INTERNATIONAL INTERGOVERNMENTAL ORGANIZATION
МЕЖДУНАРОДНАЯ МЕЖПРАВИТЕЛЬСТВЕННАЯ ОРГАНИЗАЦИЯ

JOINT INSTITUTE FOR NUCLEAR RESEARCH
ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ

THE MESHCHERYAKOV LABORATORY OF INFORMATION TECHNOLOGIES

Korenkov Vladimir

Director MLIT

DLCP, Dubna
6 july 2022

6th International Workshop on Deep Learning in Computational Physics (DLCP-2022)

6-8 July 2022

Europe/Moscow timezone

We are pleased to invite you to participate to the DLCP-2022 – **The 6th International Workshop on Deep Learning in Computational Physics** which will be held in the Meshcheryakov Laboratory of Information Technologies (MLIT) of the Joint Institute for Nuclear Research (JINR) **on July 6-8, 2022 in a mixed format.**

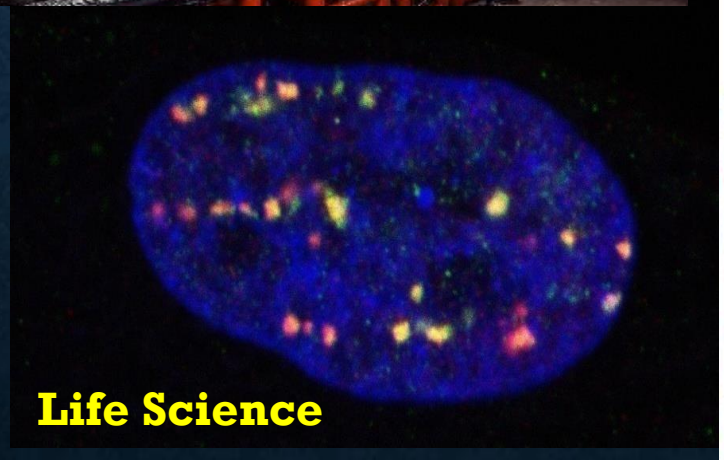
The workshop primarily focuses on the use of machine learning in particle astrophysics and high energy physics, but is not limited to this area. Topics of interest are various applications of artificial neural networks to physical problems, as well as the development of new modern machine learning methods for analyzing various scientific data, including big data.

A poster section for students will be held within the workshop only in face to face format.

The working language is English.

All relevant information on the workshop website: dlcp2022.jinr.ru.

Implementation of the JINR Development Program



International Large-scale projects

Russian research institutes and universities actively participate in international large-scale projects:

- LHC, CERN (experiments: ATLAS, ALICE, LHCb, CMS)
- XFEL, DESY (European free electron laser)
- ESRF, France (European synchrotron center)
- FAIR, GSI, Germany (CBM, PANDA experiments)
- ITER, France …

International large-scale projects are being prepared in Russia:

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

Joint Institute for Nuclear Research
SCIENCE BRINGING NATIONS TOGETHER

СКИФ
СИБИРСКИЙ КОЛЬЦЕВОЙ ИСТОЧНИК ФОТОНОВ

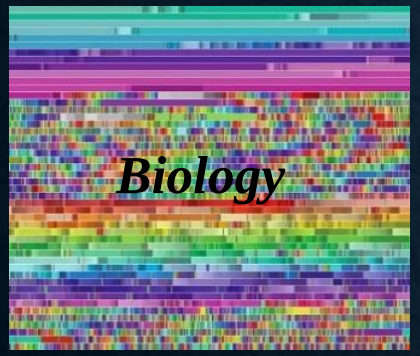
Институт ядерной физики имени Г. И. Будкера СО РАН

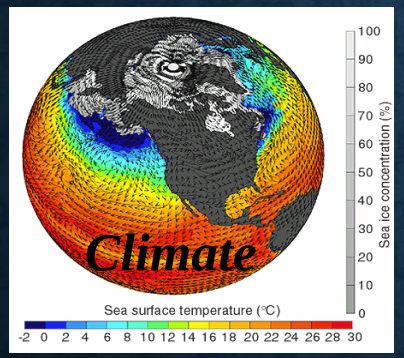
HPC+Big Data+Artificial intelligence

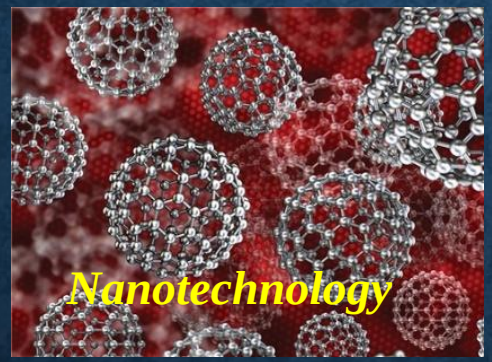
High Energy Physics

CERN

CERN Large Hadron Collider > 600 Pb/Year







Square Kilometer Array radio telescope
> 1 Eb/Year ra
data (estimatic

SKA
SQUARE KILOMETRE ARRAY

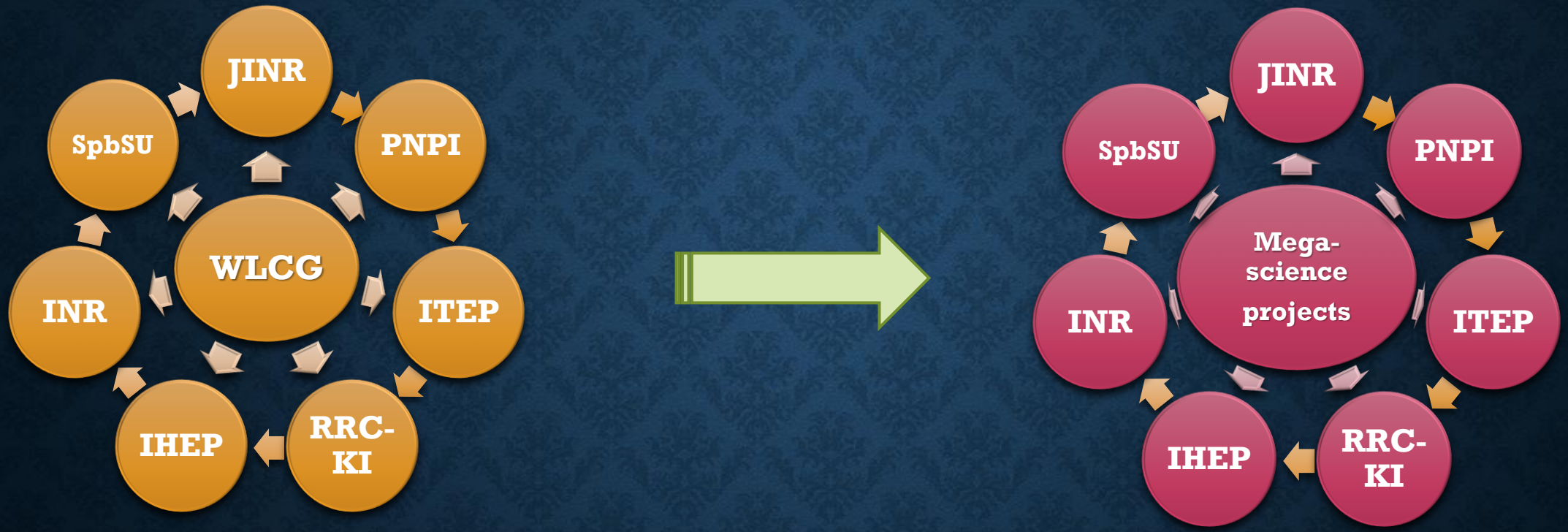
LSST
Large Synoptic Survey Telescope

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

... et cetera

From RDIG to RDIG-M

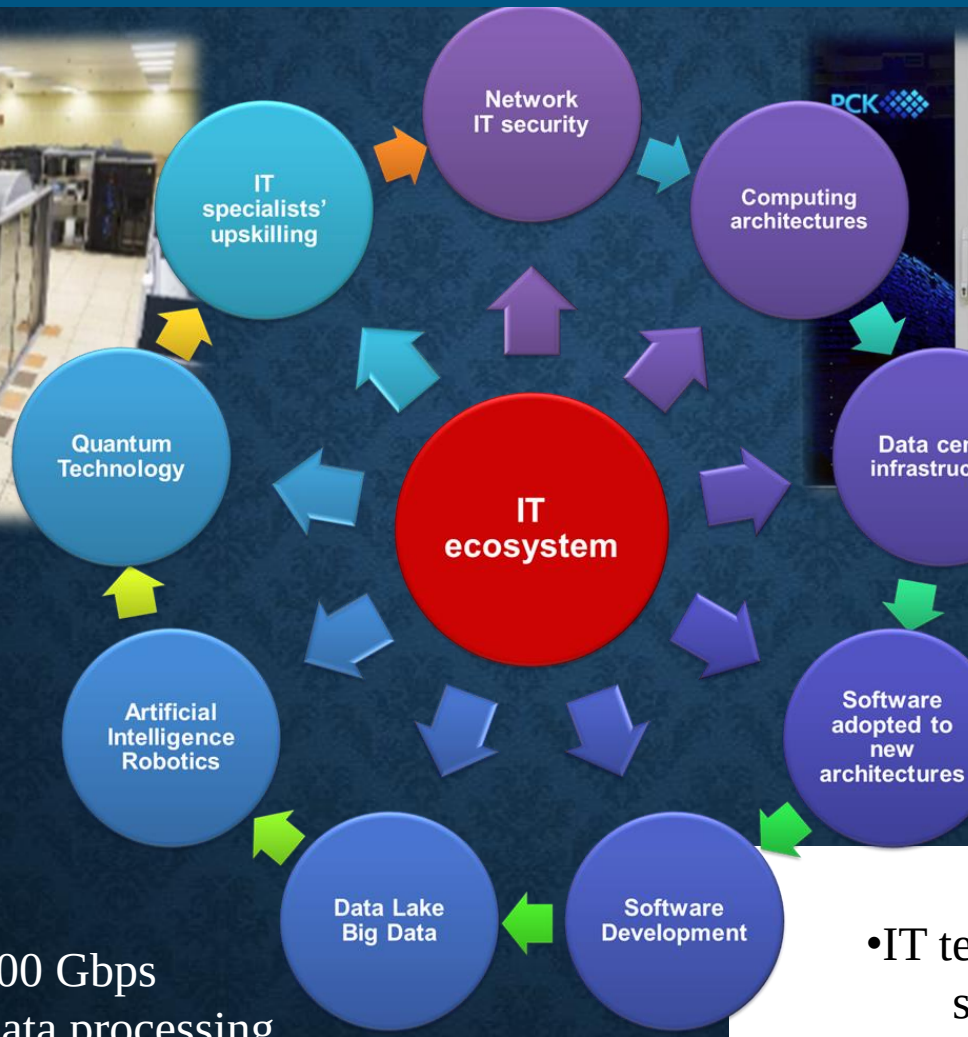
Consortium RDIG-M – Russian Data Intensive GRID for Megascience projects



Проекты класса мегасайенс в области физики высоких энергий и астрофизики, синхротронно-нейтронных исследований, нейтринной программы.

- Развитие методов и алгоритмов искусственного интеллекта для анализа данных экспериментов;
- Развитие суперкомпьютерных, грид, облачных технологий, распределенных хранилищ данных для моделирования, обработки и анализа данных;
- Развитие технологий высокоскоростной передачи научных данных.

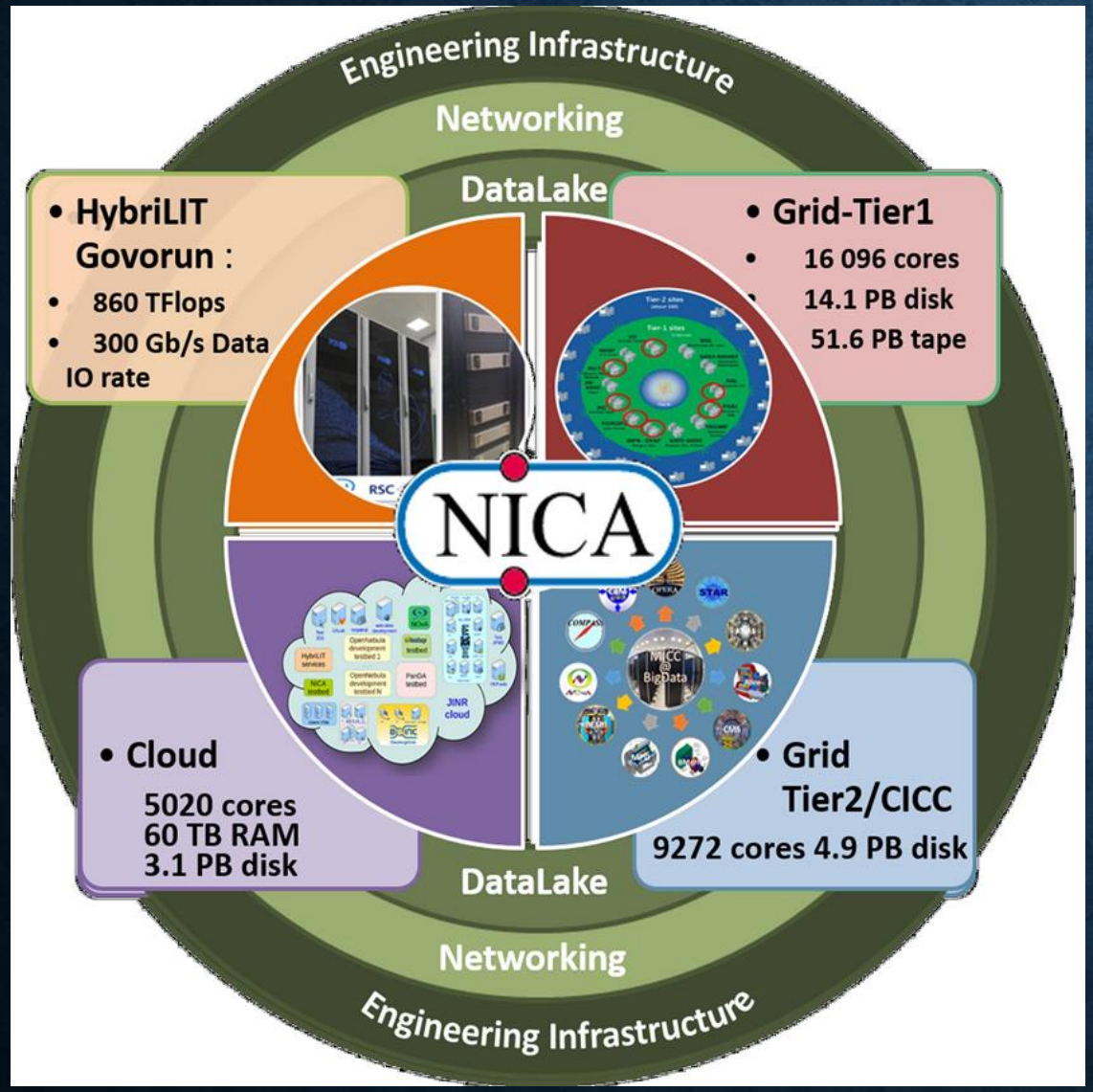
MLIT today



Staff: 325
Scientists: 100
Doctors of Science: 24
Candidates of Science: 61
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

•CONCEPT of the development of
•IT technologies and scientific computing aimed at solving the strategic tasks of JINR through the introduction and development of a whole range of advanced IT solutions, integrated into a unified computing environment that combines a variety of technological solutions, concepts and methods.

Multifunctional Information and Computing Complex at JINR



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

- multifunctionality,
- 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 IT infrastructure is one of JINR's basic facilities

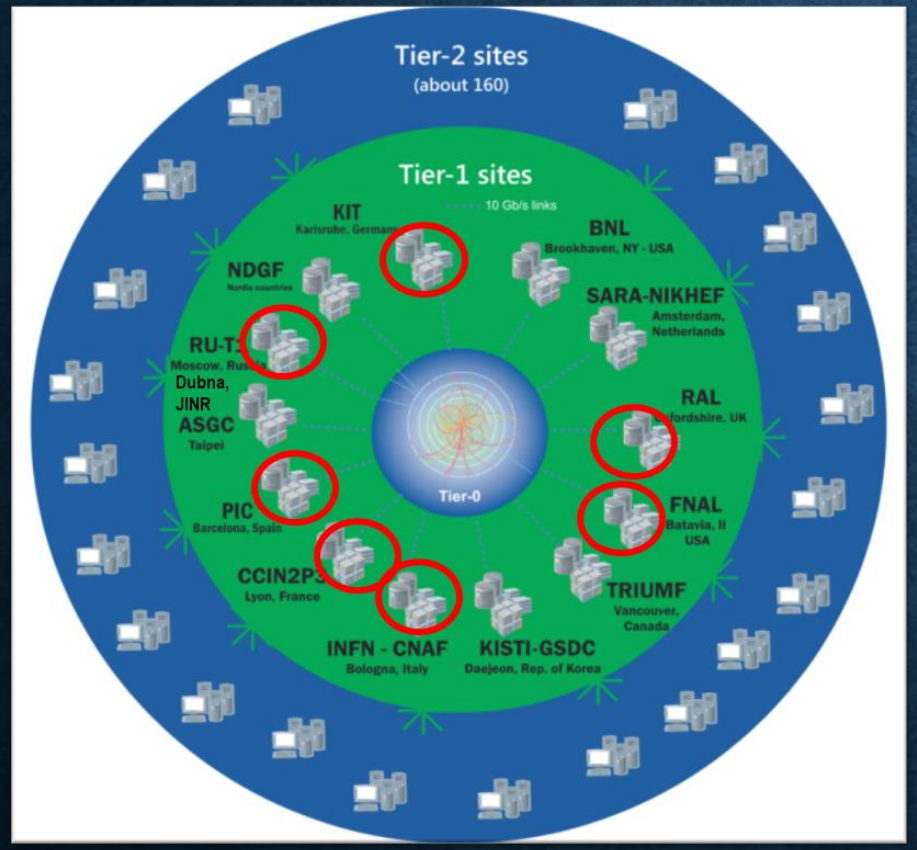
Worldwide LHC Computing Grid

WLCG: an international collaboration to distribute and analyze LHC data. Integrates computer centres worldwide that provide computing and storage resources into a unified infrastructure accessible to all LHC physicists.

Tier0 (CERN): data recording, reconstruction and distribution

Tier1: permanent storage, reprocessing, 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

~1M CPU cores

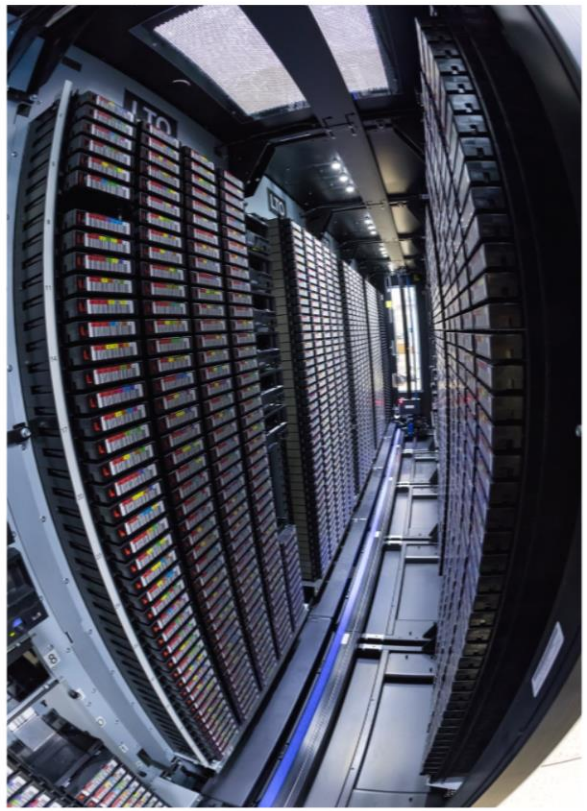
1 EB of storage

> 3 million jobs/day

10-100 Gb links

JINR is a part of the Worldwide LHC Computing Grid

JINR Tier1 for CMS



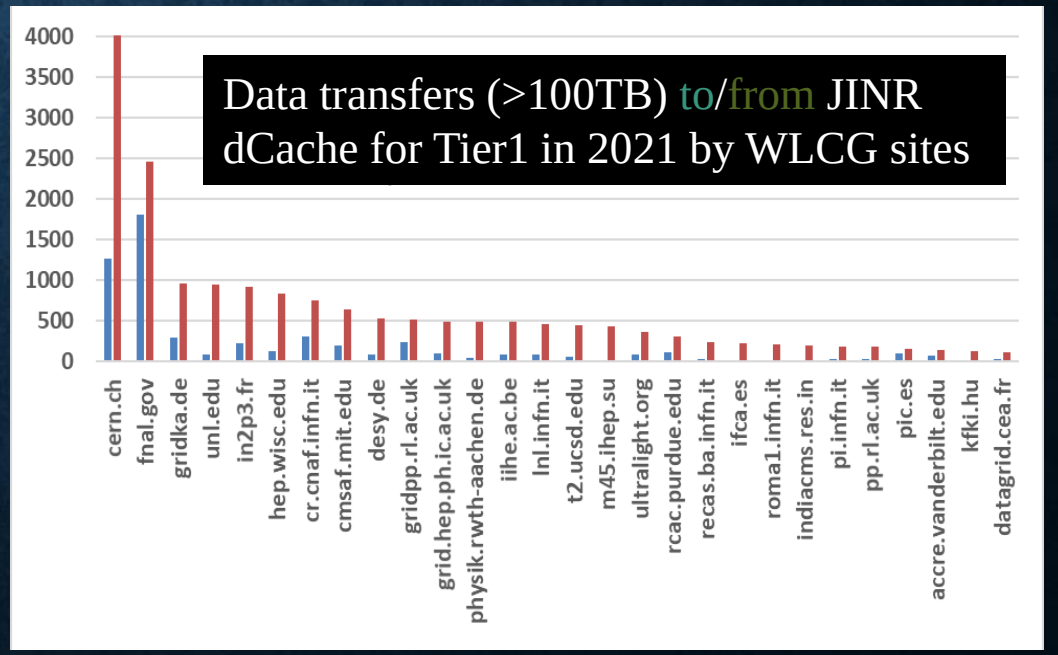
- 16096 cores
- 260
- 14.1 PB
- 52.6 PB
- 100%

The CMS Tier1 center at JINR has demonstrated stable work through the entire period since its launch into full operation. The Tier1 site for CMS is ranked first among world centers for CMS.

One of the main functions of Tier1 centers is to provide data exchange with all global sites that run CMS jobs. In 2021, more than 30.5 PB of data from more than 210 grid sites were transferred to and from our Tier1.

Tier1 — Sum CPU (HS06 hours) 2021-2022

RU-JINR-T1	2,069,498,327	27,88%
US-FNAL-CMS	1,612,815,957	21,72%
DE-KIT	1,460,440,934	19,67%
IT-INFN-CNAF	743,512,935	10,01%
FR-IN2P3	668,957,987	9,00%
UK-T1-RAL	601,454,698	8,09%
ES-PIC	263,218,979	3,54%



Tier2 for Experiments and the JINR Laboratories

Tier2 for Alice, ATLAS, CMS, LHCb, BES, BIOMED, COMPASS, MPD, NOvA, ILC, etc. is recognized the best in the Russian Data Intensive Grid (RDIG) Federation.

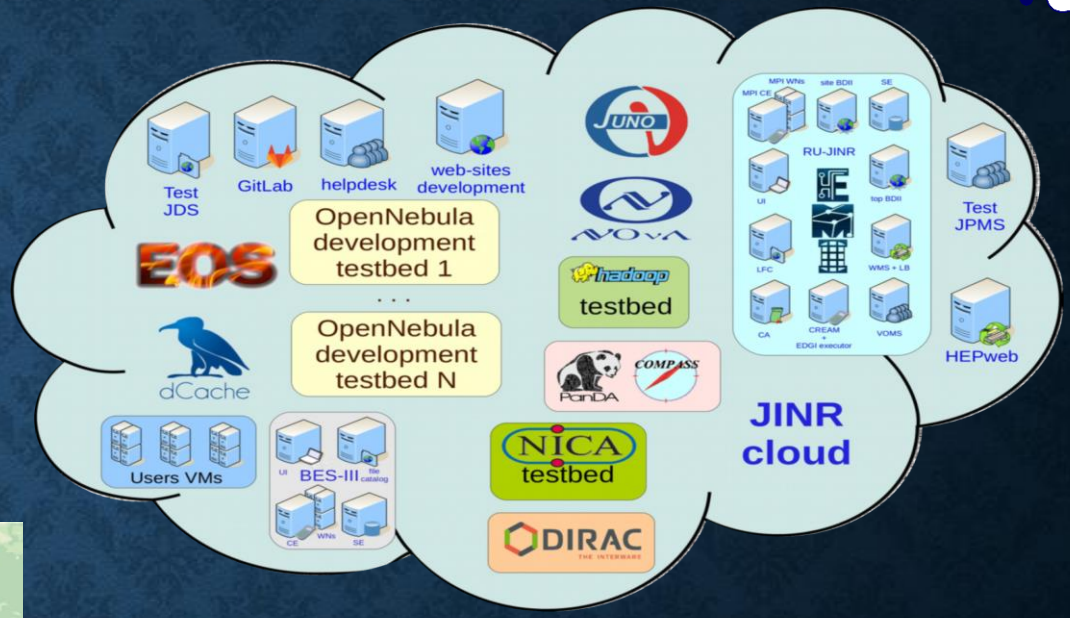
RDIG: distribution CPU time (HS2006) by organizations

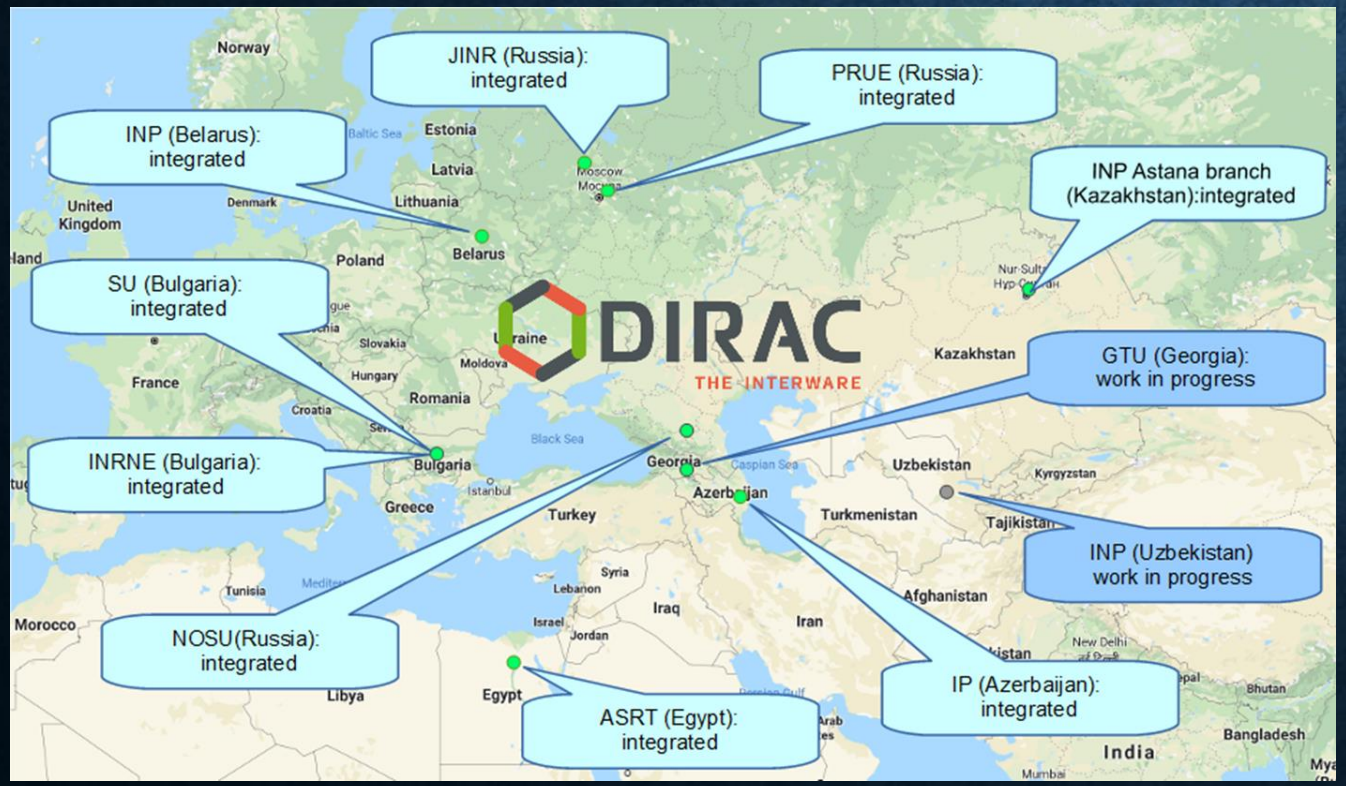
Site	Total	Percent
JINR-LCG2	1,024,918,744	81.12%
RU-Protvino-IHEP	148,746,062	11.4%
RU-SARFTI	27,863,245	2.14%
ru-PNPI	26,352,034	2.06%
RU-SPbSU	19,106,917	1.5%
Ru-Troitsk-INR-LCG2	8,943,304	0.7%
ITEP	8,263,710	0.6%

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

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





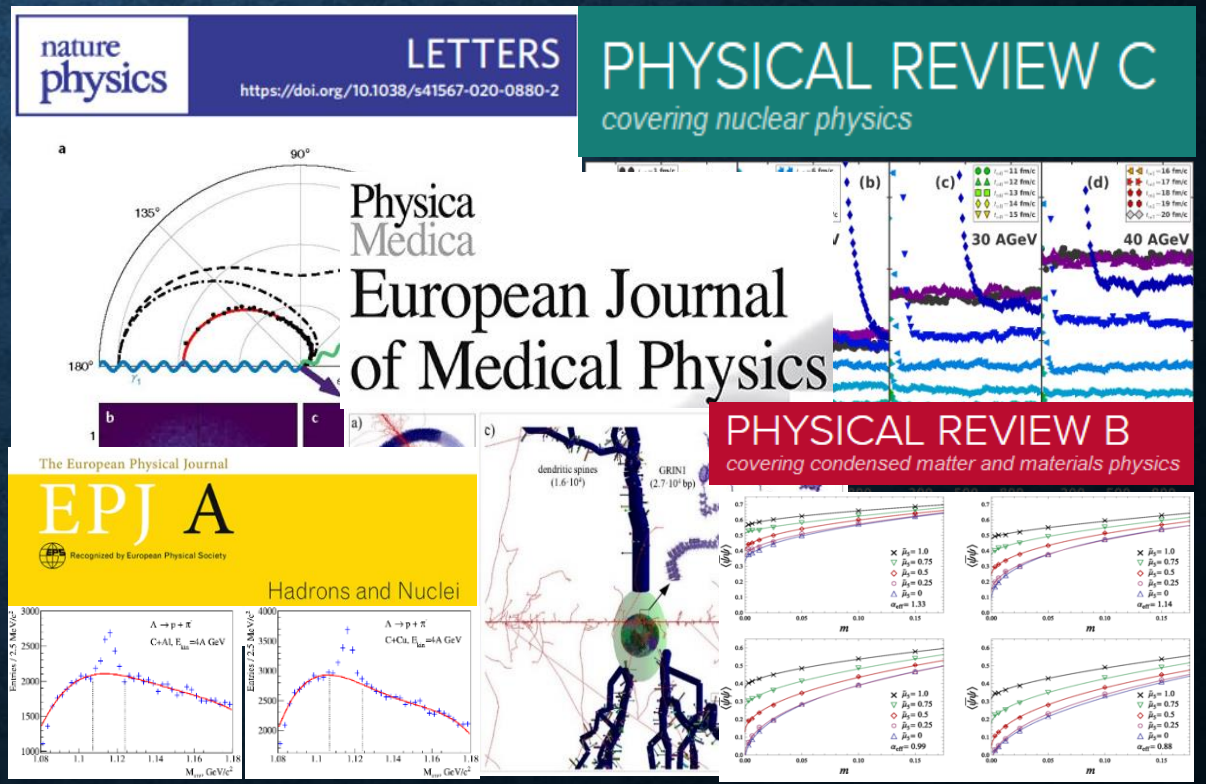
- VMs for JINR users,
- Computational resources for neutrino experiments,
- Testbeds for R&D in IT,
- COMPASS production system services,
- Data management system of the UNECE ICP Vegetation,
- Scientific and engineering computing,
- Service for data visualization based on Grafana, jupyterhub head and execute nodes for it,
- Gitlab and its runners, as well as some others.

“Govorun” Supercomputer

- Hyper-converged software-defined system
- Total peak performance: 860 TFlops DP
- Scalable solution Storage-on-demand
- Multilayered storage system for maximum efficiency
- Hot water cooling (compute, storage, interconnect)
- The most energy-efficient center in Russia (PUE = 1.06)
- Storage performance >300 GB/s
- The DAOS polygon of the “Govorun” supercomputer takes the 1st place among Russian supercomputers in the current IO500

Total number of registered users of the "Govorun" supercomputer: 517 (322 – JINR, 195 – Member States)

Key projects that use the resources of the “Govorun” supercomputer: 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.

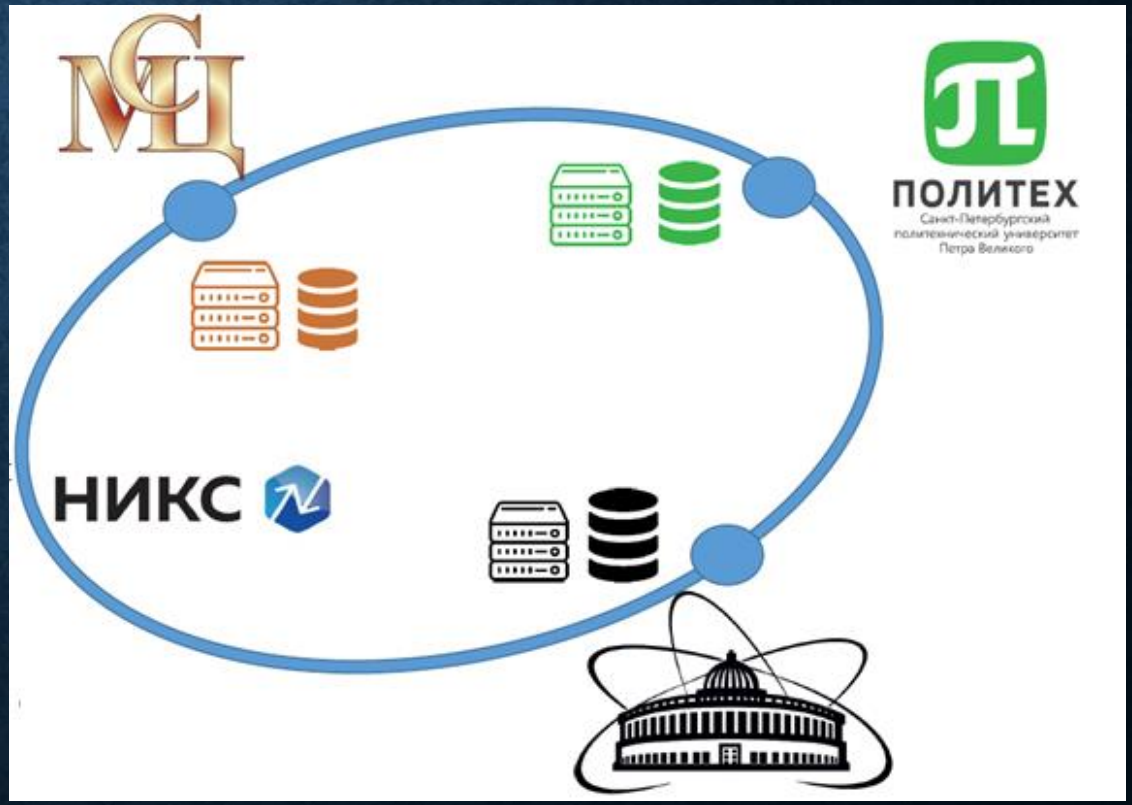


Unified Scalable Supercomputer Research Infrastructure

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 scalable research infrastructure of a new level was created. Such an infrastructure is in demand for the tasks of the NICA megascience project.

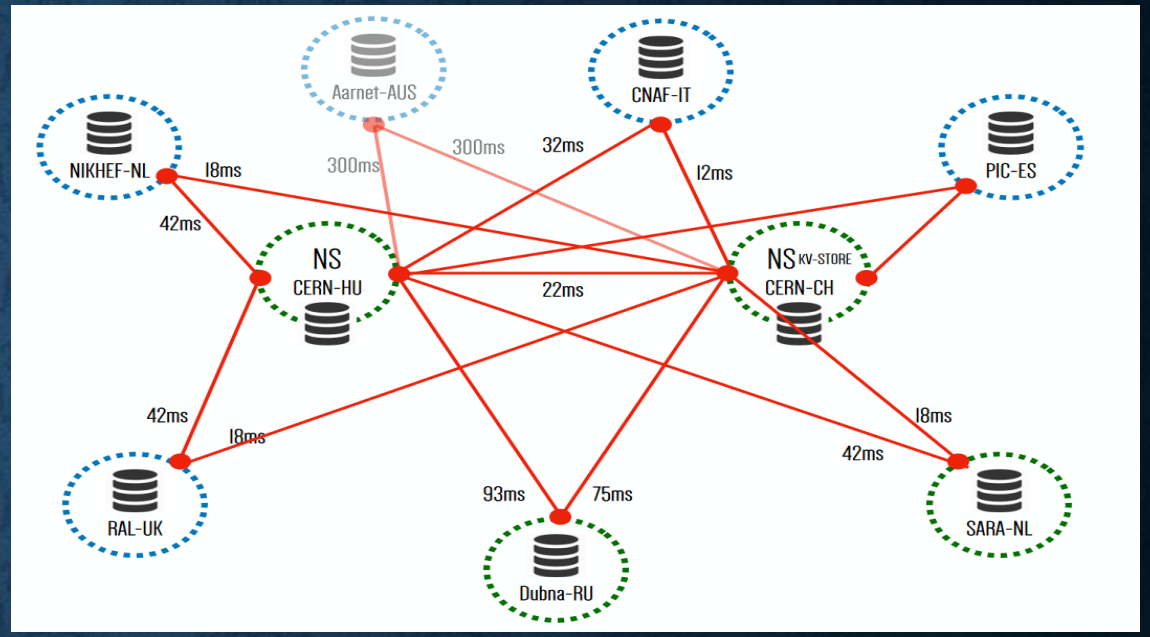
In January 2022, the first joint experiment on the use of the unified supercomputer infrastructure for the tasks of the NICA megascience project was successfully completed:

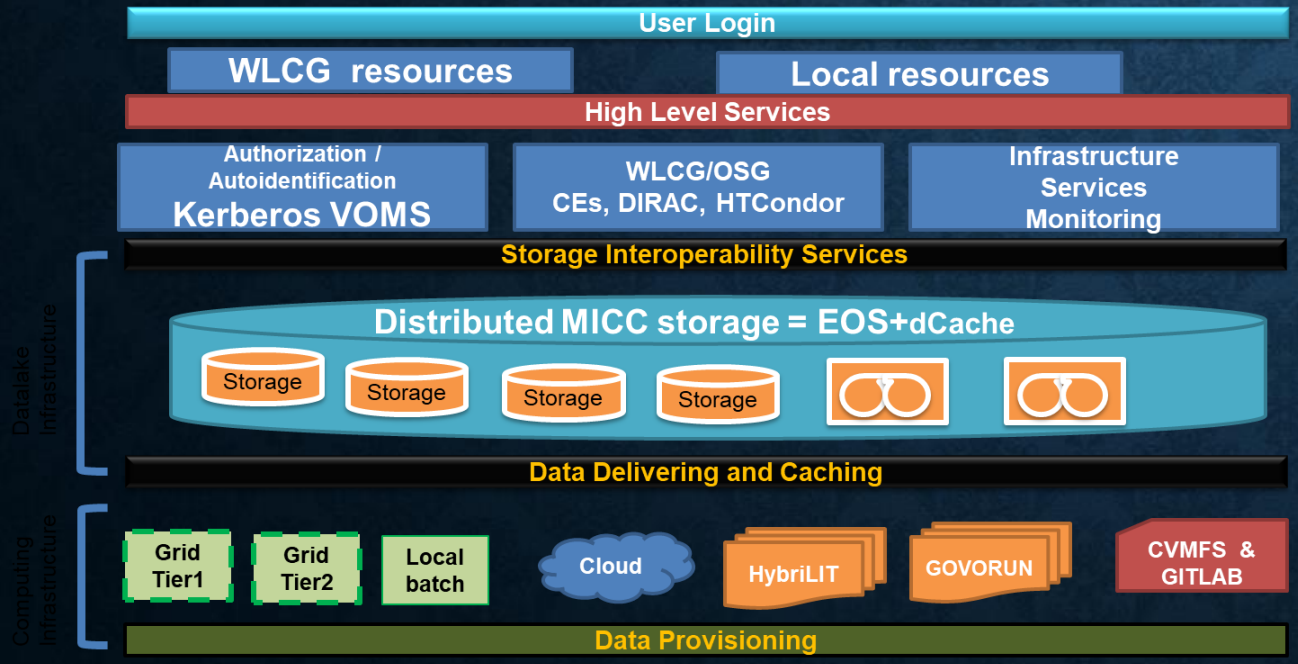
- ✓ 3,000 Monte-Carlo data generation and event reconstruction jobs were launched for the MPD experiment
- ✓ about 3 million events were generated and reconstructed
- ✓ the obtained data were transferred to Dubna for further processing and physics analysis.



JINR in DataLakes

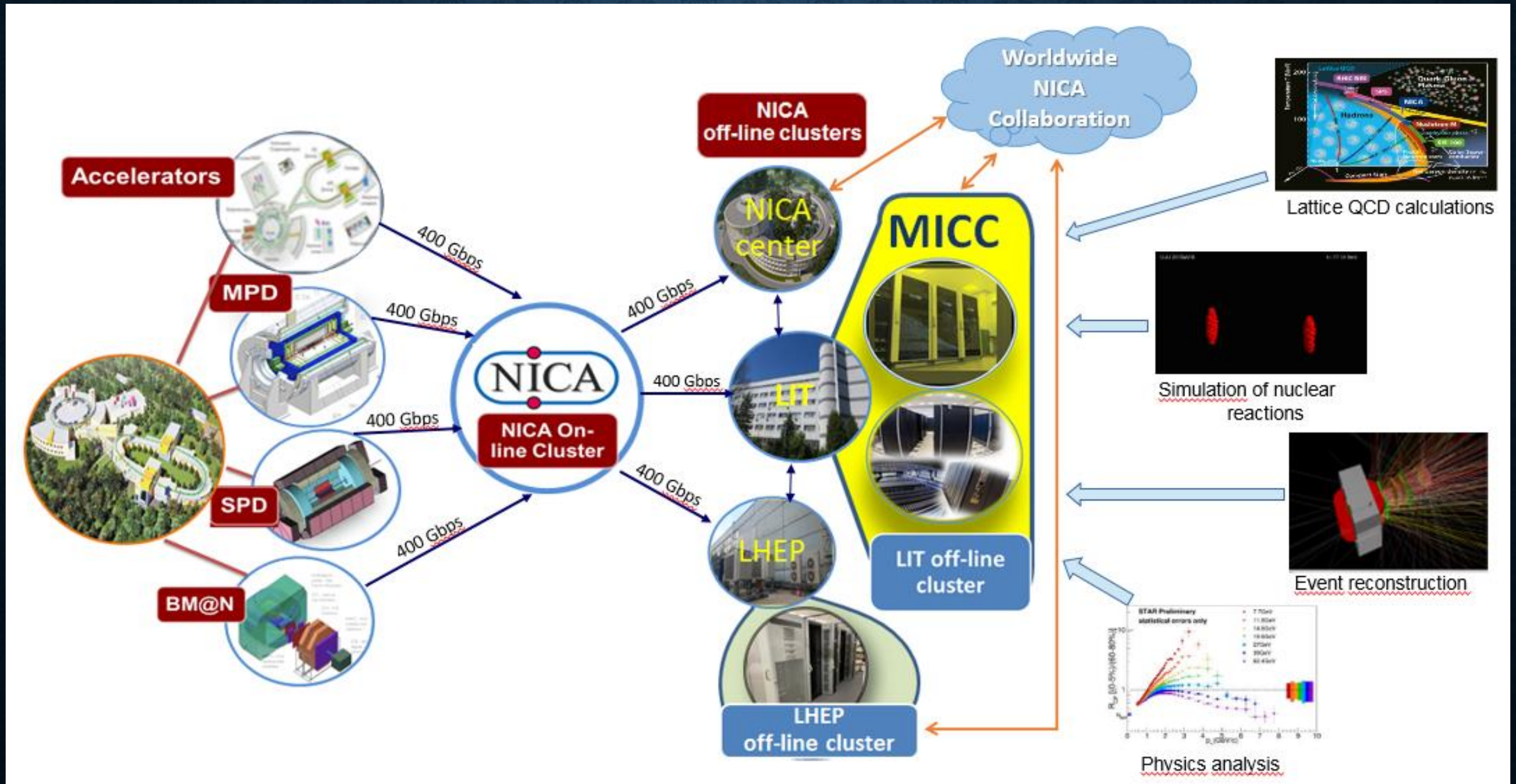
- The JINR data lake prototype was built as a distributed EOS storage system.
- EOS was successfully integrated into the MICC structure.
- EOS is used for storing and accessing big arrays of information.
- It can be applied for collective data simulation, storage of raw data gathered from experimental setups, data processing and analysis.





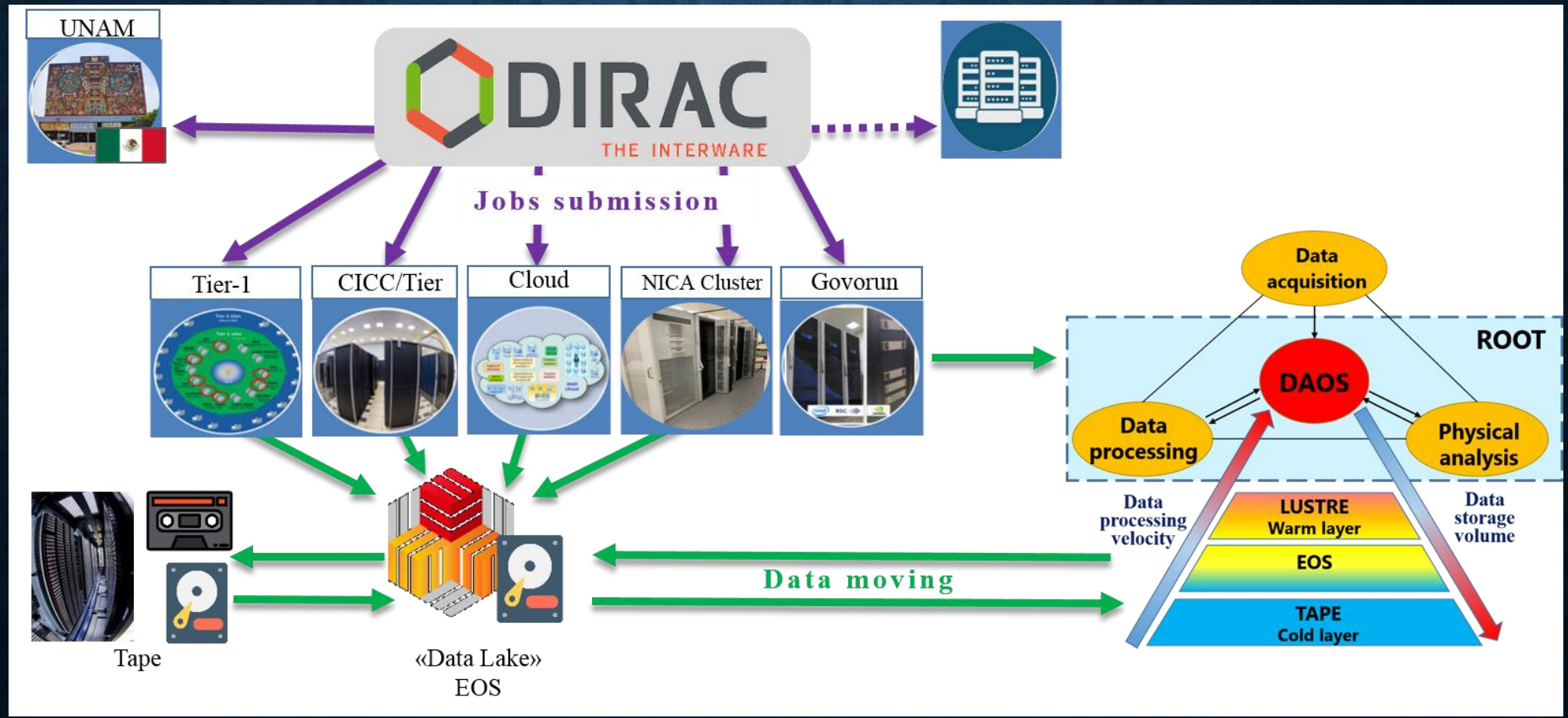
- There is currently 16.7 PB of disk space available for EOS.
- Baikal-GVD, DANSS, FOBOS, JUNO, BM@N, MPD, SPD, PANDA are its major users.
- EOS is visible as a local file system on the MICC working nodes and allows authorized users (by the kerberos5 protocol) to read and write data.

Offline Computer Complex for NICA



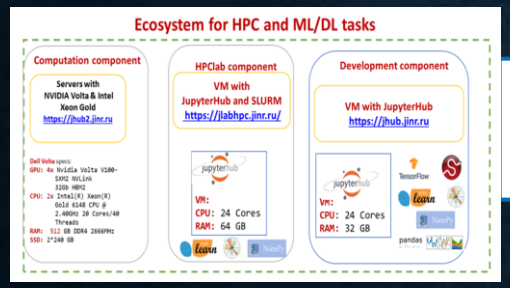
Heterogeneous Distributed Computing Environment

A heterogeneous computing environment, based on the DIRAC platform, was created for processing and storing data of the experiments conducted at JINR. Tier1, Tier2, the “Govorun” supercomputer, the clouds of the JINR Member States, the NICA cluster, as well as the cluster of the National Autonomous University of Mexico (NAUM, within the cooperation on the MPD project), were integrated into DIRAC. For the time being, the distributed infrastructure is used by the following experiments: MPD, Baikal-GVD, BM@N, SPD.



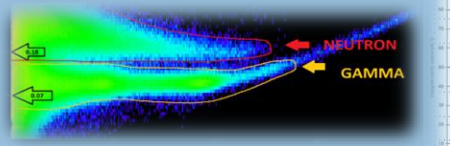
Development of tools and services for users of the “Govorun” supercomputer

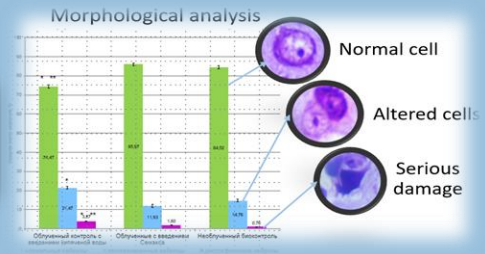
Ecosystem for Supercomputer Modeling

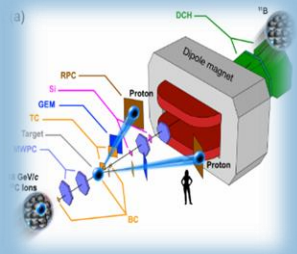


Ecosystem for Machine/Deep Learning

Ecosystem for Applied Computations





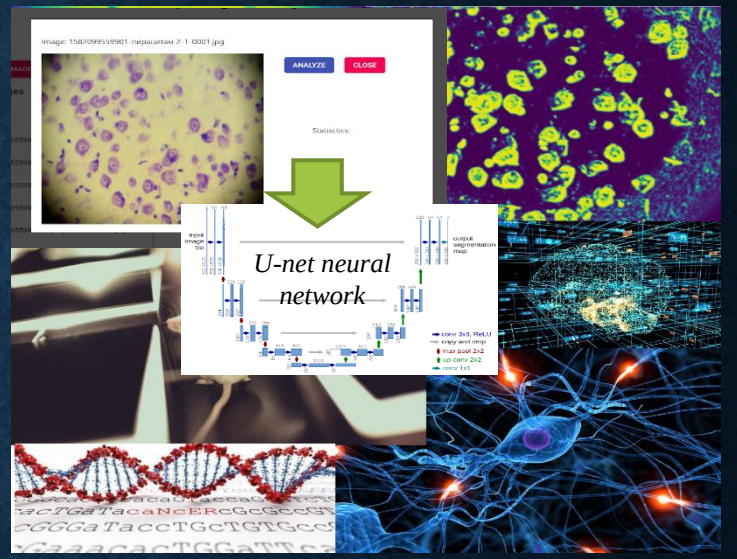


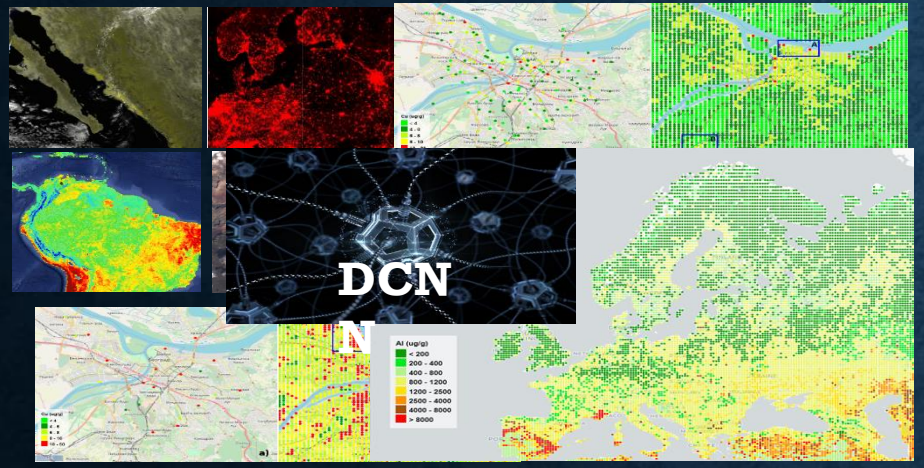
RESEARCH ENVIRONMENT FOR SOLVING RESOURCE-INTENSIVE TASKS OF JINR:

- Parallel computing
- ML/DL/AI tasks
- Quantum computing
- Tools for data analysis and visualization
- Calculations on application packages
- Web services for application programs
- Training courses

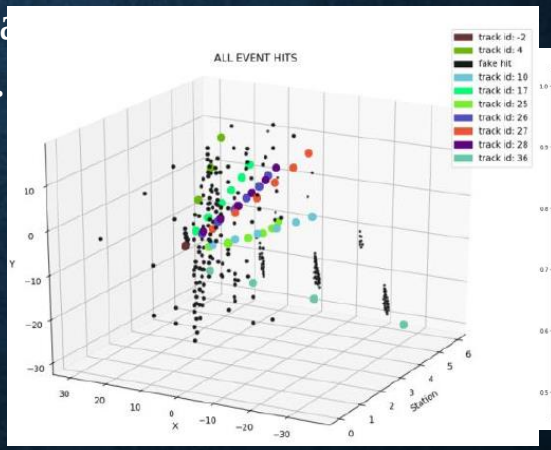
Artificial Intelligence in JINR tasks

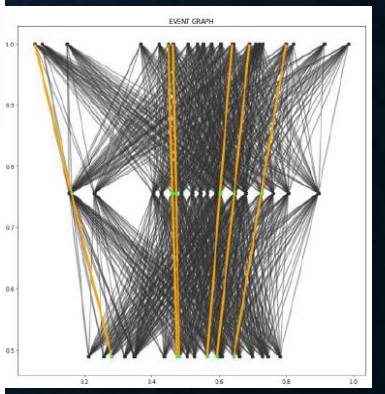
I. The joint project of LIT and LRB is focused on creating an Information System (IS) as a set of IT solutions providing the storage, analysis and visualization of data from experiments at LRB. The IS is based on machine and deep learning methods and neural network approaches.





II. Within the framework of cooperation between LIT and FLNP, the work on the prediction of air pollution by heavy metals using biomonitoring data, satellite imagery and different technologies of ma and deep learning is in progress.



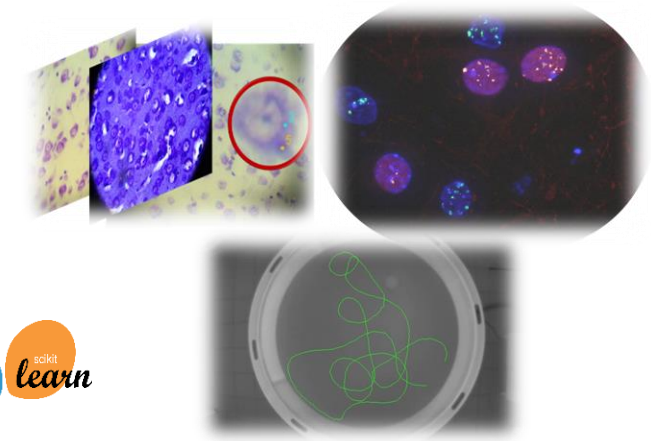


III. The application of high-performance deep neural networks for solving tracking problems in a dense environment of experiments with heavy ions (BM@N, BESIII, SPD et al.).

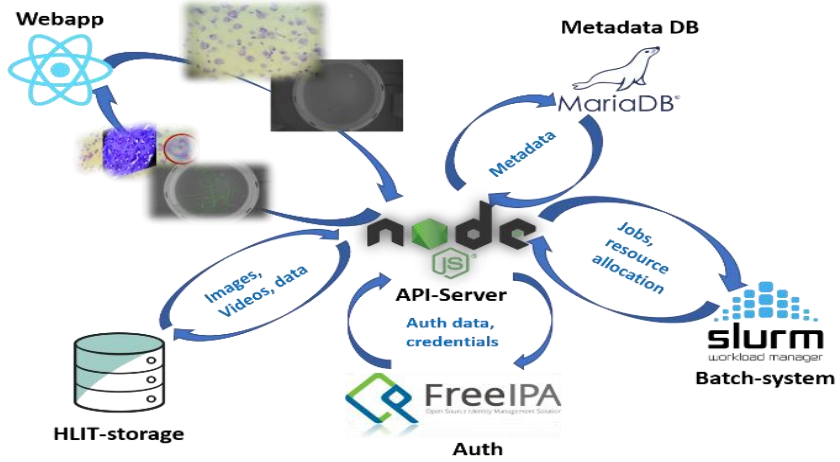
IV. Development of an intelligent automated system for providing liquid nitrogen for a cryogenic test bench for superconducting magnets (in collaboration with LHEP).

Information System for the tasks of RADIATION BIOLOGY
https://bio.jinr.ru

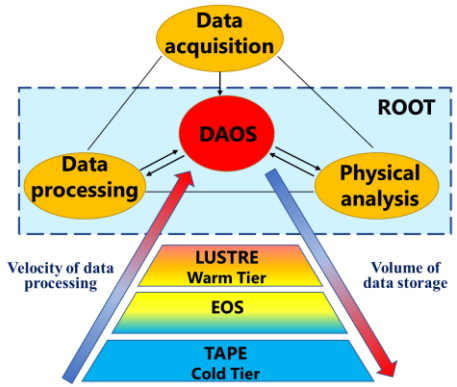
Machine and deep learning methods, computer vision and neural network approaches



Web services for morphofunctional changes in the CNS and DNA damage analysis



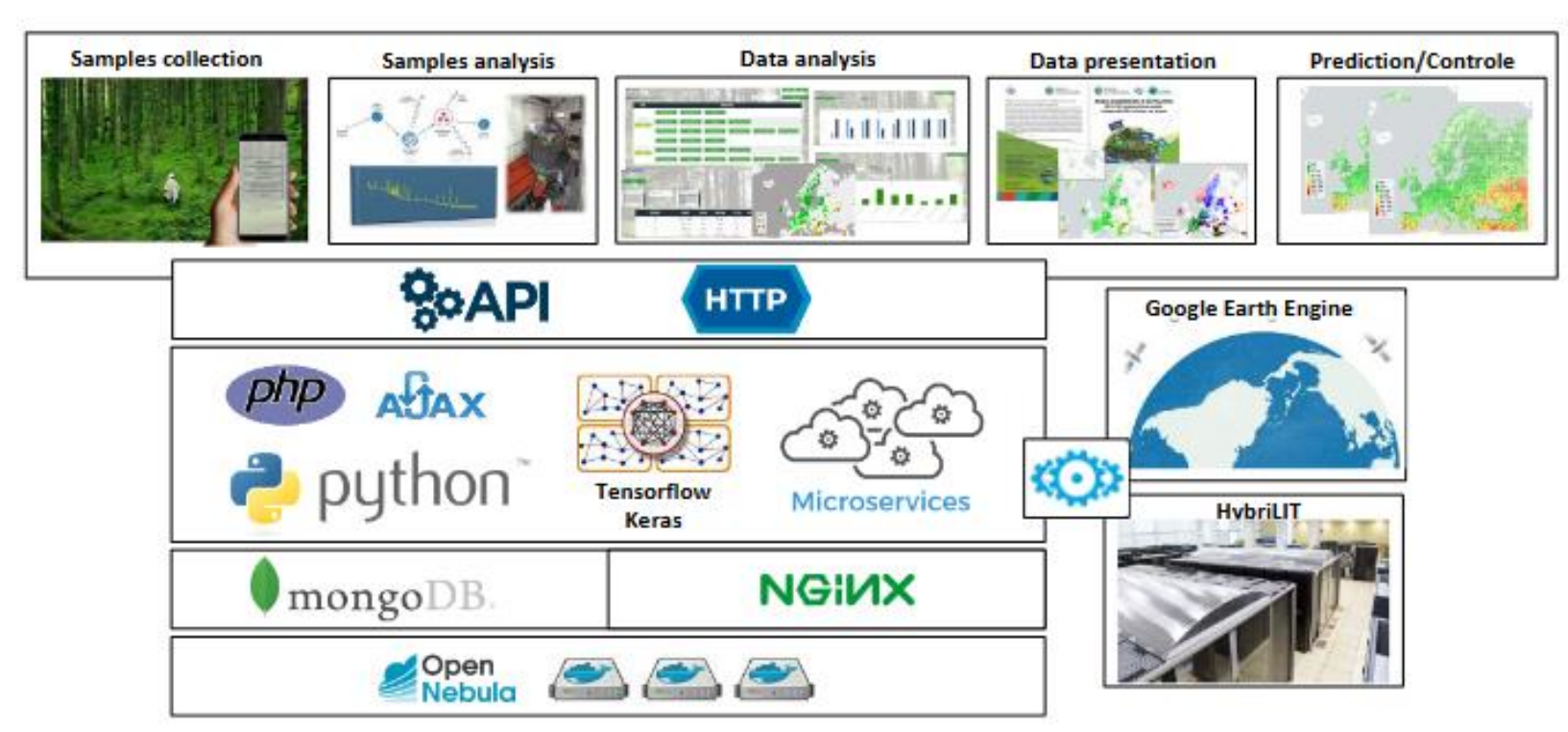
Computing resources and hierarchical storage system based on the "Govorun" SC

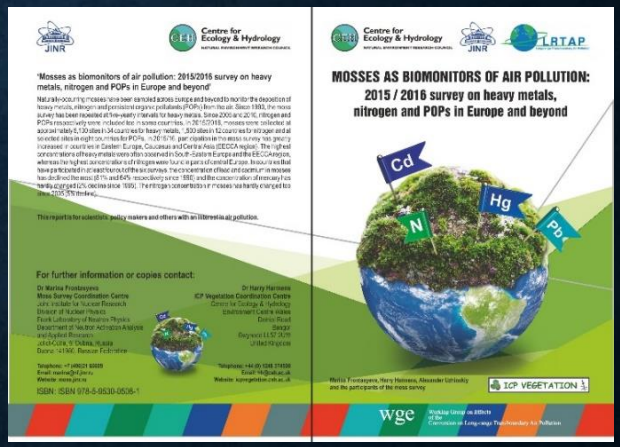


Intelligent Environmental Monitoring Platform

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 IT technologies is in progress.

On the MLIT cloud platform, a Data Management System (DMS) of the UNECE ICP Vegetation was created. The intelligent environmental monitoring platform provides a new level of service for UNECE ICP Vegetation participants. DMS is intended to provide its participants with a modern unified system of collecting, analyzing and processing biological monitoring data**.** A combination of satellite data, biomonitoring measurements, and different machine and deep learning technologies was used to predict potentially toxic elements.





A. Uzhinskiy, M. Aničić Urošević, M. Frontasyeva. Prediction of air pollution by potentially toxic elements over urban area by combining satellite imagery, Moss Biomonitoring Data and Machine Learning. Ciencia e Tecnica Vitivinicola Journal, in press

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

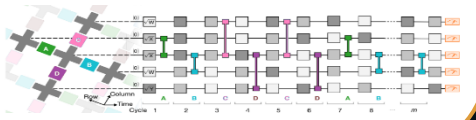
jupyterhub

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

The International Symposium Nuclear Electronics and Computing

- Detector & Nuclear Electronics
- Triggering, Data Acquisition, Control Systems
- Distributed Computing, GRID and Cloud Computing
- Research Data Infrastructures
- Machine Learning Algorithms and Big Data Analytics
- Computations with Hybrid Systems
- Computing for Large Scale Facilities
- 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