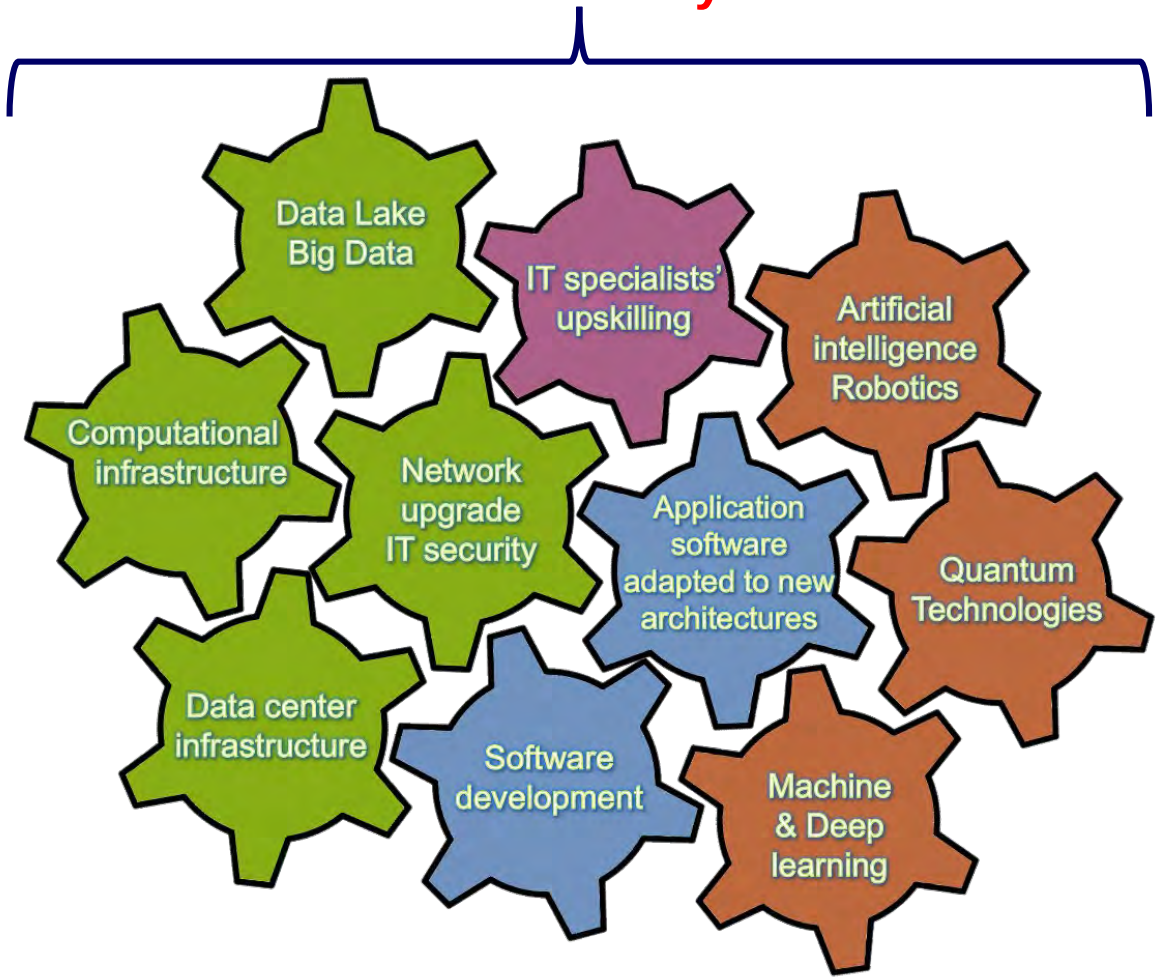
# Status and prospects of the Meshcheryakov Laboratory of Information Technologies scientific program


**Olga Derenovskaya**

Scientific secretary of the Meshcheryakov Laboratory of Information Technologies, JINR

# 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.

# Cooperation with All JINR Laboratories

## Information Technologies
(Scientific directions and information systems)

### 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
- …

### 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


400 Gbps

400 Gbps

400 Gbps

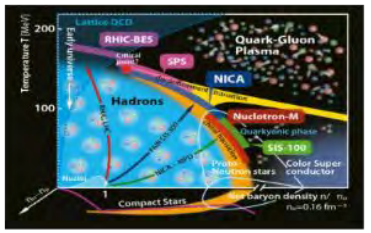
400 Gbps


400 Gbps

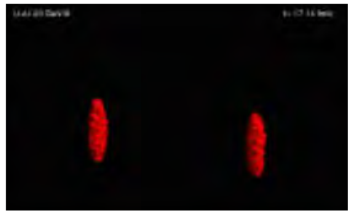
400 Gbps

400 Gbps

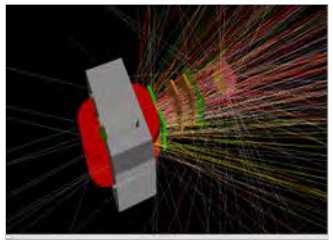




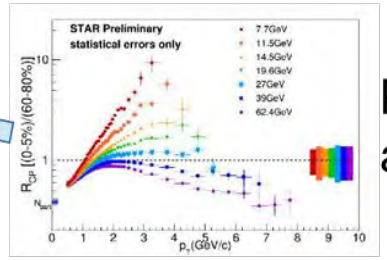

Lattice QCD calculations



Simulation of nuclear reactions



Event reconstruction




Physics analysis

# Support for the JINR Neutrino Program

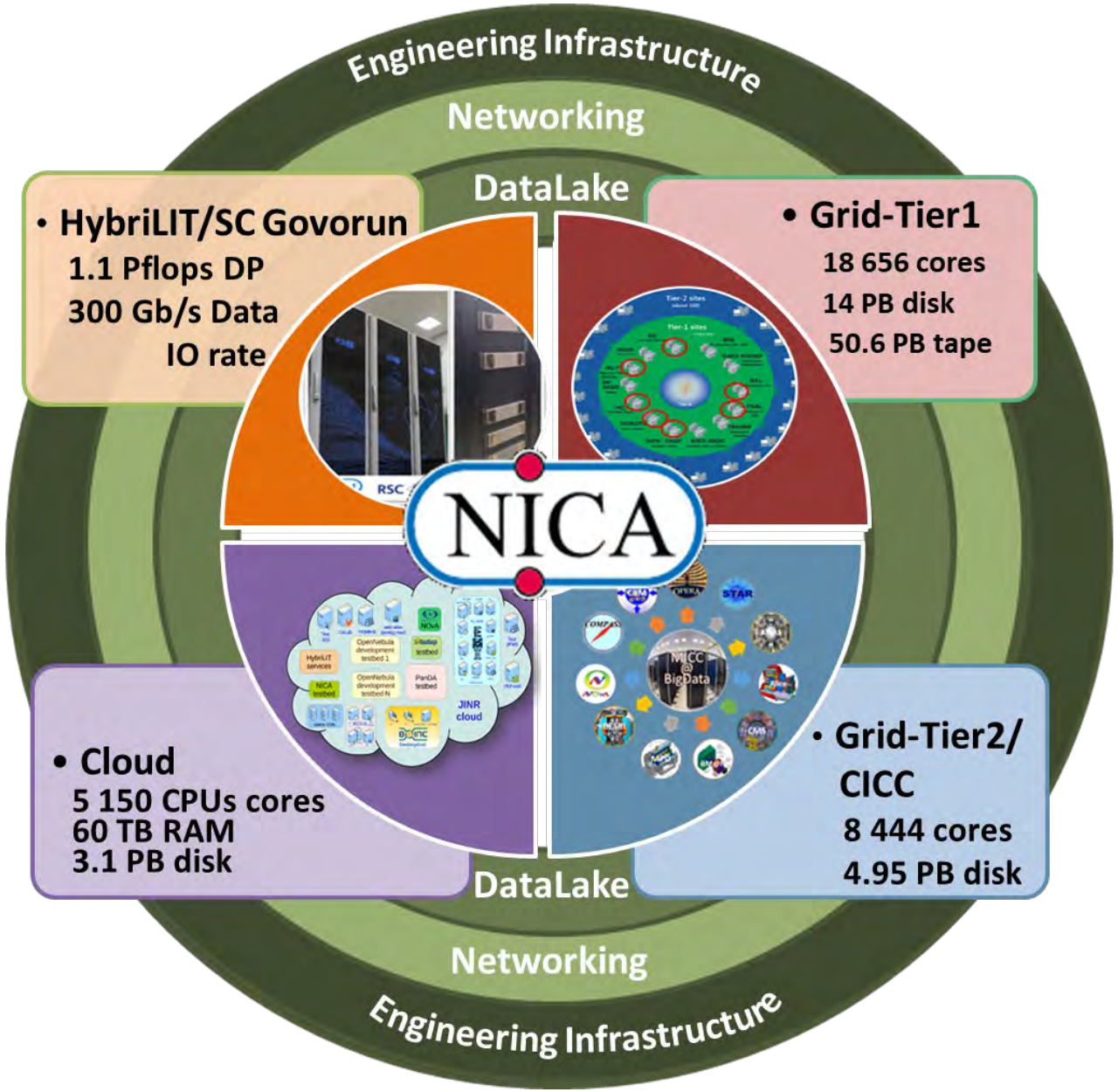

MLIT contribution:
engineering infrastructure
(electricity, UPS, cooling,
network, racks, manpower)

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

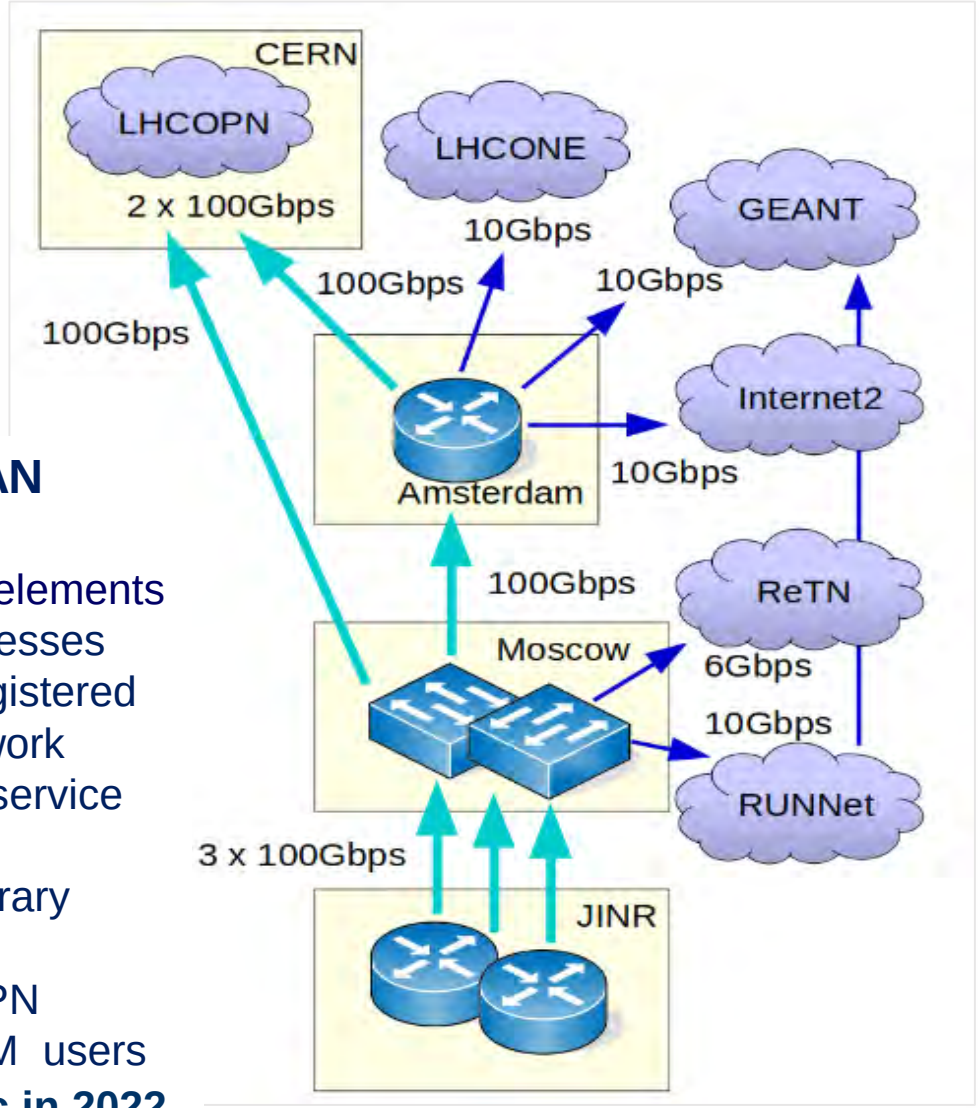



The **MICC** meets the requirements for a modern highly performant scientific computing complex:

- 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 a basic scientific infrastructure project.**

# 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

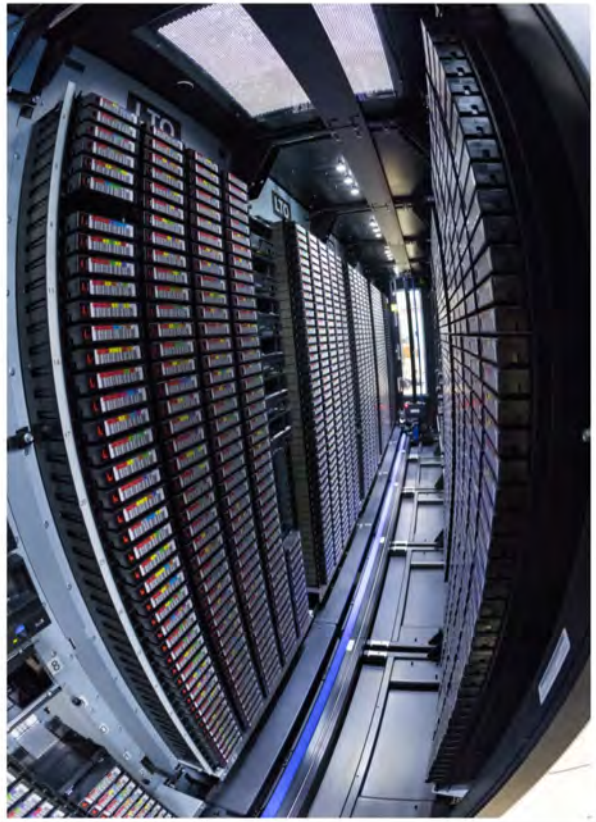
**network traffic in 2022**

- 29.56 PB - input
- 34.19 PB - output

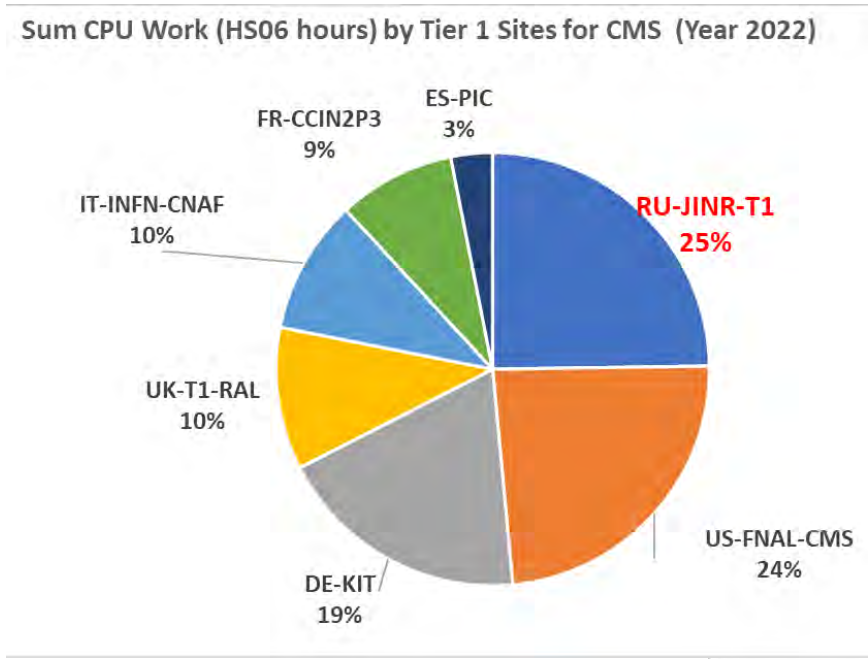
The network infrastructure is a fundamental component of the IT infrastructure of JINR and of the MICC. It provides access to the Internet, computing resources and data storage systems, as well as enables experimental data processing and computing. MLIT ensures the reliable and fault-tolerant operation 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 Gb/s

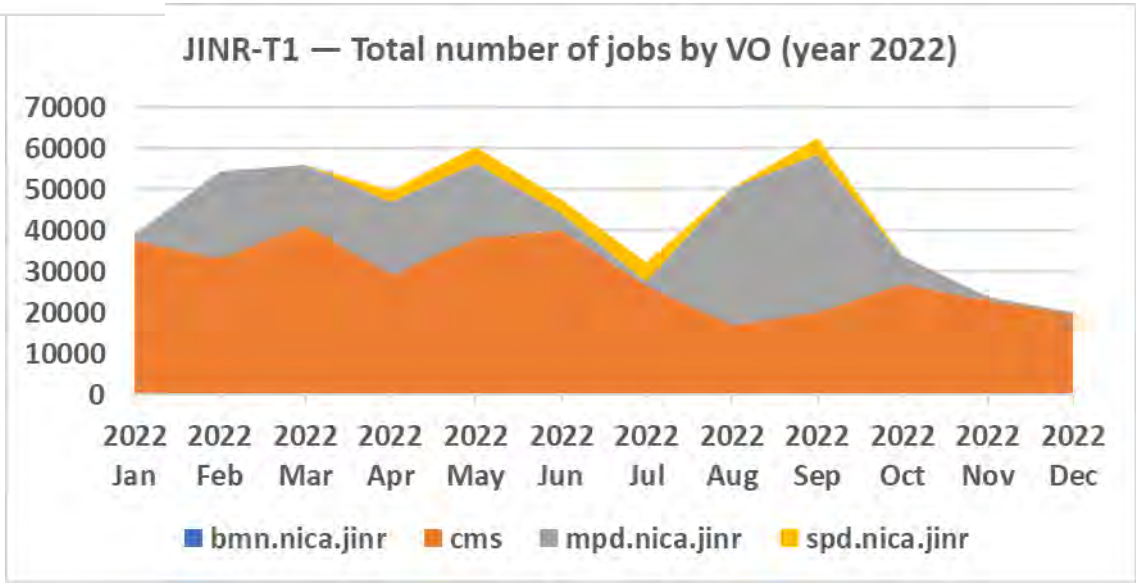
# Tier1 at JINR



- 18656 cores
- 260 kHS06
- 14 PB disks
- 50.6 PB tapes
- 100% reliability and availability

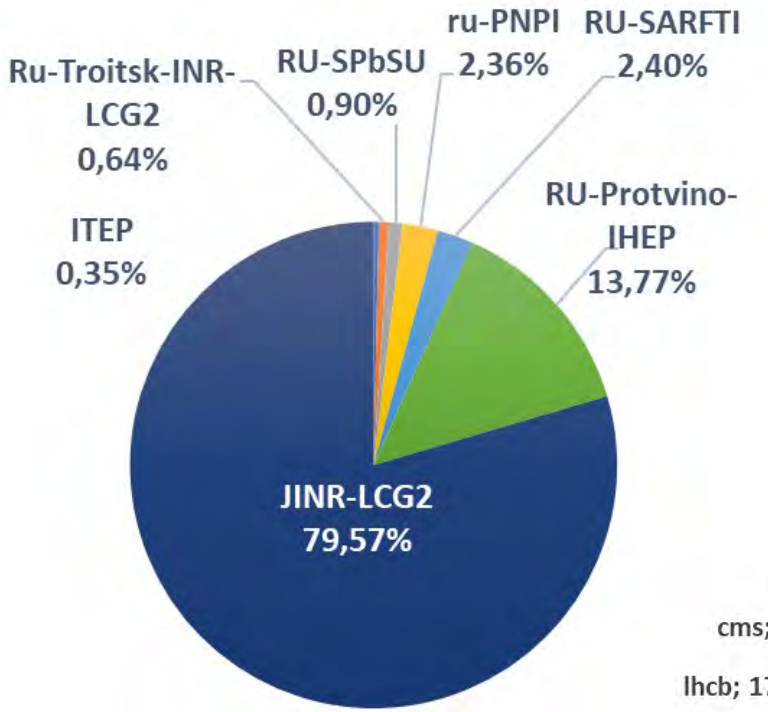



- The JINR Tier1 center has demonstrated stable work not only for CMS (LHC), but also for MPD (NICA).
- The Tier1 site for CMS is ranked first among world centers for CMS.
- 30% of all jobs executed at Tier1 JINR are NICA jobs



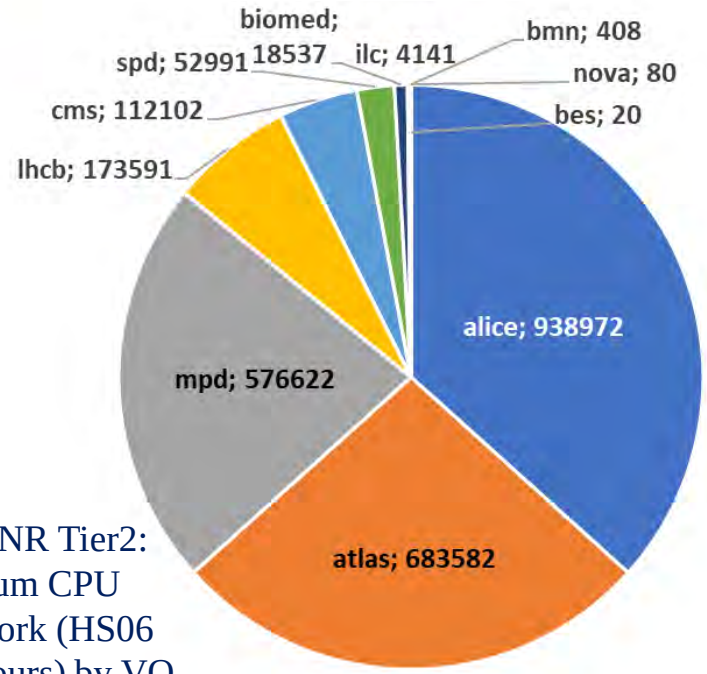
# Tier2 at JINR

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)

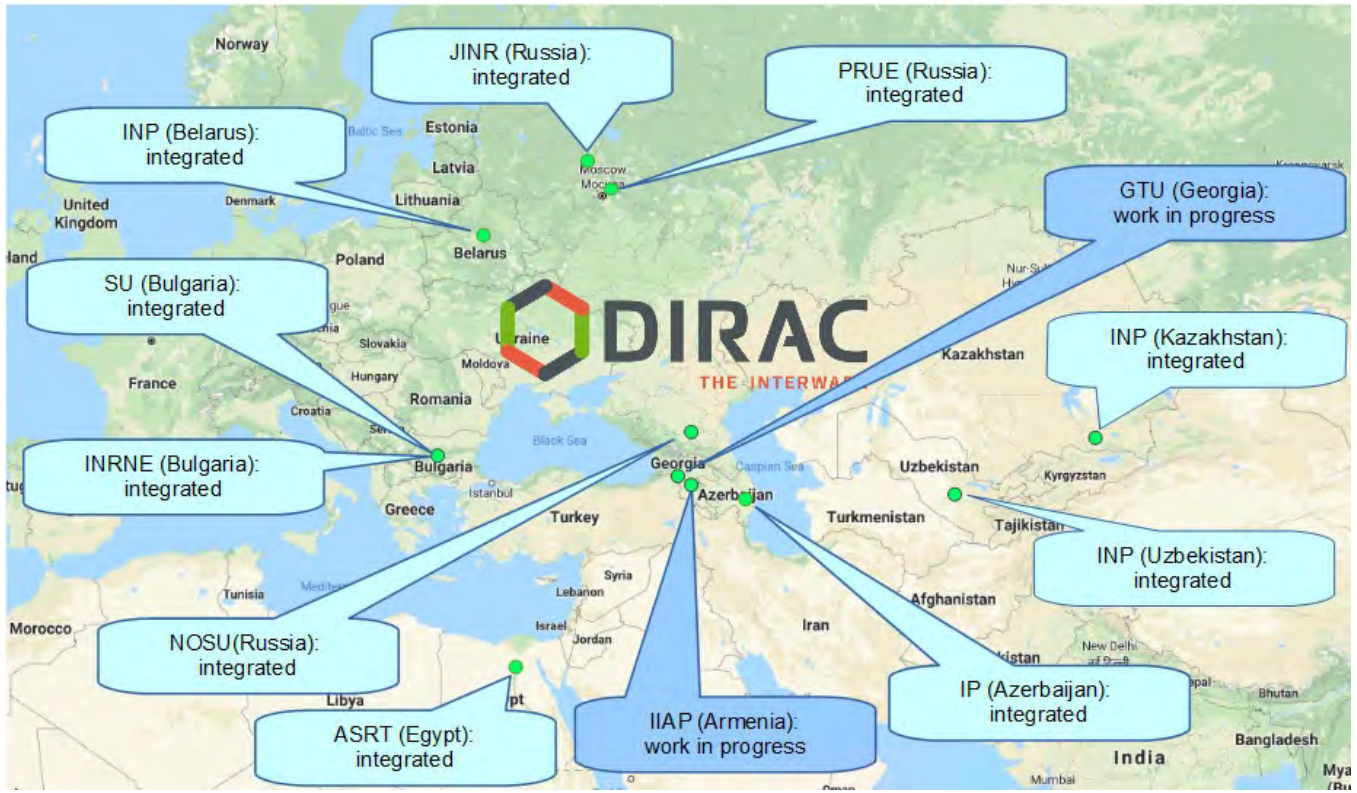
JINR Tier2 is the most productive in the Russian Data Intensive Grid (RDIG) Federation.



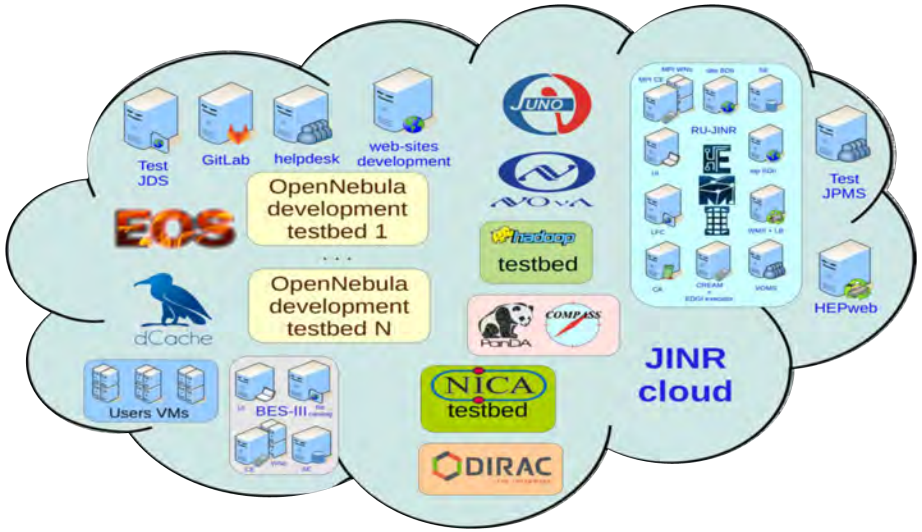

JINR Tier2: Sum CPU work (HS06 hours) by VO (year 2022)

# Cloud Infrastructure

- Cloud Platform - OpenNebula
- Virtualization - KVM
- Storage (Local disks, Ceph)
- Total Resources
  ~ 5,000 CPU cores; 60 TB RAM; 3.1 PB of raw ceph-based storage



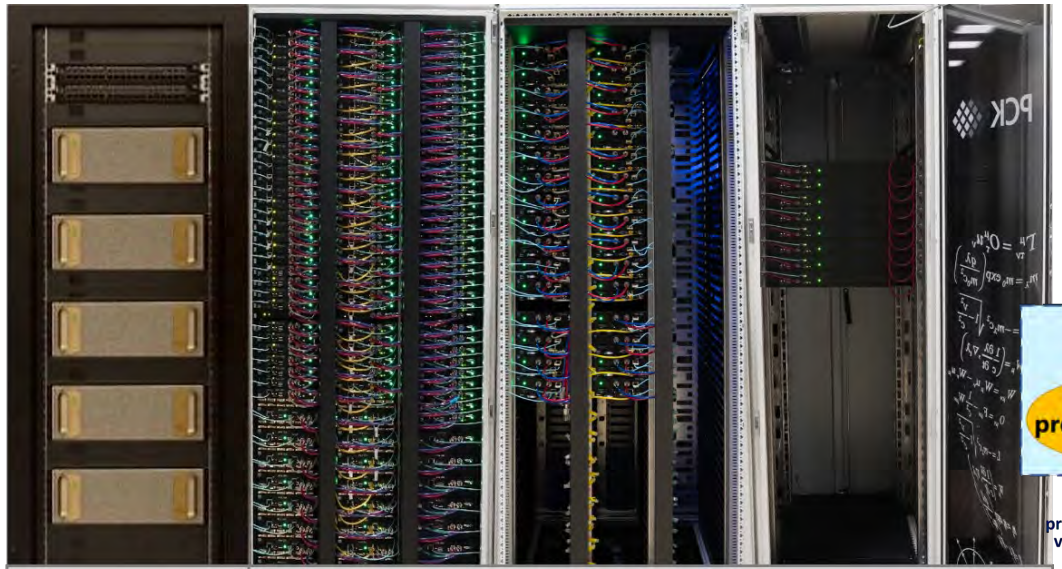

DIRAC-based distributed information and computing environment (DICE) that integrates the JINR Member State organizations' clouds



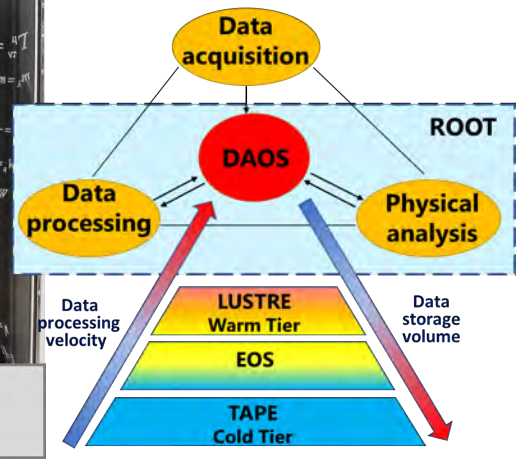

- VMs for JINR users
- Computational resources for neutrino experiments
- Testbeds for research and development in IT
- COMPASS production system services
- Data management system of the UNECE ICP Vegetation
- Scientific and engineering computing
- Service for data visualization
- Gitlab and some others

# “Govorun” Supercomputer

- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 1.1 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



GPU-accelerator | Hyperconverged CPU and Distributed Storage Nodes



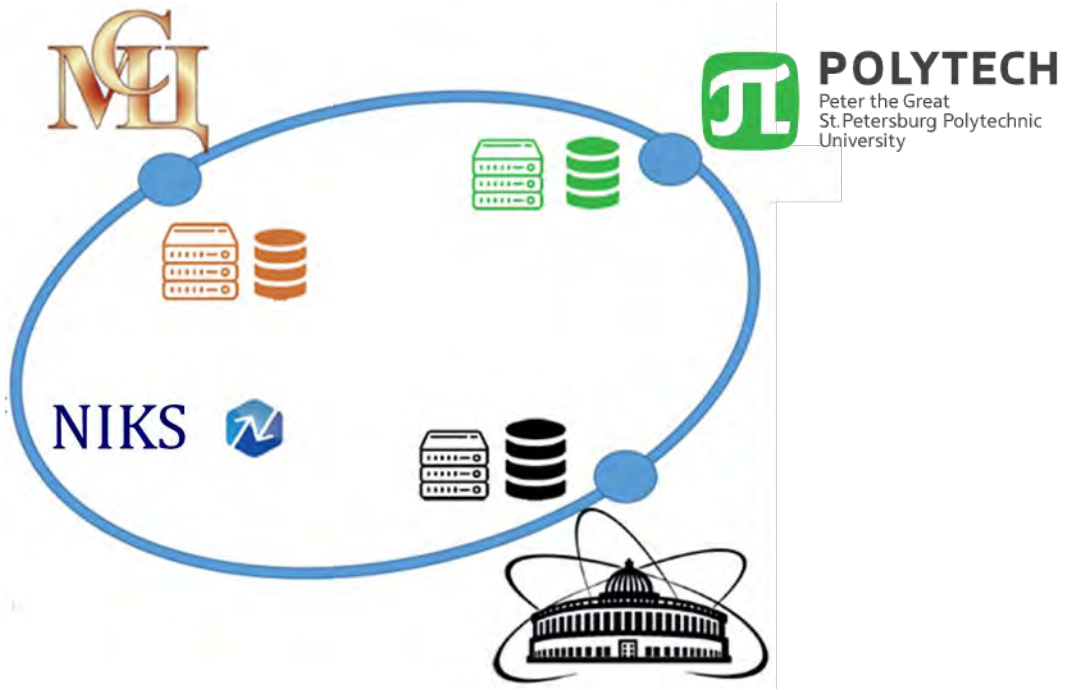

The resources of the “Govorun” supercomputer are used by scientific groups from all the Laboratories of the Institute for solving a wide range of tasks in the field of theoretical physics, as well as for physics modeling and experimental data processing.

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.

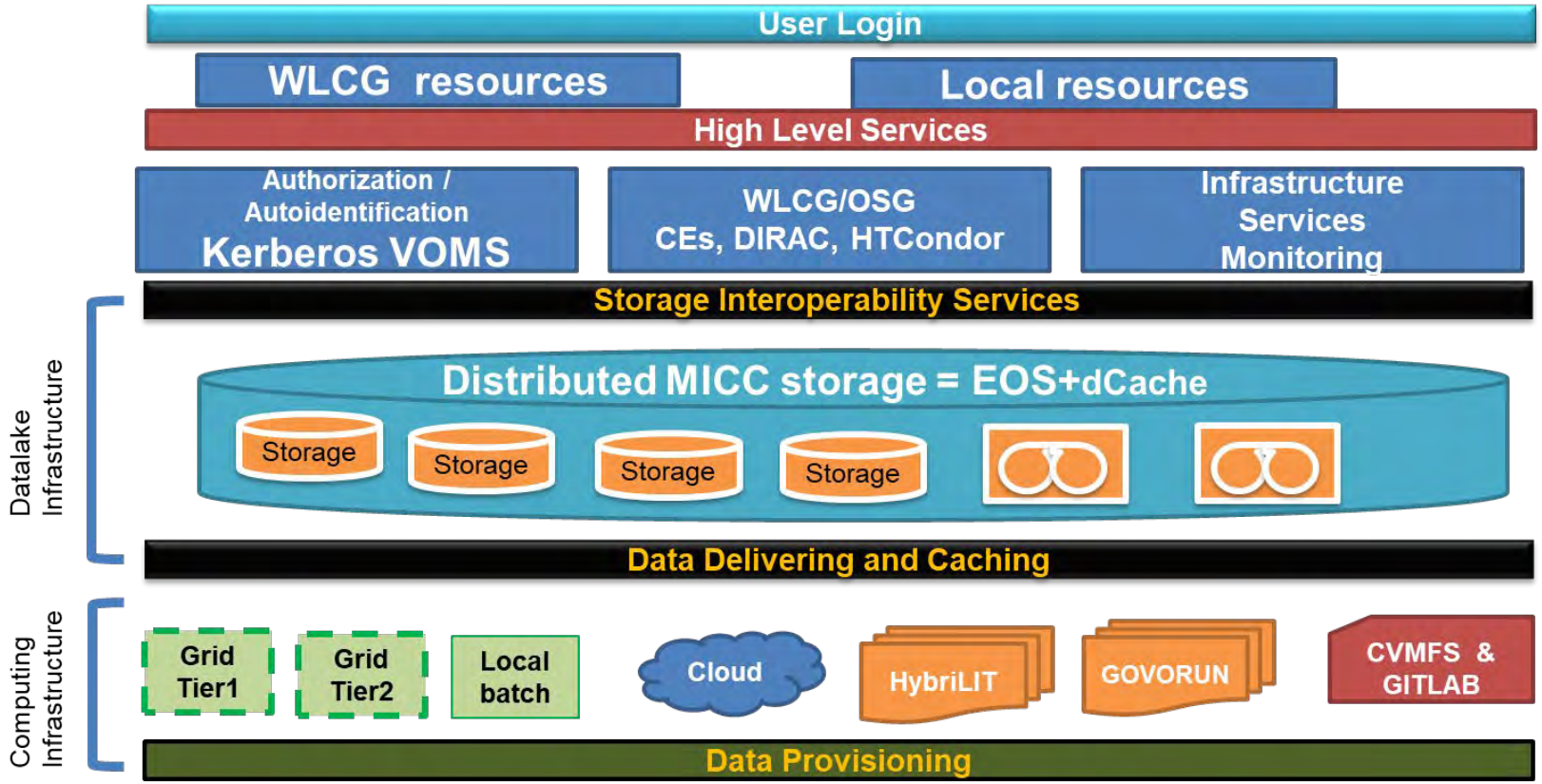
# Unified Scalable Supercomputer Research Infrastructure

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.

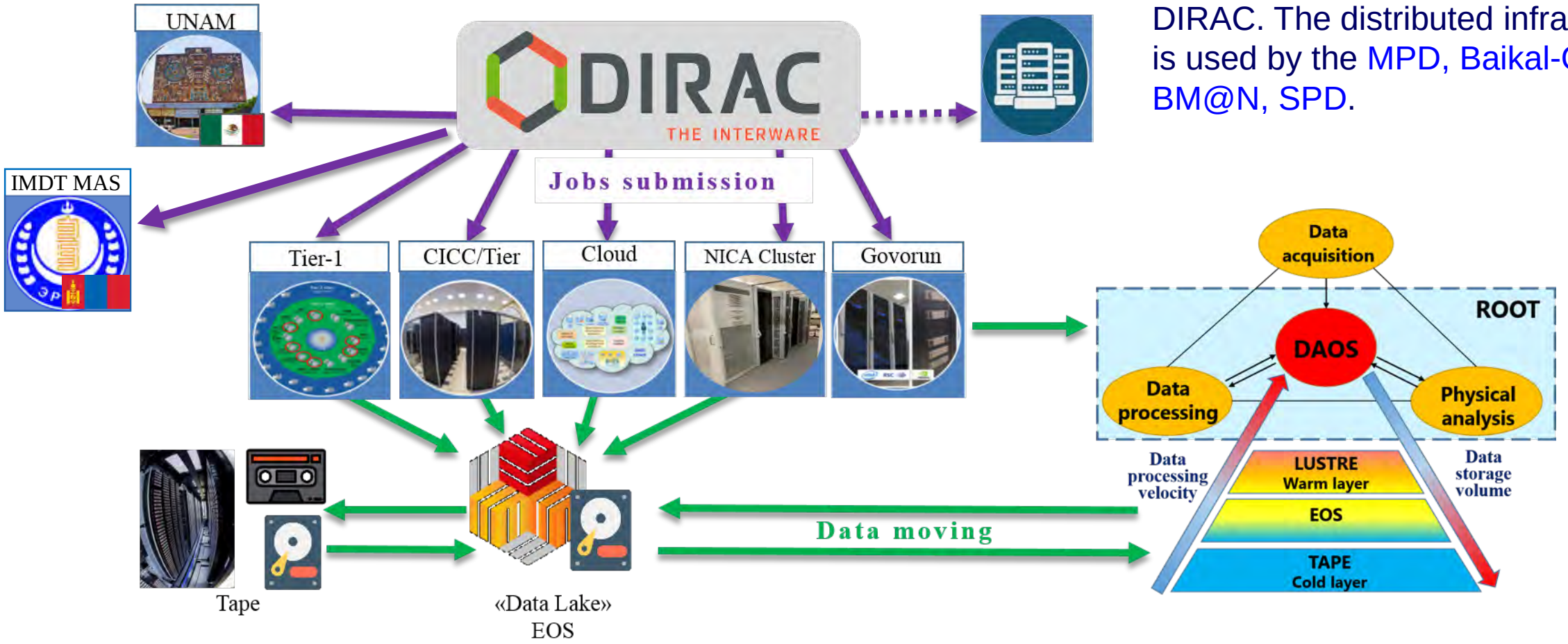
# JINR in DataLakes

- ✓ 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.

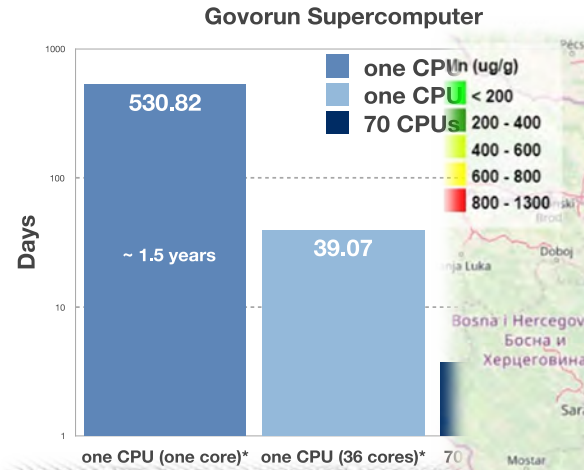


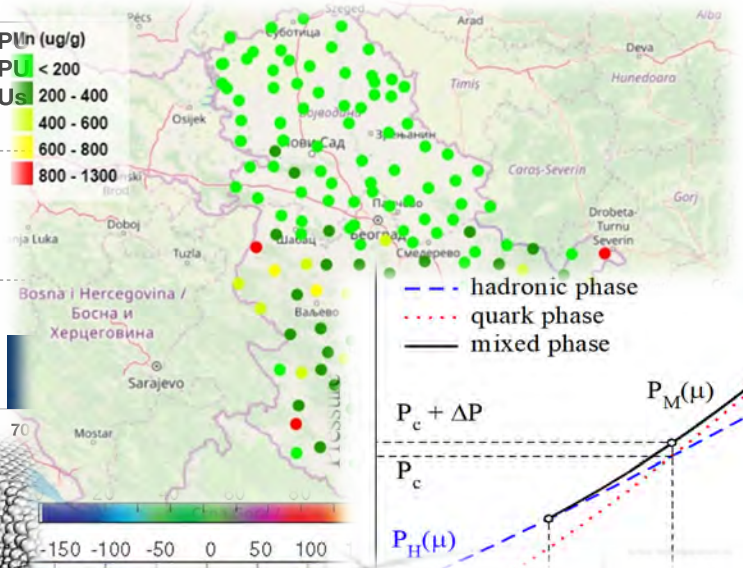
# Heterogeneous Distributed Computing Environment

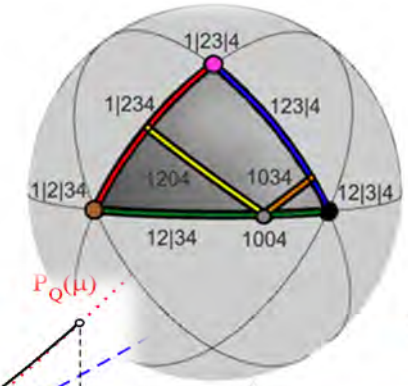
A heterogeneous computing environment, based on the DIRAC platform, was created for processing and storing data of experiments conducted at JINR. Tier1, Tier2, the “Govorun” supercomputer, the clouds of the JINR Member States, the NICA cluster, as well as the resources of the National Research Computer Network of Russia, the cluster of the National Autonomous University of Mexico (UNAM, within cooperation on the MPD project) and the cluster of Institute of Mathematics and Digital Technology (Mongolian Academy of Science), were integrated into DIRAC. The distributed infrastructure is used by the MPD, Baikal-GVD, BM@N, SPD.



# Methods, Algorithms and Software

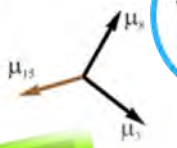


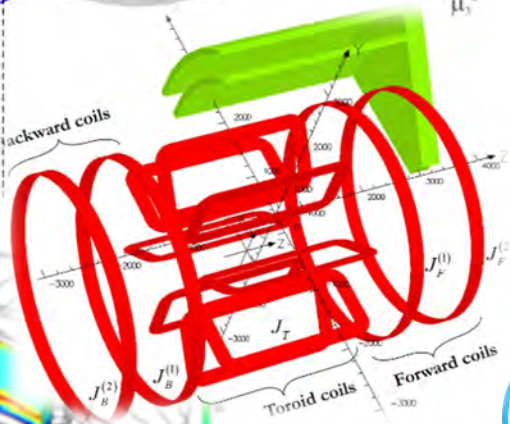



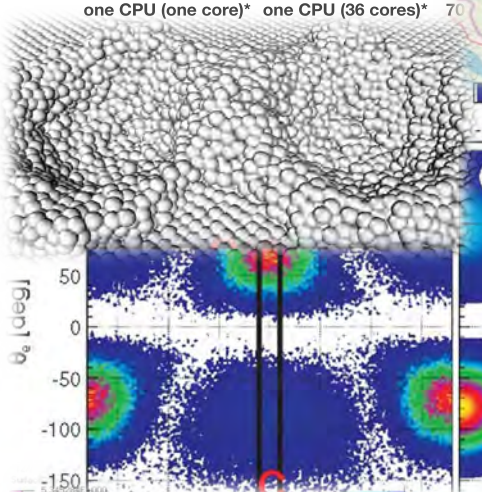



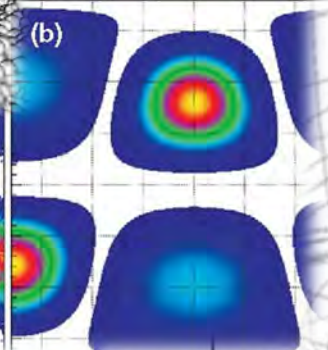


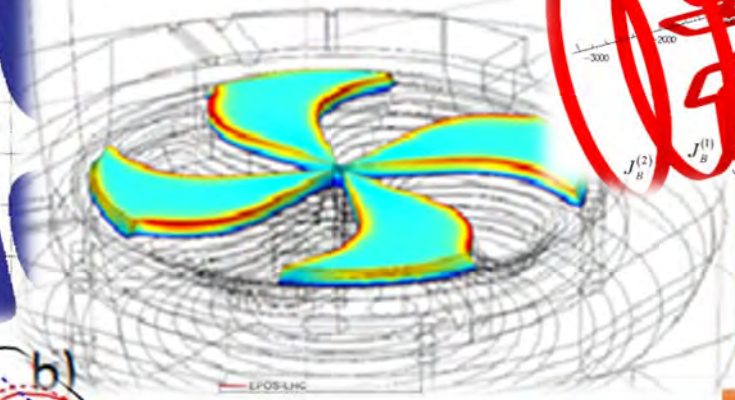




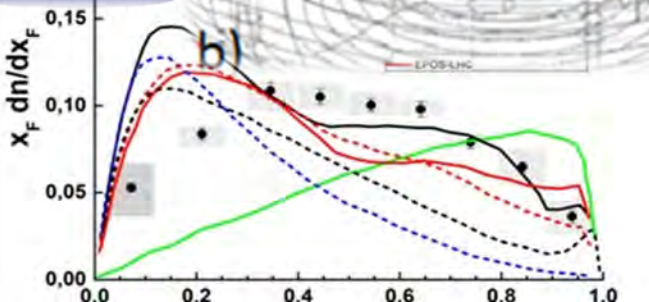



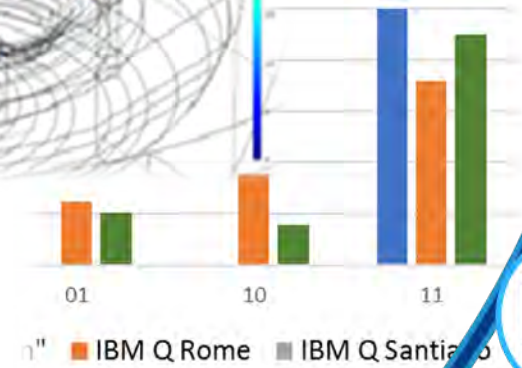






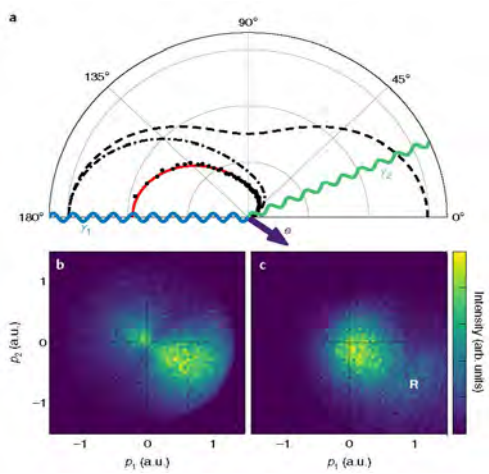


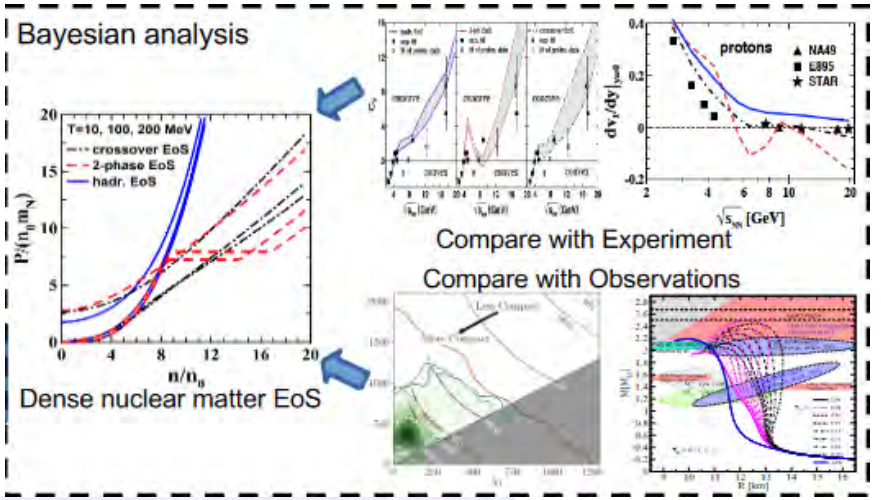

- Numerical modeling of complex physical systems
- Experimental data processing and analysis
- Big Data
- Machine and Deep learning
- AI and robotics
- Computer algebra
- Quantum computing

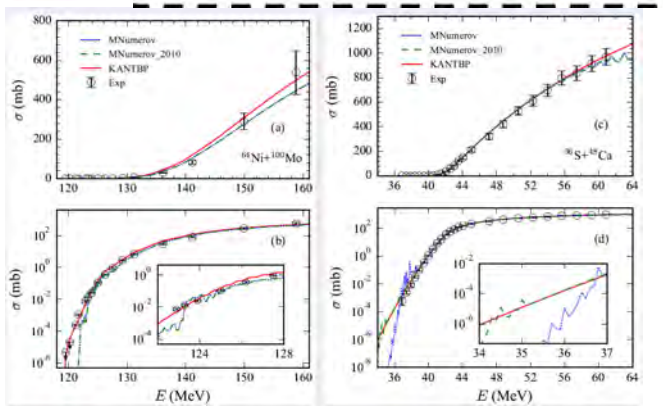
# Methods of Mathematical Modeling, Computational Physics, and High-Performance Computing for Complex System Studies at JINR

- **Simulating interactions of various types in nuclear-physical systems**, including calculations of cross sections for sub-barrier fusion/fission reactions of heavy nuclei within the coupling channel method.
- **Studies of intricate processes in models of complex systems subject to external influences**, including simulations of structural changes in materials under irradiation with charged particles and of superconducting processes in Josephson junctions.
- **Solving problems raised by the design and optimization of the operation of large experimental facilities**, including specific simulations of magnetic field configurations.
- **Modeling physical phenomena based on the state equation of dense nuclear matter**, including complex astrophysical systems and heavy ion collision processes in the NICA energy range.




Scheme of ionization by Compton scattering at hv=2.1 keV




Sub-barrier fusion and fission reactions of heavy nuclei

In 2020-2022, over 150 publications in peer-reviewed scientific journals have been done in cooperation with colleagues from other JINR Laboratories and Member States; 4 problem-oriented software packages in the JINRLIB electronic program library; 2 computer programs were published in the CPC program library.

# Mathematical Methods and Software for Experimental Data Processing and Analysis

## ✓ Physical processes modeling

- event simulations
- GEANT-simulation of experimental setups

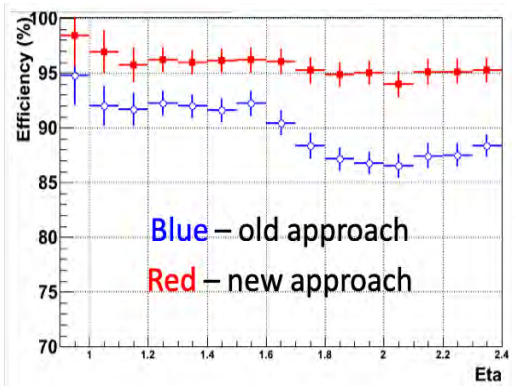
## ✓ Event reconstruction & data analysis

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

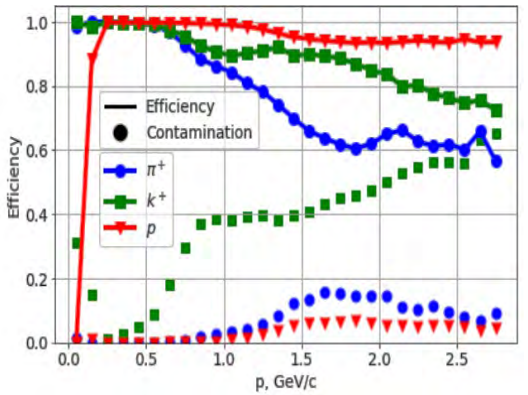
## ✓ Applied software and Data Bases

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

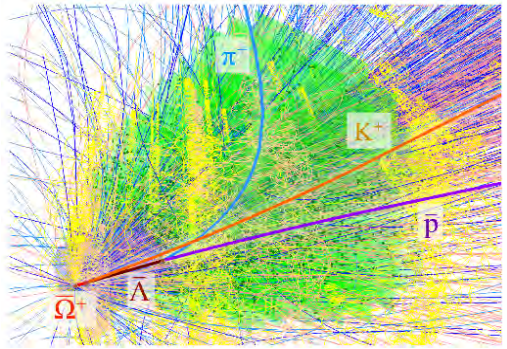
The team members published over 100 specific papers during the last 5 years. They co-authored over 500 papers as members of the international collaborations BM@N, MPD, Baikal-GVD, CMS, ATLAS, CBM, etc.



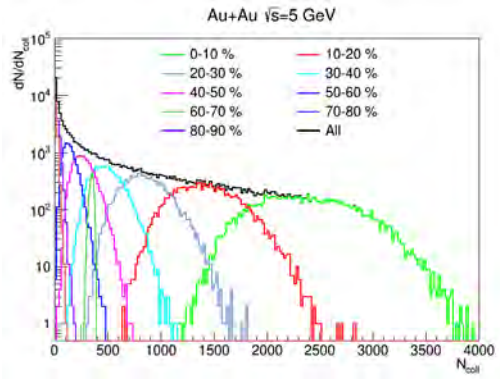

Effective algorithm for TeV muons reconstruction in CMS



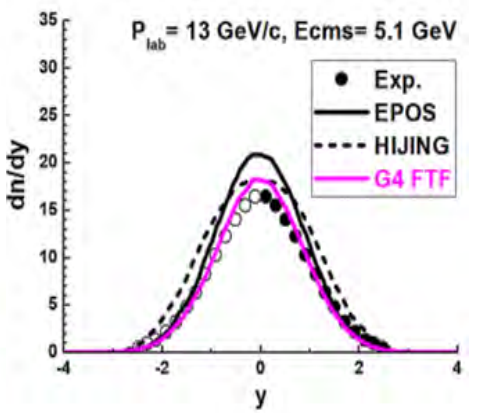

Gradient-boosted decision trees for PID in MPD




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




Monte-Carlo Generator DCM-QGSM-SMM for NICA




NA61/SHINE π−meson rapidity distributions (Geant4)

Experiments DataBases:

ATLAS EventPickingService,

ATLAS CREST,

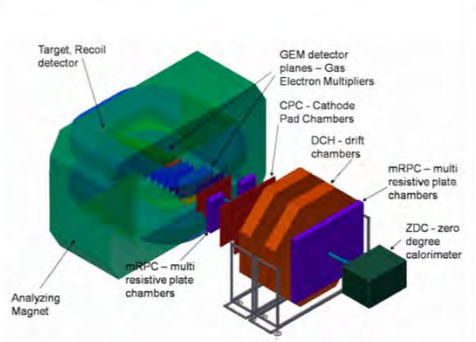
Geometry and Configuration DBs for BM@N

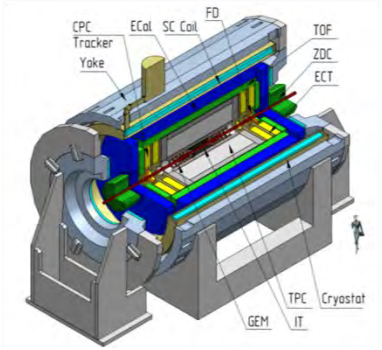
# Implementation of ML/DL Methods for the Data Processing and Analysis of the NICA Experiments BM@N, MPD and SPD

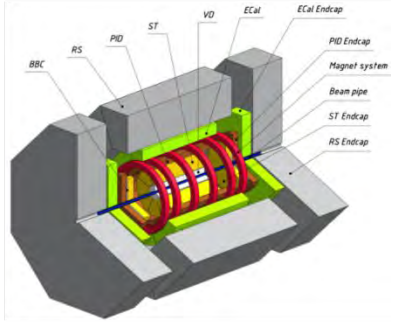
Scientific and practical significance: **expanding the scope of machine learning methods**, in particular, in high energy physics; **software for** experimental data **processing and analysis** at the NICA accelerator complex; corresponding development of root-frameworks.

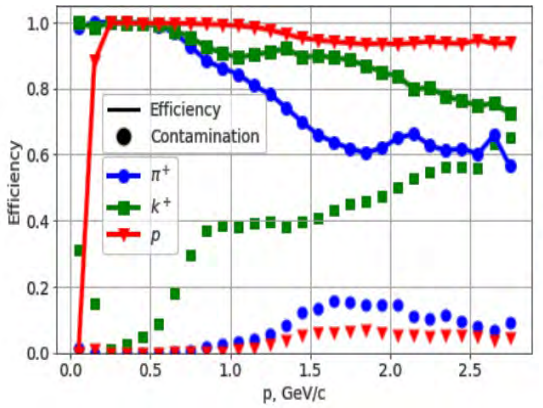
**Foreseen areas for ML/DL application:** hit finding, tracking, particle identification, decay reconstruction, global tracking.

**ML/DL methods under study:** Recurrent Neural Networks, Graph Neural Networks, Convolutional Neural Networks, Decision Trees, Gradient Boosting.

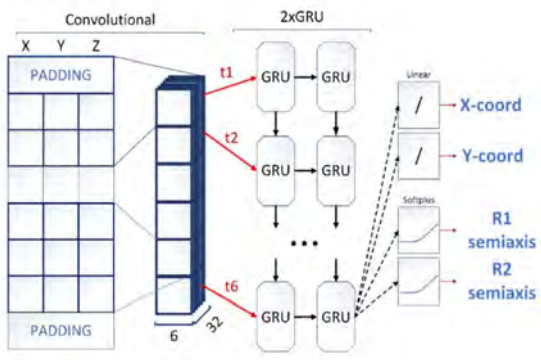


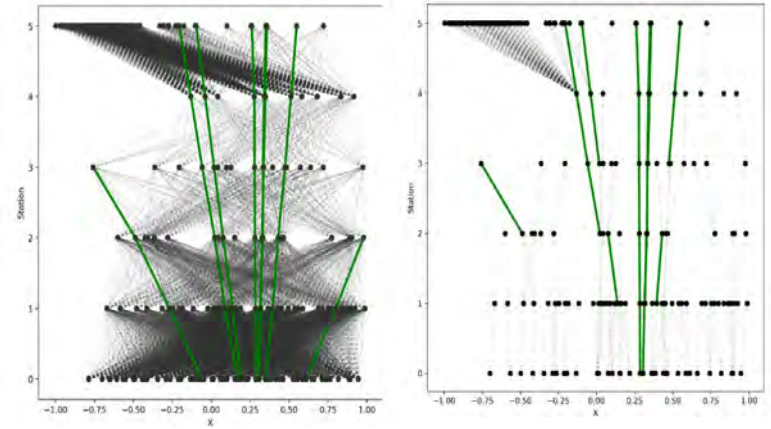




Gradient-boosted decision trees for PID in MPD




Deep GNNs for solving tracking problems in BM@N, BESIII, SPD

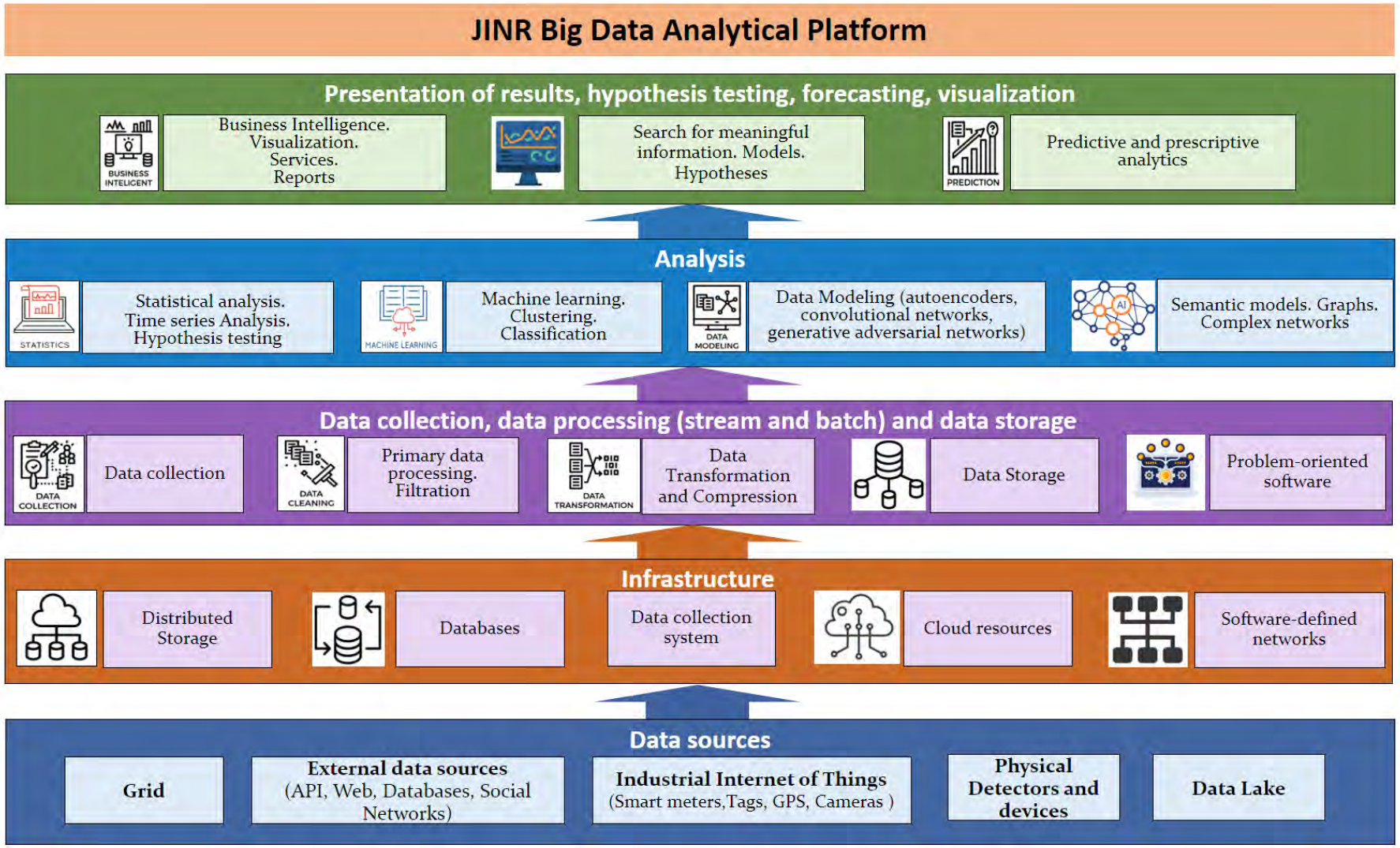


Graph Neural Networks for Tracking

The LIT team is actively involved in all the three targeted international collaborations, BM@N, MPD, SPD.

# Methods of Artificial Intelligence and Big Data Analytics

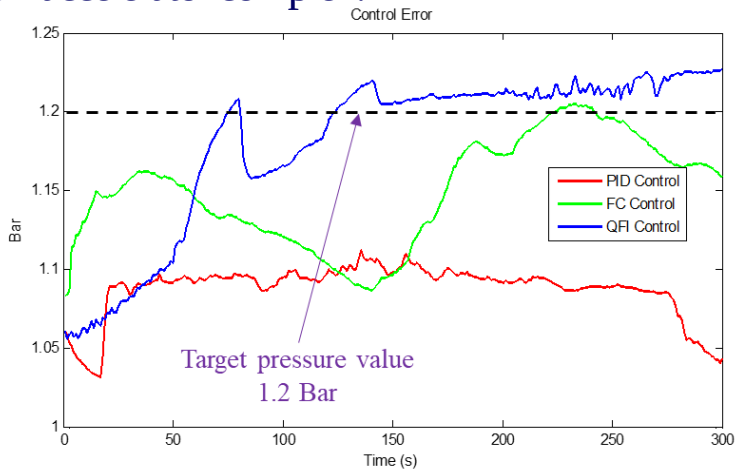
- Bringing best of Big Data approaches to JINR practices
- Providing the Big Data infrastructure for users



# Quantum Intelligent Control of Technological Processes and Physical Facilities at JINR

1. Development of built-in quantum self-organizing controllers for systems of the intelligent control of technological processes, devices and facilities at JINR (including for unforeseen and unpredictable situations).
2. Development of an artificial intelligence platform based on quantum optimization for the tasks of intelligent cognitive robotics and quantum intelligent control in JINR projects.
3. Development of quantum software engineering methods for quantum deep learning based on quantum algorithms, quantum programming, quantum genetic algorithms and quantum soft computing.

A software and hardware platform has been developed on the basis of quantum fuzzy controllers embedded into the control loop to the control of the pressure and flow of liquid nitrogen of the superconducting magnets of the cryogenic system of the NICA accelerator complex.

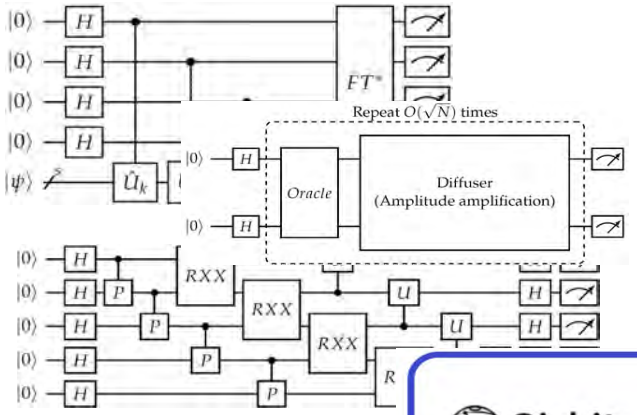



The quantum controller demonstrated the highest speed in achieving the target value, low overshoot and accuracy of achieving the control goal compared to other types of controllers.

*Butenko A.V., Zrelov P.V., Korenkov V.V., Kostromin S. A., Nikiforov D.N., Reshetnikov A.G., Semashko S.V., Trubnikov G.V., Ulyanov S.V.* Intelligent system for remote control of liquid nitrogen pressure and flow in the cryogenic system of superconducting magnets: hardware and software platform // PEPAN Letters (accepted for publication).

# 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

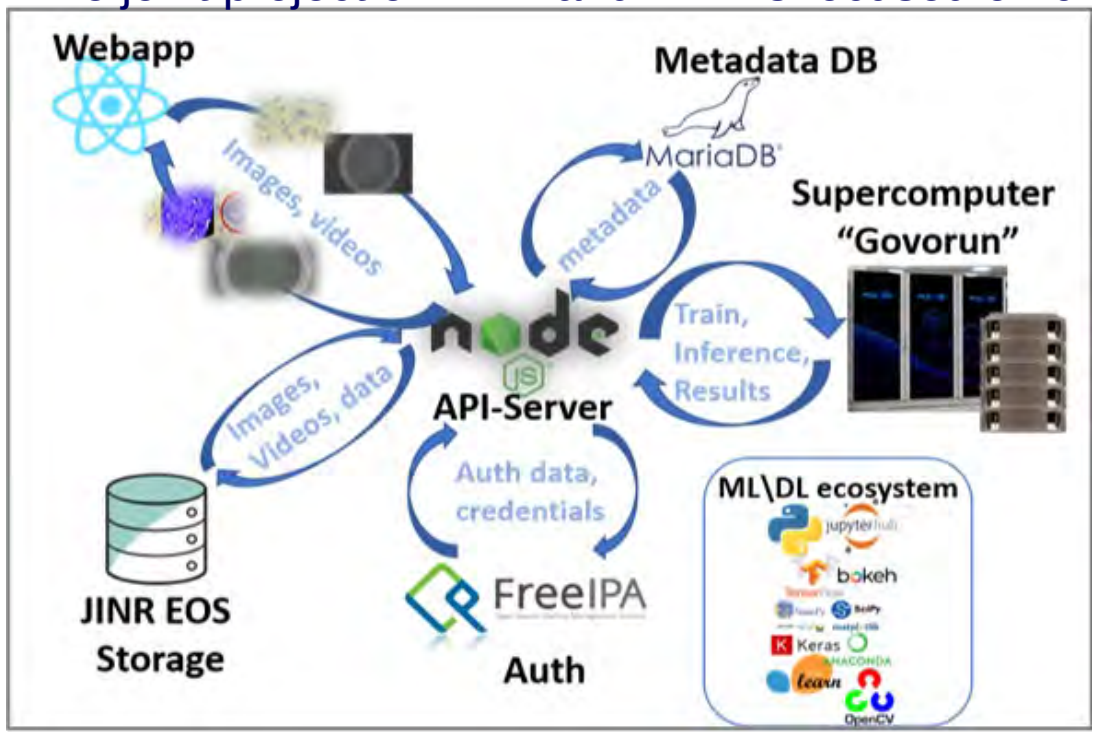
**SC “Govorun”**

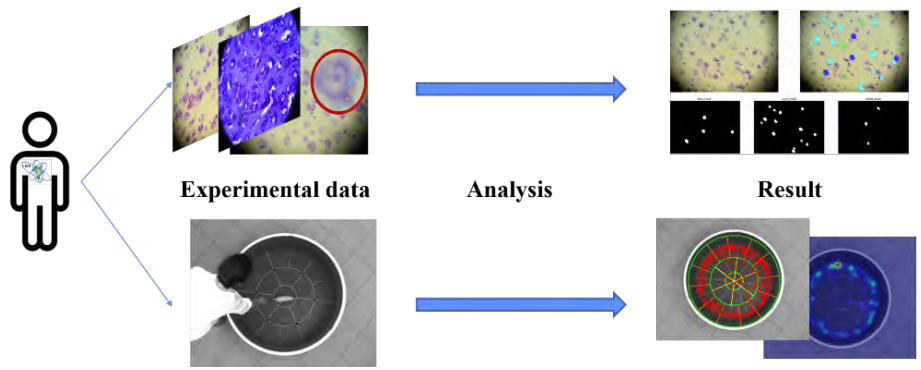

**TASKS**

- 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.



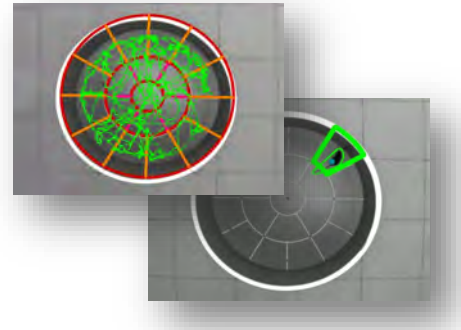



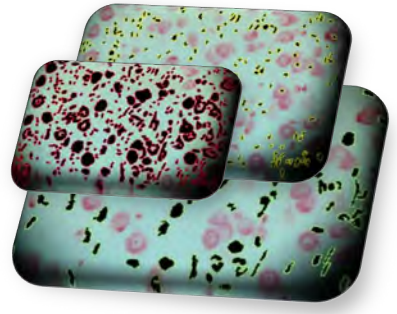

**Conceptual scheme of the service**

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.

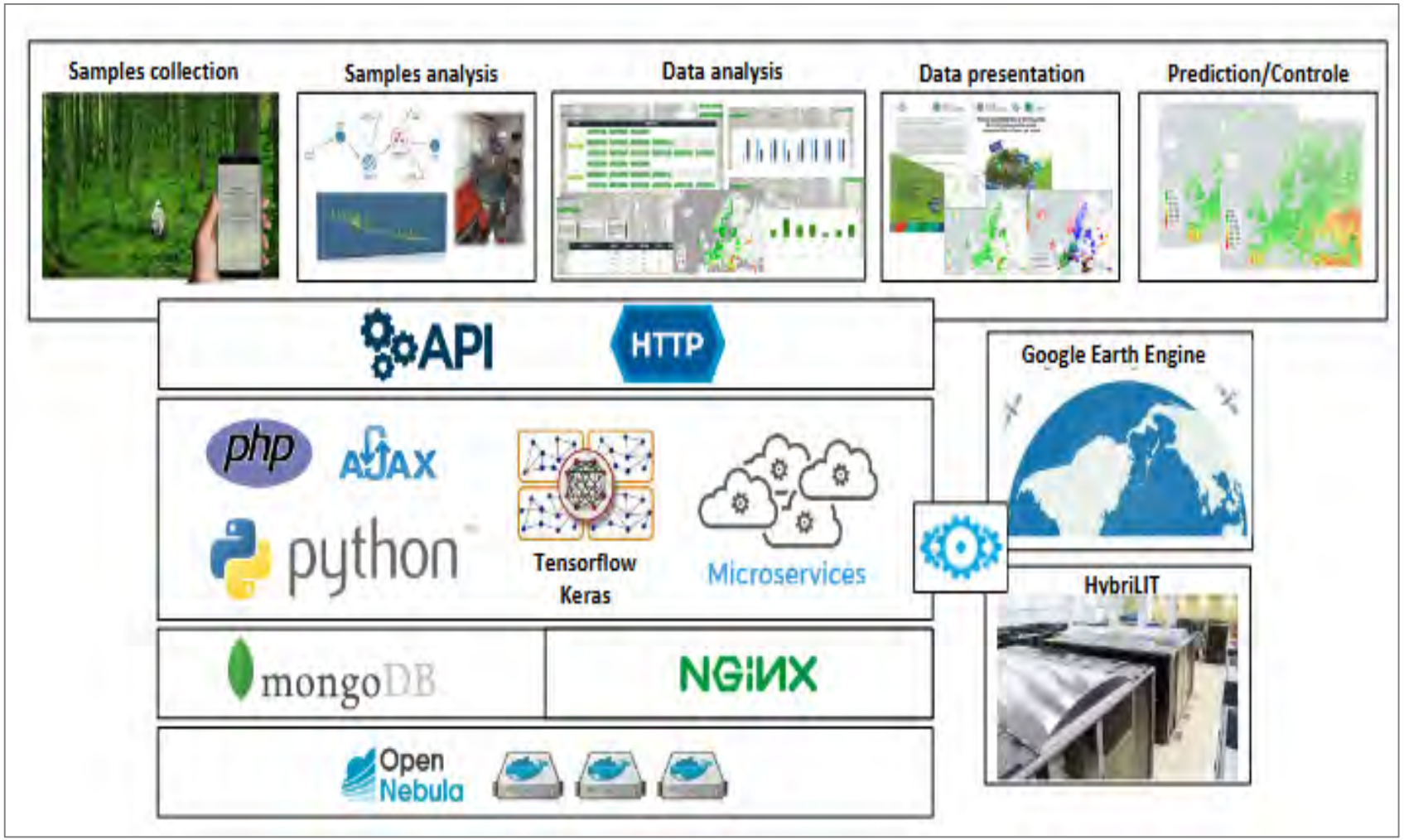
## Tasks of the IS algorithmic block

- 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





# Intelligent Environmental Monitoring Platform




Within the framework of cooperation between MLIT and FLNP, the work on the prediction of air pollution by heavy metals using biomonitoring data, satellite imagery and different technologies of machine and deep learning is in progress. On the MLIT cloud platform, the Data Management System (DMS) of the UNECE ICP Vegetation was created to provide its participants with a modern unified system of collecting, analyzing and processing biological monitoring data.

hYBRI LIT/JINR

*The studies are carried out using the HybriLIT platform.*

# JINR Digital EcoSystem

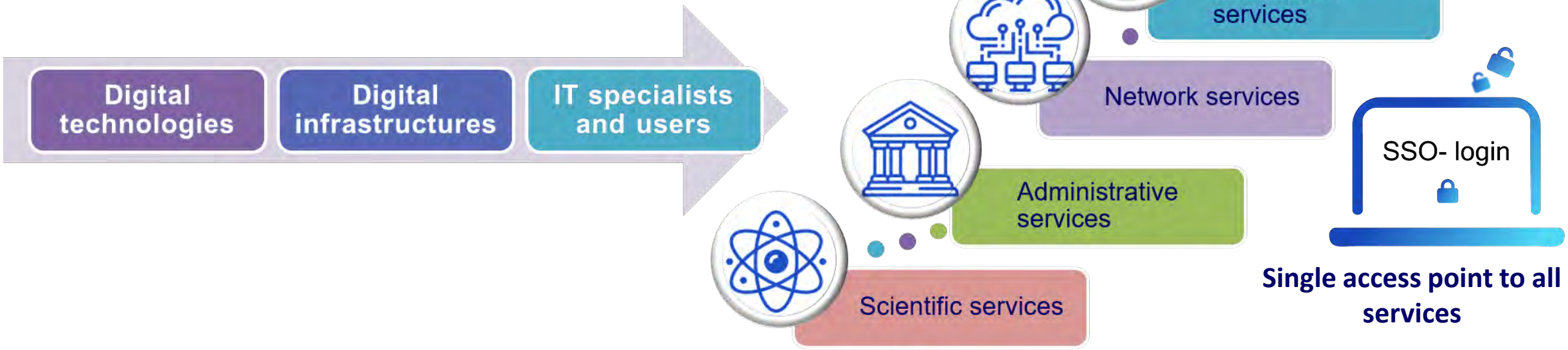
The digital platform “JINR Digital EcoSystem ” integrates existing and future services

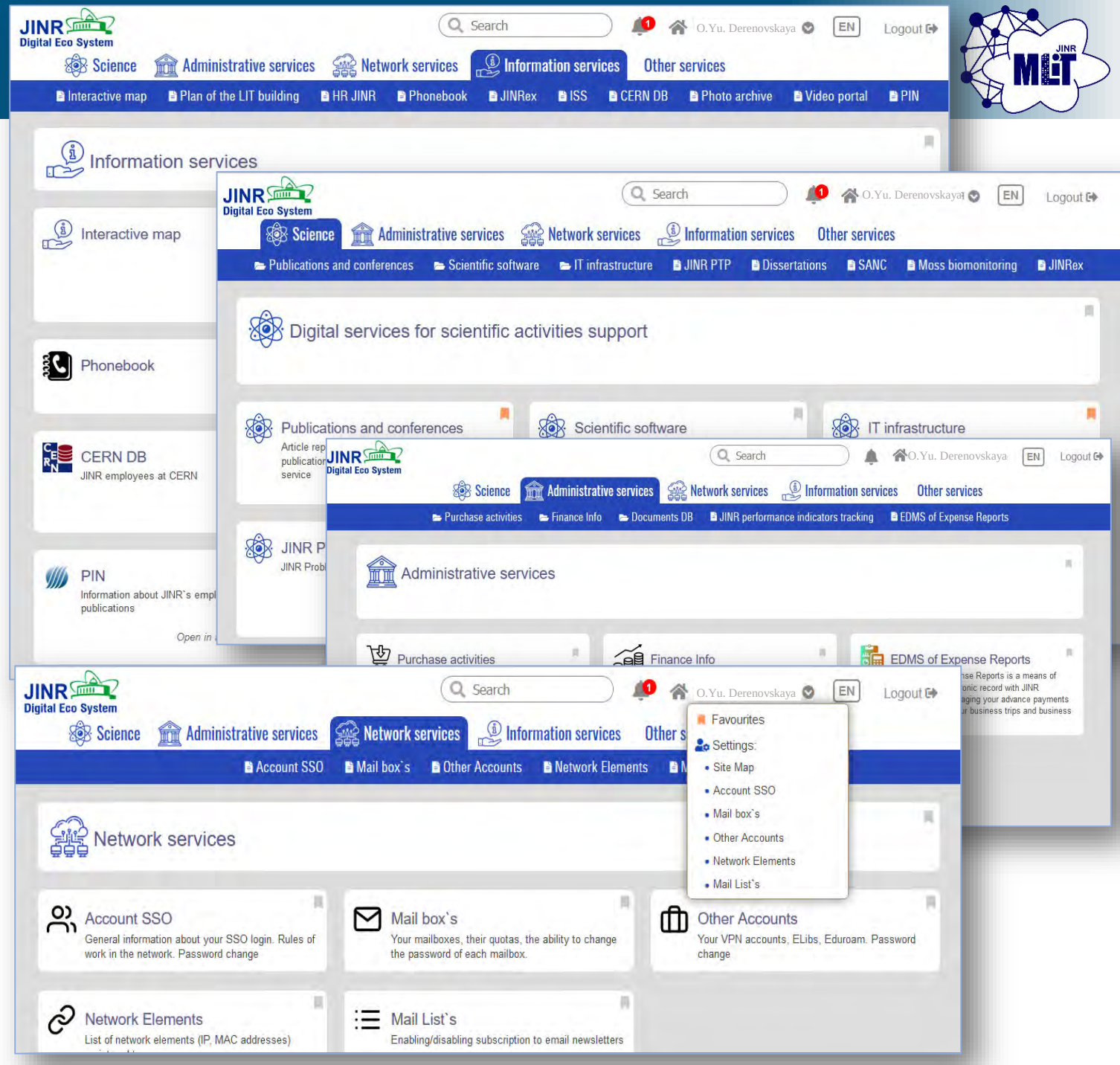
to support
scientific,
administrative and social activities,
maintenance of the engineering and IT infrastructures

to provide
reliable and secure access to various types of data

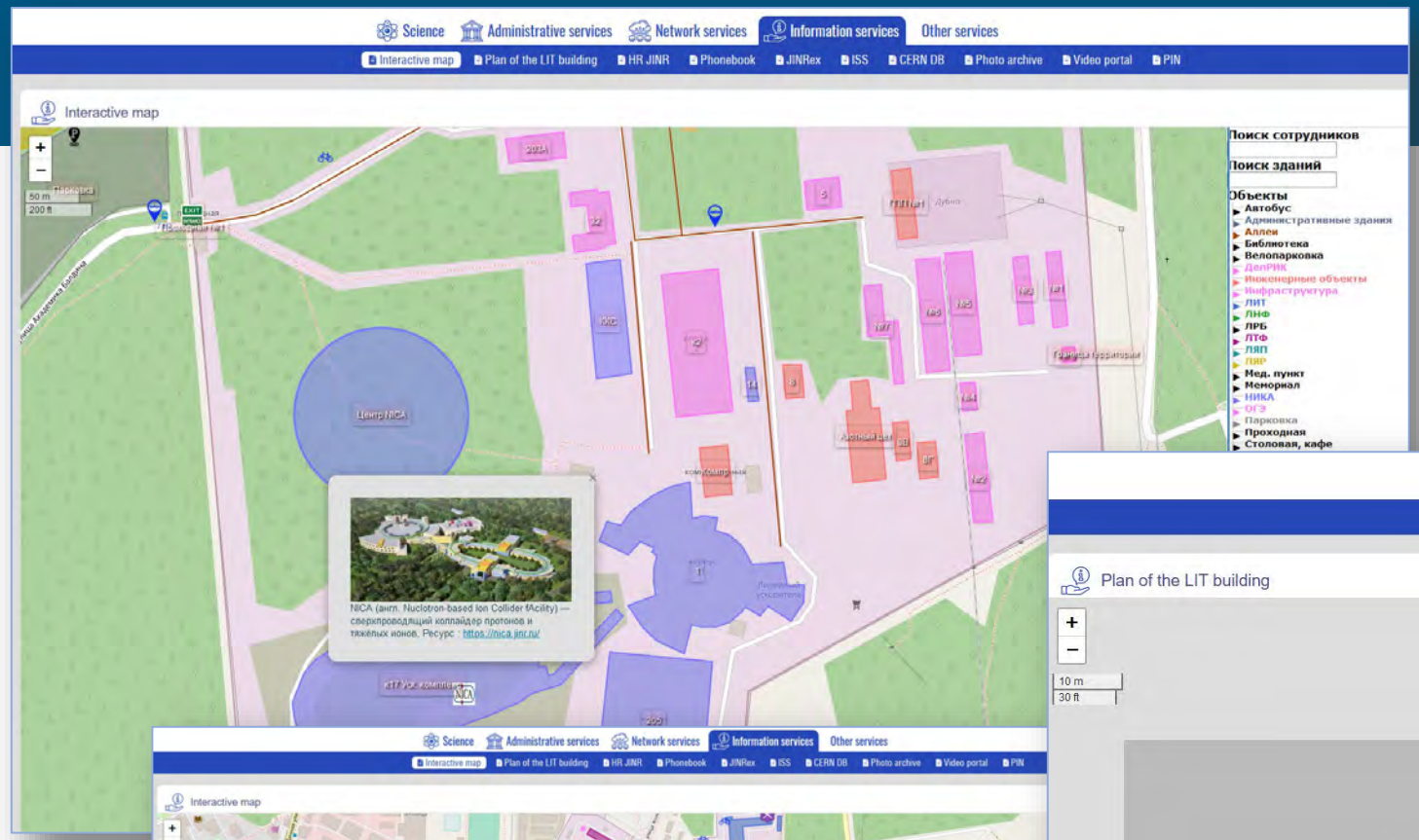
to enable
a comprehensive analysis of information

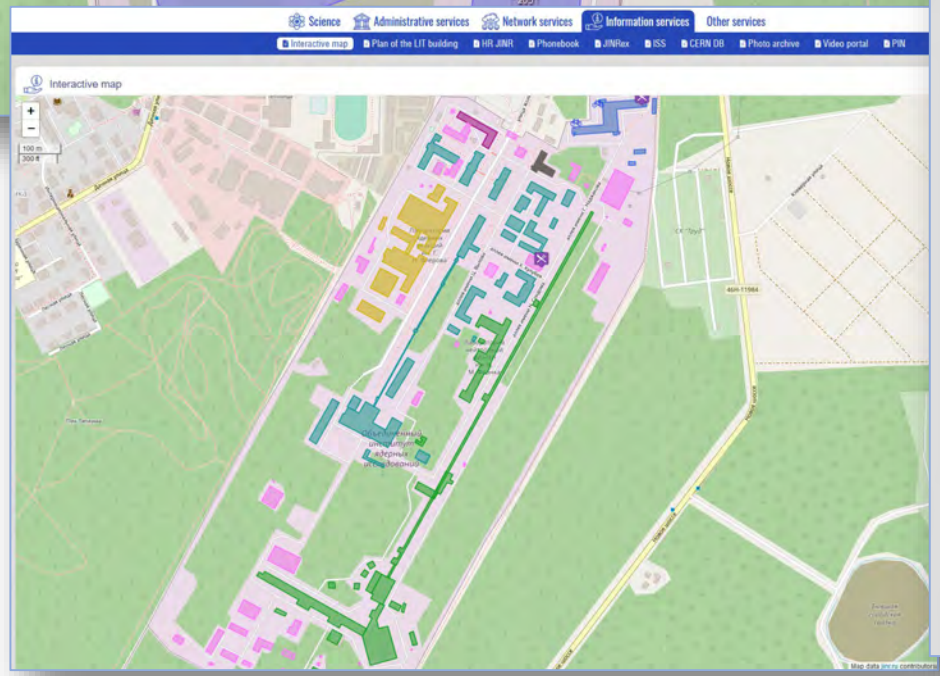
using
modern Big Data technologies and artificial intelligence.

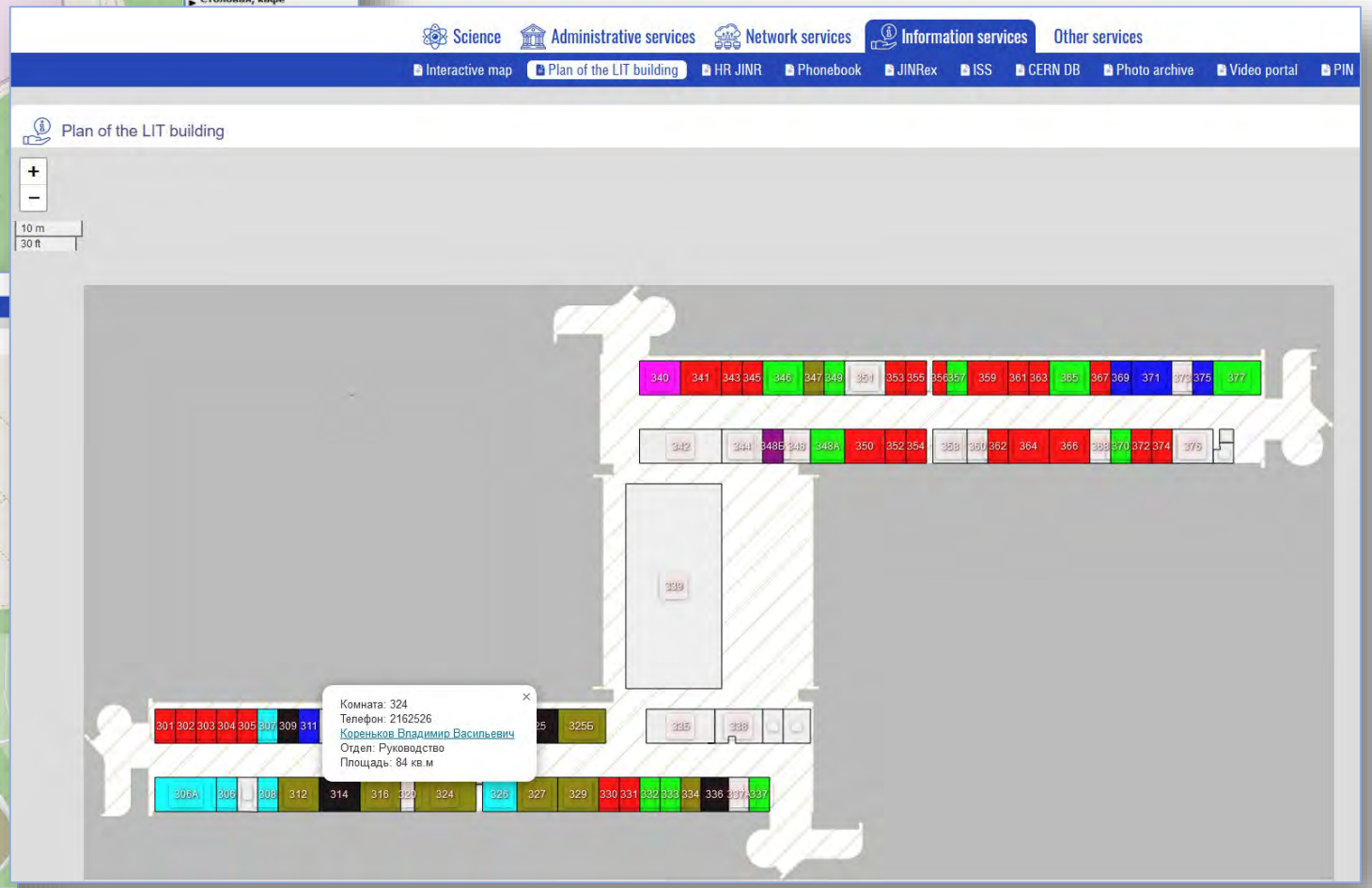



- ✓ Personal account of a JINR employee
- ✓ Notifications in a personal account
- ✓ Responsive interface, customizable by the user
- ✓ Easy access, convenient navigation and search for information on a large-scale network of a wide variety of JINR services




- ✓ Quick and easy search for information, both by services and by employees and buildings on an interactive JINR map

# 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

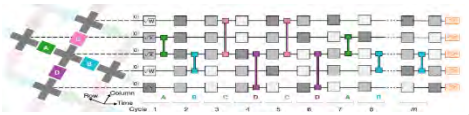
jupyterhub

TensorFlow

NumPy

scikit learn

### Quantum algorithms, quantum programming and quantum control



# Information technology projects (I)

## Critical basic scientific infrastructure project

**Multifunctional Information and Computing Complex, including the “Govorun” supercomputer (JINR Data and Networking Center):**

Engineering infrastructure (electricity and cooling)

Networking (local and worldwide connectivity)

HT Computing (Grid Tier1 and Tier2)

HP Computing (“Govorun” supercomputer)

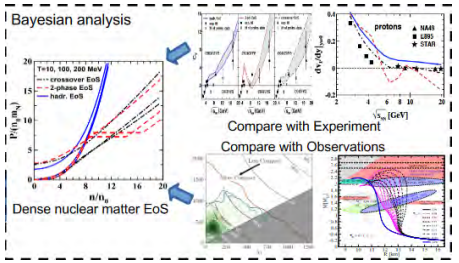
Cloud Computing

Data storage (long-term and middle-term Data lake)

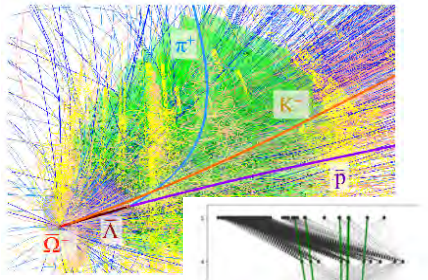

# Information technology projects (II)

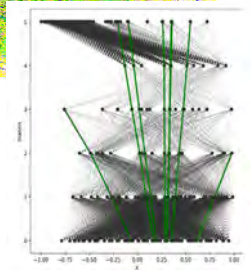
## Scientific directions and information systems

- **Methods of Mathematical Modeling, Computational Physics, and High-Performance Computing for Complex System Studies**

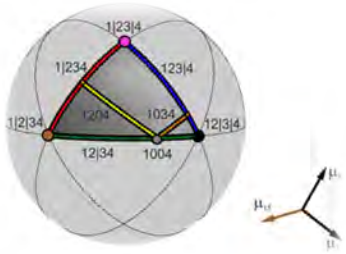



- **Mathematical Processing and Analysis of Experimental Data**
  - **Implementation of ML/DL Methods in Data Processing and Analysis at the NICA Experiments**





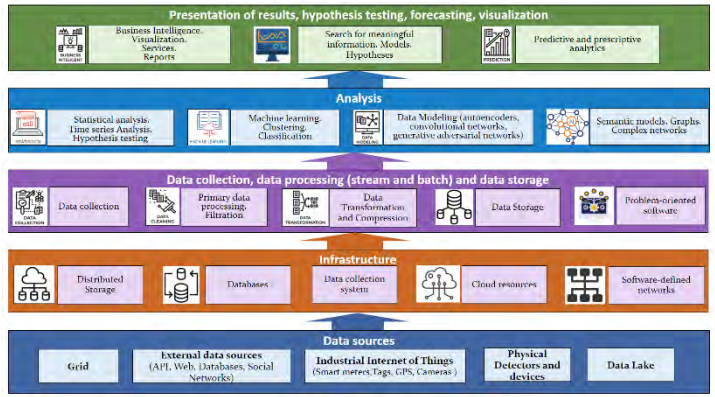
- **Development and application of methods of computational mathematics in quantum information theory**



- **Digital JINR**

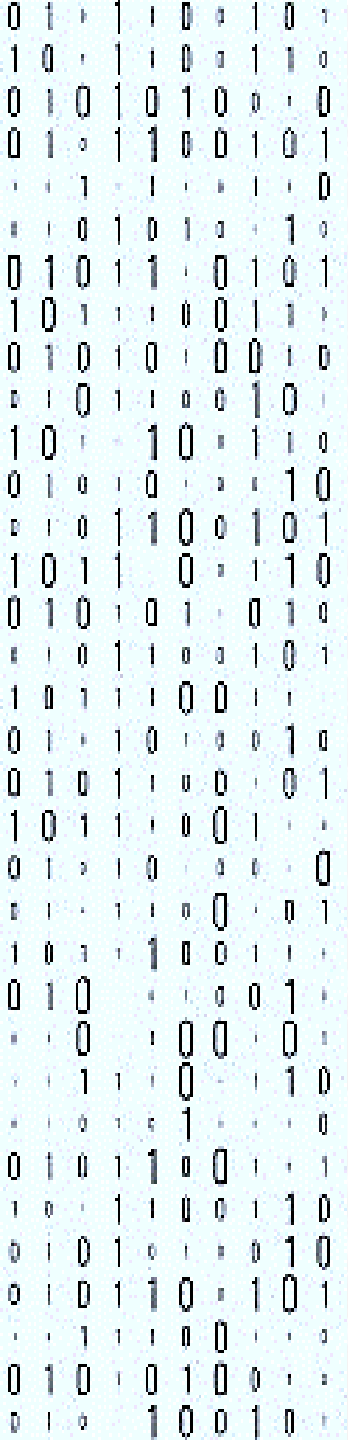
JINR Digital Eco System

- **Methods of Artificial Intelligence and Big Data Analytics**




- **Quantum intelligent control of technological processes and physical facilities at JINR**

## IT School



# Thank you for attention!