INFORMATION TECHNOLOGY

Development of computing infrastructure

To achieve the main goals of the key JINR projects, a huge amount of experimental data will have to be processed. Very roughly, these are tens of thousands of processor cores. In particular, the NICA project requires grid infrastructures of Tier0, Tier1, and Tier2 levels, and the neutrino program needs computational and storage resources. To maintain leading research at JINR, it is necessary to develop distributed multilevel heterogeneous computing environments, including those on the resources of the participants of the experiments.

The Tier0 and Tier1 centers for the NICA project are assumed to be created on JINR resources, including hundreds of petabytes of long-term raw data storage. This will allow for 25–30% of all computing resources in the distributed system, provision and support of the main services for the distributed computing system (DIRAC, PanDA, etc.).

The data storage and computing capacity of the WLCG project, aimed at solving the tasks in the scope of JINR participation in the CERN experiments, should increase annually by 10–20%, thus maintaining the required processing speed.

The elaboration of new deep and machine learning algorithms for data processing and analysis will require the support and development of a high-performance computing infrastructure. The "Govorun" supercomputer is a flexible, scalable, hyper-converged system that combines computational architectures of different types, a hierarchical data processing and storage system. The development of the "Govorun" supercomputer is aimed at the creation of an environment for supercomputer modeling and solution of resource-intensive theoretical and experimental tasks of JINR. Such a research environment is required for parallel computing, ML/DL/AI tasks, quantum computing, data analysis and visualization tools, application packages, web services for applications, training courses and practices.

One of the main priorities of the Seven-Year Plan is to expand the JINR cloud infrastructure and create an integrated cloud environment for experiments performed at JINR and its Member States based on containerization technologies. Progress in this field will depend mainly on the readiness of the experiments to adopt such a workflow.

The development of information technologies is directly related to the further development of the JINR network infrastructure. The support of modern network technologies includes software-defined networking (SDN), content delivery networks (CDN), named data networking (NDN), and the technology for building distributed data procession centers (Data Center Interconnect, DCI).

The Big Data development strategy at the Institute includes a wide range of research: preparation of the infrastructure for Big Data storage and processing (hardware and software, security); elaboration of modern methods and algorithms of Big Data for solving applied tasks; intelligent monitoring of the operation and security of distributed computing systems; providing the Big Data infrastructure to users.

The first-priority task in the field of the development and application of quantum computing, quantum software engineering and quantum intelligent control is the creation of quantum systems of intelligent control of physical experimental facilities, including those for unpredictable situations.

Based on these requirements, the main direction of the development of the MLIT IT ecosystem is connected with the update of network communication channels, the engineering and computing infrastructure of the Multifunctional Information and Computing Complex (MICC) as well as the development of data processing and storage technologies for the NICA experiments and the JINR neutrino program. To ensure a sustainable operation of the MICC, the existing infrastructure needs to be regularly upgraded and maintained.

The IT ecosystem will become a platform for training IT specialists engaged in developing algorithmic and software solutions for JINR tasks.

The work will be carried out in close cooperation with research teams and IT specialists from all JINR Laboratories and Member States.

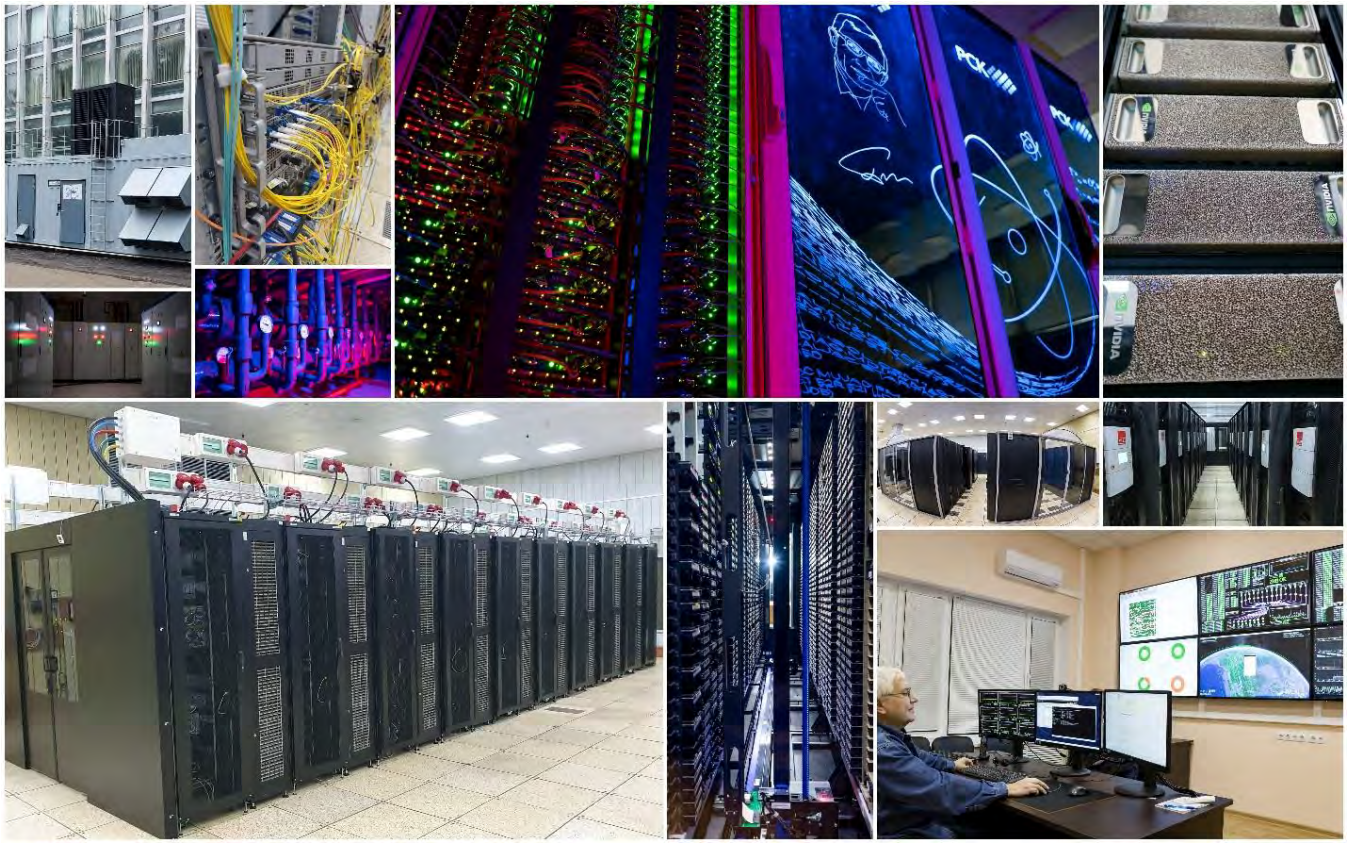

Fig. 32. Elements of the Multifunctional Information and Computing Complex

Expected results:

1. Modernization of the engineering infrastructure of the MICC JINR (reconstruction in accordance with modern requirements of the engine room of the MLIT 4th floor).
2. Modernization and development of the offline distributed computing platform for the NICA project with the involvement of NICA collaboration computing centers.
3. Creation of a Tier0 grid cluster for the NICA megaproject experiments to store experimental and simulated data. Expansion of the performance and storage capacity of the Tier1 grid cluster and Tier2/CICC integrated storage systems as data centers for the NICA megaproject experiments, the JINR neutrino program and experiments at the LHC.

4. Expansion of the JINR cloud infrastructure to increase the range of services available to users applying containerization technologies. Automation of the deployment of cloud technologies in the organizations of the JINR Member States.
5. Expansion of the heterogeneous HybriLIT platform, including the "Govorun" supercomputer as a hyper-converged software defined environment with a hierarchical data storage and processing system.
6. Creation of a multipurpose software and hardware platform of Big Data analytics based on hybrid hardware accelerators (GPU, FPGA, quantum systems); machine learning algorithms; analytics, reporting and visualization tools; support for user interfaces and tasks.
7. Creation and development of a distributed software defined high-performance computing platform combining supercomputer (heterogeneous), grid and cloud technologies for the efficient use of new computing architectures.
8. Development of a computing infrastructure protection system based on fundamentally new paradigms, including quantum cryptography, neurocognitive principles of data organization and interaction of data objects, global integration of information systems, universal access to applications, new Internet protocols, virtualization, social networks, data of mobile devices and geolocation.

Table 7. Approximate estimate of required computing resources

		2024	2025	2026	2027	2028	2029	2030
LHC Tier1 (CMS)	CPU (Pflops)	1,53	1,69	1,84	2,03	2,22	2,45	2,68
	Disk (PB)	18	20	25	28	31	34	40
	Tape (PB)	46	50	60	70	80	90	100
	Network (Gbps)	200	400	400	600	600	800	800
LHC Tier2 (ATLAS, CMS, ALICE, LHCb) + etc.	CPU (Pflops)	0,73	0,81	0,88	0,96	1,04	1,15	1,27
	Disk (PB)	7,7	8,5	9,2	10	11,	12,80	14
	Network (Gbps)	200	400	400	600	600	800	800
"Govorun" supercomputer	CPU (Pflops)	1,2	2,2	3,2	4,2	5,2	6,2	7,2
	Disk (PB)	8	9	10	11	12	13	14
DataLake	Disk (PB)	60	60	60	80	80	80	100
*NICA Tier 0,1,2	CPU (Pflops)	2,2	2,6	8,6	8,6	15,6	15,6	15,6
	Disk (PB)	17	24	47	75	96	119	142
	Tape (PB)	45	88	170	226	352	444	536
	Network (Gbps)	400	400	400	400	400	400	400
*Baikal-GVD, NOvA, JUNO, DUNE Tier 0,1,2	CPU (Pflops)	0,94	1,02	1,2	1,28	1,36	1,54	1,62
	Disk (PB)	1,9	3,2	3,5	3,8	4,6	4,9	5,2
	Tape (PB)	9	12	15	18	21	24	27
	Network (Gbps)	200	200	200	200	200	200	200

* The financing of computing resources for computing under the NICA project and the JINR neutrino program is carried out within the budgets of the corresponding directions.

Mathematical support of studies performed at JINR

An important activity is the provision of mathematical, algorithmic and software support for experimental and theoretical research performed at JINR. The aim is to simulate physical processes, to create algorithms and software systems for processing and analyzing experimental data, to develop algorithms in the field of machine and deep learning, artificial intelligence and cognitive intelligent robotics, systems of quantum intelligent control, the development of methods of computer algebra and quantum computing, as well as Big Data analytics.

Expected results:

1. Development of information and computing systems for data analysis and processing in the field of radiobiology.
2. Development of algorithms based on recurrent and convolutional neural networks for machine and deep learning as well as Big Data analytics problems, created primarily to solve various tasks of particle physics experiments, including the NICA megaproject and neutrino experiments.
3. Creation of modern research tools for international collaborations (NICA, JINR neutrino program, LHC experiments).
4. Development of new numerical and computing models, including quantum computing, for theoretical research at JINR.
5. Elaboration of algorithms for intelligent control of JINR experimental facilities based on the quantum approach.
6. Development of a system on the Big Data analytics platform for the analysis and protection of data of the JINR computer network in real time based on network traffic.
7. Development of machine learning and artificial intelligence algorithms to optimize the functioning and intelligent monitoring of distributed computing systems.
8. Creation of a new-generation analytical system based on effective methods and algorithms of formalization, knowledge extraction and Big Data processing.
9. Development of intelligent information systems for research and applications.
10. Development of quantum IT data processing technologies with-access to NISQ (Noisy Intermediate-Scale Quantum) computers/quantum computers with reliable error protection.
11. Development of scalable algorithms and software for processing multiparameter, multidimensional, hierarchical datasets of exabyte volume.

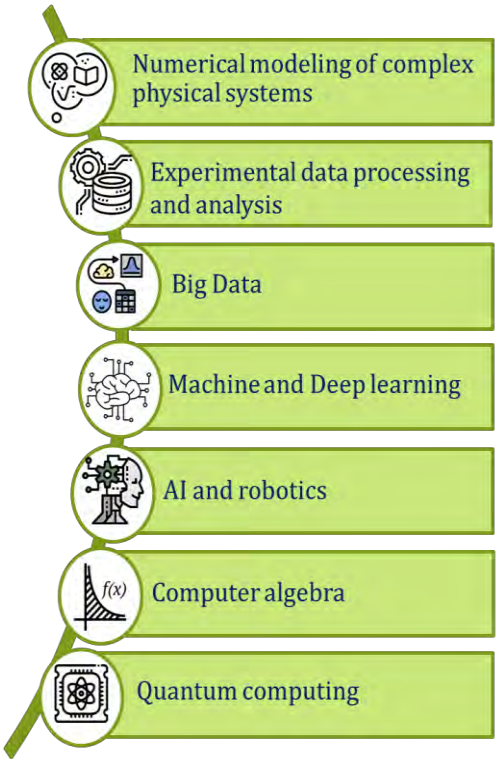

Fig. 33. Directions of development within the framework of mathematical support for research conducted at JINR

JINR Digital Ecosystem

One of the uppermost tasks of the Seven-Year Plan is the creation of the JINR-wide digital platform "JINR Digital Ecosystem". The main goal is the organization of a digital space with single access and data exchange between electronic systems, as well as the automation of actions that previously required a personal or written request. The platform should ensure the integration of existing and future services to support the Institute's scientific, administrative and social activities, as well as the engineering and IT infrastructures.

The user will have the possibility of a single entry point for the JINR digital environment, through which access to a large-scale network of different services will be provided. The "Digital Ecosystem" interface will represent a "showcase" of digital services and resources with the ability to perform a certain set of actions (for example, account management) or switch to a fully functional version of the service. Examples of services are resources for users of basic facilities, library services, document servers, MICC computing resources, 1C administrative services (finance, personnel, electronic document management), etc.

Within the platform being created, registered users (with a JINR account, i.e., Single Sign-On, SSO) will be able to draw up and approve different documents in electronic form, to register and use scientific and administrative services without filling in paper forms and personally visiting the staff members responsible for them. A system of notifications from different services (for example, about documents that await signing) will be available in the personal account. The level of access to services will depend on the position of the staff member and his functional duties. A user-friendly interface allowing one to quickly update information will be organized for service administrators. Part of the resources will also be available to unregistered users: telephone directory, information on dissertation councils, scientific software, JINR map.

The JINR geoinformation system, including an interactive map, information on JINR buildings and other objects (plans of buildings, engineering and other networks, staff accommodation, accounting and analysis of the use of premises taking into account their class, type and purpose), etc. will be developed within the digital platform. The geoinformation system will enable to perform a quick and convenient search for information on both JINR buildings and staff members. It is supposed to use the technology of mobile robots and quantum control elements to solve the tasks of premises' automatic explication (creating plans for buildings) and object localization on the map.

The platform should provide reliable and secure access to different types of data that arise in the course of the Institute's work, from open to confidential. A sample dataset from key services will be placed in storage for further joint analysis using Big Data and artificial intelligence technologies. The automated monitoring of performance indicators for both individual objects and the Institute as a whole will be possible on the basis of data such as information about staff members' publications, financial information and the use of computing resources.

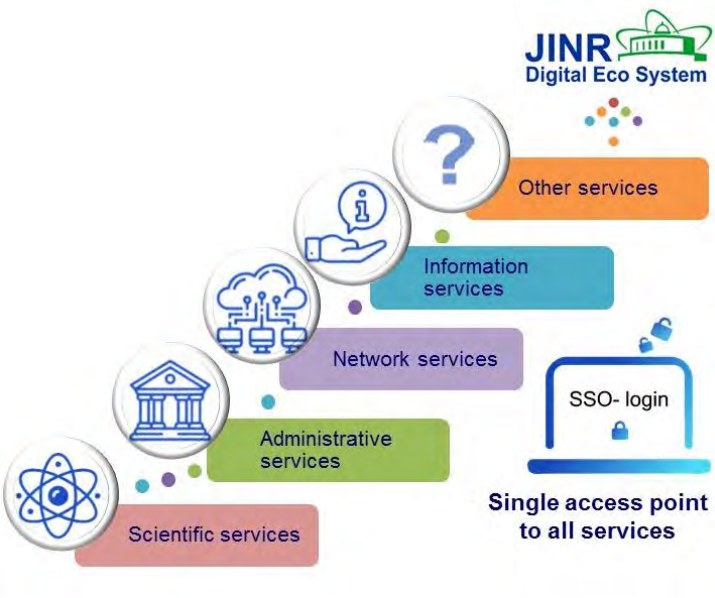

Fig. 34. JINR Digital Ecosystem

Expected results:

1. Creation of the platform "JINR Digital Ecosystem".

Table 8. MICC

(thousand US dollars)

Expenses	2024	2025	2026	2027	2028	2029	2030	Total
Staff costs (Art. 1–3)								
Creation, development, modernization costs (Art. 5, 6, 9, 10, 18, 19)								**40 000,0**
Operating and maintenance costs (Art. 5, 6, 7, 8, 10, 14)								**14 757,4**
International cooperation costs (Art. 4)								
Service costs (Art. 11, 12, 13 ,14, 15, 16, 17)								
TOTAL								