Status and prospects of the Meshcheryakov Laboratory of Information Technologies scientific program

Olga Derenovskaya

Scientific secretary of the Meshcheryakov Laboratory of Information Technologies, JINR

History

The Laboratory of Computing Techniques and Automation of the Joint Institute for Nuclear Research in Dubna was founded in August 1966.

The main directions of the activities at the Laboratory are connected with the provision of networks, computer and information resources, as well as mathematical support of a wide range of research at JINR in high energy physics, nuclear physics, condensed matter physics, etc.

Computing is an integral part of theory, experiment, technology development

(17.09.1910 – 24.05.1994)

(18.03.1930 – 21.07.1989)

On 25 March 2021 the Committee of Plenipotentiary Representatives of the Governments of the JINR Member States **decided to name the Laboratory of Information Technologies after M. G. Meshcheryakov** for his outstanding contribution to the creation and development of the network infrastructure and the Information and Computing Complex of the Laboratory, the Institute, and the Member States.

MLIT today: Scientific IT-ecosystem

Staff: 318
Scientists: 100
Doctors of Science: 26
Candidates of Science: 59
Campus network 2x100 Gbps
Multisite network 4x100 Gbps
Telecommunication channel 3x100 Gbps
Grid Tier1 and Tier2 for global data processing
JINR Cloud computing
JINR Member States' Cloud environment
"Govorun" supercomputer

- The coordinated development of interconnected IT technologies and computational methods
- Providing the IT services necessary for the fulfillment of the JINR Topical Plan on Research and International Cooperation in an efficient manner.
- Building world-class competence in IT and computational physics.
- 24x7 support of the computing infrastructure and services.

Cooperation with All JINR Laboratories

Particle Physics and HEP

- NICA computing
- Methods and algorithms for data analysis
- Intelligent control systems
- …

Nuclear Physics

- Computations of the properties of atoms of superheavy elements
- Analysis of fine structures in the mass distribution of nuclear reaction products
- Sub-barrier fusion and fission reactions of heavy nuclei
- …

Life Science

- Information System for Radiation Biology tasks
- Analysis of Small-Angle scattering data from nanodrugs
- Environmental monitoring
- …

Information Technologies

(Scientific directions and information systems)

Theoretical Physics

- Calculations of lattice QCD
- Numerical simulation within effective theories of QCD
- Compton scattering
- …

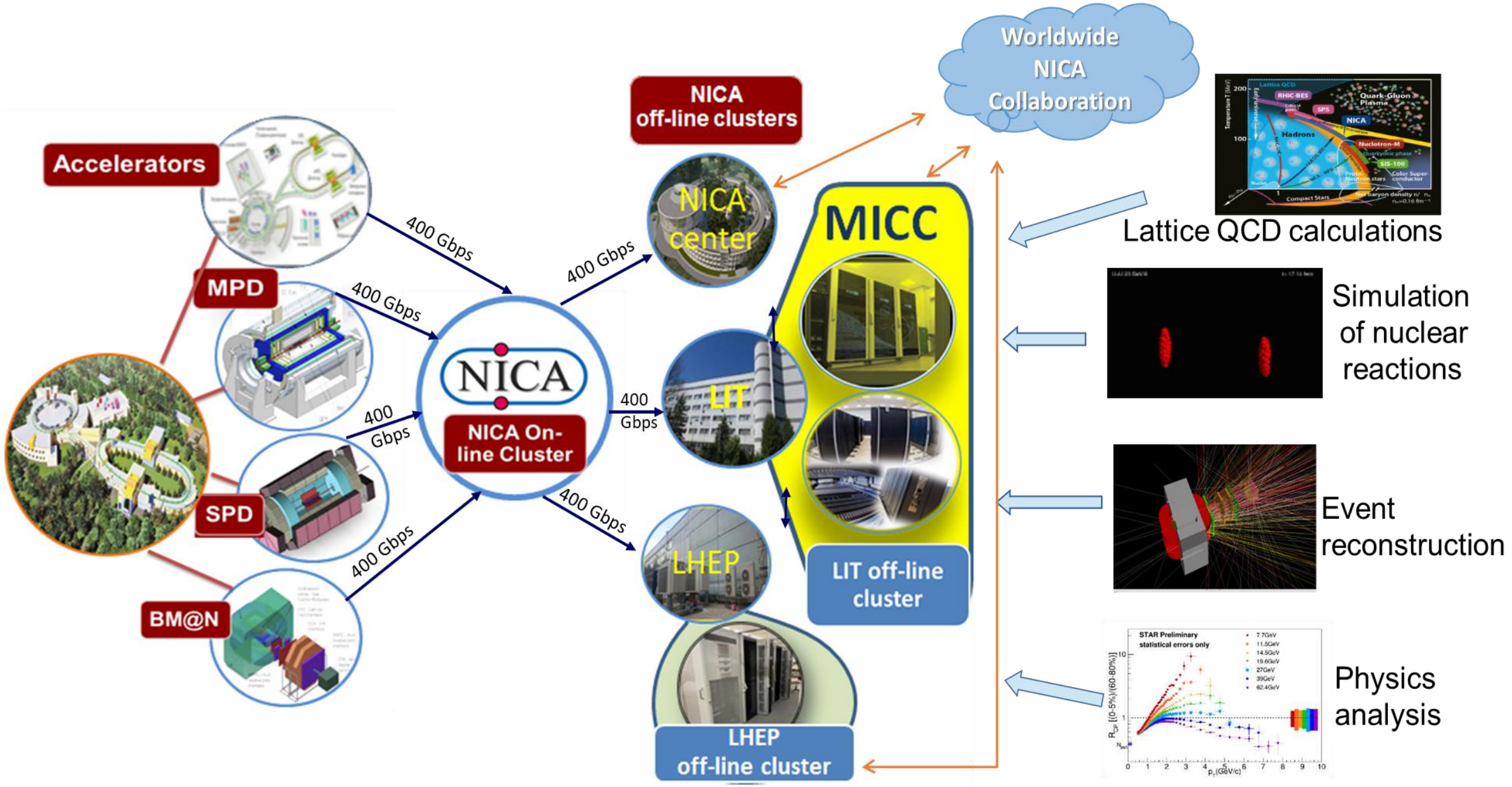
Condensed Matter

- Analysis of polydisperse populations of phospholipid vesicles
- Study of nanocomposite thin films using neutron and X-ray reflectometry methods
- Simulation of thermal processes occurring in materials
- …

Neutrino Physics and Astrophysics

- Support of the JINR neutrino program
- Data acquisition system software for Baikal-GVD
- …

NICA Computing



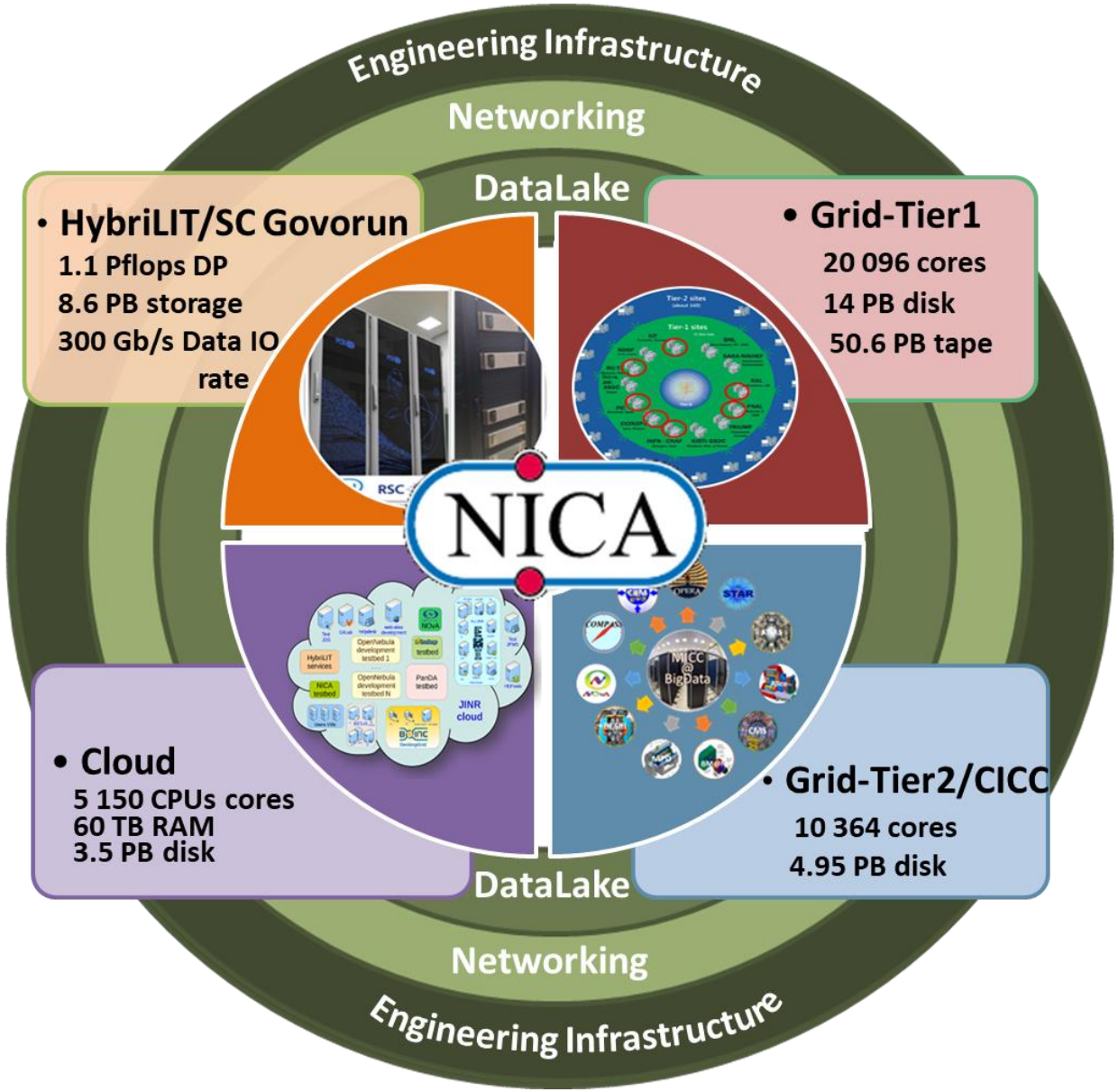
Support for the JINR Neutrino Program

MLIT contribution:
engineering infrastructure
(electricity, UPS, cooling,
network, racks, manpower)

DLNP contribution:
computing and storage
resources
(CPUs/GPUs&disks)

Computational resources for the JINR neutrino program using the cloud infrastructure of the MICC.
The NOvA, Baikal-GVD and JUNO experiments are the major users of the cloud infrastructure.

Multifunctional Information and Computing Complex at JINR



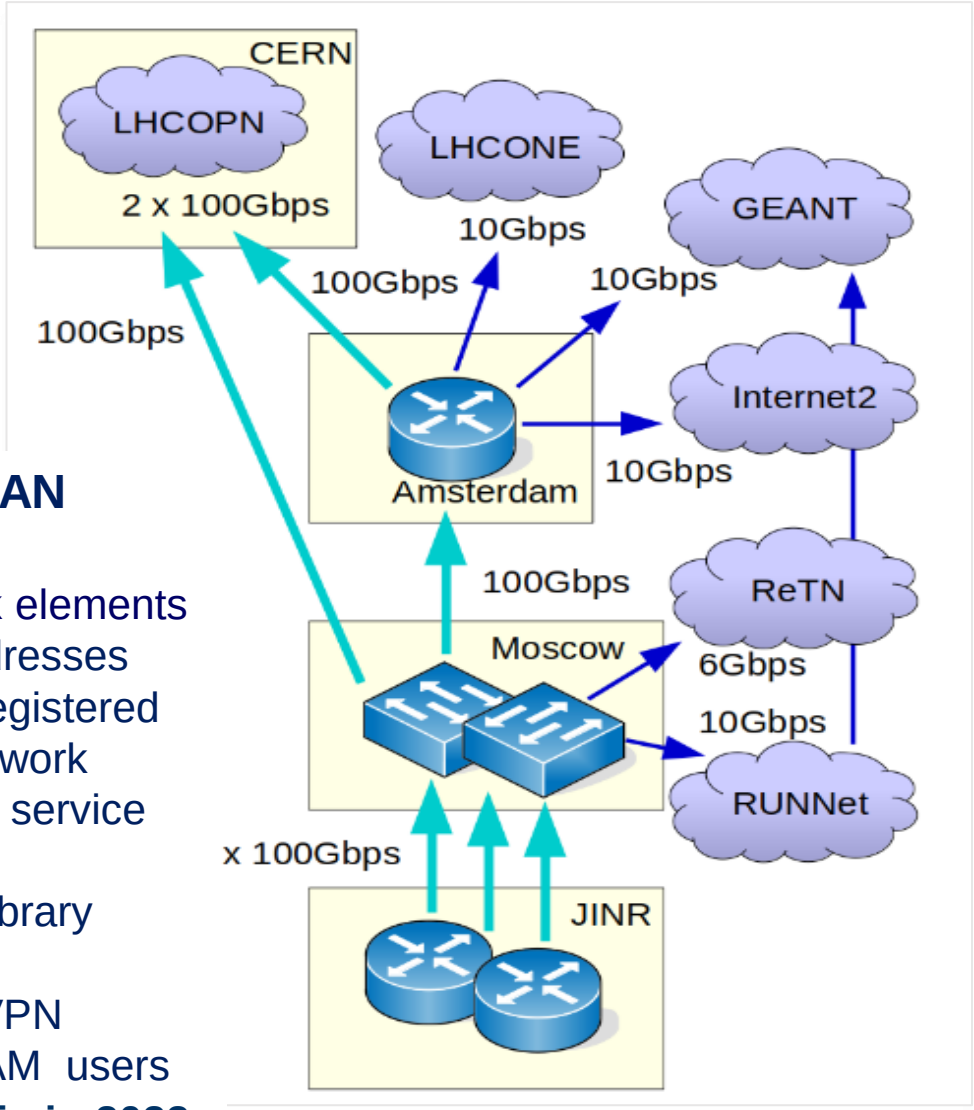
The **MICC** meets the requirements for a modern highly performant scientific computing complex:

- multi-functionality,
- high performance,
- task-adapted data storage system,
- high reliability and availability,
- information security,
- scalability,
- customized software environment for different user groups,
- high-performance telecommunications and modern local network.

The MICC should be considered as the large research infrastructure project.

micc.jinr.ru

Network Infrastructure



The JINR LAN comprises:

9291 network elements
18044 IP-addresses
6355 users registered within the network
4477 *.jinr.ru service users
1455 digital library users
837 remote VPN
111 EDUROAM users

network traffic in 2022

- 29.56 PB - input
- 34.19 PB - output

The network infrastructure is a fundamental component of the IT infrastructure of JINR and of the MICC. It provides access to the Internet, computing resources and data storage systems, as well as enables experimental data processing and computing.

MLIT ensures the reliable and fault-tolerant operation of all components of the network infrastructure:

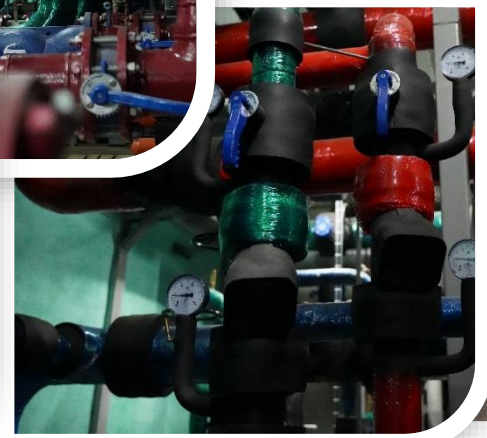
- JINR-Moscow 3x100 Gbit/s
- JINR-CERN 100 Gbit/s and JINR-Amsterdam 100 Gbit/s
- multi-site cluster network with a bandwidth of 4x100 Gbit/s for the NICA megaproject
- local area network with a bandwidth of 2x100 Gbit/s

Engineering Infrastructure









- ✓ Power supply expansion
- ✓ Cooling system for the MICC machine hall
- ✓ 100% “hot water” cooling system of the “Govorun” supercomputer
- ✓ Guaranteed power supply using diesel generators and uninterruptible power supplies

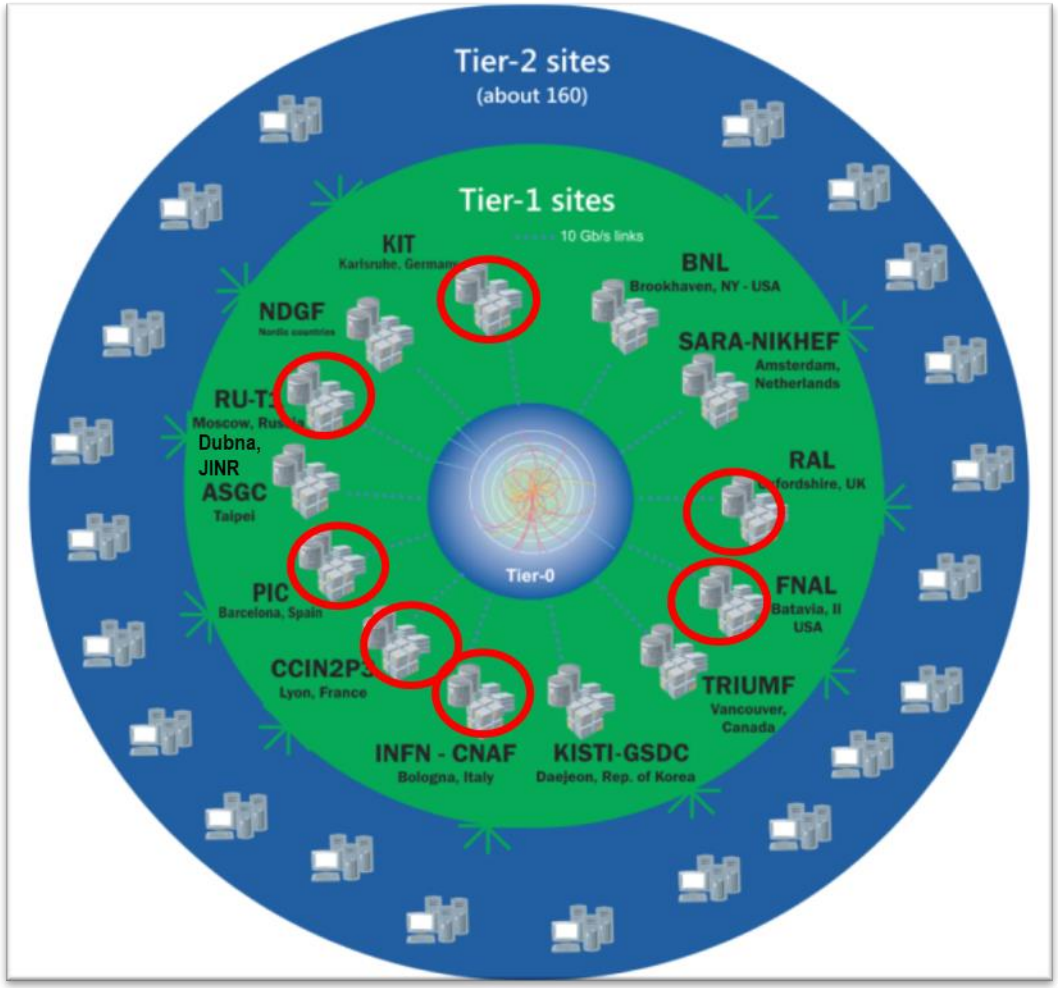
The Worldwide LHC Computing Grid

The mission of the **WLCG** is to provide global computing resources for the storage, distribution and analysis of the data generated by the LHC. **Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists**

Tier0 (CERN): data recording, reconstruction and distribution

Tier1: permanent storage, re-processing, analysis

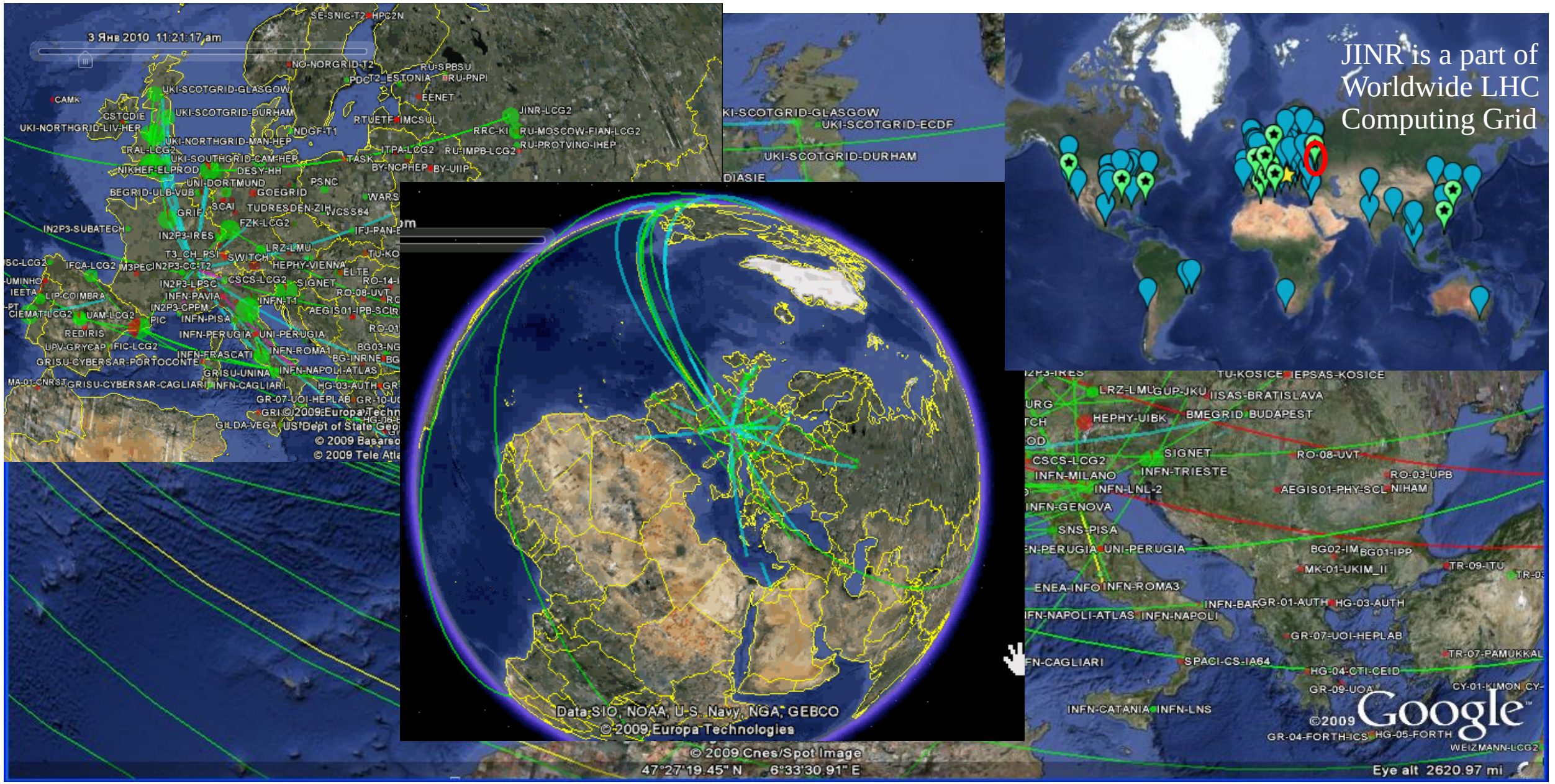
Tier2: Simulation, end-user analysis



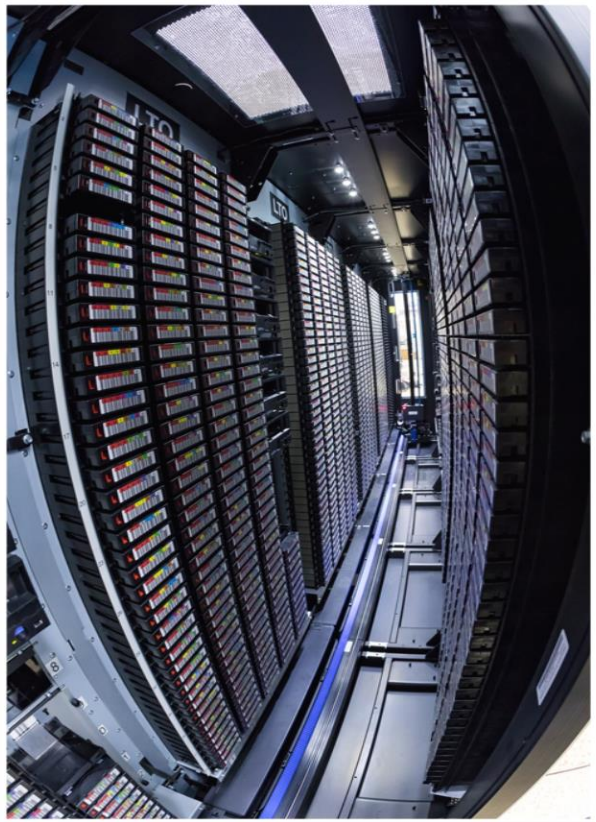
WLCG computing enabled physicists to announce the discovery of the Higgs Boson on 4 July 2012.

170 sites
42 countries
> 12k physicists
~1.4 M CPU cores
1.5 EB of storage
> 2 million jobs/day
100-250 Gb/s links

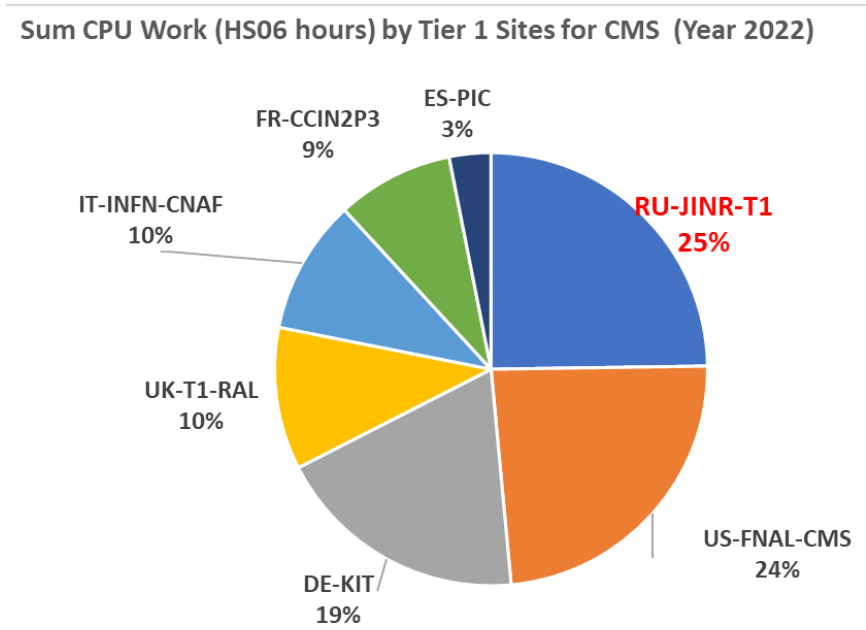
The Worldwide LHC Computing Grid (WLCG)



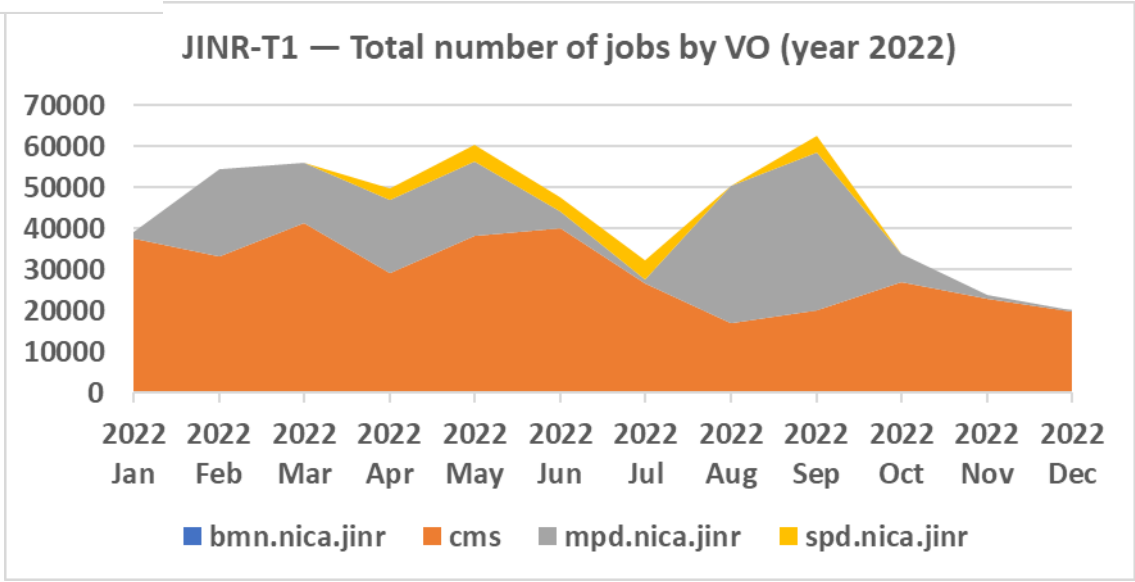
Tier1 at JINR



- 20096 cores
- 360 kHS06
- 14 PB disks
- 50.6 PB tapes
- 100% reliability and availability

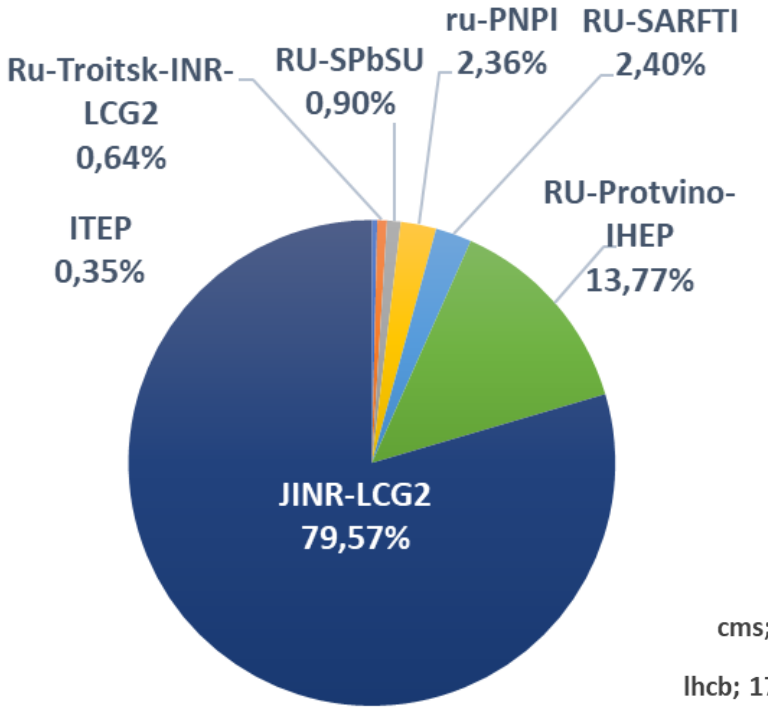


- The JINR Tier1 center has demonstrated stable work not only for CMS (LHC), but also for MPD (NICA).
- The Tier1 site for CMS is ranked first among world centers for CMS.
- 30% of all jobs executed at Tier1 JINR are NICA jobs



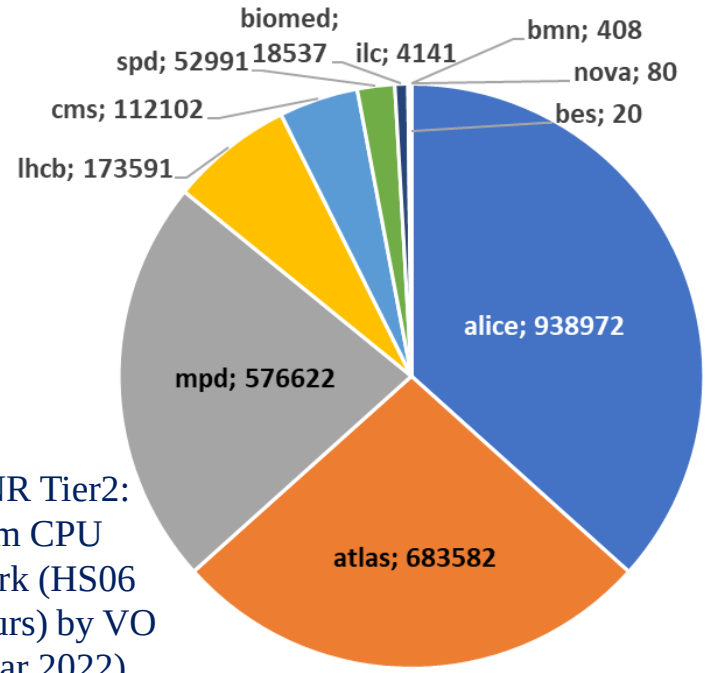
Tier2 at JINR

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 (NICA, LHC, FAIR, etc.).



RDIG: distribution by the number of jobs by websites of organizations (year 2022)

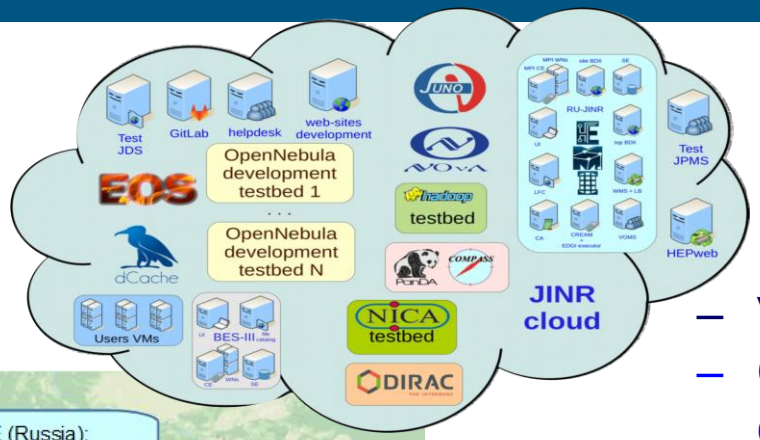
JINR Tier2 is the most productive in the Russian Data Intensive Grid (RDIG) Federation.

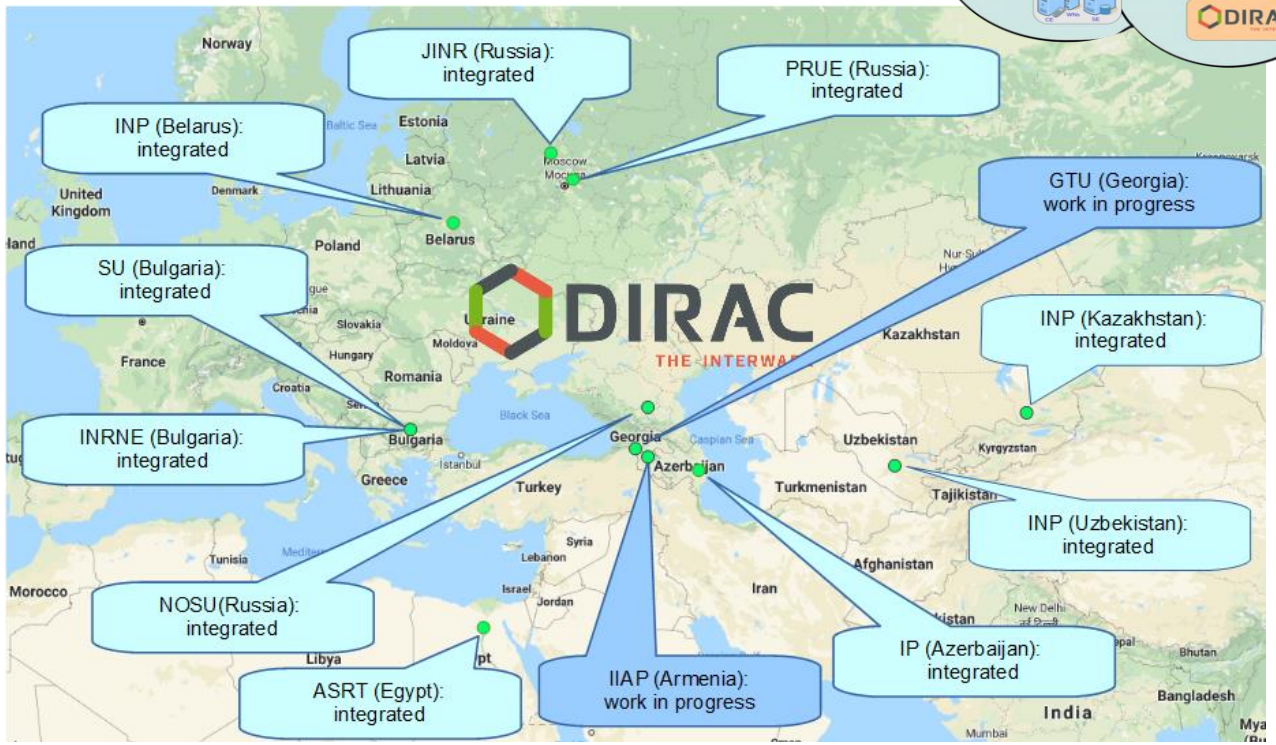


JINR Tier2: Sum CPU work (HS06 hours) by VO (year 2022)

Cloud Infrastructure

- Cloud Platform - OpenNebula
- Virtualization - KVM
- Storage (Local disks, Ceph)
- Total Resources
  ~ 5,000 CPU cores; 60 TB RAM;
  3.5 PB of raw ceph-based storage





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

- VMs for JINR users
- Computational resources for neutrino experiments

MLIT contribution: engineering infrastructure (electricity, UPS, cooling, network, racks, manpower)

DLNP contribution: computing and storage resources (CPUs/GPUs&disks)

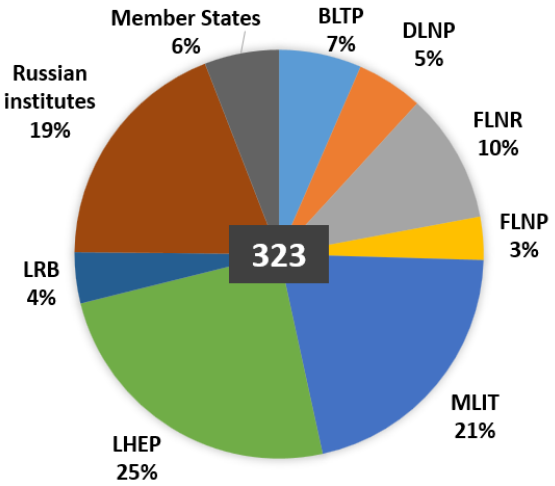
- Testbeds for research and development in IT
- COMPASS production system services
- Data management system of the UNECE ICP Vegetation
- Scientific and engineering computing
- Service for data visualization
- Gitlab and some others

“Govorun” Supercomputer

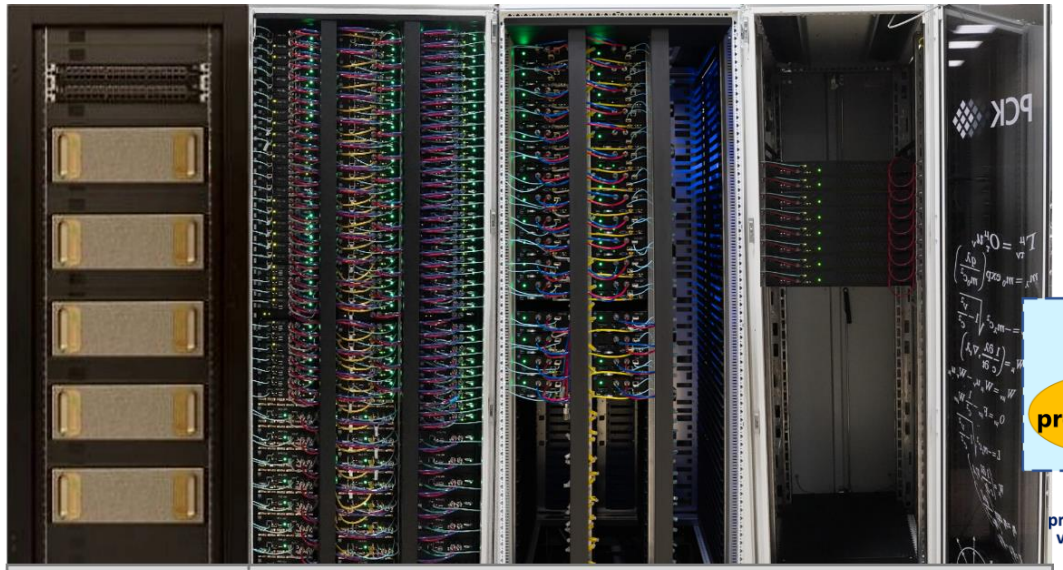
- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 1.7 PFlops DP
- GPU component based on NVIDIA
- CPU component based on RSC “Tornado” liquid cooling solutions
- The most energy-efficient center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

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

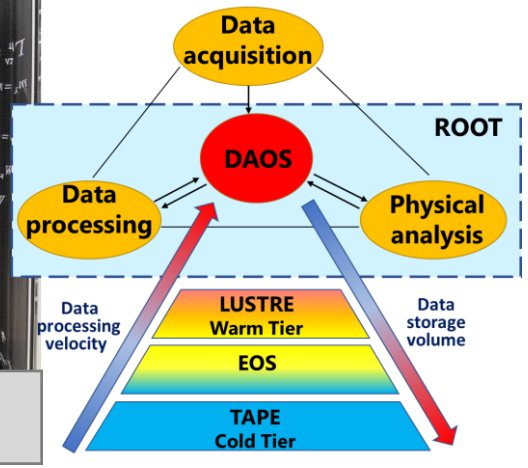
- NICA megaproject,
- calculations of lattice quantum chromodynamics,
- computations of the properties of atoms of superheavy elements,
- studies in the field of radiation biology,
- calculations of the radiation safety of JINR’s facilities.

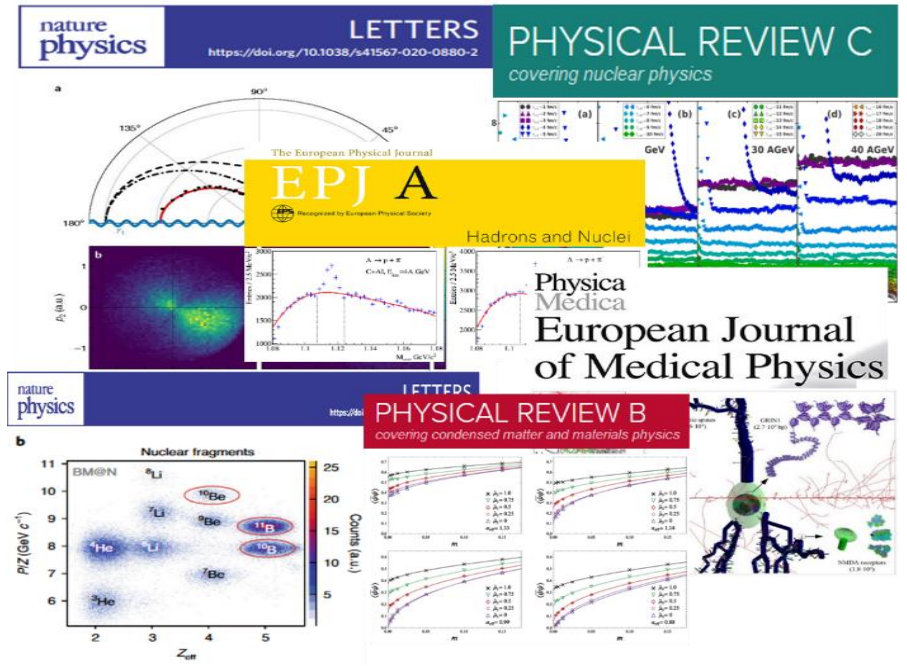


Total number of users : 323



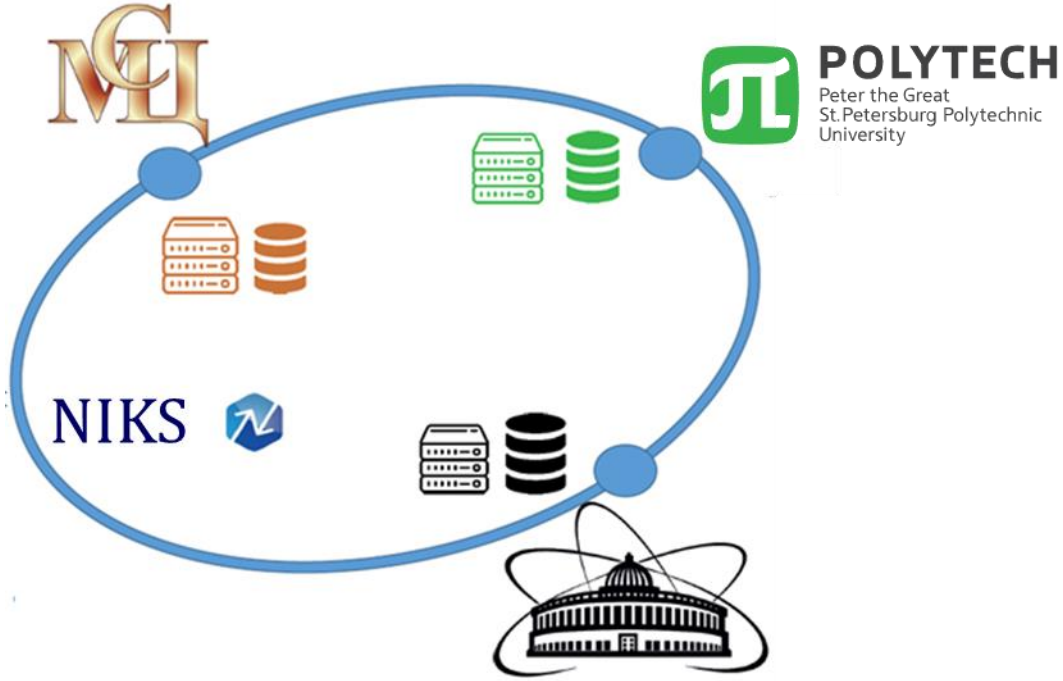
GPU-accelerator | Hyperconverged CPU and Distributed Storage Nodes





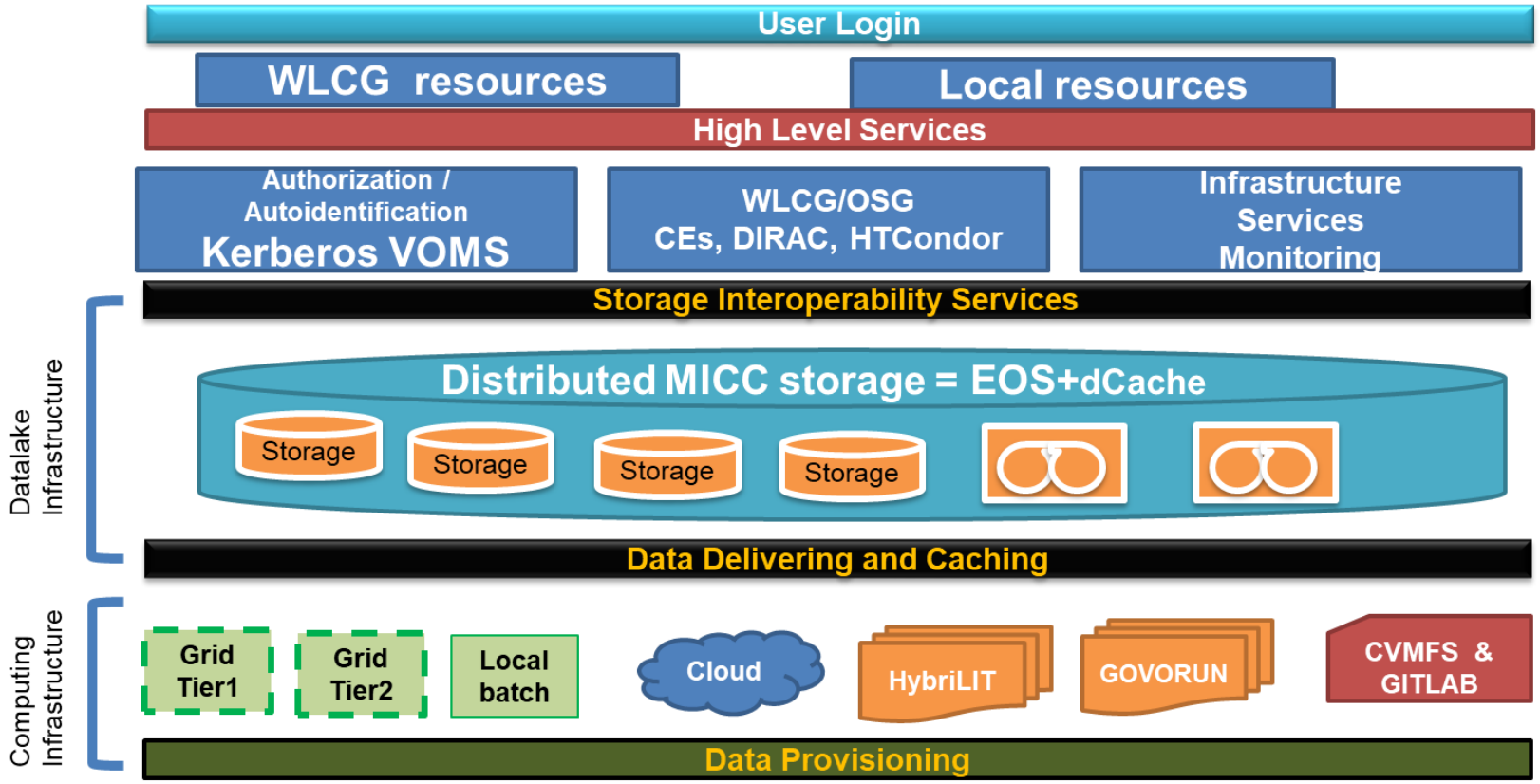
Unified Scalable Supercomputer Research Infrastructure

Based on the integration of the supercomputers of JINR, of the Interdepartmental Supercomputer Center of the Russian Academy of Sciences and of Peter the Great St. Petersburg Polytechnic University, a unified scalable supercomputer research infrastructure based on the National Research Computer Network of Russia (NIKS) was created. Such an infrastructure is in demand for the tasks of the NICA megaproject.



JINR in DataLakes

- ✓ The JINR data lake was built as a distributed EOS storage system.
- ✓ EOS is used for storing and accessing big arrays of information. It can be applied for collective data simulation, storage of raw data gathered from experimental setups, data processing and analysis.
- ✓ There is currently 17 PB of disk space available for EOS.
- ✓ Baikal-GVD, DANSS, FOBOS, JUNO, BM@N, MPD, SPD, PANDA are its major users.



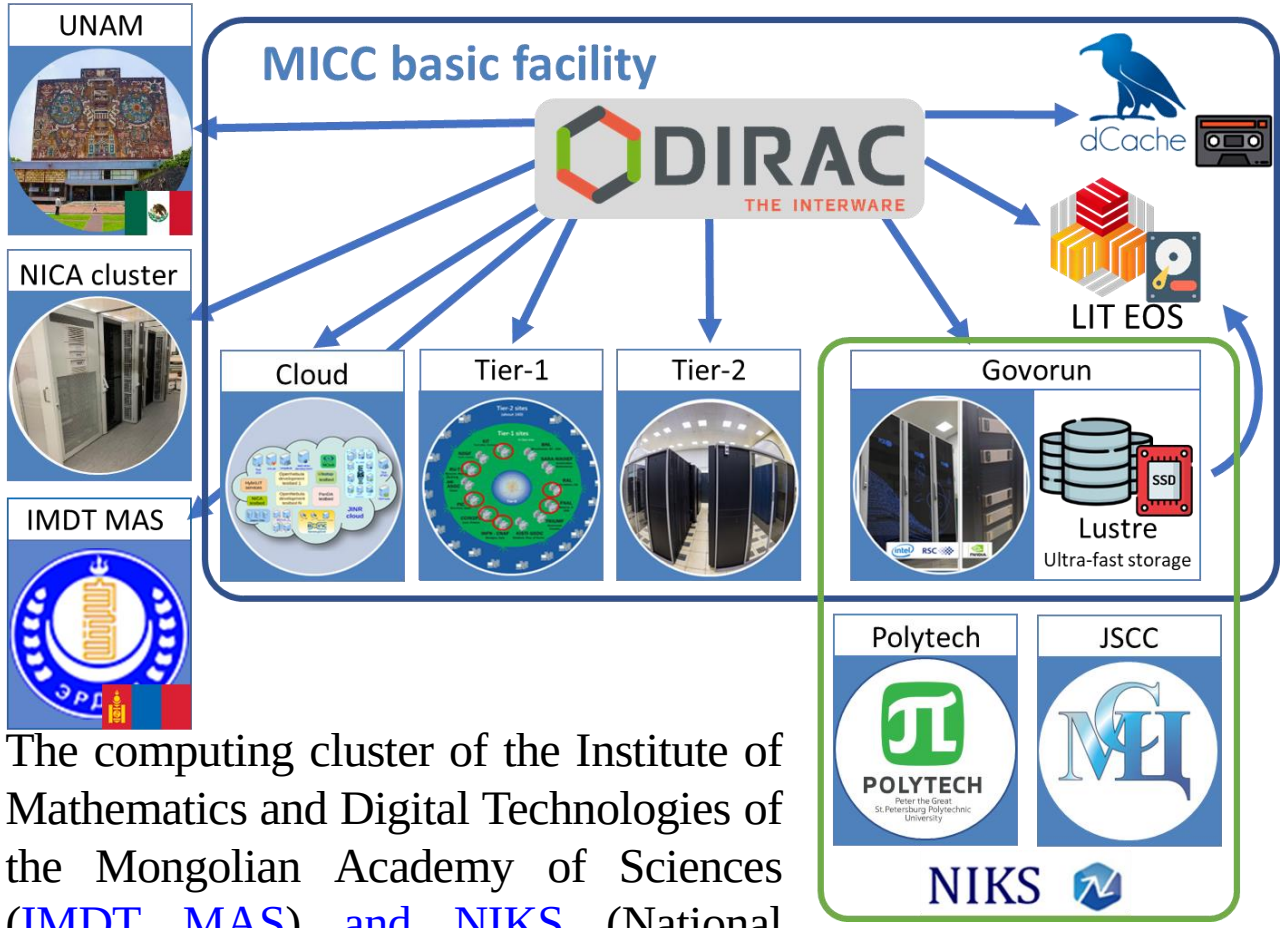
MICC Monitoring

For a robust performance of the complex it is necessary to monitor the state of all nodes and services - from the supply system to the robotized tape library.

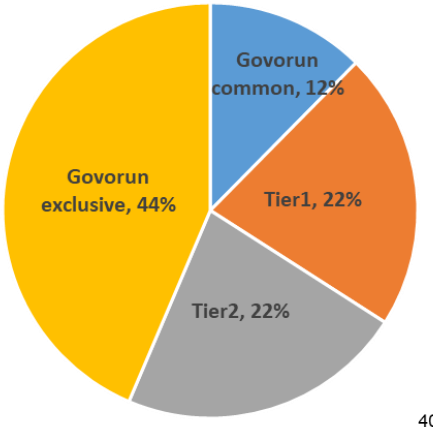
- Global real time 24x7 survey of the state of the whole computing complex
- In case of emergency, alerts are sent to users via e-mail, SMS, etc.
- ~ 1500 elements are under observation


MICC Operational Center

DIRAC-based distributed heterogeneous environment for MPD tasks

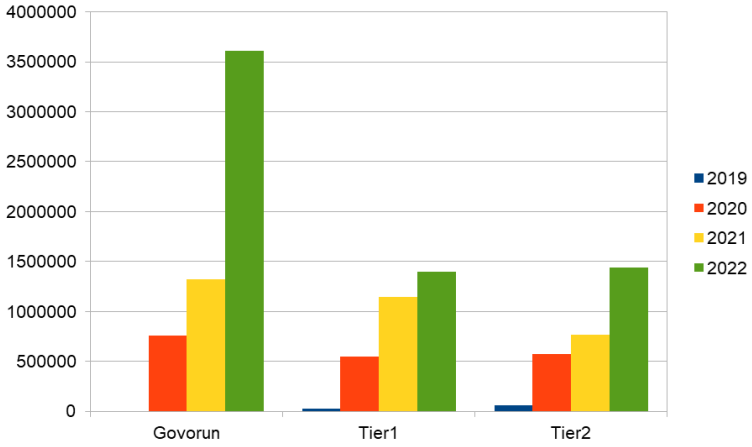


The computing cluster of the Institute of Mathematics and Digital Technologies of the Mongolian Academy of Sciences (IMDT MAS) and NIKS (National Research Computer Network, the Russia's largest research and education network) were integrated into the heterogeneous distributed environment based on the DIRAC platform.



Share of the use of different MICC components for MPD tasks in 2022: the "Govorun" SC resources are the most efficient for MPD tasks.

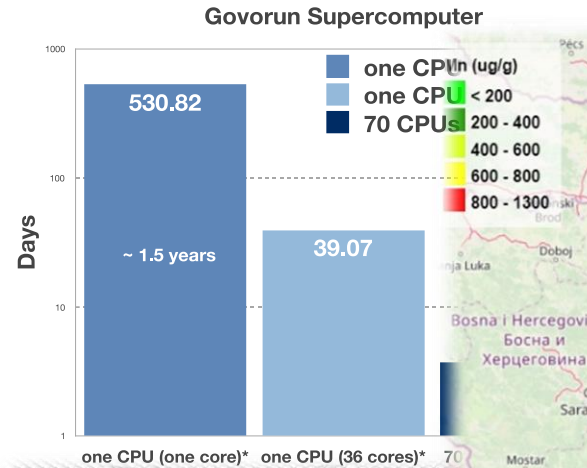
Increase in the share of the MICC computing resources on the DIRAC platform in normalized CPU HEP-SPEC06 days

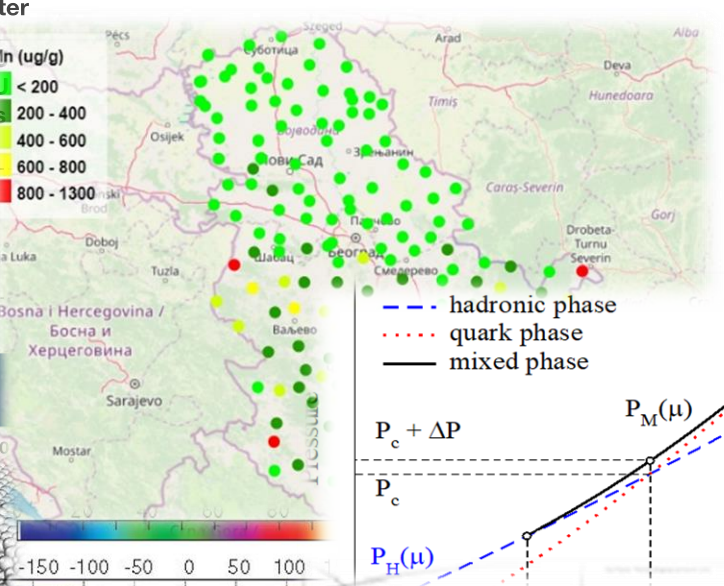


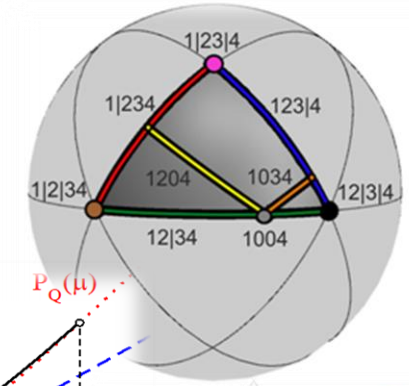
Summary statistics of using the DIRAC platform for MPD tasks in 2019-2022

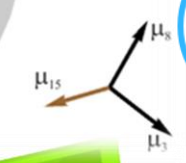
31	1.283 B	440M	1.5M	1557 years	1.3 PB
data generation campaigns	events generated	events reconstructed	jobs completed	total computation time	MPD data produced

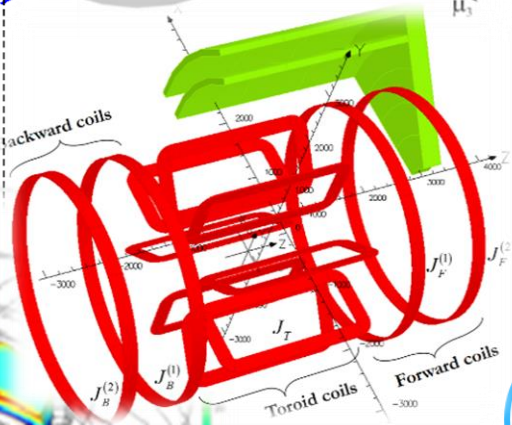
Methods, Algorithms and Software

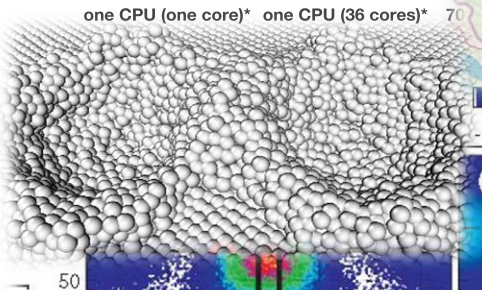


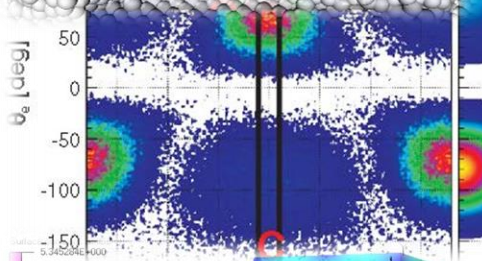


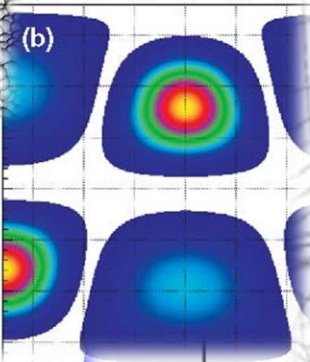


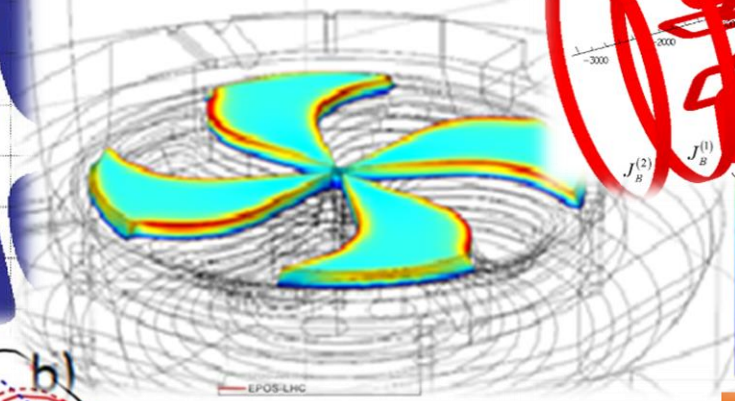


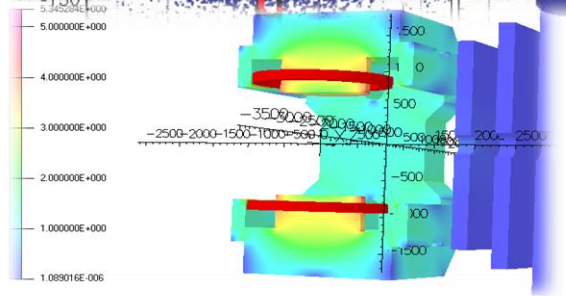


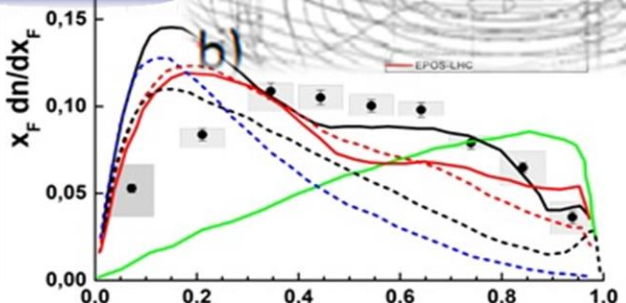


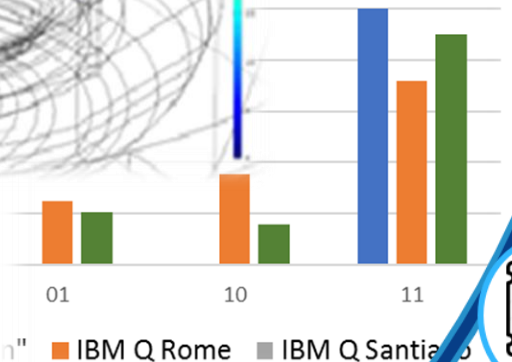












- Numerical modeling of complex physical systems
- Experimental data processing and analysis
- Big Data
- Machine and Deep learning
- AI and robotics
- Computer algebra
- Quantum computing

Mathematical Methods and Software for Experimental Data Processing and Analysis

✓ Physical processes modeling

- event simulations
- GEANT-simulation of experimental setups

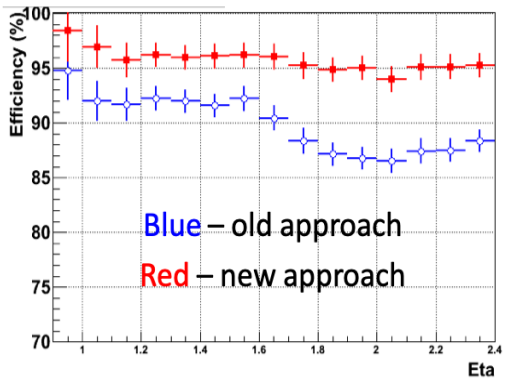
✓ Event reconstruction & data analysis

- particle trajectory reconstruction
- particle identification
- physical processes reconstruction
- data analysis

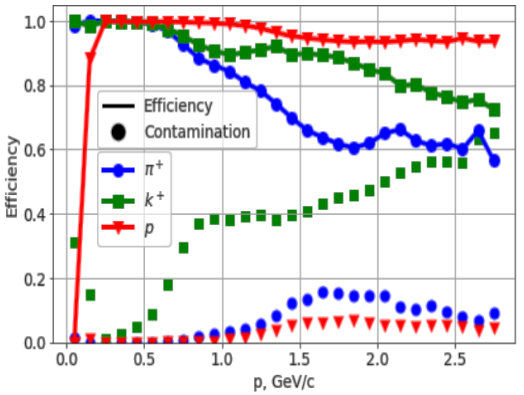
✓ Applied software and Data Bases

- DBs for experimental services
- experimental software frameworks
- data modeling and data processing
- event visualization and monitoring

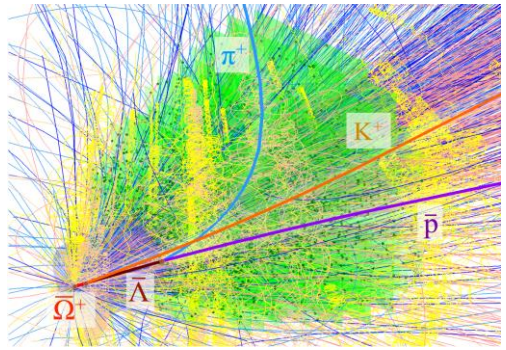
The team members published over 100 specific papers during the last 5 years. They co-authored over 500 papers as members of the international collaborations BM@N, MPD, Baikal-GVD, CMS, ATLAS, CBM, etc.



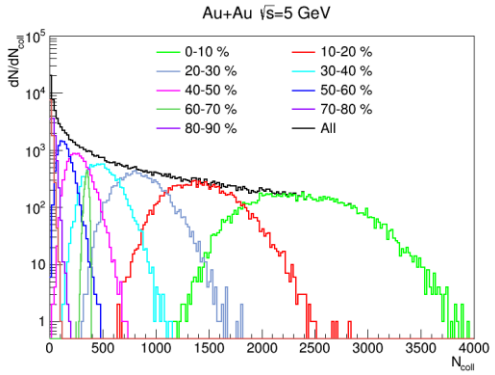
Effective algorithm for TeV muons reconstruction in CMS



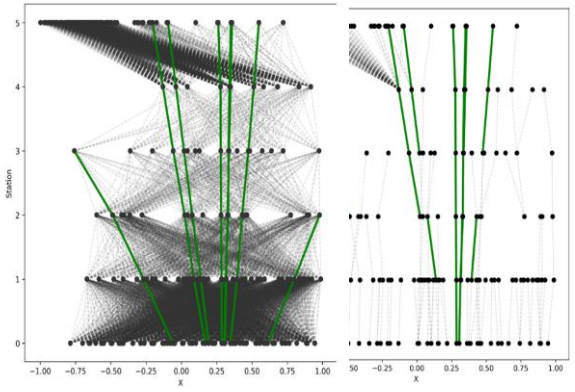
Gradient-boosted decision trees for PID in MPD



Missing mass method for the reconstruction of strange particles in CBM (FAIR) and STAR (BNL)



Monte-Carlo Generator DCM-QGSM-SMM for NICA



Graph Neural Networks for Tracking

<u>Experiments DataBases:</u>

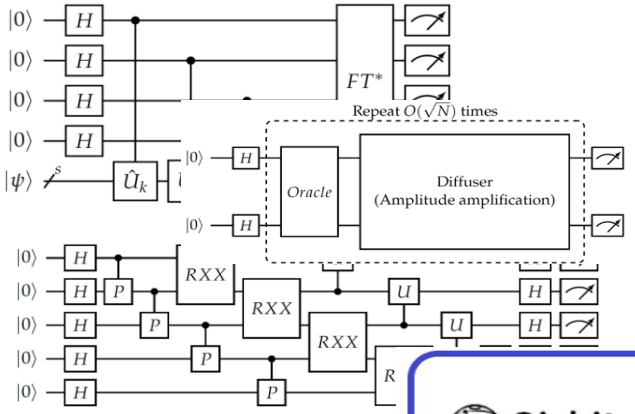
ATLAS EventPickingService,

ATLAS CREST,

Geometry and Configuration DBs for BM@N

Quantum Computing and Quantum Algorithms

Software quantum simulators for computing on computers of a classical architecture using CPUs and GPUs is of particular interest for solving a number of problems in condensed matter, high-energy physics, quantum chemistry, AI, etc.



Quantum algorithms

Quantum simulators

Qiskit
QuEST
qsim
CUQUANTUM

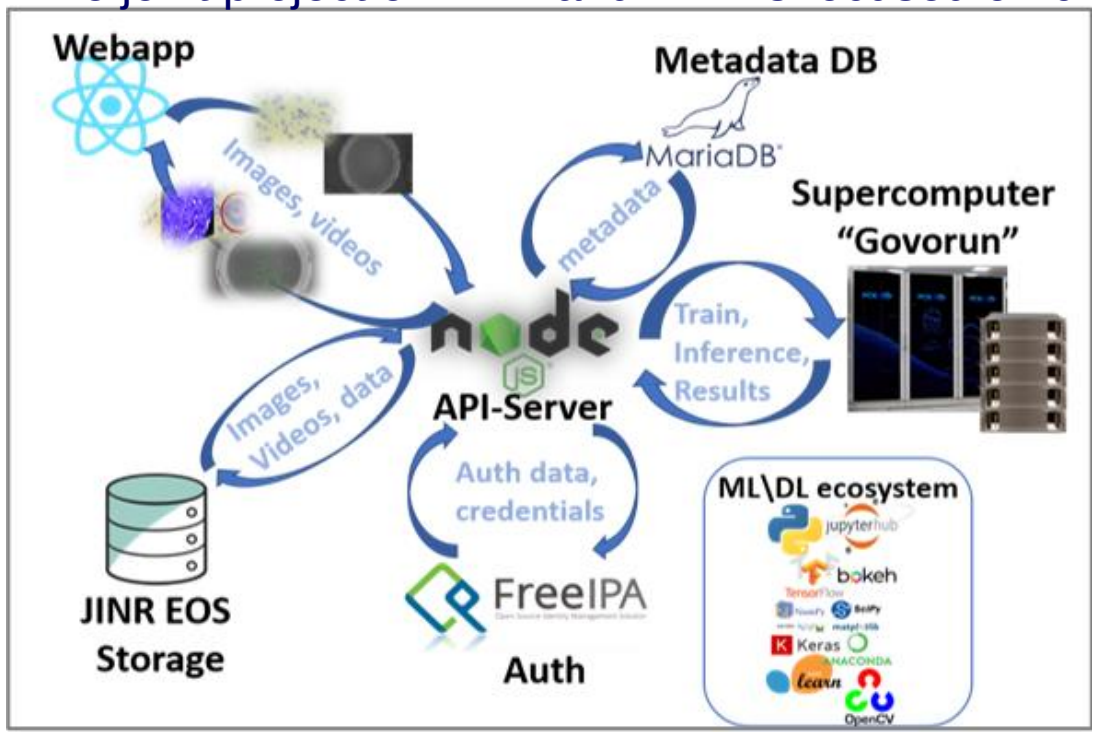
SC “Govorun”

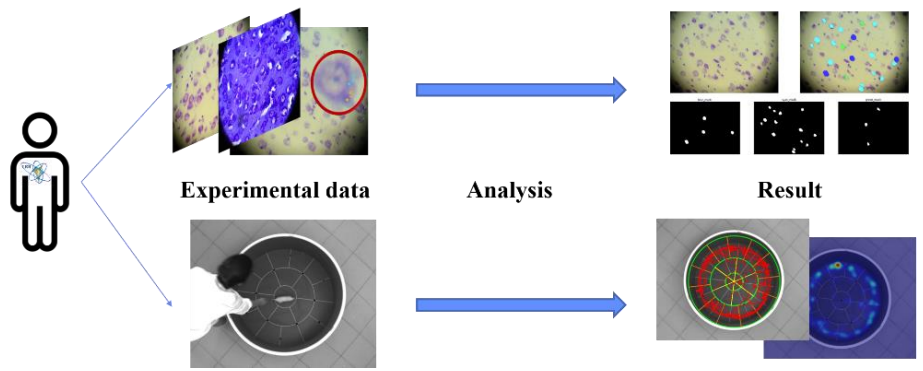
TASKS

- Form a list of QAs required to solve tasks within the studied physical models
- Select the type of quantum simulator to simulate a classical architecture on computers
- Define resources for the selected quantum-limiting capabilities of available computing simulators (number of qubits and computation time)
- Search for exact solutions to urgent problems of quantum chemistry and study the chemical properties of heavy elements

Information System for Radiation Biology Tasks

The joint project of MLIT and LRB is focused on creating an Information System (IS) as a set of IT solutions.



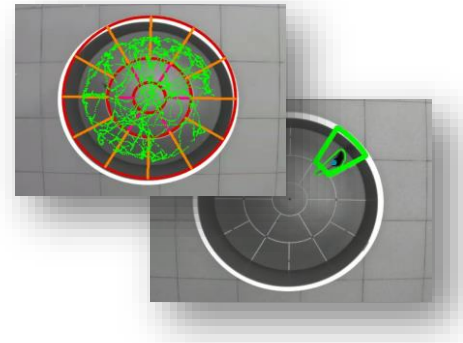


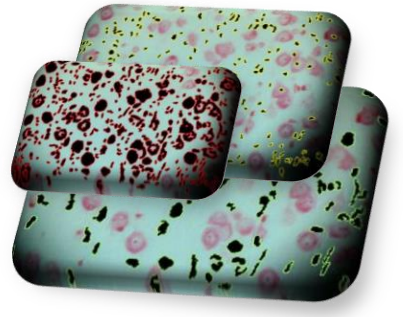
Conceptual scheme of the service

The information system allows one to store, quickly access and process data from experiments at LRB using a stack of neural network and classical algorithms of computer vision, providing a wide range of possibilities for automating routine tasks. It gives an increase in productivity, quality and speed of obtaining results.

Tasks of the IS algorithmic block

- Analysis of the experimental field markup
- Tracking the position of the animal as part of the experiment
- Classification and determination of the type of animal activity (grooming, fading, etc)
- Segmentation of neurons in images of histological slices
- Classification of neurons by type and belonging to the layer
- Statistical analysis of behavioral patterns and correlations with pathomorphological analysis





JINR Digital EcoSystem

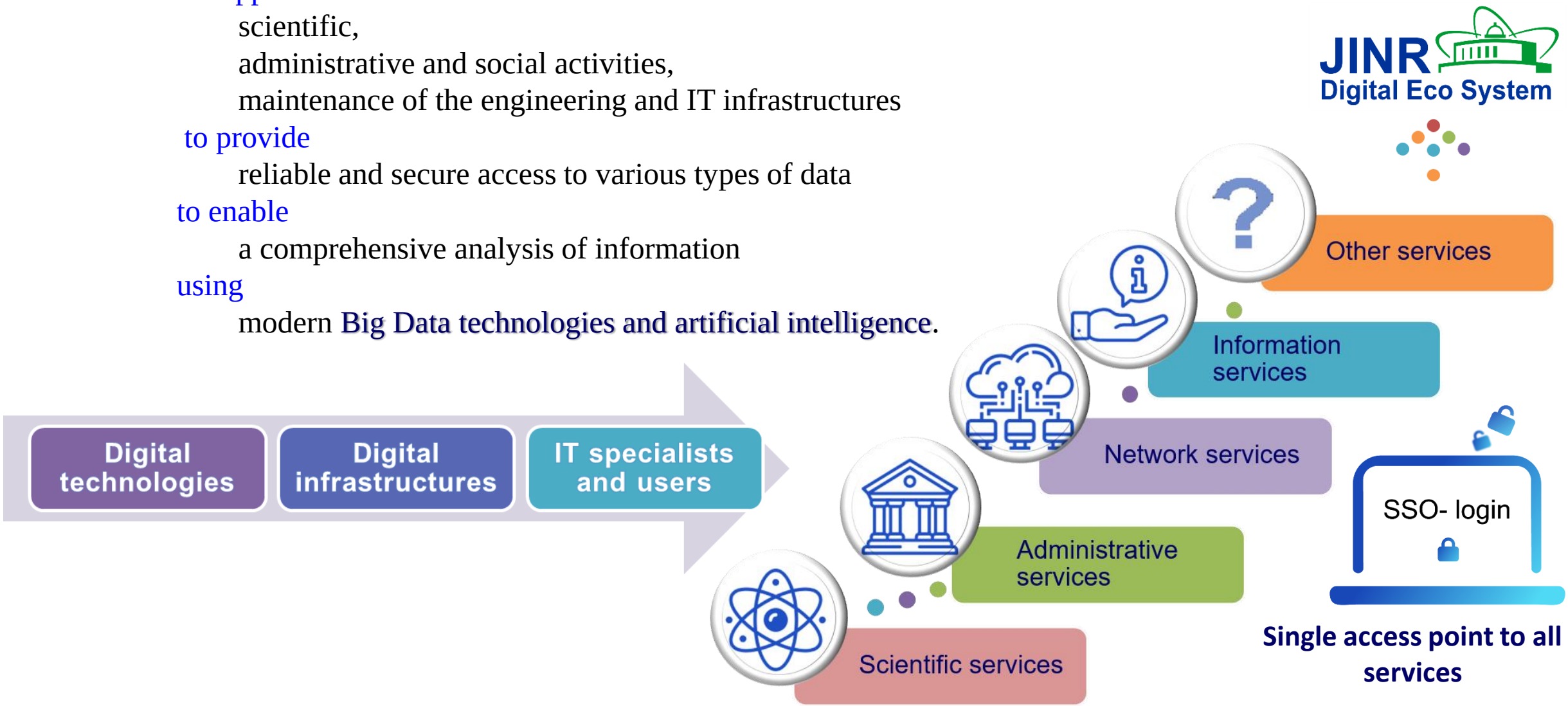
The digital platform “JINR Digital EcoSystem ” integrates existing and future services

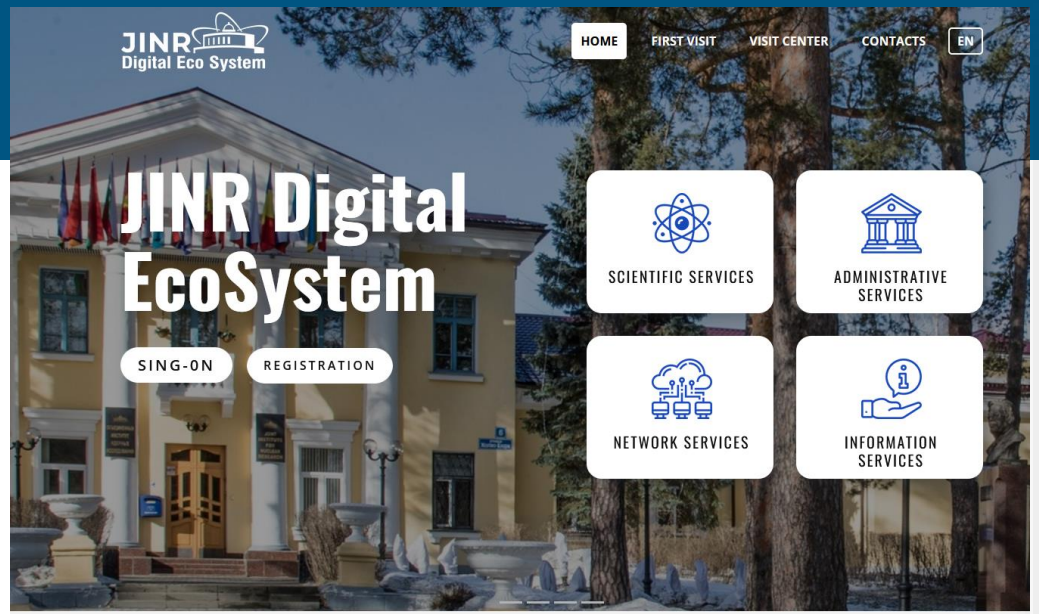
to support
scientific,
administrative and social activities,
maintenance of the engineering and IT infrastructures

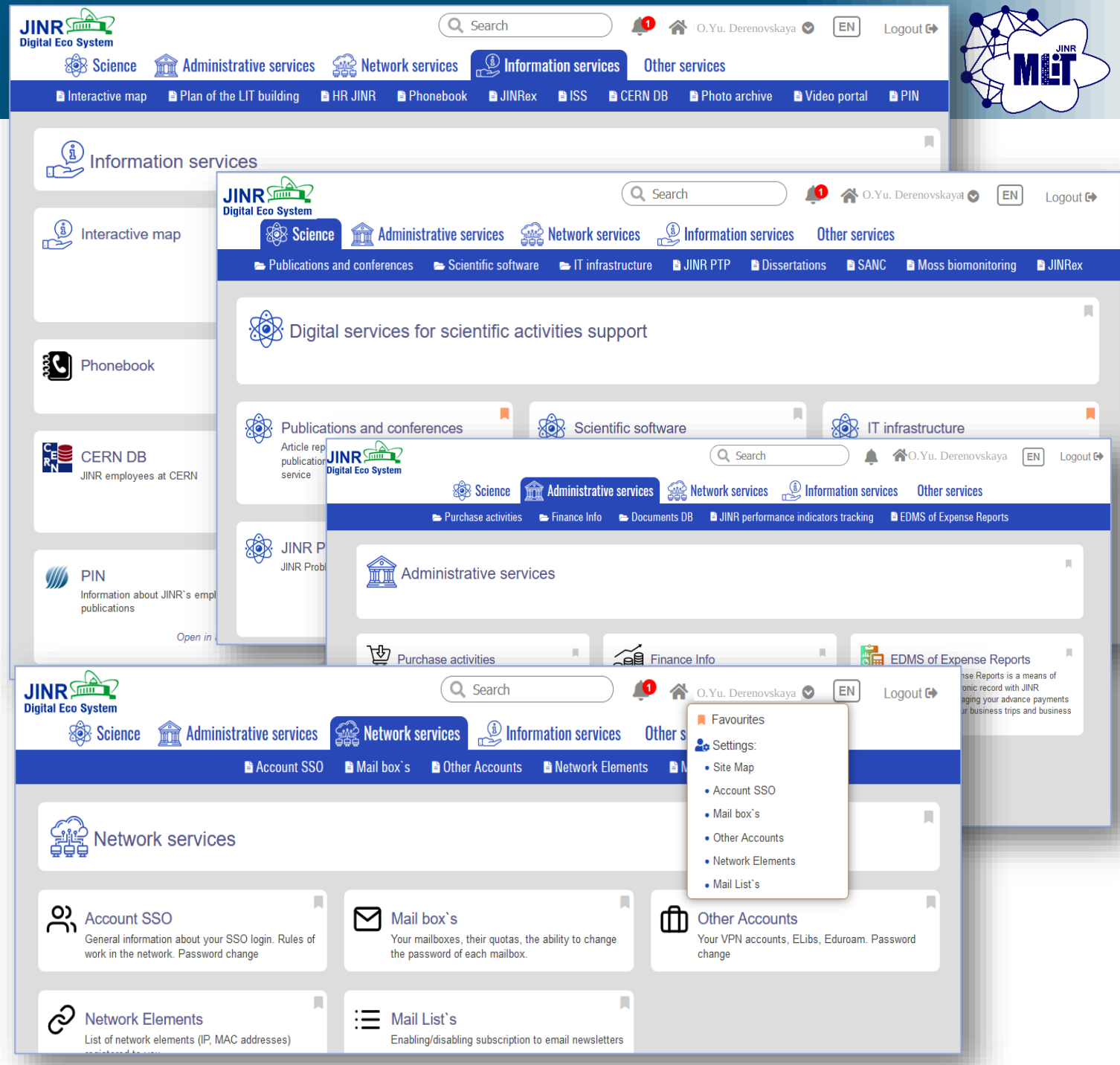
to provide
reliable and secure access to various types of data

to enable
a comprehensive analysis of information

using
modern Big Data technologies and artificial intelligence.







- ✓ Personal account of a JINR employee
- ✓ Notifications from departments and services in a personal account
- ✓ Responsive interface, customizable by the user
- ✓ Easy access, convenient navigation and search for information on a large-scale network of a wide variety of JINR services

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

OpenMP

MPI

OpenCL

NVIDIA CUDA

Tools for debugging and profiling parallel applications

Intel Cluster Studio

Work with applied software packages

COMSOL MULTIPHYSICS

Wolfram Mathematica

ROOT Data Analysis Framework

MATLAB

GEANT4

Maple

Frameworks and tools for ML/DL tasks

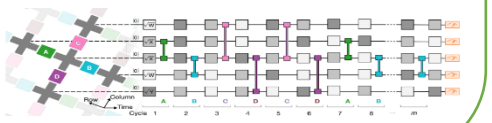
jupyterhub

TensorFlow

NumPy

scikit learn

Quantum algorithms, quantum programming and quantum control



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

- Distributed computing systems
- Computing for MegaScience Projects
- Distributed computing applications
- Data Management, Organisation and Access
- HPC
- Virtualization
- Big data Analytics and Machine learning
- Research infrastructure

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

MMCP

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.

MLIT Schools

10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)

Website: grid2023.jinr.ru

MIXED format

Conference languages – Russian and English

During the conference, there will be held the

- **workshop "Issues of organizing the Center for High-Performance Computing of the SRF- "SKIF",**
- **workshop "Computing for radiobiology and medicine",**
- **workshop "Modern approaches to the modeling of research reactors, creation of the "digital twins" of complex systems" (4-5 July),**
- **round table "RDIG-M - Russian distributed infrastructure for large-scale scientific projects in Russia".**

Conference Topics:

Research Computer Infrastructure

1. Distributed Computing Systems – technologies, architectures, models, operation and optimization, middleware and services.

2. HPC – supercomputers, CPU architectures, GPU, FPGA.

3. Cloud Technologies – cloud computing, virtualization technologies, automation of deployment of software infrastructure and applications.

4. Distributed Storage Systems

5. Distributed Computing and HPC Applications in science, education, industry and business, open data.

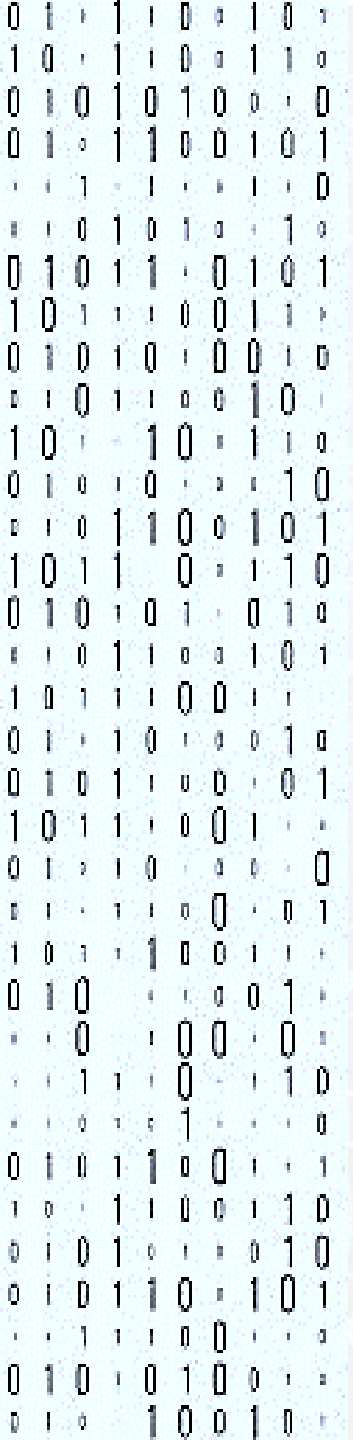
6. Computing for MegaScience Projects

Computing Science Trends

7. Quantum informatics and computing – information processing, machine learning, communication, program engineering and robotics, simulation of quantum algorithms.

8. Big Data, Machine Learning and Artificial Intelligence – big data analytics and platforms, neural networks and deep learning, intelligent control systems, decision intelligence tools and recommendation systems.

9. e-Learning – e-Learning tools, virtual labs, EdTech and HR Tech, human assets management and development systems.



Thank you for attention!