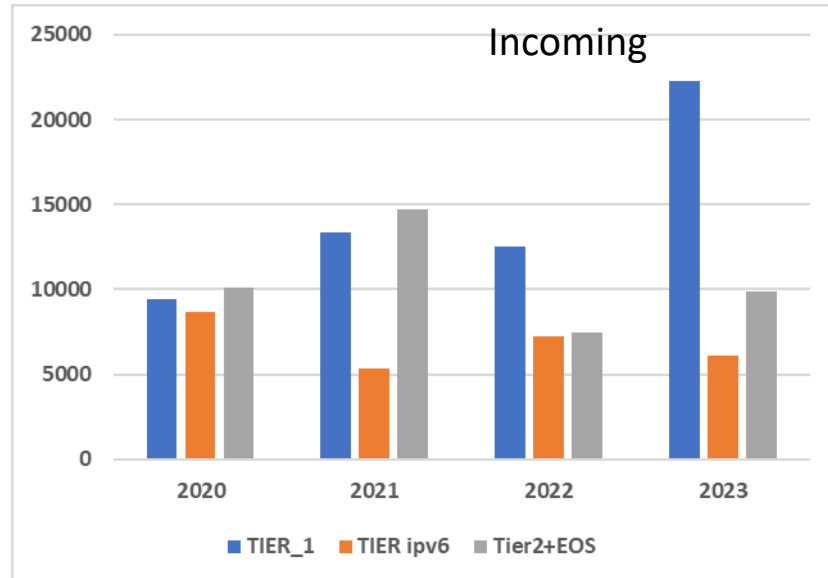
9327 network elements
18163 IP-addresses
6355 users
1464 E-library
911 remote VPN
121 VOIP
116 EDUROAM
4579 Email @jinr.ru



Networking @ Traffic



Distribution of the incoming and outgoing traffics by the JINR MICC in 2020-2023 (TB)

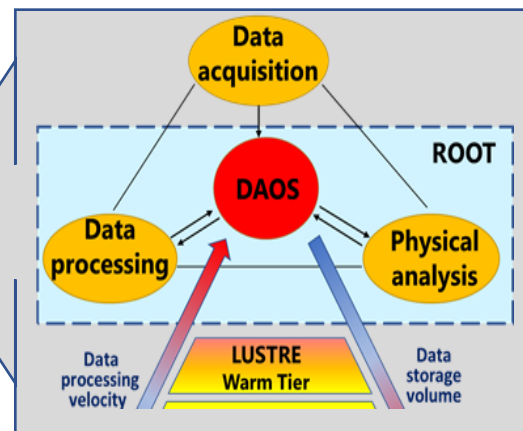


Общий входящий трафик ОИЯИ, включая сервера общего назначения, Tier1, Tier2, СК «Говорун» и облачные вычисления, составил в 2023 году 41,5 ПБ, общий исходящий – 27,5 ПБ.

Distributed Multilayered Data Storage System



- Limited data and **short-term** storage – to store OS itself, temporary user files
- AFS distributed global system – to store user home directories and software
- dCache is traditional for MICC grid sites – to large amounts of data (mainly LHC experiments) for **middle-term** period
- EOS is extended to all MICC resources – to store large amounts of data for **middle-term** period. At present, EOS is used for storage by BM@N, MPD, SPD, BaikalGVD, etc.
- Tape robotic systems – to store large amounts of data for **long-term** period. At present for CMS. BM@N, MPD, SPD, JUNO – in progress.

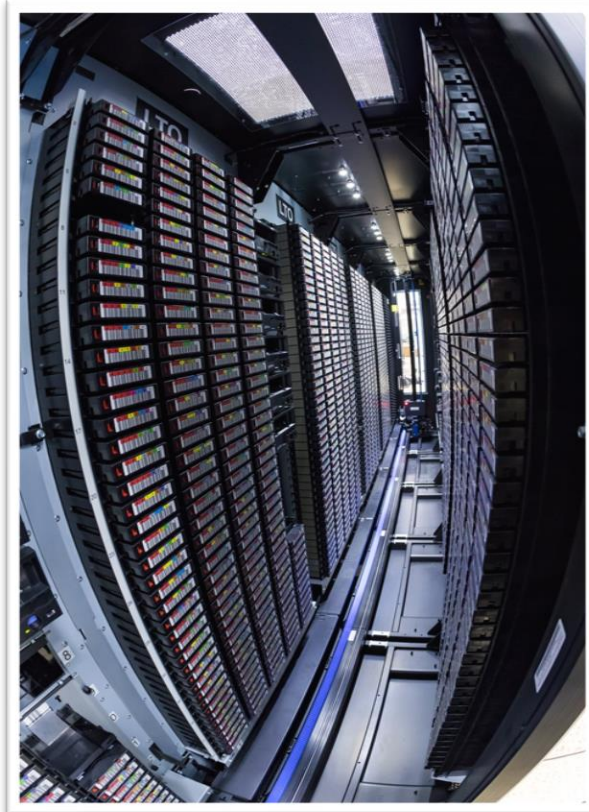


Special **hierarchical data processing and storage system** with a software-defined architecture was developed and implemented on the “Govorun” supercomputer.

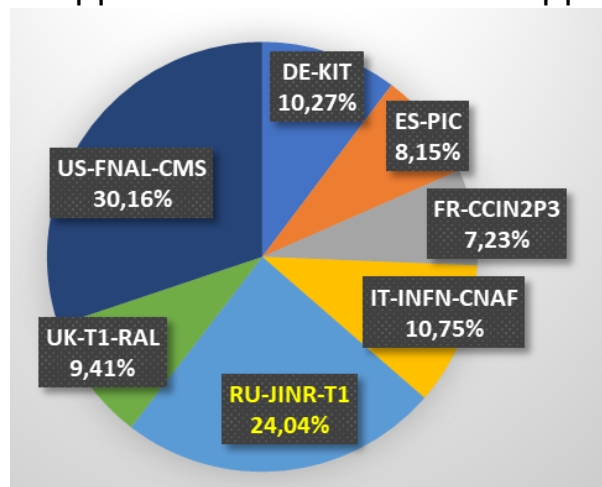
According to the speed of accessing data there are next layers:

- ✓ very hot data (DAOS (Distributed Asynchronous Object Storage)) ,
- ✓ the most demanded data (fastest access),
- ✓ hot data
- ✓ warm data (LUSTRE).

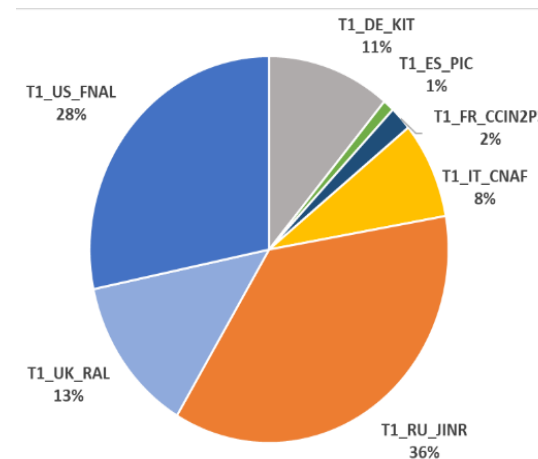
JINR Tier1 for CMS (LHC) and NICA



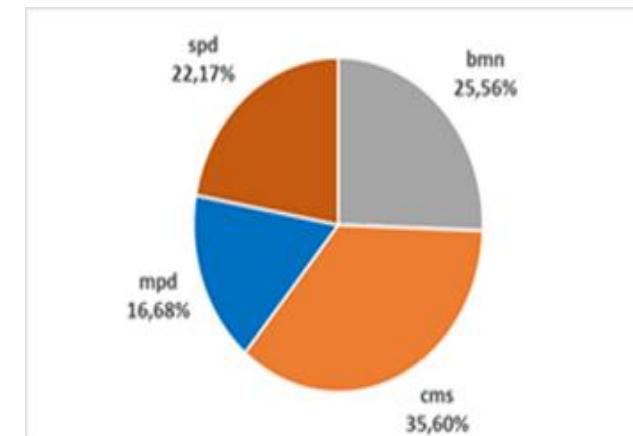
Вклад мировых Tier1 центров в обработку экспериментальных данных CMS за 2023 год:



Количество обработанных событий эксперимента CMS за 2023 год



Распределение по числу задач выполненных на Tier1 экспериментами CMS, BM@N, MPD и SPD в 2023 году



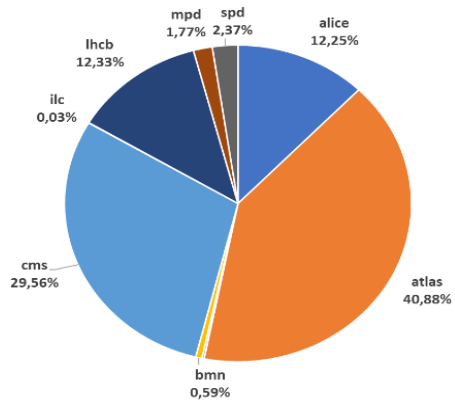
- 20064 cores
- 360 kHS06
- 12.5 PB disks
- 50.6 (+ 50) PB tapes
- 100% reliability and availability

Tier1 CMS	2023 July	2024	2024 march
US-FNAL-CMS	1,920,062,861	623,513,239	208,060,175
RU-JINR-T1	1,660,359,767	601,406,360	249,711,583
DE-KIT	790,474,346	280,425,085	130,491,459
UK-T1-RAL	680,993,976	247,032,988	102,539,577
FR-CCIN2P3	651,295,705	0	0
IT-INFN-CNAF	617,509,540	29,307,972	0
ES-PIC	491,172,615	150,204,096	57,226,844

Tier2 at JINR

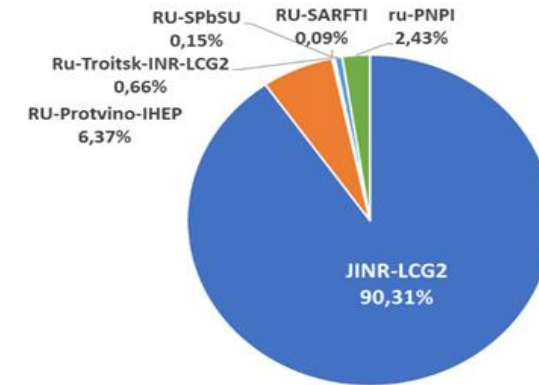


Использование Tier2 сайта ОИЯИ (JINR-LCG2) виртуальными организациями в рамках грид-проектов



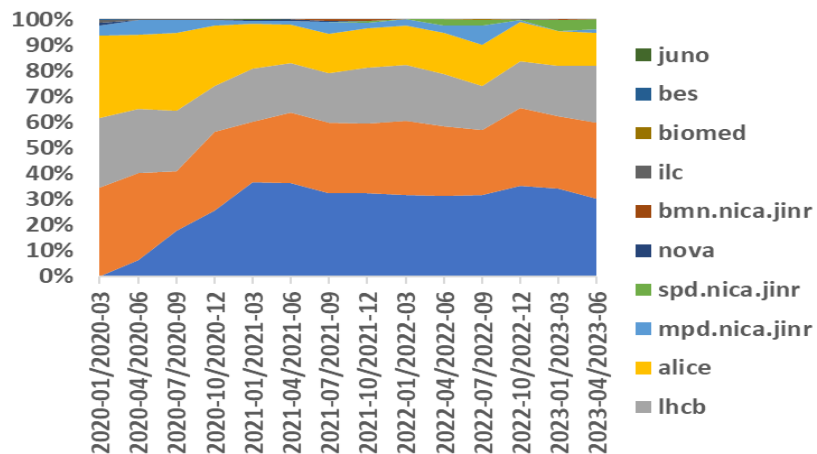
Tier2 at JINR provides computing power and data storage and access systems for the majority of JINR users and user groups, as well as for users of virtual organizations (VOs) of the grid environment (LHC, NICA, FAIR, etc.).

Распределение выполненных на грид-сайтах RDIG задач

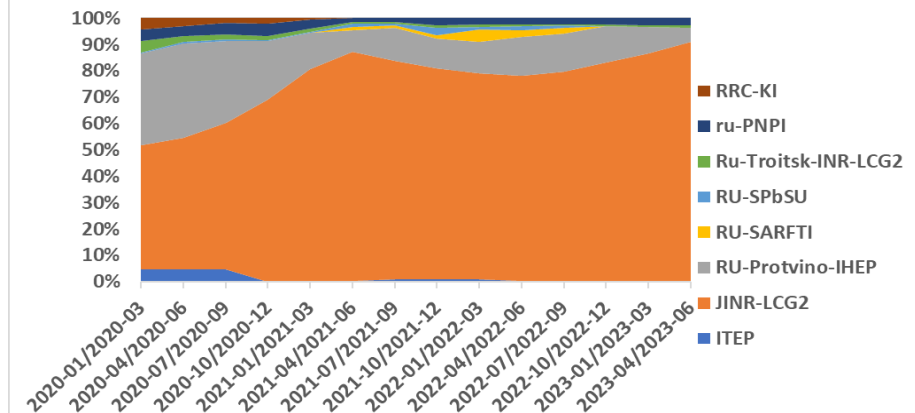


The JINR Tier2 output is the highest (90.31%) in the Russian Data Intensive Grid (RDIG) Federation.

Accounting - 2020_1 to 2023_5 normcpu on JINR Tier2 for VO and Quarter



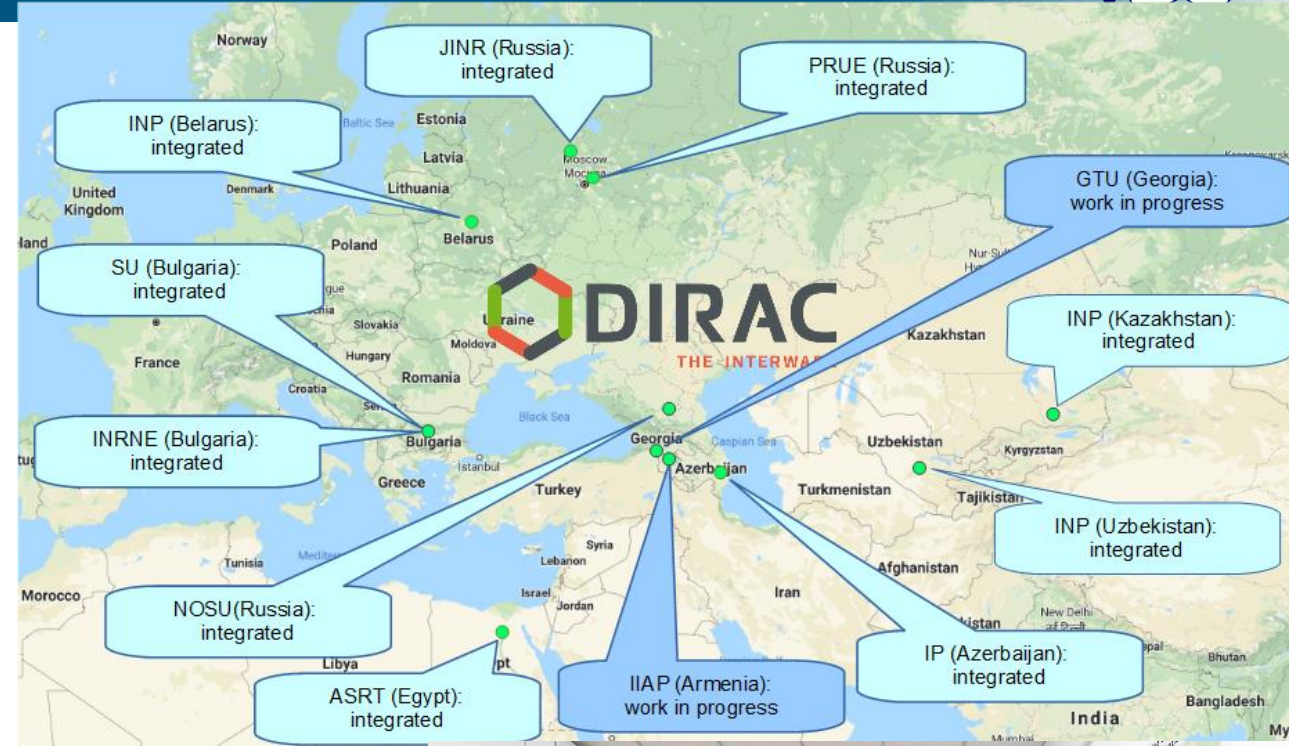
Accounting - 2020_1 to 2023_5 normcpu for RDIG Tier2 and Quarter



Cloud Infrastructure



DIRAC-based distributed information and computing environment (DICE) that integrates the JINR Member State organizations' clouds



- Cloud Platform - OpenNebula
- Virtualization - KVM
- Storage (Local disks, Ceph)
- Total Resources
 - ~ **5,152** CPU cores; 80 TB RAM;
 - 3.5 PB** of raw ceph-based storage

- Computational resources for neutrino experiments
- Testbeds for research and development in IT
- COMPASS production system services
- Data management system of the UNECE ICP Vegetation
- Scientific and engineering computing
- Service for data visualization
- VMs for JINR users

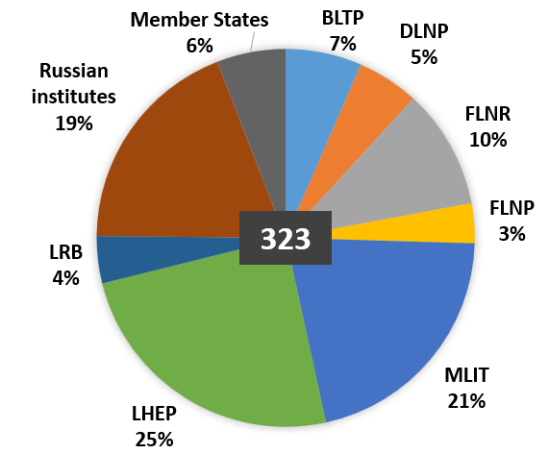
“Govorun” Supercomputer



- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 1.7 PFlops DP
- GPU component based on NVIDIA Tesla V100&A100
- CPU component based on RSC “Tornado” liquid cooling solutions
- The most energy-efficient center in Russia (PUE = 1.06)
- Storage performance >300 GB/s

Key projects that use the resources of the SC “Govorun”:

- NICA megaproject,
- calculations of lattice quantum chromodynamics,
- computations of the properties of atoms of superheavy elements,
- studies in the field of radiation biology,
- calculations of the radiation safety of JINR’s facilities.

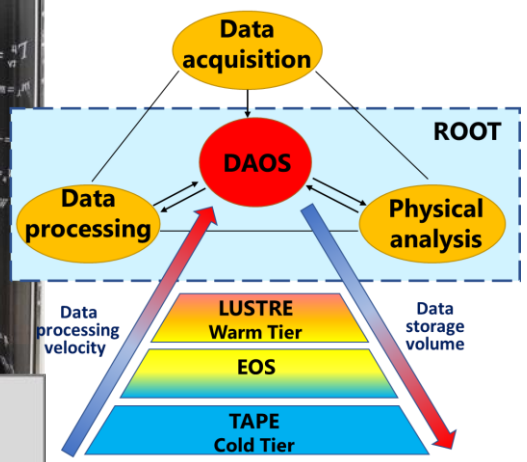


Total number of users : 323



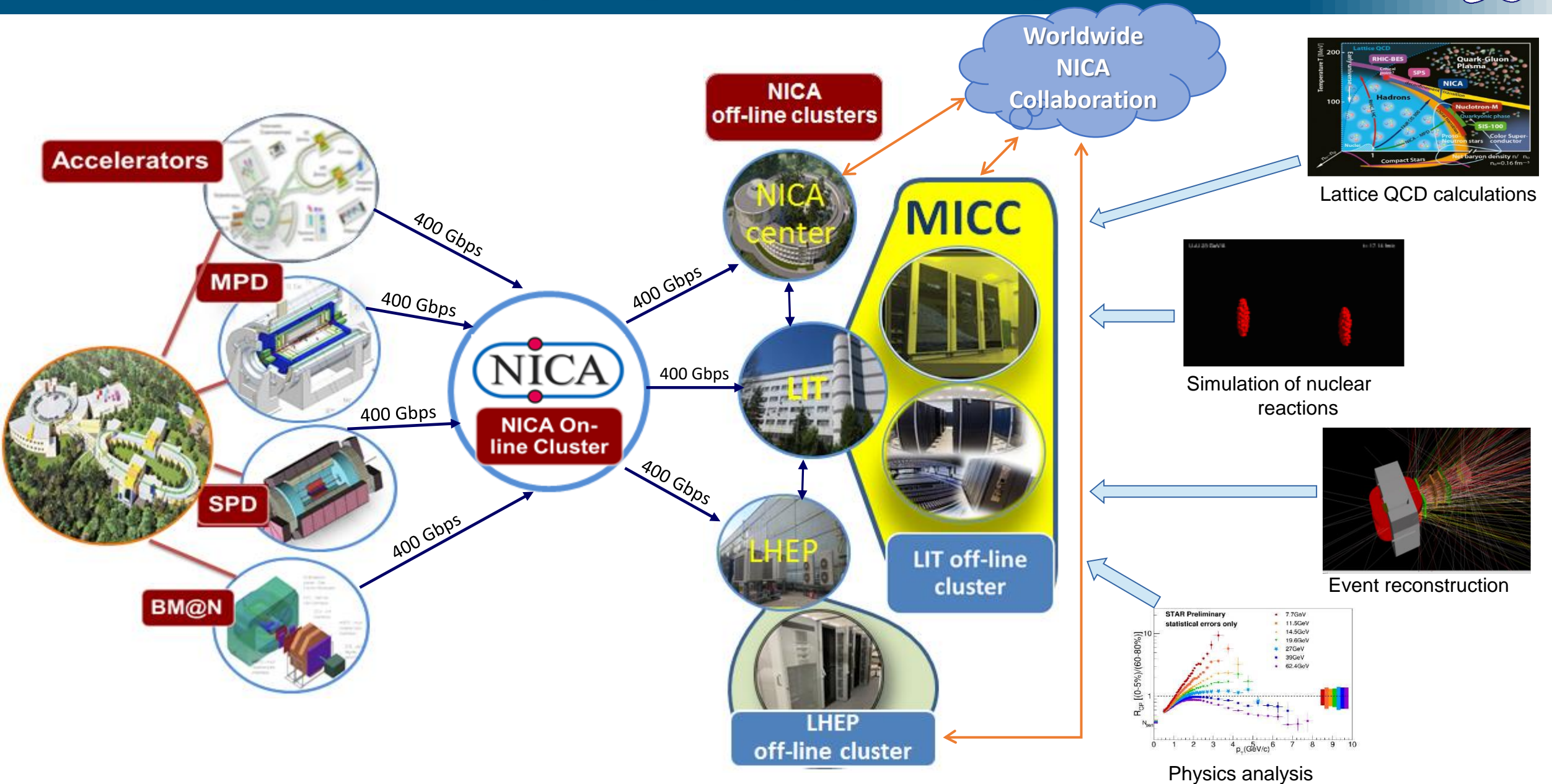
GPU-accelerator

Hyperconverged CPU and Distributed Storage Nodes

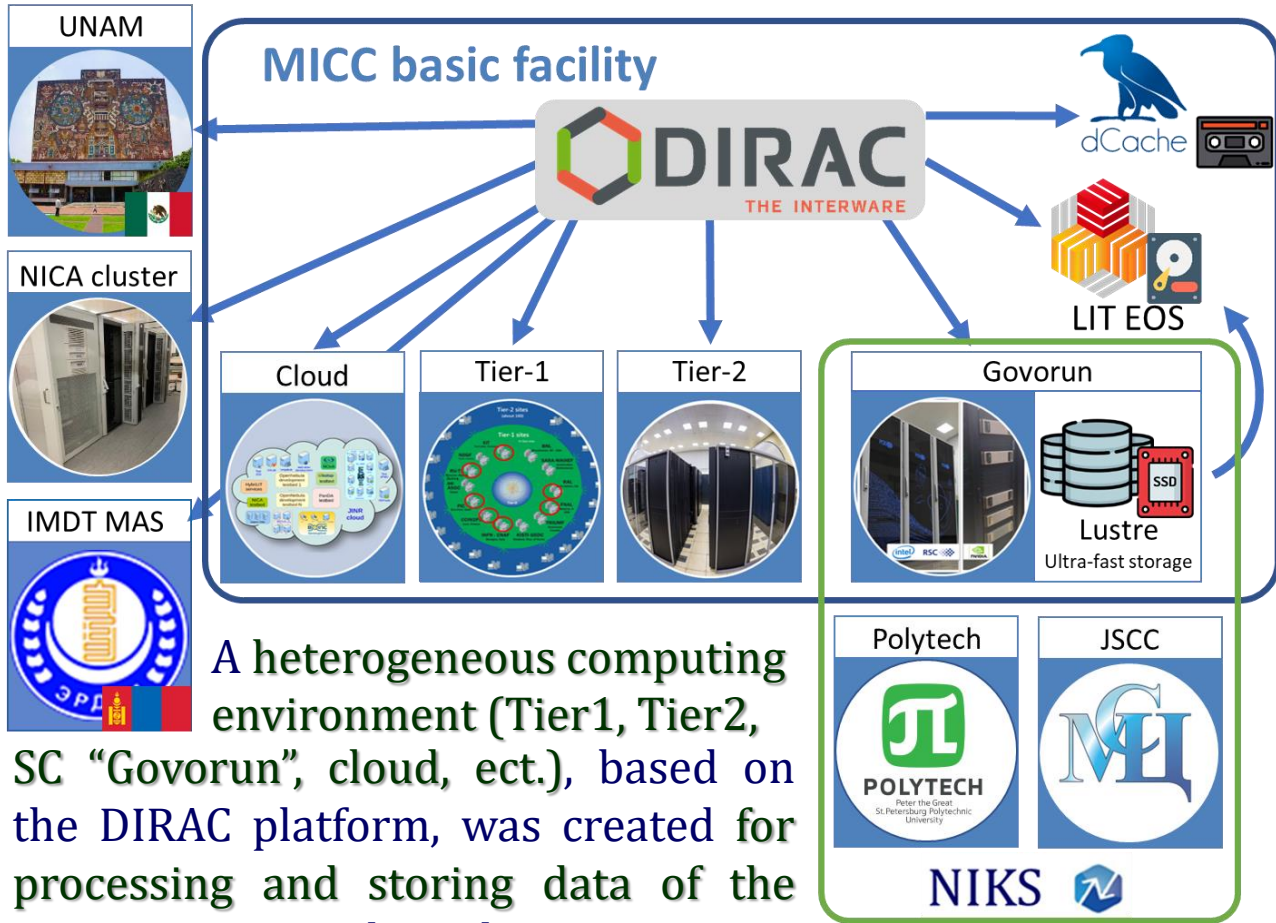


> 250 user papers (two in Nature Physics)

NICA Computing Concept & Challenges



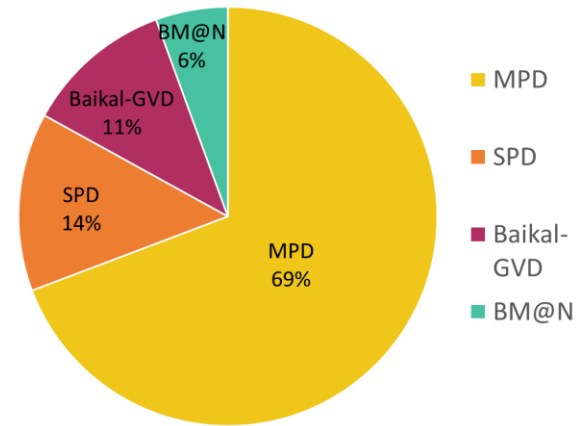
DIRAC-based distributed heterogeneous environment



A heterogeneous computing environment (Tier1, Tier2, SC “Govorun”, cloud, ect.), based on the DIRAC platform, was created for processing and storing data of the experiments conducted at JINR.

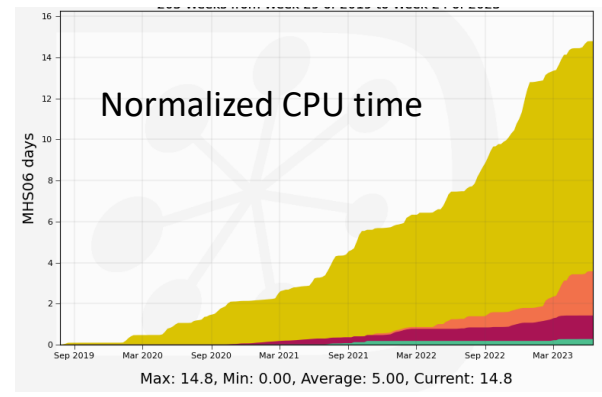
The distributed infrastructure is used by the MPD, Baikal-GVD, BM@N, SPD.

Use of DIRAC platform by experiments in 2019-2022

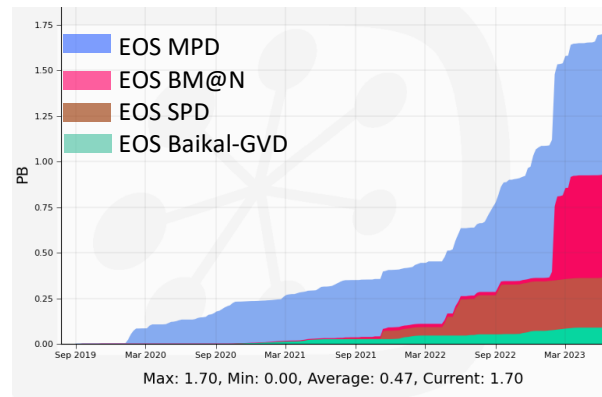


Total Number of executed jobs

The major user of the distributed platform is the MPD experiment



Data processed by experiments



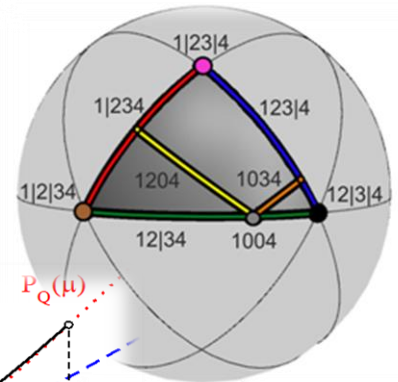
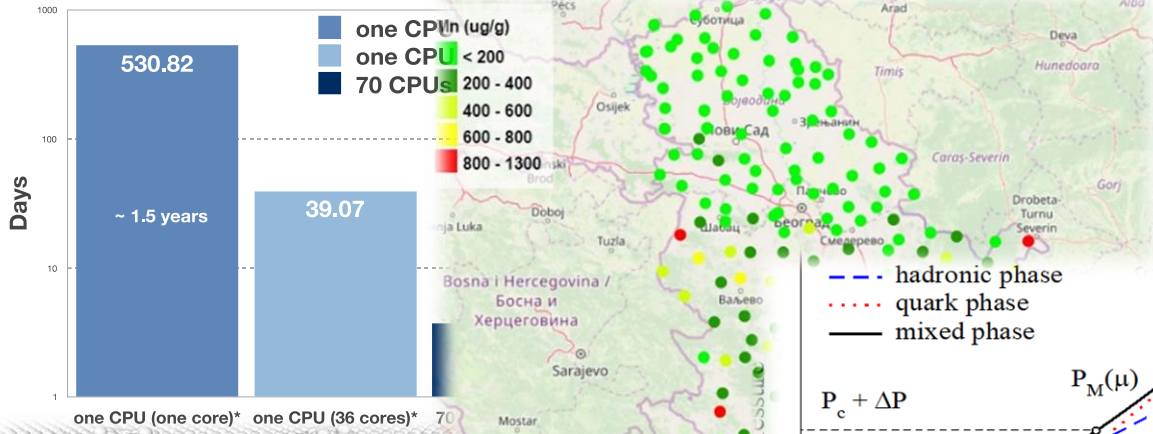
Summary statistics of using the DIRAC platform for MPD tasks in 2019-2022

<p>31 data generation campaigns</p>	<p>1.283 B events generated</p>	<p>440M events reconstructed</p>	<p>1.5M jobs completed</p>	<p>1557 years total computation time</p>	<p>1.3 PB MPD data produced</p>
-------------------------------------	---------------------------------	----------------------------------	----------------------------	--	---------------------------------

Methods, Algorithms and Software

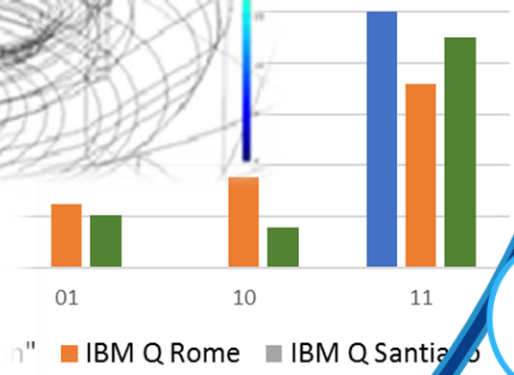
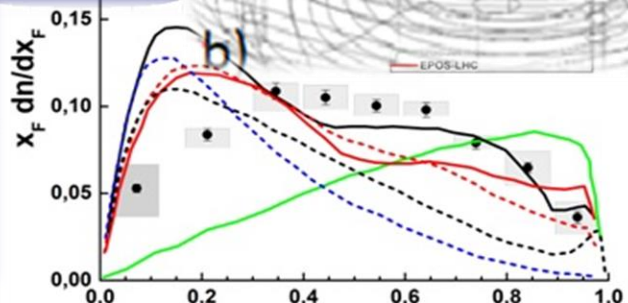
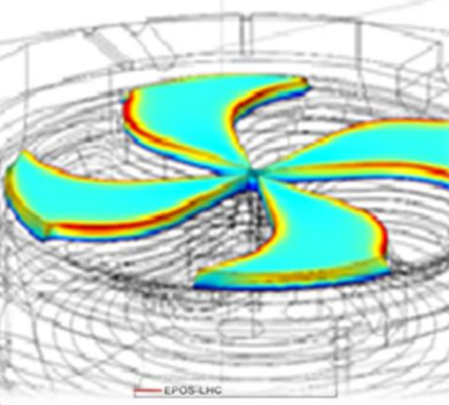
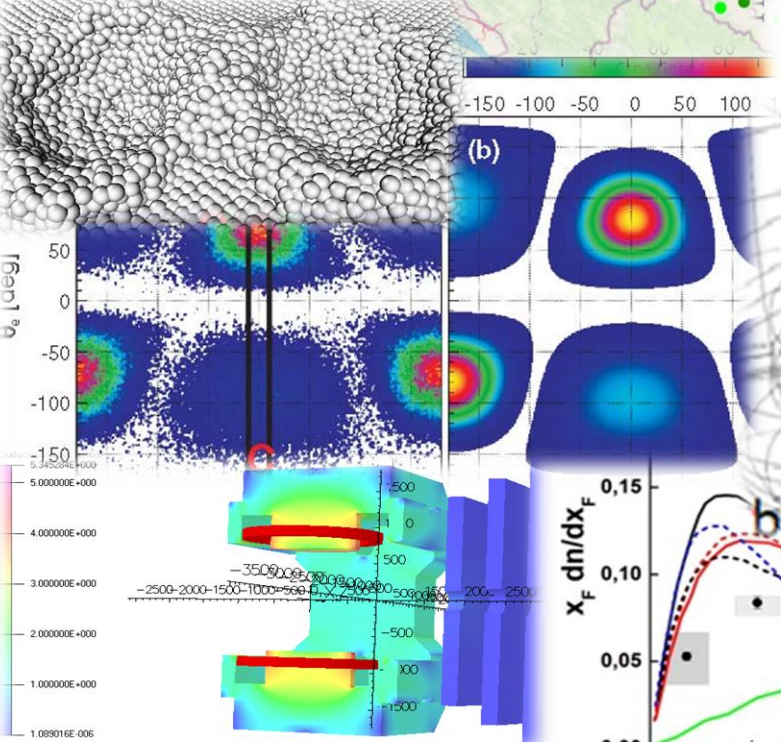


Govorun Supercomputer



— hadronic phase
 quark phase
 — mixed phase

$P_c + \Delta P$
 P_c
 $P_H(\mu)$



Numerical modeling of complex physical systems



Experimental data processing and analysis



Big Data



Machine and Deep learning



AI and robotics

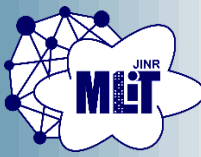


Computer algebra



Quantum computing

Activity: Digital ecosystem (Digital JINR)



The digital platform “**JINR Digital EcoSystem**” integrates existing and future services

to support

scientific,
administrative and social activities,
maintenance of the engineering and IT infrastructures

to provide

reliable and secure access to various types of data

to enable

a comprehensive analysis of information

using

modern **Big Data technologies and artificial intelligence.**

JINR 
Digital Eco System



Other services



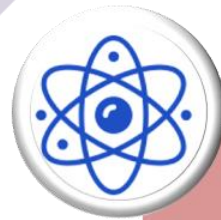
Information services



Network services



Administrative services



Scientific services



Single access point to all services

Digital technologies

Digital infrastructures

IT specialists and users

Development of the system for training and retraining IT specialists

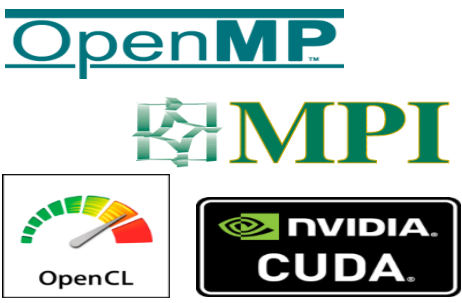


Training courses, master classes and lectures

MLIT staff and leading scientists from JINR and its Member States

Leading manufacturers of modern computing architectures and software

Parallel programming technologies



Tools for debugging and profiling parallel applications



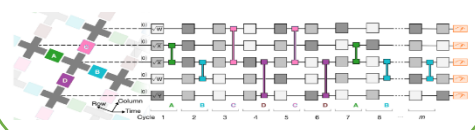
Work with applied software packages



Frameworks and tools for ML/DL tasks



Quantum algorithms, quantum programming and quantum control





International Conference "Distributed Computing and Grid Technologies in Science and Education"



- Distributed computing systems
- Computing for MegaScience Projects
- Distributed computing & HPC applications
- Distributed Storage Systems
- HPC
- Cloud technologies
- Big data Analytics and Machine learning
- Quantum informatics and computing

M M C P MATHEMATICAL MODELING AND COMPUTATIONAL PHYSICS



- ❑ methods, software and program packages for data processing and analysis;
- ❑ mathematical methods and tools for modeling complex physical and technical systems, computational biochemistry and bioinformatics;
- ❑ methods of computer algebra, quantum computing and quantum information processing;
- ❑ machine learning and big data analytics;
- ❑ algorithms for parallel and hybrid calculations.

NEC'2019

The International Symposium Nuclear Electronics and Computing



- Detector & Nuclear Electronics
- Triggering, Data Acquisition, Control Systems
- Distributed Computing, GRID and Cloud Computing
- Machine Learning Algorithms and Big Data Analytics new!

- Research Data Infrastructures
- Computations with Hybrid Systems (CPU, GPU, coprocessors)
- Computing for Large Scale Facilities (LHC, FAIR, NICA, SKA, PIC, XFEL, ELI, etc.)
- Innovative IT Education

IT SCHOOL JINR
Involvement of young specialists in solving tasks that face JINR using state-of-the-art information technologies



10

GRID2023

3-7 July 2023

10th International Conference
“Distributed Computing and Grid Technologies in
Science and Education”



Conference Topics:

1. Distributed Computing Systems
2. HPC
3. Distributed Computing and HPC Application
4. Cloud Technologies
5. Computing for MegaScience Projects
6. Quantum Informatics and Computing
7. Big Data, M/D Learning, Artificial Intelligence
8. Student session

Workshop “Computing for radiobiology and medicine”

Workshop “Modern approaches to the modeling of research reactors, creation of the “digital twins” of complex systems”

Round table “RDIG-M - Russian distributed infrastructure for large-scale scientific projects in Russia”

Round table on IT technologies in education

More than **275** participants

In person - 216

Remotely - 60

30 Plenary reports

135 Sessional reports

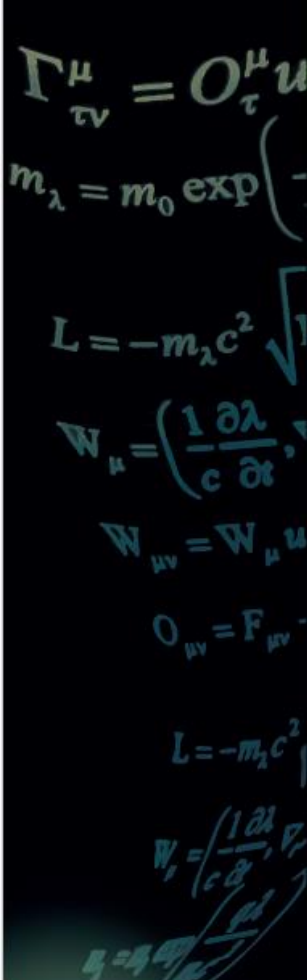
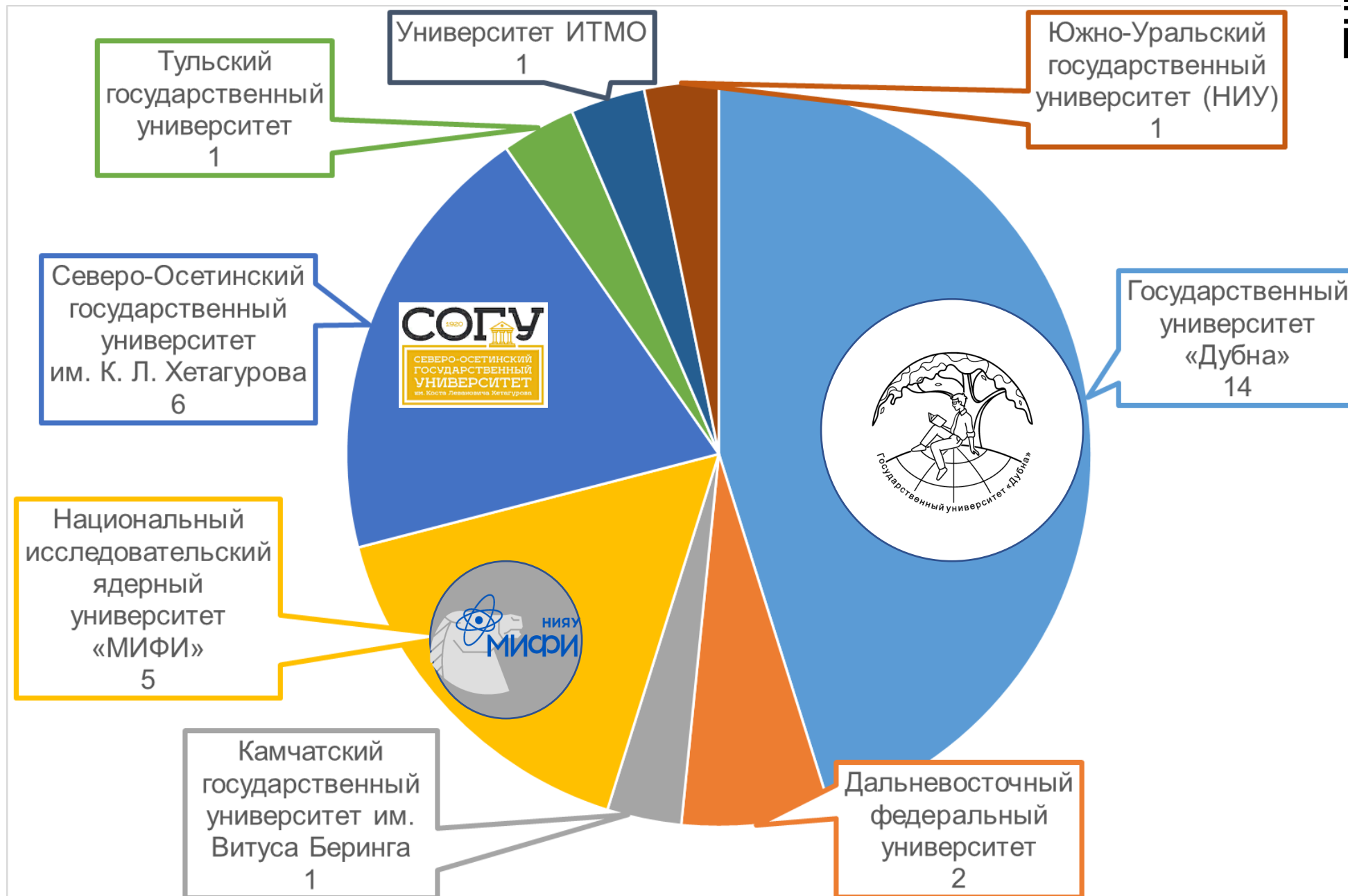
17 Countries: Azerbaijan, Armenia, Belarus, Bulgaria, the Czech Republic, Egypt, Germany, Georgia, Iran, Kazakhstan, Mexico, Moldova, Mongolia, Serbia, CERN and Uzbekistan. **Russia** was represented by participants **from 41 universities and research centers.**



JINR School of Information Technology 2023

50 students from 11 Russian universities







BMK MGU



Общие отцы-основатели ВМК МГУ и ЛВТА ОИЯИ (ныне ЛИТ им. М.Г. Мещерякова) внесли огромный вклад в развитие программного обеспечения для БЭСМ-6, численных методов, вычислительной физики и компьютеринга. И сегодня продолжает развиваться сотрудничество между МГУ и ЛИТ в области:

- математических методов и математического моделирования,
- грид-технологий в России,
- аналитики Больших данных,
- суперкомпьютерных технологий и методов параллельных вычислений,
- организации и проведения конференций и школ.



Академик Самарский Александр Андреевич



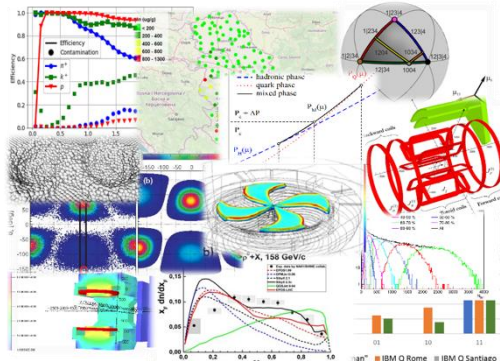
Член-корр. Говорун Николай Николаевич

Академик Тихонов Андрей Николаевич

Потребность в подготовке высококвалифицированных кадров в области математического моделирования и обработки данных проектов класса мегасайнс с применением методов аналитики Больших данных и искусственного интеллекта привела к идее создания на базе филиала МГУ в Дубне направления подготовки «Прикладная математика и информатика»

Магистерской программы

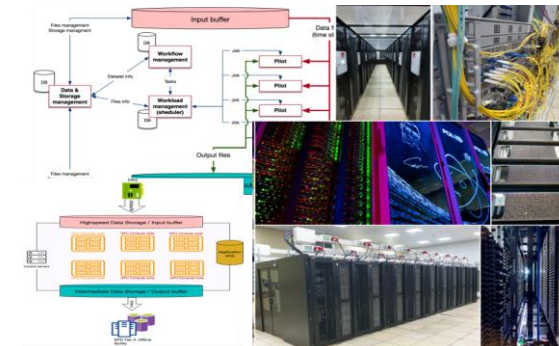
«Методы и технологии обработки данных в гетерогенных вычислительных средах»



Математическое моделирование, численные методы и комплексы программ



Глубокое машинное обучение и аналитика больших данных



Компьютинг (программные средства и модели) для проектов класса мегасайнс

Создана рабочая группа по подготовке магистерской программы, планируется создание координационного комитета совместно с ВМК МГУ

Первоначальный план организации работы консорциума RDIG-M

- формирование Координационного совета (при необходимости Управляющего Совета)
- организация работ консорциума (создание рабочих групп/коллабораций) по различным направлениям (физика высоких энергий, нейтринная программа, синхротрон-нейтронная программа, лайф-сайенс и пр.)
- организация работ консорциума по созданию сервисов, необходимых для функционирования вычислительно/сетевой инфраструктуры:
 - управление сетевой инфраструктурой (NOC)
 - единая система аутентификации и авторизации (CA)
 - интеграция, управление, распределение ресурсов (ROC, pledged resource)
 - управление распределенными хранилищами данных
 - мониторинг, статистика, учет ресурсов, (M&A)
 - надежность и доступность (R&A)
 - управление виртуальными организациями или прикладными проектами (VOMS)
 - управление tickets или запросами (GGUS, helpdesc)
 - ...
- формулировки задач консорциума в виде технического проекта RDIG-M (TDR)
- процедура включения в консорциум других организаций
- организация подготовки и переподготовки специалистов

Технический проект РДИГ-М (1)

Нужно от экспериментам

- Вычислительные модели. Модели данных. Ожидаемые ресурсы – требования к производительности
- Обработка – калибровка – повторная обработка – моделирование – распространение данных – анализ

Необходимо

- Выявление общего – выявление различий – создание глобальной картины

Вычислительная архитектура РДИГ-М

- Грид-архитектура в целом – поток данных –нескольких гридов и облаков
- Архитектура Tier 0
- Архитектура Tier 1
- Роль Tier2
- Сеть
- Эксплуатационные сервисы
- Сервисы поддержки
- Сервисы учета ресурсов и их использования

Технический проект РДИГ-М (2)

Программное обеспечение

- Операционные системы
- Промежуточное ПО
- Общие приложения и инструменты
- Поддержка физического анализа
- Базы данных – распределенное развертывание
- Поддержка жизненного цикла – управление развертыванием и управлением версиями

Технологии

- Статус и ожидаемое развитие – процессоры – компиляторы – хранилище – запоминающее устройство – сеть
- Выбор исходных решений
- Жизненный цикл оборудования

Технический проект РДИГ-М (3)

Прототипы и их оценка-на основе цифровых двойников

- Проверка работы с данных
- Проверка сервисов

Сценарий запуска

- Пилотный запуск
- Опытная эксплуатация

Ресурсы – человеческие ресурсы и материалы

- Доступные и ожидаемые
 - - в основной лаборатории и центрах Tier 0
 - - в центрах Tier 1
 - - в центрах Tier 2
 - - для разработки, внедрения и сопровождения программного обеспечения
- Требуемые финансы

Обязанности сторон – Организация управления проектом

- Соглашение на участие- организационная структура- группы по направлениям