Status and plans of the Meshcheryakov Laboratory of Information Technologies

Danila Oleynik, MLIT
Alushta 2023

LCTA foundation (1966)

П Р И К А З ПО ОБЪЕДИНЕННОМУ ИНСТИТУТУ ЯДЕРНЫХ ИССЛЕДОВАНИЙ №149

"19" августа 1966 года

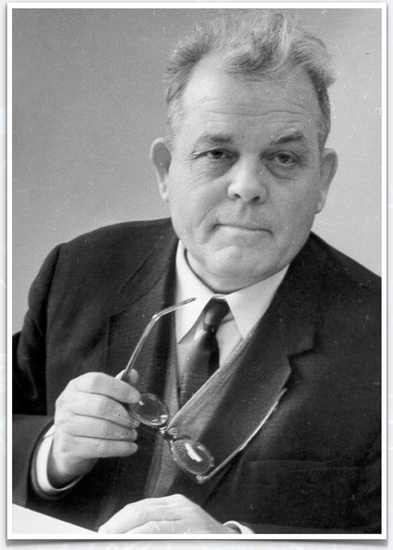
В связи с решением XX сессии Ученого Совета и Комитета Полномочных Представителей в составе Объединенного института ядерных исследований организуется Лаборатория вычислительной техники и автоматизации, на которую возлагается:

всестороннее развитие вычислительной техники и вопросов программирования в Институте, как основы автоматизации обработки экспериментальной информации и математических расчетов для теоретических и экспериментальных физических исследований;

обеспечение всего комплекса обработки экспериментальной информации на вычислительных машинах и, прежде всего, обработки фотографий с пузырьковых и искровых камер, получаемых в ОИЯИ и на ускорителе в Серпухове;

обеспечение связи и координация совместных работ стран-участниц ОИЯИ по вопросам вычислительной техники, программированию, развитию методик обработки и другим вопросам автоматизации;

координация основных работ по созданию и развитию измерительных центров в лабораториях ОИЯИ и внедрению цифровых вычислительных машин в экспериментальные методики.



LCTA - The Laboratory of Computing Techniques and Automation, reorganised to Laboratory of Information Technologies in 2000

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
- …..

Information Technologies (Scientific directions and information systems)

Life Science

- Information System for Radiation Biology study
- 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

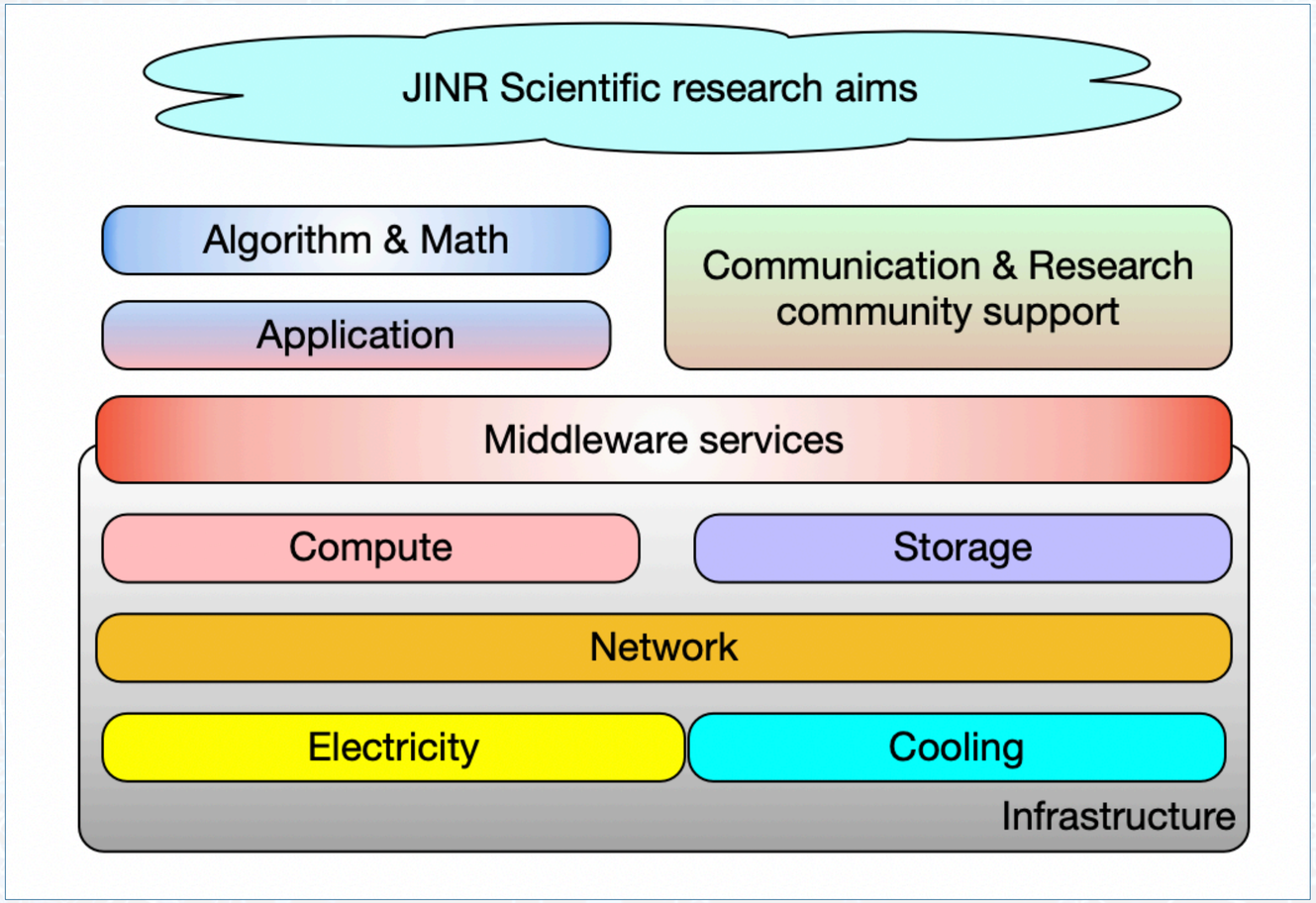
- Analysis of polydisperse populations of phospholipid vesicles
- Study of nanocomposite thin films using neutron and X-ray reflectometry methods
- Simulation of thermal processes occurring in materials
- …..

Neutrino Physics and Astrophysics

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

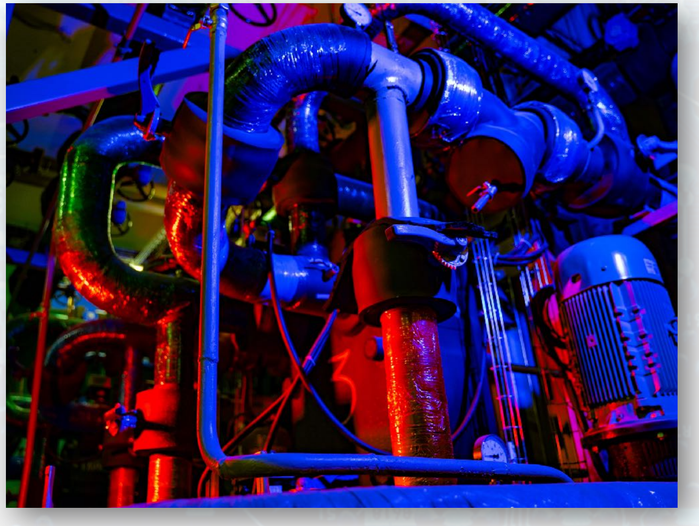
Information Technology and Scientific Computing at JINR

Coordinated development of interconnected IT technologies and computational methods



- “Layer cake” from engineering infrastructure through hardware facilities and middleware services to application
  - *Middleware services - a wide set of tools and solutions, which allows users applications employs network, compute and storage facilities*

Engineering infrastructure

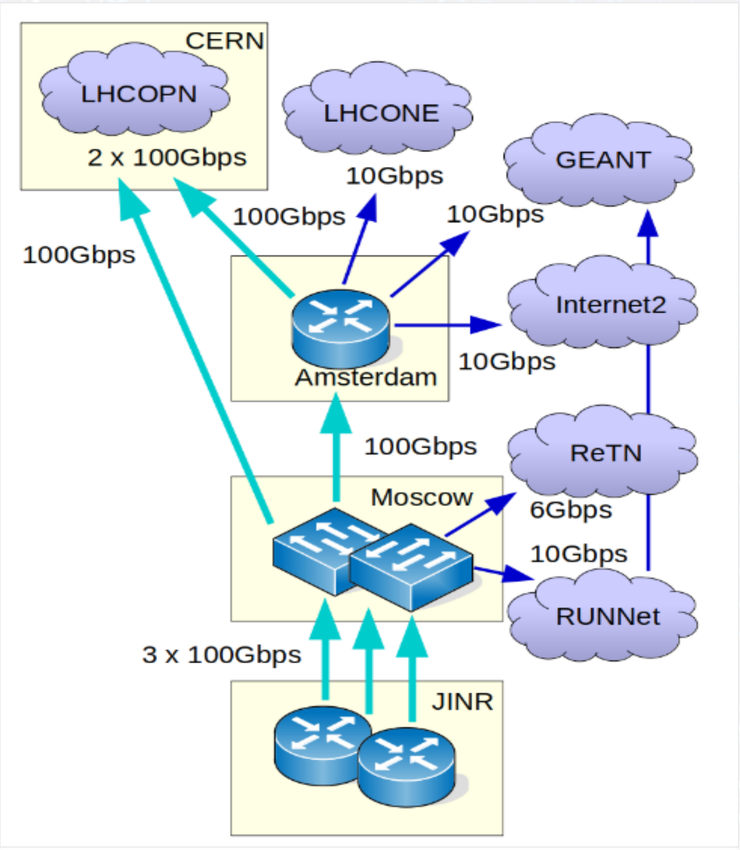


Power@Cooling 800kVA@1400kW

Network Infrastructure

The network infrastructure is a fundamental component of the IT infrastructure of JINR. 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
- dedicated campus network with a bandwidth of 4x100 Gbit/s for the NICA experiments
- local area network with a bandwidth of 2x100 Gb/s



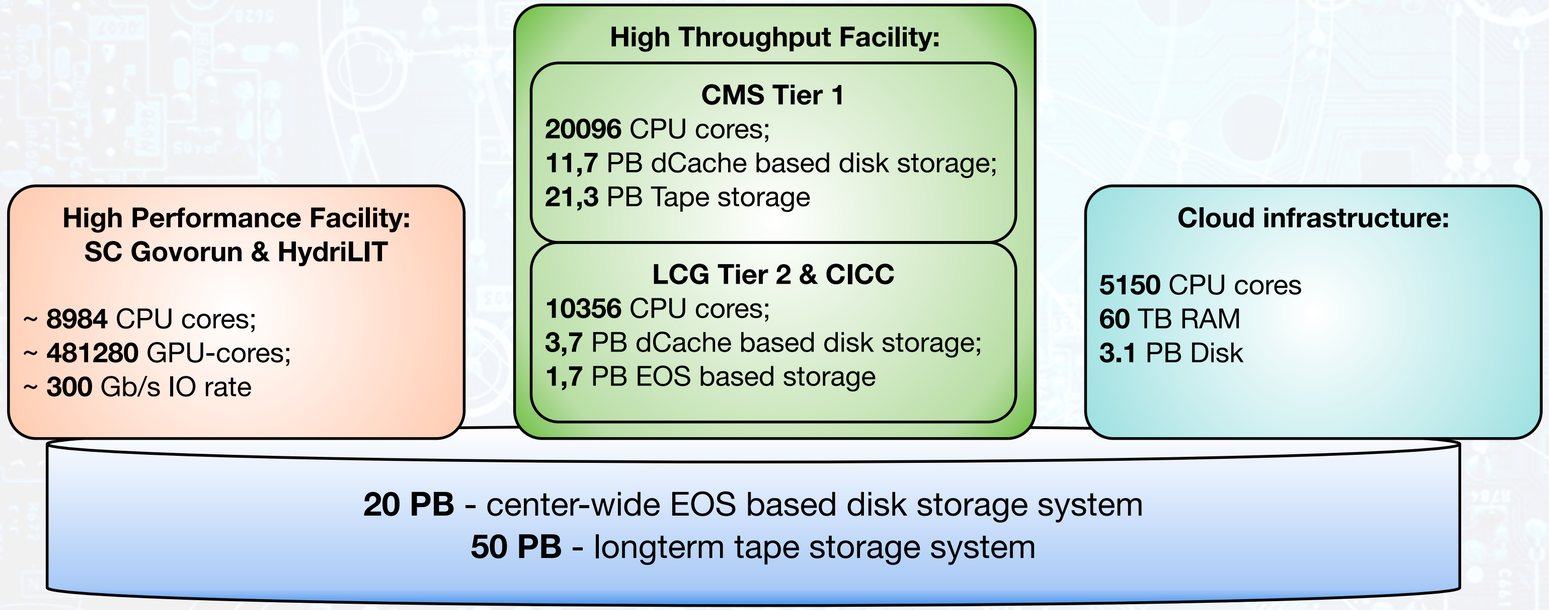
The JINR LAN comprises:

- 9291 network elements
- 18044 IP-addresses
- 6355 users registered within the network

network traffic in 2022

- 29.56 PB - input
- 34.19 PB - output

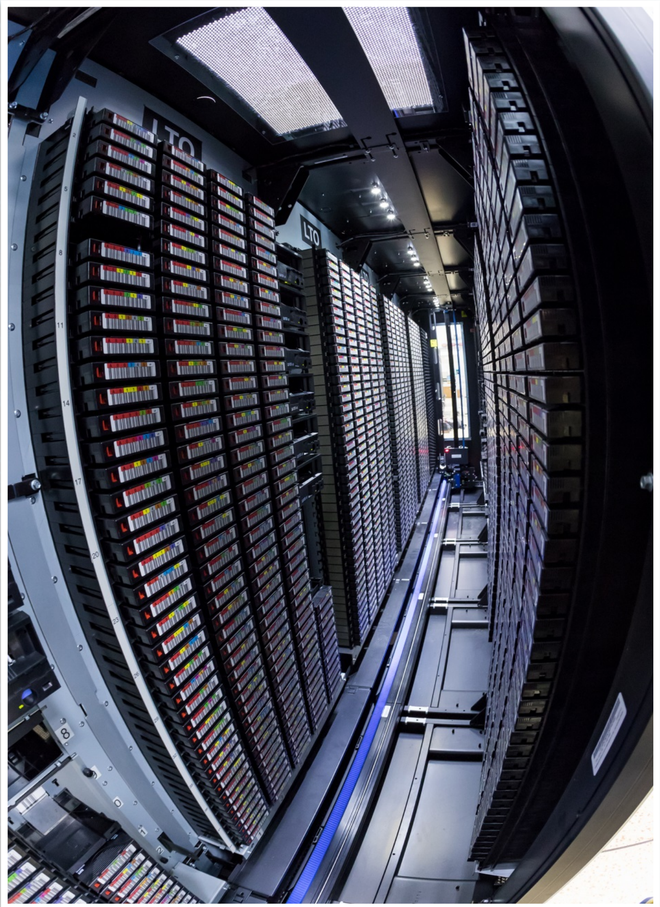
Compute and storage resources



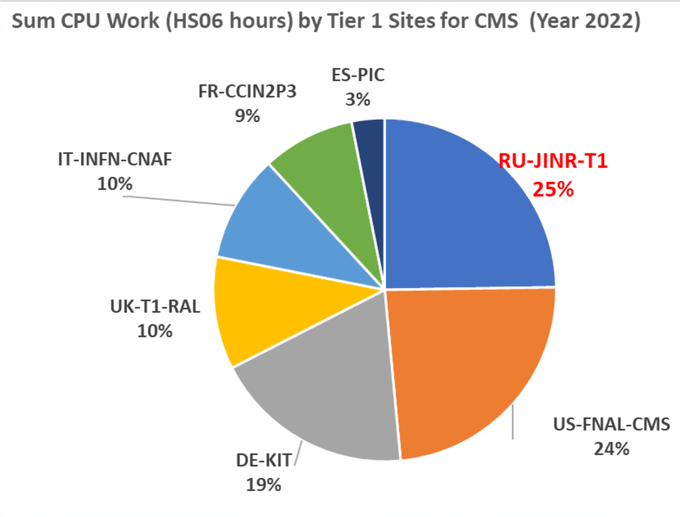
The MICC meets the requirements for a modern highly performant scientific computing complex:

- high reliability and availability,
- multi-functionality,
- scalability,
- high performance,
- high-through modern local network,
- center-wide data storage system,
- supporting of customisation of software environment for different user groups,
- information security.

CMS Tier1 at JINR



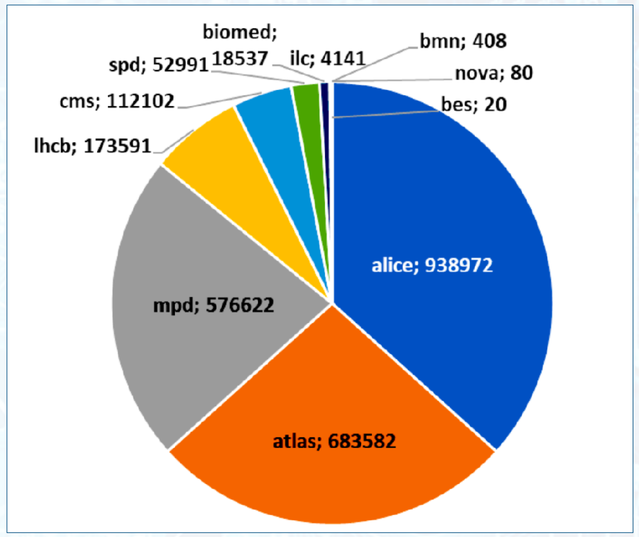
20096 CPU cores;
11,7 PB dCache based disk storage;
21,3 PB Tape storage
100% reliability and availability



- The JINR Tier1 data processing center has demonstrated stable work for CMS experiment on LHC from beginning of work in 2013.
- The Tier1 site for CMS is ranked first among world centers for CMS (by delivered CPU work).

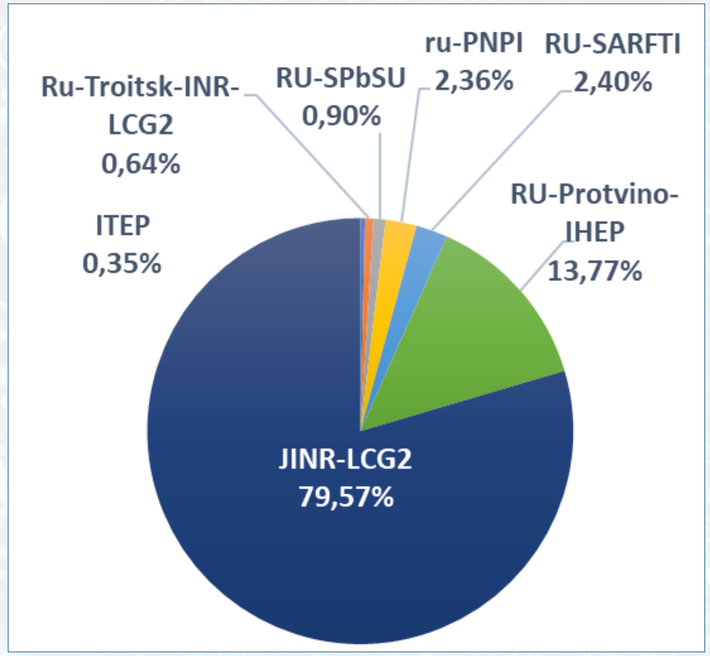
Central Information and Computer Complex - CICC

CICC provides computing power, 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.). It acts like LCG Tier 2 data processing center for all four LHC experiments.



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

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



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

“Govorun” HPC

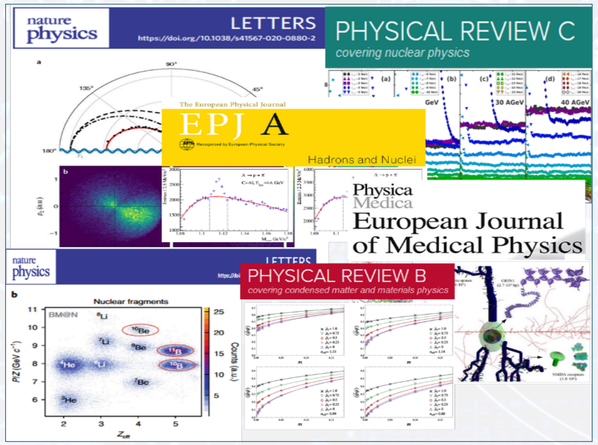
- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 1.1 PFlops DP
- GPU component based on NVIDIA
- CPU component based on RSC “Tornado” liquid cooling solutions
- The most energy-efficient center in Russia (PUE = 1.06)
- Storage performance >300 GB/s

GPU-accelerator	Hyperconverged CPU and Distributed Storage Nodes

The resources of the “Govorun” HPC 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 modelling and experimental data processing.

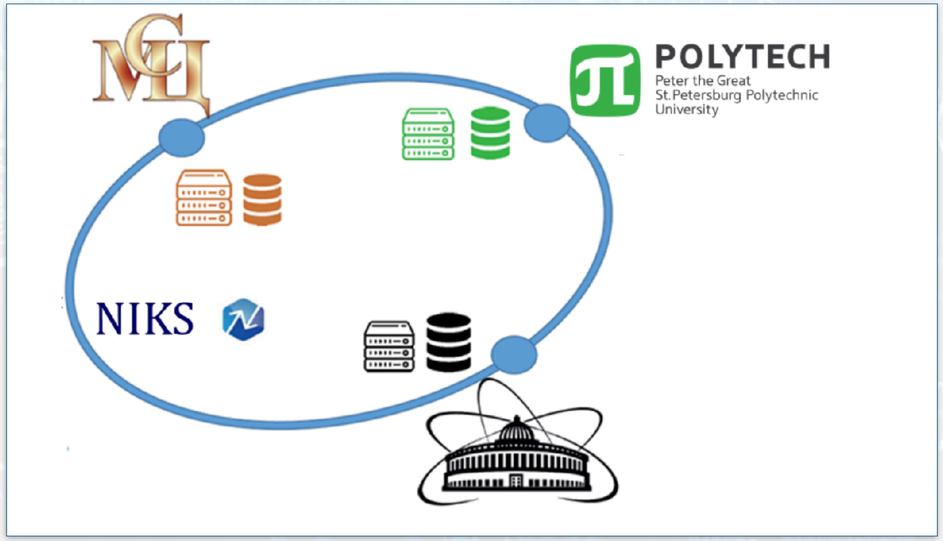
Key projects that use the resources of the HPC:

- 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

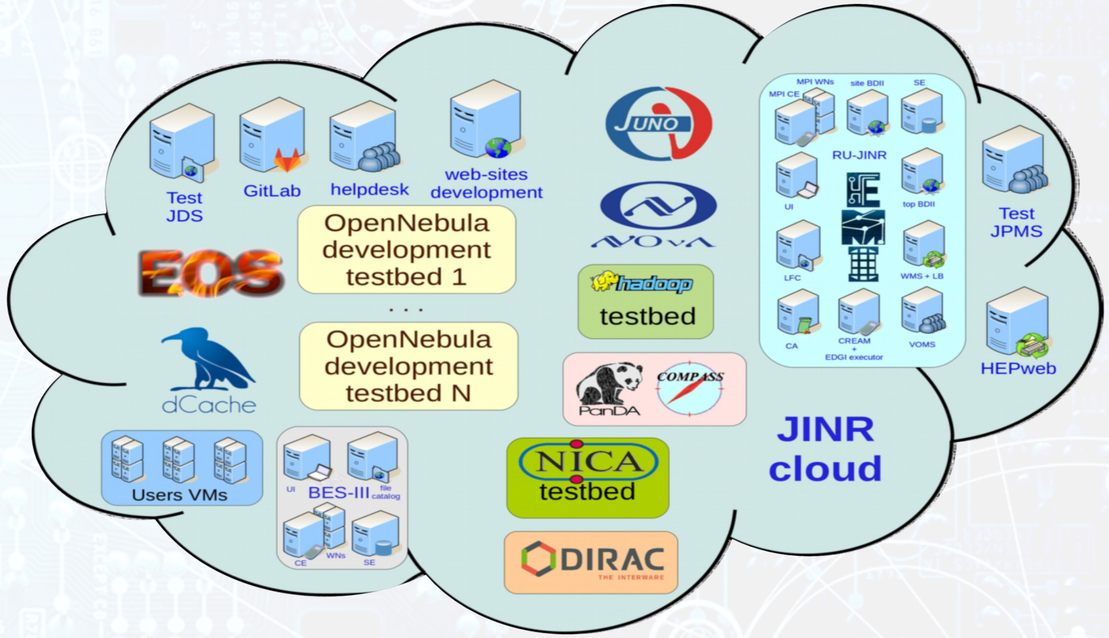
Based on the integration of the supercomputers of JINR, of the Interdepartmental Supercomputer Center of the Russian Academy of Sciences and of Peter the Great St. Petersburg Polytechnic University, a unified scalable supercomputer research infrastructure based on the National Research Computer Network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



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



- 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

Support for the JINR Neutrino Program

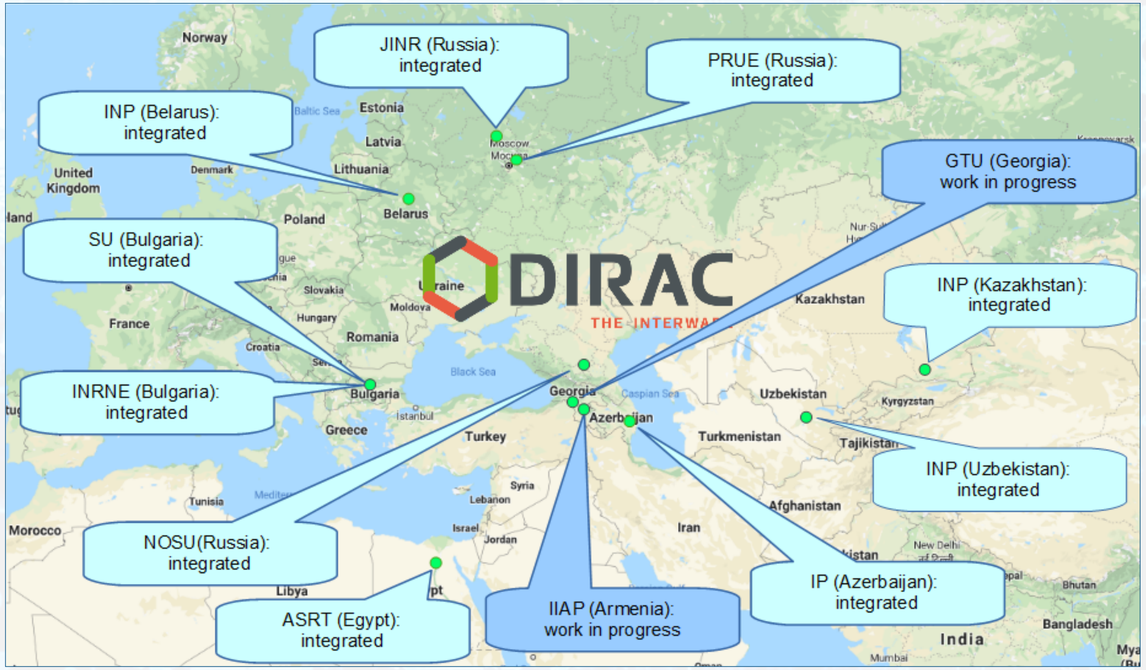
MLIT contribution:
engineering infrastructure (electricity, UPS, cooling, network, racks, manpower)

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 consumers** of the cloud infrastructure.

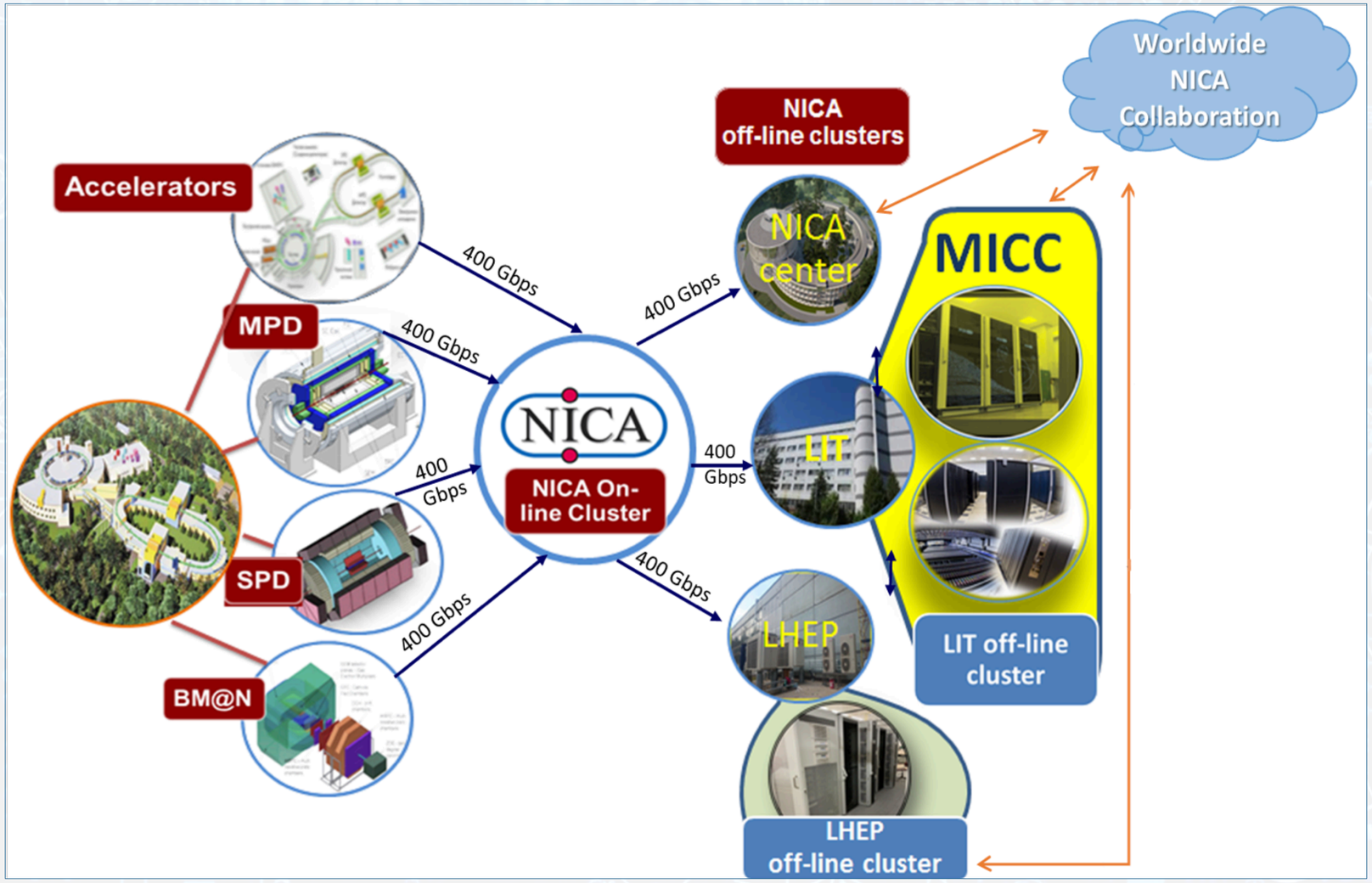
Heterogeneous Distributed Computing Environment

A heterogeneous computing environment, based on the DIRAC platform, was created for processing and storing data of experiments conducted at JINR. Computing facilities in MLIT, cloud infrastructures 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 through DIRAC. The distributed infrastructure is used by the MPD, Baikal-GVD, BM@N, SPD experiments.

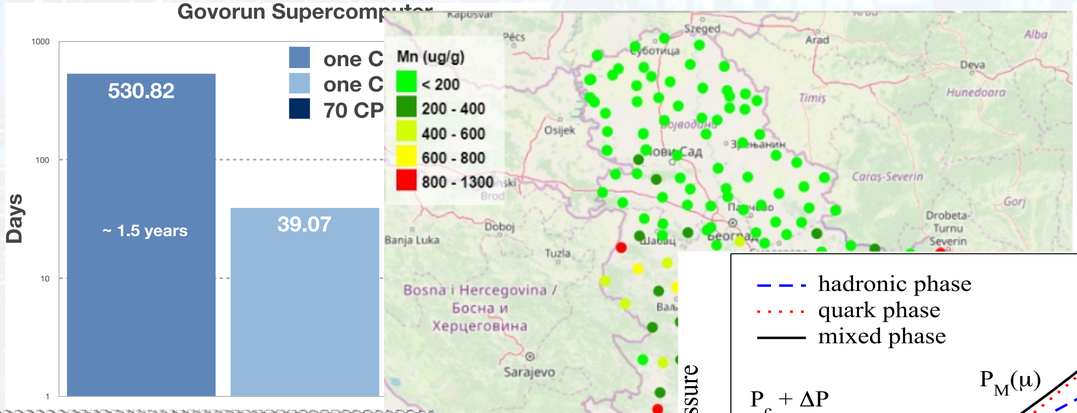


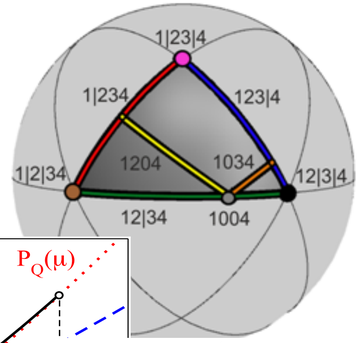
DIRAC-based distributed information and computing environment (DICE)

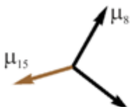
Tier 0 for NICA Computing

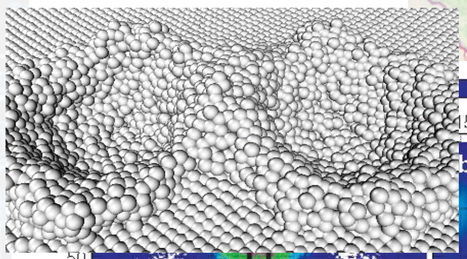


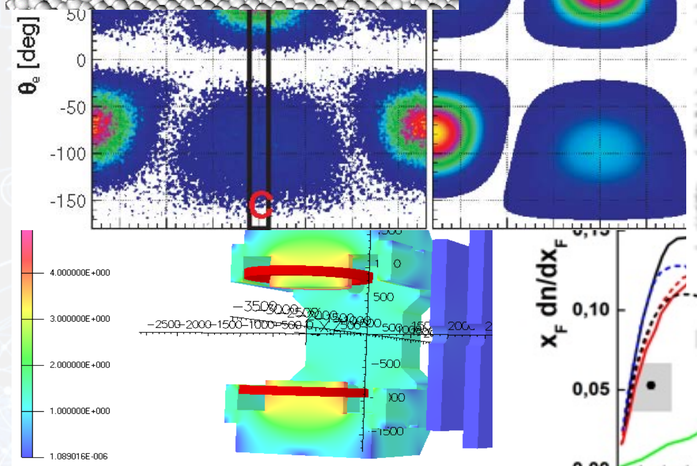
Methods, Algorithms and Software

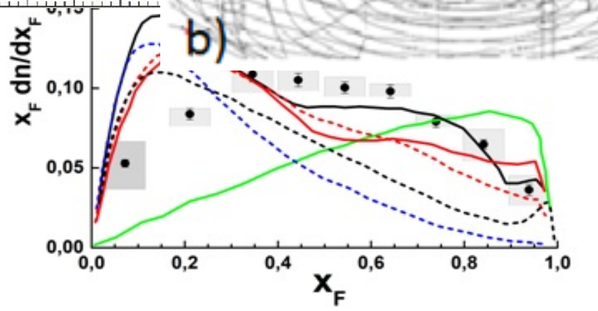


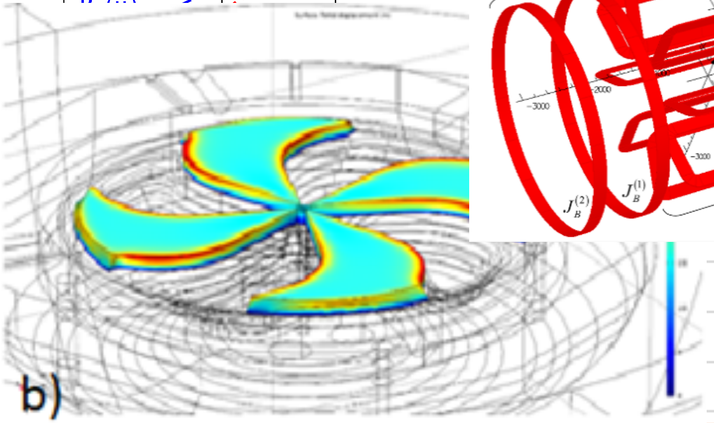


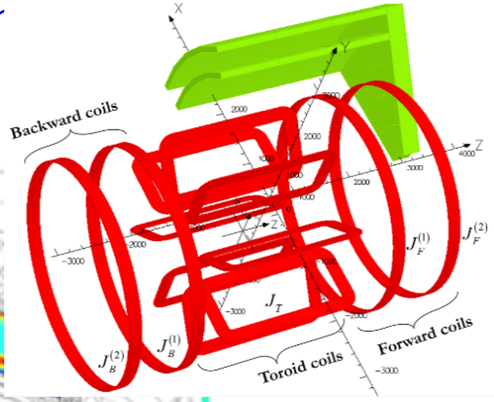


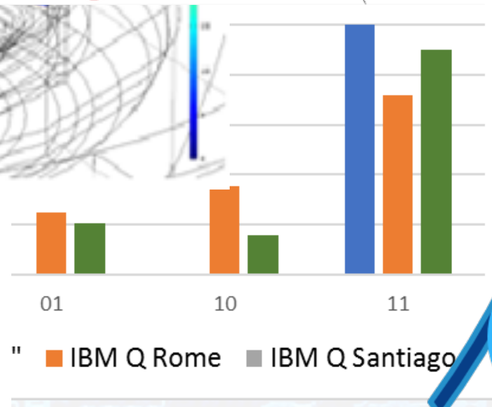








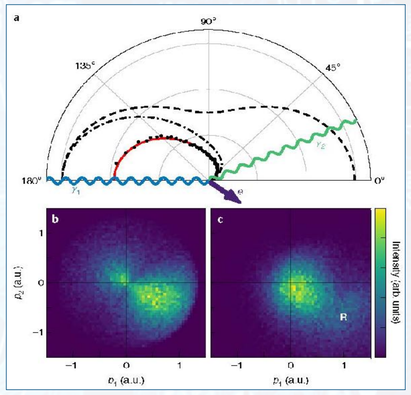




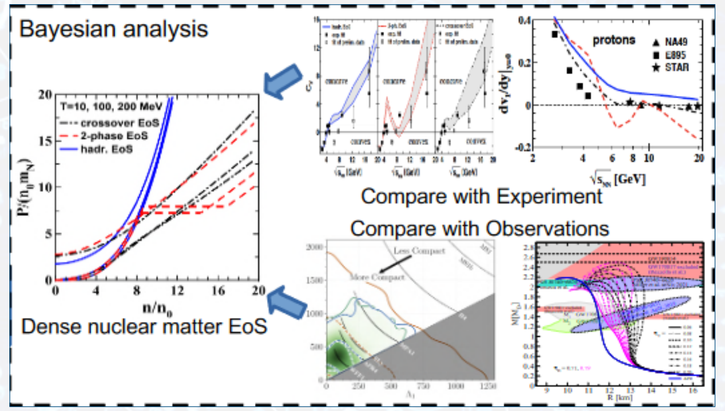
- Numerical modeling of complex physical systems
- Experimental data processing and analysis
- Big Data
- Machine and Deep learning
- AI and robotics
- Computer algebra
- Quantum computing

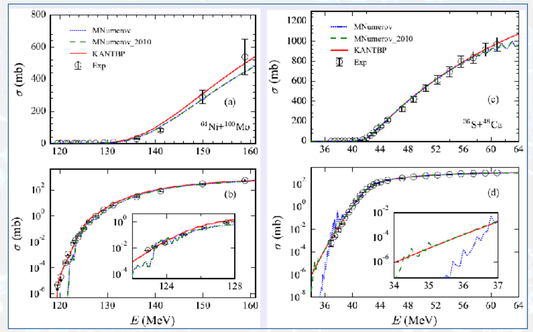
Methods of Mathematical Modelling, Computational Physics, and High-Performance Computing for Complex System Studies at JINR

- **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=2.1keV



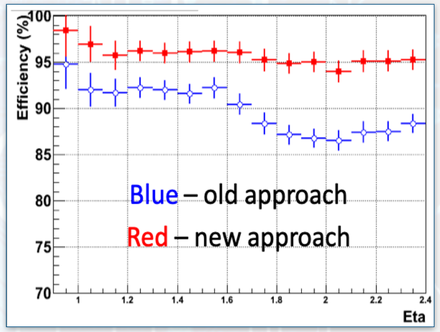


Sub-barrier fusion and fission reactions of heavy nuclei

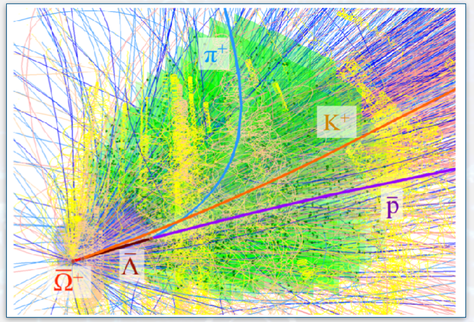
In 2020-2022, over **150** publications in peer-reviewed scientific journals have been done in cooperation with colleagues from other JINR Laboratories and Member States; 4 problem-oriented software packages in the JINRLIB electronic program library; 2 computer programs were published in the CPC program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

Mathematical Methods and Software for Experimental Data Processing and Analysis

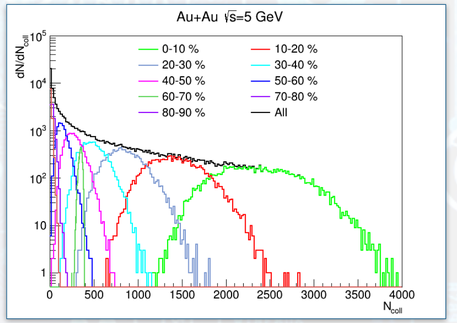
- **Physical processes modelling**
  - event generations
  - GEANT-simulation of experimental setups
- **Event reconstruction & data analysis**
  - particle trajectory reconstruction
  - particle identification
  - physical processes reconstruction
  - data analysis
- **Applied software and Data Bases**
  - Information systems for experimental services
  - experimental software frameworks
  - data modelling and data processing
  - event visualization and monitoring



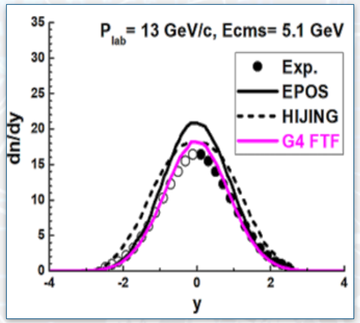
Effective algorithm for TeV muons reconstruction in CMS



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



Monte-Carlo Generator DCM-QGSM-SMM for NICA



Monte-Carlo Generator DCM-QGSM-SMM for NICA

Experiments IS:

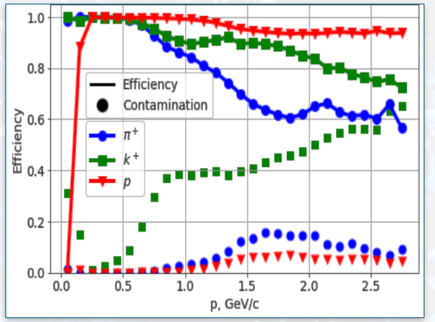
- ATLAS EventPickingService,
- ATLAS CREST,
- Geometry and Configuration DBs for BM@N

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

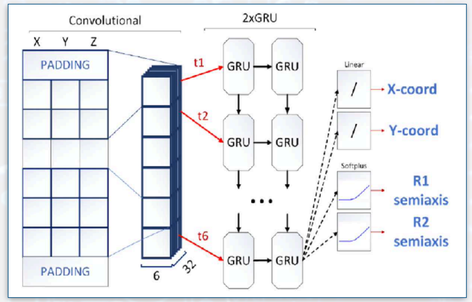
Scientific and practical significance: **expanding the scope of machine learning methods**, in particular, in high energy physics; software for experimental **data processing and analysis** at the NICA accelerator complex; corresponding development of root-frameworks.

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

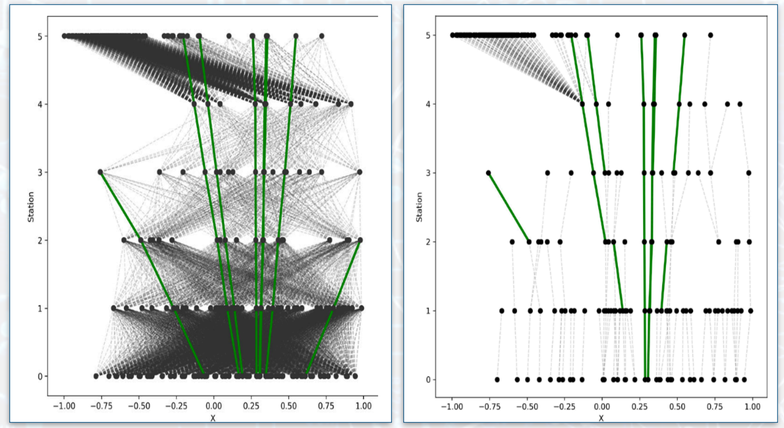
ML/DL methods under study: Recurrent Neural Networks, Graph Neural Networks, Convolutional Neural Networks, Decision Trees, Gradient Boosting.



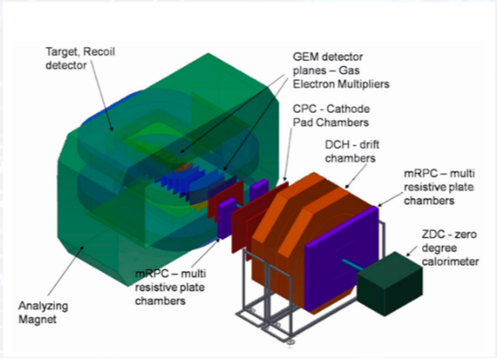
Gradient-boosted decision trees for PID in MPD

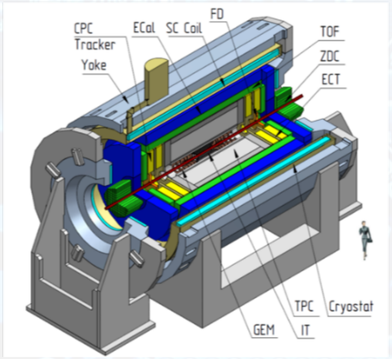


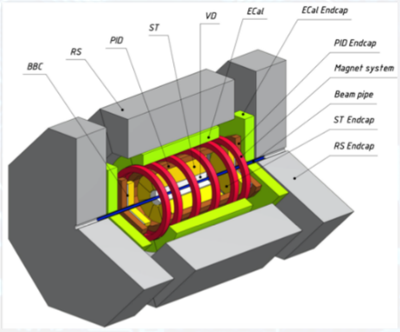
Deep GNNs for solving tracking problems in BM@N, BESIII, SPD



Graph Neural Networks for Tracking

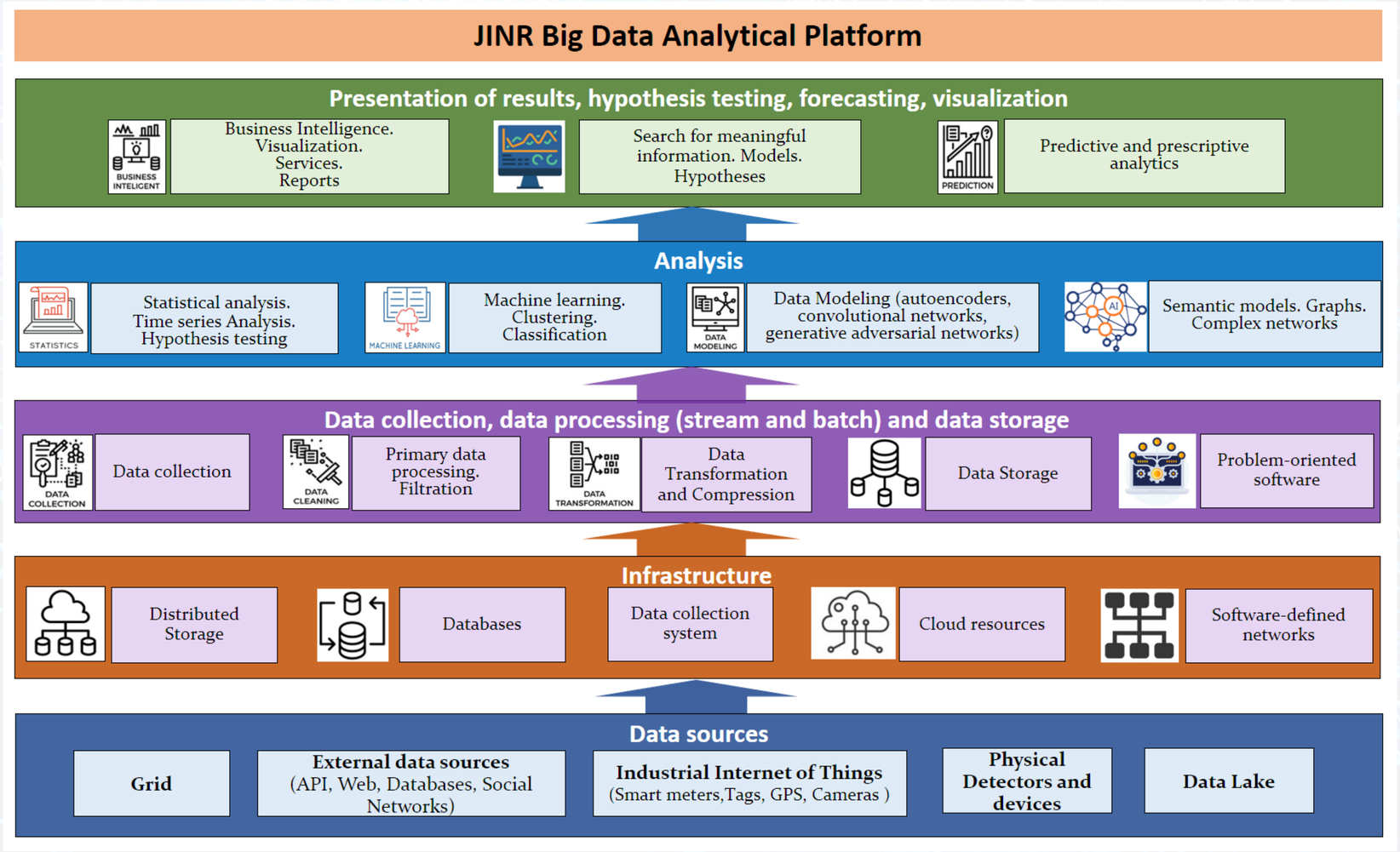






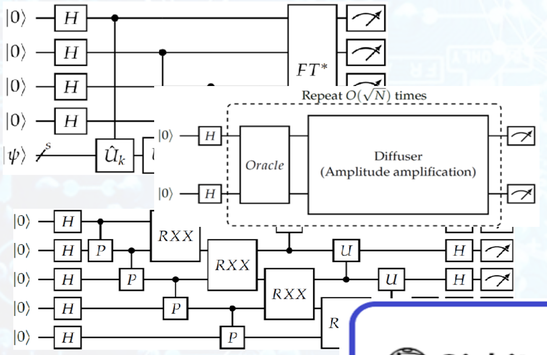
Methods of Artificial Intelligence and Big Data Analytics

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



Quantum 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.

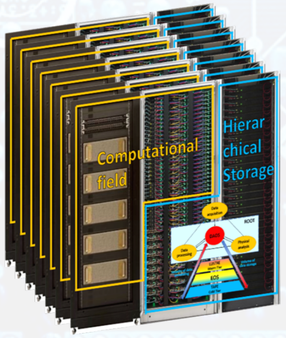


Quantum algorithms

Quantum simulators

Qiskit
QuEST
qsim
CUQUANTUM

SC “Govorun”

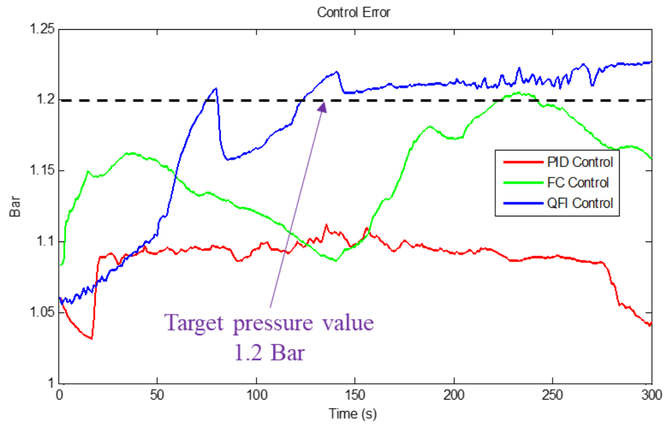


- Form a list of QAs required to solve tasks within the studied physical models
- Select the type of quantum simulator to simulate a classical architecture on computers
- Define resources for the selected quantum-limiting capabilities of available computing simulators (number of qubits and computation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

Intelligent Control of Technological Processes and Physical Facilities at JINR

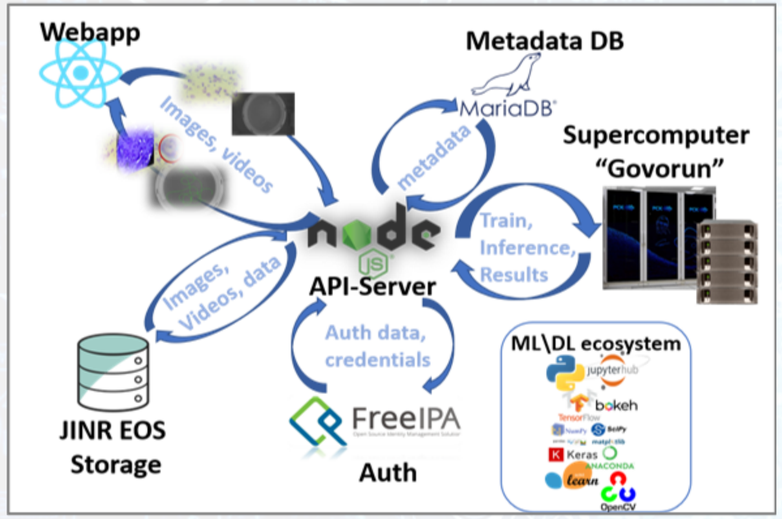
- Development of built-in self-organizing controllers for systems of the **intelligent control of technological processes, devices and facilities** at JINR (including for unforeseen and unpredictable situations).
- Development of an artificial intelligence platform based on ***quantum optimization*** for the tasks of **intelligent cognitive robotics and intelligent control** in JINR projects.
- 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.

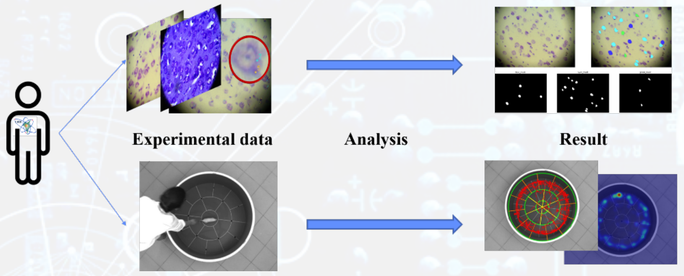


The quantum controller demonstrated the highest speed in achieving the target value, low overshoot and accuracy of achieving the control goal compared to other types of controllers.

Information System for Radiation Biology Studies

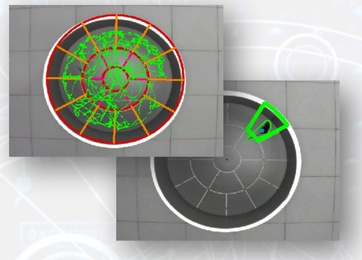


Conceptual scheme of the service



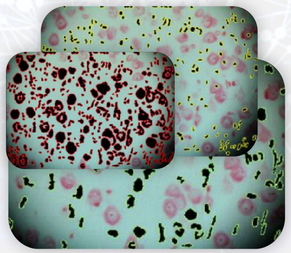
The joint project of MLIT and LRB is focused on creating an Information System (IS) as a set of IT solutions.

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.

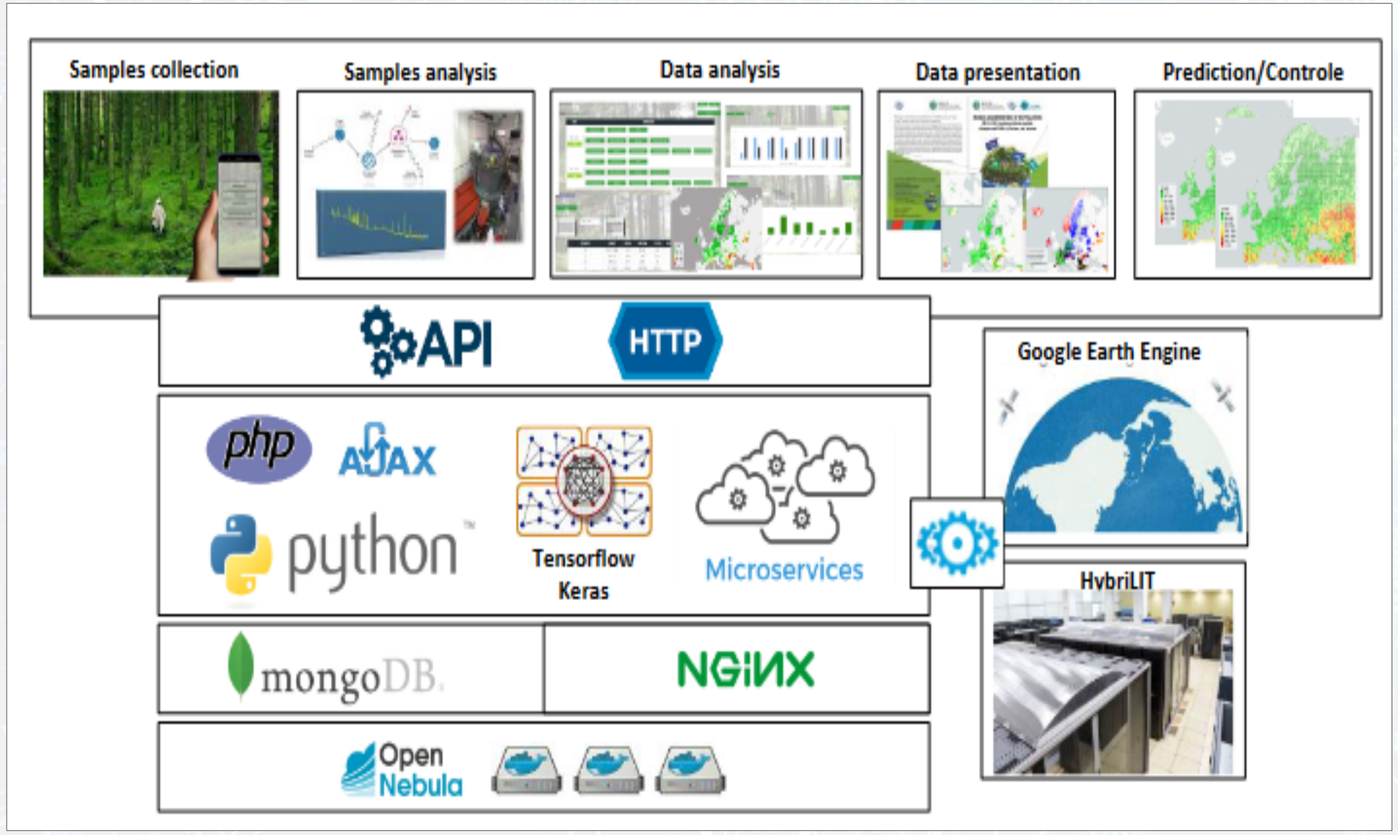


Tasks of the IS algorithmic block

- 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 research life cycle,

administrative and social activities;

to provide:

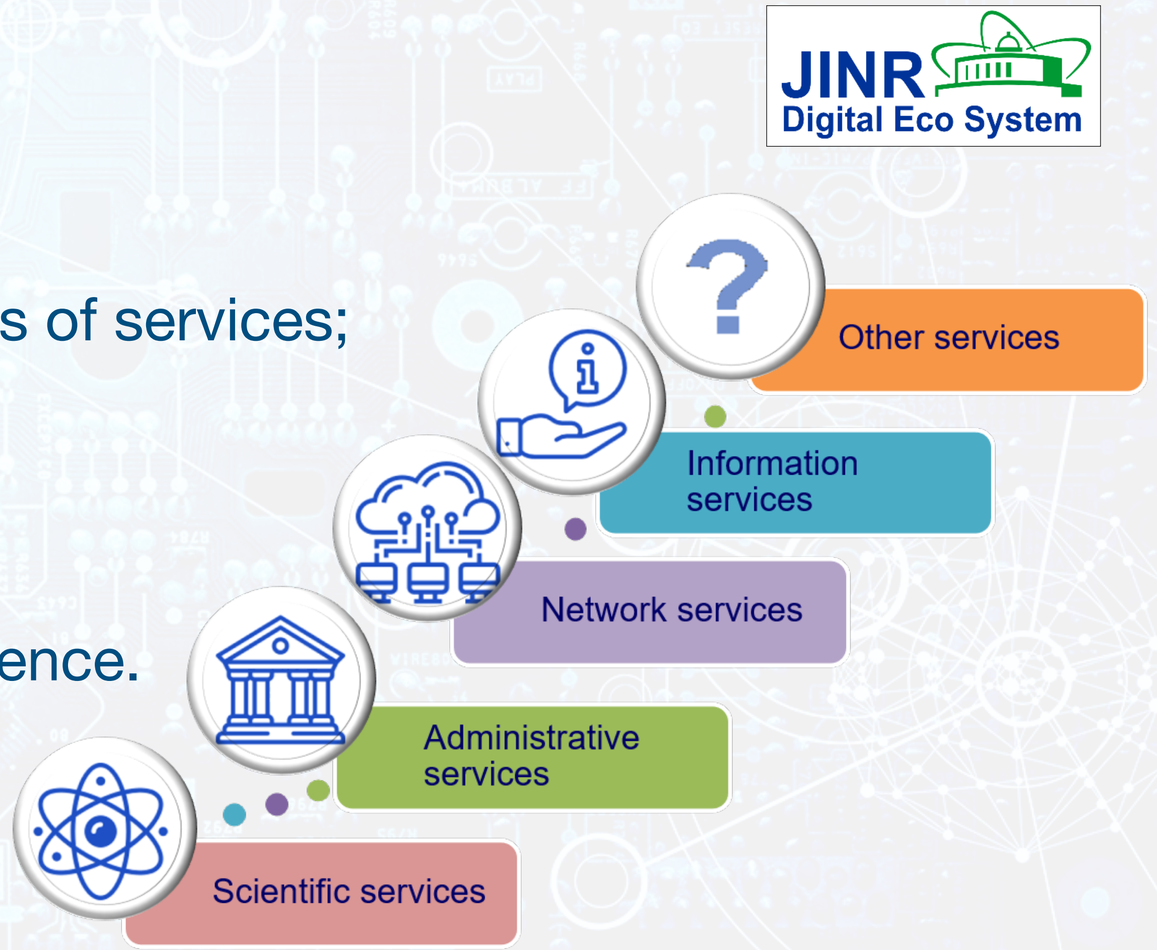
reliable and secure access to various types of services;

to enable:

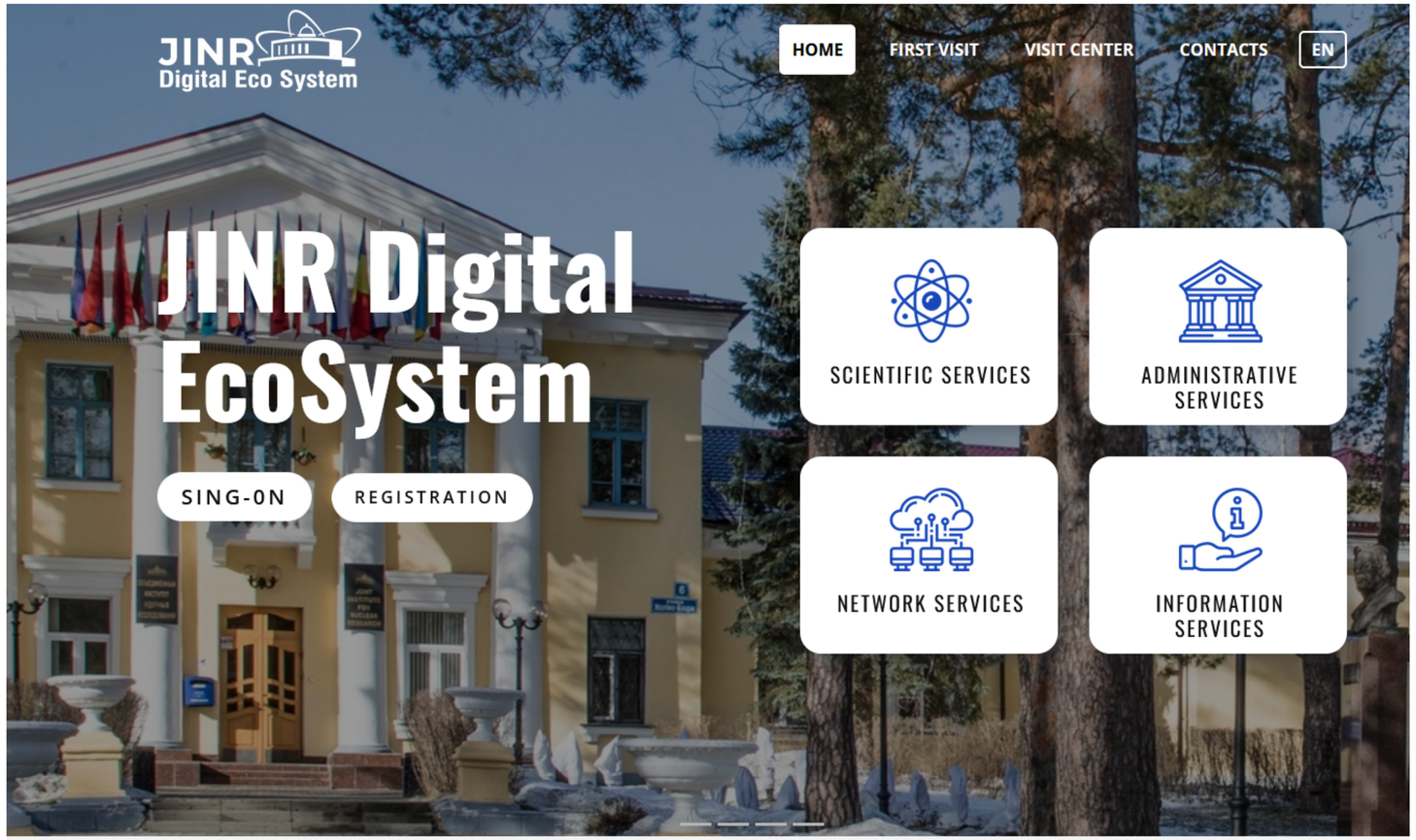
a comprehensive analysis of information;

using:

Big Data technologies and artificial intelligence.



JINR Digital EcoSystem. Entry point



Digital Ecosystem

This is a complex digital environment that combines a large number of information services and business processes based on the principles of mutually beneficial relationships (“win-win”)

- Personal account of a JINR employee
- 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

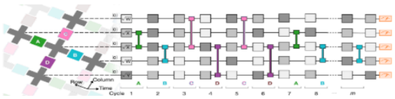
jupyterhub

TensorFlow

NumPy

scikit learn

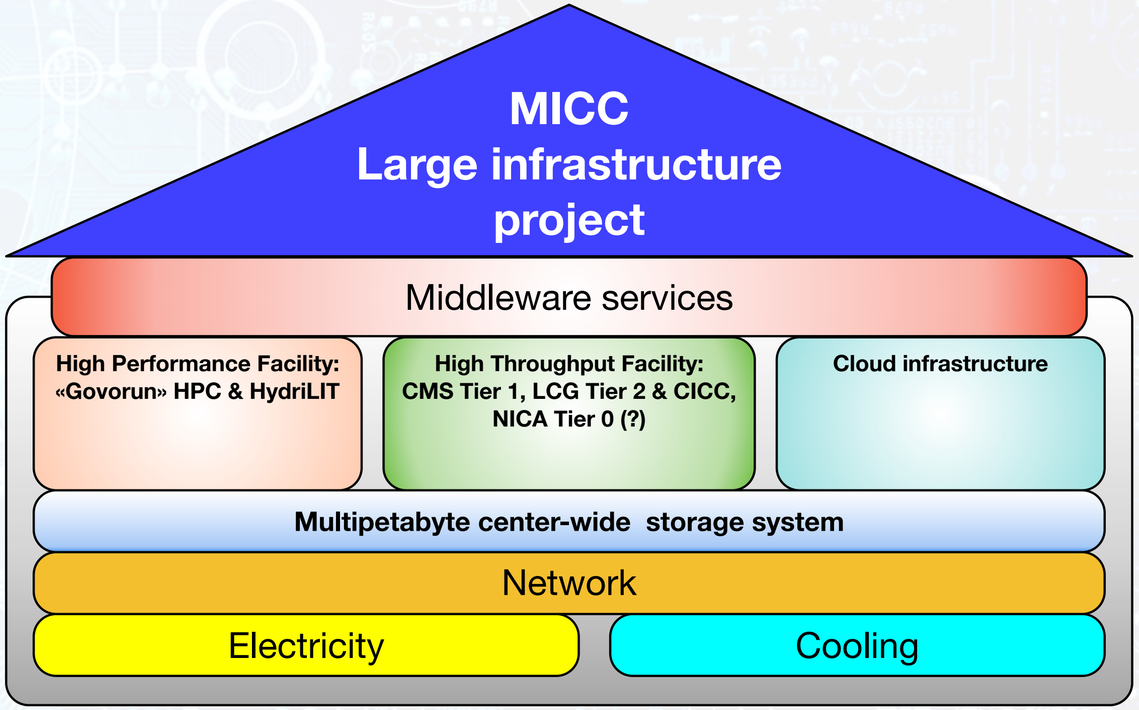
Quantum algorithms, quantum programming and quantum control



MICC - Large 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 (CMS Tier1, LCG Tier2 & CICC)
HP Computing (“Govorun” HPC)
Cloud Computing
Center-wide Data storage



Scientific directions and information systems

- Methods of Mathematical Modelling, 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
- Methods of Artificial Intelligence and Big Data Analytics
- Intelligent control of technological processes and physical facilities at JINR

Education: JINR IT School

Thank you for attention!